Article

RAD52 resolves transcription-replication conflicts to mitigate R-loop induced genome instability

Manisha Jalan [1] ✉, Aman Sharma [1], Xin Pei [1], Nils Weinhold[1], Erika S. Buechelmaier[1], Yingjie Zhu [2], Sana Ahmed-Seghir[1], Abhirami Ratnakumar[1], Melody Di Bona [1,3], Niamh McDermott[1], Joan Gomez-Aguilar[1], Kyrie S. Anderson[1], Charlotte K. Y. Ng[4,5], Pier Selenica[2], Samuel F. Bakhoum[1,3], Jorge S. Reis-Filho[2,7], Nadeem Riaz [1] & Simon N. Powell [1,6] ✉

Collisions of the transcription and replication machineries on the same DNA strand can pose a significant threat to genomic stability. These collisions occur in part due to the formation of RNA-DNA hybrids termed R-loops, in which a newly transcribed RNA molecule hybridizes with the DNA template strand. This study investigated the role of RAD52, a known DNA repair factor, in preventing collisions by directing R-loop formation and resolution. We show that RAD52 deficiency increases R-loop accumulation, exacerbating collisions and resulting in elevated DNA damage. Furthermore, RAD52's ability to interact with the transcription machinery, coupled with its capacity to facilitate R-loop dissolution, highlights its role in preventing collisions. Lastly, we provide evidence of an increased mutational burden from double-strand breaks at conserved R-loop sites in human tumor samples, which is increased in tumors with low RAD52 expression. In summary, this study underscores the importance of RAD52 in orchestrating the balance between replication and transcription processes to prevent collisions and maintain genome stability.

Transcription and replication are two tightly regulated processes necessary for gene expression and DNA duplication respectively, both of which are essential for cellular integrity. It is imperative for the cell to maintain temporal and spatial separation of these two processes to prevent them from colliding (transcription-replication conflicts (TRCs)), which can result in replication stress and DNA damage, ultimately leading to genome instability and tumorigenesis. Aberrant accumulation of secondary structures such as R-loops have been implicated as a major source of TRCs[1,2]. R-loops are three-stranded RNA-DNA hybrids that are formed transiently during transcription when the nascent RNA anneals back to the template DNA and displaces the non-template strand within the RNA polymerase active site[3]. Traditionally, these hybrids have been shown to play an important physiological role in gene activation, termination, and chromatin regulation. However, transcriptional dysregulation, both in the form of gene overexpression and aberrant RNA polymerase II (Pol II) pausing, has been associated with the accumulation and persistence of pathological R-loops[4–6]. R-loops pose a significant threat to DNA replication as transcription and replication translocate on the same DNA template.

As TRCs pose a considerable threat to genomic integrity, mechanisms to manage the collisions are required to prevent them causing undue DNA damage[4,7]. Prevention mechanisms include

[1]Department of Radiation Oncology, MSKCC, New York, NY 10065, USA. [2]Department of Pathology and Laboratory Medicine, MSKCC, New York, NY 10065, USA. [3]Human Oncology and Pathogenesis, MSKCC, New York, NY 10065, USA. [4]Department for BioMedical Research, University of Bern, Bern, CH 3008, Switzerland. [5]SIB, Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland. [6]Molecular Biology Program, MSKCC, New York, NY 10065, USA. [7]Present address: AstraZeneca, Gaithersburg, MD 20878, USA. ✉e-mail: jalanm@mskcc.org; powells@mskcc.org

limiting the accumulation of R-loops by assembling RNA-binding proteins on nascent RNA[8,9]; regulating topological stress associated with transcription and chromatin architecture[10–12]; removal of R-loops via nucleases[13,14] or RNA-DNA helicases[15,16]; and, the subsequent repair of the damage resulting from TRCs[17–21]. In contrast, recent studies have proposed that R-loops can play a major role in double-strand break (DSB) repair at transcriptionally active loci via homologous recombination (HR), by providing a scaffold for the recruitment of DNA repair factors to the site of damage[22,23]. Unexpectedly, RAD52 emerged as a common factor in all these R-loop associated DSB repair pathways[24–28].

Human RAD52, a protein known to have a DNA binding ability, has been associated with DSB repair owing to its role as a back-up HR repair factor[29] and its synthetic lethal relationship in BRCA-deficient cancers[30–32]. Surprisingly, recent studies have shown strong RNA-binding ability for RAD52 in vitro supporting a role in resolving transcription associated DSBs[24–28]. However, given the interaction of RAD52 with RNA[26,33,34] and Pol II[35], it is conceivable that RAD52 could also be involved upstream in the regulation of R-loops themselves.

To understand RAD52's role in R-loop management, we performed mass spectrometry (MS) analysis of the RAD52 protein interactome. We found that RAD52 predominantly interacts with proteins engaged in the transcription complex, suggesting that RAD52 recruitment to R-loop sites may be facilitated via this interaction. We observed that loss of RAD52 induces elevated levels of Pol II pausing and R-loop accumulation leading to increased TRCs and genomic instability. Furthermore, we identified a role for the previously uncharacterized C-terminal domain of RAD52, in that it is essential for RAD52's interaction with Pol II and helps recruit Topoisomerase IIα (TOP2A) to R-loops, in order to alleviate torsional stress and aid in resolving TRCs. Additionally, we found direct evidence of increased mutational scars at R-loop forming regions across tumor types and these were exacerbated in tumors with low levels of RAD52 expression. This study supports a role of RAD52 directly at R-loops, and its absence contributes to increased R-loop associated genomic instability.

Results

RAD52 interacts with the transcriptional complex and co-localizes with RNA POL II

We first sought to identify RAD52-interacting proteins under physiological conditions in an unbiased manner by performing an immuno-precipitation (IP) with the expression of RAD52 fused to a N-terminal HA-tag, followed by MS (Fig. 1a, b, Supplementary Fig. 1a). MS discovered 212 proteins significantly enriched over the HA-tag control. Reassuringly, RPA1, a critical sub-unit of the RPA complex and a known interactor of RAD52[36–39] was identified in this analysis (Fig. 1b, c, Supplementary Data 1). However, the majority of hits identified had an RNA-associated role as depicted by the gene ontology (GO) analysis (Supplementary Fig. 1b, c), rather than DNA repair, suggesting that RAD52 has a strong interaction with the transcription machinery (as highlighted in Fig. 1b). To validate this observation, we performed both a co-immunoprecipitation (Co-IP) and a proximity ligation assay (PLA) between endogenous RAD52 and Pol II, finding clear evidence for their interaction independent of DNA or RNA (Fig. 1c–f, Supplementary Fig. 1d), corroborating previous observations of this interaction seen with over-expressed RAD52[35].

To understand the spatial resolution of the RAD52-Pol II interaction, we performed a chromatin immunoprecipitation sequencing (ChIP-seq) of endogenous Pol II and RAD52. We observed that there is a significant coincidence of Pol II and RAD52 peaks across the genome, with increased enrichment seen at transcription start sites (TSS) and transcription end sites (TES) (Fig. 1g–k, Supplementary Fig. 1e). Approximately 40% of all RAD52 peaks were associated with Pol II, though only a subset of Pol II peaks colocalized with RAD52 (~10%), suggesting that RAD52 is recruited to only a subset of all sites of transcription (Fig. 1j).

RAD52 prevents TRCs by reducing Pol II pausing and decreasing the level of pathologic R-loops

We next sought to better understand the function of RAD52 as an accessory factor associated with the transcription complex. We observed that depletion of RAD52 leads to increased Pol II pausing at the TSS as demonstrated by increased accumulation of Pol II peaks at the TSS specifically in the subset of genes that had Pol II-RAD52 co-occupancy (Fig. 2a, b, Supplementary Fig. 2a, b), loss of RAD52 having no effect on the Pol II profiles of the other genes (Supplementary Fig. 2c). Furthermore, this did not affect global gene expression profiles (Supplementary Fig. 2d, Supplementary Data 2), suggesting RAD52 does not alter transcriptional profiles in cells.

Increased Pol II pausing has been known to be associated with R-loop accumulation[40]. We wanted to see if this holds true at RAD52 associated Pol II pausing sites (Fig. 2a, b). To this end, we performed ChIP-seq analysis of R-loop associated peaks using the S9.6 antibody that has been characterized to specifically interact with RNA-DNA hybrids[41]. We observed that 22% of RAD52 peaks associated with R-loops, albeit at a frequency lower than its association with Pol II (41%), suggesting that RAD52 may associate with Pol II independently of R-loops (Fig. 2c, d). Upon comparing the RAD52 interactome (Fig. 1b) with that of R-loops[42], it was apparent that the majority of proteins which associated with RAD52 also associated with R-loops (Supplementary Fig. 2e, f, Supplementary Data 3), indicating that RAD52 may play an important role in R-loop homeostasis.

We next set out to investigate the effect of RAD52 loss on global R-loop levels. We observed that RAD52 loss in cells led to a significant increase in global S9.6 signal (Fig. 2e, f, Supplementary Fig. 3a–h), comparable to depletion of Aquarius (AQR), a known R-loop resolution factor[15] (Supplementary Fig. 3i, j). Furthermore, the observed S9.6 signal was sensitive specifically to RNase H treatment but not RNase III (Fig. 2e, f, Supplementary Fig. 3b–d). RNase H is known to explicitly digest RNA-DNA hybrid, confirming the detection of R-loops, and no other non-specific RNA species under the given conditions[43]. Prior work has suggested that increased R-loops pose a threat to replication, leading to increased TRCs[44]. In order to understand the physiological consequences of increased R-loops in RAD52 deficient backgrounds, we performed a PLA between Pol II and PCNA, an essential component of the replisome (Fig. 2g, Supplementary Fig. 3k, l). We found a significant increase in TRCs observed with the loss of RAD52, which was further amplified with increased R-loops in an AQR-deficient context (Fig. 2h, i), implicating RAD52 as a mediator of TRC resolution. This effect of RAD52 was also found in Senataxin (SETX)[45] depleted cells (Supplementary Fig. 3m, n), confirming that the effect of RAD52-loss on TRC's was caused by the presence of increased R-loops, independent of causation. However, it is worth noting that the RAD52-Pol II interaction is not limited to the S-phase of the cell cycle, suggesting that RAD52 associates with the transcription machinery throughout the cell cycle, potentially acting as a surveyor of replication stress (Supplementary Fig. 3o–q).

RAD52 is recruited to sites of transcription-replication conflicts via its RNA-Pol II interacting C-terminal domain

A previous study demonstrated that RAD52 interacts with the transcription complex via its C-terminal domain[35]. Given our observation that RAD52 associates with Pol II (Fig. 1 & 2), we posited that RAD52's C-terminal domain would be essential for its role in resolving TRCs via its interaction with the transcription machinery. In order to test this hypothesis, we generated an HA-tagged RAD52 mutant in which we deleted amino acids 302–410 (referred to as RAD52^ΔC) (Fig. 3a, b, Supplementary Fig. 4a). This amino acid region has previously been identified as the minimum number of residues needed for RAD52 to interact with Pol II in vitro[35]. As RAD52 is a protein known for its role in DNA repair, we first confirmed that this was not disrupted by deleting the C-terminus. We tested this using functional assays of DSB repair,
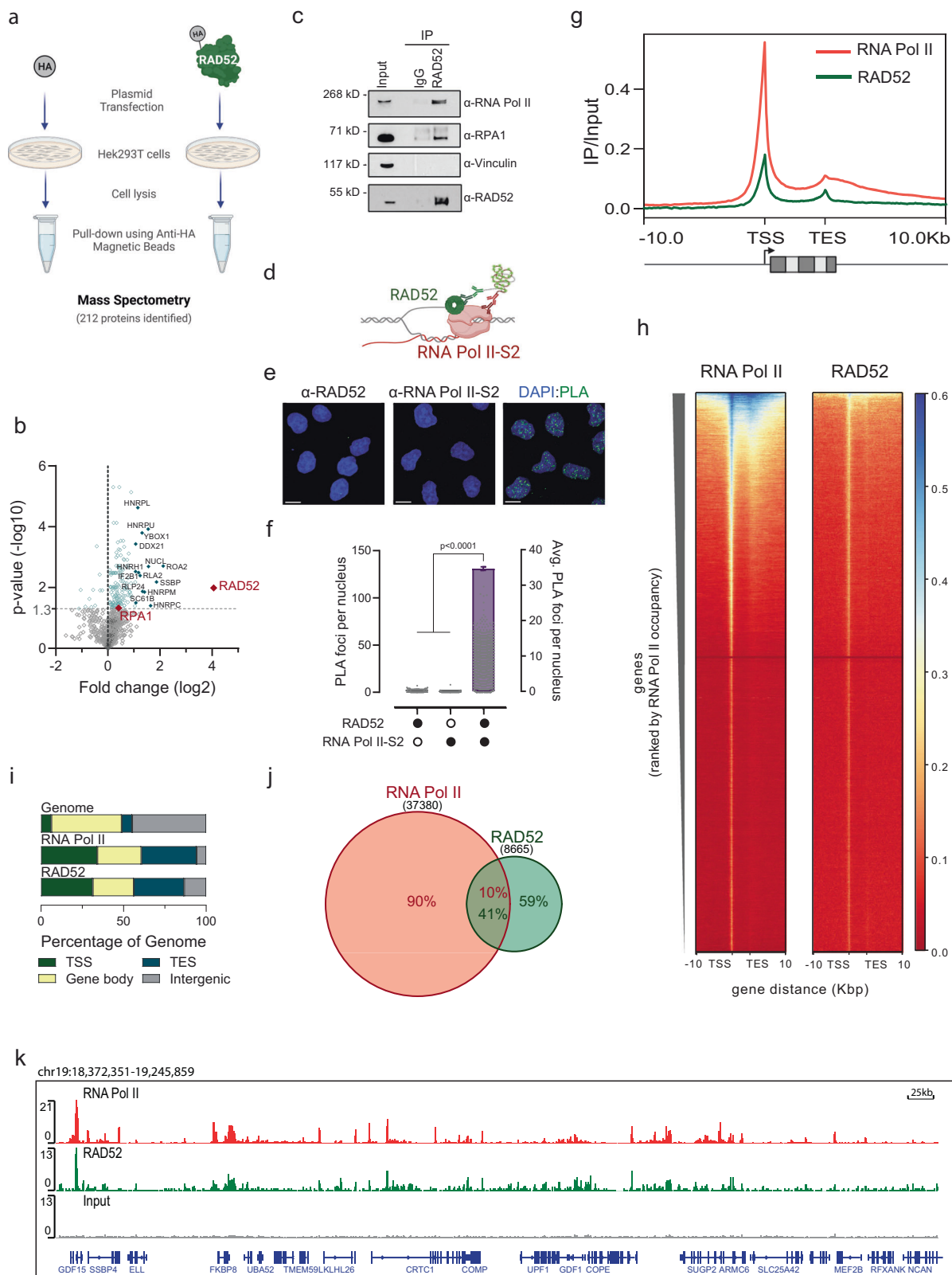
namely single strand annealing (SSA) and HR using the previously described reporters[46,47]. In *RAD52*[-/-] cells, we observed that complementation with either RAD52[WT] or RAD52[ΔC] rescued the SSA, and HR deficient phenotypes induced by RAD52 deficiency (Fig. 3c, d). This implies that the loss of the C-terminus of RAD52 does not impair its DNA repair activity.

To test if the C-terminus was indeed responsible for RAD52's interaction with Pol II, we performed a PLA between the two proteins (Fig. 3e, Supplementary Fig. 4b, c). As suggested by in vitro biochemical studies[35], RAD52[ΔC] had a reduced interaction with Pol II in human cells (Fig. 3f, g), confirming the importance of the C-terminal domain. Furthermore, RAD52[ΔC] failed to rescue the elevated levels of R-loops

Fig. 1 | RAD52 association with the transcriptional complex. a Schematic representation of the workflow for the identification of RAD52 interacting proteins. HA-control and HA-RAD52 immunoprecipitation was performed in HEK293T cells using α-HA tagged magnetic beads for the pulldown followed by Mass spectrometry (MS). **b** Volcano plot of the proteins identified in RAD52 IP-MS in $n = 3$ biologically independent experiments. Mean log2 fold change in protein intensities on the x-axis of all replicates between HA and HA-RAD52 are plotted against the −log10 adjusted p-value (Student's two-sided t-test with equal variance) on the y-axis. 212 proteins were identified to be significantly enriched. Significantly enriched proteins in blue ($p < 0.05$) and non-significant in grey. **c** Co-immunoprecipitation of endogenous RAD52 binding proteins in HeLa cells. RAD52 and IgG antibodies were used to immuno-precipitate proteins and analyzed by immunoblotting with indicated antibodies. Results reproducible for at least 2 biological replicates. **d** Schematic representation of PLA to visualize proximity of RAD52 protein and RNA Pol II. **e** Representative images of the nuclear PLA foci (α-RAD52: α-RNA Pol II S2) across stated conditions (Scale bar 10 μM). **f** Quantitative analysis of nuclear PLA foci from (**e**) Data are plotted as mean ± SEM. The data presented shows ≥ 500 nuclei from 3

biological replicates; p-values calculated using unpaired two tailed t-tests. **g** Metagene plots showing the distribution of the RNA Pol II and RAD52 Chromatin immunoprecipitation sequencing (ChIP-seq) peaks (IP/input) in HeLa cells across genes and the flanking regions (± 10 kb). TSS: Transcription Start Site, TES: Transcription End Site. **h** Heatmap representing RNA Pol II and RAD52 ChIP-seq tracks, centered at the TSS and TES ± 10 kb, and rank-ordered according to RNA Pol II occupancy. **i** Bar chart showing how RNA Pol II and RAD52 peaks are distributed across different genomic regions as indicated. Peaks were obtained with MACS2. Genome wide distribution is shown on top for comparison. **j** Venn diagram showing the overlap of peaks RNA Pol II ChIP and RAD52 ChIP according to MACS2 across the genome. **k** A representative snapshot of chromosome 19 depicting RNA Pol II (red) and RAD52 (green) ChIP binding sites in control HeLa cells. Input DNA (grey) represents a negative control for background normalization. Schematics in Fig. 1 (**a**) and (**d**) were created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license. Source data are provided as a Source Data file.

and TRCs associated with the loss of RAD52 (Fig. 3h–j, Supplementary Fig. 4d, e), implicating the essentiality of the C-terminal domain of RAD52 in the reduction of TRCs.

RAD52 recruits TOP2A to R-loops to help resolve transcription-replication conflicts

In order to elucidate the mechanism by which RAD52 facilitates resolution of TRCs, we performed an IP-MS analysis of overexpressed RAD52[WT] and RAD52[ΔC] to tease apart factors that were differentially associated with the C-terminal domain of RAD52 (Supplementary Fig. 5a). The screen was done in an AQR-depleted background so as to increase basal levels of R-loop (Supplementary Fig. 3i, j). Of the 26 R-loop associated factors that had differential interaction with RAD52[WT] and RAD52[ΔC], TOP2A stood out as a top hit owing to its previously suggested role in TRC resolution[48–50] (Supplementary Fig. 5b, c, Supplementary Data 4&5, Supplementary Note 1). PLA between RAD52 and TOP2A was performed to confirm this interaction (Fig. 4a–c), which was further elevated in the presence of increased R-loops (Supplementary Fig. 5d, e). As expected, the RAD52-TOP2A interaction was disrupted in the absence of the RAD52 C-terminal domain (Fig. 4d–f), corroborating the IP-MS results.

Consistent with previous studies[48–50], loss of TOP2A led to increased R-loops as well as elevated TRCs (Fig. 4g–k, Supplementary Fig. 5f), confirming its role in mitigating R-loop associated TRCs. To study if recruitment of TOP2A to these sites was indeed RAD52 associated, we performed a PLA between TOP2A and S9.6 (Fig. 4l, Supplementary Fig. 5g, h). There was a significant reduction in TOP2A recruitment to R-loops in the absence of RAD52, both in physiological conditions and with elevated R-loops (Fig. 4m, n), supporting our hypothesis that RAD52 helps to recruit TOP2A to R-loop associated TRCs, alleviating the duplex torsional stress and helping resolve TRCs (Fig. 4o). This observation was further supported by correlation with a previously published TOP2A ChIP-seq dataset[51], where 18% of RNA Pol II-RAD52 overlapping peaks (from Fig. 1j) co-occurred with TOP2A peaks, indicating that TOP2A is recruited to RAD52-mediated TRCs and thus consistent with our model (Supplementary Fig. 5i).

RAD52-depleted cells have increased replication stress and accumulate γH2AX at R-loop forming regions

TRCs have been shown to be an established source of replication stress in cells owing to stalled replication and increased fork collapse[1,2,5]. Given RAD52's role in TRC resolution, we hypothesized that RAD52 depletion could potentially lead to increased replication stress. We observed a mild increase in replication stress as measured by reduced DNA fiber track lengths after a sequential pulse with two thymidine analogs−5-Chloro-2′-deoxyuridine (CldU) and 5-iodo-2′-deoxyuridine (IdU)−for 30 min each (Fig. 5a–c, Supplementary

Fig. 6a–c). Furthermore, consistent with its role in the resolution of TRCs, overexpression of the RAD52[ΔC] mutant was unable to rescue the increased replication stress phenotype observed with RAD52-depletion (Fig. 5a–c). However, depletion of RAD52 did not affect global cell cycle profiles nor did it lead to an ATM and ATR-dependent checkpoint activation, suggesting that the induced local replication effects do not lead to an altered S-phase (Supplementary Fig. 6d, e). Moreover, R-loop associated TRCs have been shown to have increased levels of DNA damage[2,44].

To further assess the DNA damage at RAD52-associated R-loop forming regions (Supplementary Fig. 6b), we analyzed the distribution of γH2AX around R-loops in RAD52-depleted cells using ChIP-seq. We observed that there was increased γH2AX accumulation at R-loop forming genes, which was persistent even ±0.5 Mb around the TSS (Fig. 5d, Supplementary Fig. 6f, g). These findings were further corroborated by performing a PLA of S9.6 and γH2AX in RAD52-depleted cells, under physiological and increased R-loop conditions (Fig. 5e–g). However, loss of RAD52 does not elicit a global DNA damage response as confirmed by the unaltered levels of total γH2AX in normal versus RAD52-depleted cells (Supplementary Fig. 6h, i). These findings are also consistent with the unchanged cell cycle progression previously observed.

R-loops are a source of genome instability in tumors

R-loops have long been associated with DNA damage as a consequence of prolonged replication fork stalling and DSBs arising from collapsed replication forks[1,2]. However, most of the evidence for the damage associated with R-loops comes from indirect evidence in the form of increased γH2AX foci or comet assay tail lengths[5]. The direct consequence of R-loops on the genome in the form of mutational signatures remains poorly understood. We hypothesized that if R-loops can lead to DSBs, there should be an increased burden of genomic scars associated with conserved R-loop forming regions across human tumors.

To investigate this hypothesis, we built a consensus R-loop dataset comprised of correlated peaks from 18 published datasets[52] (Supplementary Fig. 7a, b, Supplementary Data 6). This combined dataset was assessed to confirm that it followed the established conventions of R-loops being associated with transcribed genes, with a significant increase in occurrence being observed at TSSs and TESs[53] (Fig. 6a). We next proceeded to overlay our R-loop dataset with previously identified somatic mutations from the PCAWG, ICGC and TCGA cohorts[54–56] (see methods for details). We observed a significant increase in structural alterations associated with R-loop forming regions across the genome (Fig. 6b). Not surprisingly, we observed a decrease in single nucleotide variants (SNVs) in R-loop forming regions, consistent with likely increased transcription-
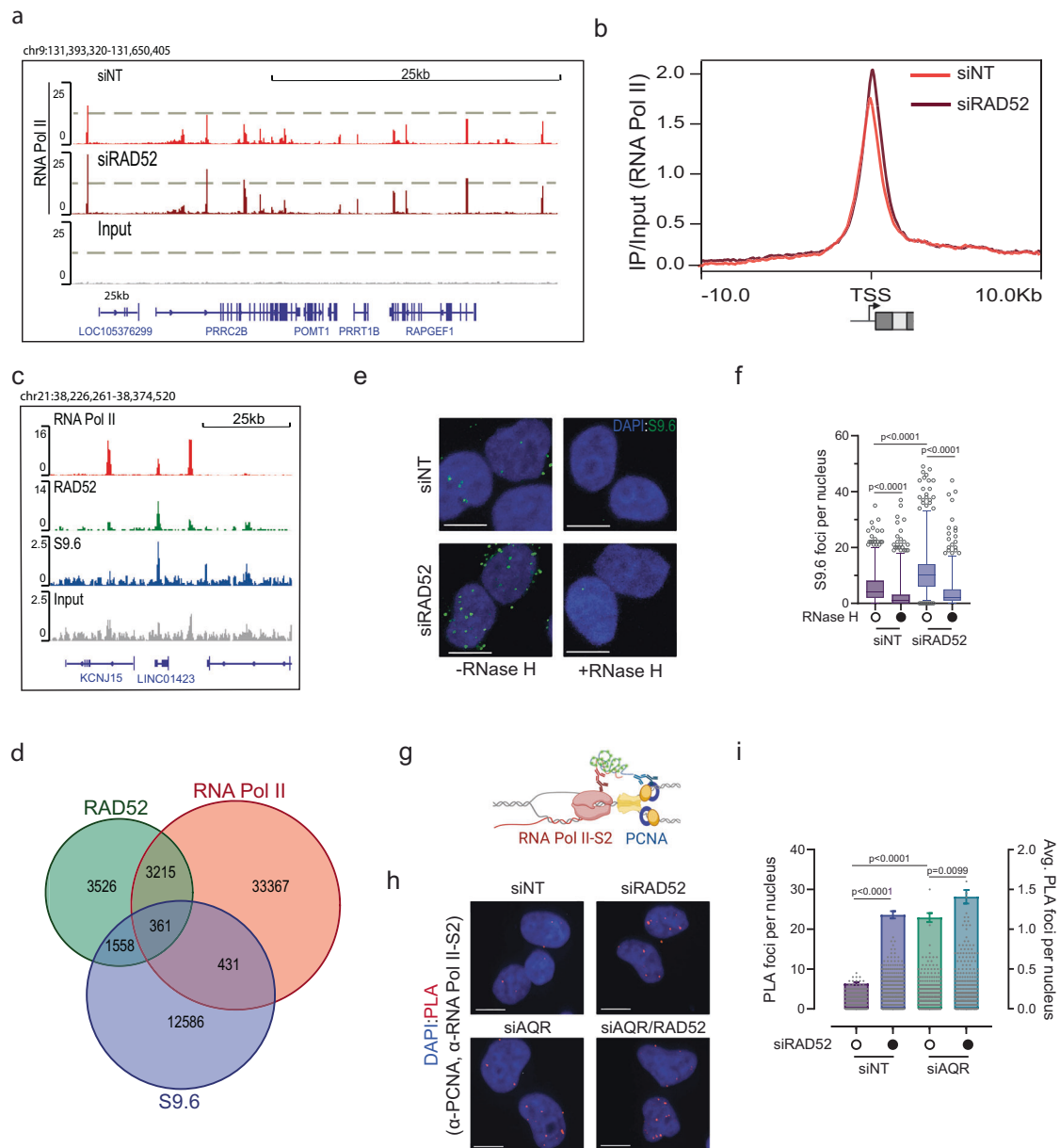
Fig. 2 | Loss of RAD52 increases R-loop formation and exacerbates transcription-replication conflicts. a Representative snapshot of chromosome 9 depicting RNA Pol II occupancy from ChIP-seq analysis (IP/input) in siNT (red) and siRAD52 (dark red) transfected HeLa cells. (**b**) Metagene plot showing the distribution of the RNA Pol II occupancy at the TSS and flanking regions (± 10 kb) of genes with overlapping RNA Pol II and RAD52 peaks. Plots shown: siNT (control) and siRAD52 transfected HeLa cells. (**c**) ChIP-seq of RNA Pol II (red), RAD52 (green) and S9.6 (R-loops; blue) occupancy in control HeLa cells. Representative snapshot of chromosomes 21 are shown. Input (grey) DNA as negative control for background normalization. **d** Venn diagram of the percentage of genes overlapping with RNA Pol II, RAD52 and S9.6 ChIP peaks (MACS2). **e** Representative images of S9.6 immunostaining to detect R-loops in siNT (control) and siRAD52 transfected HeLa cells. RNase H treatment was added as a negative control to eliminate R-loops (Scale

bar 10 μM). **f** Quantitative analysis of nuclear S9.6 foci across stated conditions from (**e**). Data plotted as box and whiskers. Boxes extend from the 25th–75th percentiles, with the median displayed as a line. The whiskers mark the minimum (1 percentile) and maximum (99th percentile). The data presented shows ≥ 500 nuclei from 3 biological replicates; *p*-values calculated using unpaired two tailed *t*-tests. **g** Schematic representation of PLA to visualize proximity of PCNA and RNA Pol II to measure TRCs. The schematic illustration was created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license. **h** Representative images of the nuclear PLA foci (α-PCNA: α-RNA Pol II S2) across stated conditions (Scale bar 10 μM). **i** Quantitative analysis of nuclear PLA foci from (**h**). Data are plotted as mean ± SEM. The data presented shows ≥ 500 nuclei from 3 biological replicates; *p*-values calculated using unpaired two tailed *t*-tests. Source data are provided as a Source Data file.

coupled repair (Fig. 6c and 6f). In contrast, genetic alterations likely to form due to a DSB were significantly increased in R-loop forming regions, including both insertions and deletions >1 bp (long InDels) and structural variants (SVs), (Fig. 6d, e, g, h, Supplementary Fig. 8a–c). Historically, SVs and indels have been associated with aberrant repair at DSBs either from NHEJ or backup pathways to HR[57]. These observations are consistent with the idea that R-loops lead to

the formation of DSBs, repair of which can result in large insertions, deletions, and translocations, as we observed.

TRCs are preferentially enriched at sites of head-on collisions (HO) as opposed to co-directional collisions (CD)[44] (Fig. 6i). To determine whether HO collisions could lead to increased accumulation of R-loop induced genomic alterations, we classified our R-loop dataset as CD or HO-associated by overlaying it with the previously
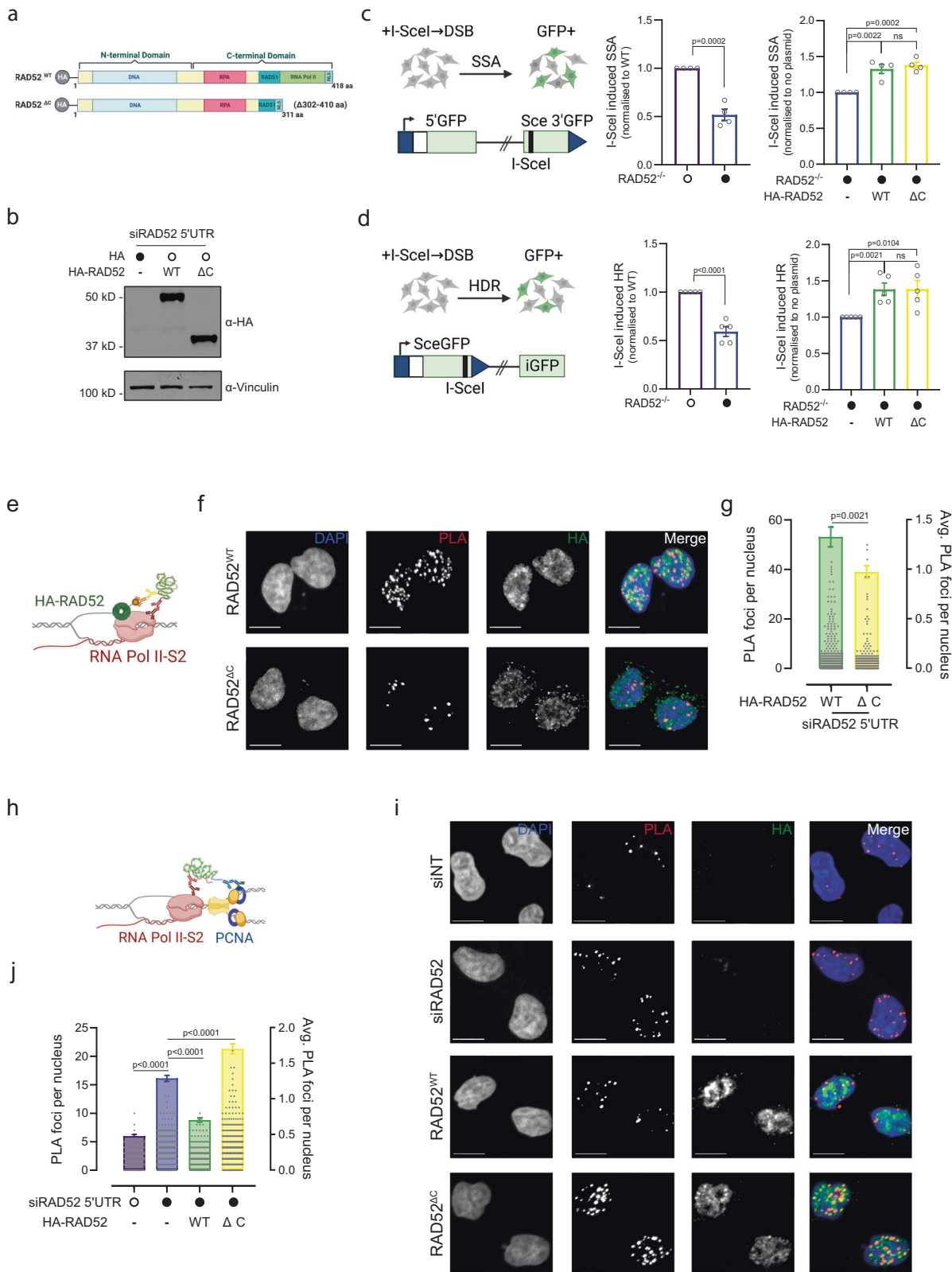
published and annotated Okazaki fragment sequencing (OK-seq) data[44,58] (Fig. 6j). As expected, we observed a significant difference between the genetic alterations at CD versus HO, with a 3-fold increase at HO (Fig. 6k). Furthermore, tumors with lower RAD52 expression levels seem to correlate with increased mutations at R-loops (Fig. 6l), supporting the idea that RAD52 acts at R-loops to prevent genomic rearrangements.

Discussion

Despite its apparent dispensability in humans, RAD52 has been characterized as an essential backup DNA repair factor for BRCA2 due to its ability to mediate HR and SSA. Recently, RAD52 was shown to be involved in genome maintenance via additional roles in Break-Induced Replication (BIR) and mitotic DNA synthesis (MiDAS) arising from replication stress[29,59]. In this study, we uncovered a previously

Fig. 3 | C-terminal domain of RAD52 is essential for the prevention of transcription-replication conflicts. a Schematics of the domain structures of wild type (WT) - RAD52 protein and C-terminal (ΔC) deleted RAD52 (Δ302-410 amino acids). From N-terminal to C-terminal, RAD52 protein has DNA binding domain, RPA binding domain, RAD52 binding domain, RNA Pol II binding domain and a nuclear localization signal (NLS). The domains are not drawn to scale. **b** Western blot confirming the expression of HA-RAD52[WT] and HA-RAD52[ΔC]. Results reproducible for at least 2 biological replicates. **c** (Left) Scheme of the single stranded annealing (SSA) reporter system: The SSA-GFP reporter contains a 5′ fragment of the GFP (5′-GFP) gene, and a 3′ fragment of the GFP (3′-GFP) with an I-SceI site. Repair of the I-SceI-induced DSB by SSA leads to formation of GFP+ cells. (Middle) Quantification of SSA repair assay in WT and RAD52[−/−] HCT116 cells. (Right) Quantification of SSA repair assay in RAD52[−/−] HCT116 cells with overexpression of either RAD52[WT] or RAD52[ΔC] (n = 4 biological replicates). **d** (Left) Scheme of the homology dependent recombination (HDR) reporter system The HDR-GFP reporter system contains the GFP gene interrupted by a I-SceI site, and a fragment of the GFP with truncated 3′- and 5′-terminus. Repair of the I-SceI-induced DSB by HDR leads to formation of GFP+ cells. (Middle) Quantification of HDR repair assay in WT and RAD52[−/−] HCT116 cells. (Right) Quantification of HDR repair assay in RAD52[−/−] HCT116 cells with overexpression of either RAD52[WT] or RAD52[ΔC]. (n = 5 biological replicates). **e** Schematic representation of PLA to visualize proximity of HA-tagged RAD52 (HA-RAD52) and RNA Pol II. **f** Representative images of the nuclear PLA foci (α-HA: α-RNA Pol II S2) across stated conditions with overexpression of either RAD52[WT] or RAD52[ΔC] (Scale bar 10 μM). **g** Quantitative analysis of nuclear PLA foci across stated conditions described in (**f**). The data presented shows ≥ 500 nuclei from 3 biological replicates. **h** Schematic representation of PLA to visualize proximity of PCNA and RNA Pol II to measure TRCs. **i** Representative images of the nuclear PLA foci (α-PCNA: α-RNA Pol II S2) across stated conditions with overexpression of either RAD52[WT] or RAD52[ΔC] in HeLa cells (Scale bar 10 μM). **j** Quantitative analysis of nuclear PLA foci from across stated conditions described in (**i**). The data presented shows ≥ 500 nuclei from 3 biological replicates. In Fig. 3 (**c**) (**d**) (**g**) and (**j**), data are plotted as mean ± SEM and p-values calculated using unpaired two tailed t-tests. Schematics in Fig. 3 (**a**) (**c**) (**d**) (**e**) and (**h**) were created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license. Source data are provided as a Source Data file.

undescribed role for human RAD52 in R-loop homeostasis via its association with the transcriptional machinery. We identified a robust RAD52-Pol II interaction (Fig. 1) and determined that RAD52 associates with Pol II predominantly at the TSS in a subset of genes. Notably, loss of RAD52 alone was sufficient to cause increased Pol II pausing at these loci (Fig. 2).

While Pol II pausing has been implicated as a rate-limiting step in transcription, it can be particularly problematic for the maintenance of genome integrity by interfering with the replication machinery, causing TRCs followed by DNA damage[60]. Pol II pausing promotes the formation of transient secondary structures such as R-loops which are the major source of such conflicts[40,44]. In this study, we present strong evidence that RAD52 helps resolve these R-loops from forming TRCs, the loss of RAD52 leading to increased TRCs and associated DNA damage (Figs. 2 and 5). Interestingly, while almost half of chromatin bound RAD52 was associated with Pol II, only half of these sites were associated with R-loop formation (Fig. 2d), suggesting that RAD52 may associate with transcription sites independently of R-loop formation. Furthermore, we found that the previously uncharacterized C-terminal domain of RAD52 is essential for its Pol II interaction and its role in TRC resolution, separate to its role in HR and SSA (Fig. 3). RAD52's involvement in R-loop resolution and collision avoidance is supplementary to its DSB-repair roles. RAD52's ability to support DNA/RNA binding is linked to its annealing abilities, but recruitment to the sites of transcription requires the C-terminal domain. We think that apart from RAD52 helping to resolve R-loops, RAD52 can also have a downstream role in repairing DSBs that arise from persistent R-loops as seen in transcription-coupled homologous recombination (TC-HR) or transcription-associated homologous recombination repair (TA-HR)[24–28].

The cell tightly maintains R-loop homeostasis by regulating pathways that control its formation and degradation[4,7]. The release of supercoiling associated with transcription and replication is mediated by topoisomerase 1 (single-strand DNA nicking) within the region of the transcription site or the site active replication[61]. However, when transcription and replication are leading to a head-on collision, the duplex DNA between the sites of transcription and replication are trapped by supercoiling of different polarities, creating a zone of conflict, where the duplexes may form "knotted" loops. Release of the accumulated duplex DNA torsional stress is required to resolve the TRC. The Topoisomerase IIα (TOP2A) cleavage complex is one such protein complex that could release the two sources of negative supercoiling coming from opposite directions during transcription and replication[62,63], the loss of which leads to increased R-loop accumulation and increased TRCs[50] (Fig. 4). We found that RAD52 promotes TOP2A recruitment to R-loop sites, thus helping resolve TRCs

and preventing the ensuing genomic instability (Fig. 4). It is therefore conceivable that the increased Pol II pausing observed in RAD52-depleted cells (Fig. 2) could be ascribed to the inability of the cell to recruit TOP2A to the TRC region, leading to increased torsional stress[63] and R-loop enrichment. However, while we demonstrate that the C-terminal domain of RAD52 seems to mediate TOP2A's recruitment to TRC's, further experiments are warranted to determine if this function is dependent or independent of RAD52's association to the Pol II complex, which is also mediated via its C-terminal domain.

Pathological R-loops have long been implicated in genome instability, albeit through indirect evidence in the form of γH2AX signal or accumulation of DSBs at R-loop forming regions[5]. Here, we provide direct evidence of R-loops acting as a driver of DSB-induced genomic instability in varied human tumor samples (Fig. 6). We observed elevated levels of structural variants and indels at R-loop forming regions across tumor types, but not single nucleotide variants, in contrast to recent observations reported to be linked to R-loops[64]. This difference likely stems from analytic differences, whereas here we used tumors only sequenced by WGS with light filtering, prior work had analyzed heterogeneously sequenced tumors with over 95% of cancers removed from analysis. Furthermore, an increased density of these mutations was observed at HO collisions when compared to CO collisions, supporting the concept that HO collisions produce DSB and are more harmful for the cell. Additionally, low RAD52 expression in tumors was associated with an increased mutational burden at R-loops, consistent with the previously uncharacterized role for RAD52. Moreover, there was no correlation between TOP2A expression level and mutational signatures, unlike that seen with RAD52 expression levels (Fig. 6l, Supplementary Fig. 9a–c). This can be explained by the fact that there is no direct correlation between the expression levels of TOP2A and RAD52 (Supplementary Fig. 9c), but the effect we see is a function of RAD52 not being able to recruit TOP2A to the sites of collisions, resulting in DNA breaks and genomic instability. This is in concordance with the long-standing view in the field that for DNA repair, the protein levels do not always play the rate-limiting step, but it is often the opportunity for the protein to get to the DNA lesion at the right time that matters. Hence, expression of TOP2A is unsurprisingly unrelated to genomic instability.

Our study finds a unique role for RAD52 in genome maintenance via its ability to resolve R-loops and TRCs. Considering that transcription induced replication stress is one of the most common endogenous sources of DSB in the cell, it is possible that this transcription associated role of RAD52 may also contribute to its synthetic lethal phenotype observed in BRCA-deficient cells, in addition to the previously characterized DSB repair activities, including RAD51-mediator function and single-strand annealing[30–32]. Furthermore, we
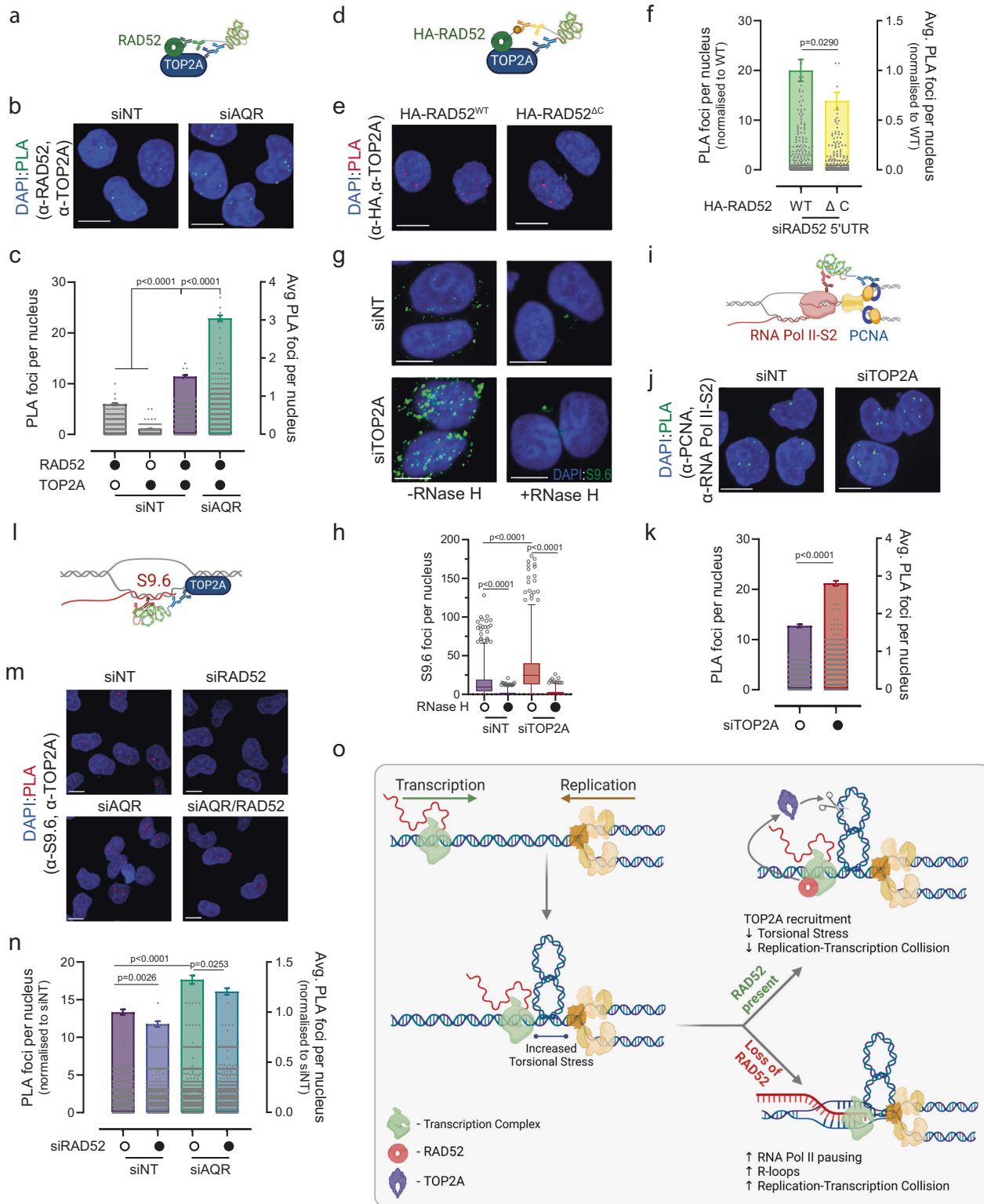
demonstrate that R-loops if left unrepaired, can lead to genomic instability resulting in mutagenesis, chromosomal rearrangements, and cancer.

Methods

Cell Culture and transfections

HeLa (ATCC, #CCL-2), HEK293T (ATCC, #CLR-3216) and U2OS (ATCC, #HTB-96) cells were grown in complete DMEM high glucose supplemented with 10% FBS, 2 mM L-glutamine, 20 mM HEPES, 100 I.U./ml Penicillin, and 100 μg/ml Streptomycin. HCT116 WT and *RAD52*[-/-] cell lines were obtained from Dr. Eric A. Hendrickson[65] and cultured in McCoy's 5 A medium supplemented with 10% FBS, 2 mM L-glutamine and 100 I.U./ml Penicillin, and 100 μg/ml Streptomycin. All cells were grown in a humidified 37 °C incubator with 5% CO_2.

0.5×10^6 cells were reverse transfected using RNAiMAX (Invitrogen, #13778150) according to the manufacturer's instructions with 40 pmol of siRNAs of Rad52 (Dharmacon ON-TARGETplus SMART-pool, #L-011760), Rad52 5'UTR (Dharmacon ON-TARGET 5'UTR, #J-

Fig. 4 | RAD52 recruits TOP2A to mitigate transcription-replication conflicts.
a Schematic representation of PLA to visualize proximity of RAD52 and TOP2A.
b Representative images of the nuclear PLA foci (α-RAD52: α-TOP2A) in siNT
(control) and siAQR transfected HeLa cells (Scale bar 10 μM). **c** Quantitative analysis
of nuclear PLA foci across stated conditions described in (**b**). The data presented
shows ≥ 500 nuclei from 3 biological replicates. **d** Schematic representation of PLA
to visualize proximity of HA-tagged RAD52 (HA-RAD52) and TOP2A.
e Representative images of the nuclear PLA foci (α-HA: α-TOP2A) in siRAD52
(5'UTR) transfected HeLa cells with overexpression of either RAD52^{WT} or RAD52^{ΔC}
(Scale bar 10 μM). **f** Quantitative analysis of nuclear PLA foci across stated condi-
tions described in (**e**). The data presented shows ≥ 500 nuclei from 3 biological
replicates. **g** Representative images of S9.6 immunostaining to detect R-loops in
siNT (control) and siTOP2A transfected HeLa cells. RNase H treatment was added as
a negative control to eliminate R-loops (Scale bar 10 μM). **h** Quantitative analysis of
nuclear S9.6 foci across stated conditions from (**g**). Data plotted as box and whis-
kers. Boxes extend from the 25th to 75th percentiles, with the median displayed as a
line. The whiskers mark the minimum (1 percentile) and maximum (99th

percentile). The data presented shows ≥ 500 nuclei from 3 biological replicates; *p*-
values calculated using unpaired two tailed *t*-tests. **i** Schematic representation of
PLA to visualize proximity of PCNA and RNA Pol II to measure TRCs.
j Representative images of the nuclear PLA foci (PCNA: RNA Pol II S2) in siNT
(control) and siTOP2A transfected HeLa cells (Scale bar 10 μM). **k** Quantitative
analysis of nuclear PLA foci across stated conditions described in (**j**). The data
presented shows ≥ 500 nuclei from 3 biological replicates (**l**) Schematic repre-
sentation of PLA to visualize proximity of S9.6 and TOP2A. **m** Representative
images of the nuclear PLA foci (α-S9.6: α-TOP2A) in siNT (control), siRAD52 and
siAQR transfected HeLa cells (Scale bar 10 μM). **n** Quantitative analysis of nuclear
PLA foci across stated conditions described in (m) normalized to siNT. The data
presented shows ≥ 500 nuclei from 3 biological replicates. **o** Mechanistic model of
RAD52 role in preventing transcription-replication conflicts. In Fig. 4 (**c–k**) and (**n**),
data are plotted as mean ± SEM and *p*-values calculated using unpaired two tailed *t*-
tests. Schematics in Fig. 4 (**a**) (**d**) (**i**) (**l**) and (**o**) were created with BioRender.com
released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0
International license. Source data are provided as a Source Data file.

011760-06), AQR (Dharmacon ON-TARGETplus SMARTpool, #L-
022214), TOP2A (Dharmacon ON-TARGETplus SMARTpool, #L-
004239) or scrambled non target siRNA (Dharmacon ON-TARGETplus
SMARTpool, #D-001810) as indicated. Cells were harvested 48 h after
transfection and processed as needed.

0.25 × 10⁶ cells/well were seeded in a six-well plates and treated
with the respective siRNA as described above, 24 h post knock-down,
cells were transfected with 2 μg of the HA-RAD52^{WT} or HA-RAD52^{ΔC}
over-expressing or HA-control plasmid (see below) using Lipofecta-
mine 3000 (Invitrogen, #L3000015).

Plasmid constructs
The plasmids used to express HA-RAD52^{WT} and HA-RAD52^{ΔC} were
derivatives of pcDNA3.1(+)-N-HA. The gblocks corresponding to full
length RAD52 and RAD52 (Δ302-410 amino acids) were cloned into the
pcDNA3.1(+)-N-HA backbone using KpnI/NotI restriction enzymes and
the plasmid was confirmed by sanger sequencing.

Immunoprecipitation and MASS spectrometry Analysis
HEK293T cells were transfected with the respective plasmids as per the
experimental conditions mentioned. Post transfection, the cells were
washed with ice cold PBS and resuspended in Lysis buffer (50 mM Tris-
HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 1X Protease Inhibitor)
and incubate for 20 min on rotor at 4 °C. After 4 cycles of water-bath
sonication for a total of 6 min, lysates were centrifugation for 10 min at
15000 x g at 4 °C. Immunoprecipitation was performed on the super-
natants using Pierce Anti-HA Magnetic Beads (Thermo Scientific,
#88836) overnight at 4 °C on the rotor. Beads were extensively washed in
the lysis buffer and stored at -80 °C, until ready for mass spectrometry.

A fraction of the beads was processed for immunoblotting to
confirm the pull-down. The beads were denatured and eluted in LDS
Non-Reducing Sample Buffer (Thermo Scientific, #84788). by boiling
for 5–10 min. Proteins were separated on 4–12% acrylamide SDS-PAGE,
transferred on Nitrocellulose membrane and detected with the indi-
cated antibodies described in the table and ECL reagents.

Sample preparation and mass spec analysis was carried out by
Poochon Scientific (Frederick, Maryland). Post IP, in order to precipitate
the protein, beads from 3 independent replicates were treated with 50 μl
of 2% SDS, heated at 95 °C for 10 min and centrifuged. Supernatant was
processed for trypsin digestion as per SOP-PS-6003 (Standard Opera-
tion of Procedure for in Solution Digestion). The digested peptide
mixture was then concentrated and desalted using SPN columns as per
SOP- PS-6005 (Standard Operation of Procedure for Desalting Digested
Peptides). Reconstituted desalted peptides in 30 μl of 0.1% formic acid.
12 μl of peptides was analyzed by 110 min LC/MS/MS run. The LC/MS/MS
analysis of samples were carried out using a Thermo Scientific Orbitrap
Exploris 240 Mass Spectrometer and a Thermo Scientific Dionex

UltiMate 3000 RSLCnano System. Peptide mixture from each sample
was loaded onto a peptide trap cartridge at a flow rate of 5 μL/min. The
trapped peptides were eluted onto a reversed-phase EasySpray C18
column (Thermo Scientific) using a linear gradient of acetonitrile
(3–36%) in 0.1% formic acid. The elution duration was 110 min at a flow
rate of 0.3 μl/min. Eluted peptides from the EasySpray column were
ionized and sprayed into the mass spectrometer, using a Nano-
EasySpray Ion Source (Thermo Scientific) under the following settings:
spray voltage, 1.6 kV, Capillary temperature, 275 °C. Raw data file
acquired from each sample was searched against human protein
sequences database and target protein sequences provided by the client
using the Proteome Discoverer 2.4 software (Thermo Scientific, San
Jose, CA) based on the SEQUEST algorithm. Carbamidomethylation
(+ 57.021 Da) of cysteines was fixed modification, and Oxidation Met
and Deamidation Q/N-deamidated (+ 0.98402 Da) were set as dynamic
modifications. The minimum peptide length was specified to be five
amino acids. The precursor mass tolerance was set to 15 ppm, whereas
fragment mass tolerance was set to 0.05 Da. The maximum false peptide
discovery rate was specified as 0.01. The resulting Proteome Discoverer
Report contains all assembled proteins with peptides sequences and
peptide spectrum match counts (PSM#).

Protein quantification/normalization used the normalized spec-
tral abundance factors (NSAFs) method to calculate the protein rela-
tive abundance[66,67]. NSAF normalization was carried out as follow.
NSAFs were calculated as follows:

$$NSAF_N = (SN/LN) / \left(\sum ni = 1 Si/Li \right) \tag{1}$$

Where N is the protein index; SN (PSM#) is the number of peptide
spectra matched to the protein; LN is the length of protein N (number
of amino acid residues); and n is the total number of proteins in the
input database (proteome profile for one cell sample). Protein
enrichment was calculated by comparing fold change between the
sample pull down and the HA-tag control.

Pathway analysis was carried out using Gene Ontology (GO)
software[68,69]. Functional protein interaction network analysis was
performed using interaction data from the STRING database[70]. Only
interactions with a score >0.15 are represented in the networks.

Co-Immunoprecipitation assay
To detect endogenous RAD52 interacting proteins, HeLa cells were
seeded in a 10 cm dish for 24 h. The cells were then washed with ice
cold PBS and resuspended in Lysis buffer (50 mM Tris-HCl pH 7.5,
150 mM NaCl, 1 mM EDTA, 1% NP-40, 1X Protease Inhibitor) and incu-
bate for 20 min on rotor at 4 °C. After 5 cycles of water-bath sonication
for a total of 10 min, lysates were centrifugation for 10 min at 15000 x g
at 4 °C. The lysate was treated with Benzonase nuclease (Millipore-

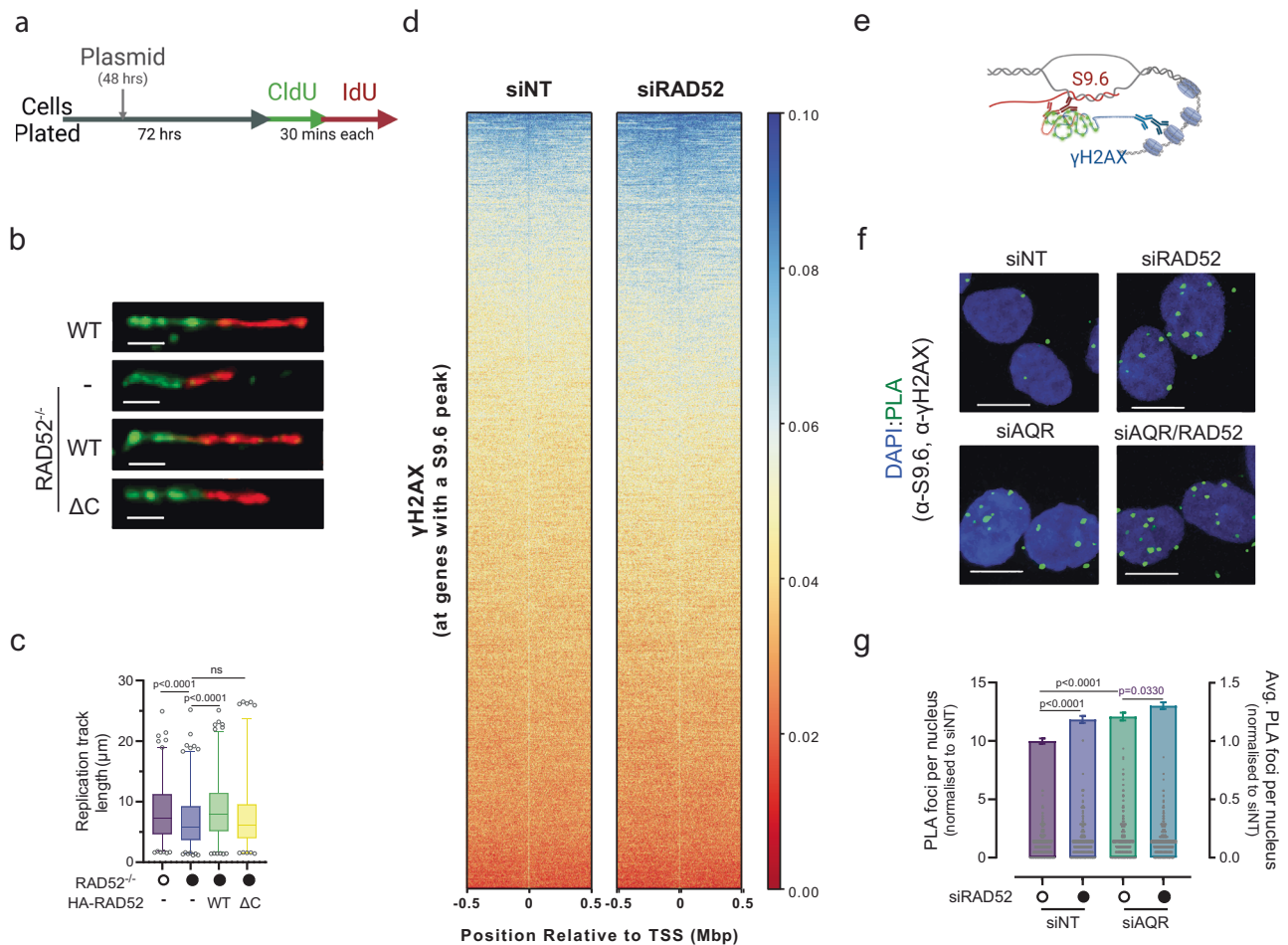**Fig. 5 | Loss of RAD52 causes replication stress and increased DNA damage.**
a Schematic representation of DNA fiber assay performed in HCT116 wild type (WT) and RAD52 knockout cells (RAD52$^{-/-}$) cells with plasmid overexpression of either RAD52^{WT} or RAD52$^{\Delta C}$ followed by incubation with 5-Chloro-2′-deoxyuridine (CldU) and 5-iodo-2′-deoxyuridine (IdU) for 30 min each to label nascent DNA.
b Representative images of DNA fiber images in HCT116 WT and *RAD52*$^{-/-}$ cells with overexpression of either RAD52^{WT} or RAD52$^{\Delta C}$ (Scale bar 2 μM). (**c**) Measurement of DNA fiber lengths across stated conditions described in (**b**) to measure replication rates. Data plotted as box and whiskers. Boxes extend from the 25th to 75th percentiles, with the median displayed as a line. The whiskers mark the minimum (1 percentile) and maximum (99th percentile). The data presented shows ≥100 DNA fibers from 3 biological replicates; *p*-values calculated using unpaired two tailed *t*-tests. **d** Heat map of the intensity of γH2AX ChIP signals (siNT and siRAD52 transfected HeLa cells) at genes that have a detectable R-loop peak as determined in Supplementary Fig. 6b. The γH2AX occupancy is displayed relative to the TSS ± 0.5 Mb. **e** Schematic representation of PLA to visualize proximity of S9.6 and γH2AX. **f** Representative images of the nuclear PLA foci (α-S9.6: α-γH2AX) in siNT (control), siRAD52 and siAQR transfected HeLa cells (Scale bar 10 μM).
**g** Quantitative analysis of nuclear PLA foci across stated conditions described in (**f**). Data are plotted as mean ± SEM. The data presented shows ≥ 500 nuclei from 3 biological replicates; *p*-values calculated using unpaired two tailed *t*-tests. Schematics in Fig. 5 (**a**) and (**e**) were created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license. Source data are provided as a Source Data file.

Sigma Novagen, #707463) at a concentration of 25U per 1 ml on the rotor for 1 h at room temperature. Immunoprecipitation was performed on the supernatants using α-RAD52 conjugated agarose beads (Santa Cruz Biotech, #sc-365341 AC) overnight at 4 °C on the rotor. A control immunoprecipitation was carried out using normal mouse IgG conjugated agarose beads (Santa Cruz Biotech, #sc-2343). Beads were extensively washed in the lysis buffer and processed for immunoblotting to confirm the pull-down. The beads were denatured and eluted in LDS Non-Reducing Sample Buffer (Thermo Scientific, #84788). by boiling for 5–10 min. Proteins were separated on 4–12% acrylamide SDS-PAGE, transferred on PVDF membrane and detected with the indicated antibodies and ECL reagents.

Proximity ligation assay
Cells were seeded on poly-L-lysine-coated coverslips and reverse transfected with indicated siRNAs on the same day. 48 h after transfection, cells were washed with PBS and pre-extracted with 0.25%

TritonX-100 for 5 min on ice. Cells were fixed with 4% paraformaldehyde for 20 min and then washed with PBS twice for 5 min. Cells were incubated with 100% methanol for 30 s and then washed with PBS three times for 5 min. Cells were blocked with blocking solution overnight at 4 °C provided in the PLA kit. Coverslips were incubated with primary antibodies (Supplementary Table 1) diluted in antibody diluent for 1 h at room temperature (RT). Next, coverslips were incubated with mouse/rabbit secondary probes Duolink® In Situ PLA® Probe Anti-Rabbit PLUS and Duolink® In Situ PLA® Probe Anti-Mouse MINUS. Proximity ligation was performed using either Duolink In Situ Red Kit Mouse/Rabbit (Millipore-Sigma, #DUO92008), Duolink® In Situ Detection Reagents FarRed (Millipore-Sigma, #DUO92013), and Duolink® In Situ Detection Reagents Green (Millipore-Sigma, DUO92014) according to the manufacturer's protocol. The oligonucleotides and antibody-nucleic acid conjugates used were those provided in the Millipore-Sigma PLA kit. For EdU (5-ethynyl-2′-deoxyuridine) staining in PLA experiments, cells were incubated with
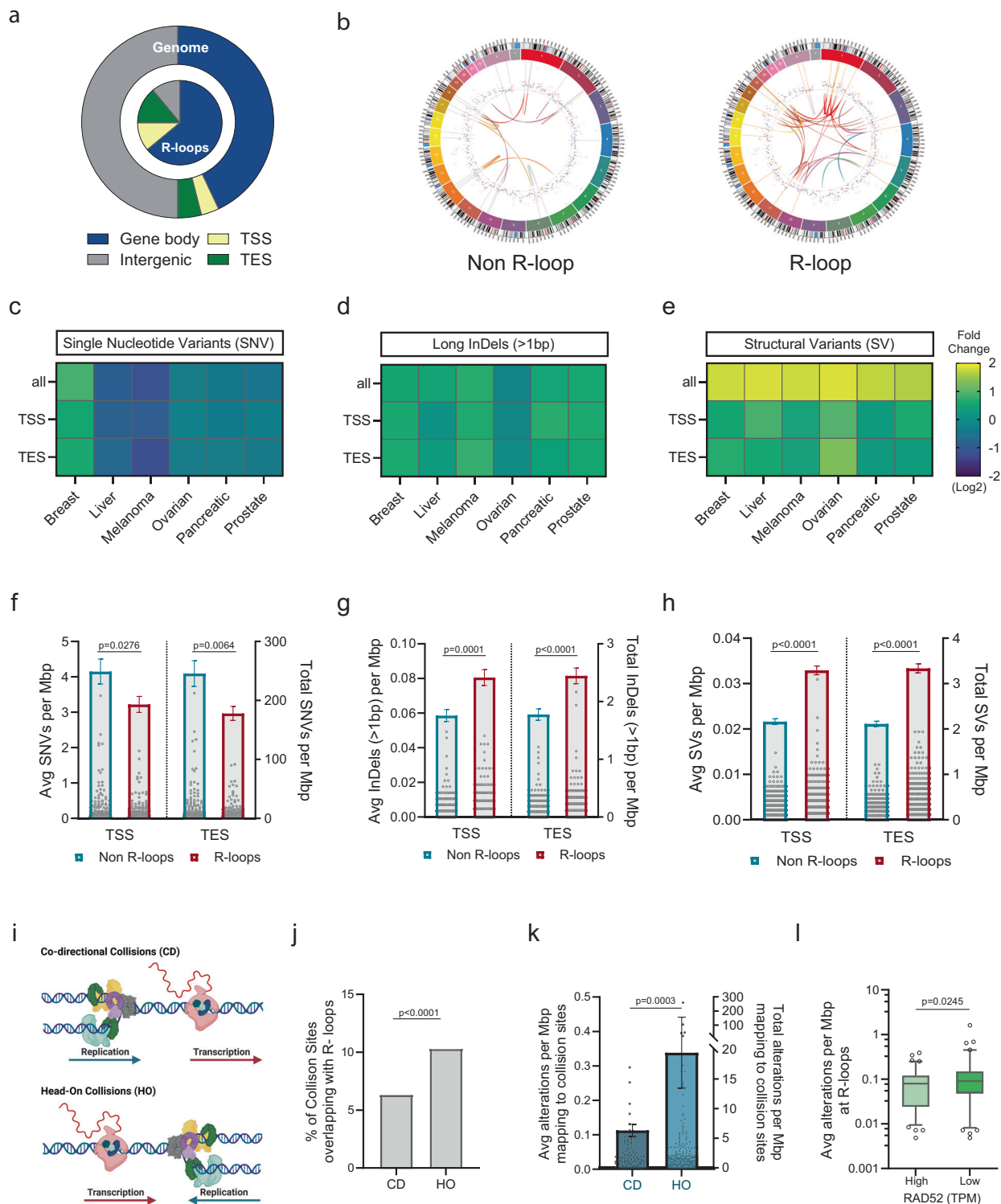
10 μM EdU for 30 min before pre-extraction step. EdU detection was performed via Click-IT reaction mixture provided in the kit (Thermo Scientific, #C10086) according to manufacturer's instructions. Stained cells were mounted with mounting medium containing DAPI. Samples were visualized on the Nikon Spinning disk confocal microscope at 60X, and images were collected and then analyzed with the Nikon Elements AR Analysis Explorer (version 5.21.03).

Immunofluorescence

For R-loop staining, Experiments were performed similar to reported procedures[71,72] with details as follows. Cells were fixed with ice cold methanol for 10 min and permeabilized with acetone for 3 min. Cells were washed 3 times with ice cold PBS (5 min each). For RNase III treated samples: cells were incubated with RNase III enzyme (New England Biolabs, #M0245S) with a dilution of 1:200 in 1X RNase III buffer supplemented with manganese chloride @37 °C for 30 min. For RNase H treated samples: cells were incubated with RNase H enzyme (New England Biolabs, #M0297L) with a dilution of 1:50 in 1X RNase H buffer. Cells were incubated in RNase H enzyme @37 °C for 2 h. After incubation, cells were washed with cold PBS for 5 min (3 times). Cells were incubated in blocking buffer (3% BSA, 1% goat serum 0.1% Triton in 4X SSC buffer) overnight at 4 C. Cells were incubated with the S9.6

Fig. 6 | Increased mutational burden and genomic instability associated with R-loops were observed in human tumor samples. a The genomic distribution of the consensus R-loop dataset as identified in Supplementary Fig. 7b. Various genomic regions are color coded according to the labels on the bottom. The expected distribution in case peaks were randomly positioned in the genome is shown for comparison. TTS and TES are significantly enriched in the R-loop dataset ($P < 0.001$) as determined by the Fisher's exact test. **b** Circos plots showing structural variations and genomic alterations caused by breakpoints enriched in R-loop (right) forming regions versus non-R-loop regions (left). **c−e** Genomic windows depicting the frequencies of single nucleotide variants (SNV-left), long InDels > 1 bp (middle) and structural variants (SV-right), analyzed at R-loop vs non-R-loop across various cancer types. The horizontal coordinate represents different types of cancers and vertical coordinates represents coverage at all genomic regions, TSS and TES. Data is quantified by log fold change between mutational burden at R-loop versus non-R-loop regions. **f−h** Quantification of the average number of SNVs, Long indels, SVs per Mb of genome at TSS and TES in R-loop versus non-R-loop forming regions. Data are plotted as mean ± SEM; p-values calculated using unpaired two tailed t-tests. **i** Schematic to show the two types of TRCs: co-directional collisions (top) and Head on collision (bottom). The schematic illustration was created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license. **j** Quantification of the percentage of collisions occur at R-loop sites in terms of co-directional collisions and head-on collisions. Data are plotted as a bar graph with absolute percentage. (Fisher's exact test). **k** Quantification of the comparison of average number of alterations per Mb of genome which are mapped to collision sites between CD and HO. Data are plotted as mean ± SEM. p-values were calculated by two-sided non-parametric Mann–Whitney test. **l** Quantification of the comparison of average number of alterations per Mb of genome at R-loop sites between tumors with high and low expression of RAD52. Tumors were categorized as expressing low (RAD52 low; bottom quartile) or high levels of RAD52 mRNA (RAD52 high; top quartile). Data plotted as box and whiskers. Boxes extend from the 25th to 75th percentiles, with the median displayed as a line. The whiskers mark the minimum (5th percentile) and maximum (95th percentile). ($n = 95$ (RAD52 high), $n = 94$ (RAD52 low)); p-values calculated using unpaired two tailed t-tests. Source data are provided as a Source Data file.

primary antibody (1:500, Millipore, #MABE1095) diluted in blocking buffer for 2 h at RT and washed 3 times with 4X SSC buffer for 5 min each. Cells were incubated with the Alexa Fluor Plus 488 secondary antibody (1:1500, Invitrogen, #A48255) diluted in blocking buffer for 1 h at RT (dark storage) and washed 3 times with 4X SSC buffer for 5 min each. For mitochondrial staining, cells were incubated with 250 nM MitoTracker Deep Red FM probe (Thermo Scientific, #M22426) for 30 min prior to the fixation step. Cells were mounted with mounting medium containing DAPI (Millipore-Sigma, #DUO82040) for 30 min at RT (dark storage) and the slides were stored at 4 °C. For counter staining the nucleolus, the nucleolin antibody (Cell Signaling, #14574; 1:1000 dilution) was combined with the S9.6 antibody and staining was carried out as described above, The Alexa Fluor Plus 555 secondary antibody (1:1500, Invitrogen, #A32732) was used.

For γH2AX staining, cells were fixed with 4% paraformaldehyde (EMS, #15710) for 20 min and washed with 1X PBS twice for 5 min each. Cells were permeabilized with 0.5% Triton X-100 for 10 min at RT and washed with 1X PBS three times for 5 min each. Cells were incubated in blocking buffer (3% BSA, 0.1% Triton in 1X PBS buffer) overnight at 4 C. Primary antibody incubation was performed with mouse monoclonal anti-phospho-H2A.X (Ser139) Antibody (1:1500) (Millipore-Sigma, #05-636) for 2 h at RT. Cells were washed three times with 0.1% TritonX-100 in 1X PBS for 5 min each. Secondary antibody incubation was performed with Goat anti-mouse Alexa Fluor Plus 488 (Invitrogen, #A32723) for 1 h in dark at RT. Cells were washed three times with 0.1% TritonX-100 in 1X PBS for 5 min each.

Samples were visualized on the Nikon Spinning disk confocal microscope at 60X, and images were collected and then analyzed with the Nikon Elements AR Analysis Explorer (version 5.21.03).

Dot blot

Total nucleic acid was extracted using the DNeasy Blood and Tissue Kit (Qiagen, #69504) and RNA:DNA hybrids were detected and quantified by dot blot assay. Samples were spotted on Amersham Hybond-N+ membrane (Cytiva, #RPN119B) in duplicates using the Bio-Dot Apparatus (BioRad, #1706545) and vacuum suction, dried and UV cross-linked. For the RNase H treatment, the genomic DNA was incubated in with the enzyme (New England Biolabs, #M0297) at a concentration of 10U of RNase H /μg of DNA at 37 °C for 20 h prior to spotting. Blots were blocked with 5% nonfat dried milk in TBST, then incubated overnight at 4 °C with an anti-RNA:DNA hybrid antibody S9.6 against RNA:DNA hybrids (Millipore, #05-636, 1:500 dilution) and double-stranded DNA (Novus Biologicals,#NBP3-07302,1:500 dilution) in TBST. Blots were washed 3 times for 15 min each in TBST, incubated in

either α-mouse IgG, HRP-linked Antibody (Cell Signaling, #7076; 1:5000 dilution) or α-rabbit IgG, HRP-linked Antibody (Cell Signaling, #7074; 1:5000 dilution) respectively for 1 hr at room temperature. Images were taken after incubation with SuperSignal West Pico PLUS (Fisher Pierce, PI34578).

Western Blotting

Cells were collected by trypsinization, lysed on ice in RIPA buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% SDS, 1% NP-40, 1% sodium deoxycholate) supplemented with 1 tablet/10 ml lysis buffer of cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche, #11-873-580-001), 1 tablet/10 ml lysis buffer of phosphatase inhibitor cocktail PhosSTOP (Millipore-Sigma, #4906845001) and 10 mM PMSF for 30 min, sonicated, and clarified by centrifugation for 20 min at 10,000 RPM at 4 °C. The supernatant was quantified using the Pierce BCA protocol (Thermo Scientific, #23225). Equivalent amounts of proteins were separated by SDS−PAGE and transferred onto a nitrocellulose membrane. Membranes were blocked in 5% milk in TBST (137 mM NaCl, 2.7 mM KCl, 19 mM Tris-Base and 0.05% Tween-20) for at least 1 h at room temperature. Incubation with primary antibodies was performed overnight at 4 °C. Membranes were washed in TBST and incubated with HRP-conjugated secondary antibodies for 1 h at room temperature, and developed with Pierce ECL (Thermo Scientific, #32106). The primary antibodies used for Western blotting included α-RAD52 (Santa Cruz Biotech, #sc-365341; 1:500 dilution), α-GAPDH (Abcam, #ab8245c; 1:1000 dilution), α-AQR (Bethyl Laboratories Inc, #A302547A; 1:500 dilution), α-Lamin A/C (Cell Signaling, #4777; 1:1000 dilution), α-HA (Santa Cruz Biotech, #sc-7392; 1:1000 dilution), α-Vinculin (Cell Signaling, #13901; 1:1000 dilution), α-TOP2A (Santa Cruz Biotech, #sc-365916; 1:500 dilution), α-βTubulin (GeneTex, #GTX107175; 1:1000 dilution), α-ATR (phospho Thr1989) (GeneTex, #GTX128145; 1:1000 dilution), α-ATM (phospho S1981) (Abcam, #ab81292; 1:2000 dilution), α-RPA2 (phospho S4/S8) (Abcam, #ab87277; 1:2000 dilution), α-RPA2 (phospho S33) (Bethyl Laboratories Inc, #A300246A; 1:1000 dilution), α-ATR (Santa Cruz Biotech, sc-1887; 1:500 dilution), α-ATM (Abcam, #ab78; 1:1000 dilution), α-RPA2 (Cell Signaling, #2208; 1:1000 dilution), α-RPA1 (Cell Signaling, #2267; 1:1000 dilution), α-Rbp1 CTD (Cell Signaling, #2629; 1:4000 dilution) and α-SETX (Novus Biologicals, #NB100-57542; 1:1000 dilution). The secondary antibodies were α-mouse IgG, HRP-linked Antibody (Cell Signaling, #7076; 1:5000 dilution), α-rabbit IgG, HRP-linked Antibody (Cell Signaling, #7074; 1:5000 dilution), α-rat IgG, HRP-linked Antibody (Cell Signaling, #7077; 1:5000 dilution) or α-goat IgG (H + L), HRP-linked Antibody (Invitrogen, #PA1-28664; 1:5000 dilution).

Chromatin Immunoprecipitation

HeLa cells were transfected with the respective siRNAs as per the experimental conditions mentioned. 48 h post transfection, the cells were cross-linked with formaldehyde at 1% final concentration for 10 min at room temperature or cross-linked according to dual cross-linking protocol as described previously[73]. Chromatin immunoprecipitation (ChIP) assays were conducted using the Zymo-Spin ChIP kit (Zymo Research Corp., #D5210) following manufacturer's instructions. Sonication was performed at high power setting for 80 cycles (30 s ON, 30 s OFF) using a Bioruptor Plus (Diagenode Inc., Denville, NJ), yielding a modal fragment size of <600 bp. Antibodies used in ChIP assays included: α-RNA Pol II (Cell Signaling 2629S), α-γH2AX (Abcam, #ab2893), α-RAD52 (Santa Cruz Biotechnology, #sc-365341), or α-S9.6 (Kerafast, #ENH001) and normal mouse IgG (Millipore-Sigma, #12-371). Approximately 20 μg of chromatin was used in each ChIP assay with 5 μg of antibodies, or 100 μg of chromatin was used in each ChIP assay with 10 μg of antibodies. IgG negative control was included with each assay. DNA libraries were prepared by Zymo Research Epigenetics Services and were sequenced on a NovaSeq sequencer.

ChIP-seq analysis

ChIP sequencing reads were trimmed using Cutadapt[74] and aligned to the human reference genome (hg19) using BWA[75]. We applied read filtering to remove reads that were marked as duplicates (picard) ("Picard Toolkit." 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/; Broad Institute), reads that were not primary alignments, unmapped, mapped to multiple locations, or contained >4 mismatches (samtools)[76].

Deeptools[77] bamCoverage was used to create normalized bigWig files using CPM (counts per million) normalization. We further used deeptools computeMatrix, plotHeatmap and plotProfile for visualization of ChIP-Seq data at TSS and TTS.

MACS2 callpeak[78] was used to identify ChIP-Seq peaks relative to input data. To compare peaks from different conditions we retained peaks with P-values >2 (-log10) from the narrowPeak output files. The ChIPpeakAnno package[79] was used in R to create Venn diagrams of colocalized peaks in different conditions.

RNA-seq

HeLa cells were transfected with the respective siRNAs as per the experimental conditions mentioned. 48 h post transfection, samples were sent to Zymo Research for Total RNA-Seq Service. Total RNA-Seq libraries were constructed from 500 ng of total RNA. Libraries were prepared using the Zymo-Seq RiboFree Total RNA Library Prep Kit (Zymo Research Corp., #R3000) according to the manufacturer's instructions. Briefly, RNA was reverse transcribed into cDNA, which was followed by ribosomal RNA depletion. After that partial P7 adapter sequence was ligated at 3' end of cDNAs, followed by second strand synthesis and partial P5 adapter ligation to 5' end of the double stranded DNAs. Lastly, libraries were amplified to incorporate full length adapters under the following conditions: initial denaturation at 95 °C for 10 min; 10–16 cycles of denaturation at 95 °C for 30 sec, annealing at 60 °C for 30 sec, and extension at 72 °C for 60 sec; and final extension at 72 °C for 7 min. Successful library construction was confirmed with Agilent's D1000 ScreenTape Assay on TapeStation. RNA-Seq libraries were sequenced on an Illumina NovaSeq to a sequencing depth of at least 30 million read pairs (150 bp paired-end sequencing) per sample.

RNA-seq analysis

RNA-Seq reads were aligned to GRCh37 human genome using STAR RNA-Seq aligner[80], and then reads from transcripts were counted using GenomicAlignments package in Bioconductor[81,82]. Fold changes between siRAD52 and siNT cells were obtained from DESeq2[83], which was performed using raw read counts in siRAD52/siNT as pairs.

P-values obtained from DESeq2 were corrected for multiple testing using Benjamini and Hochberg method.

According to Chip-seq RNA Pol II and RAD52 data, we found 10% of peaks in RNA Pol II data overlapped with RAD52 peaks. Thus, we tried to tell whether the expression changes in genes near these overlapped peaks were influenced by RAD52. Here, we retrieved genes overlapped with these peaks covering transcription start sites for the analysis. We first examined gene expression levels for the overlapped genes according to DESeq2 normalized counts. Most of genes did not display significant change between RAD52 and NT cells. Next, we compared gene fold changes between RAD52 and NT cells for genes overlapped with RAD52 peaks and not overlapped with RAD52 peaks. However, by examining the relationship between averaged expression with fold changes included in DESeq2 output, we observed higher variance of fold changes for lowly expressed genes, therefore, we decided to exclude lowly expressed genes with averaged expression level <50 for the comparison. Student's t-test was used to test the differences of gene fold changes. At last, pathway enrichment analysis was performed for the genes overlapped with RAD52 peaks using ClusterProfiler[84].

DSB repair assays

The SSA reporter plasmid hprtSAGFP (Addgene, #41594) and the HR reporter plasmid pDRGFP (Addgene, #26475) were gifts from Maria Jasin.

HCT116 WT or RAD52[-/-] cells were transfected with 0.5 μg DR-GFP or SA-GFP and 1.5 μg pCBASceI (Addgene, #26477) using Lipofectamine 3000 as described in the cell culture and transfection section. 72 h later, the cells were harvested, and percentages of GFP-positive cells per 100,000 cells were determined by flow cytometry (HTFC Screening System, IntelliCyt). For each experiment, the percentage of GFP positive cells in the empty vector control was subtracted from the I-SceI-transfected cells. Flow cytometry data were analyzed using BD FlowJo (v.10.6.2).

DNA fiber analysis

Cells were seeded onto 6 well plate and allowed to grow for 24 h. Cells were sequentially labeled with thymidine analogs: 25 μM CldU (5-Chloro-2′-deoxyuridine) and 250 μM IdU (iododeoxyuridine) for 30 min each. The reaction was terminated by addition of ice-cold PBS and cells were trypsinized. Cells were lysed with lysis buffer (50 mM EDTA pH 8.0, 0.5% SDS, 200 mM Tris-Cl pH 7.5) on a clean slide and incubated for 5–7 min. After cell lysis, DNA was spread on glass slides and slides were tilted at an angle of 25°. Slides were air dried. Cells were fixed in methanol/acetic acid (3:1) for 10 min, denatured with 2.5 M HCl for 1 h. Cells were blocked with 5% BSA in PBST (10 mM sodium phosphate, 0.15 M NaCl, 0.1% Tween™ 20 buffer at pH 7.5) for 1 h and stained with primary antibody Anti-BrdU (5-bromo-2′-deoxyuridine) monoclonal antibody from rat (abcam #ab6326; 1:50) and Anti-BrdU (5-bromo-2′-deoxyuridine) monoclonal antibody from mouse (BD Biosciences #347580; 1:100) for 2 h. Slides were washed 3 times with PBST. Slides were incubated with Chicken anti-rat Alexa Fluor 488 (Invitrogen, #A21470; 1:300) and Rabbit anti-Mouse Alexa fluor 594-conjugated (Invitrogen, #A11062; 1:200) for 1 h. Slides were washed 3 times with PBST and coverslips were mounted on the slides with the mounting medium. DNA fibers were visualized on the Nikon Spinning disk confocal microscope at 60X, and images were collected and then analyzed with ImageJ.

Cell cycle analysis

HeLa cells were transfected as described and harvested 48 h post transfection. Cells were then washed with PBS and fixed in 1 ml cold 70% ethanol for at least 30 min on ice. Cells were pelleted and washed with PBS. The cells were resuspended in the staining solution (0.1% TrittonX-100, 200 μg/mL RNase A and 50 μg/mL propidium iodide in

PBS) and incubated for 15 min at 37 °C in the dark. 50,000 cells per condition were analyzed by flow cytometry using the LSR Fortessa instrument (BD Biosciences). BD FACSDiva software was used with the BD Biosciences LSR Fortessa Analyzer for flow cytometry data acquisition. Flow cytometry data were analyzed using BD FlowJo (v.10.6.2).

R-loop consensus analysis

To identify consensus regions of R-loop in the human genome, we sourced 23 published R-loop bigwig files from the UCSC genome browser[52]. These files were subsequently converted to bed format utilizing the 'bigWigToBedGraph' tool (https://genome.ucsc.edu/goldenPath/help/bigWig.html). Strands were merged, and replicates were consolidated with the 'bedtools unionbedg' function[85]. Broad peaks were then identified using the 'macs2 bdgpeakcall' function[86]. The average intensity score for the bed regions was ascertained with our custom 'calculate_mean_intensity_score.pl' script. On assessing the correlation of peak scores across the 23 tracks, five tracks exhibited discrepancies and were consequently excluded from further analyses. The remaining 18 bed files were merged using the 'bedtools multiinter' function, including all R-loop regions without any filters. These regions were subtracted from the entire human hg19 reference genome, to obtain genomic regions devoid of any R-loop. To pinpoint the consensus R-loop region, we filtered out peak scores below 200, merged regions from varying tracks with the 'bedtools multiinter' function, and further refined this merged file to capture regions with a minimum of five overlapping tracks and a maximum length of 5000 bp. The resulting consensus R-loop region and control areas were employed to analyze R-loop enrichment across different genomic sections and to evaluate breakpoint density within R-loop regions subsequently. All scripts used in processing and the derived consensus R-loop and control regions can be obtained at https://github.com/ipstone/rloop_genome_instability.

Consensus overlap analysis with genomic mutational signatures

The mutational calls (SNVs, indels and structural variants) were downloaded from PCAWG (264 liver, 239 pancreatic, 189 prostate, 71 ovarian and 70 melanoma tumors)[56], ICGC/BRCA-EU project (320 ER+ breast tumors)[55] and TCGA cohorts (377 liver, 185 pancreatic, 500 prostate, 587 ovarian, 470 melanoma, and 713 ER+ breast tumors)[54]. The gene annotations for transcription start site (TSS) and transcription end site (TES) were downloaded from GENCODE for 'genecode.v19.annotation.gtf' data file. A window of 1 kb (+/-) TSS and TES were added to these sites to intersected with the previously prepared R-loop positive regions as well as the R-loop negative regions. The resulting bed files was used to intersect with the genomics mutation data to calculate the density of SNVs, indels and structural variant breakpoints on these regions. For SNVs and indels, the count of mutations were calculated using the 'SigProfilerMatrixGenerator' function from SigProfiler package (https://github.com/AlexandrovLab/SigProfilerMatrixGenerator), these numbers were further divided by the genomic region length to get mutation densities for the respective regions. Similarly, for the structural variant breakpoints, the downloaded structure variants data were intersected with the respective R-loop positive and negative genomic regions using 'bedtools pairtobed' function. After intersection, the SV breakpoints densities were calculated similarly as SNVs and indels. The density calculations and comparisons were all done in the R statistical software and visualized in Prism.

For the head-on (HO) collision and the co-directional (CD) collision double-strand break density comparison, the HO/CD regions published in the afore mentioned paper[44] were downloaded and intersected with R loop consensus positive regions and negative regions. A +/- 1 kb window was added to the R-loop consensus region before intersection. The double strand breaks density (indels and SVs) was calculated similarly as for the TSS/TES regions above. All the analysis code and accompanying input files are available at the following github repository: https://github.com/ipstone/rloop_genome_instability.

Evaluating association between RAD52 RNA expression and genomic alterations at R-loop

To investigate whether down-regulation of RAD52 is associated with increased alterations in R-loops, we performed an analysis of tumor RNA-seq data in the PCAWG cohort. Specifically, we downloaded PCAWG RNA-seq data from ICGC and extracted RAD52 Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values, which were further transformed into Transcripts Per Million (TPM) values. To eliminate confounding factors stemming from homologous recombination deficiency (HRD), HRD cases were excluded based on the results defined by Nguyen et al[87].

Utilizing the RAD52 TPM values, tumors were categorized into RAD52-high and RAD52-low groups, representing the top 25% and bottom 25% of samples.

To compare structural variants (SVs) and Indels in R-loop regions, we calculated the averaged SVs and Indels per Mbp for each sample. This calculation was performed using the formula: $(N_{sv} + N_{indel})/$R-loop segment size $* 1{,}000{,}000$. Finally, an unpaired t-test with Welch's correction was used to compare the levels of SVs and Indels in R-loop regions for RAD52-high and RAD52-low tumors.

Image processing and data analysis

For PLA and S9.6, γH2AX experiments, slides were imaged at 60X (immersion oil) with Nikon spinning disk confocal microscope. PLA foci per nucleus, S9.6 and γH2AX foci per nucleus were calculated using Nikon Elements AR Analysis Explorer (version 5.21.03), where DAPI was used as a mask for the nucleus. The number of PLA foci, S9.6 foci, γH2AX foci was counted for each DAPI to obtain the average number of gH2AX foci in each condition.

Statistical analysis

Statistical analysis was carried out by unpaired two-tailed t-test (unless stated otherwise) using GraphPad Prism Version 10.2.1 for Windows (GraphPad Software, San Diego, CA, USA). All values are expressed as mean ± standard error of the mean (SEM). p-values <0.05 were considered statistically significant. ns: non-significant.

Reporting summary

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

Data availability

The ChIP-seq data generated in this study have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE266483. The RNA -seq data generated in this study have been deposited with the GEO series accession number GSE272608. The mass spectrometry proteomics data are described in Supplementary Table 2 and have been deposited to the ProteomeXchange Consortium via the PRIDE[88] partner repository with the dataset identifier PXD054611. RNA/DNA interactome[42] was obtained from the previously published dataset PXD002960. TOP2A ChIP-seq data was obtained from the previously published dataset GSE181450. Requests for materials and/or questions regarding any of the constructs, cell lines, microscopy results or other data described should be addressed to the corresponding authors. Source data are provided with this paper.

Code availability

All scripts used in processing and the derived consensus R-loop and control regions are available at https://github.com/ipstone/rloop_genome_instability.

References

1. Gómez-González, B. & Aguilera, A. Transcription-mediated replication hindrance: a major driver of genome instability. *Genes Dev.* **33**, 1008–1026 (2019).
2. Goehring, L., Huang, T. T. & Smith, D. J. Transcription–replication conflicts as a source of genome instability. *Annu. Rev. Genet.* **57**, 157–179 (2023).
3. Thomas, M., White, R. L. & Davis, R. W. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc. Natl Acad. Sci. USA* **73**, 2294–2298 (1976).
4. 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).
5. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-loops as cellular regulators and genomic threats. *Mol. Cell* **73**, 398–411 (2019).
6. Castillo-Guzman, D. & Chédin, F. Defining R-loop classes and their contributions to genome instability. *DNA Repair* **106**, 103182 (2021).
7. Brickner, J. R., Garzon, J. L. & Cimprich, K. A. Walking a tightrope: the complex balancing act of R-loops in genome stability. *Mol. Cell* **82**, 2267–2297 (2022).
8. Domínguez-Sánchez, M. S., Barroso, S., Gómez-González, B., Luna, R. & Aguilera, A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7**, e1002386 (2011).
9. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
10. Bayona-Feliu, A. & Aguilera, A. The role of chromatin at transcription-replication conflicts as a genome safeguard. *Biochem Soc. Trans.* **49**, 2727–2736 (2021).
11. Tuduri, S. et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **11**, 1315–1324 (2009).
12. Promonet, A. et al. Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nat. Commun.* **11**, 3940 (2020).
13. Hyjek, M., Figiel, M. & Nowotny, M. RNases H: structure and mechanism. *DNA Repair* **84**, 102672 (2019).
14. Freudenreich, C. H. R-loops: targets for nuclease cleavage and repeat instability. *Curr. Genet* **64**, 789–794 (2018).
15. Sollier, J. et al. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56**, 777–785 (2014).
16. Yang, S., Winstone, L., Mondal, S. & Wu, Y. Helicases in R-loop formation and resolution. *J. Biol. Chem.* **299**, 105307 (2023).
17. 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).
18. Bhatia, V. et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
19. García-Rubio, M. L. et al. The fanconi anemia pathway protects genome integrity from R-loops. *PLoS Genet* **11**, e1005674 (2015).
20. Schwab, R. A. et al. The fanconi anemia pathway maintains genome stability by coordinating replication and transcription. *Mol. Cell* **60**, 351–361 (2015).
21. Chappidi, N. et al. Fork cleavage-religation cycle and active transcription mediate replication restart after fork Stalling at co-transcriptional R-loops. *Mol. Cell* **77**, 528–541.e528 (2020).
22. Ohle, C. et al. Transient RNA-DNA hybrids are required for efficient double-strand break repair. *Cell* **167**, 1001–1013.e1007 (2016).
23. Aymard, F. et al. Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat. Struct. Mol. Biol.* **21**, 366–374 (2014).
24. Wei, L. et al. DNA damage during the G0/G1 phase triggers RNA-templated, cockayne syndrome B-dependent homologous recombination. *Proc. Natl Acad. Sci. USA* **112**, E3495–E3504 (2015).
25. Teng, Y. et al. ROS-induced R loops trigger a transcription-coupled but BRCA1/2-independent homologous recombination pathway through CSB. *Nat. Commun.* **9**, 4115 (2018).
26. Welty, S. et al. RAD52 is required for RNA-templated recombination repair in post-mitotic neurons. *J. Biol. Chem.* **293**, 1353–1362 (2018).
27. Yasuhara, T. et al. Human Rad52 promotes XPG-mediated R-loop pcessing to initiate transcription-associated homologous recombination repair. *Cell* **175**, 558–570.e511 (2018).
28. Davó-Martínez, C. et al. Different SWI/SNF complexes coordinately promote R-loop- and RAD52-dependent transcription-coupled homologous recombination. *Nucleic Acids Res.* **51**, 9055–9074 (2023).
29. Gottifredi, V. & Wiesmüller, L. Current understanding of RAD52 functions: fundamental and therapeutic insights. *Cancers* **12**, 705 (2020).
30. Feng, Z. et al. Rad52 inactivation is synthetically lethal with BRCA2 deficiency. *Proc. Natl Acad. Sci. USA* **108**, 686–691 (2011).
31. Lok, B. H., Carley, A. C., Tchang, B. & Powell, S. N. RAD52 inactivation is synthetically lethal with deficiencies in BRCA1 and PALB2 in addition to BRCA2 through RAD51-mediated homologous recombination. *Oncogene* **32**, 3552–3558 (2013).
32. Lok, B. H. & Powell, S. N. Molecular pathways: understanding the role of Rad52 in homologous recombination for therapeutic advancement. *Clin. Cancer Res.* **18**, 6400–6406 (2012).
33. Mazina, O. M., Keskin, H., Hanamshet, K., Storici, F. & Mazin, A. V. Rad52 inverse strand exchange drives RNA-templated DNA double-strand break repair. *Mol. Cell* **67**, 19–29.e13 (2017).
34. Keskin, H. et al. Transcript-RNA-templated DNA recombination and repair. *Nature* **515**, 436–439 (2014).
35. Liu, J., Meng, X. & Shen, Z. Association of human RAD52 protein with transcription factors. *Biochem. Biophys. Res. Commun.* **297**, 1191–1196 (2002).
36. Park, M. S., Ludwig, D. L., Stigger, E. & Lee, S. H. Physical interaction between human RAD52 and RPA is required for homologous recombination in mammalian cells. *J. Biol. Chem.* **271**, 18996–19000 (1996).
37. Plate, I. et al. Interaction with RPA is necessary for Rad52 repair center formation and for its mediator activity. *J. Biol. Chem.* **283**, 29077–29085 (2008).
38. Ma, C. J., Kwon, Y., Sung, P. & Greene, E. C. Human RAD52 interactions with replication protein A and the RAD51 presynaptic complex. *J. Biol. Chem.* **292**, 11702–11713 (2017).
39. Sugiyama, T., New, J. H. & Kowalczykowski, S. C. DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc. Natl Acad. Sci. USA* **95**, 6049–6054 (1998).
40. Chédin, F. Nascent connections: R-loops and chromatin patterning. *Trends Genet.* **32**, 828–838 (2016).
41. Boguslawski, S. J. et al. Characterization of monoclonal antibody to DNA · RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89**, 123–130 (1986).
42. 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).
43. 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).
44. 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.e719 (2017).

45. 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).

46. Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* **13**, 2633–2638 (1999).

47. Stark, J. M., Pierce, A. J., Oh, J., Pastink, A. & Jasin, M. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Mol. Cell Biol.* **24**, 9305–9316 (2004).

48. Dykhuizen, E. C. et al. BAF complexes facilitate decatenation of DNA by topoisomerase IIα. *Nature* **497**, 624–627 (2013).

49. Yeo, C. Q. X. et al. p53 maintains genomic stability by preventing interference between transcription and replication. *Cell Rep.* **15**, 132–146 (2016).

50. Tsai, S. et al. ARID1A regulates R-loop associated DNA replication stress. *PLoS Genet* **17**, e1009238 (2021).

51. Das, S. K. et al. MYC assembles and stimulates topoisomerases 1 and 2 in a "topoisome. *Mol. Cell* **82**, 140–158.e112 (2022).

52. Chédin, F., Hartono, S. R., Sanz, L. A. & Vanoosthuyse, V. Best practices for the visualization, mapping, and manipulation of R-loops. *EMBO J.* **40**, e106394 (2021).

53. 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).

54. Ellrott, K. et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst.* **6**, 271–281.e277 (2018).

55. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).

56. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).

57. Koh, G., Degasperi, A., Zou, X., Momen, S. & Nik-Zainal, S. Mutational signatures: emerging concepts, caveats and clinical applications. *Nat. Rev. Cancer* **21**, 619–637 (2021).

58. Petryk, N. et al. Replication landscape of the human genome. *Nat. Commun.* **7**, 10208 (2016).

59. Jalan M., Olsen K. S., Powell S. N. Emerging roles of RAD52 in genome maintenance. *Cancers* **11**, 1038 (2019).

60. Noe Gonzalez, M., Blears, D. & Svejstrup, J. Q. Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nat. Rev. Mol. Cell Biol.* **22**, 3–21 (2021).

61. Pommier, Y., Nussenzweig, A., Takeda, S. & Austin, C. Human topoisomerases and their roles in genome stability and organization. *Nat. Rev. Mol. Cell Biol.* **23**, 407–427 (2022).

62. Nitiss, J. L. DNA topoisomerase II and its growing repertoire of biological functions. *Nat. Rev. Cancer* **9**, 327–337 (2009).

63. Herrero-Ruiz, A. et al. Topoisomerase IIα represses transcription by enforcing promoter-proximal pausing. *Cell Rep.* **35**, 108977 (2021).

64. Bayona-Feliu, A. et al. The chromatin network helps prevent cancer-associated mutagenesis at transcription-replication conflicts. *Nat. Commun.* **14**, 6890 (2023).

65. Kan, Y., Batada, N. N. & Hendrickson, E. A. Human somatic cells deficient for RAD52 are impaired for viral integration and compromised for most aspects of homology-directed repair. *DNA Repair* **55**, 64–75 (2017).

66. Florens, L. et al. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40**, 303–311 (2006).

67. Paoletti, A. C. et al. Quantitative proteomic analysis of distinct mammalian mediator complexes using normalized spectral abundance factors. *Proc. Natl Acad. Sci. USA* **103**, 18928–18933 (2006).

68. Ashburner, M. et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

69. The Gene Ontology C. et al. The gene ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).

70. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–d613 (2019).

71. De Magis, A. et al. DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl Acad. Sci. USA* **116**, 816–825 (2019).

72. Miglietta, G., Marinello, J., Capranico, G. Immunofluorescence microscopy of G-quadruplexes and R-loops. In: *Methods in Enzymology* (eds Raney, K. D., Eoff, R. L., Byrd, A. K., Kendrick, S.) Ch. 6 (Academic Press, 2024).

73. Tian, B., Yang, J. & Brasier, A. R. Two-step cross-linking for analysis of protein-chromatin interactions. *Methods Mol. Biol.* **809**, 105–120 (2012).

74. Kechin, A., Boyarskikh, U., Kel, A. & Filipenko, M. cutPrimers: A new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* **24**, 1138–1143 (2017).

75. Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

76. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

77. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

78. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

79. Zhu, L. J. et al. ChIPpeakAnno: a bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinform.* **11**, 237 (2010).

80. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

81. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

82. Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

84. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).

85. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

86. Gaspar, J. M. Improved peak-calling with MACS2. *bioRxiv* 496521 https://doi.org/10.1101/496521 (2018).

87. Nguyen, L., Martens, J. W. M., Van Hoeck, A. & Cuppen, E. Pan-cancer landscape of homologous recombination deficiency. *Nat. Commun.* **11**, 5584 (2020).

88. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–d552 (2022).

Acknowledgements

Author contributions

M.J. and S.N.P. conceived and designed the studies. A.S., X.P., and N.W. made equal secondary contributions. M.J. performed experiments with help from A.S., E.S.B., S.A.S., N.M.D., J.G.A., K.S.A. N.W. analyzed the ChIP-seq data along with M.J., X.P., N.R. X.P. analyzed the human tumor data along with A.R., Y.Z., N.R., N.W. and M.J. C.K.Y.N. and P.S. helped in the bioinformatics analysis. M.D.B. helped in quantifying immuno-fluorescence images. S.F.B, J.S.R-F., N.R. and S.N.P. provided critical insights into experimental design and interpretation. M.J. and S.N.P. drafted the manuscript with help from A.S., N.M.D. and E.S.B. All of the authors commented on and edited the manuscript.

Competing interests

S.F.B. owns equity in, receives compensation from, and serves as a consultant for and serves on the Scientific Advisory Board (SAB) and Board of Directors (BOD) of Volastra Therapeutics Inc and he serves on the scientific advisory board of Meliora Therapeutics. J.S.R.-F. reports current employment at AstraZeneca and stocks in AstraZeneca, Repare Therapeutics, Paige.AI; J.S.R.-F. previously held a fiduciary role in Grupo Oncoclinicas and was a consultant with Goldman Sachs Merchant Banking, Bain Capital, Repare Therapeutics, Paige.AI, Volition Rx and MultiplexDx. S.N.P. reports consulting fees from the following companies outside the scope of this study: Varian Medical Systems, Philips Healthcare and AstraZeneca, as well as research funding from Philips Healthcare and Varian Medical Systems. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51784-x.

Correspondence and requests for materials should be addressed to Manisha Jalan or Simon N. Powell.

Peer review information *Nature Communications* thanks Stephan Hamperl, Li Lan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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