

R-loop-derived cytoplasmic RNA–DNA hybrids activate an immune response

<https://doi.org/10.1038/s41586-022-05545-9>

Received: 30 June 2021

Accepted: 8 November 2022

Published online: 21 December 2022

 Check for updates

Magdalena P. Crossley^{1,8}, Chenlin Song^{1,8}, Michael J. Bocek¹, Jun-Hyuk Choi^{1,6,7}, Joseph N. Kousouros¹, Ataya Sathirachinda¹, Cindy Lin^{2,3,4}, Joshua R. Brickner¹, Gongshi Bai¹, Hannes Lans⁵, Wim Vermeulen⁵, Monther Abu-Remaileh^{2,3,4} & Karlene A. Cimprich¹✉

R-loops are RNA–DNA-hybrid-containing nucleic acids with important cellular roles. Deregulation of R-loop dynamics can lead to DNA damage and genome instability¹, which has been linked to the action of endonucleases such as XPG^{2–4}. However, the mechanisms and cellular consequences of such processing have remained unclear. Here we identify a new population of RNA–DNA hybrids in the cytoplasm that are R-loop-processing products. When nuclear R-loops were perturbed by depleting the RNA–DNA helicase senataxin (*SETX*) or the breast cancer gene *BRCA1* (refs. ^{5–7}), we observed XPG- and XPF-dependent cytoplasmic hybrid formation. We identify their source as a subset of stable, overlapping nuclear hybrids with a specific nucleotide signature. Cytoplasmic hybrids bind to the pattern recognition receptors cGAS and TLR3 (ref. ⁸), activating IRF3 and inducing apoptosis. Excised hybrids and an R-loop-induced innate immune response were also observed in *SETX*-mutated cells from patients with ataxia oculomotor apraxia type 2 (ref. ⁹) and in *BRCA1*-mutated cancer cells¹⁰. These findings establish RNA–DNA hybrids as immunogenic species that aberrantly accumulate in the cytoplasm after R-loop processing, linking R-loop accumulation to cell death through the innate immune response. Aberrant R-loop processing and subsequent innate immune activation may contribute to many diseases, such as neurodegeneration and cancer.

R-loops are three-stranded nucleic acid structures that form during transcription. Unscheduled R-loop formation can interfere with productive DNA replication and transcription, and has been linked to the formation of double-stranded breaks, genome instability, senescence and cell death in several disease states^{1,5,11}. Many factors suppress R-loop formation in human cells, including the helicase *SETX*^{3,7}, which can unwind the RNA–DNA hybrid portion of the R-loop, and the breast-cancer-predisposition gene *BRCA1* (ref. ⁶), which is involved in DNA repair, DNA replication and transcription. Although DNA breaks are known to result from the endonucleolytic processing of R-loops^{2–4}, the fate of these processed nucleic acids and their impact on the cell remains unclear.

Cytoplasmic RNA–DNA hybrid accumulation

To study R-loop processing, we used recombinant, GFP-tagged, catalytically inactive RNase H1 D210N (GFP–dRH) to visualize RNA–DNA hybrids throughout the cell¹². Notably, we observed that short interfering RNA (siRNA)-mediated knockdown of two factors that affect R-loop levels, *SETX* or *BRCA1*, not only led to an increase in nuclear GFP–dRH signal as previously observed¹², but also an increase in cytoplasmic GFP–dRH signal (Fig. 1a and Extended Data Fig. 1a–c). This signal was

sensitive to the pretreatment of the cells with RNase H, which degrades the RNA moiety of RNA–DNA hybrids, indicating that RNA–DNA hybrids accumulate in the cytoplasm of cells. To characterize these nucleic acids, we developed a method to biochemically purify and visualize cytoplasmic RNA–DNA hybrids—termed cytoplasmic DNA–RNA hybrid immunoprecipitation (cytoDRIP) (Fig. 1b). Using this approach, we found that depletion of *SETX* or *BRCA1* (Fig. 1c,d and Extended Data Fig. 1d–g), as well as splicing inhibition using pladienolide B¹³ (PlaB) (Extended Data Fig. 1h), resulted in an increased accumulation of cytoplasmic RNA–DNA hybrid fragments, ranging in size from 100 bp to several kilobases.

XPG and XPF excise RNA–DNA hybrids

As the endonucleases XPG and XPF have been implicated in R-loop processing^{2–4}, we investigated whether they contribute to cytoplasmic hybrid formation. Notably, cytoplasmic RNA–DNA hybrid accumulation was abrogated by siRNA-mediated depletion of XPG or XPF (Fig. 1e,f and Extended Data Fig. 1i–l), or by degradation of auxin-induced degron (AID)-tagged XPG (Extended Data Fig. 1m–o). These results indicate that R-loop deregulation leads to an XPG- and XPF-dependent increase in the formation of cytoplasmic RNA–DNA

¹Department of Chemical and Systems Biology, Stanford University, Stanford, CA, USA. ²Department of Chemical Engineering, Stanford University, Stanford, CA, USA. ³Department of Genetics, Stanford University, Stanford, CA, USA. ⁴The Institute for Chemistry, Engineering & Medicine for Human Health (ChEM-H), Stanford University, Stanford, CA, USA. ⁵Department of Molecular Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. ⁶Present address: Biometrology Group, Division of Chemical and Biological Metrology, Korea Research Institute of Standards and Science, Daejeon, South Korea. ⁷Present address: Department of Bio-Analytical Science, University of Science & Technology, Daejeon, South Korea. ⁸These authors contributed equally: Magdalena P. Crossley, Chenlin Song. ✉e-mail: cimprich@stanford.edu

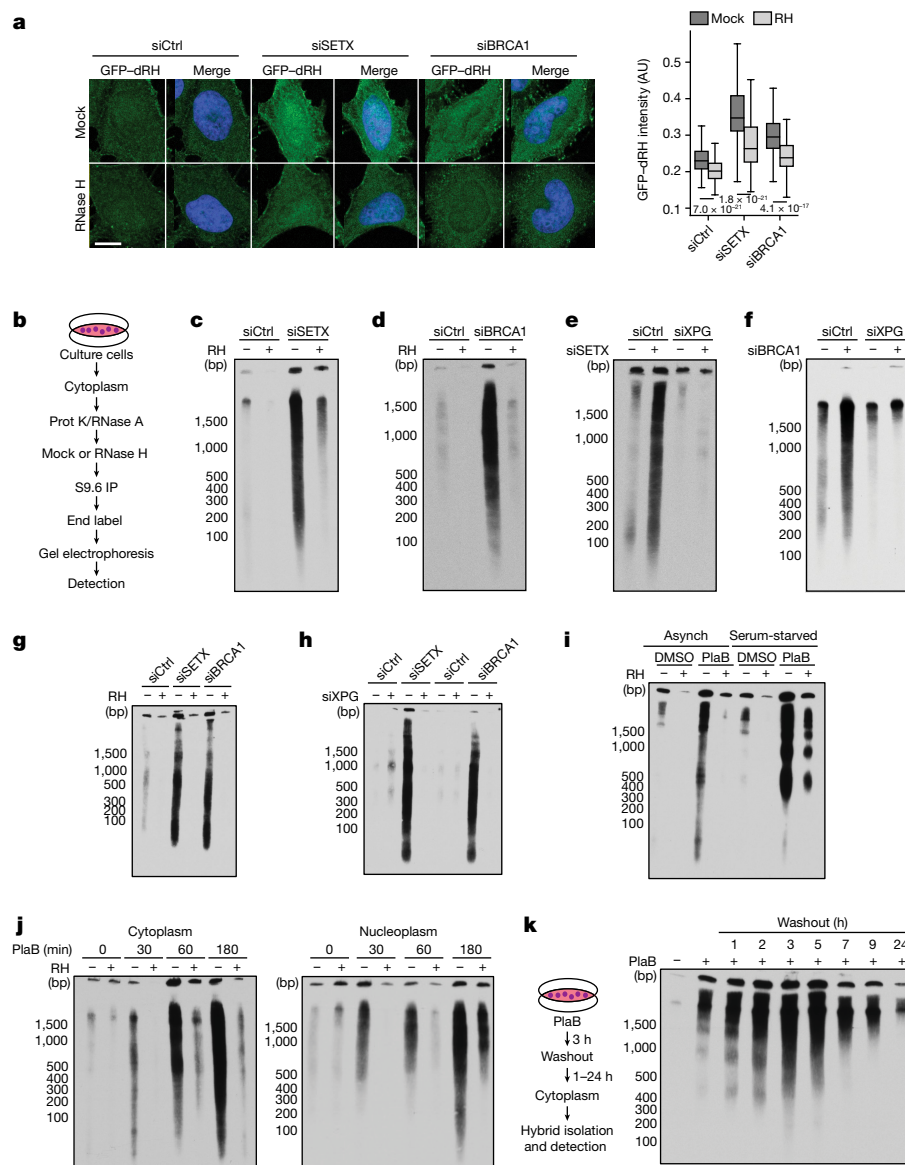


Fig. 1 | The loss of R-loop resolution factors leads to XPG-dependent cytoplasmic RNA-DNA hybrid accumulation. a, Left, images of *SETX*- or *BRCA1*-depleted HeLa cells probed with GFP-dRH protein after fixation, after mock or RNase H (RH) pretreatment. Scale bar, 10 μ m. Right, quantification of cytoplasmic GFP-dRH intensities. *P* values calculated using two-sided Mann-Whitney U-tests are shown in the figure. From left to right, $n = 463, 249, 397, 237, 286$ and 224 . The centre line shows the median, the box limits show the 75th and 25th percentiles, and the whiskers show the minimum and maximum values. AU, arbitrary units. **b**, Schematic of the cytoDIP experimental workflow. Prot K, proteinase K. **c**, Cytoplasmic RNA-DNA hybrids extracted from control (siCtrl) and siSETX-treated HeLa cells, with mock and RNase H treatment before pull-down. Size markers are indicated in base pairs (bp). **d**, As described in

c, but for control and siBRCA1-treated cells. **e**, As described in **c**, but with siXPG. **f**, As described in **e**, but after *BRCA1* knockdown. **g**, Cytoplasmic hybrid levels in *BAX*^{-/-} *BAK*^{-/-} HeLa cells after treatment with siCtrl, siSETX or siBRCA1, with mock and RNase H treatment before pull-down. **h**, As described in **g**, with or without XPG knockdown. **i**, Cytoplasmic hybrids extracted from asynchronous (asynch) or serum-starved MCF10A cells after DMSO or PlaB treatment (500 nM, 3 h), with mock and RNase H treatment before pull-down. **j**, Cytoplasmic (left) or soluble nuclear (right) hybrids from HeLa cells after a time course of PlaB, with mock and RNase H treatment before pull-down. **k**, Cytoplasmic hybrids extracted from HeLa cells after PlaB treatment (500 nM, 3 h) and then washout for up to 24 h.

hybrids. To ensure that cytoplasmic hybrids were not simply generated as a result of apoptosis, we generated apoptosis-deficient *BAX*^{-/-} *BAK*^{-/-} double-knockout cells^{14,15}. We observed R-loop-induced, XPG-dependent accumulation of cytoplasmic hybrids in these cells (Fig. 1g,h and Extended Data Fig. 2a). We also observed cytoplasmic hybrids in serum-starved cells (Fig. 1i and Extended Data Fig. 2b-k), indicating that cytoplasmic hybrid production does not require DNA replication and that these hybrids are not simply released into the cytoplasm through breakdown of the nuclear envelope. The nuclear transport receptor exportin-1, which is involved in the export of

nucleic acids¹⁶⁻¹⁸, also had a partial role in regulating the localization of the hybrids induced by SETX loss or PlaB treatment (Extended Data Fig. 2l-n). This finding suggests that cytoplasmic hybrid formation is an active process involving nuclear export. Finally, we took advantage of the rapid action and reversibility of PlaB (Extended Data Fig. 2o,p) to study the dynamics and stability of cytoplasmic RNA-DNA hybrids. We found that cytoplasmic and nucleoplasmic hybrids formed within 30 min of PlaB addition, and their levels accumulated over time (Fig. 1j). Three hours after PlaB withdrawal, hybrid levels began to decrease, exhibiting a half-life of approximately 4 h and returning to the baseline

by 24 h (Fig. 1k). Thus, R-loop processing results in the rapid formation and active export of RNA–DNA hybrids to the cytoplasm, from which they are eventually cleared.

Hybrids originate from genomic R-loops

To trace the origin of cytoplasmic RNA–DNA hybrids, we combined cytoDRIP with strand-specific RNA–DNA hybrid sequencing (cytoDRIP–seq) (Fig. 2a,b) using control and *SETX*-depleted cells. We sequenced the single-stranded DNA moiety of the hybrids to prevent known issues of the S9.6 antibody binding to dsRNA^{12,19}, obtaining a highly reproducible signal (Extended Data Fig. 3a). The sequencing reads were primarily derived from the nucleus, with a small fraction mapping to the mitochondrial genome (Extended Data Fig. 3b). We identified 866 peaks in control cells and 5,726 peaks in *SETX*-depleted cells, representing 0.04% and 0.19% of genome space, respectively (Extended Data Fig. 3c). Most of these sites were not present in the control cells (Extended Data Fig. 3d) and were enriched above the IgG control (Extended Data Fig. 3e). Importantly, we also demonstrated sensitivity to RNase H (Extended Data Fig. 3f) and validated the role of XPG in cytoplasmic hybrid formation (Fig. 2c). The cytoDRIP sites mapped to both genic and intergenic regions (Fig. 2d and Extended Data Fig. 3g,h). Within genes, most cytoDRIP sites occurred within gene bodies (Extended Data Fig. 4a), and within intergenic regions, there was notable enrichment at enhancers (Fig. 2d). As nuclear R-loops are enriched for certain repetitive DNA sequences²⁰, we examined whether repeats overlapped cytoplasmic hybrid sites more than expected. We found that cytoDRIP regions were elevated for several types of repeat, in particular, simple and low-complexity repeats (Extended Data Fig. 4b,c), centromeres and rDNA (Extended Data Fig. 4d).

By comparing cytoDRIP–seq and nuclear DRIP–seq signals²¹, we found that most cytoDRIP regions overlapped sites that form nuclear RNA–DNA hybrids (Fig. 2a and Extended Data Fig. 4e), as expected. However, cytoDRIP regions collectively occupied a much smaller area of the genome (Extended Data Fig. 3c), and peak lengths were smaller compared with those for nuclear R-loops (Extended Data Fig. 4f), indicating that cytoplasmic hybrids are derived from a small subset of nuclear R-loops, and that only a portion of nuclear R-loops may be susceptible to processing. The cytoDRIP–seq peak strength was not correlated with nuclear DRIP–seq levels (Extended Data Fig. 4g), or nascent transcription levels as measured using global run-on sequencing (Extended Data Fig. 4h), and well-studied sites of abundant nuclear R-loop formation did not generate cytoplasmic hybrids (Extended Data Fig. 4i,j). Thus, highly transcribed, abundant R-loops are not necessarily susceptible to processing and cytoplasmic accumulation. We also investigated whether XPG was preferentially recruited to genomic R-loops corresponding to cytoDRIP sites, as compared to nuclear R-loop sites that are not represented in the cytoplasm. Using a knockin cell line expressing GFP-tagged XPG (Extended Data Fig. 4k), we performed chromatin immunoprecipitation with quantitative PCR (ChIP–qPCR) against GFP. XPG binding increased after *SETX* loss specifically at hybrid sites found in the cytoplasm, but not at other nuclear R-loop sites (Extended Data Fig. 4l). Overall, these data suggest that certain genomic R-loops become more susceptible to XPG-dependent processing in the absence of *SETX*.

Hybrid sequences and stability

We next examined whether the stability of genomic hybrids affects the likelihood of cytoplasmic hybrid accumulation. Previous modelling revealed a range of nuclear hybrid lifetimes on the genome, with an average half-life of 11 min (refs. ^{21,22}). Using actinomycin D to inhibit transcription and new R-loop formation, we examined the lifetimes of nuclear R-loops from which the cytoDRIP signal was derived, using for comparison previously identified²¹ short-, medium- and long-lived

nuclear hybrids (Fig. 2e and Extended Data Fig. 5a,b). Notably, we estimated R-loop half-lives of 43–67 min from the cytoDRIP–seq sites tested, indicating that these R-loops are particularly long-lived on the genome (Fig. 2e). We also observed a strong association between cytoDRIP sites and a subset of nuclear RNA–DNA hybrids that were previously identified as partially resistant to RNase H treatment^{21,23} (Fig. 2a and Extended Data Figs. 4e and 5c). An *in vitro* RNase H titration combined with high-resolution nuclear DRIP–qPCR confirmed that cytoplasmic hybrids map to nuclear R-loop regions that are less sensitive to RNase H and require longer treatment for degradation (Fig. 2f). However, the long genomic half-life and RNase H resistance were not sufficient determinants of R-loop processing, as multiple long-lived or RNase-H-resistant R-loops were not identified in the cytoplasm by cytoDRIP–seq (Extended Data Figs. 3c and 5d).

Interestingly, when averaging across all cytoDRIP peaks, we observed that the sense and antisense cytoplasmic and nuclear hybrid signals formed two distinct peaks, with the antisense signal shifted approximately 100 nucleotides downstream (Fig. 2a,g). This suggests that cytoplasmic hybrids are derived from genomic regions that have adjacent and potentially overlapping nuclear RNA–DNA hybrids on both strands in a convergent (that is, head-on) orientation (Extended Data Fig. 6a). To test this, we calculated the ratio between the sense and antisense nuclear hybrid signal within each cytoDRIP peak and, for comparison, within each nuclear DRIP peak. As expected, the nuclear hybrid ratios within nuclear DRIP peaks were much higher than one, reflecting that the hybrid signal within nuclear R-loops is predominantly derived from the sense strand (Fig. 2h). However, the ratios within cytoDRIP peaks were smaller and close to one, indicating that these are sites in which both sense and antisense hybrids form. This is consistent with hybrid formation associated with sites of convergent transcription. We next analysed nucleotide features of the cytoDRIP peaks. Although cytoDRIP regions had overall similar GC and AT content relative to nuclear R-loop regions (Extended Data Fig. 6b), they exhibited abrupt switches in the polarity of GC and AT skew (asymmetry in G content and A content between DNA strands, respectively), shifting from high to low skew at the centre of the peak (Fig. 2i). Similar patterns in nucleotide skew were not observed for nuclear DRIP–seq^{24–26} (Extended Data Fig. 6c). As R-loops are known to form preferentially at and be stabilized by GC-skewed regions²⁴, the overlapping, convergent hybrid signal and nucleotide skew observed at cytoDRIP peaks may promote the formation of particularly stable nuclear R-loops.

Notably, the small number of cytoDRIP sites identified in control cells exhibited similar characteristics to those induced by *SETX* loss (Extended Data Fig. 6d–f), suggesting that there is a low basal level of hybrid processing that occurs on the genome that is substantially increased when nuclear R-loop dynamics are perturbed. Taken together, our results indicate that the cytoplasmic hybrids observed after *SETX* loss are derived from a small subset of nuclear R-loop regions that are partially RNase H resistant, relatively long lived and may result from convergent transcription at sites of nucleotide skew.

Hybrid-activated innate immune responses

Cellular nucleic acids can stimulate immune responses through pattern recognition receptors (PRRs)^{8,10,15,27–29}, and deregulation of R-loops has been linked to this signalling^{17,30–33}. Whether cytoplasmic hybrids resulting from nuclear R-loop processing contribute to this response is unknown. We found that depletion of *SETX* or *BRCA1*, or *PlaB* treatment, triggered an increase in phosphorylation of IRF3 at Ser386 (pIRF3) (Fig. 3a and Extended Data Fig. 7a–d)—a marker of IRF3 immune signalling. This phosphorylation was reduced by the expression of nuclear-localized wild-type RNase H1 (NLS–RH), but not catalytically inactive NLS–dRH (Fig. 3a and Extended Data Fig. 7e). Several IRF3 effectors, including interferon beta (*IFNB1*) and several interferon-stimulated

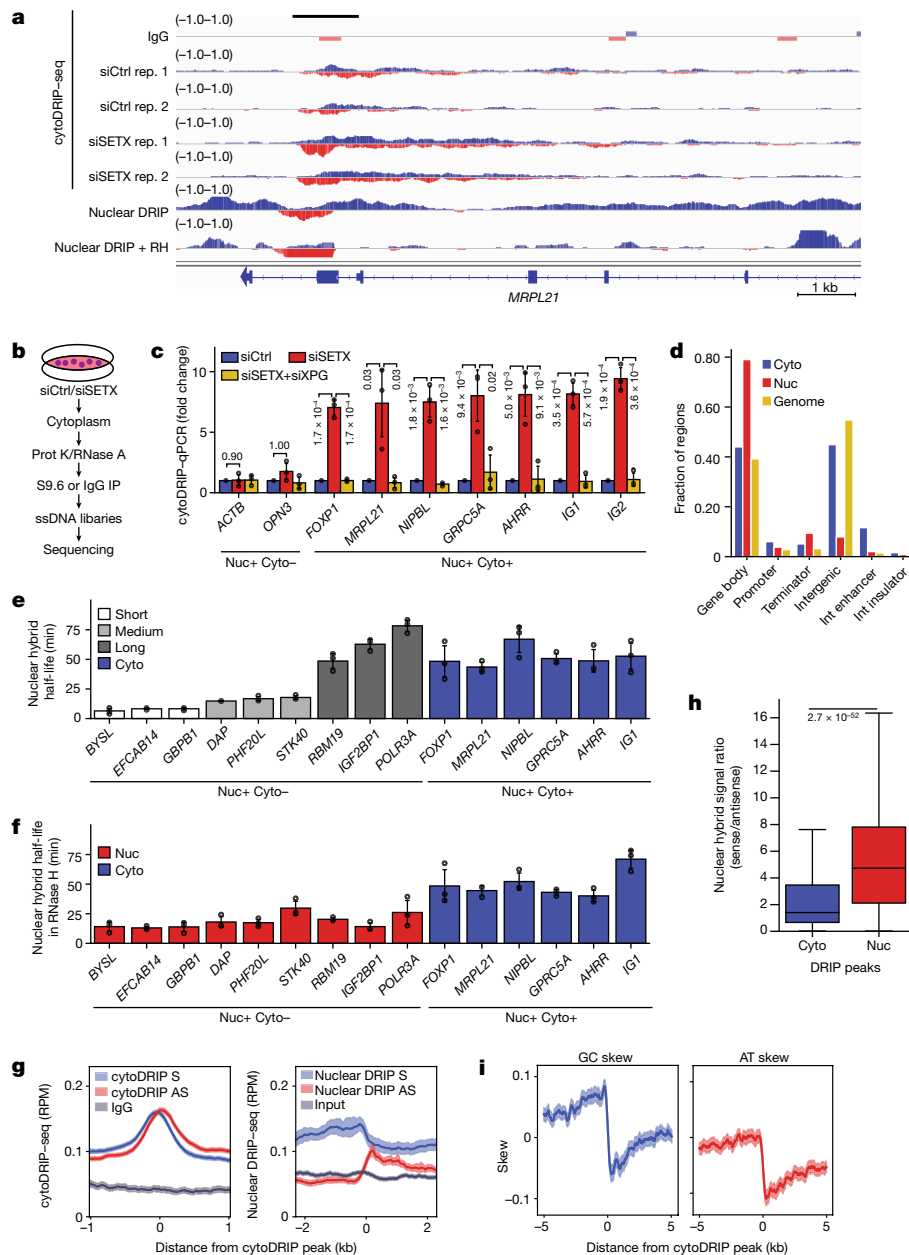


Fig. 2 | cytoDRIP-seq shows that cytoplasmic RNA-DNA hybrids are derived from a subset of nuclear R-loops. a, Genomic tracks showing stranded signal (red, positive; blue, negative) for cytoDRIP-seq (IgG, siCtrl, siSETX), nuclear DRIP-seq and nuclear DRIP-seq + RNase H²¹. The black bar shows the peak. Rep, replicate. **b**, The cytoDRIP-seq workflow. **c**, cytoDRIP-qPCR after depletion of SETX and/or XPG. P-values calculated using unpaired two-tailed *t*-tests are shown in the figure. **d**, The genomic distributions of peaks from cytoDRIP-seq (Cyto), nuclear DRIP-seq (Nuc) and genome values (Genome). Int, intergenic. Statistical analysis was performed using a Kruskal-Wallis test; $P = 1.0 \times 10^{-16}$ between Cyto, Nuc and Genome. **e**, Nuclear hybrid half-lives after transcription inhibition. Short, medium and long R-loop lifetimes were derived previously²¹. Statistical analysis was performed using a two-sided Mann-Whitney *U*-test; $P = 1.6 \times 10^{-7}$ between nuclear DRIP sites with short or medium lifetimes and cytoDRIP sites. **f**, Nuclear hybrid half-lives after in vitro RNase H treatment. Statistical analysis was performed using a two-sided

Mann-Whitney *U*-test; $P = 2.3 \times 10^{-7}$ between nuclear DRIP and cytoDRIP sites. **g**, Plots around siSETX cytoDRIP regions showing cytoDRIP-seq (left) and nuclear DRIP-seq (right) defined by transcription direction: sense-strand signal (S, blue), antisense-strand signal (AS, red), IgG or input (grey). Data are the mean of 1,762 genic peaks ($n = 1,762$). **h**, The ratios of sense/antisense nuclear hybrid signal in siSETX cytoDRIP (Cyto, $n = 984$) and nuclear R-loop (Nuc, $n = 30,446$) peaks. P-values calculated using two-sided Mann-Whitney *U*-tests are shown. The centre line shows the median, the box limits show the 75th and 25th percentiles, and the whiskers show the minimum and maximum values. **i**, Plots around siSETX cytoDRIP regions from **g** showing GC skew (blue) and AT skew (red). For **c**, **e** and **f**, Nuc+ Cyto+ sites (blue) were found in the nucleus and cytoplasm, Nuc+ Cyto- sites (red) were found only in the nucleus. Gene names are indicated; IG1/IG2 are intergenic sites. Data are mean $\pm$ s.d. from three biological replicates ($n = 3$) (**c**, **e** and **f**). For **g** and **i**, error bands show 95% confidence interval of the mean.

genes (ISGs), were also upregulated in an RNase-H-reversible manner (Fig. 3b and Extended Data Fig. 7f). Importantly, depletion of XPG or XPF by siRNA (Fig. 3c,d) or using AID-tagged XPG (Fig. 3e and Extended Data Fig. 7g) reversed this signalling. These observations couple IRF3 signalling to nuclear R-loop processing.

IFN β and some ISGs upregulated by IRF3 signalling are known to induce apoptosis³⁴. We found that cleaved PARP (C-PARP)³⁵ and caspase 3 activity³⁶ were induced after the loss of SETX or BRCA1 in a manner that was blocked either by knockdown of the R-loop processing factors XPG and XPF, or expression of NLS-RH (Fig. 3f-h and Extended Data Fig. 7h,i).

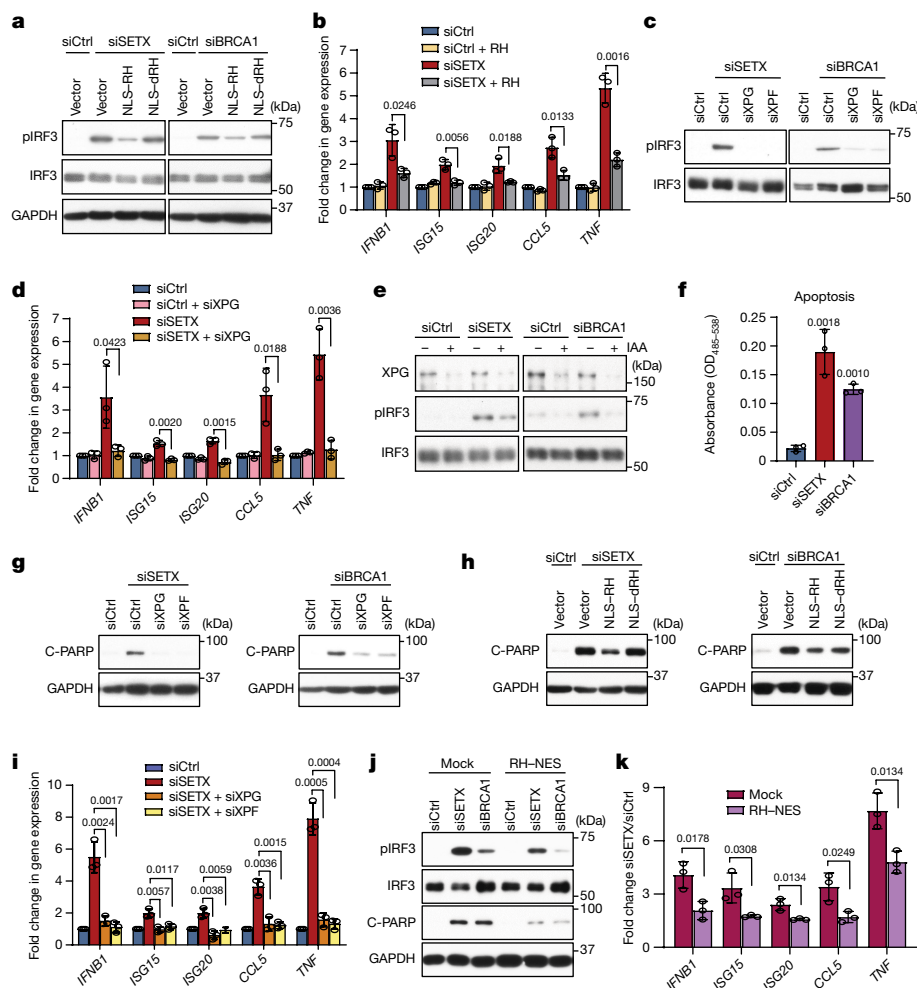


Fig. 3 | Cytoplasmic RNA–DNA hybrids derived from R-loop processing activate IRF3 signalling and induce apoptosis. **a**, Western blot analysis of pIRF3 after transfection of empty vector, human NLS-tagged RNase H1 or NLS-tagged RNase H1 catalytically dead mutant D210N (dRH) in SETX- and BRCA1-deficient HeLa cells. GAPDH is the loading control. **b**, RT–qPCR measurements of IRF3 effectors after knockdown of *SETX* and overexpression of RNase H in HeLa cells. **c**, As described in **a** with *XPG* or *XPF* knockdown in HeLa cells. **d**, As described in **b** with *XPG* knockdown in HeLa cells. **e**, pIRF3 levels after knockdown of *SETX* or *BRCA1* with IAA treatment to induce *XPG* degradation in HeLa AID-tagged *XPG* degenon cells. **f**, Caspase-3 activity after depletion of either *SETX* or *BRCA1* in HeLa cells. OD_{485–538}, optical density at

485–538 nm. **g**, Western blot analysis of C-PARP after depletion of *XPG* or *XPF* in siSETX- or siBRCA1-treated HeLa cells. GAPDH was used as the loading control. **h**, As described in **g** after expression of empty vector, RNase H or dRH in HeLa cells. **i**, RT–qPCR analysis of IRF3 effectors after knockdown of *SETX* and *XPG* or *XPF* in *BAX*^{−/−} *BAK*^{−/−} HeLa cells. **j**, Western blot analysis of pIRF3 and C-PARP after knockdown of *SETX* or *BRCA1* in HeLa cells stably expressing GFP (Mock) or RH–NES. **k**, RT–qPCR measurements of IRF3 effectors in mock and RH–NES HeLa stable cell lines after *SETX* knockdown. For **b**, **d**, **i** and **k**, data are mean ± s.d. *n* = 3 independent biological replicates. Statistical analysis was performed using two-tailed *t*-tests with confidence intervals of 95%; *P* values are shown at the top of the graphs.

Moreover, the proinflammatory and apoptosis factor TNF³⁷ was elevated after R-loop processing (Fig. 3d), and its knockdown reduced C-PARP levels (Extended Data Fig. 7j). These findings indicate that apoptosis, mediated in part by TNF, is a consequence of R-loop processing. Importantly, R-loop-induced, *XPG*- and *XPF*-dependent IRF3 signalling was also observed in *BAX*^{−/−} *BAK*^{−/−} cells (Fig. 3i and Extended Data Fig. 7k). These results demonstrate that the innate immune response can be triggered by R-loop processing independent of apoptosis.

Finally, to establish whether cytoplasmic RNA–DNA hybrids can directly induce the immune response observed in SETX- and BRCA1-deficient cells, we stably expressed cytoplasmically localized RNase H (RH–NES) in these cells. We observed efficient digestion of cytoplasmic hybrids (Extended Data Fig. 7l,m) as well as diminished innate immune signalling and apoptosis when RH–NES was expressed (Fig. 3j,k). Thus, although other nucleic acids could contribute to activation of the innate immune response, our findings strongly suggest that cytoplasmic RNA–DNA hybrids directly contribute to its activation and to apoptosis in these cells.

cGAS and TLR3 sense cytoplasmic hybrids

To test which innate immune sensor mediates the activation of IRF3 when R-loops are induced, we knocked down *TLR3*, *RIGI* or *MDAS*, or inhibited cGAS using RU.521 (ref. 38) in SETX-deficient cells. Inhibition of cGAS, depletion of *TLR3* or knockout of either cGAS and *TLR3* strongly reduced pIRF3 and downstream effectors, whereas depletion of *RIGI* or *MDAS* had only a modest effect (Fig. 4a and Extended Data Fig. 8a–f). Consistently, combined cGAS inhibition and *TLR3* knockdown fully suppressed the activation of IRF3 downstream effectors (Fig. 4a,b and Extended Data Fig. 8g,h) and apoptosis (Fig. 4c and Extended Data Fig. 8i) in SETX/BRCA1-deficient cells. We also excluded the possibility that cGAS or *TLR3* depletion regulates pIRF3 levels indirectly by reciprocally affecting protein levels (Extended Data Fig. 8j). These observations indicate that R-loop-induced IRF3 signalling is mediated primarily by cGAS and TLR3.

We next sought to elucidate whether cGAS and TLR3 recognize cytoplasmic RNA–DNA hybrids in SETX-deficient cells. Consistent

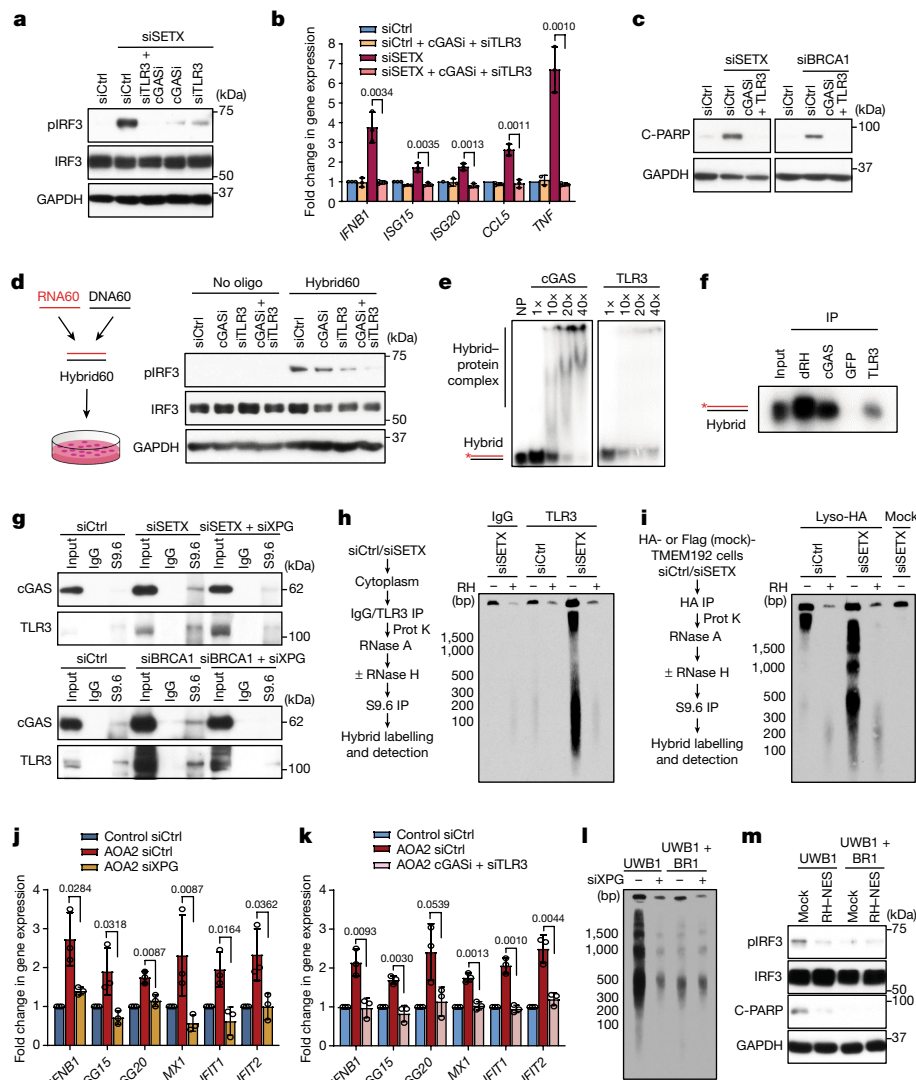


Fig. 4 | R-loop-derived cytoplasmic RNA–DNA hybrids trigger IRF3 signalling through the cGAS and TLR3 receptors. **a**, Dependence of *SETX*-knockdown-induced pIRF3 level on cGAS or TLR3 in HeLa cells. **b**, RT–qPCR measurements of IRF3 effectors after knockdown of *SETX* and *TLR3* and cGAS inhibitor (cGASi) treatment in HeLa cells. **c**, Western blot analysis of C-PARP after perturbation of immune receptors in *SETX*- or *BRCA1*-depleted HeLa cells. **d**, Western blot analysis of pIRF3 following TLR3 or cGAS loss after synthetic RNA–DNA hybrid transfection. **e**, Gel shift assay showing in vitro binding of cGAS and TLR3 to RNA–DNA hybrids. NP, no protein. **f**, In vitro pull-down assay showing the hybrid-binding activity of purified cGAS and TLR3. dRH and GFP are positive and negative controls, respectively. **g**, Co-IP analysis of cGAS or TLR3 with cytoplasmic hybrids after knockdown of *SETX* or *BRCA1*, with or without *XPG* knockdown. **h**, cytoDRIP blot of TLR3-associated hybrids in the cytoplasm of control and *SETX*-depleted HeLa cells, with mock and RNase H

treatment before hybrid pull-down. dsDNA markers are indicated in bp. **i**, The lysosome immunoprecipitation (lysoIP) blot shows RNA–DNA hybrid levels in the endolysosome of control and *SETX*-depleted HA–TMEM192 HEK293T cells, with or without RNase H treatment before hybrid pull-down. Flag–TMEM192 HEK293T cells were used for the mock control. **j**, RT–qPCR analysis of *IFNβ1* and ISGs in control and AOA2 fibroblasts after knockdown of *XPG*. **k**, As described in **j**, after cGAS inhibition and *TLR3* knockdown. **l**, cytoDRIP blot showing cytoplasmic hybrid production after *XPG* knockdown in UWB1.289 and UWB1.289 + *BRCA1* cells. **m**, Western blot analysis of pIRF3 and C-PARP in UWB1.289 and UWB1.289 + *BRCA1* cells stably expressing GFP (mock) and RH–NES. For **b**, **j** and **k**, data are mean ± s.d. $n = 3$ independent biological replicates. Statistical analysis was performed using two-tailed *t*-tests with confidence intervals of 95%; *P* values are shown at the top of the graphs.

with previous reports, cGAS-dependent IRF3 signalling could be activated by synthetic RNA–DNA hybrids^{32,39}. Notably, RNA–DNA hybrids also induced TLR3-dependent IRF3 signalling (Fig. 4d and Extended Data Fig. 8k). Both TLR3 and cGAS also bound to RNA–DNA hybrids directly in vitro, as well as their canonical ligands, double-stranded RNA (dsRNA) and dsDNA⁸, respectively (Fig. 4e,f and Extended Data Fig. 8l,m). Furthermore, the knockdown of *SETX* or *BRCA1* led to the association of cGAS and TLR3 with cytoplasmic hybrids in a manner that was abrogated by the knockdown of *XPG*, removal of hybrids by in vitro RNase H treatment or competition with a synthetic hybrid (Fig. 4g

and Extended Data Fig. 9a–d). These results suggest that cGAS and TLR3 directly recognize endogenous, R-loop-derived cytoplasmic RNA–DNA hybrids.

To confirm the interaction between TLR3 and cytoplasmic hybrids, for which previous reports are lacking, we immunoprecipitated TLR3 from the cytoplasm and probed for hybrids. We observed an RNase-H-sensitive increase in TLR3-co-associated cytoplasmic hybrids in *SETX*-deficient cells (Fig. 4h and Extended Data Fig. 9e), consistent with the S9.6 co-immunoprecipitation (co-IP) results (Fig. 4g). We also examined whether we could observe RNA–DNA hybrids in acidified

endolysosomal compartments in which TLR3 is enriched⁴⁰ by isolating lysosomes using the HA-tagged lysosomal transmembrane protein TMEM192 (ref. ⁴¹) and then performing an S9.6 immunoprecipitation (Extended Data Fig. 9f). We observed an increase in RNA–DNA hybrids in endolysosomes after *SETX* depletion (Fig. 4i and Extended Data Fig. 9g–i). Moreover, we found that cGAS did not interact with TLR3 in the cytoplasm (Extended Data Fig. 9j,k). These observations strongly suggest cGAS and TLR3 directly sense R-loop-derived cytoplasmic RNA–DNA hybrids in the cytosol and endolysosomes, respectively, and cooperate to activate IRF3-mediated signalling (Extended Data Fig. 9l).

Pathological hybrids and disease

Finally, we examined whether IRF3 signalling can be triggered by R-loops that accumulate under pathological conditions. We used a cell line derived from an individual with the neurodegenerative disease ataxia oculomotor apraxia type 2 (AOA2), with a loss of function *SETX* mutation^{9,42,43}, and the UWB1.289 human ovarian cancer cell line, in which *BRCA1* is mutated¹⁰. Cytoplasmic hybrids were induced in fibroblasts derived from the patient with AOA2 in an XPG-dependent manner (Extended Data Fig. 10a–c). Furthermore, *IFNβ1* and several ISGs were increased in an XPG- and cGAS/TLR3-dependent manner in AOA2 fibroblasts (Fig. 4j,k and Extended Data Fig. 10d,e), as well as in control fibroblasts after *SETX* depletion (Extended Data Fig. 10f,g). Similarly, in UWB1.289 *BRCA1*-deficient cells, we observed increased cytoplasmic hybrids (Fig. 4l and Extended Data Fig. 10h,i) and an NES–RH-sensitive immune response and apoptosis, compared with isogenic controls in which *BRCA1* was restored (Fig. 4m and Extended Data Fig. 10j,k). Notably, cytoplasmic hybrids, as well as an XPG-dependent and NES–RH-sensitive immune response and apoptosis were also observed after the knockdown of *SAMHD1* (Extended Data Fig. 10l–o), which is mutated in the autoimmune disease Aicardi–Goutières syndrome²⁸. Together, these data indicate that R-loop-induced accumulation of cytoplasmic RNA–DNA hybrids and the subsequent activation of the innate immune response and apoptosis can be observed in several models of human disease.

Discussion

Here we identified that cytoplasmic RNA–DNA hybrids are immunogenic products of R-loop processing (Extended Data Fig. 9l). We show that cytoplasmic hybrids accumulate when nuclear R-loop metabolism is deregulated, and that this accumulation depends on the endonucleases XPG and XPF. Importantly, we found that endogenous cytoplasmic hybrids are sensed by the immune receptors cGAS and TLR3, of which the canonical activation has been ascribed to DNA and RNA, respectively. We therefore reveal an arm of the innate immune response and establish cellular RNA–DNA hybrids as drivers of IRF3 signalling that, when accumulated aberrantly at high levels, can induce apoptosis. Notably, low but detectable levels of cytoplasmic hybrids are present in unperturbed cells, suggesting that XPG-mediated excision may be a mechanism used to remove a small subset of persistent R-loops. However, when R-loops are deregulated or resolution pathways are disrupted, some genomic R-loops become susceptible to nucleolytic processing, raising the levels of cytoplasmic hybrids above a critical threshold for IRF3 activation.

How R-loop processing is regulated remains unclear, as is the mechanism by which these hybrids leave the nucleus, but our findings suggest that these pathways occur throughout the cell cycle and are focused on a subset of nuclear R-loops. Using next-generation sequencing on biochemically purified cytoplasmic hybrids from *SETX*-deficient cells, we traced their origin to genomic R-loops that are highly stable and exhibit distinct sequence properties, consistent with convergent transcription and hybrid formation. We envision that the increased stability of hybrids formed at these sites, stalled RNA polymerases

and potential secondary structure formation may render them more prone to XPG-mediated processing. Many hybrids were enriched in the cytoplasm only after loss of *SETX*, suggesting that these R-loop sites are normally resolved by this RNA–DNA helicase and are therefore not usually vulnerable to processing. Furthermore, XPG-dependent cytoplasmic hybrid accumulation, immune activation and apoptosis were observed in cells derived from a patient with AOA2 and in UWB1 human ovarian cancer cells, which contain mutated *SETX* and *BRCA1*. These results suggest that aberrant R-loop processing and subsequent innate immune activation may be pathological processes that could differentially affect disease outcome based on the cellular context. For example, such processing may contribute to cell death when associated with neurodegenerative diseases but act as a protective mechanism during early oncogenesis to remove genomically unstable cells in cancer cells with mutated *BRCA1* or deregulated splicing^{8,44}.

Our findings therefore reveal a new mechanistic connection linking R-loop deregulation and processing with innate immune activation that could be relevant to many human diseases. They also suggest that the innate immune response may represent a second, distinct pathological response to R-loops beyond canonical DNA damage mechanisms.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-05545-9>.

- Crossley, M. P., Bocek, M. & Cimprich, K. A. R-loops as cellular regulators and genomic threats. *Mol. Cell* **73**, 398–411 (2019).
- Sollier, J. et al. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56**, 777–785 (2014).
- Makharashvili, N. et al. Sae2/CtIP prevents R-loop accumulation in eukaryotic cells. *eLife* **7**, e42733 (2018).
- Cristini, A. et al. Dual processing of R-loops and topoisomerase I induces transcription-dependent DNA double-strand breaks. *Cell Rep.* **28**, 3167–3181 (2019).
- Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.* **16**, 583–597 (2015).
- Hatchi, E. et al. *BRCA1* recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol. Cell* **57**, 636–647 (2015).
- Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
- Schlee, M. & Hartmann, G. Discriminating self from non-self in nucleic acid sensing. *Nat. Rev. Immunol.* **16**, 566–580 (2016).
- Suraweera, A. et al. Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. *J. Cell Biol.* **177**, 969–979 (2007).
- Harding, S. M. et al. Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature* **548**, 466–470 (2017).
- He, Y. et al. NF-κB-induced R-loop accumulation and DNA damage select for nucleotide excision repair deficiencies in adult T cell leukemia. *Proc. Natl Acad. Sci. USA* **118**, e2005568118 (2021).
- Crossley, M. P. et al. Catalytically-inactive, purified RNaseH1: a specific and sensitive probe for RNA–DNA hybrid imaging. *J. Cell Biol.* **220**, e202101092 (2021).
- Nguyen, H. D. et al. Functions of replication protein A as a sensor of R loops and a regulator of RNaseH1. *Mol. Cell* **65**, 832–847 (2017).
- Bernardini, J. P. et al. Parkin inhibits BAK and BAX apoptotic function by distinct mechanisms during mitophagy. *EMBO J.* **38**, e99916 (2019).
- Tigano, M., Vargas, D. C., Tremblay-Belzile, S., Fu, Y. & Sfeir, A. Nuclear sensing of breaks in mitochondrial DNA enhances immune surveillance. *Nature* **591**, 477–481 (2021).
- Köhler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. *Nat. Rev. Mol. Cell Biol.* **8**, 761–773 (2007).
- Chatzidoukaki, O. et al. R-loops trigger the release of cytoplasmic ssDNAs leading to chronic inflammation upon DNA damage. *Sci. Adv.* **7**, eabj5769 (2021).
- Koo, C. X. et al. RNA polymerase III regulates cytosolic RNA:DNA hybrids and intracellular microRNA expression. *J. Biol. Chem.* **290**, 7463–7473 (2015).
- Smolka, J. A., Sanz, L. A., Hartono, S. R. & Chédin, F. Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. *J. Cell Biol.* **220**, e202004079 (2021).
- Petermann, E., Lan, L. & Zou, L. Sources, resolution and physiological relevance of R-loops and RNA–DNA hybrids. *Nat. Rev. Mol. Cell Biol.* **23**, 521–540 (2022).
- Crossley, M. P., Bocek, M. J., Hamperl, S., Swigut, T. & Cimprich, K. A. qDRIP: a method to quantitatively assess RNA–DNA hybrid formation genome-wide. *Nucleic Acids Res.* **48**, e84 (2020).
- Sanz, L. A. et al. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).

23. Lim, G. & Hohng, S. Single-molecule fluorescence studies on cotranscriptional G-quadruplex formation coupled with R-loop formation. *Nucleic Acids Res.* **48**, 9195–9203 (2020).
24. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chédin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).
25. Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I. & Chédin, F. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* **23**, 1590–1600 (2013).
26. Xu, W. et al. The R-loop is a common chromatin feature of the *Arabidopsis* genome. *Nat. Plants* **3**, 704–714 (2017).
27. Shen, Y. J. et al. Genome-derived cytosolic DNA mediates type I interferon-dependent rejection of B cell lymphoma cells. *Cell Rep.* **11**, 460–473 (2015).
28. Coquel, F. et al. SAMHD1 acts at stalled replication forks to prevent interferon induction. *Nature* **557**, 57–61 (2018).
29. Mackenzie, K. J. et al. cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465 (2017).
30. Weinreb, J. T. et al. Excessive R-loops trigger an inflammatory cascade leading to increased HSPC production. *Dev. Cell* **56**, 627–640 (2021).
31. Giordano, A. M. S. et al. DNA damage contributes to neurotoxic inflammation in Aicardi–Goutières syndrome astrocytes. *J. Exp. Med.* **219**, e20211121 (2022).
32. Suter, M. A. et al. cGAS-STING cytosolic DNA sensing pathway is suppressed by JAK2-STAT3 in tumor cells. *Sci. Rep.* **11**, 7243 (2021).
33. Cristini, A. et al. RNase H2, mutated in Aicardi–Goutières syndrome, resolves co-transcriptional R-loops to prevent DNA breaks and inflammation. *Nat. Commun.* **13**, 2961 (2022).
34. Chawla-Sarkar, M. et al. Apoptosis and interferons: role of interferon-stimulated genes as mediators of apoptosis. *Apoptosis* **8**, 237–249 (2003).
35. Boulares, A. H. et al. Role of poly(ADP-ribose) polymerase (PARP) cleavage in apoptosis. Caspase 3-resistant PARP mutant increases rates of apoptosis in transfected cells. *J. Biol. Chem.* **274**, 22932–22940 (1999).
36. Porter, A. G. & Jänicke, R. U. Emerging roles of caspase-3 in apoptosis. *Cell Death Differ.* **6**, 99–104 (1999).
37. Wajant, H., Pfizenmaier, K. & Scheurich, P. Tumor necrosis factor signaling. *Cell Death Differ.* **10**, 45–65 (2003).
38. Zhou, Y., He, C., Wang, L. & Ge, B. Post-translational regulation of antiviral innate signaling. *Eur. J. Immunol.* **47**, 1414–1426 (2017).
39. Mankan, A. K. et al. Cytosolic RNA:DNA hybrids activate the cGAS-STING axis. *EMBO J.* **33**, 2937–2946 (2014).
40. Blasius, A. L. & Beutler, B. Intracellular toll-like receptors. *Immunity* **32**, 305–315 (2010).
41. Abu-Remaileh, M. et al. Lysosomal metabolomics reveals V-ATPase- and mTOR-dependent regulation of amino acid efflux from lysosomes. *Science* **358**, 807–813 (2017).
42. Yeo, A. J. et al. R-loops in proliferating cells but not in the brain: implications for AOA2 and other autosomal recessive ataxias. *PLoS ONE* **9**, e90219 (2014).
43. Kanagaraj, R. et al. Integrated genome and transcriptome analyses reveal the mechanism of genome instability in ataxia with oculomotor apraxia 2. *Proc. Natl Acad. Sci. USA* **119**, e2114314119 (2022).
44. Semmler, L., Reiter-Brennan, C. & Klein, A. *BRCA1* and breast cancer: a review of the underlying mechanisms resulting in the tissue-specific tumorigenesis in mutation carriers. *J. Breast Cancer* **22**, 1–14 (2019).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022, corrected publication 2024

Methods

Cell culture and transfection

HeLa, HCT116, MCF10A and HEK293T cells were obtained from ATCC, where they were tested for mycoplasma and verified by STR profiling, and grown in DMEM (Gibco) supplemented with 10% FBS and 1% penicillin–streptomycin–glutamine (PSG). Control normal foreskin fibroblasts and fibroblasts from a patient with AOA2 (SETX-1RM)⁹ (gifts from S. West) were cultured in DMEM supplemented with 15% FBS and 1% PSG (lacking FBS for serum starvation for 3 days). UWB1.289 (UWB1) or UWB1.289+BRCA1 (UWB1+BR1) reconstituted cells (gifts from R. Greenberg¹⁰) were cultured in 1:1 RPMI1640 and MEGM (BulletKit, Lonza) with 10% FBS, penicillin and streptomycin. MCF10A cells were cultured in 1:1 DMEM and F12 medium, 5% horse serum, 0.5 $\mu\text{g ml}^{-1}$ hydrocortisone, 10 $\mu\text{g ml}^{-1}$ insulin, 20 ng ml^{-1} EGF, 100 ng ml^{-1} cholera toxin and 1% PSG (for serum starvation, medium lacking horse serum and EGF was used for 48 h). All cells were grown in a 37 °C humid incubator with 5% CO_2 . siRNA transfections were performed using Lipofectamine RNAiMax (Thermo Fisher Scientific) and 20 nM siRNA (Supplementary Table 1). Plasmid DNA transfections were performed with FuGENE HD (Promega) for 48 h or as indicated. For transfection into AID-fused XPG degran cells, 4 mM indole-3-acetic acid (IAA, Sigma-Aldrich) or an equal volume of DMSO was added to the culture medium immediately after transfection. The following inhibitors were used for the times indicated: 2 $\mu\text{g ml}^{-1}$ cGAS inhibitor RU.521 (InvivoGen), 500 nM Plab (Cayman Chemicals) and 5 nM leptomycin B (LMB) (Cayman Chemicals).

Immunofluorescence

Immunofluorescence experiments with GFP–dRH protein were performed as described previously¹². For cyclin B1, cells were fixed with 4% paraformaldehyde for 20 min at room temperature and permeabilized with 0.25% Triton X-100 and incubated with cyclin B1 antibodies at 4 °C overnight, then finally incubated with 5 ng ml^{-1} DAPI, 0.2 μl HCS Cellmask Deep Red (Thermo Fisher Scientific) and Alexa Fluor 488. Coverslips were mounted onto glass slides using Prolong Glass antifade mountant (Thermo Fisher Scientific). A list of antibody dilutions is provided in Supplementary Table 1.

Image acquisition and analysis

For GFP–dRH analysis, images were acquired as described previously¹². Using CellProfiler (v.4.2.1), the DAPI channel was used to identify nuclei using the IdentifyPrimaryObjects module, primary objects and whole-cell stain were then used to identify cells as secondary objects. The cytoplasmic area was identified as a tertiary object from the whole cells shrunk by one pixel and nuclei expanded by three pixels. The mean intensity for each cytoplasmic area was calculated and exported. For epifluorescence imaging, the Zeiss OBSERVER.Z1 INVERTED microscope was used with a Plan-APO $\times 40/1.4$ NA oil-immersion DIC (UV) VIS-IR objective. Images were adjusted equally in ImageJ (v.2.0.0).

Cell cycle analysis

To monitor cell cycle synchronization in MCF10As, cells were incubated with 10 μM 5-bromo-2'-deoxyuridine (BrdU) and processed according to the manufacturer's guidelines (BD Biosciences). Data analysis was performed using FlowJo v.3.05 and the gating strategy is shown in Supplementary Fig. 2. For fibroblasts, cells were incubated with 10 μM EdU for 30 min and processed⁴⁵.

cytoDRIP

Cells ($10\text{--}50 \times 10^6$) were collected using trypsin, washed in PBS and pelleted by centrifugation and fractionated using the Nuclear and Cytoplasmic Extraction kit (Thermo Fisher Scientific). The cytoplasmic or nucleoplasmic fractions were recovered and incubated in 0.4% SDS and 40 $\mu\text{g ml}^{-1}$ of proteinase K (Thermo Fisher Scientific) for 90 min at 37 °C. The samples were resuspended in ultrapure

water, normalized by protein concentration in the cytoplasmic or nucleoplasmic extract or to equal cell counts, adjusted to 550 mM NaCl and treated with RNase A (1 $\mu\text{g ml}^{-1}$) for 25–45 min. For RNase H treatment, the samples were digested overnight in 1 $\times$ NEB RNase H buffer and RNase H (0.4 U μl^{-1}). For immunoprecipitation, 16 μg of S9.6 antibody or mouse IgG was bound to Dynabeads Protein G beads (Thermo Fisher Scientific) in 1 $\times$ binding buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 1% Triton X-100, 150 mM NaCl, 0.5% sodium deoxycholate) for 4–6 h at 4 °C. In parallel, the samples were resuspended in 1 $\times$ TE buffer and then precleared with Dynabeads Protein G for 1–2 h in 1 $\times$ binding buffer. Precleared genomic samples were then added to S9.6-antibody- or IgG-bound beads and incubated overnight with rotation at 4 °C. Bound beads were washed with TSE buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 150 mM NaCl), and then with TE buffer. Elution was performed in 200 μl elution buffer (50 mM Tris pH 8, 10 mM EDTA, 0.5% SDS, 8 μl proteinase K 20 mg ml^{-1}) for 50 min at 50 °C. For cytoplasmic hybrid association with TLR3, hybrids were first enriched by TLR3 immunoprecipitation, eluted in elution buffer as described above, before the second immunoprecipitation with S9.6. For cytoDRIP–qPCR, the samples were resuspended in ultrapure water and analysed. For cytoDRIP blotting, the eluted samples were resuspended in TE buffer and 3'-end labelled with 2 μM biotin-11-dUTP (Biotium) and 0.2 U μl^{-1} of TdT (NEB) in 1 $\times$ TdT reaction buffer supplemented with 0.25 mM CoCl_2 for 45 min at 37 °C. The labelling reactions were stopped by addition of 20 mM EDTA and put on ice. The reactions were then incubated with 0.4% SDS and 40 $\mu\text{g ml}^{-1}$ of proteinase K for 40 min at 37 °C. About 100 ng of 100 bp DNA ladder (NEB) was labelled in an identical reaction for use as a molecular mass marker during gel electrophoresis. The labelled samples were separated on 4–20% TBE gels, transferred onto Biodyne B Nylon membranes (Thermo Fisher Scientific) and fixed by ultraviolet cross-linking. Membranes were processed as described previously⁴⁶ and chemiluminescence was detected by X-ray film. The uncropped gel images are provided in Supplementary Fig. 1.

Nuclear DRIP–qPCR

Nuclear DRIP–qPCR was performed as described previously²¹. For hybrid lifetime analysis, actinomycin D (2 $\mu\text{g ml}^{-1}$, Cayman Chemical Company) was added before collection. For RNase H titration, the samples were treated after cell lysis and before immunoprecipitation in 200 μl reaction volumes as follows: 13 μg DNA with 0.5 U RNase H for 5 min; 13 μg DNA with 0.5 U RNase H for 15 min; 13 μg DNA with 2.5 U RNase H for 45 min; 5.5 μg DNA with 100 U RNase H for 16 h; 5.5 μg DNA with 100 U RNase H for 40 h.

Library preparation and sequencing for cytoDRIP–seq

After elution, genomic material was resuspended in TE buffer and sonicated to a peak fragment size of 300 bp, performed on the Covaris machine (E220 evolution) (10% duty factor, 200 cycles per burst, 140 peak incident power, 30 s per tube). DNA libraries were synthesized from ssDNA using the Accel-NGS1S DNA library kit (Swift Biosciences) as described previously²¹. Library DNA was sequenced on the HiSeq 4000 (Illumina) system at the Stanford Genome Sequencing Service Center, using 2 $\times$ 150 bp sequencing.

ChIP analysis

A total of 5–15 million cells were cross-linked per ChIP sample in 25 ml PBS with 1% methanol-free formaldehyde for 10 min and quenched with a final concentration of 0.125 M glycine for 5 min with nutation. The samples were processed as described previously⁴⁷. A total of 7.5 μg GFP antibody was added per ChIP sample and incubated overnight at 4 °C. Protein G Dynabeads (50 μl , Thermo Fisher Scientific) were blocked with block solution (0.5% BSA (w/v) in 1 $\times$ PBS) and then added to antibody-bound chromatin for 4 h, washed and eluted. ChIP and input samples were purified by phenol–chloroform–isoamylalcohol

Article

extraction and ethanol precipitated. A list of the antibodies used is provided in Supplementary Table 1.

qPCR

Cells were collected 48 h after transfection and lysed with Trizol (Invitrogen). RNA was isolated by phenol–chloroform extraction and converted to cDNA using the SuperScript III First-Strand Synthesis System (Invitrogen). qPCR was performed on the Roche LightCycler 480 Instrument II using the SYBR-Green master mix (Bio-Rad Laboratories). To measure the transcription level, primers recognizing the transcript of genes of interest and *ACTB*, serving as the internal control, were designed, and the RNA level of each target gene was normalized to that of *ACTB*. For splicing inhibition, cDNA was amplified with primers within different introns to monitor mRNA splicing efficiency. qPCR data were analysed using Roche LightCycler (v.1.5.1). A list of the primers used for qPCR is provided in Supplementary Table 1. C_t values from qPCR analysis of immune genes are provided in the Source Data.

cytoDRIP-seq analysis

Trimmed reads (using cutadapt v.1.16) were aligned to human genome reference hg38 using bowtie2 (v.2.3.4). Reads were separated into positive- and negative-stranded files using SAMtools (v.1.10) and Unix text-processing utilities. Genome browser tracks were produced with the BEDTools genomecov utility, normalized to reads per million mapped, and visualized using IGV (v.2.8.2). Tracks for nuclear DRIP-seq and RNase-H-resistant DRIP-seq signal in HeLa cells were generated previously²¹.

Peak calling

Peaks were called against a merged BAM file from all IgG samples using MACS2 with narrow peak settings. BEDTools (v.2.29.2) was then used to obtain coverage in each experiment over these consensus peaks. Using these read counts, we filtered out peaks that were highest in IgG coverage (top 5% measured by reads per million), and then filtered only for regions with a peak score of >50. The remaining peaks from two biological replicates were merged for siCtrl and siSETX samples. This resulted in 2,911 peaks that were used for subsequent meta analyses. Strand annotations were assigned by intersecting peaks with genes expressed in HeLa cells²¹.

Metaplots

Metaplots around cytoDRIP peaks and other genome features were produced using deepTools (v.3.2.1). Tracks for GC and AT skew were generated as described previously²¹. GC and AT content within peaks was calculated using bedtools nuc. Ratios of sense and antisense hybrid signal were calculated from the coverage of plus and minus strand reads within the peaksets indicated. Only peaks with at least one sense and one antisense read were included. Data processing for all genomic plots was performed using Python v.3.7.13, NumPy v.1.21.5 and Pandas v.1.3.5. Data were visualized using the Python packages Matplotlib v.3.5.1 and Seaborn v.0.11.2. Statistical analysis was performed in Python using SciPy v.1.7.3.

Analysis of genome features

Analysis of genome compartments overlapping cytoDRIP peaks was performed using the *cis*-regulatory element annotation system. For intergenic enhancer and insulator annotations, ChromHMM annotations for HeLa cells (from the UCSC Table Browser) were intersected with intergenic cytoDRIP peaks using bedtools intersect. For DNA repeat elements, RepeatMasker annotations were used. The cytoDRIP peaks were subsampled 10,000 times (bedtools shuffle) from within all nuclear DRIP peaks. A Jaccard coefficient (bedtools Jaccard) was calculated for each randomized peak set and the Z-score was calculated from the resulting distribution. Consensus rDNA regions (5.8S:NG_054872.1, 18S:NG_054871.1, 28S:M11167, total rDNA:U13369.1)

and alpha satellite (M95601.1) were indexed and sequencing reads were then aligned to these using bwa mem. For telomere sequence analysis, R1 and R2 reads with ≥ 3 instances of AATCCC or TTAGGG were counted.

R-loop lifetime and RNase H half-life analysis

Percentage inputs were obtained by qPCR for each hybrid region from a time course with actinomycin D treatment or a time course of RNase H treatment and half-lives were derived by fitting an exponential decay function to these measurements. Short-, average- and long-lived nuclear R-loop sites were identified previously²¹.

Electrophoretic mobility shift assay

RNA60 and DNA60 oligos were 5'-end labelled with γ -³²P-ATP by T4 polynucleotide kinase (NEB) and purified using the Illustra G-25 microspin column (GE Lifesciences). The RNA60 oligo was annealed with unlabelled DNA60 or antisense RNA60 (asRNA60) oligo to form a 60 bp-long hybrid or dsRNA substrate. DNA60 oligo and unlabelled asDNA60 oligo were annealed together to form a dsDNA substrate. For annealing, two oligonucleotides were mixed at a 1:1 molar ratio in buffer containing 1.25 mM EDTA and 12.5 mM Tris-HCl pH 7.6. The samples were heated at 95 °C for 5 min and were allowed to slowly cool to room temperature. A list of the oligo sequences is provided in Supplementary Table 1. For the electrophoretic mobility shift assay reaction, proteins purified from mammalian cells were mixed with 1 nM labelled substrate at molar ratios ranging from 1:1 to 40:1. The samples were incubated for 30 min at 37 °C in a total volume of 10 μ l containing 25 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA pH 8.0, 1 mM DTT, 6% glycerol, 0.1% NP-40, 0.1 mg ml⁻¹ BSA, RNaseOUT and 0.5 mM PMSF. Afterwards, the samples were run on a non-denaturing 6% polyacrylamide gel in 0.5 $\times$ TBE buffer at 4 °C and visualized using the Typhoon imager scanner. Uncropped gel images are in Supplementary Fig. 1.

In vitro pull-down

To measure the hybrid-protein binding affinities, we performed in vitro pull-down using the radioactively labelled 60-bp-long RNA–DNA hybrid described above. Flag-tagged proteins were purified from mammalian cells and immobilized to M2 magnetic beads (Sigma-Aldrich). In each reaction, 62.5 fmol hybrid was incubated with a specific protein at 1:10 molar ratio, and 1 $\times$ binding buffer (25 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA pH 8.0, 1 mM DTT, 0.1% NP-40, 0.1 mg ml⁻¹ BSA) was added into the reaction to bring the total volume up to 150 μ l. The samples were held on a rotator for 3 h at room temperature. After the binding reaction, magnetic beads were washed three times with 1 $\times$ binding buffer to remove the unassociated hybrids. The hybrids that were still associated with the beads were eluted by proteinase K digestion, which was performed in 10 μ l 1 $\times$ binding buffer with 0.1% SDS for 30 min at 37 °C. Eluted samples were analysed by non-denaturing 6% polyacrylamide gel in 0.5 $\times$ TBE buffer at 4 °C followed by autoradiography imaging. Human RNase H1 catalytic inactive mutant (D210N) and GFP were used as positive and negative controls, respectively. Uncropped gel images are provided in Supplementary Fig. 1.

Cytoplasmic S9.6 and TLR3 immunoprecipitation

HeLa cells were fractionated as described previously⁴⁸. The purity of the cytoplasmic portion was confirmed by analysing lamin B1 protein using western blotting. After fractionation, hybrids in the cytoplasm were enriched by the S9.6 antibody or the TLR3 antibody and 30 μ l protein A/G agarose (Life Technologies) in IP buffer (150 mM KCl, 0.2 mM EDTA, 5 mM MgCl₂, 20 mM Tris-HCl pH 8.0, 10% glycerol) at 4 °C overnight. The beads were washed three times in the IP buffer and immunoprecipitated material was eluted in Laemmli buffer. For hybrid competition S9.6 co-IP, a 15 bp synthetic hybrid annealed by Com-RNA15 and Com-DNA15 was added into the samples at the S9.6 immunoprecipitation step as described previously⁴⁹, at the indicated concentration. A list of the oligonucleotide sequences and antibody

amounts used for each immunoprecipitation reaction is provided in Supplementary Table 1. For RNase H digestion, 50 U ml⁻¹ RNase H was added to the sample for 1 h at 37 °C before the immunoprecipitation step. Detailed information on the S9.6 and TLR3 antibodies used in IPs is provided in Supplementary Table 1.

LysolP

HEK293T HA-Lyso cells expressing TMEM192-3×HA and Control-Lyso cells expressing TMEM192-2×FLAG⁴¹ were cultured as indicated. For each condition, 6 × 10⁷ cells were processed per immunoprecipitation and 6 immunoprecipitation reactions were performed in parallel⁴¹. Cells were rinsed twice with PBS, scraped in 1 ml of PBS, pooled based on condition and then centrifuged at 1,000 g for 2 min at 4 °C. Cells were resuspended and divided into six parts of 950 µl PBS. The cells were homogenized with 25 strokes of a 2 ml homogenizer. The homogenate was centrifuged at 1,000 g for 2 min at 4 °C and the supernatant containing the cellular organelles including lysosomes was incubated with 100 µl of PBS prewashed anti-HA magnetic beads (Thermo Fisher Scientific) on a rotator shaker for 3 min at 4 °C. The immunoprecipitates were washed three times with PBS on a DynaMag Spin Magnet. Beads with bound lysosomes were then resuspended in 400 µl DRIP elution buffer with 20 µl proteinase K, incubated for 1.5 h at 50 °C and nucleic acids were purified by phenol–chloroform extraction, before continuing with the cytoDRIP protocol. For protein lysates, 2.5% total cells were reserved for the whole-cell fraction. The remaining cells were used for lysolP as described above except that, after the final wash, beads with bound lysosomes were resuspended in 80 µl of lysis buffer (50 mM HEPES pH 7.4, 1% Triton X-100, 10 mM β-glycerol phosphate, 10 mM pyrophosphate, 1.5 mM NaF, 40 mM NaCl, 2 mM EDTA and cOmplete Protease Inhibitor Cocktail), incubated for 30 min on ice and the supernatant was recovered. For the whole-cell fraction, lysis buffer was added to the samples for 30 min and protein extracts were recovered after centrifugation at maximum speed for 10 min at 4 °C. Uncropped gel images are provided in Supplementary Fig. 1. A list of the antibodies used in lysolP is provided in Supplementary Table 1.

Caspase-3 activity

A total of 7,000 HeLa cells were seeded into triplicate wells of a 96-well plate containing the transfection reagent–siRNA mixture. The medium was refreshed the next morning. Then, 48 h after siRNA transfection, caspase-3 activity was measured using the Apo-ONE Homogeneous Caspase-3/7 Assay (Promega) according to the manufacturer's instructions.

Cloning of NLS- and NES-tagged RNase H

For NLS-tagged RNase H, cDNA fragments containing wild-type (WT) human RNase H1 and the D201N catalytic dead mutant without the first 27 amino acid mitochondrial localization signal were amplified by PCR. Two SV40 nuclear localization signals (NLS) (CCCAAAAGAAACGCAAAGTG) were introduced by the forward primer at the PCR step. A list of the primers used in PCR amplification is provided in Supplementary Table 1. For NES-tagged RNase H, the sequence encoding *Escherichia coli* RNase H1 was tagged with a nuclear export signal (NES) (CTGTCTCCCACTTCCAGG AGCTGTCCATC) at the C-terminal end. Both NLS- and NES-tagged DNA fragments were then cloned into the pEGFP-N2 vector. The DNA fragments containing RNase H and eGFP were then moved into the pLVX-puro vector, enabling us to establish stable cell lines with these plasmids through lentiviral infection. Nuclear and cytoplasmic localization of the proteins was confirmed by fluorescence microscopy (Extended Data Figs. 7e,i and 10j). A vector only expressing eGFP was used as mock control.

Western blot

Cells were lysed in lysis buffer (300 mM KCl, 0.2 mM EDTA, 5 mM MgCl₂, 20 mM Tris-HCl pH 8.0, 10% glycerol) with proteinase and phosphatase inhibitors (Sigma-Aldrich). Protein samples were then sonicated with a probe-type ultrasonicator and analysed by SDS–PAGE as described

previously⁵⁰. Chemiluminescence was detected by exposure to an X-ray film. Representative blots are shown in the figures. A list of the antibodies used in western blotting is provided in Supplementary Table 1. Uncropped gel images are provided in Supplementary Fig. 1.

Construction of HeLa auxin-inducible XPG degron

To insert an AID downstream of the *XPG* coding sequence in the genomic DNA, guide DNAs (gDNAs) targeting the 3' untranslated region (UTR) of the human *XPG* gene were designed using Zhang Lab design resources (<https://zlab.bio/guide-design-resources>). gDNA oligos with BbsI overhangs were annealed and ligated into pSpCas9(BB)-2A-Puro(PX459) V2.0 (62988, Addgene). The homology arms flanking the *XPG* 3' UTR and overwriting the stop codon were produced by PCR with Phusion High-Fidelity DNA polymerase (NEB) using the *XPG* 5' arm or *XPG* 3' arm forward and reverse primers (Supplementary Table 1). The amplified *XPG* 5' and 3' arms were then assembled together with a fragment encoding eGFP–AID and DNA backbone using HiFi DNA Assembly (NEB). The eGFP–AID fragment was inserted between the *XPG* 5' arm and *XPG* 3' arm designed to be inserted into the genome by homology-directed repair. All of these plasmids were transfected into HeLa cells using FuGENE HD. At 24 h after transfection, cells were selected with 1 µg ml⁻¹ puromycin for 48 h and then, 14 days after transfection, GFP-positive cells were selected by fluorescence-activated cell sorting (FACS). Homozygous clones were identified by PCR and western blotting with the anti-XPG antibody. To stably express *Oryza sativa* TIR1, homozygous clones with eGFP–AID successfully inserted were infected with lentivirus expressing both *OsTIR1* and blue fluorescent protein (BFP). BFP-positive cells were obtained by FACS.

Construction of HCT116 auxin-inducible XPG degron

HCT116 cells expressing degron-tagged XPG were generated by transfecting HCT116-*OsTIR1* cells⁵¹ with pLentiCRISPR-V2 plasmid⁵² expressing Cas9 and an sgRNA targeting the *XPG* C terminus (AAGGAACTAAG ACGTGCGA) and with two homology-directed repair constructs based on pMK289 and pMK290 plasmids (a gift from M. Kanemaki⁵¹) containing the mAID-mClover sequence and a hygromycin- or neomycin-resistance cassette flanked by the 200 bp *XPG* homology arms. A homozygous knockin cell line was obtained after selection with G418 and hygromycin, which was verified by genotyping and sequencing.

Establishing *BAX*^{-/-}*BAK*^{-/-} double-knockout cell lines

Cell lines were generated as previously described¹⁵. In brief, HeLa and MCF10A cells were co-transfected with two plasmids, gifts from the Sfeir laboratory (Addgene plasmids 167296 and 167295). Each of them contains the sequence encoding Cas9 and two sgRNAs targeting *BAX* (GCTGCAGGATGATTGCCGCCG and GTCTGACGGCACTTCAACTG) or *BAK* (GCATGAAGTCGACCACGAAG and GCCCATGCTGGTAGACGTGT), respectively. Then, 48 h after transfection, cells were then treated with three drugs—A-1331852, ABT-199 and S63845 (MedChem Express). HeLa cells were treated with 1 µM each of drug and MCF10A cells with 0.5 µM each of drug, and the surviving cells were selected.

CRISPR–Cas9 knockout cell lines

To establish *cGAS*- and *TLR3*-knockout cell lines, we first generated HeLa cells stably expressing Cas9. For this purpose, cells were infected with a lentivirus encoding Cas9 and a blasticidin-resistance marker (Lentiviral Prep 52962-LV, Addgene). Next, negative control sgRNA (TrueGuide sgRNA A35526, Thermo Fisher Scientific) and sgRNAs targeting *cGAS* or *TLR3* (Supplementary Table 1) were transfected into the prepared Cas9 stable cell line using Lipofectamine RNAiMax. Cells were then split into 15 cm dishes and single clones were selected and validated by western blotting. Two clones from each group were used in experiments.

Statistical analysis and reproducibility

All box and whisker plots demarcate the median (centre line), 75th and 25th percentiles (upper and lower bounds, respectively), and

Article

minimum and maximum values (whiskers). Two-tailed Mann–Whitney *U*-tests were performed to determine statistical significance of three biologically independent replicates in aggregate. Cell numbers per condition are provided in the figure legends. For all bar graphs, statistical analysis was performed using GraphPad Prism v.9.3.1 or Python v.3.7.13 using SciPy v.1.7.3. Bar graphs represent the mean $\pm$ s.d. When comparing two samples (the aggregate of three biologically independent experiments), unpaired two-tailed *t*-tests were performed. Each western blot, cytoDRIP blot, in vitro binding assay or co-IP was repeated at least three times as biologically independent experiments (Figs. 1c–k, 3a,c,e,g,h,j and 4a,c–i,l,m and Extended Data Figs. 1a,d–j,m, 2a,c,e,h,j–m, 4k, 7a–d,g–k,m, 8a,c,e,f,j–m, 9a–k and 10b,c,h,i,m–o). The immunofluorescence experiments shown in Fig. 1a and Extended Data Fig. 1c,k–l depict representative images and the quantifications are the aggregate of three biological replicates. Other micrographs (Extended Data Figs. 1b, 2n, 7e,l and 10j) were performed once as proof-of-concept examples. No statistical methods were used to determine sample size. Experiments were not randomized, nor were the investigators blinded to allocation. All data were assembled into figures with Adobe Illustrator CS6.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All sequencing data generated in this work have been deposited at the Gene Expression Omnibus (GEO) under accession number GSE178841. For nuclear DRIP-seq, datasets under accession number GSE134084 were used from <https://doi.org/10.1093/nar/gkaa500>. Source data are provided with this paper.

Code availability

Further code information is available on request from the corresponding author.

45. Bai, G. et al. HLTf promotes fork reversal, limiting replication stress resistance and preventing multiple mechanisms of unrestrained DNA synthesis. *Mol. Cell* **78**, 1237–1251 (2020).

46. Choi, J.-H., Kim, S.-Y., Kim, S.-K., Kemp, M. G. & Sancar, A. An integrated approach for analysis of the DNA damage response in mammalian cells: nucleotide excision repair, DNA damage checkpoint, and apoptosis. *J. Biol. Chem.* **290**, 28812–28821 (2015).
47. Long, H. K. et al. Loss of extreme long-range enhancers in human neural crest drives a craniofacial disorder. *Cell Stem Cell* **27**, 765–783 (2020).
48. Guerra, J. et al. Lysyl-tRNA synthetase produces diadenosine tetraphosphate to curb STING-dependent inflammation. *Sci. Adv.* **6**, eaax3333 (2020).
49. Cristini, A., Groh, M., Kristiansen, M. S. & Gromak, N. RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell Rep.* **23**, 1891–1905 (2018).
50. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **170**, 774–786 (2017).
51. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep.* **15**, 210–218 (2016).
52. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

Acknowledgements We thank J. Wysocka, S. Hamperl and J. Sollier for discussions and comments; and M.-S. Tsai (for GFP-dRH) and the members of the Straight laboratory for help in designing AID-XPG in HeLa cells. This work was supported by the Leukemia and Lymphoma Society (5455-17 to M.P.C.); the National Institutes of Health (GM119334 to K.A.C., S10OD018220 to the Stanford Functional Genomics Facility, T32-CA09302 to M.J.B., T32-HG000044 to C.L., DP2-CA271386 to M.A.-R.); Stanford Cancer Institute, an NCI-designated Comprehensive Cancer Center, to M.A.-R. and M.P.C.; the Korea Research Institute of Standards and Science (KRISS-GP2021-0003-10 to J.-H.C.), the National Research Foundation of Korea (MSIT) (NRF-2020R1A2C1101575 to J.-H.C.); the National Science Foundation (GRFP to C.L.), Jane Coffin Childs Memorial Fund for Medical Research (61-1755 to J.R.B.), the Gravitation Program CancerGenomics.nl from the Netherlands Organisation for Scientific Research (NOW) and the Oncode Institute, which is partly financed by the Dutch Cancer Society to W.V., and the V Foundation (D2018-017 to K.A.C.). M.A.-R. is a Terman Fellow and Pew-Stewart Scholar. K.A.C. is an ACS research professor.

Author contributions M.P.C., C.S., M.J.B. and K.A.C. designed the study. M.P.C., C.S., J.-H.C., M.J.B., J.R.B., G.B. and C.L. performed the experiments and data analyses. M.P.C. and M.J.B. performed the bioinformatic analyses. K.A.C. and M.A.-R. supervised the experiments and data analyses. J.N.K. and A.S. provided technical support. H.L. and W.V. designed and prepared HCT116 XPG-AID cells. M.P.C., C.S. and K.A.C. prepared the manuscript with contributions from the other authors.

Competing interests K.A.C. is a scientific advisory board member of RADD Pharmaceuticals and IDEAYA Biosciences. M.A.-R. is a scientific advisory board member of Lycia Therapeutics. The other authors declare no competing interests.

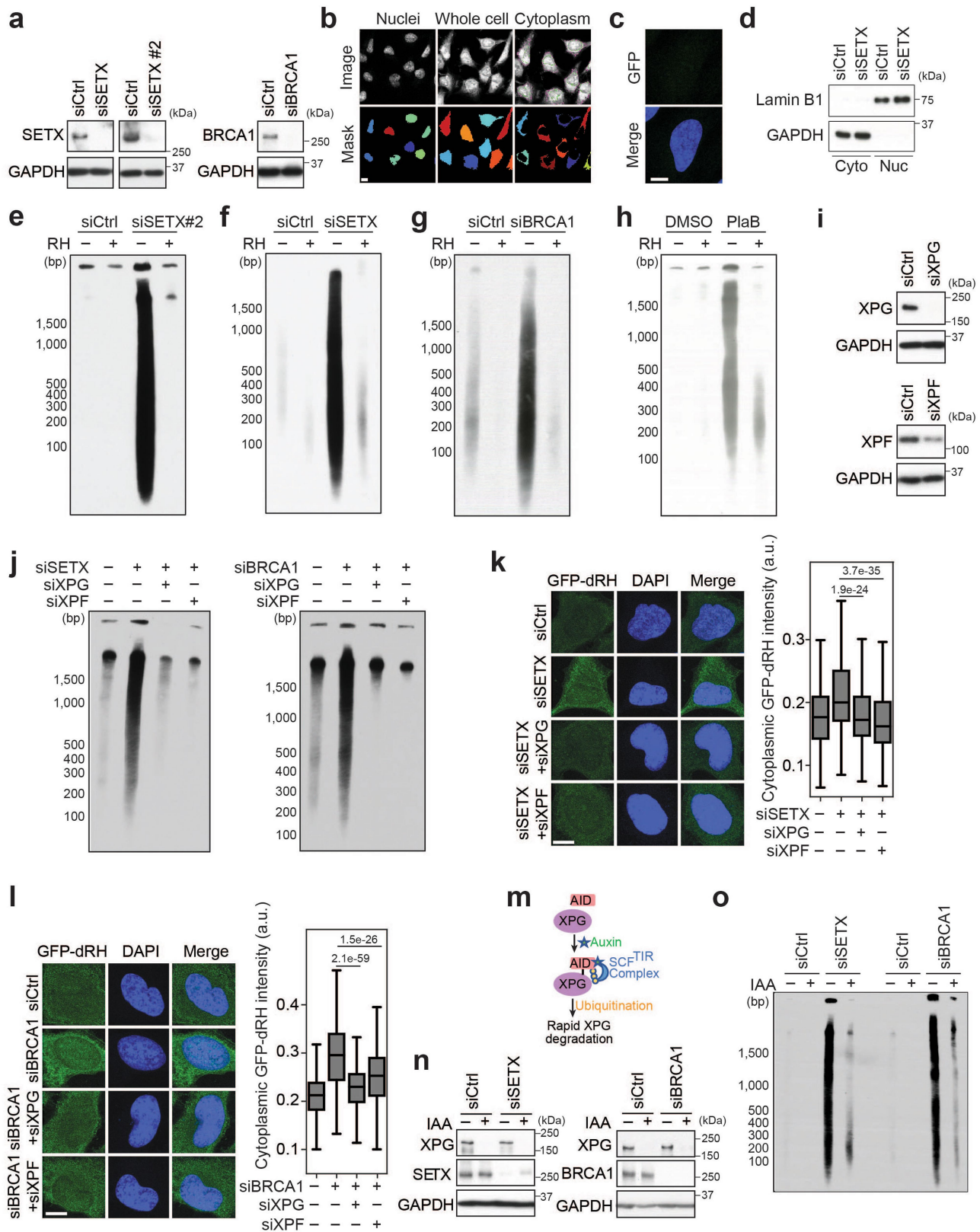
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-05545-9>.

Correspondence and requests for materials should be addressed to Karlene A. Cimprich.

Peer review information Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

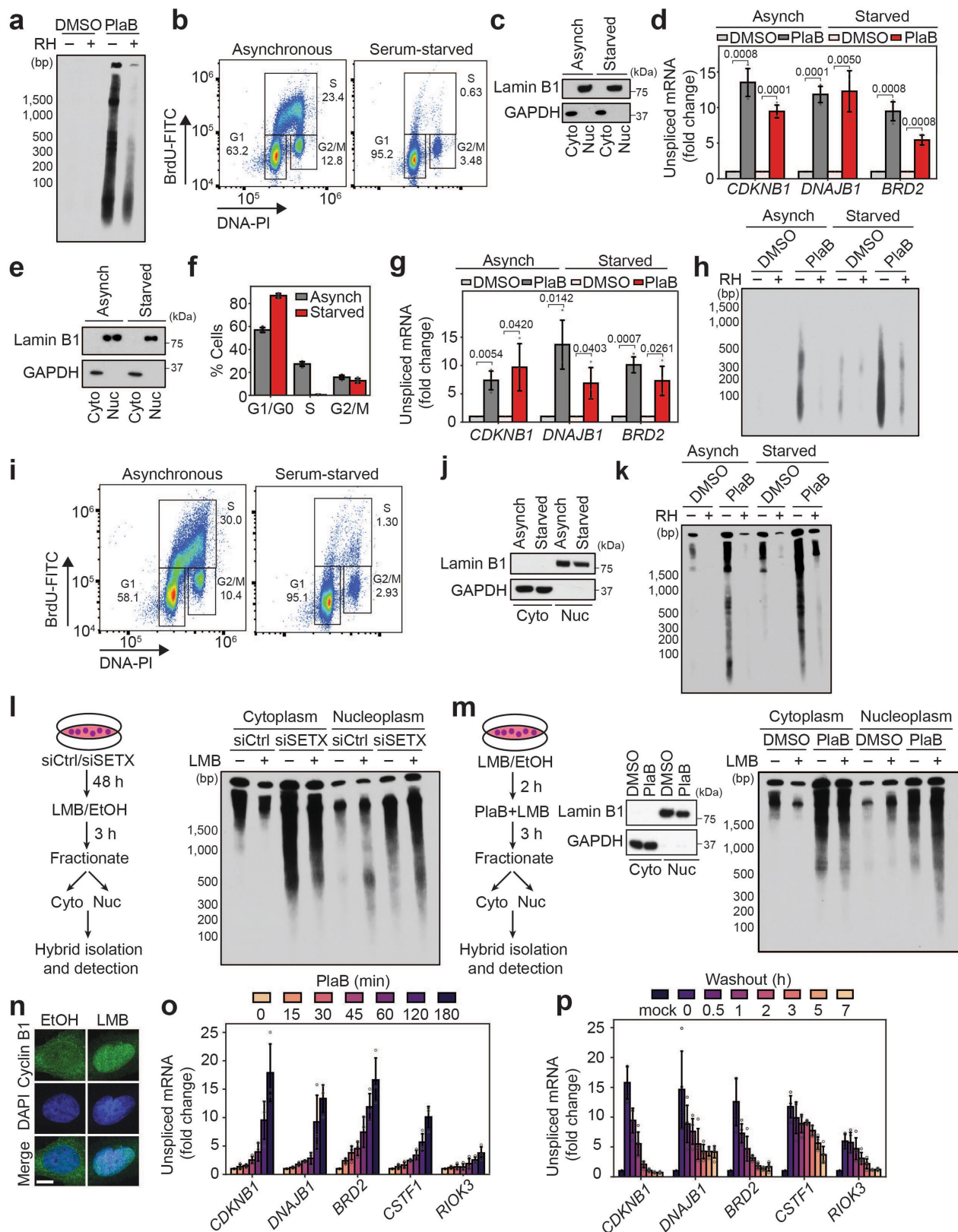


Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | Cytoplasmic RNA–DNA hybrids are induced upon multiple cellular perturbations. (a) Western blot showing knockdown efficiency of siRNAs targeting *SETX* and *BRCA1* in HeLa cells. (b) Images showing segmentation of nuclear and cytoplasmic compartments in HeLa cells, using DAPI and whole cell stain as masks, respectively. Scale bar is 20 μm . (c) Images showing the lack of GFP protein binding on fixed HeLa cells. Scale bar is 10 μm . (d) Western blot showing fractionation of siCtrl and siSETX-treated HeLa cells into soluble nuclear and cytoplasmic compartments with Lamin B1 and GAPDH as markers, respectively. (e) cytoDRIP blot showing cytoplasmic hybrid accumulation in HeLa cells following *SETX* knockdown using a second siRNA. In vitro RH treatment was performed prior to pull-down. (f) cytoDRIP blot showing cytoplasmic hybrid accumulation in siSETX-treated HCT116 cells. In vitro RH treatment was performed prior to pull-down. (g) cytoDRIP blot showing cytoplasmic hybrid accumulation in siBRCA1-treated HCT116 cells. In vitro RH treatment was performed prior to pull-down. (h) cytoDRIP blot showing cytoplasmic hybrid accumulation in PlaB-treated (500nM, 3h) HeLa cells. In vitro RH treatment was performed prior to pull-down.

(i) Western blot showing knockdown efficiency of siRNAs targeting *XPG* and *XPF* in HeLa cells. (j) cytoDRIP blot showing the role of XPG and XPF in cytoplasmic hybrid production after *SETX* or *BRCA1* knockdown in HeLa cells. (k) Left, images of HeLa cells after *SETX* and *XPG* or *XPF* knockdown probed with GFP–dRH protein after fixation, following mock or RH pre-treatment. Scale bar is 10 μm . Right, quantification of cytoplasmic GFP–dRH intensities; p-values are shown; two-sided Mann Whitney U test: n-values from left to right: 611, 573, 659, 686. Centre line, median; box limits, 75 and 25 percentiles, whiskers, min and max values. (l) As in (k) but after *BRCA1* knockdown in HeLa cells. Two-sided Mann Whitney U test: n-values from left to right: 526, 502, 633, 653. Centre line, median; box limits, 75 and 25 percentiles; whiskers, min and max values. (m) Schematic of the XPG auxin-inducible degron (AID) system. (n) Western blots showing XPG degradation after knockdown of *SETX* (left) and *BRCA1* (right) in HCT116 cells. (o) cytoDRIP blot showing cytoplasmic hybrid accumulation after knockdown of *SETX* or *BRCA1* and impact of auxin-induced XPG degradation in HCT116 cells.



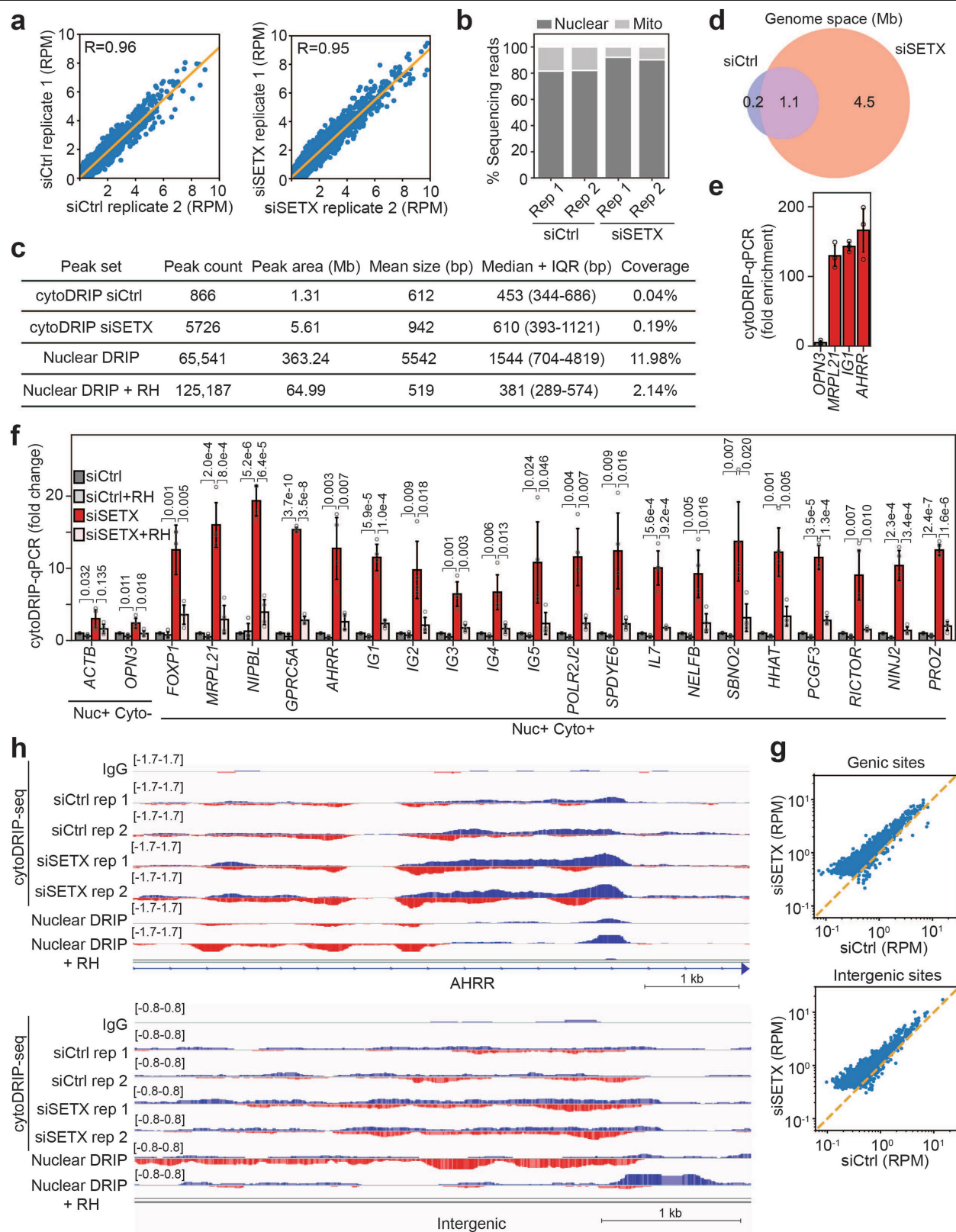
Extended Data Fig. 2 | See next page for caption.

Article

Extended Data Fig. 2 | Dynamics of cytoplasmic hybrid production.

(a) cytoDRIP blot showing cytoplasmic hybrids in mock or PlaB treated (500 nM, 3 h) $BAX^{-/-}BAK^{-/-}$ HeLa cells with or without in vitro RH treatment. (b) Flow cytometry analysis of asynchronous or serum-starved MCF10A cells following incubation with BrdU. Cells were segmented based on DNA content (propidium iodide staining) and BrdU intensity. The percentage of cells in G1, S and G2/M are indicated. At least 50,000 cells were quantified per condition. (c) Western blot showing fractionation of asynchronous (asynch) and serum-starved (starved) MCF10A cells into soluble nuclear and cytoplasmic compartments with Lamin B1 and GAPDH as markers, respectively. (d) RT-qPCR from asynchronous or serum-starved MCF10A cells showing increased unspliced mRNA following PlaB treatment (500 nM, 3 h). Shown is the mean $\pm$ s.d. from three independent biological replicates ($n = 3$), p-values are indicated in the figure; unpaired, two-tailed t -test. (e) As in (c) but for foreskin fibroblasts. (f) Cell cycle quantification from high-content imaging of foreskin fibroblasts after EdU incorporation, using DAPI staining for DNA content. Shown is the mean $\pm$ s.d. from three independent biological replicates ($n = 3$). (g) As in (d) but for foreskin fibroblasts. (h) cytoDRIP blot showing cytoplasmic hybrids extracted from equal numbers of asynchronous or serum-starved foreskin fibroblasts following DMSO or PlaB treatment (500 nM, 3 h), with mock and RH treatment prior to pull-down. (i) As in (b) but for $BAX^{-/-}BAK^{-/-}$ MCF10A cells.

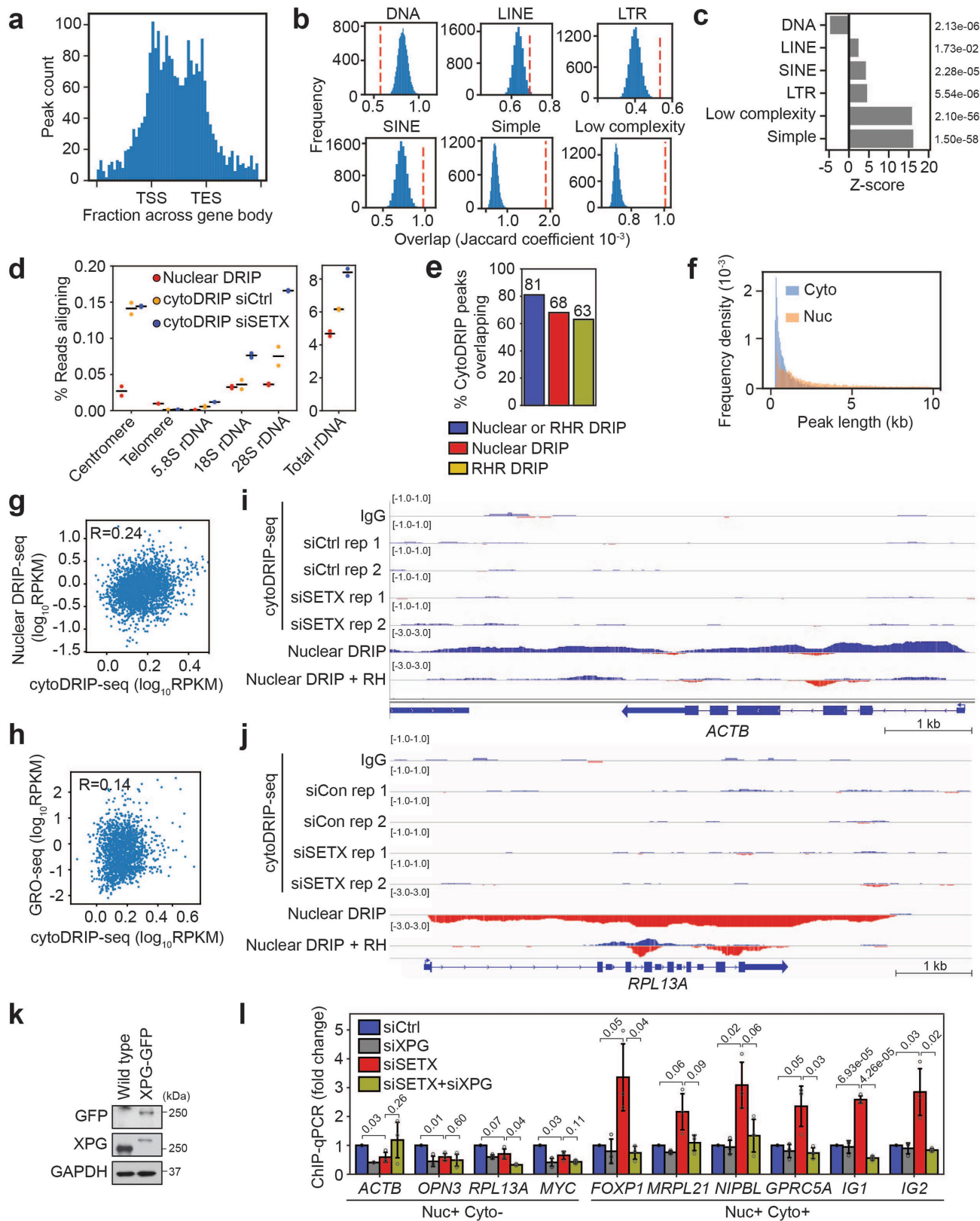
(j) Western blot showing fractionation of asynchronous and serum-starved $BAX^{-/-}BAK^{-/-}$ MCF10A cells into soluble nuclear and cytoplasmic compartments with Lamin B1 and GAPDH as markers, respectively. (k) cytoDRIP blot showing cytoplasmic hybrids in asynchronous and serum-starved $BAX^{-/-}BAK^{-/-}$ MCF10A cells with DMSO or PlaB treatment (500 nM, 3 h). Each sample was treated with RH in vitro to confirm the specificity of the hybrid IP. (l) Left, experimental workflow. Right, blots showing hybrids isolated from the cytoplasm or nucleoplasm of siCtrl or siSETX-treated HeLa cells, with mock or LMB treatment (3 h, 5 nM) prior to harvest. (m) Left, experimental workflow. Middle, western blot as in (c) from HeLa cells treated with vehicle control (DMSO) or PlaB (500 nM, 3 h). Right, blot showing hybrids as in (i) but in HeLa cells treated with LMB (2 h, 5 nM) followed by PlaB + LMB for a further 3 h. (n) Representative images showing cyclin B1 localization in fixed HeLa cells treated with LMB (5 h, 5 nM) or vehicle control (EtOH). Scale bar is 10 μ m. (o) RT-qPCR from HeLa cells showing increased unspliced mRNA following treatment with PlaB (500 nM) for the times indicated. Shown is the mean $\pm$ s.d. from three independent biological replicates ($n = 3$). (p) As in (o) but cells were treated with PlaB (500 nM, 3 h) and then fresh media was added following PlaB withdrawal for the times indicated. Shown is the mean $\pm$ s.d. from four independent biological replicates ($n = 4$).



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Characteristics of cytoDRIP peaks. (a) Scatter plots showing high reproducibility between cytoDRIP-seq siCtrl (left) and siSETX (right) replicates; Pearson's correlation: $R = 0.96$ and 0.95 respectively; $p < 1e-16$ (machine precision limit). (b) Bar blot showing proportion of deduplicated sequencing reads mapping to the nuclear and mitochondrial (mito) genomes in cytoDRIP-seq samples. Data from two biological replicates are shown. (c) Table showing peak characteristics in cytoDRIP-seq, nuclear DRIP-seq, and nuclear DRIP-seq following RH treatment (RHR DRIP). Numbers of peaks (peak count), genomic space covered by peaks (peak area), size of peaks (mean and median), percent of genome covered by peaks (coverage) are shown. IQR is interquartile range. (d) Venn diagram of genome areas (in megabases) occupied by peaks identified in siCtrl and siSETX cytoDRIP-seq samples. (e) Bar plot showing enrichment of cytoplasmic hybrid sites by qPCR after S9.6 pull-down, relative to IgG. *OPN3* was only found in the nucleus; the other sites

were found in the nucleus and cytoplasm. Shown is the mean $\pm$ s.d. from three independent biological replicates ($n = 3$). (f) cytoDRIP-qPCR in HeLa cells after depletion of *SETX* at cytoDRIP-seq sites and nuclear R-loop forming sites. RH treatment was performed in vitro, prior to hybrid pull-down. 'Nuc+ Cyto+' sites were found in the nucleus and cytoplasm, while 'Nuc+ Cyto-' sites were only found in the nucleus. Gene names are shown; *IG1-IG5* are intergenic sites. Shown is the mean $\pm$ s.d. from four independent biological replicates ($n = 4$); p-values are shown in the figure, unpaired two-tailed *t*-test. (g) Scatter plots showing increased cytoDRIP-seq signal upon depletion of *SETX* in genic sites (upper) and intergenic sites (lower). Dashed line represents $x = y$. (h) Genome browser views of genic (top) and intergenic (bottom) cytoDRIP-seq sites. From top to bottom normalized tracks are: IgG, siCtrl (2 replicates), siSETX (2 replicates), nuclear DRIP-seq, nuclear DRIP-seq + RH. Red indicates negative strand signal, blue indicates positive strand signal.

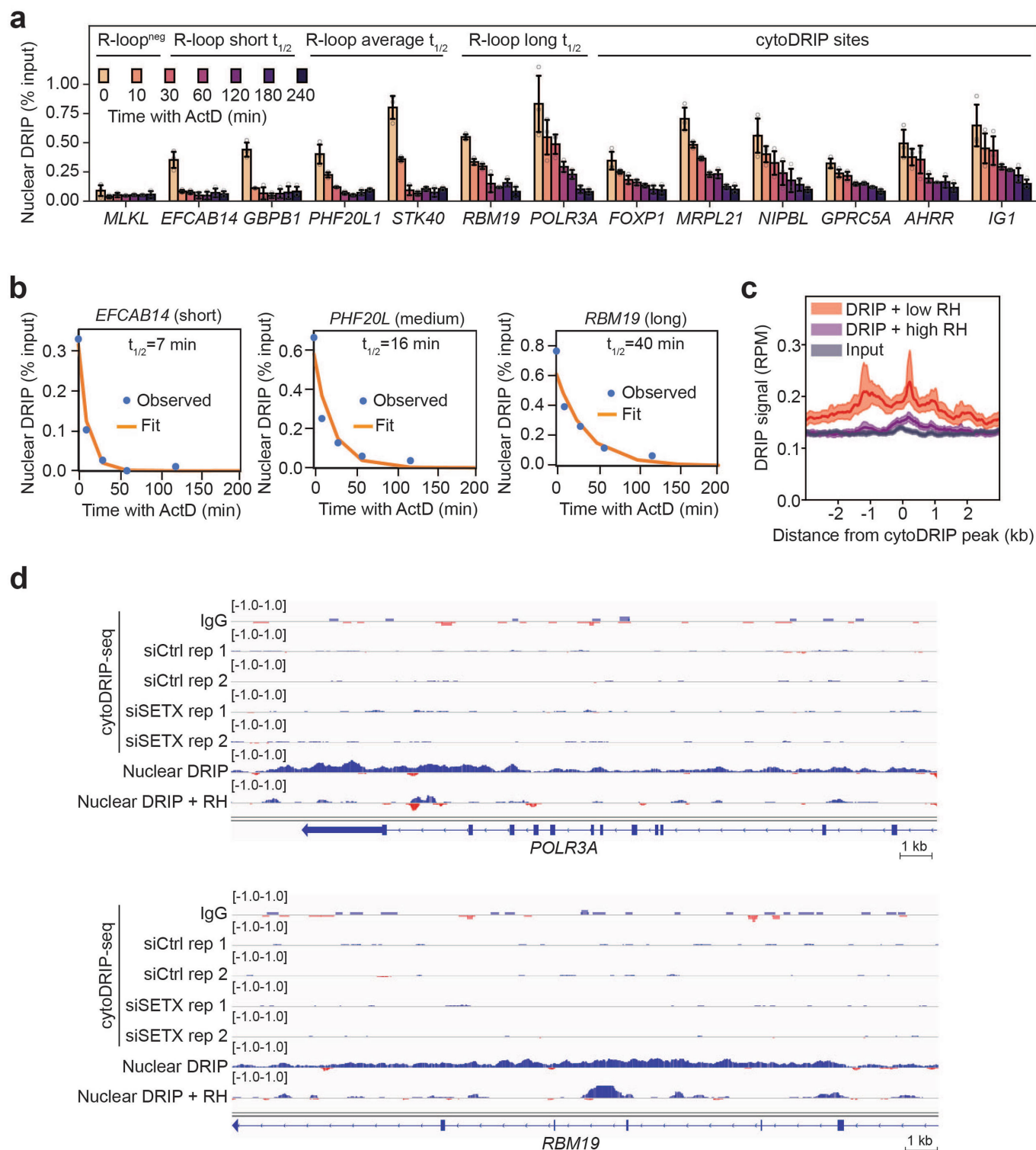


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | cytoDRIP sites map to genic and intergenic regions.

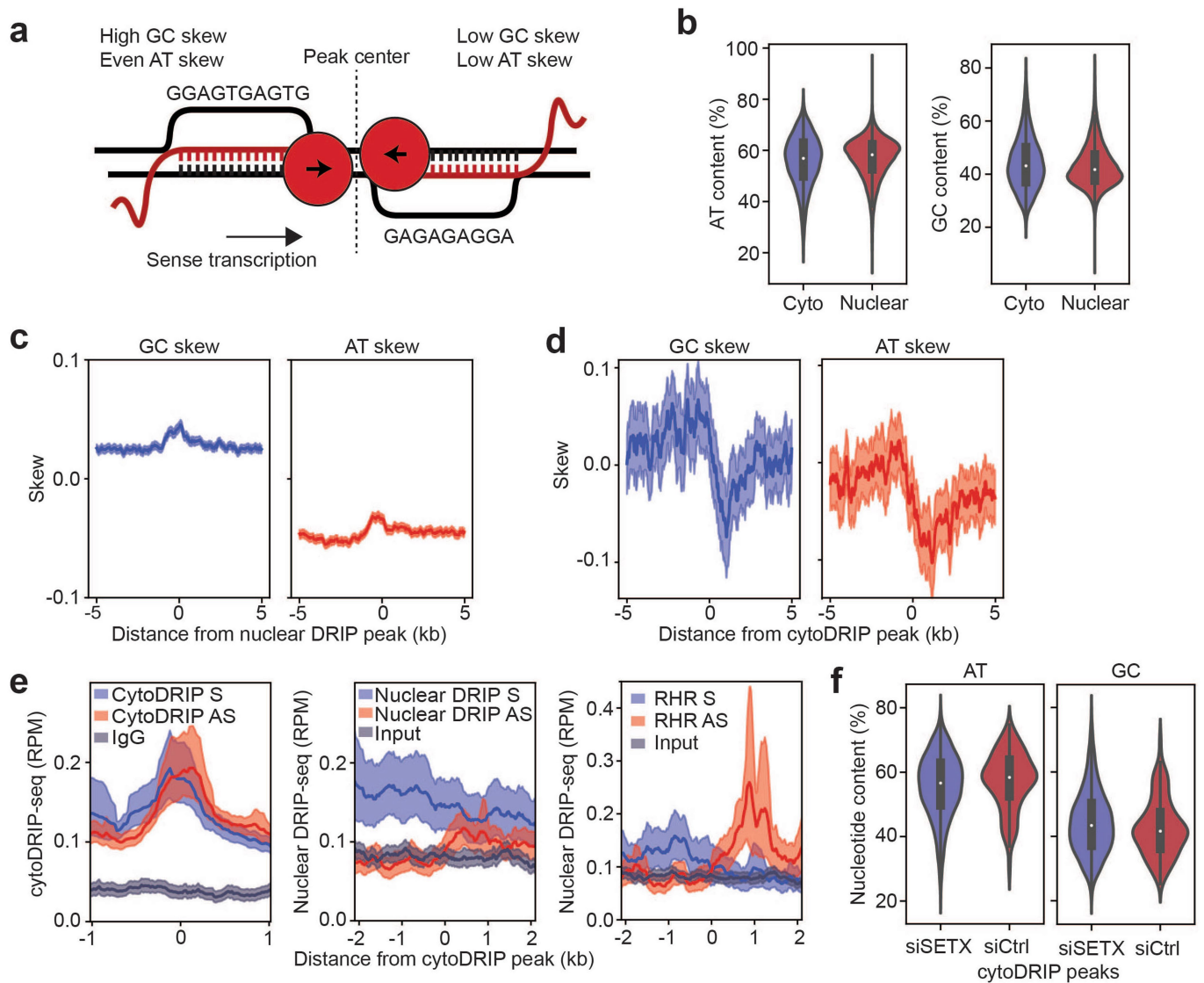
(a) Histogram showing distribution of cytoDRIP peaks over genes. Positions of transcription start site (TSS) and transcription end site (TES) are indicated. (b) Blue histograms show expected overlaps between repeat elements and randomly sampled peak sets (matched in size and number from cytoDRIP peaks) from within all nuclear R-loop peaks. Red dashed line indicates the observed overlap for cytoDRIP peaks. (c) Z-scores for the overlaps calculated in (b); individual p-values are shown on the right. (d) Proportion of nuclear DRIP-seq and cytoDRIP-seq reads aligning to consensus regions for rDNA, alpha satellite for centromeres and telomeric repeats. Data are from two independent biological replicates ($n = 2$) per condition, black lines show the mean. (e) Bar plot showing proportion of cytoDRIP peaks overlapping nuclear DRIP and/or RNase H resistant hybrid (RHR) sites. (f) Histogram of peak lengths comparing cytoDRIP-seq (siSETX condition) (Cyto) and nuclear DRIP-seq (Nuc) peaks. (g) Scatter plot correlating nuclear DRIP-seq signal

at cytoDRIP regions with cytoDRIP-seq signal, Pearson's correlation: $R = 0.24$. (h) Scatter plot correlating nascent transcription by global run-on sequencing at cytoDRIP regions with cytoDRIP-seq signal, Pearson's correlation: $R = 0.14$. (i,j) Genome browser views showing lack of cytoDRIP signal at sites with robust nuclear R-loop formation (i) *ACTB*, (j) *RPL13A*. From top to bottom normalized tracks are: IgG, siCtrl (2 replicates), siSETX (2 replicates), nuclear DRIP-seq, nuclear DRIP-seq + RH. Red indicates negative strand signal, blue indicates positive strand signal. (k) Western blot showing HeLa cells stably expressing GFP-tagged XPG. GAPDH is the loading control. (l) GFP ChIP-qPCR in HeLa cells following knockdown of *SETX* and/or *XPG*, showing GFP-XPG binding at hybrid sites. 'Nuc+ Cyto+' sites were found in the nucleus and cytoplasm, while 'Nuc+ Cyto-' sites were only found in the nucleus. Gene names are shown; *IG1* and *IG2* are intergenic sites. Shown is the mean $\pm$ s.d. from three independent biological replicates ($n = 3$); unpaired two-tailed *t*-test; p-values are shown.



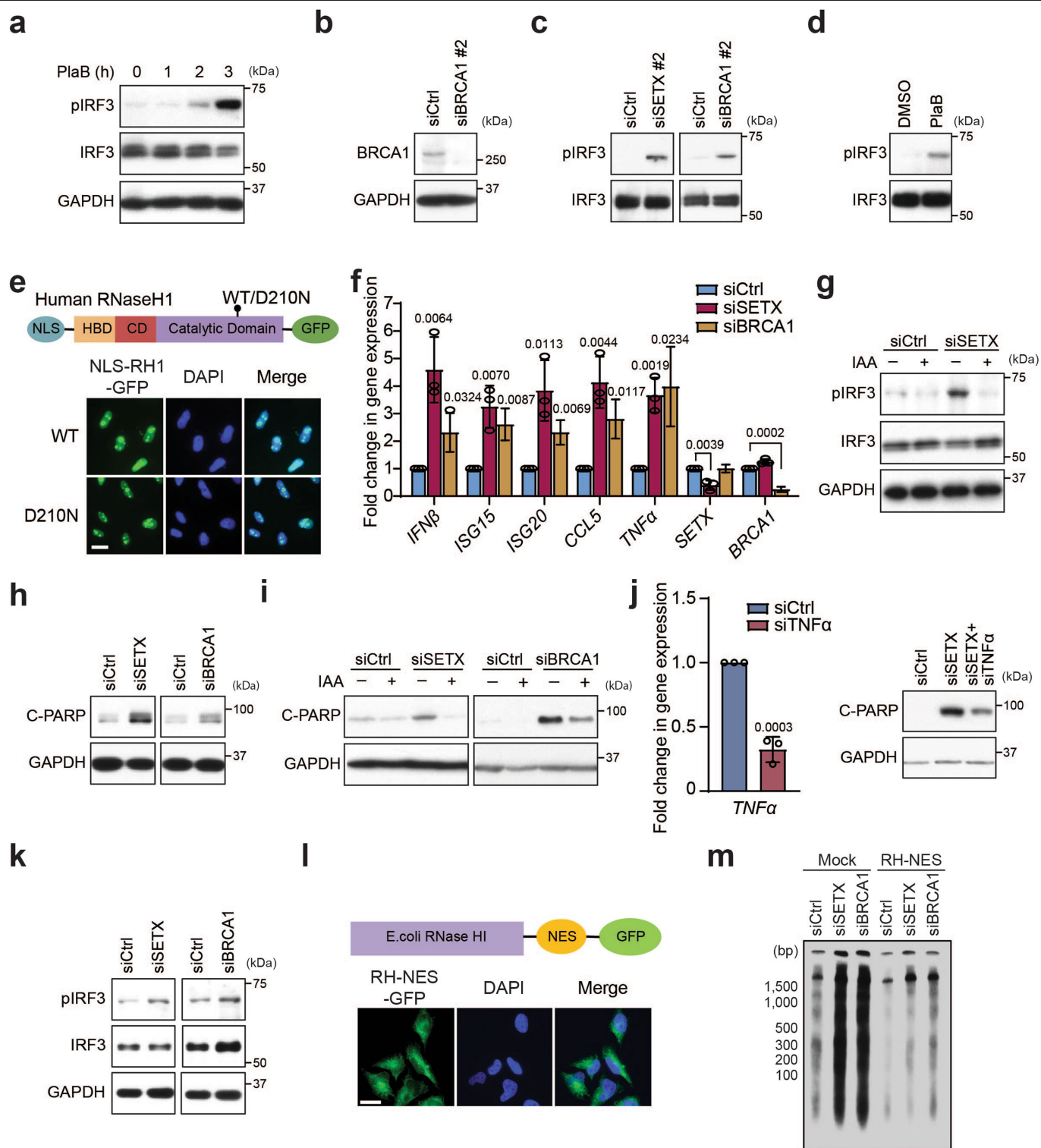
Extended Data Fig. 5 | Cytoplasmic hybrids are derived from long-lived and partially RNase H-resistant nuclear R-loops. (a) Nuclear DRIP-qPCR after actinomycin D treatment. Nuclear R-loop sites with short, average and long half-lives²¹ are indicated, as well as cytoDRIP sites. R-loop^{neg} indicates a nuclear site with low R-loop abundance. Gene names are indicated. *IG1* is an intergenic site. Shown is the mean $\pm$ s.d. from three independent biological replicates ($n = 3$). $P = 9.81e-12$ between nuclear sites with short or medium lifetimes and cytoDRIP sites (two-tailed Mann Whitney U test). (b) Nuclear DRIP-qPCR after actinomycin D treatment showing example fits of exponential decay to derive

RNA-DNA hybrid half-lives. Shown is the mean from three independent biological replicates ($n = 3$). (c) Aggregate plots around cytoDRIP regions showing nuclear DRIP-seq signal following low (red) or high (purple) RH treatment in vitro. Input signal is grey. Each line is the mean of 1762 genic peaks ($n = 1762$). Error bands represent 95% CI of the mean. (d) Genome browser views of previously identified²¹ long-lived nuclear R-loop sites. From top to bottom normalized tracks are: IgG, siCtrl (2 replicates), siSETX (2 replicates), nuclear DRIP-seq, nuclear DRIP-seq + in vitro RH. Red indicates negative strand signal, blue indicates positive strand signal.



Extended Data Fig. 6 | Cytoplasmic hybrids are characterized by switches in nucleotide skew. (a) Model of converging transcription at cytoDRIP peaks. RNA-DNA hybrids form as a result of sense and antisense transcription in regions of high GC-skew on the non-template strand. Example nucleotide sequences that fit the observed skew pattern are shown. (b) Violin plots of AT content (left) and GC content (right) for cytoDRIP ($n = 2911$) and nuclear DRIP ($n = 65,541$) regions, two-tailed Mann Whitney U between cytoDRIP and nuclear DRIP regions: $p = 3.6 \times 10^{-8}$ (left), $p = 3.5 \times 10^{-16}$ (right). Centre line, median; box limits, 75 and 25 percentiles, whiskers, min and max values. (c) Aggregate plots around genic nuclear DRIP regions ($n = 56,433$) showing GC skew (left) and AT skew (right). Error bands represent 95% CI of the mean. (d) Aggregate plots around

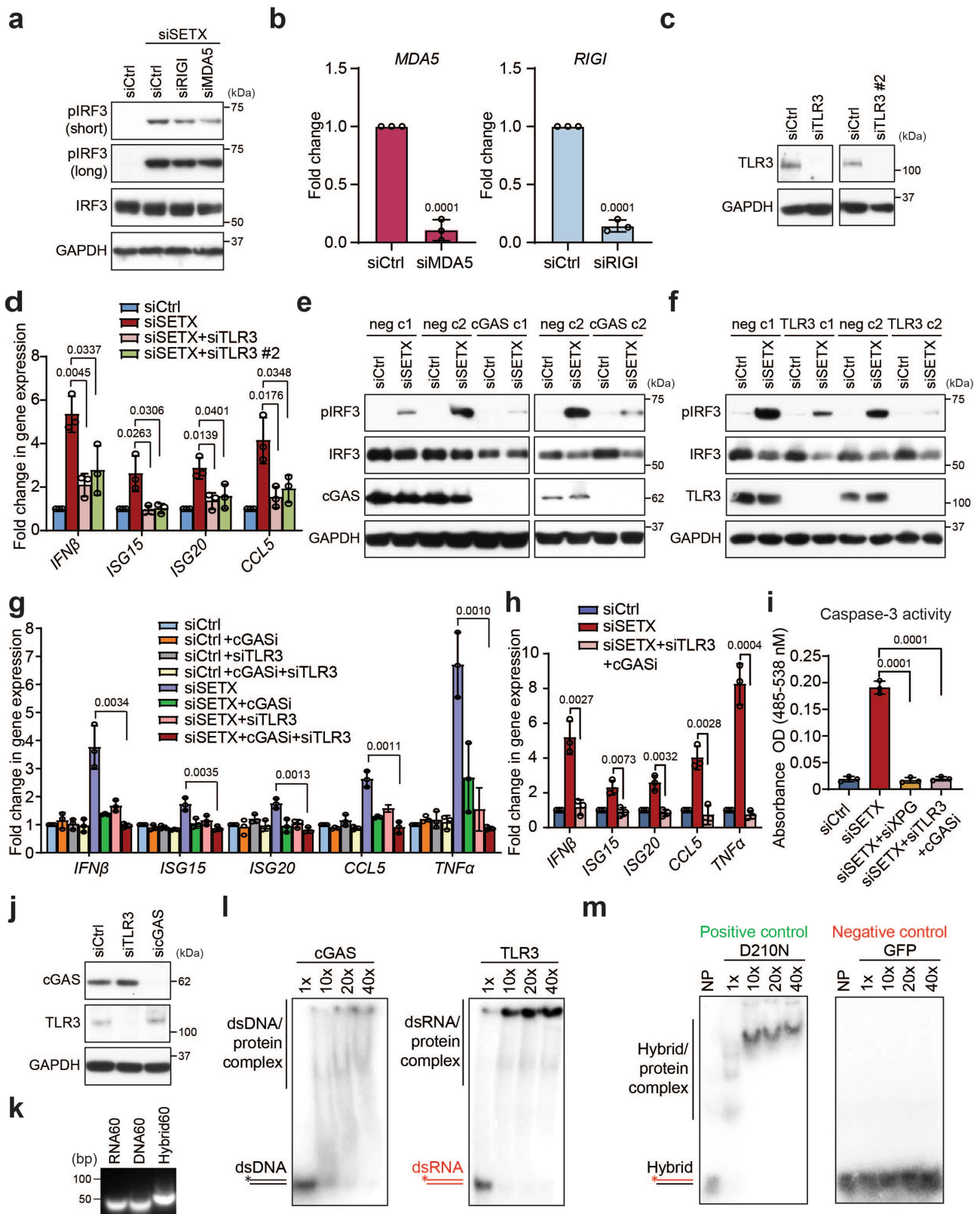
genic siCtrl cytoDRIP regions showing GC skew (left) and AT skew (right). Means of 282 peaks ($n = 282$) are shown; error bands represent 95% CI of the mean. (e) Aggregate plots around genic siCtrl cytoDRIP regions showing cytoDRIP-seq signal (left), nuclear DRIP-seq signal (middle) and nuclear RHR signal (right). Means of 282 peaks ($n = 282$) are shown; error bands represent 95% CI of the mean. (f) Violin plots of AT content (left) and GC content (right) for siSETX ($n = 2629$) and siCtrl ($n = 282$) cytoDRIP regions, two-tailed Mann Whitney U between cytoDRIP and nuclear DRIP regions: $p = 0.003$ (left), $p = 0.003$ (right). Centre line, median; box limits, 75 and 25 percentiles, whiskers, min and max values.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Different perturbations inducing cellular R-loops trigger IRF3 signalling and apoptosis. (a) Western blot of pIRF3 upon PlaB treatment (500 nM) in HeLa cells. GAPDH is the loading control. (b) Western blot showing knockdown efficiency of a second siRNA to *BRCA1*. GAPDH serves as the loading control. (c) Western blot showing pIRF3 levels upon knockdown of *SETX* or *BRCA1* using a second siRNA in HeLa cells. (d) Effect of PlaB treatment (500 nM, 3 h) on pIRF3 in HCT116 cells. (e) Top: Schematic of nuclear localization signal (NLS)-tagged wild-type (WT) or catalytically-inactive (D210N) RNase H1. HBD = hybrid binding domain, CD = connection domain. Bottom: cellular localization of GFP-tagged NLS-RNase H1 WT/D210N. Scale bar, 20 μ m. (f) RT-qPCR measurements of IRF3 effectors upon knockdown of *SETX* or *BRCA1* in MCF10A cells. (g) Western blot of pIRF3 after auxin-induced XPG degradation and *SETX* knockdown in HCT116 cells. GAPDH is the loading control. (h) Western blot showing C-PARP levels upon knockdown of *SETX* or

BRCA1 in MCF10A cells. (i) Western blot showing the impact of auxin-induced XPG degradation on C-PARP in siSETX- or siBRCA1-treated HCT116 cells. The same GAPDH blot, which is the loading control, is used in Extended Data Fig. 1n. (j) Left, RT-qPCR showing the knockdown efficiency of *TNF α* in HeLa cells. Right, western blots showing levels of C-PARP upon knockdown of *TNF α* in siSETX-treated HeLa cells. (k) Western blot showing levels of pIRF3 upon knockdown of *SETX* or *BRCA1* in *BAX*^{-/-}*BAK*^{-/-} HeLa cells. (l) Top: Schematic of nuclear export signal (NES)-tagged RNase H1. Bottom: cellular localization of GFP-tagged RNase H1-NES in HeLa cells. Scale bar, 20 μ m. (m) cytoDRIP blot showing cytoplasmic hybrids upon knockdown of *SETX* or *BRCA1* in mock-treated HeLa cells and HeLa cells stably expressing NES-tagged RNase H1. Bar graphs are mean $\pm$ s.d. from 3 independent biological replicates (n = 3) (unpaired two-tailed t-test with CI = 95%). P values are shown at the top of the graphs.

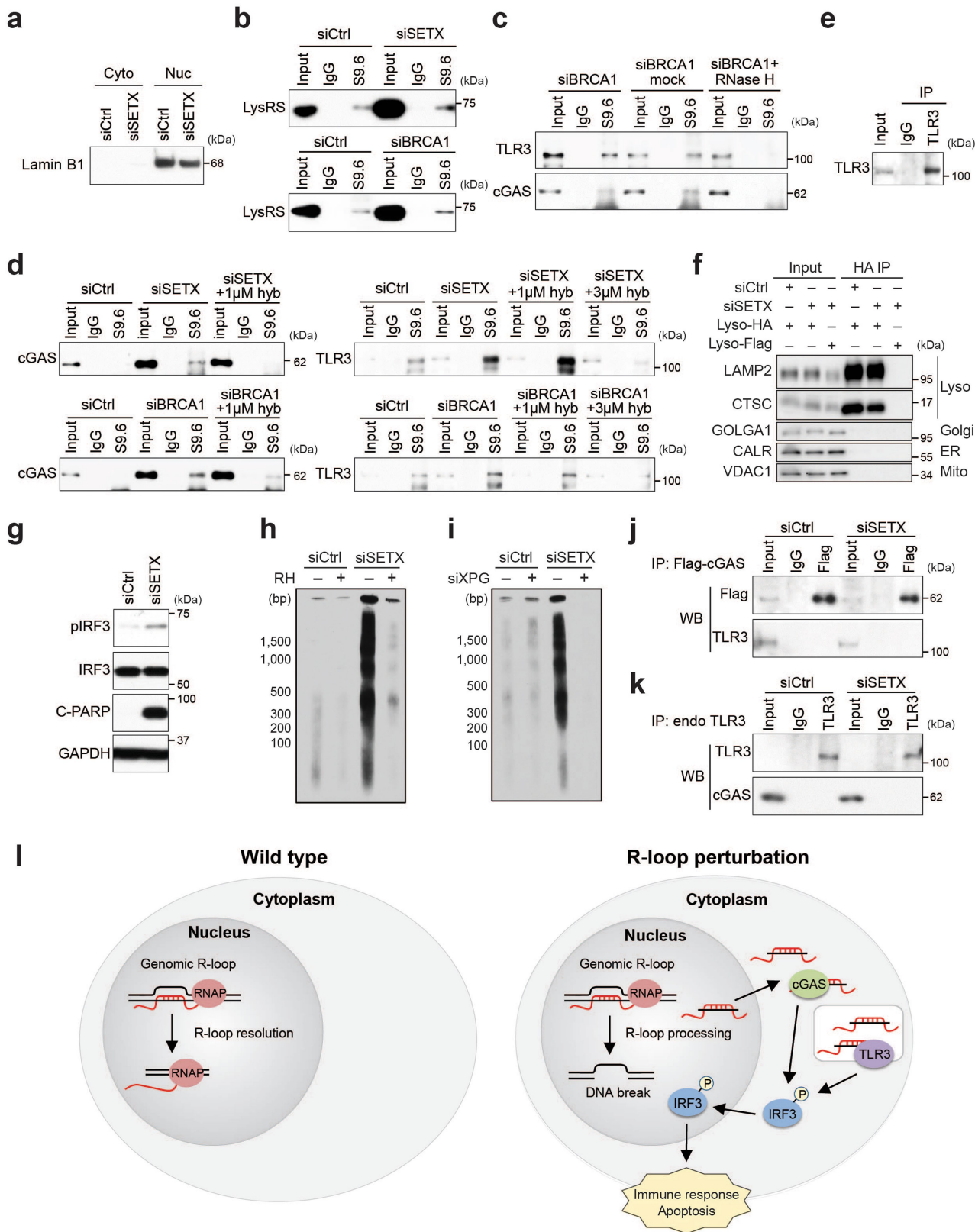


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | cGAS and TLR3 cooperate to activate IRF3 signalling.

(a) Western blot showing pIRF3 levels upon siRNA-mediated knockdown of *SETX* and either *RIGI* or *MDA5* in HeLa cells. GAPDH is the loading control. (b) RT-qPCR showing the knockdown efficiency of *RIGI* and *MDA5* in HeLa cells. (c) Western blot showing the knockdown efficiency of two different TLR3 siRNAs in HeLa cells. (d) RT-qPCR measurements of IRF3 effectors upon *TLR3* knockdown with two different siRNAs in siSETX-treated *BAX*^{-/-} *BAK*^{-/-} HeLa cells. (e) and (f) Western blot showing levels of pIRF3 in two negative control (neg) clones and either *cGAS* knockout clones (e) or *TLR3* knockout clones (f) generated using the CRISPR-Cas9 system in HeLa cells. c1=clone 1, c2=clone 2. GAPDH serves as the loading control. (g) RT-qPCR measurements of IRF3 effectors upon single or combined inhibition/knockdown of *cGAS* and *TLR3* in control and siSETX-treated HeLa cells. (h) As in (g) but in *BAX*^{-/-} *BAK*^{-/-} HeLa

cells. (i) Caspase 3 activity assay after knockdown of *SETX* and either *XPG* knockdown or the combination of *cGAS* inhibition and *TLR3* knockdown. (j) *cGAS* and TLR3 protein levels upon siRNA-mediated knockdown of *TLR3* or *cGAS* in HeLa cells. (k) Agarose gel showing DNA (60 nt) and RNA (60 nt) oligonucleotides can anneal to form a DNA-RNA hybrid. (l) Gel shift assays of *cGAS* binding to double-stranded DNA (dsDNA) (left) and TLR3 binding to double-stranded RNA (dsRNA) (right). (m) Gel shift assays show binding of human RNaseH1 D210N catalytically-inactive mutant and GFP protein to RNA-DNA hybrids which are used as positive and negative controls, respectively. NP stands for no protein. Bar graphs are mean ± s.d. from three independent biological replicates (n = 3) (unpaired two-tailed t-test with CI = 95%). P values are shown at the top of the graphs.

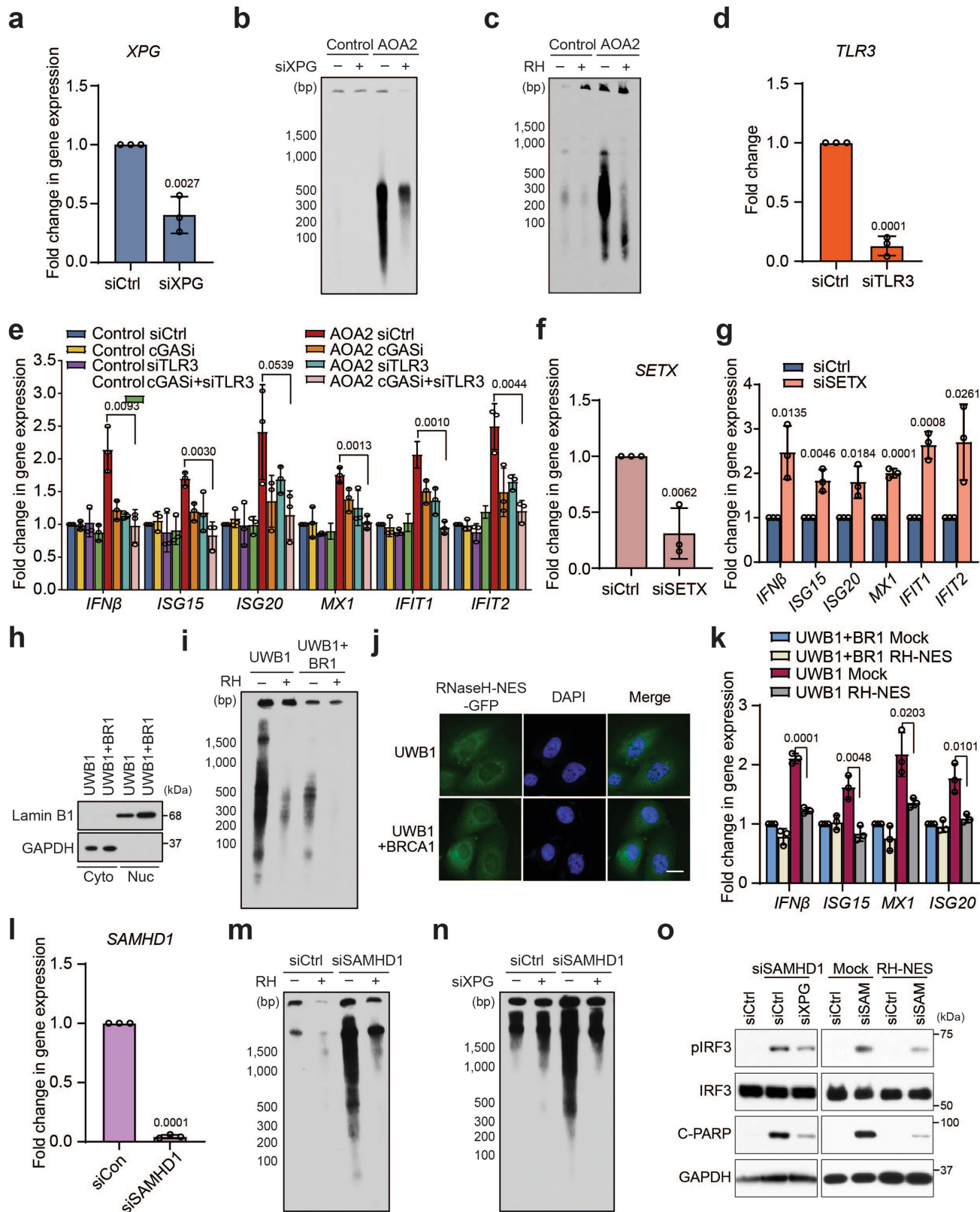


Extended Data Fig. 9 | See next page for caption.

Article

Extended Data Fig. 9 | cGAS and TLR3 bind directly to cytoplasmic RNA–DNA hybrids. (a) The purity of the cytoplasmic fraction used for the S9.6 co-IP was assessed by western blot. (b) S9.6 co-IP from the cytoplasmic fraction showing LysRS binds to cytoplasmic hybrids in our methods. LysRS has been reported to interact with cytoplasmic hybrids and serves as the positive control. (c) S9.6 co-IP from the cytoplasmic fraction showing cGAS and TLR3 associate with RNA–DNA hybrids isolated from siBRCA1-treated cells, as well as the impact of 37 °C no enzyme mock control and in vitro RNase H treatment before the IP step. RNase H treatment, 50 U ml⁻¹ for 1 h at 37 °C. (d) S9.6 co-IP from cytoplasmic fraction showing cGAS binding to hybrids induced by siSETX is disrupted by 1 μM hybrid competitor in IP reaction, and TLR3 binding to hybrids is disrupted by 3 μM hybrid competitor in an IP reaction. hyb = hybrid. (e) Western blot validating the TLR3 IP efficiency in experiments to detect TLR3-associated cytoplasmic hybrids by performing TLR3 IP followed by S9.6 IP (Fig. 4h). (f) Western blot assessing the purity of the endolysosomal fraction after isolation following HA immunoprecipitation in control or *SETX*-depleted HA-TMEM192 HEK293T cells. Flag-TMEM192 HEK293T cells were used as a negative control for the LysoIP. Proteins marking the lysosome (Lyso), Golgi apparatus (Golgi), endoplasmic reticulum (ER) and mitochondria (Mito) are indicated. (g) Western blot showing pIRF3 and C-PARP levels induced by *SETX* knockdown

in HA-TMEM192 HEK293T cells, as was observed in HeLa cells. This result suggests this cell line is suitable for the study of R-loop-induced immune activation. This experiment is a control for the LysoIP (Fig. 4i). (h) cytoDRIP blot showing cytoplasmic hybrids levels are elevated upon knockdown of *SETX* in HA-TMEM192 HEK293T cells. In vitro RNase H digestion was used to ensure IP specificity. This experiment is also a control for the LysoIP. (i) cytoDRIP blot showing cytoplasmic hybrids upon knockdown of *SETX* in HA-TMEM192 HEK293T cells with or without knockdown of *XPG*. (j) co-IP testing the interaction between Flag-tagged cGAS and endogenous TLR3. (k) co-IP testing the interaction between endogenous TLR3 and cGAS. (l) Working model. Left: in wild-type cells, nuclear R-loops are efficiently resolved by RNase H or RNA–DNA helicases, such as SETX. Only a small number of R-loops are processed by XPG and converted to cytoplasmic hybrids, so that cytoplasmic hybrid levels are below the threshold required for activation of IRF3 signalling. Right: under certain perturbations, including depletion of SETX/BRCA1, or under pathological conditions that deregulate R-loops, a subset of nuclear R-loops that may not be efficiently resolved are processed by XPG, leading to RNA–DNA hybrid accumulation in the cytoplasm. These hybrids are then recognized by cGAS and TLR3 in the cytosol and endolysosome, activating IRF3-mediated immune signalling and apoptosis.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | R-loop-induced cytoplasmic RNA–DNA hybrid accumulation and innate immune activation in patient-derived disease cell models. (a) RT-qPCR showing the XPG siRNA knockdown efficiency in AOA2 patient-derived fibroblasts. (b) cytoDRIP blot showing cytoplasmic hybrids in control and AOA2 patient-derived fibroblasts with or without knockdown of *XPG*. (c) cytoDRIP blot showing cytoplasmic hybrids in control and AOA2 patient-derived fibroblasts with or without in vitro RNase H treatment prior to hybrid IP. (d) RT-qPCR showing the knockdown efficiency of *TLR3* in AOA2 patient-derived fibroblasts. (e) RT-qPCR measurements of immune effectors upon single or combined inhibition and knockdown of cGAS and *TLR3*, respectively, in control and AOA2 patient-derived fibroblasts. (f) RT-qPCR showing the SETX siRNA knockdown efficiency in control fibroblasts. (g) RT-qPCR measurements of *IFN β* and ISGs upon knockdown of *SETX* in control fibroblasts. (h) Western blot showing the fractionation of UWB1.289 and UWB1.289+BRCA1 cells into soluble nuclear and cytoplasmic compartments with Lamin B1 and GAPDH as markers, respectively.

(i) cytoDRIP blot showing cytoplasmic hybrids in UWB1.289 and UWB1.289+BRCA1 cells with or without in vitro RNaseH treatment prior to hybrid IP. (j) Cellular localization of GFP-tagged NES-RNaseH1 in UWB1.289 and UWB1.289+BRCA1 cells. Scale bar, 20 μ m. (k) RT-qPCR measurements of immune effectors in UWB1.289 and UWB1.289+BRCA1 cells stably expressing GFP (mock) or NES-tagged RH (RH-NES). (l) RT-qPCR showing the SAMHD1 siRNA knockdown efficiency in HeLa cells. (m) cytoDRIP blot showing cytoplasmic hybrids in control and SAMHD1-deficient HeLa cells with or without in vitro RNase H treatment. (n) cytoDRIP blot showing cytoplasmic hybrids in control and SAMHD1-deficient HeLa cells with or without *XPG* knockdown. (o) Left: western blots showing levels of pIRF3 and C-PARP upon knockdown of *XPG* in siSAMHD1-treated HeLa cells. Right: western blots showing pIRF3 and C-PARP level upon knockdown of *SAMHD1* in mock and RH-NES HeLa stable cell lines. siSAM = siSAMHD1. Bar graphs are mean $\pm$ s.d. from three independent biological replicates (n = 3) (unpaired, two-tailed t-test with CI = 95%). P values are shown at the top of the graphs.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Sonication was performed on a Covaris machine (E220 evolution) using SonoLab 7.3 software. A Zeiss LSM800 confocal microscope equipped with an AxioImager.Z2 and a Zeiss OBSERVER.Z1 INVERTED microscope which use ZEN (ZEISS Efficient Navigation) 2.6 blue edition software were used. Real Time quantitative PCR data was collected on Roche LightCycler480 running on Roche Light Cycler version 1.5.1 Microscopy images were acquired with Nikon Imaging Software NIS Elements. Images of Radioactivity experiments were collected by Typhoon 9410.
Data analysis	Data analysis was performed using publicly available software packages: cutadapt v1.16, bowtie2 (v2.3.4), BEDTools (v2.29.2), SAMtools (v1.10), deepTools (v3.2.1), unix text-processing utilities. Peaks were called against a merged bam file from all IgG samples using MACS2 with narrow peak settings. Aggregate plots around cytoDRIP peaks and other genome features were produced from genome browser tracks with deepTools. Genome alignment files were viewed using IGV Genome Browser. Analysis of genome compartments overlapping cytoDRIP peaks was performed using CEAS (Cis-Regulatory Element Annotation System). For intergenic enhancer and insulator annotations, ChromHMM annotations for HeLa cells (from UCSC Table Browser) were intersected with intergenic cytoDRIP peaks using bedtools. Data processing for all genomic plots was performed with Python v3.7.13, NumPy v1.21.5 and Pandas v1.3.5. Data were visualized with the Python packages Matplotlib v3.5.1 and Seaborn v0.11.2. Statistical analysis was primarily performed in Python using SciPy v1.7.3. Microsoft Excel (v16.50) was used to fit exponential decay functions. GraphPad Prism (v9.1.0) was also used to display data and perform statistical tests. Custom code will be provided upon request. Microscopy images were exported as 16-bit, all cropped and adjusted equally in ImageJ (version 2.0.0). Image analysis was performed using CellProfiler (version 4.2.1). The DAPI channel was used to identify nuclei using the IdentifyPrimaryObjects module, with manual thresholding.

The Primary Objects and whole cell stain were then used to identify cells as Secondary Objects. The cytoplasmic area was identified as a Tertiary Object from the whole cells shrunk by one pixel and nuclei expanded by three pixels. The mean intensity for each cytoplasmic area was calculated and exported.

The following software was used for data analysis:
 ImageJ version 2.0.0 for immunofluorescence images
 Roche Light Cycler version 1.5.1 for gene expression analysis after real-time qPCR
 Figures assembled with Adobe Illustrator CS6.
 GraphPad Prism v9.3.1 for general statistical analysis and graphing.
 FlowJo v3.05 for flow cytometry analysis

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All sequencing data generated in this work have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE178841. For nuclear DRIP-seq, datasets under accession number GSE134084 were used.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size

No statistical methods were used to determine the sample size. The sample size follows common standards ($n = 3$ or more biological replicates for qPCR, blots and flow cytometry) and is reported in legends for main and Extended figures. Imaging experiments included at least 100 cells/sample. Two repeats of each sequencing experiment were performed and are indicated in the text. This sample size selection is consistent with similar publications in the field (Coquel et al., PMID: 29670289 and Harding et al., 28759889, Wu et al, PMID: 33767446) and also based on our previous experience (Crossley, MP et al, PMID: 32544226.; Crossley, et al PMID: 34232287).

Data exclusions

No data were excluded from analysis.

Replication	For most of our figures, at least three independent experiments were performed and have not encountered discrepancy between results and conclusions. We have indicated the sample number in the figure legends/figures and last paragraph of the methods section.
Randomization	For immunofluorescent staining, fields of view were randomly selected. For other experiments where samples were not allocated randomly, group allocation and randomization were unnecessary because all samples were measured independently in the same way in an internally controlled manner.
Blinding	Blinding was not feasible during this study. Knowledge of cell line identify was required due to different culture conditions required. Perturbations resulted in differences in morphology and viability which were highly telling.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used

Antibodies (Species - Rb/rabbit; Ms/mouse), catalog number, company and dilution used for western blotting (WB), immunoprecipitation (IP) or other noted method

XPG (Rb); A301-484A; Bethyl; WB: 1:1000
 SETX (Rb); NB100-57543; Novus Biologicals; WB: 1:1000
 BRCA1 (Ms); sc6954; Santa Cruz; WB: 1:500
 IRF3 (Rb); ab76409; Abcam; WB: 1:1000
 pIRF3 (Rb); ab76493; Abcam; WB: 1:1000
 GAPDH (Ms); ab8245; Abcam; WB: 1:10,000
 TLR3 (Rb; Western); Ab62566; Abcam; WB: 1:1000
 TLR3 (Ms; Western); MAB1487(Clon # 512505); R&D Systems; WB: 1:1000
 TLR3 (Rb; IP); A11778; Abclonal; co-IP: 7.5ug, IP-cytoDRIP: 15ug
 cGAS (Rb); 15102S; Cell Signalling Technology; WB: 1:1000
 C-PARP (Rb); #5625; Cell Signalling Technology; WB: 1:1000
 Lamin B1 (Rb); ab133741; Abcam; WB: 1:5000
 S9.6 (Ms); Derived from ATCC hybridoma cell line HB-8730 (purified by Antibodies Incorporated); cytoDRIP: 16ug, co-IP: 20ug
 Mouse IgG; 12-371; Millipore; cytoDRIP: 16ug
 Cathepsin C (Ms); sc-74590; Santa Cruz; WB: 1:500
 Golgin-97 (Rb); 13192; Cell Signaling Technology; WB 1:1000
 Calreticulin (Rb); 12238; Cell Signaling Technology; WB: 1:3000
 VDAC (Rb); 4661; Cell Signaling Technology; WB: 1:1000
 LysRS (Rb); A300-630A-M; Bethyl Laboratories; WB: 1:1000
 FLAG M2 (Ms); F1804; Millipore-Sigma; WB: 1:1000
 CRISPR-Cas9 (Ms); ab191468; Abcam; WB: 1:1000
 GFP (Rb); ab290; Abcam; ChIP: 7.5ug, WB: 1:1000
 Cyclin B1 (Ms); sc-245; Santa Cruz; IF: 1:500
 Alexa Fluor 488 (Goat anti-mouse, IgG H+L); A11001; Invitrogen; IF: 1:1000
 BrdU (Ms); Clone B44; 347580; BD Biosciences; Flow: 1:50
 Goat anti-Rabbit IgG (H+L) HRP; G-21234; Thermo Fisher Scientific; WB: 1:5000
 Goat anti-Mouse IgG (H+L) HRP; 31430; Thermo Fisher Scientific; WB: 1:5000

Validation

All antibodies used in this study were acquired from commercial sources and were validated for specificity and species reactivity by the manufacturer. The information are readily available on the website of the manufacturer listed below. Additional validation is as follows: SETX, BRCA1, IRF3, pIRF3, XPG, cGAS and TLR3 antibodies were validated by siRNA-mediated knockdown. TLR3 and cGAS antibodies were additionally validated by CRISPR knockout approach. XPG antibody was further validated in XPG AID-degron cells. pIRF3 and C-PARP show increase of signal when cells are treated with Poly I:C to induce immune response and apoptosis. The specificity of S9.6 antibody is validated by RNase H digestion.

Manufacturer's Information:

XPG: <https://www.thermofisher.com/antibody/product/ERCC5-XPG-Antibody-Polyclonal/A301-484A>
 SETX: https://www.novusbio.com/products/senataxin-antibody_nb100-57543

BRCA1: <https://www.scbt.com/p/brca1-antibody-d-9>
 IRF3: <https://www.abcam.com/irf3-antibody-ep2419y-ab76409.html>
 pIRF3: <https://www.abcam.com/irf3-phospho-s386-antibody-epr2346-ab76493.html>
 GAPDH: <https://www.abcam.com/gapdh-antibody-6c5-loading-control-ab8245.html>
 TLR3 (Rb, Western): <https://www.abcam.com/tlr3-antibody-ab62566.html>
 TLR3 (Ms, Western): https://www.rndsystems.com/products/human-tlr3-antibody-512505_mab1487
 TLR3 (Rb; IP): <https://abclonal.com/catalog-antibodies/TLR3RabbitpAb/A11778>
 cGAS: <https://www.cellsignal.com/products/primary-antibodies/cgas-d1d3g-rabbit-mab/15102>
 C-PARP: <https://www.cellsignal.com/products/primary-antibodies/cleaved-parp-asp214-d64e10-xp-rabbit-mab/5625>
 Lamin B1: <https://www.abcam.com/lamin-b1-antibody-epr8985b-ab133741.html>
 Mouse IgG: https://www.emdmillipore.com/US/en/product/Normal-Mouse-IgG-MM_NF-12-371
 Cathepsin C: <https://www.scbt.com/p/cathepsin-c-antibody-d-6>
 GOLGIN-97: https://www.cellsignal.com/products/primary-antibodies/golgin-97-d8p2k-rabbit-mab/13192?site-search-type=Products&N=4294956287&Ntt=golgin-97&fromPage=plp&_requestid=789344
 Calreticulin: <https://www.cellsignal.com/products/primary-antibodies/calreticulin-d3e6-xp-rabbit-mab/12238>
 VDAC1: <https://www.cellsignal.com/products/primary-antibodies/vdac-d73d12-rabbit-mab/4661>
 LysRS: <https://www.fortislife.com/products/primary-antibodies/rabbit-anti-kars-antibody/BETHYL-A300-630>
 FLAG M2: <https://www.sigmaaldrich.com/US/en/product/sigma/f1804>
 CRISPR-Cas9: <https://www.abcam.com/crispr-cas9-antibody-7a9-3a3-ab191468.html>
 GFP: <https://www.abcam.com/gfp-antibody-ab290.html>
 Cyclin B1: <https://www.scbt.com/p/cyclin-b1-antibody-gns1>
 Alexa Fluor 488 (Goat anti-mouse, IgG H+L): <https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11001>
 BrdU: <https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/clinical-discovery-research/single-color-antibodies-ruo-gmp/purified-mouse-anti-brdu.347580>
 Goat anti-Rabbit IgG (H+L) HRP: <https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/G-21234>
 Goat anti-Mouse IgG (H+L) HRP: <https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgG-H-L-Secondary-Antibody-Polyclonal/31430>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

HeLa, HCT116, HEK293T and MCF10A were obtained from ATCC. Control normal foreskin (CTRL) fibroblasts and AOA2 patient-derived (SETX-1RM) fibroblasts were gifts from Stephen West's lab. UWB1.289 (UWB1) and UWB1.289+ BRCA1 (UWB1+BR1) reconstituted cells were gifts from the Greenberg's lab. Flag- and HA-tagged TMEM192 293T cells used for LysoLP were given by Abu-Remaileh's lab.

Authentication

STR profiling is used to authenticate the cell lines that were purchased from ATCC. AOA2 cell lines were subjected to STR profiling. UWB1.289 cells and the complemented line were verified by testing for loss of BRCA1 expression in UWB1.289 cells and ectopic BRCA1 expression in the reconstituted line. Flag- and HA-tagged TMEM192 293T cells were authenticated by STR profiling and western blotting (Abu-Remaileh et al., PMID: 29074583)

Mycoplasma contamination

Cell lines were tested regulatory and were negative for mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified lines were used.

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

To monitor S-phase progression, cells were pulse-labeled with 10 μ M BrdU for 30 min in pre-warmed media, trypsinized and washed twice with PBS. After fixing samples with ice-cold 70% ethanol, DNA was denatured with 2N HCl and 0.5% Triton X-100 for 30 min at room temperature. Cells were washed twice with PBS and then incubated in staining solution (1% BSA, 0.5% Tween20 in PBS) and BrdU antibody for 60 min at room temperature with rotation. Supernatant was removed after centrifugation and then cells were incubated in staining solution and secondary antibody at room temperature and shielded from light for 60 min. Cells were washed once with PBS and then resuspended in propidium iodide (10 μ g/mL), RNase A at (200 μ g/mL) in 1X PBS for 45 min at room temperature. Cells were analyzed on a BD Accuri C6 Plus machine. Cell cycle

	profiles were determined using FlowJo software.
Instrument	BD Accuri C6 Plus
Software	FlowJo v3.05
Cell population abundance	Over 100,000 cells were counted for each sample Live cells were determined by FSC/SSC plots. For Extended Data Fig. 2b 81.4% (asynch) and 85.9% (synch) cells were considered viable. For Extended Data Figure 2i 69.9% (asynch) and 70.2% (synch) were considered viable.
Gating strategy	Cells were separated into G1, S and G2/M phases. Cells were gated using BrdU-FITC and 2n and 4n DNA content. For Extended Data Fig. 2b, S-phase cells were those with BrdU-FITC above 90,000. G1 cells had BrdU FITC below 90,000 and DAPI-PI below 325,000. G2/M cells had BrdU FITC below 90,000 and DAPI-PI above 360,000. For Extended Data Fig. 2i, S-phase cells were those with BrdU-FITC above 150,000. G1 cells had BrdU FITC below 150,000 and DAPI-PI below 390,000. G2/M cells had BrdU FITC below 150,000 and DAPI-PI above 430,000.

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