Break-induced replication is activated to repair R-loop-associated double-strand breaks in SETX-deficient cells

Tong Wu[1], Youhang Li[1], Linda Z. Shi[2], and Xiaohua Wu[1,*]

[1]Department of Molecular and Cell Biology, The Scripps Research Institute, La Jolla, CA 92037

[2]The Institute of Engineering in Medicine, University of California, San Diego, California 92093

[*]The Correspondence Author

Telephone: 858-784-7910, Fax: 858-784-7978, email: xiaohwu@scripps.edu

Running title: Hyper BIR in SETX-deficient cells

Key words: Senataxin (SETX), R-loop, double-strand break (DSB), Break-induced replication (BIR), homologous recombination (HR), end resection, template switching, PIF1, MRE11, XPF, RAD52

Abstract

The primary role of break-induced replication (BIR) is to repair single-ended double strand breaks (seDSBs) generated at broken replication forks and eroding telomeres. In this study, we demonstrated that when senataxin (SETX), an RNA/DNA helicase, is defective, hyperrecombination using the BIR mechanism is induced at R-loops/hybrids-accumulated double-ended DNA double strand breaks (deDSBs), suggesting a potential role of BIR in repair of R-loops/hybrids-associated deDSBs. Intriguingly, while loss of SETX initiates a non-canonical hyper end resection pathway requiring RAD52 and XPF, it also causes collisions of DNA repair synthesis with R-loops/hybrids, leading to ubiquitination of PCNA and recruitment of PIF1 to establish BIR at R-loops/hybrids-associated deDSBs. Strikingly, SETX-deficiency does not only induce hyper BIR but also leads to a drastic increase in template switching during BIR, uncovering an important mechanism causing genome instability through replicative chromosomal rearrangements due to SETX deficiency. Furthermore, SETX is synthetic lethal with PIF1, RAD52 and XPF. Collectively, our study sheds light on how R-loops/hybrids influence the utilization and fidelity of DSB repair pathways and offers new strategies for targeted treatment of SETX-deficient tumors.

Significance statement

R-loops are widespread, occupying 5-10 % of the human genome, and thus understanding the DSB repair mechanisms associated with R-loops is of great importance. Defect in senataxin (SETX), an RNA-DNA helicase, is associated with neurodegenerative disorders and tumorigenesis. Our study reveals that R-loop accumulation at DNA double strand breaks (DSBs) due to SETX deficiency triggers hyperrecombination through break-induced replication (BIR), accompanied by a drastic increase in template switching during BIR. This uncovers a significant source of replication-associated complex genome rearrangements (CGRs) linked to prevalent R-loops, providing new insights into cancer-associated genome instability. Furthermore, this work not only uncovers the repair mechanisms utilized at DSBs in the context of dynamic R-loop formation, but also opens new therapeutic opportunities for targeting SETX-deficient tumors.

Introduction

DNA double strand breaks (DSBs) represent the most dangerous form of DNA damage, which could lead to genome instability and, more severely, to cell death (1). DSBs can be repaired by two major pathways; homologous recombination (HR) and non-homologous end joining (NHEJ) (2). Homology-dependent HR serves as an error-free mechanism to repair DSB (3-5), whereas homology-independent NHEJ could be error-prone in some cases (6). While NHEJ remains active throughout the cell cycle, HR is activated when cells enter S and G2 phases when sister chromatids are available to serve as templates (7). Break-induced replication (BIR) is a subtype of HR, which operates at one-ended DSBs, such as collapsed replication forks and eroding telomeres (8-11). BIR and HR share similar initiation steps, including end resection and strand invasion, but diverge at the DNA synthesis step, where BIR exhibits a much longer DNA synthesis tract than HR (12). To achieve long-tract DNA synthesis, BIR employs a unique feature distinct from HR; BIR replication is dependent on POLD3/Pol32, a subunit of DNA polymerase δ, and helicase PIF1 (13-17). It is also noteworthy that BIR forks often are unstable, disassociating from the original templates and reinvading the same sites or dispersed repeated sequences elsewhere in the genome (18). This phenomenon is called template switching and can generate replication-associated complex chromosome rearrangements (18-20), which are frequently observed in the cancer genomes. While BIR-associated template switching has been studied in yeast (18, 21), little is known about template switching linked to BIR in mammalian cells and its contribution to cancer-like genome instability. In addition, it has been reported that the mutation frequency associated with BIR is significantly higher than that of S-phase DNA replication (22, 23). Given the highly mutagenic nature of BIR, delicate regulation pathways have been described to restrict BIR activation in yeast (24), but the mechanisms that regulate BIR usage in mammalian cells remain to be explored.

R-loops are three-stranded nucleic acid structures formed by extended annealing of nascent RNA to the DNA template, composing an RNA/DNA hybrid and displaced single-strand DNA (ssDNA). While R-loops possess important biological functions such as transcription regulation, DNA repair, mitochondrial replication and maintaining telomere and centromeric integrity, their persistence can be detrimental, leading to genome instability (25, 26). Abundant evidence has suggested that R-loops and RNA/DNA hybrids (hybrids) are induced at DSB sites and actively involved in DSB repair (27-30), but a comprehensive understanding of the underlying mechanisms is still elusive and under debate. Some studies propose that forming R-loops/hybrids at DSB ends interferes with the repair process. For

instance, upon loss of R-loop resolution factors DDX5, DHX9, and HNRNPD, hybrids accumulate at induced DSBs, which impede end resection, causing diminished RPA and RAD51 loading and defective HR (31-34). However, inactivation of an exosome catalytic subunit EXOSC10 leads to increased hybrid levels but strongly stimulates end resection, despite HR still being suppressed due to impaired RPA recruitment (35). Conversely, hybrids at DSBs have also been found to promote DSB repair by HR. DSB-induced small RNAs (diRNAs) and damage-induced long non-coding RNA (dilncRNA) facilitate the recruitment of HR proteins such as BRCA1, BRCA2 and RAD51 to DSBs (36-38). Additionally, in G2 cells exposed to IR, RAD52 is recruited to DNA damaging sites in a transcription and R-loop-dependent manner to promote HR (39). Interestingly, in yeast, defects in forming or removing RNA/DNA hybrids both impede HR efficiency, suggesting that transient RNA/DNA formation at DSBs is essential for HR (40). The controversial findings in the repair outcomes likely stem from the complex and intricate regulation exerted by R-loops/hybrids at different DSB repair steps, involving multiple players. Understanding the detailed mechanisms of how various R-loop resolving factors modulate the DSB repair process is critical for uncovering the biological significance of R-loops/hybrids at DSB ends.

SETX encodes a putative RNA-DNA helicase capable of unwinding RNA-DNA hybrids and plays an important role in R-loop homeostasis (41-43). Human SETX defects are documented in the neurodegenerative disorder ataxia with oculomotor apraxia 2 (AOA-2) (44) and autosomal dominant amyotrophic lateral sclerosis type 4 (ALS4) (45). Substantial evidence suggests an important role of SETX in the maintenance of genome stability (41-43). SETX and its yeast orthologue Sen1 interact with RNA Pol II (46-48) and are involved in transcription regulation and termination (42, 43). SETX and Sen1 associate with replication forks, promoting replication through transcription active genes (49). SETX forms foci colocalizing with DNA repair proteins upon replication stress and is believed to participate in resolving transcription and replication conflict (TRC) (50, 51). SETX in complex with BRCA1 is also involved in dismantling R-loops at transcription pause sites in transcription termination regions to prevent DNA damage (52). Furthermore, it has been demonstrated that SETX and Sen1 are recruited to DSBs when they occur in transcriptionally active loci to limit accumulation of R-loops/hybrids at DSB ends (52-55).

In this study, using EGFP-based repair reporters, which specifically monitor HR and BIR pathways in mammalian cells (56), we found a significant increase in both HR and BIR when SETX is deficient. Interestingly, short tract gene conversion (STGC), which is usually referred to as HR, alters

its genetic dependence in SETX-deficient cells and shows characteristics of BIR, requiring PIF1 and POLD3. End resection is substantially stimulated in an R-loops/hybrids-dependent manner at DSBs when SETX is deficient, leading to the onset of BIR-like mechanism for homology-directed repair. This hyper end resection is initiated by RAD52/XPF-mediated cleavage, distinct from the canonical end resection pathway which involves CtIP and the endonuclease activity of MRE11. Furthermore, analyzing BIR repair products revealed a significant increase in template switching during BIR in SETX-deficient cells, providing insights into the mechanisms underlying replication-associated chromosomal rearrangements, which are commonly observed in cancer genomes.

Results

SETX loss induces hyperrecombination for both HR and BIR in an R-loop-dependent manner.

In our previous study, we established EGFP-based reporter systems to determine the efficiency of DSB repair mediated by the HR (STGC) pathway (Fig 1A) and the BIR (long tract gene conversion: LTGC) pathway (Fig 1B) (56). In yeast, gene conversion with tract length longer than 1 kb, uses the BIR mechanism (57, 58). In our EGFP-BIR reporter (56), if BIR DNA synthesis continues for 3.8 kb, reaching the homology outside of the reporter on the right side of the break, BIR is completed by synthesis-dependent strand annealing (SDSA, BIR-SDSA), whereas if the BIR synthesizing strand is prematurely disassociated from its template and is ligated to the other side of the break, BIR is completed by end joining (BIR-EJ) (Fig S1A). If BIR DNA synthesis proceeds sufficiently to complete the FP region, green cells can be produced. Therefore, the gene conversion tract length ranges from 1.2 to 3.8 kb for BIR-EJ and is 3.8 kb or longer for BIR-SDSA (56). We showed that both BIR-SDSA and BIR-EJ depend on POLD3 and PIF1 (56).

Interestingly, we found that SETX depletion or knockout (KO) results in a significantly higher percentages of EGFP-positive cells in both U2OS (EGFP-HR) and U2OS (EGFP-BIR) reporter cell lines after I-SceI cleavage (Fig 1C and 1D), suggesting that SETX deficiency induces hyperrecombination for both HR and BIR. To determine whether the helicase activity of SETX is required for suppressing hyperrecombination, we expressed SETX wildtype (WT) allele or helicase-dead mutant SETX-P-loopΔ, containing deletion of eight amino acids (Δ1963-1970aa) in the SETX P-Loop GTP/ATP binding motif (59) in both WT and *SETX*-KO U2OS (EGFP-BIR) reporter cells. While expressing the SETX-WT allele suppresses hyper BIR in *SETX*-KO cells, the helicase-dead mutant failed to do so (Fig 1E). Consistent with the role of SETX in resolving R-loops (60, 61), we

showed stronger R-loop staining in *SETX*-KO cells compared to WT cells, with RNASEH1 overexpression attenuating the increase (Fig 1F). Overexpression of RNASEH1 in SETX-depleted cells also leads to suppression of hyperrecombination for BIR (Fig 1G) and HR (Fig 1H). These results suggest that SETX deficiency promotes hyperrecombination for both HR and BIR in an R-loop-dependent manner.

HR in SETX-deficient cells resembles BIR.

It has been well established that PIF1 helicase is required for BIR (LTGC) but is dispensable for HR (STGC) (56). We showed that both BIR in wildtype (WT) cells and hyper BIR in *SETX*-KO cells are strongly impaired upon PIF1 depletion by shRNAs, confirming the use of BIR (Fig 2A and Fig S1B). Interestingly, however, while HR in SETX-proficient WT cells is not dependent on PIF1, depleting PIF1 in *SETX*-KO cells leads to a significant reduction of HR, suggesting that PIF1 is required for HR in SETX-deficient cells (Fig 2B and Fig S1C). This suggests that SETX deficiency disrupts the regulation of the onset of HR and BIR and promotes the establishment of using the BIR mechanism even for HR/STGC. To investigate how the BIR mechanism is established at DSBs for HR/STGC, we performed chromatin immunoprecipitation (ChIP) and observed that SETX depletion causes increased recruitment of Flag-PIF1 to DSBs generated by I-SceI cleavage in the BIR reporter (Fig 2C). We speculate that enhanced PIF1 recruitment to DSBs in SETX-deficient cells may facilitate the onset of the BIR mechanism for DSB repair.

Based on our observation that SETX deficiency results in hyperrecombination, which is dependent on PIF1 for both BIR (LTGC) and HR (STGC), we reason that repair of DSBs spontaneously formed in SETX-deficient cells would heavily rely on PIF1. Indeed, γ-H2AX is significantly increased in cells depleted for both PIF1 and SETX (Fig 2D), and SETX and PIF1 exhibit synthetic lethal interactions (Fig 2E). Collectively, we conclude that hyperrecombination induced in SETX-deficient cells primarily occurs through activating the BIR mechanism rather than HR, and that BIR becomes an essential DSB repair pathway for the survival of SETX-deficient cells.

SETX loss induces excessive end resection mediated by a non-canonical pathway.

The impact of accumulated R-loops on DSB repair often involves the regulation of end resection (31-35). To investigate the effect of SETX loss on end resection, we first examined the level of RPA2 phosphorylation at S4 and S8 (pRPA2) after ionizing radiation (IR), known to be induced by end

resection (62). As shown in Fig 3A, *SETX*-KO cells manifest stronger pRPA2 signals after IR compared to WT cells, indicative of overactive end resection. Overexpression of RNASEH1 suppresses increased pRPA2 signals in *SETX*-KO cells, suggesting that hyper-resection is indeed caused by accumulation of R-loops/hybrids when SETX is deficient (Fig 3B).

By DNA/RNA immunoprecipitation (DRIP) analysis, we showed that R-loops/hybrids are accumulated at the DSBs in the BIR reporter, which can be suppressed by overexpressing RNASEH1 (Fig 3C). Using a well-established quantitative end resection assay (63), we demonstrated that end resection is significantly enhanced in *SETX*-KO cells at I-SceI-induced DSBs in the BIR reporter (Fig 3D), and overexpression of RNASEH1 suppresses hyper-resection due to SETX deficiency (Fig 3E). These data suggest that hyper-resection in SETX-deficient cells is induced by accumulated R-loops.

To investigate how hyper end resection is induced in SETX-deficient cells, we performed BIR assay after depleting MRE11 or CtIP. While BIR is impaired in SETX-proficient cells after depletion of CtIP or MRE11, surprisingly, BIR in *SETX*-KO cells is impeded only after depletion of MRE11 but not CtIP (Fig 4A and Fig S2A and S2B). We also found that hyper-resection observed in *SETX*-KO cells after I-SceI cleavage is largely dependent on MRE11 but not CtIP (Fig 4B). Similarly, IR-induced pRPA2 is dependent on MRE11 but not CtIP in *SETX*-KO cells, while RPA2 phosphorylation requires both MRE11 and CtIP in WT cells (Fig 4C). It is well established that the MRE11-RAD50-NBS1 (MRN) complex initiates the end resection by using the endonuclease activity of MRE11 to make a nick on the 5' strand with some distance from the break end, followed by using the MRE11 3' to 5' exonuclease activity to degrade DNA from the nick towards the DSB end (64-66). It has also been shown that CtIP is important for stimulating the MRE11 endonuclease activity (67-70). The dispensable role of CtIP for hyper-resection in SETX-deficient cells implies that the initial MRN-mediated endonuclease cleavage at DSB ends may not be involved. To test this possibility, we used an inhibitor PFM01 that specifically inhibits the endonuclease activity of MRE11 (64). Significantly, as shown in Fig 4D, treatment with PFM01 results in a reduction of end resection in SETX-proficient cells but fails to do so in SETX-deficient cells, and instead even stimulates it, while XPF depletion still leads to reduction of end resection in the presence of PFM01 in *SETX*-KO cells but not in WT cells. On the other hand, when we used the inhibitor PFM39, which specifically inhibits the exonuclease activity of MRE11 (64), we observed a clear reduction of end resection close to the cleavage site in both WT and *SETX*-KO cells (Fig 4E left, L1, 145 bp left to I-SceI site, Fig 3D top). However, at the site L3, which is further away from the cleavage site (L3, 2989 bp left to I-SceI site,

Fig 3D top), PFM39 treatment causes a minor decrease of end resection in *SETX*-KO cells with almost no effect in WT cells (Fig 4E right), consistent with the notion that extended 5' to 3' end resection involves BLM/DNA2 and EXO1, but not the MRE11 exonuclease activity. Collectively, these results suggest that the initiation of abnormal hyper end resection in SETX-deficient cells is not through the canonical pathway that requires MRE11- and CtIP-dependent endonucleolytic cleavage at DSB ends. However, the MRE11 exonuclease activity is still required for 3' to 5' strand removal towards the DSB ends (Fig 7F).

R-loop accumulation activates RAD52-XPF to initiate end resection in SETX-deficient cells.

The observation that CtIP and the MRE11 endonuclease activity is dispensable for hyper-resection in SETX-deficient cells suggests the involvement of a non-canonical end resection mechanism. It was shown previously that flap endonucleases XPF and XPG are involved in removing R-loops (39, 60, 71, 72). We found that depleting XPF leads to reduction of BIR in *SETX*-KO cells but not in WT cells (Fig 5A and Fig S3A). However, hyper BIR in *SETX*-KO cells is not affected after depleting XPG (Fig S3B and S3C). The end resection assay also indicates that hyper-resection in *SETX*-KO cells, but not normal end resection in SETX-proficient WT cells, is dependent on XPF (Fig 5B). It has been shown that XPF physically interacts with RAD52 (73), and RAD52 is recruited to DSBs containing R-loops at transcriptionally active loci (39, 74). We speculate that XPF may be recruited through the R-loop-RAD52 axis. Indeed, hyper BIR and hyper end resection induced by SETX depletion is diminished in *RAD52*-KO cells compared to WT cells (Fig 5C, 5D and Fig S3D), suggesting a role of RAD52 in promoting hyper-resection and hyper BIR specifically when SETX is deficient. In addition, by ChIP analysis, we observed that RAD52 accumulation at DSBs is significantly increased in the BIR reporter after I-SceI cleavage in *SETX*-KO cells compared to WT cells, which can be suppressed by RNASEH1 overexpression (Fig 5E). We also found that SETX depletion causes XPF accumulation at DSBs generated by I-SceI cleavage in WT cells but not in *RAD52*-KO cells (Fig 5F). These data support the model that RAD52 is associated with R-loops/hybrids at DSBs and subsequently recruits XPF to cleave R-loops/hybrids-associated DSB ends, initiating a non-canonical end resection pathway.

In the canonical MRN-initiated end resection, BLM, DNA2 and EXO1 are required for further long-range end resection. We depleted BLM, DNA2 and EXO1 and found that they all are indispensable for end resection in both WT and *SETX*-KO cells, as revealed by quantitative end

resection assay (Fig 4F and Fig S2C, S2D and S2E) and IR-induced RPA2 phosphorylation (Fig 4G). This suggests that BLM, DNA2 and EXO1 are also involved in the RAD52-XPF hyper end resection pathway, which is activated when R-loops are accumulated at DSBs due to SETX deficiency. We anticipate that bidirectional end resection (66) with 3' to 5' by the MRE11 exonuclease activity and 5' to 3' by BLM/DNA2 and EXO1 are still employed in the RAD52-XPF hyper end resection pathway (Fig 7F).

As RAD52 and XPF are required for hyper-resection in SETX-deficient cells, we examined whether RAD52 and XPF would exhibit synthetic lethal interactions with SETX. We found that depletion of SETX in *RAD52*- or *XPF*-KO cells significantly reduces cell viability (Fig 5G and Fig S3D, S3E), supporting the critical role of RAD52 and XPF in repairing DSBs with accumulated R-loops at the ends.

SETX deficiency leads to PCNA ubiquitination at DSBs, promoting PIF1 loading and BIR onset.

We showed that SETX loss results in a significant increase of PIF1 recruitment to DSBs (Fig 2C), thereby promoting the onset of BIR mechanism at deDSBs in SETX-deficient cells. In yeast, Pif1 interacts with PCNA and this interaction is important for promoting BIR DNA synthesis (75). In mammalian cells, we showed that both PCNA and PIF1 are required for BIR (17), and PCNA ubiquitination (PCNA-Ub) is a key step to trigger BIR onset (76). By performing an *in situ* proximity ligation assay (PLA), we showed that in *SETX*-KO cells, both PIF1 and PCNA recruitment to DSBs, marked by γH2AX after IR, is increased (Fig 6A, 6B). Using the antibody that specifically recognizes PCNA-Ub at K164 (77), we observed substantial accumulation of PCNA-Ub at DSBs in *SETX*-KO cells after IR, in contrast to minimal PCNA-Ub at DSBs in WT cells (Fig 6C). PCNA-Ub at DSBs in *SETX*-KO is suppressed by overexpressing RNASEH1 and inhibiting transcription with 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) (Fig 6C and 6D). These data suggest that accumulated R-loops/hybrids are the trigger to induce PCNA-Ub at DSBs in SETX-deficient cells. Furthermore, we showed that PIF1 accumulation at DSBs in *SETX*-KO cells can be suppressed by RNASEH1 overexpression (Fig 6A) and is dependent on PCNA and transcription (Fig 6E). We propose that R-loops/hybrids accumulated on ssDNA overhangs at DSBs as well as in the repair templates in SETX-deficient cells could causes collision of repair-associated DNA synthesis with R-loops/hybrids, leading to PCNA accumulation and ubiquitination, which in turn promotes PIF1 recruitment and the onset of BIR (see Discussion, FigS7).

SETX deficiency promotes template switching during BIR in an R-loop-dependent manner.

In yeast, it has been found that BIR is often associated with template switching (18), which could be mutagenic. We took advantage of our BIR reporter system which allows recovering BIR events by sorting EGFP-positive cells. We performed PCR analysis on isolated single green clones from U2OS (EGFP-BIR-5085) WT and *SETX*-KO cells after I-SceI cleavage to assess the BIR repair products (Fig S4A). Interestingly, we found that all green clones derived from WT cells can be analyzed using the PCR strategy (42 out of 42), yet fewer than half of the green clones (17/41) from *SETX*-KO cells succeeded in PCR amplification (Fig S4C). Among these resolved BIR events, we observed that 5 out of 42 events (11.9 %) in WT cells and 7 out of 17 events (41.2 %) have switched templates locally within the BIR reporter during BIR DNA synthesis (Fig S4C). The substantial increase of local template switching observed in *SETX*-KO cells raises the possibility that BIR template switching to other chromosomes may also be high in *SETX*-KO cells, making PCR analysis challenging. To facilitate the analysis of BIR products, we adapted the linear amplification-mediated high-throughput genome-wide sequencing (LAM-HTGTS) assay (78) to our BIR analysis (Fig S4B). We used the right side of the I-SceI cleavage site (primer R1, Fig S4B) as the bait to identify the BIR repair junctions. LAM-HTGTS-based assay offers a simple and quick way to pinpoint the BIR repair junctions for determining BIR synthesis tract length without involving long range PCR (e.g. F1/R1 PCR product: 3 kb for the BIR-SDSA product, Fig S4A). In addition to local jumping (Fig S5Aa), LAM-HTGTS can also catch the BIR events with template switching to other genomic sites but joining back to the second end of the DSB (Fig S5Ab). These jumping-out events, although with end joining back to the cleavage sites, if containing large-sized insertions, could fail the PCR analysis strategy due to too big PCR products and complex rearrangements. LAM-HTGTS cannot recover the events with template switching to other genomic loci without ligating back to the original DSB site on the reporter (Fig S5Ac).

We pooled the EGFP-positive single clones derived from U2OS (EGFP-BIR-5085) *SETX*-KO cell line after I-SceI cleavage. With LAM-HTGTS, we found two more events with local jumping in addition to the seven identified by the PCR assay from *SETX*-KO cells [total 9 events by HTGTS, Fig 7A and Fig S4C]. Most local jumping events in WT cells involve a single template switch, while *SETX*-KO cells often exhibit multiple rounds of copy-and-paste events (Fig 7D and Fig S4C). Moreover, we found that template switching to other chromosomes is drastically increased in *SETX*-

KO cells (WT: 2.4 %, *SETX*-KO: 31.7 %; Fig 7A and 7B). Since our HTGTS assay using the R1 primer only recovers events joining back to the second end of the DSB (Fig S4B and S5A), the successful identification of repair junctions in all BIR events by HTGTS suggests that all BIR events associated with template switching, recovered in our BIR assay as EGFP-positive events, have completed BIR by joining back to the original break without causing translocation. The overall frequency of template switching (including switching within and outside of the BIR reporter) in WT cells is 14.3 % (Fig 7A), and *SETX*-KO cells exhibit a markedly higher template switching frequency of 53.6 %, nearly four times that of WT cells (Fig 7A). Therefore, SETX deficiency does not only lead to hyperactive BIR, but also induces significantly more template switching, which could cause genome instability.

Transcription and the associated R-loop accumulation can block BIR replication fork progression and promote template switching (79). In our BIR reporter, CMV-promoter-driven transcription occurs in the donor template (Fig 1B). We hypothesize that SETX deficiency may induce R-loop formation in the BIR donor, causing conflict of BIR replication with R-loops at transcription sites (Fig 7G). Indeed, we found that SETX depletion enhances accumulation of R-loops/hybrids as revealed by DRIP analysis in the CMV surrounding regions in the EGFP-BIR-5085 reporter, which can be suppressed by RNASEH1 overexpression (Fig 7C). Significantly, RNASEH1 overexpression leads to a reduction of template switching frequency in *SETX*-KO cells to the level comparable to WT cells (Fig 7A and 7B), suggesting that templating switching is indeed caused by R-loop accumulation. Notably, the CMV promoter and its adjacent regions are a hot spot in the BIR reporter for initiating template switching (Fig 7D and Fig S4D), supporting the model that the collision of BIR fork with R-loops formed in the CMV and its surroundings in the BIR donor template, due to SETX deficiency, causes the disassociation of the BIR synthesizing strand from the donor to initiate template switching.

In addition, we found that the overall BIR tract length (Fig S1A) is increased upon RNASEH1 overexpression in both WT and *SETX*-KO cells (Fig 7E), suggesting that reducing R-loop levels permits more processive BIR DNA synthesis without disassociation. Collectively, these findings support the model that R-loops present on the BIR donor templets can disrupt BIR fork migration and induce template switching.

Discussion

In this report, we identified a new role of SETX in suppressing mutagenic BIR, contributing to the maintenance of genome stability. Mechanistically, loss of SETX causes accumulation of R-loops/hybrids at DSB ends, inducing a non-canonical hyperactive end resection pathway that requires RAD52-XPF. The resulting hyper end resection activity, combined with conflict of repair-associated DNA synthesis with R-loops/hybrids (see discussion below), leads to PCNA ubiquitination, PIF1 recruitment and the onset of BIR mechanism for hyperrecombination at deDSBs. Importantly, not only BIR frequency, but also BIR-associated template switching, is significantly elevated in SETX-deficient cells, revealing an unrecognized role of SETX in preventing replication-based complex genome rearrangements (CGRs), which are commonly observed in cancer genomes.

The mechanism underlying hyper end resection in SETX-deficient cells.

We observed that in SETX-deficient cells, end resection is hyperactive, mediated by a non-canonical end resection pathway that is independent of the MRE11 endonuclease activity and CtIP, but requires RAD52 and XPF. In one scenario, persistence of R-loops/hybrids at DSBs may inhibit the MRE11 nuclease activity, preventing the endonucleolytic cleavage to initiate end resection. This is consistent with the observation that RNA-DNA hybrids inhibit the MRE11 nuclease activity in an *in vitro* assay (80). Alternatively, XPF, recruited by RAD52 to R-loops/hybrids at DSBs, may override the function of MRE11 and CtIP in endonucleolytic cleavage by creating nicks more internally on the 5' strands (Fig 7F). In our study, we found that XPF, but not XPG, is involved in hyper end resection at deDSBs that are associated with R-loops/hybrids in SETX-deficient cells. Both XPF and XPG have been implicated in processing R-loops to generate DSBs (71, 81, 82), but during transcription-associated HR repair (TA-HRR), RAD52 recognizes RNA-DNA hybrids and promotes XPG-mediated R-loop processing (39). In this study, we found that RAD52 recognizes R-loops/hybrids but recruits XPF, not XPG, to cleave R-loops at deDSBs, initiating end resection in SETX-deficient cells, (Fig 7F). The use of different endonucleases could possibly be attributed to structural differences and chromatin modifications of R-loops and hybrids formed prior to DSB formation (R-loop processing) and at DSBs under normal conditions (TA-HRR) or with high levels of R-loops/hybrids due to SETX deficiency. It is worth mentioning that in an *in vitro* R-loop cleavage assay, XPF exhibits a strong activity to cleave the 5' of R-loops, but XPG almost has no cleavage activity at the 3' of R-loops (72) (Fig S5B). This observation supports our model that XPF, but not XPG, acts to process the 5' strand of DSBs for promoting end resection.

Notably, in a previous study using SETX depletion, hyper end resection was not detected despite the accumulation of R-loops/hybrids at DSB ends (53), and this discrepancy could be related to the level of SETX depletion. However, hyper end resection was also observed in yeast Sen1-deficient cells, which is consistent with our study (54). In the yeast Sen1-deficient model, Mre11 endonuclease is proposed to create a nick to initiate end resection, followed by Dna2 cleaving long 5' flaps of ssDNA in R-loops to promote extensive end resection, while Sgs1 (BLM homologue) and Exo1 are dispensable (54). The mechanism underlying hyper end resection in human SETX-deficient cells, which is independent of MRE11 endonuclease activity and CtIP, still differs from that in yeast. We propose that when SETX is deficient, RAD52 recognizes accumulated R-loops/hybrids at DSBs, and recruits XPF to cleave the 5' side of R-loops on the 5' strands of DSBs to initiate end resection (Fig 7F). With anticipated multiple transcription forks at transcriptional active DSBs, XPF could make a series of cleavages on the 5' strand of a DSB, thereby robustly accelerating end resection and resulting in hyperactive end resection phenotype. The MRE11 exonuclease activity, as well as BLM/DNA2 and EXO1, are still required for bidirectional processing to complete end resection.

Establishing BIR mechanism at R-loops/hybrids-accumulated DSBs in SETX-deficient cells.

Importantly, the hyperrecombination we observed in SETX-deficient cells is not merely an increase of recombination activity but involves a shift of pathway usage; upon loss of SETX, elevated HR (STGC) shows BIR characteristics, requiring PIF1 (Fig 2B). BIR is primarily used to repair seDSBs at broken forks and eroding telomeres (12, 83). In mammalian cells, BIR has been shown to be involved in mitotic DNA synthesis (MiDAS) (11, 84, 85), alternative telomere lengthening (ALT) (86-88) and fork repair upon cyclin E overexpression (89).

BIR is mutagenic (22, 23), and its utilization at deDSBs is regulated and restricted by multiple mechanisms as described in yeast (24). In our previous studies, we demonstrated that PIF1 binds preferentially to ubiquitinated PCNA, and PCNA accumulation and ubiquitination at DSBs is a key step to induce PIF1 loading to assemble BIR replisomes for BIR activation (56, 76). At broken forks, PCNA is accumulated and ubiquitinated at seDSBs, recruiting PIF1 to activate BIR, which is used for both BIR/LTGC and HR/STGC (Fig S6A left). This differs from the repair at deDSBs generated by endonucleases, where PIF1- and POLD3-independent HR/STGC is predominantly utilized over BIR/LTGC at deDSBs (56). This is largely attributed to the regulated assembly of BIR replisomes at deDSBs only when BIR is in need (Fig S6A right). At deDSBs, PCNA ubiquitination and PIF1

loading are minimal, in a sharp contrast to high levels of PCNA ubiquitination and PIF1 loading at seDSBs upon fork breakage (76). However, at deDSBs in 53BP1-deficient cells, stalled end-filling DNA synthesis on the ssDNA overhangs triggers PCNA ubiquitination and PIF1 loading, leading to the BIR replisome assembly and BIR activation at deDSBs (76). In this study, we demonstrated that the loss of SETX also overrides the control that restricts BIR at deDSBs (Fig S6B). We speculate that repair-associated DNA synthesis is stalled in SETX-deficient cells due to increased R-loops/hybrids. On the ssDNA overhangs, end fill-in DNA synthesis primed by Polα-primase occurs at ssDNA overhangs of deDSBs (76, 90, 91). De novo RNA synthesis and RNA/DNA hybrid formation also take place on the 3' ssDNA overhangs (30), which could accumulate in SETX-deficient cells, causing the stalling of end fill-in DNA synthesis (Fig S7A). In addition, if transcription is active in the donor templates as in the EGFP-HR and EGFP-BIR reporters, R-loops would accumulate in the templates when SETX is deficient. After strand invasion, HR/BIR DNA synthesis may stall upon encountering R-loops in the donor (Fig S7B). We propose that stalled repair DNA synthesis in SETX-deficient cells induces PCNA ubiquitination and subsequent PIF1 loading to deDSBs, leading to the onset of BIR (Fig S6B).

Hyper-recombination observed in SETX-deficient cells could be a result of hyper end resection coupled with BIR activation. Elongated ssDNA overhangs could provide extended homology to stabilize D-loops, thereby promoting recombination. Additionally, these longer ssDNA overhangs may harbor more stalled end-filling replicons loaded with ubiquitinated PCNA, potentially triggering stronger BIR activation, which could also contribute to hyper-recombination phenotype.

On the other hand, although being restricted, BIR is still used at deDSBs despite with lower frequency compared to HR (17). It remains unresolved what triggers the use of BIR at deDSBs and the biological significance of its utilization at deDSBs. We speculate that uncontrolled accumulation of R-loops/hybrids at deDSBs could be one of the signals to induce BIR onset at deDSBs. In WT cells under normal situations, R-loops/hybrids are present at deDSBs in the transcriptionally active regions but with an appropriate level, and transcription-associated homologous recombination repair (TA-HRR) can be engaged to promote HR (STGC) (39), whereas in SETX-deficient cells, R-loops/hybrids are accumulated, which induces PCNA ubiquitination at deDSBs, triggering the onset of BIR (Fig S6C). We propose that R-loop formation at deDSBs could be one mechanism to modulate the selection of HR (STGC) and BIR (LTGC) at deDSBs for DSB repair. Therefore, in addition to repairing seDSBs at collapsed forks, BIR is also involved at deDSBs in the transcriptionally active regions when R-

loops are accumulated. MiDAS, using BIR mechanism, often occurs at common fragile sites (CFSs) (85), where R-loops tend to accumulate due to TRC at large genes (92). Accumulated R-loops at CFSs may signal the onset of BIR to reduce the use of end joining, which would generate deletions, thereby safeguarding genome integrity by using BIR to complete replication of the unreplicated DNA regions (Fig S8).

R-loops/hybrids regulate HR in different steps.

Studies on the roles of RNA/DNA hybrids at deDSBs in end resection and HR yield quite diverse and controversial results. Functional impairment of different R-loop resolvases, leading to the accumulation of R-loops/hybrids at DSBs, has resulted in varying outcomes, including impaired end resection (31-34, 40), no discernible influence on end resection (93), or accelerated end resection [(35, 54), this study]. This is likely attributed to the distinct functions of different R-loop resolvases, with their impaired activities causing variations in the structure, extent, and level of R-loops/hybrids formed at deDSBs. We showed that SETX deficiency generates R-loops at DSBs that can be efficiently recognized and cleaved by RAD52-XPF, triggering a non-canonical hyper end resection mechanism, which overrides the requirement of MRE11 endonuclease activity and CtIP. However, accumulated R-loops or hybrids due to defects in other hybrid resolvases hinder end resection, which could possibly be induced by obstructing the enzymatic activities of MRE11, BLM, or EXO1, as suggested by *in vitro* experiments (80, 94).

We showed that SETX deficiency resulting in hyper recombination, but HR could also be impaired when R-loops/hybrids are accumulated at deDSBs (31-34, 38, 40). It has been proposed that hybrid formation on the 3' ssDNA overhangs aids in protecting the 3' ends, but it is necessary to remove hybrids afterwards to facilitate RPA and RAD51 loading (30, 40, 95). In yeast, both depletion of Rnh1 and Rnh201 and overexpression of Rnh1 result in HR defects (40), suggesting a delicate balance in R-loops/hybrid formation for regulating HR. In this study, we propose that R-loops/hybrids accumulated on the 3' ssDNA overhangs at deDSBs in SETX-deficient cells can be removed subsequently by other resolvases, allowing RPA and RAD51 to access ssDNA overhangs and facilitate strand invasion, thereby promoting hyper HR and BIR. Since both formation and resolution of R-loops/hybrids at DSBs are crucial for HR, different hybrid resolvases may function at distinct stages and could positively or negatively influence HR.

The role of SETX in the maintenance of genome stability.

SETX, like its counterpart Sen1, plays an important role in the maintenance of genome stability through multiple mechanisms (41-43). Our study reveals another important role of SETX in preventing replication-associated genomic rearrangements through suppressing template switching during BIR. Strikingly, overall BIR template switching increased more than four-fold in SETX-deficient cells, often with multiple template switches in a single BIR event, and a more than 10-fold increase in jumping to other chromosomes.

In our EGFP-BIR reporter, transcription is active, and SETX-deficiency could result in R-loop accumulation in the donor templates. The hyper recombination phenotype of SETX-deficient cells suggests that the strand invasion step is not impaired. However, the BIR DNA synthesis tract length is shorter upon SETX loss, but can be restored by RNASEH1 overexpression, suggesting that BIR DNA synthesis is likely disrupted due to R-loop formation in the donor, causing disassociation of invading strands from the donor and template switching (Fig S7B). Moreover, as strand invasion is from the left-side of the I-SceI cleavage site in the EGFP-BIR reporter, BIR DNA synthesis and transcription in the donor occur co-directionally in the same direction. Co-directional collisions of replication with R-loops in an episomal system has been shown to activate checkpoint and induce DNA damage (96). Co-directional conflict of replication and R-loops could also cause RNA polymerase backtracking to block replication (28). While our study showed that co-directional collision of BIR DNA synthesis with R-loops induces hyper templates switching, we predict that head-on collisions could be more detrimental and cause even greater disturbance of BIR. Therefore, SETX plays a critical role in the fidelity of BIR by reducing template switching, particularly at DSBs with active transcription. Although our study focuses on deDSBs, we speculate that on broken forks, if BIR templates are transcriptionally active, SETX is likely also involved in suppressing template switching.

Recent analysis of cancer genomes has uncovered CGRs with microhomology-mediated BIR (MMBIR) features (19, 97, 98). In a yeast study, local templated switching using microhomology sequences is observed when BIR is disrupted (21). In this study, we showed that in mammalian cells, R-loop accumulation at BIR donor templates, could drastically induce template switching during BIR not only locally but also to other genome sites, resulting in genomic instability with MMBIR features. Given that R-loop formation is widespread, occupying 5-10 % of the genome (28), BIR-mediated repair at R-loop-associated DSBs could significantly contribute to MMBIR through microhomology-mediated template switching, serving as a key driver of replication-associated chromosomal

rearrangements.

SETX and cancer connection

The critical role of SETX in R-loop homeostasis and the preservation of genome stability suggests its potential involvement in tumor suppression (42). SETX gene is frequently mutated in cancers, and its expression is often downregulated in a variety of cancers, akin to other tumor suppressor genes (99). The newly identified role of SETX in suppressing mutagenic BIR and preventing template switching during BIR in this study reveals an unrecognized aspect of SETX in maintaining genome stability and provides new insights into the connection of SETX deficiency with tumorigenesis. Moreover, we uncovered a pathway shift of HR to utilizing the BIR mechanism at deDSBs in the transcriptionally active regions and demonstrated synthetic lethal interactions of SETX with PIF1, RAD52 and XPF. The pronounced reliance of SETX-deficient cells on PIF1, RAD52 and XPF for cell viability lays a groundwork for establishing new treatment strategies through inhibiting PIF1, RAD52 and XPF activity to selectively eradicate cancers with compromised SETX function, while preserving viability of normal cells.

Materials and Methods

Cell cultures

U2OS (ATCC), 293T (ATCC) and U2OS cells with *PIF1*-KO (56) or *RAD52*-KO (100) cells were cultured in DMEM media supplemented with 10 % fetal bovine serum (FBS) and 1 % penicillin/streptomycin at 37°C, 5% CO2.

Plasmid construction

The N-terminal Flag-tagged WT and helicase-dead (P-loopΔ) SETX-expressing plasmids are gifts from Dr. Craig Bennett (59). Their coding sequences were cloned into pCDH-CMV-MCS-EF1-puro lentiviral vector (System Biosciences). The C-terminal V5-tagged WT RNASEH1 plasmids are gifts from Dr. Xiangdong Fu (101). Their coding sequences were cloned into pCDH-CMV-MCS-EF1-puro lentiviral vector (System Biosciences).

Generation of knock out (KO) cell line by CRISPR

Two sgRNAs sequences targeting coding sequence of specific genes were cloned into sgRNA/Cas9-mCherry vector derived from pSpCas9(BB)-2A-Puro (PX459) V2.0 vector (Addgene #62988). The two sgRNA plasmids were transfected together into target cell lines using standard PEI protocol. Two days later, transfected cells were sorted for mCherry-positive events and plated for single clones. These single clones were screened with PCR or Western blotting.

sgRNA sequences for *SETX*-KO:

aggctaaatccgttgatcac and TGTTGAAGCACTTTGTCGGA

sgRNA sequences for *XPF*-KO:

GGGCTAGTAGTGTGCGCCCG and GCTGGTGCTCAACACGCAGC

shRNA interference

Oligoes containing shRNA target sequences were cloned into pLKO.1-blast lentiviral vector (Addgene #26655). shRNA sequences used in this study are listed below.

SETXsh CGGCAGAAGGATTGTGTTATT

MRE11sh GATGAGAACTCTTGGTTTATT

CtIPsh GAGCAGACCTTTCTCAGTATA

BLMsh GAGCACATCTGTAAATTAATT

DNA2sh AGTTTGTGATGGGCAATATTT

EXO1shACTCGGATCTCCTAGCTTTTG

XPFsh AAGACGAGCTCACGAGTATTC

XPGsh GAACGCACCTGCTGCTGTAGA

PIF1sh GAAGACAGGTGCTCCGGAAGC

PCNAsh GTGGAGAACTTGGAAATGGAA

Knockdown efficiency of individual shRNA was confirmed by Western blotting of whole-cell lysate or RT-qPCR with primers listed below.

EXO1_quanti_F TCGGATCTCCTAGCTTTTGGCTG

EXO1_quanti_R AGCTGTCTGCACATTCCTAGCC

DNA2_quanti_F GATTTCTGGCACCAGCATAGCC

DNA2_quanti_R ACACCTCATGGAGAACCGTACC

CtIP_quanti_F TGGCAGACAGTTTCTCCCAAGC

CtIP_quanti_R GGCTCCACAAACGCTTTCTGCT

PIF1_quanti_F CATCCACAAGAGCCAAGGCAT

PIF1_quanti_R GGTGGCATAGAAGTGCAGCA

BLM_quanti_F GGTGATAAGACTGACTCAGAAGC

BLM_quanti_R AACGTGCCAAGAGCTTCCTCTC

XPG_quanti_F GATGCGGAAGATTCGCTCCATG

XPG_quanti_R GCTTTCCAGGAACATCTGTCCG

GAPDH_quanti_F GTCTCCTCTGACTTCAACAGCG

GAPDH_quanti_R ACCACCCTGTTGCTGTAGCCAA

PCNA_quanti_F CCTGCTGGGATATTAGCTCCA

PCNA_quanti_R CAGCGGTAGGTGTCGAAGC

Colony formation assay

Colony formation assay was carried out to examine the potential synthetic lethality effect between SETX and other tested genes. Cells were infected with lentivirus to knock down indicated genes. Three days after infection, cells were resuspended and counted. For each condition, 5000 cells were

plated on 10cm plate. Two weeks later, plates with growing colonies were stained with 0.5 % crystal violet in 20 % methanol.

Western blotting

Cells cultured in 6-well plate were lysed in RIPA buffer (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 5 mM EDTA, 1 % NP-40, 0.5 % sodium deoxycholate, and 0.1 % SDS). Fresh lysates were supplemented with eqaul amount of 2X loading buffer (100 mM Tris–Cl, pH 6.8, 4 % SDS, 0.2 % bromophenol blue, 20 % glycerol, 200 mM DTT) and then boiled at 95°C for 5 min. Next, sampels were separated on 6–15 % SDS–PAGE and proteins were transferred to nitrocellulose membrane (1620115, Bio-Rad) using the Trans-Blot Turbo Transfer System from Bio-Rad. Antibodies used were followed: SETX (NB100-57542, Novus Biologicals), KU70 (sc-17789, Santa Cruz Biotechnology), FLAG (F1804, Sigma-Aldrich), V5 (R96025, Invitrogen), XPF (sc-136153, Santa Cruz Biotechnology), RPA32 (Ser4/Ser8) (A300-245A, Bethyl Laboratories), γH2AX (2577, Cell Signaling Technology), PCNA (SC-56, Santa Cruz), GAPDH (AC002, ABclonal), RAD52 (A5186, ABclonal).

Immunostaining

Cells were seeded on coverslips in 6-well plate and cultured overnight. The next day, cells were first pre-extracted with 0.5 % Trion X-100 in cold PBS for 5mins, then fixed with 2 % formaldehyde for 10 mins at room temperature. Fixed cells were blocked with 5 % milk in PBS for 30 mins and then incubated overnight with indicated primary antibodies diluted in PBS at 4°C. The following day, cells were washed with PBS for three times and further incubated with secondary antibody and DAPI. Finally, slides were sealed with mounting media (P36930, Invitrogen) to prevent quenching. Images were taken on Nikon ECLIPSE Ni-L microscope. S9.6 antibody (ENH001, Kerafast) was used.

In situ proximity ligation assay (PLA)

Cells were seeded on coverslips in 6-well plate and cultured overnight. The next day, cells were treated with 4 Gy IR. Two hours later, cells were fixed and blocked the same way as immunostaining and then incubated with primary antibody overnight at 4°C. The following day, PLA was performed using the Duolink PLA kit (DUO92001 and DUO92005, Sigma-Aldrich) following the manufacturer's protocol. Finally, slides were sealed with mounting media (P36930, Invitrogen) to prevent quenching. Images were taken on Nikon ECLIPSE Ni-L microscope. Primary antibodies used were as followed:

FLAG (F1804, Sigma-Aldrich), PCNA (SC-56, Santa Cruz), γH2AX (07164, Upstate) and PCNA-Ub (13439, Cell Signaling Technology).

Chromatin immunoprecipitation (ChIP)

Cells growing on 10 cm plates were fixed with 1 % formaldehyde for 10 min at room temperature and glycine was added to the final concentration of 0.125 M. Cells were scraped from the plates and lysed with 0.5 % NP-40 buffer (10 mM Tris-HCl, 10 mM NaCl, 0.5 % NP-40, freshly added Roche cOmplete protease inhibitor cocktail) for 15mins at 4°C. Whole lysates were sonicated with Bioruptor for 15 cycles (30s on/30s off) and then centrifuged to remove debris. Supernatent was collected and incubated with Protein G dynabeads which have been pre-loaded with indicated antibody, for 4 hours at 4°C. After IP, beads were washed three times with 0.5 % NP-40 buffer and once with TE buffer. Finally, beads was incubated in elution buffer (10 mM Tris-HCl, 1.0 % SDS, 1.0 mM EDTA) overnight at 65°C to reverse the crosslinking. The next day, IP products and input were subjected to DNA extraction using phenol-chloroform. Recovered DNA was checked by qPCR to calculate the relative enrichement of target protein on specific locus. Primers used for ChIP were listed below.

At I-SceI-induced DSB end:

ChIP_L1F GAATCCATCTTGCTCCAACACCCC

ChIP_L1R CACAAACACAACTCCTCCGCGC

At repair template near CMV promoter:

ChIP_L5F CCCTGAAGTTCATCTGCACCACCG

ChIP_L5R cgagacttacCTGGACGTAGCCTTC

Antibodies used were as followed: S9.6 (ENH001, Kerafast), RAD52 (A5186, ABclonal), XPF (sc-136153, Santa Cruz Biotechnology).

End resection quantification assay

The end resection assay in reporter cells was adopted from publications of Dr. Tanya Paull's lab (63) and modified according to our BIR reporter sequence. The BIR reporter cells were infected with lentivirus expressing I-SceI to induce DSBs. Two days after infection, cells were harvested for genomic DNA extraction using Puregene Kits from QIAGEN. 1 ug genomic DNA was digested separately with NEB restriction enzyme BsiHKAI (left DSB end), ApoI (right DSB end) or SacII (mock) overnight. Restriction enzyme digested genomic DNA was used as template for qPCR with

indicated primer listed below.

L1:

End-Resection_L1F GCTCCAACACCCCAACATCTTCGAC

End-Resection_L1R CGGTACTTCGTCCACAAACACAACTCC

L3:

End-Resection_L3F GGTGAACTTCAAGATCCGCCACAAC

End-Resection_L3R CAGCTCGTCCATGCCGAGAG

R1:

End-Resection_R1F gaactagtggatcCCCTACCGTTCG

End-Resection_R1R ggtggatgtggaatgtgtgcgagg

R3:

End-Resection_R3F CGAGATCAGCAGCCTCTGTTCCACA

End-Resection_R3R CTGTCACCAATCCTGTCCCTAGTGCA

LAM-HTGTS assay

This assay was adopted from a previous publication (78) and modified to fit our BIR reporter system. EGFP-BIR-5085 reporter cells were infected with lentivirus expressing I-SceI to induce DSBs. Five days after infection, EGFP-positive cells were collected by FACS sorting and plated for single clones. Two weeks later, single clones were picked out and cultured in 6-well plate. When cells were confluent enough, genomic DNA was extracted with Puregene Kits from QIAGEN for each single clone. Equal amount of genomic DNA from indicated number of single clones was mixed together, and 500 ng mixed genomic DNA was used as template for LAM PCR, where biotin-labeled primer (Bio-R1) was supplemented together with 2x NEB Next Ultra II Q5 Master Mix (M0544L, NEB). LAM products were purified with Streptavidin C1 dynabeads (65001, Thermo Scientific). A customized linker (rand10_3Tr3) was added to the 3' end of on-bead LAM products using T4 RNA ligase I (M0204S, NEB). Nested PCR was performed using indicated primer set (3Tr3-Adaptor_R and JM_F) with the ligated on-bead LAM products as templates. Products of nested PCR were further amplified using the Illumina indexing primer set (i501-508, i701-712). Final PCR products were sent for 1x100bp deep sequencing. CD-HIT (102) was used to cluster the sequencing results. Recovered sequences of each cluster were checked for template switching.

Sequences of oligoes used for PCR analysis and LAM-HTGTS assay are listed below.

PCR analysis:

F1 GGCATGGACGAGCTGTACAAGTAA

R1 gGTAGAATTTCGAggtcgaGGATC

LAM-HTGTS:

Bio-R1 /5Biosg/GAGGATCTGGGCTAGCCCTGATCAATAAC

rand10_3Tr3/5Phos/NNNNNNNNNNAGATCGGAAGAGCACACGTCTG/3SpC3/

JM_F CTTTCCCTACACGACGCTCTTCCGATCTGCTATACGAACGGTAGGGgatcc

3Tr3-Adaptor_R GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Statistical analysis

Student's t-test was run in two-tailed non-paired mode to examine the significance of differences between samples. In all experiments, error bars show standard deviation (SD) of at least three independent experiments. The P value is labeled as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, and ns (not significant) $P > 0.05$.

Acknowledgement

We would like thank Dr. Craig Bennett (University of Lincoln, UK) for providing the cDNA of WT and helicase-dead (P-loopΔ) mutant SETX, Dr. Xiangdong Fu (UCSD) for providing the RNASEH1 constructs. Plasmids pLKO.1-blast vector (26655) and pSpCas9(BB)-2A-Puro (PX459) V2.0 (62988) are from Addgene. This study is funded by NIH grants CA244912, GM141868 and CA187052 to X.W.

Author contributions

T.W. designed and performed experiments and analyzed the data; Y.L. and L.S. conducted some experiments. X.W. is responsible for the overall project's planning and experimental design. T.W. and X.W. wrote the manuscript.

Disclosure and competing interests statement

The authors declare no competing financial interests.

Reference

1. K. K. Khanna, S. P. Jackson, DNA double-strand breaks- signaling, repair and the cancer connection. *Nature Genet.* **27**, 247–254 (2001).
2. R. Ceccaldi, B. Rondinelli, A. D. D'Andrea, Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.* **26**, 52-64 (2016).
3. W. D. Heyer, Regulation of recombination and genomic maintenance. *Cold Spring Harbor perspectives in biology* **7**, a016501 (2015).
4. F. Paques, J. E. Haber, Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **63**, 349-404 (1999).
5. M. Jasin, R. Rothstein, Repair of strand breaks by homologous recombination. *Cold Spring Harbor perspectives in biology* **5** (2013).
6. B. Zhao, E. Rothenberg, D. A. Ramsden, M. R. Lieber, The molecular basis and disease relevance of non-homologous DNA end joining. *Nat Rev Mol Cell Biol* **21**, 765-781 (2020).
7. N. Hustedt, D. Durocher, The control of DNA repair by the cell cycle. *Nat. Cell Biol.* **19**, 1-9 (2016).
8. L. Liu, A. Malkova, Break-induced replication: unraveling each step. *Trends Genet* **38**, 752-765 (2022).
9. R. P. Anand, S. T. Lovett, J. E. Haber, Break-induced DNA replication. *Cold Spring Harbor perspectives in biology* **5**, a010397 (2013).
10. B. Llorente, C. E. Smith, L. S. Symington, Break-induced replication: what is it and what is it for? *Cell Cycle* **7**, 859-864 (2008).
11. X. Wu, A. Malkova, Break-induced replication mechanisms in yeast and mammals. *Curr Opin Genet Dev* **71**, 163-170 (2021).
12. J. Kramara, B. Osia, A. Malkova, Break-Induced Replication: The Where, The Why, and The How. *Trends Genet* **34**, 518-531 (2018).
13. J. R. Lydeard, S. Jain, M. Yamaguchi, J. E. Haber, Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* **448**, 820-823 (2007).
14. R. A. Donnianni, L. S. Symington, Break-induced replication occurs by conservative DNA synthesis. *Proc Natl Acad Sci U S A* **110**, 13475-13480 (2013).
15. N. Saini *et al.*, Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**, 389-392 (2013).
16. M. A. Wilson *et al.*, Pif1 helicase and Poldelta promote recombination-coupled DNA synthesis via bubble migration. *Nature* **502**, 393-396 (2013).
17. S. Li *et al.*, PIF1 helicase promotes break-induced replication in mammalian cells. *EMBO J* 10.15252/embj.2020104509, e104509 (2021).
18. C. E. Smith, B. Llorente, L. S. Symington, Template switching during break-induced replication. *Nature* **447**, 102-105 (2007).
19. Y. Li *et al.*, Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112-121 (2020).
20. K. Hadi *et al.*, Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. *Cell* **183**, 197-210 e132 (2020).
21. C. J. Sakofsky *et al.*, Translesion Polymerases Drive Microhomology-Mediated Break-Induced Replication Leading to Complex Chromosomal Rearrangements. *Mol Cell* **60**, 860-872 (2015).
22. C. J. Sakofsky *et al.*, Break-induced replication is a source of mutation clusters underlying kataegis. *Cell reports* **7**, 1640-1648 (2014).
23. C. J. Sakofsky, S. Ayyar, A. Malkova, Break-induced replication and genome stability.

Biomolecules **2**, 483-504 (2012).
24. N. Pham *et al.*, Mechanisms restraining break-induced replication at two-ended DNA double-strand breaks. *EMBO J.* **40**, e104847 (2021).
25. T. Garcia-Muse, A. Aguilera, R Loops: From Physiological to Pathological Roles. *Cell* **179**, 604-618 (2019).
26. C. Niehrs, B. Luke, Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol* **21**, 167-178 (2020).
27. E. Petermann, L. Lan, L. Zou, Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. *Nat Rev Mol Cell Biol* **23**, 521-540 (2022).
28. J. R. Brickner, J. L. Garzon, K. A. Cimprich, Walking a tightrope: The complex balancing act of R-loops in genome stability. *Mol Cell* **82**, 2267-2297 (2022).
29. A. Marnef, G. Legube, R-loops as Janus-faced modulators of DNA repair. *Nat Cell Biol* **23**, 305-313 (2021).
30. B. Gomez-Gonzalez, A. Aguilera, Break-induced RNA-DNA hybrids (BIRDHs) in homologous recombination: friend or foe? *EMBO Rep* **24**, e57801 (2023).
31. L. Alfano *et al.*, Depletion of the RNA binding protein HNRNPD impairs homologous recombination by inhibiting DNA-end resection and inducing R-loop accumulation. *Nucleic Acids Res* **47**, 4068-4085 (2019).
32. M. Matsui *et al.*, USP42 enhances homologous recombination repair by promoting R-loop resolution with a DNA-RNA helicase DHX9. *Oncogenesis* **9**, 60 (2020).
33. G. Sessa *et al.*, BRCA2 promotes DNA-RNA hybrid resolution by DDX5 helicase at DNA breaks to facilitate their repairdouble dagger. *EMBO J* **40**, e106018 (2021).
34. Z. Yu *et al.*, DDX5 resolves R-loops at DNA double-strand breaks to promote DNA repair and avoid chromosomal deletions. *NAR Cancer* **2**, zcaa028 (2020).
35. J. Domingo-Prim *et al.*, EXOSC10 is required for RPA assembly and controlled DNA end resection at DNA double-strand breaks. *Nat Commun* **10**, 2135 (2019).
36. M. Gao *et al.*, Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Res* **24**, 532-541 (2014).
37. F. Michelini *et al.*, Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. *Nat Cell Biol* **19**, 1400-1411 (2017).
38. G. D'Alessandro *et al.*, BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. *Nat Commun* **9**, 5376 (2018).
39. T. Yasuhara *et al.*, Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. *Cell* **175**, 558-570 e511 (2018).
40. C. Ohle *et al.*, Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell* **167**, 1001-1013 e1007 (2016).
41. M. Groh, L. O. Albulescu, A. Cristini, N. Gromak, Senataxin: Genome Guardian at the Interface of Transcription and Neurodegeneration. *J Mol Biol* **429**, 3181-3195 (2017).
42. V. Gatti, S. De Domenico, G. Melino, A. Peschiaroli, Senataxin and R-loops homeostasis: multifaced implications in carcinogenesis. *Cell Death Discov* **9**, 145 (2023).
43. M. Giannini, O. Porrua, Senataxin: A key actor in RNA metabolism, genome integrity and neurodegeneration. *Biochimie* **217**, 10-19 (2024).
44. M. C. Moreira *et al.*, Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat. Genet.* **36**, 225-227 (2004).
45. Y. Z. Chen *et al.*, DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral

sclerosis (ALS4). *Am. J. Hum. Genet.* **74**, 1128-1135 (2004).
46. D. Ursic, K. Chinchilla, J. S. Finkel, M. R. Culbertson, Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res* **32**, 2441-2452 (2004).
47. Y. Z. Chen *et al.*, Senataxin, the yeast Sen1p orthologue: characterization of a unique protein in which recessive mutations cause ataxia and dominant mutations cause motor neuron disease. *Neurobiol Dis* **23**, 97-108 (2006).
48. A. Suraweera *et al.*, Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Human molecular genetics* **18**, 3384-3396 (2009).
49. A. Alzu *et al.*, Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell* **151**, 835-846 (2012).
50. O. Yuce, S. C. West, Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol Cell Biol* **33**, 406-417 (2013).
51. P. Richard, S. Feng, J. L. Manley, A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes Dev* **27**, 2227-2232 (2013).
52. E. Hatchi *et al.*, BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* **57**, 636-647 (2015).
53. S. Cohen *et al.*, Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nat Commun* **9**, 533 (2018).
54. C. C. Rawal *et al.*, Senataxin Ortholog Sen1 Limits DNA:RNA Hybrid Accumulation at DNA Double-Strand Breaks to Control End Resection and Repair Fidelity. *Cell reports* **31**, 107603 (2020).
55. S. Cohen *et al.*, A POLD3/BLM dependent pathway handles DSBs in transcribed chromatin upon excessive RNA:DNA hybrid accumulation. *Nat Commun* **13**, 2012 (2022).
56. S. Li *et al.*, PIF1 helicase promotes break-induced replication in mammalian cells. *EMBO J.* **40**, e104509 (2021).
57. S. Jain *et al.*, A recombination execution checkpoint regulates the choice of homologous recombination pathway during DNA double-strand break repair. *Genes Dev* **23**, 291-303 (2009).
58. A. Mehta, A. Beach, J. E. Haber, Homology Requirements and Competition between Gene Conversion and Break-Induced Replication during Double-Strand Break Repair. *Mol Cell* **65**, 515-526 e513 (2017).
59. C. L. Bennett, B. L. Sopher, A. R. La Spada, Tight expression regulation of senataxin, linked to motor neuron disease and ataxia, is required to avert cell-cycle block and nucleolus disassembly. *Heliyon* **6**, e04165 (2020).
60. M. P. Crossley *et al.*, R-loop-derived cytoplasmic RNA-DNA hybrids activate an immune response. *Nature* **613**, 187-194 (2023).
61. H. Zhao *et al.*, Senataxin and RNase H2 act redundantly to suppress genome instability during class switch recombination. *Elife* **11** (2022).
62. H. Liaw, D. Lee, K. Myung, DNA-PK-dependent RPA2 hyperphosphorylation facilitates DNA repair and suppresses sister chromatid exchange. *PLoS One* **6**, e21424 (2011).
63. Y. Zhou, P. Caron, G. Legube, T. T. Paull, Quantitation of DNA double-strand break resection intermediates in human cells. *Nucleic Acids Res* **42**, e19 (2014).
64. A. Shibata *et al.*, DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. *Mol Cell* **53**, 7-18 (2014).

65. R. A. Deshpande, J. H. Lee, S. Arora, T. T. Paull, Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. *Mol Cell* **64**, 593-606 (2016).
66. P. Cejka, L. S. Symington, DNA End Resection: Mechanism and Control. *Annu Rev Genet* **55**, 285-307 (2021).
67. E. Cannavo, P. Cejka, Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. *Nature* **514**, 122-125 (2014).
68. A. Zdravkovic *et al.*, A conserved Ctp1/CtIP C-terminal peptide stimulates Mre11 endonuclease activity. *Proc Natl Acad Sci U S A* **118** (2021).
69. R. Anand *et al.*, NBS1 promotes the endonuclease activity of the MRE11-RAD50 complex by sensing CtIP phosphorylation. *EMBO J* **38** (2019).
70. R. Anand, L. Ranjha, E. Cannavo, P. Cejka, Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol Cell* **64**, 940-950 (2016).
71. J. Sollier *et al.*, Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* **56**, 777-785 (2014).
72. M. Tian, F. W. Alt, Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. *J Biol Chem* **275**, 24163-24172 (2000).
73. T. A. Motycka, T. Bessho, S. M. Post, P. Sung, A. E. Tomkinson, Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. *J Biol Chem* **279**, 13634-13639 (2004).
74. J. Tan *et al.*, An R-loop-initiated CSB-RAD52-POLD3 pathway suppresses ROS-induced telomeric DNA breaks. *Nucleic Acids Res* **48**, 1285-1300 (2020).
75. O. Buzovetsky *et al.*, Role of the Pif1-PCNA Complex in Pol δ-Dependent Strand Displacement DNA Synthesis and Break-Induced Replication. *Cell Reports* **21**, 1707-1714 (2017).
76. S. B. Shah *et al.*, 53BP1 deficiency leads to hyperrecombination using break-induced replication (BIR). *Nature Communications* **15** (2024).
77. T. Thakar *et al.*, Ubiquitinated-PCNA protects replication forks from DNA2-mediated degradation by regulating Okazaki fragment maturation and chromatin assembly. *Nat Commun* **11**, 2147 (2020).
78. J. Hu *et al.*, Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat Protoc* **11**, 853-871 (2016).
79. L. Liu *et al.*, Tracking break-induced replication shows that it stalls at roadblocks. *Nature* 10.1038/s41586-020-03172-w (2021).
80. E. Y. Chang *et al.*, MRE11-RAD50-NBS1 promotes Fanconi Anemia R-loop suppression at transcription-replication conflicts. *Nat Commun* **10**, 4265 (2019).
81. N. Makharashvili *et al.*, Sae2/CtIP prevents R-loop accumulation in eukaryotic cells. *Elife* **7** (2018).
82. C. T. Stork *et al.*, Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *Elife* **5** (2016).
83. X. Wu, A. Malkova, Break-induced replication mechanisms in yeast and mammals. *Curr. Opin. Genet. Dev.* **71**, 163-170 (2021).
84. E. A. Epum, J. E. Haber, DNA replication: the recombination connection. *Trends Cell Biol* **32**, 45-57 (2022).
85. R. Bhowmick, I. D. Hickson, Y. Liu, Completing genome replication outside of S phase. *Mol Cell* **83**, 3596-3607 (2023).

86. R. L. Dilley *et al.*, Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature* **539**, 54-58 (2016).
87. J. M. Zhang, T. Yadav, J. Ouyang, L. Lan, L. Zou, Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. *Cell Rep* **26**, 955-968 e953 (2019).
88. F. M. Roumelioti *et al.*, Alternative lengthening of human telomeres is a conservative DNA replication process with features of break-induced replication. *EMBO Rep* **17**, 1731-1737 (2016).
89. L. Costantino *et al.*, Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* **343**, 88-91 (2014).
90. Z. Mirman *et al.*, 53BP1-RIF1-shieldin counteracts DSB resection through CST- and Polalpha-dependent fill-in. *Nature* **560**, 112-116 (2018).
91. Z. Mirman, N. K. Sasi, A. King, J. R. Chapman, T. de Lange, 53BP1-shieldin-dependent DSB processing in BRCA1-deficient cells requires CST-Polalpha-primase fill-in synthesis. *Nat Cell Biol* **24**, 51-61 (2022).
92. A. Helmrich, M. Ballarino, L. Tora, Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* **44**, 966-977 (2011).
93. L. Li *et al.*, DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. *Mol Cell Biol* **36**, 2794-2810 (2016).
94. J. M. Daley *et al.*, Specificity of end resection pathways for double-strand break regions containing ribonucleotides and base lesions. *Nat Commun* **11**, 3088 (2020).
95. S. Liu *et al.*, RNA polymerase III is required for the repair of DNA double-strand breaks by homologous recombination. *Cell* **184**, 1314-1329 e1310 (2021).
96. S. Hamperl, M. J. Bocek, J. C. Saldivar, T. Swigut, K. A. Cimprich, Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170**, 774-786 e719 (2017).
97. B. Osia *et al.*, Cancer cells are highly susceptible to accumulation of templated insertions linked to MMBIR. *Nucleic Acids Res* **49**, 8714-8731 (2021).
98. I. Cortes-Ciriano *et al.*, Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* **52**, 331-341 (2020).
99. D. R. Rhodes *et al.*, Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166-180 (2007).
100. S. Liu *et al.*, DNA repair protein RAD52 is required for protecting G-quadruplexes in mammalian cells. *J Biol Chem* **299**, 102770 (2023).
101. J. Y. Chen, X. Zhang, X. D. Fu, L. Chen, R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1. *Nat. Protoc.* **14**, 1661-1685 (2019).
102. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).

Figure Legends

Figure 1. SETX loss induces hyperrecombination for both HR and BIR in an R-loop-dependent manner.

A, B. Schematic drawing of EGFP-HR-5633 (A) and EGFP-BIR-5085 (B) reporters.

C, D. U2OS WT and *SETX*-KO EGFP-HR-5633 (C) and EGFP-BIR-5085 (D) reporter cells expressing SETX shRNA or vector, were infected with lentivirus expressing I-SceI to induce DSBs. The percentage of EGFP-positive cells was quantified by FACS analysis five days later (top). The efficiency of SETX knockdown or knockout were determined by Western blotting (bottom).

E. U2OS WT or *SETX*-KO EGFP-BIR-5085 reporter cells were infected with lentivirus expressing Flag-tagged WT SETX or mutated SETX, then infected with lentivirus expressing I-SceI to induce DSBs. The percentage of EGFP-positive cells was quantified by FACS analysis five days later (Left). The expression of Flag-SETX was determined by Western blotting (Right).

F. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells were infected with lentiviruses expressing RNASEH1. Control cells and RNASEH1-expressing cells were fixed and subjected to immunostaining with S9.6 antibody to quantify R-loop levels.

G, H. U2OS EGFP-HR-5633 (G) or EGFP-BIR-5085 (H) reporter cells with or without RNASEH1 expression were infected with lentiviruses expressing SETX shRNA or vector. Three days later, all cells were infected with lentiviruses expressing I-SceI to induce DSBs. The percentage of EGFP-positive cells was quantified by FACS analysis five days later (Left). The expression level of SETX and RNASEH1-V5 was determined by Western blotting (Right).

Figure 2. HR in SETX-deficient cells resembles BIR.

A, B. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells (A) or EGFP-HR-5633 reporter cells (B) were infected with lentiviruses expressing PIF1 shRNA or vector. After three days, these cells were infected with I-SceI-expressing lentiviruses to induce DSBs. The percentage of EGFP-positive cells was quantified by FACS analysis five days later.

C. U2OS WT EGFP-BIR-5085 reporter cells were infected with Flag-PIF1 lentiviruses. Cells with Flag-PIF1 expression were infected with SETX shRNA or vector. Three days later, cells were further infected with I-SceI-expressing lentivirus to induce DSBs. Two days after infection, cells were harvested for ChIP analysis of Flag-PIF1 at indicated locus.

D. U2OS cells were infected with lentiviruses expressing vector, SETX shRNA, PIF1 shRNA, or both

SETX and PIF1 shRNA. Cell lysates were collected four days after virus infection and checked with Western blotting.

E. U2OS cells were infected with lentiviruses expressing vector, SETX shRNA, PIF1 shRNA, or both SETX and PIF1 shRNAs. Three days after infection, cells were resuspended and counted. 5000 cells were seeded on plates and cultured for two weeks, then stained with crystal violet.

Figure 3. SETX loss induces excessive end resection.

A. WT and *SETX*-KO U2OS cells were treated with 10 Gy IR. Cell lysates were collected at indicated time points and Western blotting was performed using indicated antibodies.

B. WT and *SETX*-KO U2OS cells with or without RNASEH1 expression were treated with 10 Gy IR. Cell lysates were collected one hour after IR and Western blotting was performed using indicated antibodies.

C. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells, with or without RNASEH1 expression, were infected with I-SceI-expressing lentiviruses to induce DSBs. Two days after infection, cells were harvested for DRIP analysis at indicated locus.

D. Illustration of site-specific end resection assay in EGFP-BIR-5085 reporter (top). U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells were infected with I-SceI-expressing lentiviruses to induce DSBs (bottom). After two days, genomic DNA (gDNA) is extracted and digested with indicated restriction enzyme (RE): BsiHKAI for the left DSB ends, ApoI for the right DSB ends, or SacII as mock for both ends. RE-resistant ssDNA is quantified by qPCR using primer sets around the RE recognition sites.

E. U2OS EGFP-BIR-5085 reporter cells with or without RNASEH1 expression were infected with SETX shRNA or vector. Two days after infection, cells were infected with I-SceI-expressing lentivirus to induce DSBs. Three days after the second infection, cells were harvested for gDNA for end resection assay.

Figure 4. SETX loss activates non-canonical resection pathway.

A, B, C. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells were infected with lentiviruses expressing MRE11 shRNA, CtIP shRNA, or vector. Three days were given to allow shRNA expression to achieve effective knockdown of each gene. For BIR efficiency analysis (A), shRNA-treated cells were infected with I-SceI-expressing lentiviruses to induce DSBs. The percentage of EGFP-positive

cells was quantified by FACS analysis five days after I-SceI infection. For end resection assay (B), shRNA-treated cells were infected with I-SceI-expressing lentiviruses to induce DSBs. Two days after I-SceI infection, cells were harvested for end resection assay. For examining the pRPA2 level (C), shRNA-treated cells were treated with 10 Gy IR. Cell lysates were collected one hour after IR and Western blotting was performed using indicated antibodies.

D. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells were infected with lentiviruses expressing XPF shRNA or vector. Three days later, cells were infected with I-SceI-expressing lentiviruses to induce DSBs. Inhibitor PFM01 (100 uM) or DMSO was applied to cells together with I-SceI lentiviruses. Two days after I-SceI infection, cells were harvested for end resection assay.

E. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells were infected with I-SceI-expressing lentiviruses to induce DSBs, simultaneously treated with DMSO or inhibitor PFM39 (100 uM). Two days after I-SceI induction, cells were harvested for end resection assay.

F, G. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells were infected with lentiviruses encoding BLM shRNA, DNA2 shRNA, EXO1 shRNA, or vector. Three days later, cells were then infected with I-SceI-expressing lentiviruses to induce DSBs for end resection assay (F), or treated with 10 Gy IR to examine the pRPA2 level (G).

Figure 5. R-loop accumulation activates RAD52-XPF to initiate end resection in SETX-deficient cells.

A, B. U2OS WT and *SETX-KO* EGFP-BIR-5085 reporter cells were infected with lentiviruses encoding XPF shRNA or vector. Three days after infection, cells were infected with I-SceI-expressing lentiviruses to induce DSBs. The percentage of EGFP-positive cells was quantified by FACS analysis five days after I-SceI induction (A), and end resection assay was performed two days after I-SceI induction (B).

C, D. U2OS WT and *RAD52*-KO EGFP-BIR-5085 reporter cells were infected with lentiviruses encoding SETX shRNA or vector. Three days after infection, cells were infected with I-SceI-expressing lentiviruses to induce DSBs. The percentage of EGFP-positive cells was quantified by FACS analysis five days after I-SceI induction (C), and end resection assay was performed two days after I-SceI induction (D).

E. U2OS WT and *SETX-KO* EGFP-BIR-5085 reporter cells, with or without RNASEH1 expression, were harvested for ChIP analysis of RAD52 at the indicated locus on the EGFP-BIR-5085 reporter.

F. U2OS WT and *RAD52*-KO EGFP-BIR-5085 reporter cells were infected with lentiviruses encoding

SETX shRNA or vector. Four days later, cells were harvested for ChIP analysis of XPF at the indicated locus on the EGFP-BIR-5085 reporter.

G. WT, *RAD52*-KO, and *XPF*-KO U2OS cells were infected with lentiviruses encoding SETX shRNA or vector. Three days after infection, cells were resuspended and counted. 5000 cells were seeded on plates and cultured for two weeks, followed by staining with crystal violet.

Figure 6. PCNA accumulation and ubiquitination at DSB ends underlie the BIR activation in SETX-deficient cells.

A. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells, with or without RNASEH1 expression, were infected with Flag-PIF1-expressing lentiviruses. All cells were treated with 4 Gy IR. Two hours later, the co-localization of Flag-PIF1 and γ-H2AX was analyzed by PLA.

B. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells, with or without RNASEH1 expression, were treated with 4 Gy IR. Two hours later, the co-localization of PCNA and γ-H2AX was analyzed by PLA.

C. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells, with or without RNASEH1 expression, were treated with 4 Gy IR. Two hours later, the co-localization of ubiquitinated PCNA and γ-H2AX was analyzed by PLA.

D. U2OS *SETX*-KO EGFP-BIR-5085 reporter cells were treated with DMSO or 20 μM DRB for 16 hours prior to 4Gy IR. Two hours after IR, the co-localization of ubiquitinated PCNA and γ-H2AX was analyzed by PLA.

E. U2OS *SETX*-KO EGFP-BIR-5085 reporter cells with Flag-PIF1 expression were infected with lentiviruses encoding PCNA shRNA or vector (left), treated with DMSO or 20 μM DRB for 16 hours (right) prior to 4Gy IR. Two hours after IR, the co-localization of Flag-PIF1 and γ-H2AX was analyzed by PLA.

Figure 7. SETX deficiency promotes template switching during BIR in an R-loop-dependent manner.

A. Summary of template switching events identified from both PCR analysis and LAM-HTGTS. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells with or without RNASEH1 expression were infected with I-SceI lentiviruses to induce DSBs. Five days after infection, EGFP-positive cells were sorted as BIR-positive events. BIR repair products were analyzed by PCR or LAM-HTGTS (Fig S4A

and S4B) to determine the occurrence of template switching. Template switching events are classified as “within reporter” (template switching occurs within the BIR reporter), “outside reporter” (template switching copies sequences outside of the BIR reporter on the same chromosome or from other chromosomes), and “plasmid” (template switching copies sequences from lentiviral plasmids that have been delivered to cells). The frequency of template switching was calculated as the number of total template-switching events, including “within reporter”, “outside reporter” and “plasmid”, divided by the total number of EGFP-positive clones.

B. Template switching events defined as “outside reporter” in A are displayed in Circos plot.

C. U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells with or without RNASEH1 expression were infected with I-SceI-expressing lentiviruses to induce DSBs. Two days after infection, cells were harvested for DRIP analysis at indicated locus.

D. Examples of “within reporter” template switching events in U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells as described in A.

E. Summary of tract length data of recovered BIR repair products in U2OS WT and *SETX*-KO EGFP-BIR-5085 reporter cells with or without RNASEH1 overexpression. Red lines mark the median tract length in each group.

F. The working model for hyper end resection at DSBs in SETX-deficient cells. SETX deficiency causes more R-loop formation on DSB ends and RAD52 recruitment, which in turn recruits XPF. Instead of MRE11, XPF generates nicks on the 5’ strand of the R-loops, followed by removal of the remaining 5’ strand towards DSB ends by MRE11 with its 3’ to 5’ exonuclease activity, and long-range resection of 5’ strand away from DSB ends by EXO1 and BLM/DNA2.

G. Illustration of template switching in SETX-deficient cells. SETX deficiency causes increased R-loop formation on the BIR repair templates. After strand invasion, BIR DNA synthesis would encounter R-loops on the repair templates accumulated due to SETX deficiency, which could stall BIR fork progression and cause displacement of the synthesizing strand from D-loops. Reassembly of collapsed BIR forks using templates at other genomic loci gives rise to template switching.

Fig 1

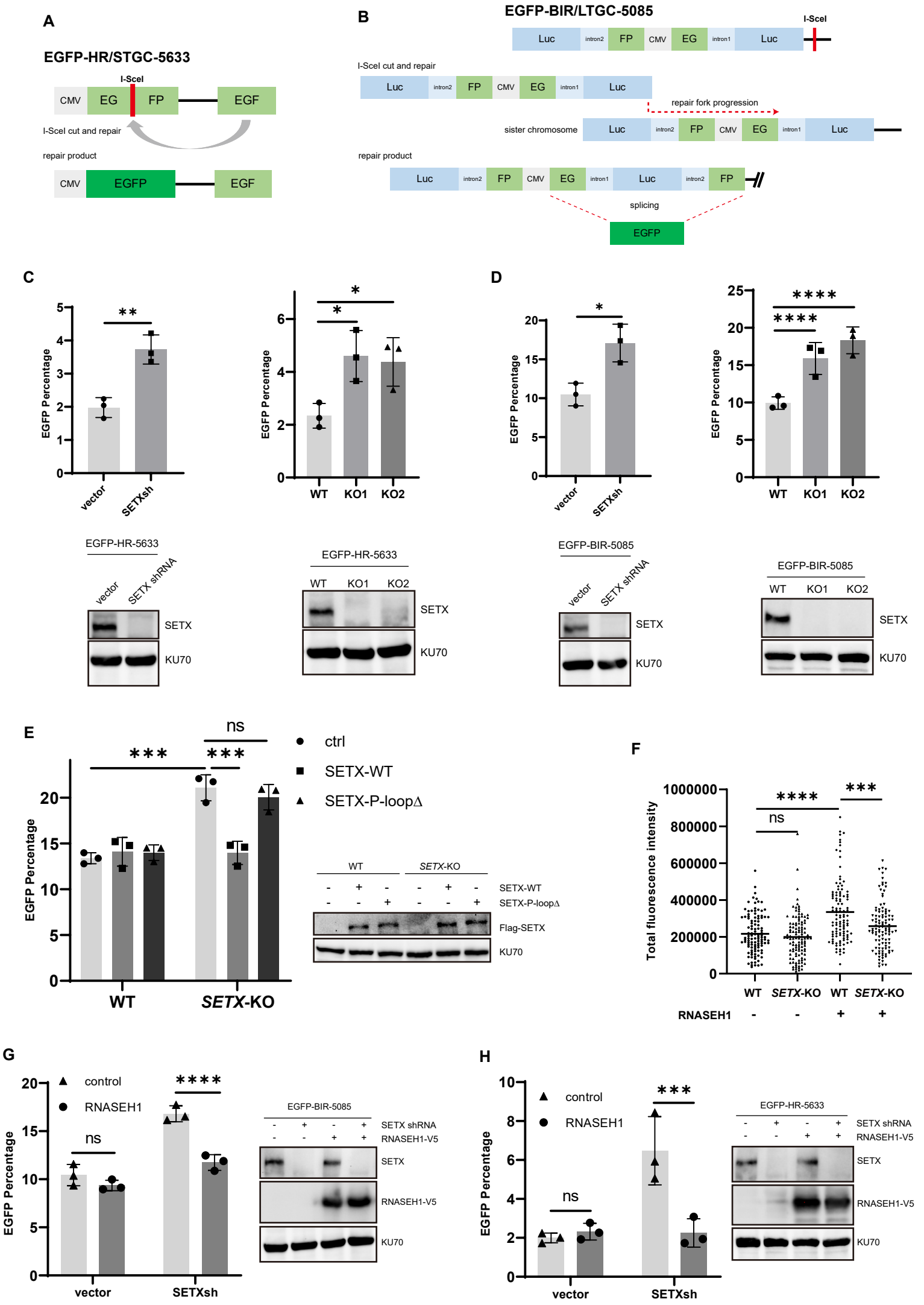

A
EGFP-HR/STGC-5633
I-SceI
CMV
EG
FP
EGF
I-SceI cut and repair
repair product
CMV
EGFP
EGF
B
EGFP-BIR/LTGC-5085
I-SceI
Luc
intron2
FP
CMV
EG
intron1
Luc
I-SceI cut and repair
repair fork progression
sister chromosome
repair product
splicing
EGFP
C
EGFP Percentage
vector
SETXsh
WT
KO1
KO2
EGFP-HR-5633
SETX shRNA
SETX
KU70
D
EGFP-BIR-5085
E
ns
ctrl
SETX-WT
SETX-P-loopΔ
WT
SETX-KO
Flag-SETX
F
Total fluorescence intensity
RNASEH1
G
control
RNASEH1
RNASEH1-V5
H

Fig 2

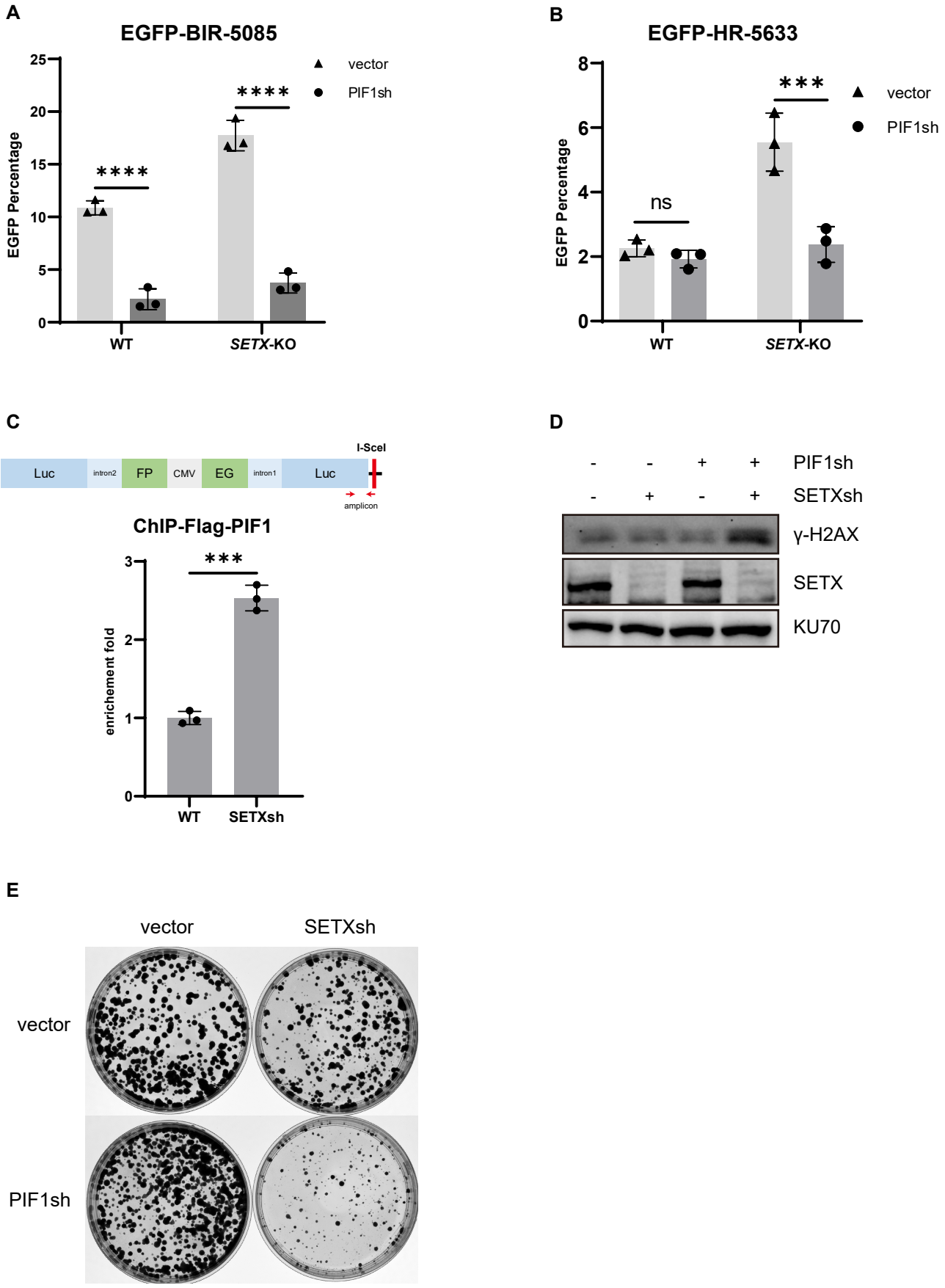

A
EGFP-BIR-5085
EGFP Percentage
25
20
15
10
5
0
****
****
vector
PIF1sh
WT
SETX-KO
B
EGFP-HR-5633
EGFP Percentage
8
6
4
2
0
ns
***
vector
PIF1sh
WT
SETX-KO
C
I-SceI
Luc
intron2
FP
CMV
EG
intron1
Luc
amplicon
ChIP-Flag-PIF1
enrichement fold
3
2
1
0
***
WT
SETXsh
D
PIF1sh
SETXsh
γ-H2AX
SETX
KU70
E
vector
SETXsh
vector
PIF1sh

Fig 3

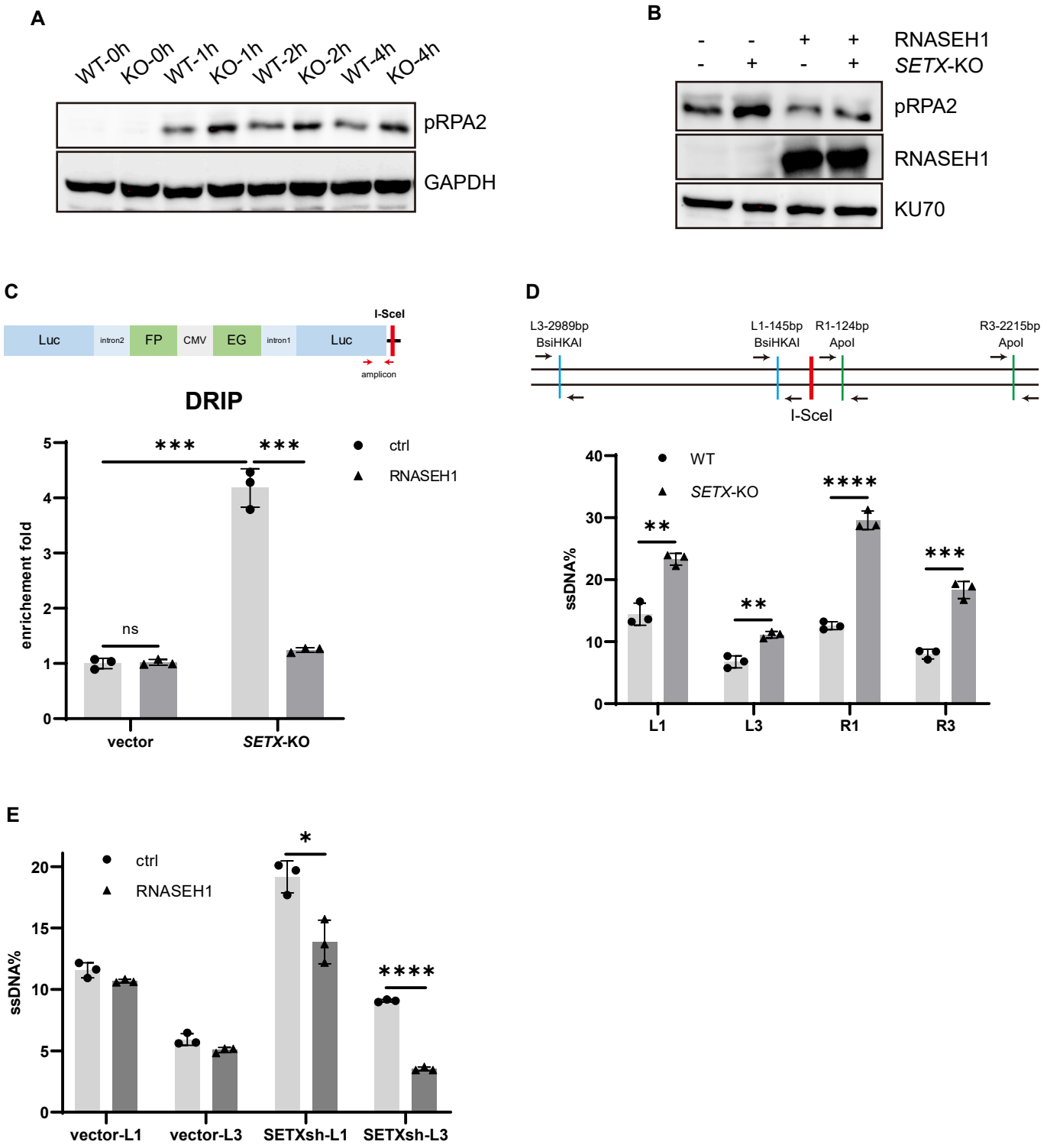

A
WT-0h
KO-0h
WT-1h
KO-1h
WT-2h
KO-2h
WT-4h
KO-4h
pRPA2
GAPDH
B
RNASEH1
SETX-KO
pRPA2
RNASEH1
KU70
C
I-SceI
Luc
intron2
FP
CMV
EG
intron1
Luc
amplicon
DRIP
enrichement fold
ctrl
RNASEH1
ns
vector
SETX-KO
D
L3-2989bp
BsiHKAI
L1-145bp
BsiHKAI
R1-124bp
ApoI
R3-2215bp
ApoI
I-SceI
ssDNA%
WT
SETX-KO
L1
L3
R1
R3
E
ssDNA%
ctrl
RNASEH1
vector-L1
vector-L3
SETXsh-L1
SETXsh-L3

Fig 4

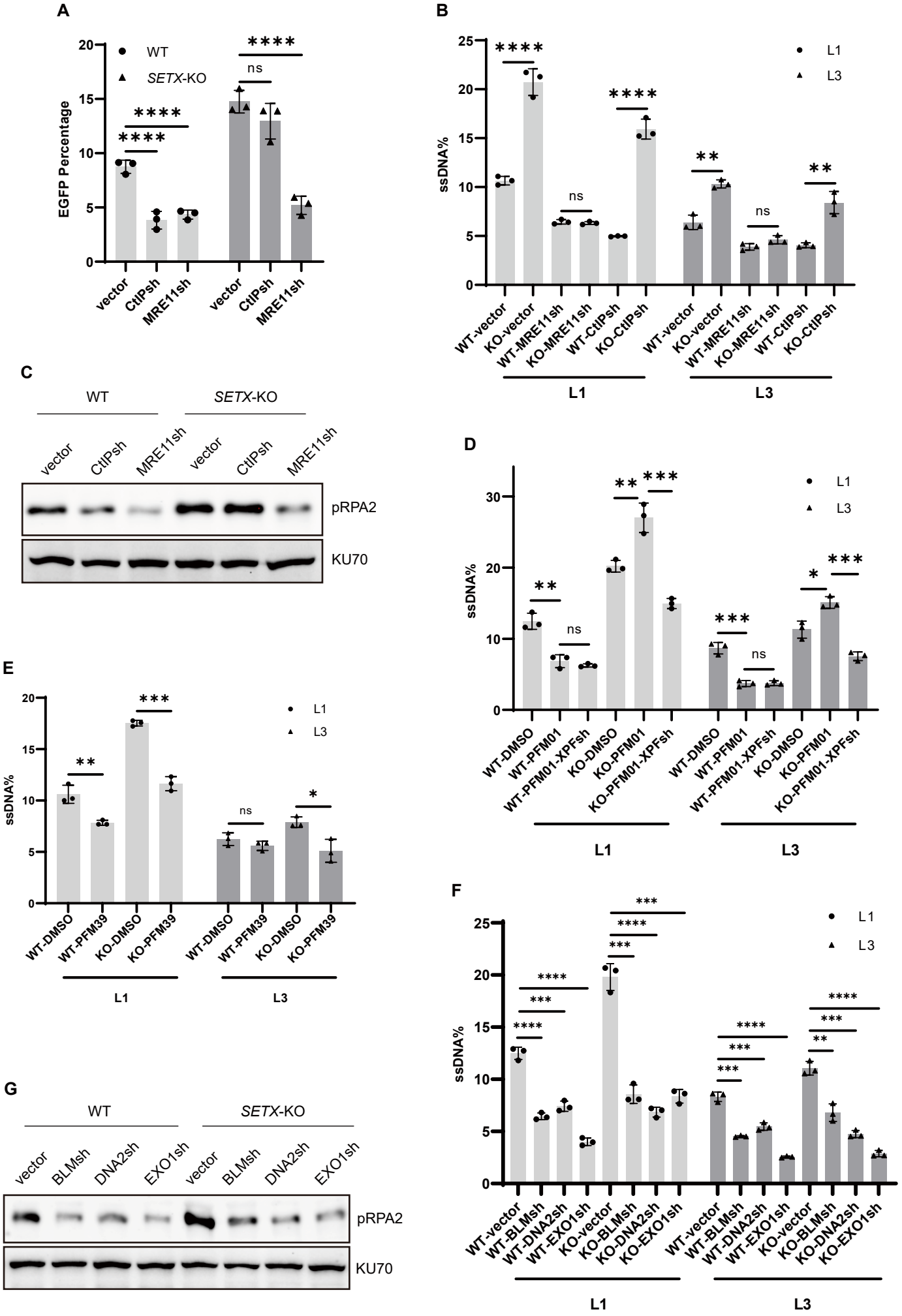

A
WT
SETX-KO
EGFP Percentage
vector
CtlPsh
MRE11sh
B
L1
L3
ssDNA%
WT-vector
KO-vector
WT-MRE11sh
KO-MRE11sh
WT-CtlPsh
KO-CtlPsh
C
pRPA2
KU70
D
WT-DMSO
WT-PFM01
WT-PFM01-XPFsh
KO-DMSO
KO-PFM01
KO-PFM01-XPFsh
E
WT-PFM39
KO-PFM39
F
WT-BLMsh
WT-DNA2sh
WT-EXO1sh
KO-BLMsh
KO-DNA2sh
KO-EXO1sh
G
BLMsh
DNA2sh
EXO1sh

Fig 5

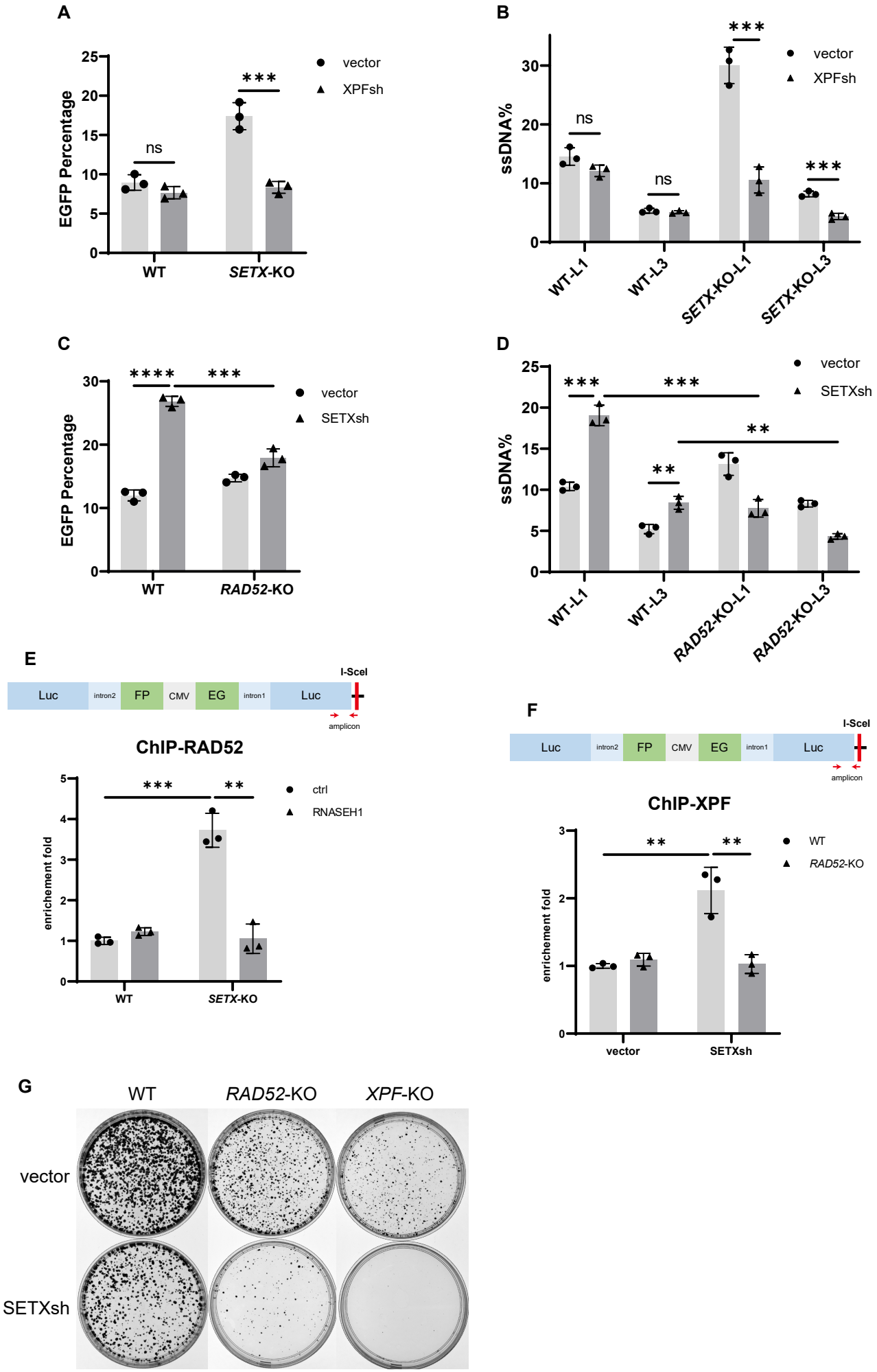

A
EGFP Percentage
vector
XPFsh
ns
***
WT
SETX-KO
B
ssDNA%
vector
XPFsh
ns
ns
***
***
WT-L1
WT-L3
SETX-KO-L1
SETX-KO-L3
C
EGFP Percentage
vector
SETXsh
****
***
WT
RAD52-KO
D
ssDNA%
vector
SETXsh
***
***
**
**
WT-L1
WT-L3
RAD52-KO-L1
RAD52-KO-L3
E
I-SceI
Luc
intron2
FP
CMV
EG
intron1
Luc
amplicon
ChIP-RAD52
enrichement fold
ctrl
RNASEH1
***
**
WT
SETX-KO
F
I-SceI
Luc
intron2
FP
CMV
EG
intron1
Luc
amplicon
ChIP-XPF
enrichement fold
WT
RAD52-KO
**
**
vector
SETXsh
G
WT
RAD52-KO
XPF-KO
vector
SETXsh

Fig 6

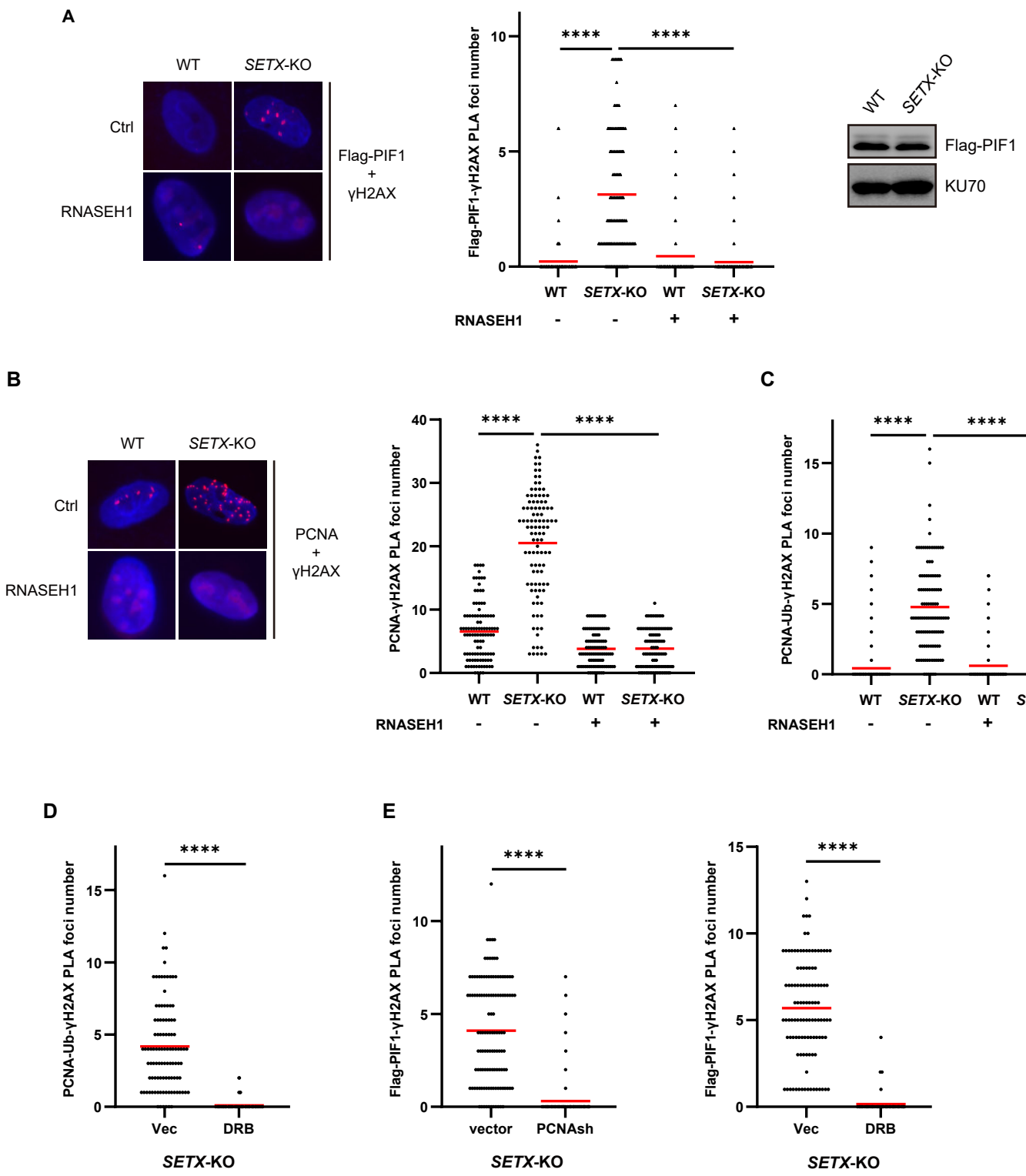
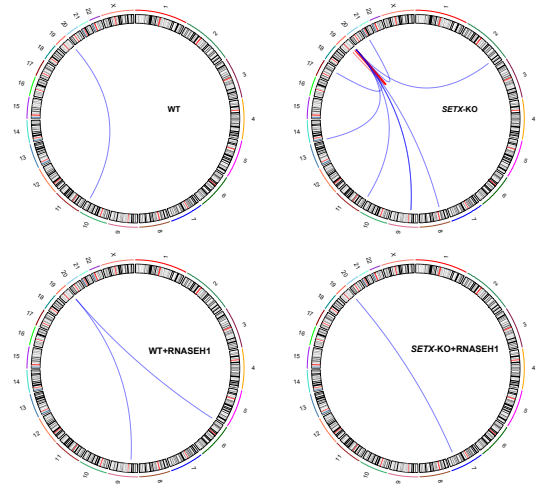

A
WT
SETX-KO
Ctrl
RNASEH1
Flag-PIF1
+
γH2AX
Flag-PIF1-γH2AX PLA foci number
****
****
WT
SETX-KO
WT
SETX-KO
RNASEH1
-
-
+
+
WT
SETX-KO
Flag-PIF1
KU70
B
WT
SETX-KO
Ctrl
RNASEH1
PCNA
+
γH2AX
PCNA-γH2AX PLA foci number
C
PCNA-Ub-γH2AX PLA foci number
D
PCNA-Ub-γH2AX PLA foci number
Vec
DRB
SETX-KO
E
Flag-PIF1-γH2AX PLA foci number
vector
PCNAsh
SETX-KO
Flag-PIF1-γH2AX PLA foci number
Vec
DRB
SETX-KO

Fig 7

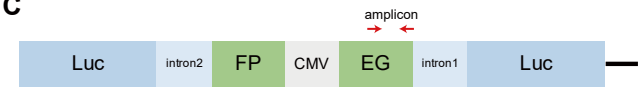
A

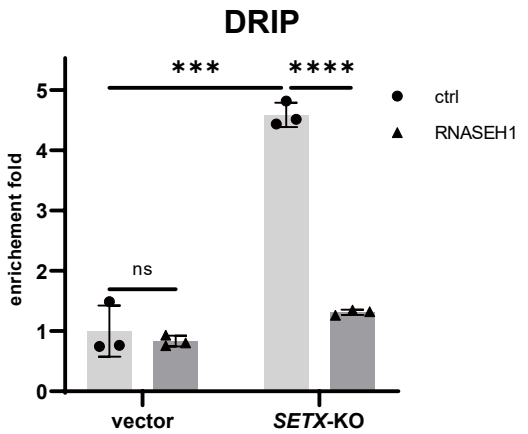
	total events	within reporter	outsite reporter	plasmid	frequency
WT	42	5 (11.9%)	1 (2.38%)	0	14.29%
KO	41	9 (21.95%)	13 (31.71%)	1	53.66%
WT+RNASEH1	48	2 (4.17%)	2 (4.17%)	3	14.58%
KO+RNASEH1	41	1 (2.44%)	1 (2.44%)	4	14.63%

B

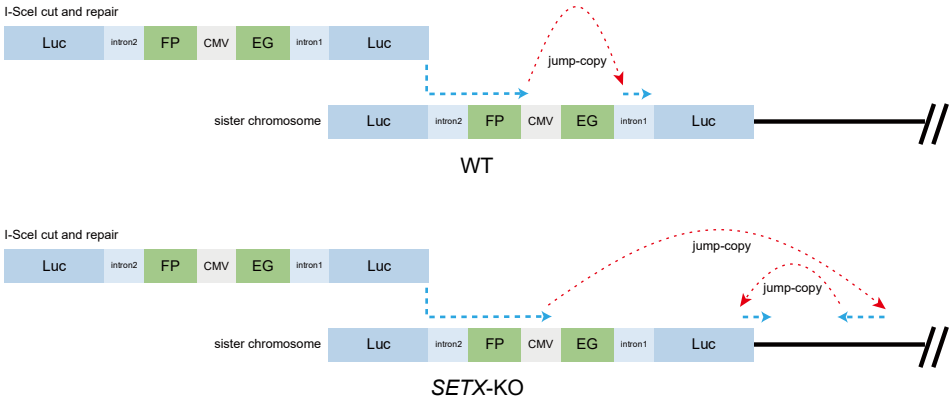


C

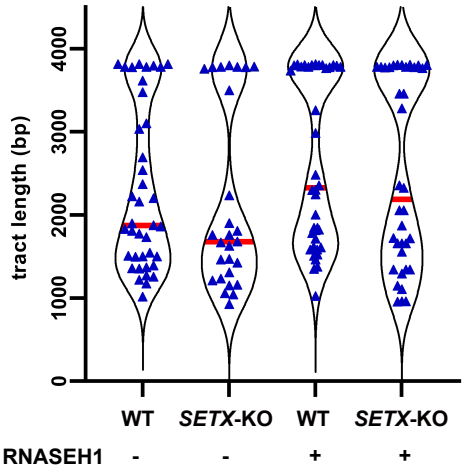




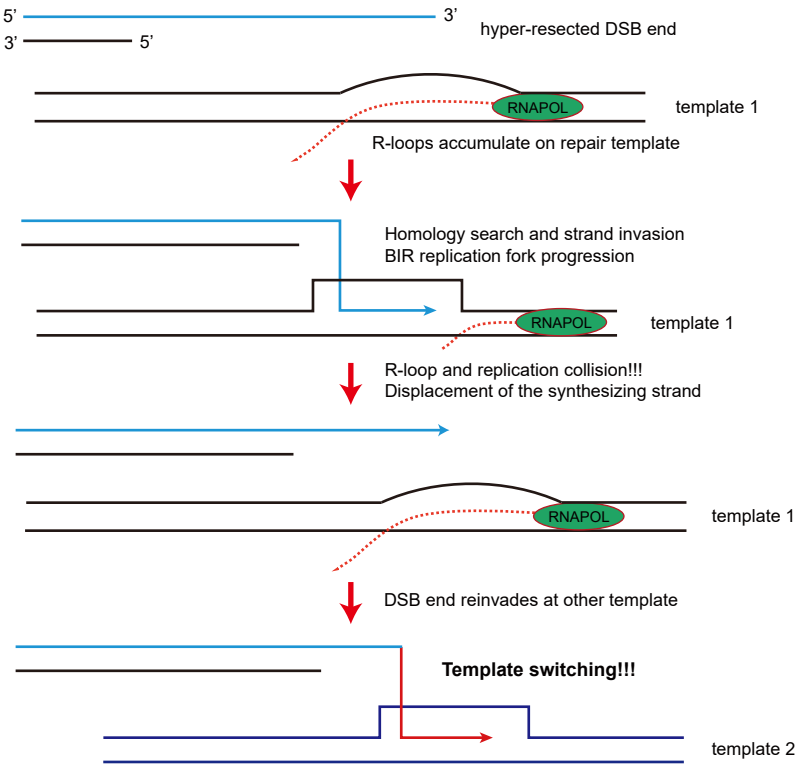
D



E



G



F

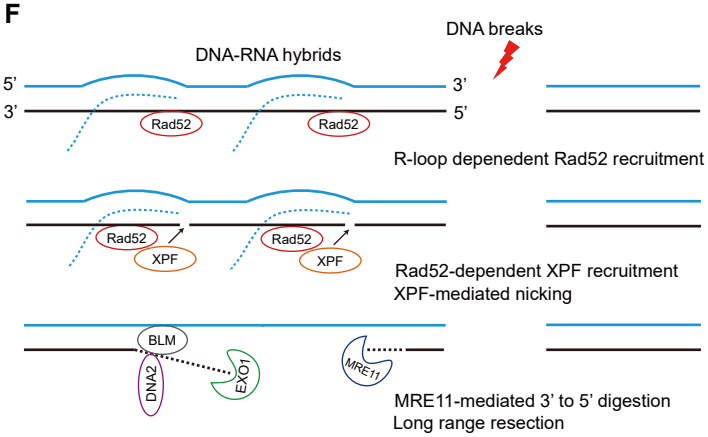


Fig S1

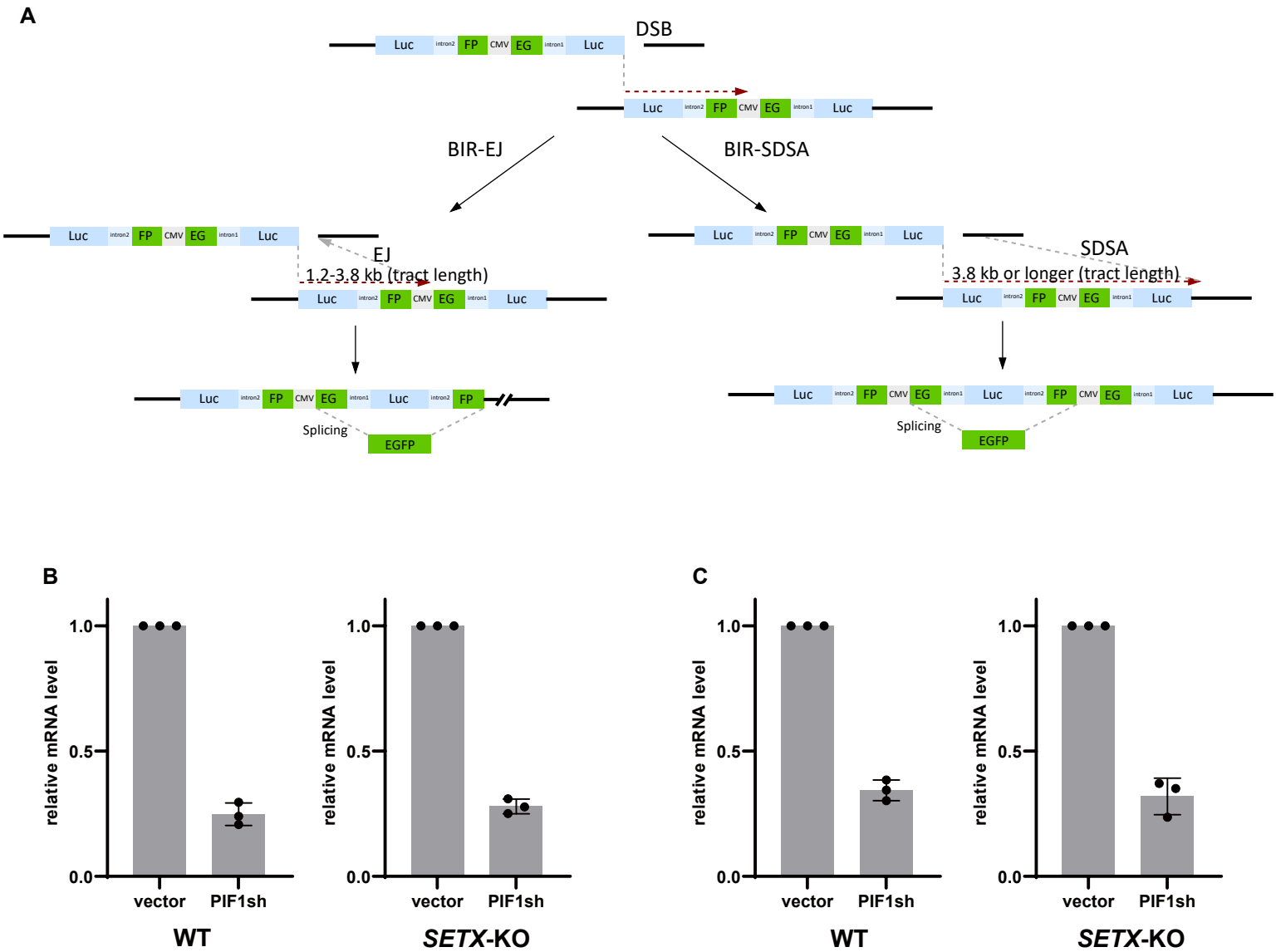


Fig S1 related to Fig 1 and Fig 2

Determining the dependence on essential BIR genes after SETX inactivation

A. Schematic drawing of repair products (BIR-EJ or BIR-SDSA) in EGFP-BIR-5085 reporter.

B,C qPCR results showing the efficiency of PIF1 shRNA in WT and *SETX*-KO U2OS EGFP-BIR-5085 and EGFP-HR-5633 reporter cells.

Fig S2

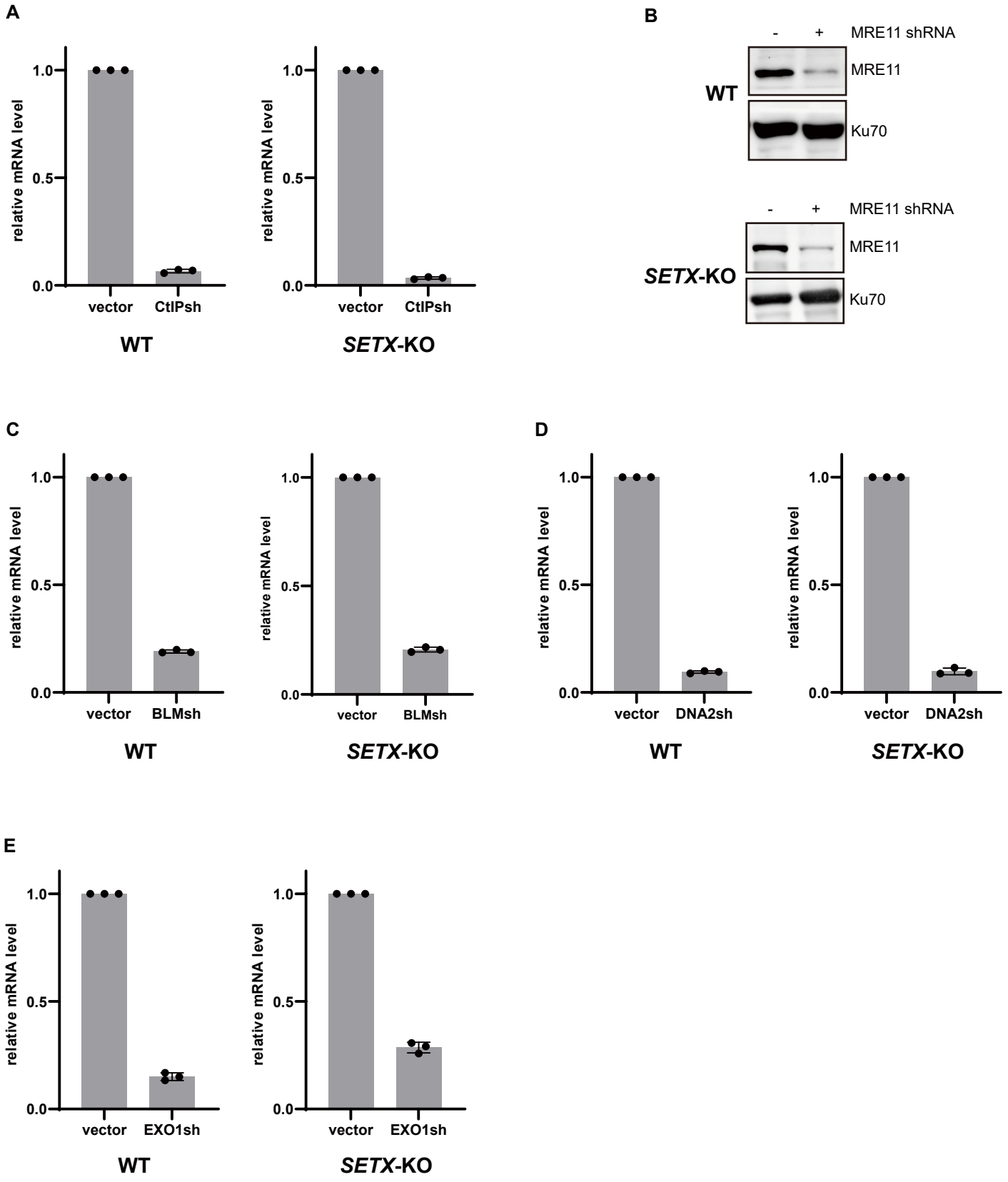


Fig S2 related to Fig 3 and Fig 4

Determining the dependence on canonical end resection pathway in SETX-deficient cells

A. qPCR results showing the efficiency of CtIP shRNA in WT and *SETX*-KO U2OS EGFP-BIR-5085 reporter cells.

B. Western blotting showing the efficiency of MRE11 shRNA in WT and *SETX*-KO U2OS EGFP-BIR-5085 reporter cells.

C,D,E. qPCR results showing the efficiency of BLM shRNA, DNA2 shRNA, and EXO1 shRNA in WT and *SETX*-KO U2OS EGFP-BIR-5085 reporter cells.

Fig S3

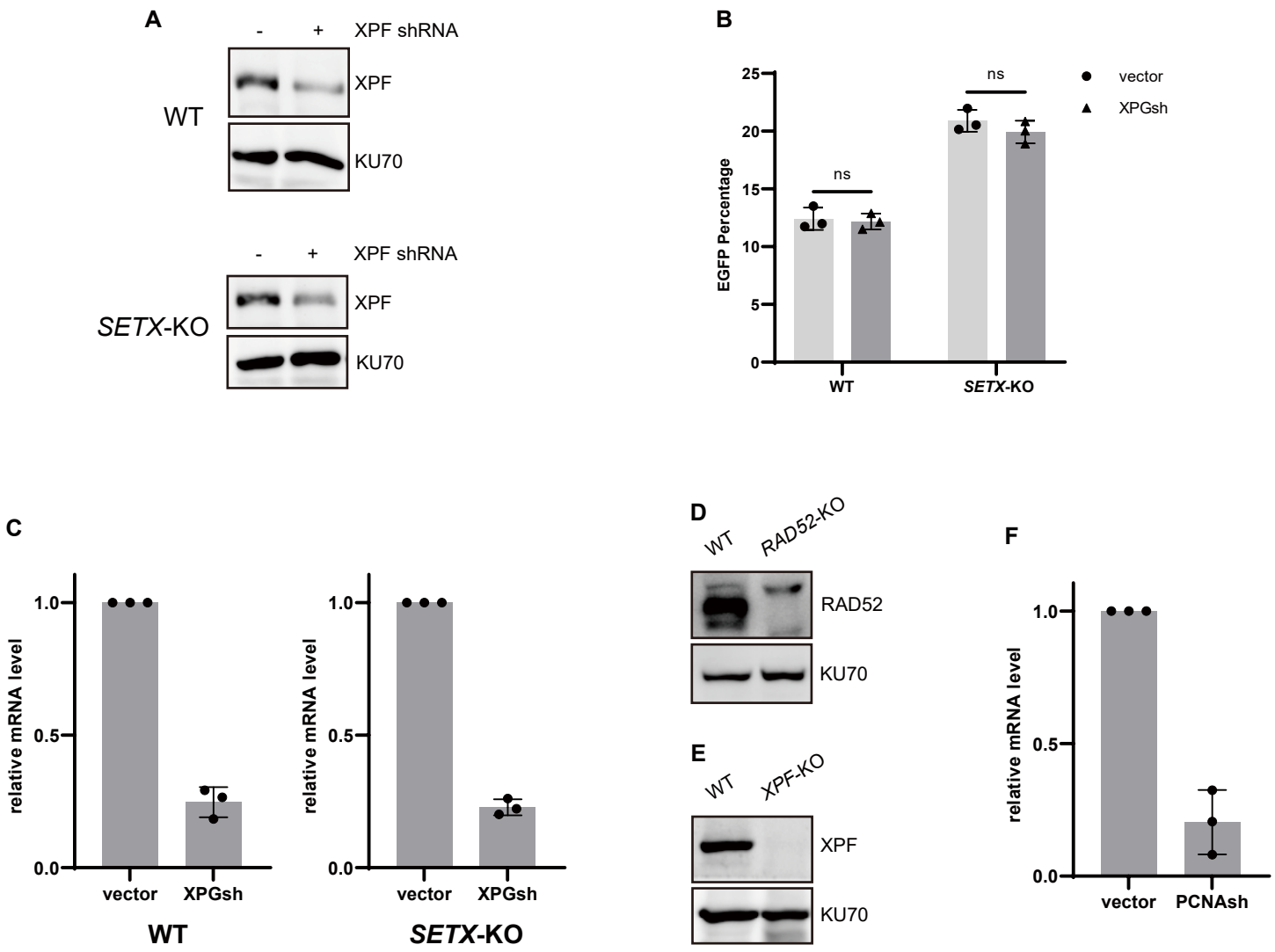


Fig S3 related to Fig 5

Determining the BIR dependence on XPF/RAD52 in SETX-deficient cells

A. Western blotting showing the efficiency of XPF shRNA in WT and *SETX*-KO U2OS EGFP-BIR-5085 reporter cells.

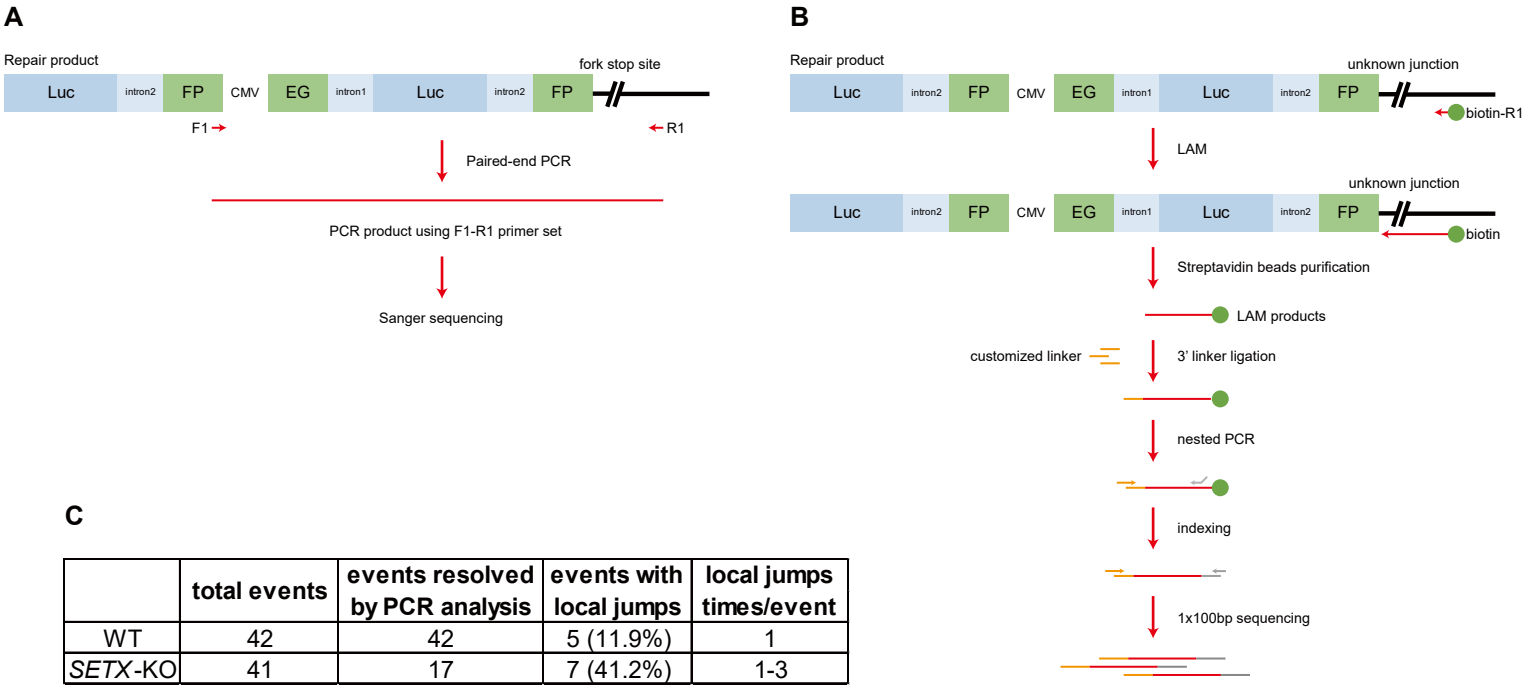
B. WT and *SETX*-KO EGFP-BIR-5085 reporter cells were infected with lentivirus encoding XPG shRNA or vector. Three days after infection, cells were infected with I-SceI-expressing lentivirus to induce DSBs. The percentage of EGFP-positive cells was quantified by FACS analysis five days after I-SceI induction.

C. qPCR results showing the efficiency of XPG shRNA in WT and *SETX*-KO U2OS EGFP-BIR-5085 reporter cells.

D,E. Western blotting showing the efficiency of RAD52 knockout and XPF knockout in EGFP-BIR-5085 reporter cells.

F. qPCR results showing the efficiency of PCNA shRNA in EGFP-BIR-5085 reporter cells.

Fig S4



C

	total events	events resolved by PCR analysis	events with local jumps	local jumps times/event
WT	42	42	5 (11.9%)	1
SETX-KO	41	17	7 (41.2%)	1-3

D

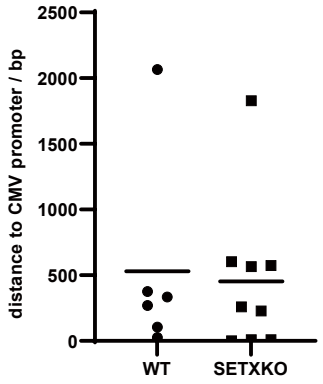


Fig S4 related to Fig 7

Development of LAM-HTGTS to analyze BIR repair product from U2OS EGFP-BIR-5085 reporter cells

A. Schematic drawing of PCR analysis in EGFP-BIR-5085 reporter cells.

B. Schematic drawing of LAM-HTGTS assay in EGFP-BIR-5085 reporter cells.

C. Summary of repair products which are resolved by PCR analysis.

D. Summary of the distance between the first jumping site to the CMV promoter in template-switching events. If the jumping happens within the CMV promoter, the distance will be scored as 0 bp. Black line marks the mean value.

Fig S5

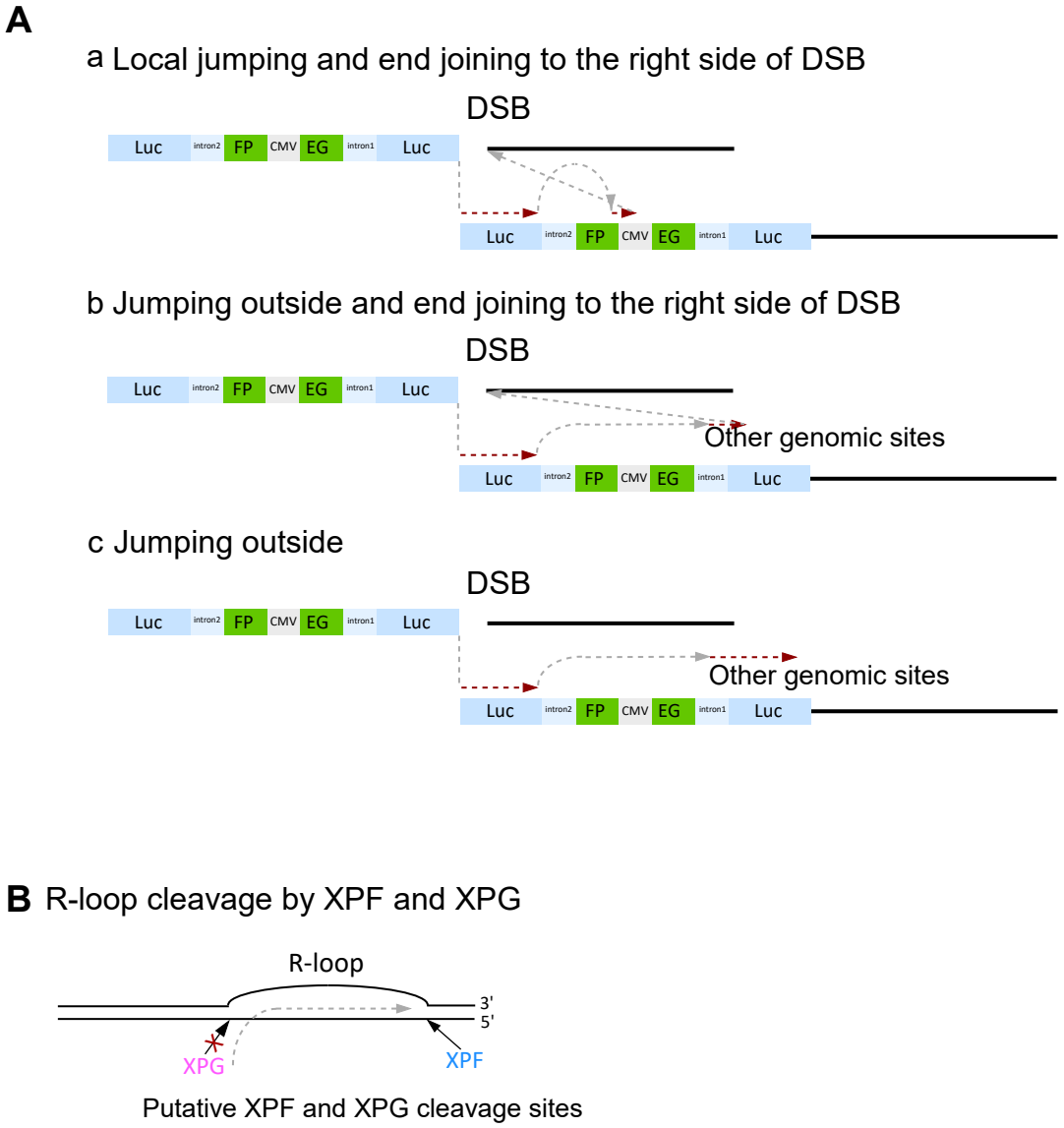


Fig S5 LAM-HTGTS application and XPF/XPG cleavage pattern on the R-loop substrate.

A. Different types of template switching during BIR scored by the EGFP-BIR-5085 reporter: a, local jumping within the BIR reporter; b, long-distance jumping to the genomic sites outside of the BIR reporter region with BIR completed by end joining back to the original cleavage site; c, long-distance jumping to the genomic sites outside of the BIR reporter region, causing translocation to other genomic sites.

B. Schematic drawing of the putative XPF and XPG cleavage sites on R-loop-associated DSB ends to initiate end resection.

Fig S6

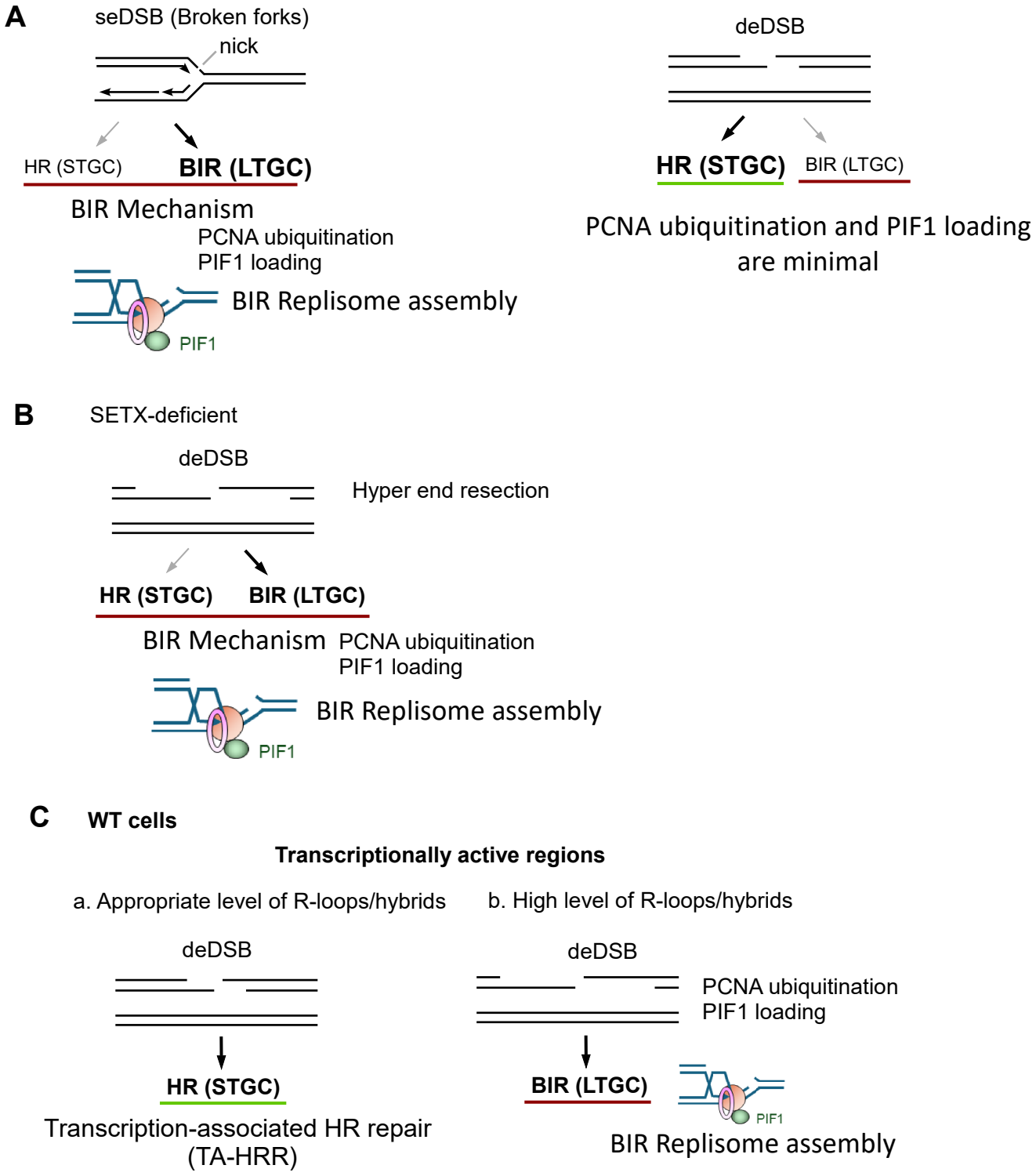


Fig S6 SETX and R-loop modulate DSB repair pathway choice.
A. Models showing BIR activation at seDSB on broken forks (left) and at deDSB (right).
B. Models showing BIR activation at deDSB when SETX is deficient.
C. Models showing how R-loop causes BIR activation at deDSB.

Fig S7

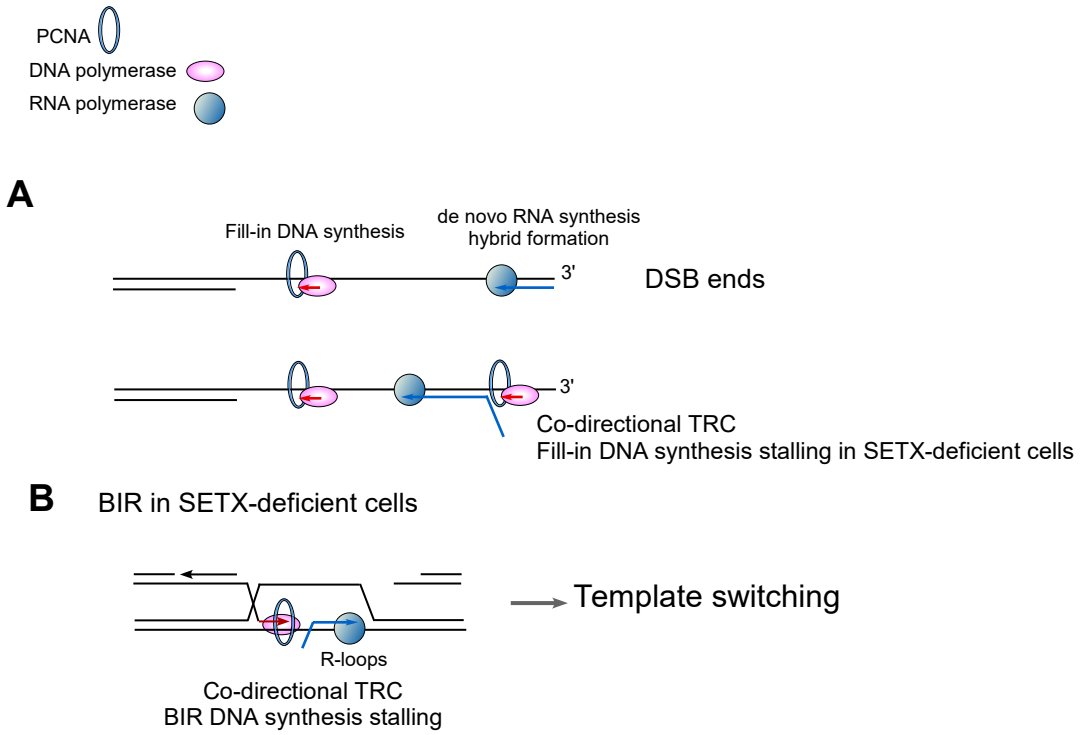


Fig S7 Co-directional DNA synthesis and transcription occur at DSB ends and during BIR.

A. Fill-in DNA synthesis primed by Pola-primase occurs at 3' ssDNA overhangs at DSBs, which is stalled in SETX-deficient cells.

B. BIR DNA synthesis is co-directional with transcription in the donor, causing replication stalling in SETX-deficient cells. Co-directional transcription and replication conflict (TRC) in the BIR donor leads to a marked increase of template switching in SETX-deficient cells.

Fig S8

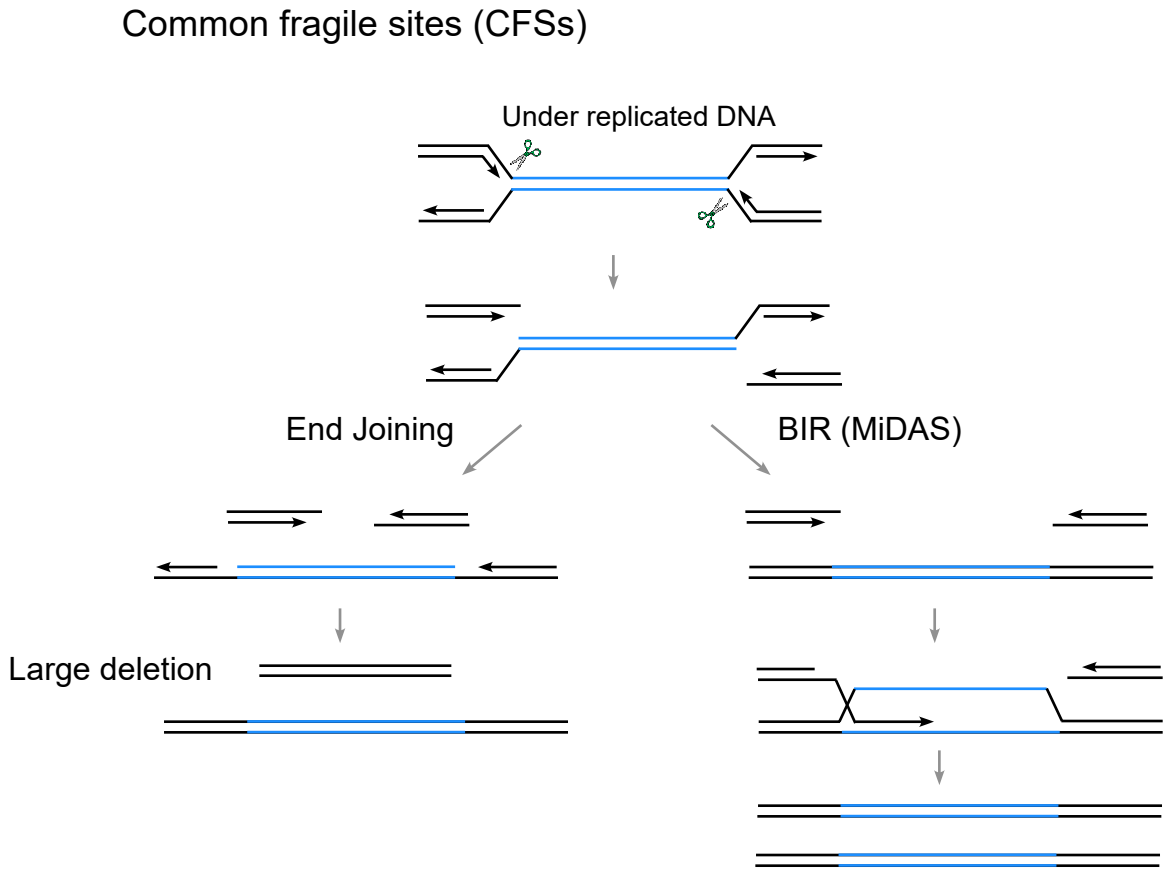


Fig S8 Models showing BIR prevents large deletions at CFSs through MiDAS. Schematic drawing of repair products by end joining and BIR to repair DSBs at CFSs in mitosis.