Conditional knockout of RAD51-related genes in *Leishmania major* reveals a critical role for homologous recombination during genome replication

Jeziel D. Damasceno[1*], João Reis-Cunha[2], Kathryn Crouch[1], Craig Lapsley[1], Luiz R. O. Tosi[3], Daniella Bartholomeu[2] and Richard McCulloch[1*]

1. The Wellcome Centre for Integrative Parasitology, University of Glasgow, Institute of Infection, Immunity and Inflammation, Sir Graeme Davies Building, 120 University Place, Glasgow, G12 8TA, United Kingdom.

2. Laboratório de Imunologia e Genômica de Parasitos, Departamento de Parasitologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil

3. Department of Cell and Molecular Biology, Ribeirão Preto Medical School, University of São Paulo; Ribeirão Preto, SP, Brazil.

* To whom correspondence should be sent:
Jeziel.Damasceno@glasgow.ac.uk
Richard.mcculloch@glasgow.ac.uk

Keywords: *Leishmania*; DNA replication; homologous recombination; Rad51; Rad51 paralogue; CRISPR-Cas9; DiCre

Abstract

Homologous recombination (HR) has an intimate relationship with genome replication, both during repair of DNA lesions that might prevent DNA synthesis and in tackling stalls to the replication fork. Recent studies led us to ask if HR might have a more central role in replicating the genome of *Leishmania*, a eukaryotic parasite. Conflicting evidence has emerged regarding whether or not HR genes are essential, and genome-wide mapping has provided evidence for an unorthodox organisation of DNA replication initiation sites, termed origins. To answer this question, we have employed a combined CRISPR/Cas9 and DiCre approach to rapidly generate and assess the effect of conditional ablation of RAD51 and three RAD51-related proteins in *Leishmania major*. Using this approach, we demonstrate that loss of any of these HR factors is not immediately lethal, but in each case growth slows with time and leads to DNA damage, accumulation of cells with aberrant DNA content, and genome-wide mutation. Despite these similarities, we show that only loss of RAD51 and RAD51-3 impairs DNA synthesis, and that the factors act in distinct ways. Finally, we reveal that loss of RAD51 has a profound effect on DNA replication, causing loss of initiation at the major origins and increased DNA synthesis at subtelomeres. Our work clarifies questions regarding the importance of HR to survival of *Leishmania* and reveals an unanticipated, central role for RAD51 in the programme of genome replication in a microbial eukaryote.

Introduction

Homologous recombination (HR) has critical roles in the genome maintenance of all organisms, mainly through repair of double stranded DNA breaks (1). HR is a multistep repair process initiated by resection of the ends of double-stranded DNA breaks to generate single stranded DNA overhangs. This processing provides access to a key player in HR: the Rad51 recombinase (RecA in bacteria, RadA in archaea)(2), which catalyses invasion of the single-stranded DNA into intact homologous duplex DNA, allowing template-directed repair of the broken DNA site. During evolution, duplications of the Rad51 gene have resulted in so-called Rad51 paralogues (3), a set of factors that are found in variable numbers in different organisms and whose spectrum of roles remain somewhat undefined, at least in part because they can belong to a number of protein complexes. Nonetheless, Rad51 paralogues have been implicated in directly modulating HR, acting on Rad51 HR intermediates (4-6), and in wider repair activities for cell cycle progression (7, 8). HR reactions mediated by Rad51 (9-12) and modulated by the Rad51 paralogues (13) are also required for resolving DNA replication impediments, by promoting protection and restart of stalled replication forks during replication stress. An even more intimate association between HR and DNA replication has been described in bacteria and archaea, where RecA (14-17) and RadA (18) can mediate DNA replication when origins (the genome sites where DNA synthesis begins during replication) have been removed.

In addition to its roles in promoting genome stability, HR can drive to genome variation, which can cause diseases (19), as well as being a means for targeted sequence change during growth, such as during mating type switching in yeast (20). Genome variation due to HR is found widely in trypanosomatid parasites, which are single-celled microbes that cause human and animal diseases worldwide. In *Trypanosoma brucei*, HR factors have been clearly implicated in the directed recombination of Variant Surface Glycoprotein (VSG) genes during host immune evasion by antigenic variation (21), as well as in maintenance of the massive subtelomeric VSG gene archive (22, 23). In *T. cruzi*, HR has been suggested to be a driver of variability in multigene families (24, 25) and in cell hybridisation (26). Finally, in *Leishmania,* HR related factors have been implicated in mediating the formation or maintenance of episomes, which appear to form stochastically, arise genome-wide and have been implicated in acquisition of drug resistance (27-32). Whether the same roles for HR extend to widespread, stochastic formation of aneuploidy is unknown (33), but this other form of genome-wide variation has also been implicated in adaptation of the parasite, such as during life cycle transitions and in response to drug pressure (34-38).

Despite emerging evidence for HR roles in *Leishmania* genome change, is it possible that the reaction has wider and deeper functions in genome maintenance and transmission in the parasite? One reason for asking this question is recent observations suggesting that origin number and distribution in *Leishmania* is unusual, since one study detected only a single site of DNA replication initiation per chromosome (39), while a later study suggested >5000 sites (40). These data indicate either a pronounced paucity or huge overabundance of origins relative to all other eukaryotes. Alternatively, the disparity in the datasets may be due to a widespread, unconventional route for initiation of DNA synthesis, acting alongside a small number of conventional origins, perhaps indicating novel strategies for DNA replication that may link with genome plasticity (41). A second reason for asking about the importance of HR for genome transmission in *Leishmania* is other work that has led to uncertainty about the importance of HR factors for survival of the parasite. *Leishmania* encodes a highly conserved, canonical Rad51 recombinase (30, 42, 43), as well three Rad51 paralogues, referred to as RAD51-3, RAD51-4 and RAD5-6 (29), a slightly smaller repertoire of non-meiotic Rad51-related proteins than is found in *T. brucei* (44-47). In *L. donovani* (48), unlike in *L. infantum* (30), it has proved impossible to make RAD51 null mutants. Furthermore, while null mutants of RAD51-4 are viable in *L. infantum*, RAD51-3 has been described as being essential, and RAD51-6 nulls were not recovered in the same experiments (29). Mutation in Rad51 or its relatives has never been shown to be lethal in any single celled eukaryote, notably including *T. brucei* (49), or in prokaryotes, making these observations in *Leishmania* particularly striking.

In this work we sought to resolve the question marks over essentiality of HR factors in *Leishmania* and to test for roles in DNA replication by using conditional gene knockout (KO), comparing the short- and long-term effects of ablating RAD51 and each of its three RAD51 paralogues. Our data show that loss of each gene is, over time, increasingly detrimental to *Leishmania* fecundity, demonstrating that black and white

definitions of essential or non-essential are too limiting for HR genes in the parasite. In addition, we show that the functions provided by the RAD51 paralogues are non-overlapping in *Leishmania*, and we reveal that RAD51 plays an unexpected, central role in genome replication, where it is required for proper activation of DNA replication origins.

Results

A combined CRISPR/Cas9 and DiCre approach for assessment of gene function in *L. major*

In order to compare the effects of ablating RAD51 and each of the three known *L. major* RAD51 paralogues, we adopted a rapid approach to generate cell lines for conditional induction of a gene KO. For this, we used a cell line constitutively expressing Cas9 and DiCre (Fig.1A). In this strategy, we first used CRISPR/Cas9-mediated HR to exchange the endogenous copy of the genes by a copy flanked by *loxP* sites. In addition, each construct translationally fused copies of the HA epitope to the C-terminus of the gene's ORF. PCR showed this approach to be very efficient for RAD51 and the three RAD51 paralogues, since selection using only puromycin resulted in all wild type (WT) copies of each gene being replaced by floxed and tagged versions after a single transformation (Fig.S1). Because RAD51 and the RAD51 paralogue mutants may generate similar phenotypes (44, 46), since each contributes to HR (29), we used the same approach to modify the *L. major* gene encoding the orthologue of *T. brucei* PIF6. This factor has not been functionally examined in *Leishmania*, but in *T. brucei* PIF6 is the sole known nuclear Pif1 helicase homologue (50). Since Pif1 helicases have been implicated in modulating DNA replication passage through barriers and during termination(51, 52), and thus operate in distinct aspects of nuclear genome maintenance compared with Rad51 paralogues, we considered this gene could provide a valuable control for the effects of conditional gene KO of the Rad51-related proteins. A number of attempts failed to replace all WT endogenous *L. major PIF6* gene copies with HA-tagged versions, whereas each copy could be floxed with untagged gene variants (Fig.S1). Growth curves showed that addition of *loxP* sites or the HA tag did not lead to any significant growth impairment for any of the five genes (Fig.S2).

Next, KO induction of each gene was attempted by rapamycin-mediated DiCre activation in logarithmically growing cultures of each cell line (Fig.1A, Fig.S3). PCR analysis using DNA from cells after a number of induction rounds ('passages'), where cells were grown from low to high density and repeated by dilution, showed that complete gene excision was achieved after passage 2 for all genes (Fig.S3). Controls without addition of rapamycin showed no gene excision, and unexcised gene copies were undetectable even after >15 passages in the presence of rapamycin (Fig.S3). The rapidity of DiCre mediated loss of the gene products was confirmed by western blotting (RAD51 and paralogues) and RT-PCR (PIF6): signal for all four HA-tagged proteins was no longer detectable after 48 h of the second round of KO induction (Fig.1B), and *PIF6* cDNA could not be PCR-amplified (Fig.1C). KO induction of *RAD51* and of each of the *RAD51*

paralogues, but not of *PIF6*, resulted in increased levels of γH2A (53) in western blotting analysis (Fig.1D), suggesting accumulation of nuclear DNA damage after loss of any *L .major* RAD51-like protein, but not after ablation of PIF6 (at least during unperturbed growth; see below). To attempt to answer the question of whether or not RAD51 and the RAD51 paralogues are essential in *Leishmania* (29, 48), we measured growth of the parasites for a prolonged period after DiCre-induced gene excision (Fig.1E). At least until passage 4, no growth defect was seen due to loss of RAD51, any of the RAD51 paralogues or PIF6. However, when kept in culture for longer periods (Fig.1E), the *RAD51* KO cells and each of the *RAD51* paralogues KO cells, but not the *PIF6* KO cells, showed marked growth defects, suggesting HR factors that contribute to the catalysis of homology-directed strand exchange might be critical for long-term *Leishmania* viability in culture. Accordingly, flow cytometry analysis showed that prolonged cultivation after KO induction of *RAD51* and the *RAD51* paralogues, but not *PIF6*, resulted in an increased proportion of cells with less than 1C DNA (Fig.S4), suggesting increased genomic instability, perhaps reflecting the increased levels of γH2A. Taken together, the phenotypes seen after induced loss of the five genes suggest some overlap in functions of RAD51 and its relatives, but a distinct role for PIF6. In addition, the PIF6 data demonstrate that prolonged exposure to rapamycin, or effects of DiCre and Cas9 expression, are negligible in these conditions.

Loss of RAD51 or RAD51-3, but not RAD51-4 or RAD51-6, impairs DNA synthesis in *L. major*

To ask if the impaired growth seen in four of the five induced KO cells is due to a common defect, we tested the extent of DNA synthesis after each gene deletion. To do this, rapamycin induced and uninduced cells were subjected to a short pulse of IdU labelling followed by immunostaining under denaturing conditions and flow cytometry detection, allowing us to track the level and pattern of DNA synthesis in each cell cycle stage (Fig.2A). Loss of RAD51 or RAD51-3, but not loss of RAD51-4, RAD51-6 or PIF6, resulted in a reduced percentage of IdU-positive cells in the population. Quantification of IdU signal in individual S-phase cells at 48 and 72 h of the second round of KO induction confirmed these effects (Fig.2B): a significant reduction in IdU fluorescence was found at both time points in the rapamycin-induced *RAD51* and *RAD51-3* KO cells compared with their cognate uninduced controls, whereas no such reduction was seen after KO of *RAD51-4*, *RAD51-6* or *PIF6*. These data suggest that only loss of RAD51 or RAD51-3 affects DNA synthesis, meaning the growth impairment seen after loss of RAD51 and its relatives, though similar in extent, might not have a common basis, or loss of DNA synthesis is not the main reason for growth reduction after the induced KO of RAD51 or RAD51-3.

Next, we asked if the observed increases in DNA damage after induced KO of the *RAD51*-like genes all had the same basis by examining γH2A levels across the cell cycle. For this, rapamycin induced and uninduced cells were arrested in G1 using 5 mM hydroxyurea (HU) and then released from arrest by removing HU, sampling at the point of arrest and at various times after release for western blotting (Fig.2C). The patterns of γH2A accumulation revealed notable differences in the cell cycle functions of the five genes (Fig.2C). KO

induction of *RAD51* or *RAD51-3* resulted in a pronounced increase of γH2A levels in cells navigating through S-phase up to G2/M, suggesting roles related to the resolution, before cell division, of genome injuries that arise during DNA replication. In contrast, loss of *RAD51-4* or *RAD51-6* did not show any clear evidence for increased γ H2A signal compared with uninduced controls after HU release, suggesting a more limited contribution to tackling replication-associated damage, which seems consistent with the absence of changes in IdU uptake after KO. *PIF6* KO displayed a yet further difference, with increased levels of γH2A only ~6 h after HU release (Fig.2C), when much of the population had passed through S-phase (Fig.S5). These data indicate that loss of PIF6 does in fact result in nuclear damage, but this is more limited than is seen after loss of the RAD51-like genes, and suggests that if the putative helicase has a role in resolving replication problems, this is concentrated in the final steps of DNA replication or even in post-replication steps of the cell cycle.

DNA content in all the samples analysed for γH2A levels was next analysed by flow cytometry (Fig.S5). Intriguingly, no pronounced changes in cell cycle progression after release from HU arrest were observed upon KO induction of any of the HR genes or *PIF6* (Fig.S5). Thus, *L. major* cell cycle checkpoints do not seem to be enacted by the gene KOs, despite clear DNA damage accumulation after loss of RAD51 or its paralogues. Altogether, these data suggest that *Leishmania* RAD51 and RAD51-3, specifically amongst the genes examined, have roles in promoting effective DNA synthesis and their absence results in increased nuclear genome damage during S-phase.

Ablation of RAD51, RAD51-3 or RAD51-4 results in genome-wide mutagenesis

We next sought to determine if loss of the HR genes results in genome instability by using short-read Illumina sequencing of DNA from *RAD51*, *RAD51-3* and *RAD51-4* KO cells after growth for two and six passages in the presence of rapamycin, as well as in the same cells grown without rapamycin. In each case, mapping of reads to the genome showed specific loss of sequence around the *loxP*-flanked gene, confirming KO induction (Fig.3A). By measuring the number of single nucleotide polymorphisms (SNPs) in the induced and uninduced cells after passage two and six, we then calculated the number of new SNPs that arose during the intervening growth period (Figs.3B,C). For each KO, loss of the HR gene resulted in an increased number of new SNPs over four passages relative to the uninduced control cells (Fig.3B). Strikingly, when new SNP density was plotted individually for each of the 36 chromosomes, it became apparent that the increase in SNPs after gene KO was not random across the genome, as the smaller chromosomes accumulated a higher density of new SNPs than the larger chromosomes (Fig.3C). In addition, there was a notable peak in new SNP density proximal to the 'strand switch regions' (SSRs) within each chromosome where multigene transcription initiation and/or termination occurs (Fig.3D), and a subset of which are where MFA-seq has mapped DNA replication initiation (i.e. are predicted replication origins)(39). This SSR proximity mapping also revealed a difference between the three gene KOs: following loss of *RAD51* a decrease in SNP accumulation was seen at both origin-active and -inactive SSRs relative to

uninduced controls, whereas the opposite effect was observed after *RAD51-3* KO, and no change in SSR-proximal SNP density was found after *RAD51-4* KO. Thus, the two HR factors whose loss was found to affect global DNA synthesis were also seen to affect SNP accumulation around SSRs, unlike *RAD51-4* KO, which did not show an effect on DNA synthesis.

To ask if the effects described above are limited to SNPs, we also measured insertions and deletions (InDels) in the same cells (Fig.S6). In this case, only KO of *RAD51-3* was found to result in an increase in new InDels across the genome (Fig.S6A), though even here the effect was modest. Nonetheless, the same pattern of increased new mutation density in the small chromosomes relative to the large was detectable in all cells (Fig.S6B). Around the SSRs, a clear peak of new InDels in the uninduced cells was less apparent than was seen for new SNPs (compare Fig.S6C with Fig.3D), though InDels appeared to accumulate around both origin-active and -inactive SSRs upon *RAD51* KO (Fig.S6C). In contrast, such a trend was seen only at origin-active SSRs upon *RAD51-3* KO, and no difference was seen in *RAD51-4* KO cells compared with controls (Fig.S6C). In total, therefore, InDel accumulation upon HR KO was more modest than SNPs, but any distinct accumulation around SSRs was most clearly detected after loss of RAD51 or RAD51-3, the two factors our data implicate in global DNA synthesis.

To examine the effects of loss of either RAD51 or RAD51-3 further, we examined survival of the KO cells relative to uninduced controls in the presence of increasing concentrations of phleomycin and camptothecin, both of which cause DNA double-strand breaks, and HU, which inhibits ribonucleotide reductase and impairs DNA synthesis (Fig.S7). As expected for predicted DNA double-strand break repair factors, KO induction of either *RAD51* or *RAD51-3* lead to increased sensitivity to phleomycin and camptothecin. However, only *RAD51-3* KO led to increased sensitivity to HU, further suggesting distinct roles for RAD51 and RAD51-3 roles during DNA synthesis. Consistent with this interpretation, differences were also observed in the patterns of SNP and InDel accumulation after exposure to and release from HU treatment (Fig.S8). Both genome-wide (Figs.S8A,B) and in each chromosome (Figs.S8C,D), fewer HU-induced new SNPs or InDels were detected in *RAD51* KO cells compared with uninduced, whereas KO of *RAD51-3* resulted in higher levels of HU-induced SNPs or InDels relative to uninduced cells. Despite the global decrease in HU-induced SNPs after loss of RAD51, SNPs were found to increase around SSRs in *RAD51* KO cells after HU treatment compared with uninduced cells (Fig.S8E), whereas no such effect was seen for InDels (Fig.S8F). Again, this phenotype differed from that seen after RAD51-3 loss: no accumulation of SSR-proximal HU-induced SNPs was seen in the *RAD51-3* KO cells compared with uninduced (Fig.S8E), whereas a very modest enrichment of InDels was observed (Fig.S8F). Taken together, these HU data reinforce the view that RAD51 and RAD51-3 play distinct roles in maintenance of the *L. major* genome, despite loss of either causing impaired DNA synthesis.

Generation of conditional double gene mutants by CRISPR/Cas9 and DiCre in *Leishmania*

To date, the analysis of single gene conditional KOs has implicated only RAD51 and RAD51-3 amongst the four HR factors in DNA synthesis. One explanation may be that RAD51-4 and RAD51-6 act redundantly in *Leishmania*. To test this, we used the combined CRISPR/Cas9, DiCre approach to attempt to make conditional double gene KOs of the Rad51-paralogues in all possible combinations (Fig.4A). First, CRISPR/Cas9 was used to generate floxed copies of either *RAD51-3* or *RAD51-4* using gene-specific puromycin-resistance constructs (described above). Next, the resulting cell lines were subjected to a second round of CRISPR/Cas9 engineering to delete the endogenous copies of *RAD51-4* or *RAD51-6* in the *RAD51-3-HA*^{flox} cells, or *RAD51-6* in the *RAD51-4-HA*^{flox} cells. PCR on G418-resistant clones for each of the three cell lines (Fig.S9) showed that it was possible to delete all copies of either *RAD51-4* or *RAD51-6*, retaining floxed copies of *RAD51-3-HA* or *RAD51-4-HA*. DiCre-mediated KO of the floxed gene was then induced by addition of rapamycin. PCR analysis (Fig.S10) showed complete excision of *RAD51-3-HA* or *RAD51-4-HA* after the second round of rapamycin-mediated DiCre induction in all three cell lines, thereby generating cells devoid of two *L. major* Rad51-paralogues genes simultaneously (*RAD51-4* and *RAD51-3*; *RAD51-6* and *RAD51-3*; or *RAD51-6* and *RAD51-4).* Complete loss of the genes was maintained after cells were grown for more than 15 passages (Fig.S10).

Western blotting (Fig.4B) showed that HA-tagged RAD51-3 or RAD51-4 protein in the null mutants was undetectable after rapamycin induction of gene KO. The *RAD51-4* and *RAD51-6* null mutants, prior to induction of conditional excision of the second gene, showed increased levels of γH2A compared with the background cell line (Fig.4C). DiCre-mediated gene excision, leading to the three different double gene KOs, did not result in further increases in the levels of the phosphorylated histone. The lack of further impairment when two genes are lost relative to one was broadly consistent with growth curves (Fig.S11A) and with flow cytometry analysis of DNA content (Fig.S11B): conditional excision of the second gene did not lead to worsening of growth or to increased numbers of aberrant cells with less than 1C content compared with uninduced single gene null cells. Intriguingly, the *RAD51-4* and *RAD51-6* null mutants, even without induction of the second gene KO and having been selected as clones, appeared to recapitulate the long-term growth impairment seen upon prolonged cultivation after DiCre excision of a single HR gene (compare Fig.S11A with Fig.1E).

IdU labelling followed by FACS analysis of each of the three cell lines (Fig.4D), before and after rapamycin induction, did not show a change in the proportion of cells that incoporated the nucleotide analogue. Quantification of IdU fluorescence levels of individual S-phase cells (Fig.4E) did not reveal a consistent pattern of fluorescence decrease upon simultaneous KO of two *RAD51* paralogs. However,a signal reduction was detected 72 h after *RAD51-3* KO induction in both the *RAD51-4* and *RAD51-6* null backgrounds, consistent with loss of RAD51-3 alone impeding DNA replication (Fig.2B). In contrast, though a mild loss of fluorescence was seen at 48 h when KO of *RAD51-4* was induced in the *RAD51-6* null background, this was not seen at 72 h (Fig.4E). These data reinforce the suggestion that RAD51-3 alone

among the three RAD51 paralogues plays a role in *L. major* DNA replication, and indicate that RAD51-4 and RAD51-6 do not obviously act redundantly in such a function.

To ask if the induction of a double *RAD51* paralogue gene KO had similar effects on genome instability relative to what was seen after induction of single gene KOs (Fig.3), we again performed Illumina sequencing of DNA from the *RAD51-4* null *RAD51-3-HA*^{flox} cells after two and six passages of growth with and without rapamycin. Read mapping to the genome showed the *RAD51-4* gene was absent and that DiCre induction lead to removal of *RAD51-3* (Fig.5A). SNPs accumulated to a greater extent across the genome (Fig.5B) and in each chromosome (Fig.5C) in the *RAD51-3 RAD51-4* double KO cells compared with the single null. However, both levels appeared less pronounced that in the single gene induced KOs (compare increase in SNPs between control and induced cells in Fig. 5B,C and Fig. 3A). Moreover, distinct from the single *RAD51-3* induced KO (Fig.3D), we did not detect pronounced changes in the pattern of SNP accumulation around SSRs upon induction of double a *RAD51-4 RAD51-3* KO compared with *RAD51-4* null cells (Fig.5D). Indeed, analysis of InDel accumulation provided no evidence for further increases in these types of mutation after *RAD51-3* induced KO relative to the uninduced *RAD51-4* null mutant cells (Fig.S12A-C). Thus, these data, allied to previous DNA synthesis and cell cycle-dependent damage analysis, suggest a complex mixture of common and separate functions for these two RAD51 paralogues, probably related to epistatic interactions among them.

Loss of RAD51 impairs initiation of DNA replication at the main origins in *Leishmania*

Our analyses until this point indicated impairment of DNA synthesis after loss of either RAD51 or RAD51-3, including evidence for distinct effects of the gene KOs. However, if and how the two factors might contribute to the programme of DNA replication in *L. major* was not known. To address this question, we performed MFA-seq analysis, comparing read depth across the chromosomes in DNA extracted from replicating and non-replicating cells, thereby identifying sites of replication initiation and patterns of fork progression (39, 54, 55). To do this, we used a modified version of MFA-seq compared to what we have described previously: rather than using FACS to isolate S- and G2-phase cells, we isolated DNA from *L. major* cells in logarithmic (log) or in stationary phase growth, sequenced each using Illumina technology, and mapped the ratio of reads in the two populations. Fig.6A shows these data as Z-scores across two chromosomes, where positive signal represents regions where the mean read depth in the log phase cells is greater than the mean of the stationary cells (Fig.S13 shows further chromosomes). This modified MFA-seq approach will be described in detail elsewhere (BioRXiv 10.1101/799429), but two primary features of *Leishmania* replication are revealed relative to MFA-seq based on S/G2 phase read depth ratios using cell cycle-sorted cells (39). First, we confirm the predominant use of single origins in each chromosome, most of which are centrally disposed in the molecules and each coinciding with SSRs. Second, we now detect weaker sites of replication initiation that were not seen previously and are proximal to one or both

telomeres in the chromosomes (see -RAP data in Fig.6A, which are highly comparable with WT cells; BioRXiv 10.1101/799429).

To ask about the effect of RAD51 or RAD51-3 loss on DNA replication, we performed MFA-seq in cells induced for KO by growth in rapamycin (48 h, second round of induction), as well as in control cells without rapamycin grown for the same time. Comparing induced *RAD51* KO cells with uninduced cells revealed a striking variation in MFA-seq pattern: loss of MFA signal was seen at the main origin that had previously been mapped within each chromosome (39), while an increase in MFA-seq signal was seen at the extremities of the chromosomes (Fig.6A, Fig.S12). In the induced *RAD51-3* KO cells, the same differential effect was not obvious: though there was potentially some loss of MFA-seq signal at the main origin, increase in subtelomeric signal was not apparent. To examine this genome-wide, we generated metaplots of the MFA-seq signal in the different cells (Fig.6B,C). Profiling of MFA signal around the main origin (Fig.6B) revealed considerable consistency in amplitude and width of the peaks, both for the 36 origins within one cell and between the two uninduced cells, confirming there is little variation in timing or efficiency of DNA synthesis initiation at all main origins in the WT *L. major* population (39). The same profiling confirmed that loss of MFA-seq signal around the main origins was a genome-wide effect upon *RAD51* KO, with lowered amplitude and width, and was less pronounced upon KO of *RAD51-3*. These data indicate that loss of RAD51 leads to a more pronounced decrease in replication initiation activity at the main origins compared with RAD51-3 loss, which appears consistent with decreased SNP accumulation at SSRs after *RAD51* KO but not after *RAD51-3* KO (Fig 3D). Profiling of the MFA-seq mapping at the chromosome ends (Fig.6C) showed that loss of RAD51, but not RAD51-3, resulted in a gain of signal in the subtelomeres, which again was strikingly consistent in amplitude and width across all chromosomes. These data indicate that loss of DNA synthesis at the main origins due to ablation of RAD51 is accompanied by increased DNA synthesis in the subtelomeres, a shift in the replication programme that is not clearly seen after loss of RAD51-3. In addition, these observations further confirm that RAD51 and RAD51-3 play distinct roles in *Leishmania* DNA replication.

Discussion

Homologous recombination is known to be important for episome formation in *Leishmania* (29-31), but the depth and breadth of how the process contributes to genome plasticity and transmission has not been fully explored. Here, we sought to answer two main questions. First, are RAD51 and RAD51 paralogues essential in *Leishmania*, given the conflicting data on the ability to generate and propagate null mutants in different species? Second, given the potentially novel distribution of mapped origins in the *Leishmania* genome, does HR play a central role in *Leishmania* genome duplication? Using a rapid, conditional knockout approach that combines CRISPR/Cas9 gene modification and DiCre gene excision, we show that loss of any RAD51-like

gene is not immediately lethal to promastigote *L. major* but is increasingly detrimental over time. Thus, we suggest that binary definitions of essential or non-essential for HR-related genes inadequately describe their contribution to parasite biology. In addition, our data reveal that loss of RAD51 and RAD51-3, uniquely among the four HR proteins examined, impairs DNA replication, though their roles are distinct, with RAD51 playing an unexpected and central role in DNA replication initiation.

Null mutants of *RAD51* and *RAD51-4* have been described in *L. infantum* (29, 30), whereas *RAD51* mutants could not be recovered by CRISPR/Cas9 gene targeting in *L. donovani* (48). In addition, *RAD51-3* and *RAD51-6* null mutants were not recovered by two-step homology-directed gene deletion in *L. infantum* (29). The conditional KO approach used here aids understanding of this complex biology, since we show that excision of any of these HR factors has no immediate impact of parasite fitness, but rather causes a progressive slowing of growth, presumably due to accrual of problems. Nonetheless, it is clear that *Leishmania* null mutants of *RAD51* (in *L. infantum*)(30), and the RAD51 paralogues *RAD51-4* and *RAD51-6* (this work, in *L. major*), can be generated. Similarly, null mutants of each *RAD51*-related gene have been described in *T. brucei* (44, 46). It is possible such variation in importance reflects species-specific aspects of HR function. However, it is also conceivable that conventional, two step transformation approaches to generate null mutant clones allows for selection of cells possessing compensatory changes that lead to survival - an adaptation that may not always be recovered when using CRISPR-Cas9 to simultaneously ablate both alleles (48), or that emerged in the timeframe we have used during conditional gene excision (this work). What such adaptations might be is unclear, but the abundant mutations we describe after loss of HR genes may provide a genetic basis for their generation. In addition, RAD51-directed HR is not the sole route for repair of DSBs in any trypanosomatid (56-62), though whether such pathways can increase in activity in the absence of RAD51-directed HR has not been tested. What aspects of *Leishmania* genome function degrade with time after loss of RAD51 and its relatives remain to be fully characterised, though our data suggest these effects may differ for the different RAD51 paralogues, and for RAD51.

Conditional gene KO loss of RAD51, RAD51-3 and RAD51-4 led to genome-wide increases in SNPs, as well as increased amounts of in InDels after RAD51-3 loss. Altogether, these data demonstrate a widespread role for these related proteins in *Leishmania* genome stability. Though we did not test for similar effects after *RAD51-6* KO, similar outcomes might be expected. In addition, we did not attempt to test for larger genome changes, but translocations have been described in MRE11 (27), which can guide RAD51 function. Our mapping of SNPs also revealed two unusual features of the patterns of mutagenesis in *L. major*. First, there was a notable chromosome size-dependence on levels of SNP accumulation, with smaller chromosomes tending to have higher SNP densities when compared with the larger chromosomes. This was common to each conditional gene KO, indicating it is general feature of *L. major* chromosome biology. However, the basis of this effect is unclear. Might it relate to differences in gene expression or nucleosome occupancy? No data that we are aware of supports such a suggestion, and the commonality of multigene

transcription in all chromosomes appears to argue against it (63). If not gene expression, then the chromosome size-dependence of SNP density might reflect the limitations of predominant DNA replication initiation at a single origin (39). The second feature was pronounced accumulation of SNPs around SSRs. Here, loss of the HR factors has somewhat different effects: for *RAD51* KO, a decrease in SSR-proximal SNP accumulation was seen, whereas an increase was seen upon *RAD51-3* KO and no change was found after KO of *RAD51-4*. These data suggest that the SNPs reflect the effects of differing roles of the proteins on damage repair. Given that both RAD51 and RAD51-3 loss affects DNA synthesis, while loss of RAD51-4 does not, an explanation could be that the SNPs are generated due to clashes between the transcription and replication machineries at SSRs, providing a source for such pronounced mutation rates at these sites. In *T. brucei* it is known that the Origin Recognition Complex (ORC) binds to potentially all SSRs and its loss affects levels of RNA at these loci (54). In addition, RNA-DNA hybrids, which have been mapped to sites of replication-transcription clashes in other eukaryotes (64), form prominently at transcription start sites in *T. brucei* (65). Thus, it seems conceivable that SSRs are also sites of such interaction in *Leishmania*, though no equivalent mapping of ORC or RNA-DNA hybrids has been reported.

Loss of IdU uptake after induced KO of *RAD51* or *RAD51-3* indicates a role for both HR factors in DNA synthesis and DNA replication, but several lines of evidence suggest these roles are not the same: distinct cell cycle timing of γH2A accumulation, distinct patterns of SNP accumulation around SSRs, and differing changes in DNA replication dynamics after their loss. Impaired nucleotide uptake after RAD51 loss appears to be explained by a shift in the programme of *L. major* DNA replication, with loss of efficient DNA replication initiation at the single primary origin in each chromosome and increased subtelomeric DNA replication initiation. The finding that RAD51 KO cells are impaired in growth and nucleotide uptake argues that increased replication from the subtelomeres is insufficient to compensate for loss of the primary initiation events. In the case of RAD51-3, the more modest loss of replication at the main origin, and no clear increase in subtelomeric replication, seems most readily explained by a widespread impairment in genome replication, perhaps because the RAD51 paralogue is needed to guide processes involved in promoting replication in the face of genome-wide impediments. In this regard, the replication phenotypes after loss of RAD51-3 may be comparable with effects seen after mutation of *T. brucei* MCM-BP (66), a poorly understood factor that modulates activity of the replicative MCM helicase (67, 68). In addition, such a role for *L. major* RAD51-3 may be akin to roles for mammalian Rad51 paralogues on stalled DNA replication forks, with the differing effects of loss of *L. major* RAD51-4 or RAD51-6 suggesting similar functional compartmentalisation amongst the paralogues (13).

How might RAD51 provide a central role in *Leishmania* DNA replication? One possibility is that DNA replication initiation at the single SSR-localised origin in each chromosome is directly driven by RAD51, perhaps even by catalysing HR. Such an origin-specific function for RAD51 or RAD51-directed HR has no precedent and, indeed, would be distinct from origin activity in *T. brucei*, where these sites are

conventionally defined by binding ORC (54, 69). Nonetheless, ORC binding has not been mapped in any *Leishmania* genome and precedents for recombination-directed replication initiation exist in viruses (70-72), bacteria (17), polypoid archaea (18) in *Tetrahymena* (73), albeit normally without a focus on defined genome sites. In addition, human Rad51 acts with MCM8-9, an alternate MCM helicase complex, to initiate DNA replication (74); though this DNA synthesis pathway is origin-independent in humans, no work has tested MCM8 or MCM9 function in any trypanosomatid. Alternatively, and perhaps more likely, RAD51 may play a more indirect role in *Leishmania* genome replication. The earliest acting origin-active SSRs in *T. brucei* co-localize with centromeres (54, 75), and recent work has mapped one putative component of the centromere-binding kinetochore, KKT1, to each MFAseq-mapped origin SSR in *L. major* (76). These loci might then be vulnerable to breakage, such as during chromosome segregation, and RAD51-directed HR may be required to repair any breaks, such as after mitosis in order to allow proper licensing and firing of origins at these loci. Such a scenario could be compatible with data from other eukaryotes of important roles for RAD51-directed HR is maintaining centromere function (77-79). In *T. brucei*, the presence of further ORC-defined origins in each chromosome may compensate for loss of RAD51 causing impaired centromere-focused origin function, but such a mutation in *Leishmania* might be more detrimental, given the presence of only a single major centromere-focused origin in each chromosome (39). A distinct suggestion for an indirect role of RAD51 is that the recombinase does not play an origin- or centromere-focused role, but instead is needed to support DNA replication genome-wide, given previous suggestions that replication from a single major origin in each *Leishmania* chromosome would be insufficient to replicate all chromosomes in S-phase (39, 41). Such a function could be compatible with origin-independent roles of HR-directed replication (discussed above) and, in the absence of RAD51, complete chromosome replication is lost during S-phase, leading to impaired mitosis and reduced numbers of cells that license the main origins. This scenario may also explain the increased levels of MFA-seq reads in the subtelomeres. As we did not detect this replication signature when previously sorting S-phase cells (39), it is likely that the subtelomeric DNA replication reaction occurs either late in S-phase, or even post-S-phase. Thus, the increased subtelomeric MFA-seq signal in *RAD51* KO cells may reflect greater numbers of cells stalling in late- or post-S phase. Alternatively, subtelomeric DNA synthesis may be a back-up strategy for the primary SSR/origin-focused DNA replication reaction and, when the primary reaction is compromised by loss of RAD51, it assumes greater importance. As it stands, we do not know the nature of the subtelomeric DNA synthesis, but these data indicate it is distinct from DNA replication emanating from the main SSR-focused origins and is RAD51-independent. Whether or not this subtelomeric DNA replication relates to a recently proposed form of telomere maintenance (34) is worthy of further investigation.

Methods

Parasite culture. Cell lines were derived from *Leishmania major* strain LT252 (MHOM/IR/1983/IR). Promastigotes were cultured at 26 °C in M199 or HOMEN medium supplemented with 10% heat-inactivated fetal bovine serum. Transfections were performed with exponentially growing cells with Amaxa Nucleofactor™ II, using the pre-set program X-001. Transfectants were selected by limiting dilution in 96-well plates in the presence of appropriate antibiotic. DiCre-expressing cells were selected with 10 µg.mL^{-1} blasticidin. DiCre/Cas9/T7-expressing cells were selected with 10 µg.mL^{-1} blasticidin and 20 µg.mL^{-1} hygromycin. Cells expressing the gene of interest (GOI) flanked by LoxP sites (GOI^{Flox} cells) were selected with 10 µg.mL^{-1} blasticidin, 20 µg.mL^{-1} hygromycin and µg.mL^{-1} puromycin.

DNA constructs and cell line generation. A background cell line was established in which DiCre is expressed from the ribosomal locus, while both Cas9 and T7 polymerase are expressed from the *tubulin* array. For this, WT cells were transfected with plasmid pGL2339 (80), previously digested with *Pac*I and *Pme*I, to generate DiCre-expressing cells. Correct integration into the ribosomal locus was confirmed by PCR. Then, the DiCre-expressing cells were transfected with plasmid pTB007 (81), previously digested with *Pac*I, to generate the DiCre/Cas9/T7-expressing cell line. Correct integration of Cas9/T7-encoding cassette was confirmed by PCR. In this way, the Cas9/T7 system, as previously described (81), was used to flank all copies of a GOI with LoxP sites, in a single round of transfection to generate the GOI^{Flox} cell lines used here. Deletion of the GOI was induced by rapamycin-mediated DiCre activation, as previously reported (82, 83). Donor fragments for Cas9-mediated genome editing were generated by PCR (Table S1). For this, the ORFs encoding each GOI were PCR-amplified using genomic DNA as template. PCR products of *RAD51* (LmjF.28.0550), *RAD51-4* (LmjF.11.0230), *RAD51-6* (LmjF.29.0450) and *PIF6* (LmjF.21.1190) were cloned between *Nde*I and *Spe*I restriction sites in the vector pGL2314 (83). PCR product of *RAD51-3* (LmjF.33.2490) was cloned into the *Spe*I restriction site of vector pGL2314. The resulting constructs contained the GOI flanked by LoxP sites (pGL2314*GOI*^{Flox}) and were used as templates in PCR reactions to generate the donor fragments flanked by sequences homologues (30 nucleotides) to the targeting integration sites. PCR products were ethanol precipitated and transfected into the DiCre/Cas9/T7-expressing cell line. Correct integration into the expected locus was confirmed by PCR analysis. The strategy used to generate sgRNAs was essentially as previously described(81). Briefly, sgRNAs were generated *in vivo* upon transfection with appropriate DNA fragment generated by PCR. These fragments contained the sequence for the T7 polymerase promoter, followed by the 20 nucleotides of sgRNA targeting site and 60 nucleotides of sgRNA scaffold sequence. The Eukaryotic Pathogen CRISPR guide RNA/DNA Design Tool (http://grna.ctegd.uga.edu) was used to generate the 20 nucleotide sequences for sgRNA targeting sites. The default parameters and the highest scoring 20 nucleotide sgRNA sequences were chosen.

Western blotting. Whole cell extracts were prepared by collecting cells and boiling them in NuPAGE™ LDS Sample Buffer (ThermoFisher). Extracts were resolved on 4-12 % gradient Bis-Tris Protein Gels (ThermoFisher) and then transferred to Polyvinylidene difluoride (PVDF) membranes (GE Life Sciences).

Before probing for specific proteins, membranes were blocked with 10% (w/v) non-fat dry milk in phosphate-buffered saline supplemented with 0.05% Tween-20 (PBS-T). Primary antibody incubation was performed for 2 h at room temperature with PBS-T supplemented with 5% non-fat dry milk. Membranes were washed with PBS-T and then incubated with secondary antibodies in the same conditions as the primary antibodies. For HRP-conjugated secondary antibodies, ECL Prime Western Blotting Detection Reagent (GE Life Sciences) was used for band detection as visualized with Hyperfilm ECL (GE Life Sciences). For IRDye-conjugates secondary antibodies, Odyssey Imaging Systems (Li-COR Biosciences) was used for band detection and visualization.

Antibodies. Generation of affinity-purified antibodies against γH2A (1:1000) from rabbit serum was previously described (84). Commercial primary antibodies used here were: mouse anti-HA (1: 5000, Sigma), anti-EF1α (1: 40 000, Merck Millipore) and anti-BrdU clone B44 (1: 500, BD Bioscience). Commercial secondary antibodies used here were: goat anti-Rabbit IgG HRP-conjugated (ThermoFisher), goat anti-Rabbit IgG HRP-conjugated (ThermoFisher), goat anti-Rabbit IgG Alexa Fluor 488-conjugated (ThermoFisher), goat anti-Rabbit IgG IRDye 800CW-conjugated (Li-COR Biosciences) and goat anti-Mouse IgG IRDye 680CW-conjugated (Li-COR Biosciences).

Genome sequencing and bioinformatics analysis. Total DNA extraction was performed with a DNeasy Blood & Tissue Kit (QIAGEN) following the manufacturer's instructions. Genomic DNA libraries were prepared using a Nextera™ NGS Library Preparation Kit (QIAGEN). Libraries were sequenced at Glasgow Polyomics (www.polyomics.gla.ac.uk/index.html), using a NextSeq™ 500 Illumina platform, generating paired end reads of 75 nucleotides. Processing of sequencing data was performed at the Galaxy web platform (*usegalaxy.org)(85).* FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimomatic (86) were used for quality control and adapter sequence removal, respectively. BWA-mem(87) was used to map processed reads to the reference genome (*Leishmania major Friedlin* v39, available at Tritrypdb - http://tritrypdb.org/tritrypdb/). Reads with mapping quality score < 30 were discarded using SAMtools (88). Single nucleotide polymorphisms (SNPs) and InDels were detected using GATK(89) and freebayes (90). Only those SNPs with at least 5 supporting reads and map quality 30 were considered. Heatmaps, violin plots and metaplots were generated using Prism Graphpad. Underlying data for metaplots and coverage tracks were generated using deepTools(91).

Marker Frequency Analysis (MFAseq). After processing, reads were compared essentially using methods described previously (39), though with modifications. Briefly, the number of reads in 0.5 kb windows along chromosomes was determined. The number of reads in each bin was then used to calculate the ratio between exponentially growing and stationary phase cells, scaled for the total size of the read library. Ratio values were converted into Z scores values in a 5 kb sliding window (steps of 500bp), for each individual chromosome. MFAseq profiles for each chromosome were represented in a graphical form using Gviz(92).

Detection of cells in S phase. Exponentially growing cells were incubated for 30 min with 150 µM IdU. Cells were collected by centrifugation and washed with 1x PBS. Fixation was performed at -20 ºC with a mixture

(7:3) of ethanol and 1x PBS for at least 16 h. Then, cells were collected by centrifugation and rinsed with washing buffer (1x PBS supplemented with 1% BSA). DNA denaturation was performed for 30 minutes with 2N HCL and reaction was neutralized with phosphate buffer (0.2 M Na_2HPO_4, 0.2 M KH_2PO_4, pH 7.4). Cells were collected by centrifugation, further incubated in phosphate buffer for 30 mins at room temperature and centrifuged again. To detect incorporated IdU, cells were incubated for 1h at room temperature with anti-BrdU antibody (diluted in washing buffer supplemented with 0.2% Tween-20), collected by centrifugation and washed with washing buffer. Cells were incubated for 1 h at room temperature with anti-mouse secondary antibody conjugated with Alexa Fluor 488 (diluted in washing buffer supplemented with 0.2% Tween-20), collected by centrifugation and washed with washing buffer. Finally, cells were re-suspended in 1xPBS supplemented with 10 $\mu g.mL^{-1}$ Propidium Iodide and 10 $\mu g.mL^{-1}$ RNAse A and passed through a 35 µm nylon mesh. Data was acquired with FACSCelesta (BD Biosciences) and further analysed with FlowJo software. Negative control cells, in which anti-BrdU antibody was omitted during IdU detection step, were included in each experiment. Negative control cells were used to draw gates to discriminate positive and negative events.

Figure legends

Figure 1. Combining CRISPR/Cas9 and DiCre allows rapid assessment of homologous recombination gene function by conditional excision. (A) A cell line was engineered to express both Cas9 and DiCre; i) Cas9 was used to rapidly replace all copies of a gene of interest (GOI) by a version of the same GOI flanked by LoxP sites (GOI^{Flox}); ii) KO induction was achieved by rapamycin-mediated activation of DiCre, which catalyses excision of GOI^{Flox}; please refer to Supplementary Figure S3A for the rapamycin induction strategy. **(B)** Western blotting analysis of whole cell extracts of the indicated cell lines after 48 h growth without addition (-RAP) or after addition (+RAP) of rapamycin, leading to DiCre induction; extracts were probed with anti-HA antiserum and anti-EF1α was used as loading control (predicted protein sizes are indicated, kDa). **(C)** RT-PCR analysis of cDNA from the *PIF6* ^{Flox} cell line after 72 h hours of growth with (+) or without (-) addition of RAP; R.T. (+) and R.T.(-) indicate addition or omission, respectively, of reverse transcriptase in the cDNA synthesis step; amplification of *RAD51-3* was used as an RT control and gDNA indicates a control PCR reaction using genomic DNA as template. **(D)** Western blotting analysis of whole cell extracts from the indicated cell lines after 48 h growth, –RAP and +RAP; extracts were probed with anti-γH2A antiserum and anti-EF1α was used as loading control (protein sizes are indicted, kDa). **(E)** Growth curves of the indicated cell lines in the presence (+, red) or absence (-, black) of RAP; cells were seeded at ~10^5 $cells.ml^{-1}$ at day 0 and diluted back to that density every 4 - 5 days to complete five passages (P1 to P5); growth profile was also evaluated after cells were kept in culture for more than 15 passages (>P15); cell density was assessed every 24 h, and error bars depict standard deviation from three replicate experiments.

Figure 2. Analysis of DNA synthesis and cell cycle-dependent accumulation of DNA damage upon induced knockout of homologous recombination factors. (A) Representative dot plots from flow cytometry analysis to detect DNA synthesis in the indicated cell lines; after 48 h growth without (-RAP) or with (+RAP) addition of rapamycin, inducing DiCre, cells were incubated with IdU for 30 min and IdU fluorescence was detected under denaturing conditions; 30,000 cells were analysed per sample; 1C and 2C indicate single DNA content (G1) and double DNA content (G2/M), respectively; dashed red lines indicate the threshold used to discriminate negative (black dots) from IdU-positive (blue dots) events; inset numbers indicate total percentage of IdU-positive events relative to the whole population. **(B)** Quantitative analysis of IdU fluorescence following gene excision in the indicated cell lines; after the indicated times of -RAP or +RAP growth, cells were labelled as in (A); fluorescence from IdU positive S phase cells is plotted as arbitrary units (a.u.); at least 2,000 cells were analysed in each time point and horizontal white lines indicate the mean; differences between -RAP or +RAP cells were tested with a Kruskal–Wallis test (**** denotes P< 0.0001). **(C)** Western blot analysis of whole cell extracts; 48 h after rapamycin DiCre induction (+RAP), or in controls cells with induction (-RAP), the indicated cell lines were left untreated (N.T.) or were treated with addition of 5 mM hydroxyurea (HU) for 8 hrs; cells were collected at the indicated time points after HU removal; extracts were probed with anti-γH2A antiserum and anti-EF1α was used as loading control (predicted protein sizes are shown, kDa).

Figure 3. Whole genome sequencing reveals mutagenesis upon induction of homologous recombination factor gene knockouts. (A) Sequence read depth around the targeted gene loci in the indicated cell lines; genomic DNA from rapamycin-induced (+RAP) and uninduced (-RAP) cells at the indicated passages (P) were prepared and subjected to deep sequencing (Illumina); coverage tracks were generated with deepTools, using the bamCoverage tool (91) and ignoring duplicated reads; RPKM normalization was used to allow comparison across samples. **(B)** Quantification of the number of new single nucleotide polymorphisms (SNPs) detected after 4 passages (between P6 and P2) in -RAP and +RAP cells; data are represented as violin plots, where the shape indicates data distribution and horizontal dotted white lines indicate median; differences were tested with a Mann-Whitney test (**** denotes P< 0.0001; ** denotes P = 0.0014). **(C)** Heatmaps showing density of new SNPs (SNPs/Kb) detected after 4 passages in each chromosome (01-36). **(D)** Metaplots of normalized density of new SNPs (SNPs/Kb) in the indicated cell line after 4 passages; data is shown for 60 kb of every chromosome centred around the MFAseq-mapped replication origins (SSR^{ORI}, n= 36), and for 60 kb around all other strand switch regions that did not show origin activity ($SSR^{non\text{-}ORI}$, n= 95).

Figure 4. Combining CRISPR/Cas9 and DiCre to generate double null mutant cells. (A) In the cell line expressing both Cas9 and DiCre, the following approach was used: i) Cas9 was used to replace all copies of the first GOI (GOI-1) by GOI-1^{Flox}; ii) Cas9 was used to replace all copies of the second GOI (GOI-2) by a neomycin resistance cassette (NEO), making this gene a null mutant (-/- panel B); iii) KO induction of GOI-

1^{Flox} is achieved by rapamycin-mediated activation of DiCre, generating a double null mutant cell line for both GOIs. **(B)** Western blotting analysis of whole cell extracts after 48 h growth without addition of rapamycin (-RAP) or after addition of rapamycin (+RAP), leading to DiCre induction; extracts were probed with anti-HA antiserum and anti-EF1α was used as loading control (predicted protein sizes are indicted, kDa). **(C)** Western blotting analysis of whole cell extracts from the indicated cell lines after 48 h of –RAP and +RAP growth; extracts were probed with anti-γH2A antiserum (green) and anti-EF1α (red) was used as loading control (protein sizes are indicted, kDa). **(D)** Representative dot plots from flow cytometry analysis to detect DNA synthesis in the indicated cell lines; after 48 h of –RAP +RAP growth, cells were incubated with IdU for 30 min and IdU fluorescence was detected under denaturing conditions; 30,000 cells were analysed per sample; 1C and 2C indicate single DNA content (G1) and double DNA content (G2/M), respectively; dashed red lines indicate the threshold used to discriminate negative (black dots) from IdU-positive (blue dots) events, and inset numbers indicate total percentage of IdU-positive events relative to the whole population **(E)** Quantitative analysis of IdU fluorescence following gene excision in the indicated cell lines; after the indicated times of -RAP or +RAP growth, cells were labelled as in (D); fluorescence from IdU positive cells is plotted as arbitrary units (a.u.); at least 2,000 cells were analysed in each time point; horizontal white lines indicate the mean; differences were tested with a Kruskal–Wallis test (**** denotes $P< 0.0001$).

Figure 5. Mutagenesis upon induction of single or double mutants of two RAD51 paralogues. (A) Sequence read depth around the targeted gene loci in a cell line that is null mutant (-/-) for *RAD51-4* and contains a variant of *RAD51-3* translationally fused to HA epitope sequence and surrounded by loxP sites (RAD51_3-HA^{flox}); genomic DNA from rapamycin-induced (+RAP) and uninduced (-RAP) cells at the indicated passages (P) were was subjected to deep sequencing (Illumina); coverage tracks were generated with deepTools, using the bamCoverage tool (48) and ignoring duplicated reads; RPKM normalization was used to allow comparison across samples. **(B)** Quantification of the number of new single nucleotide polymorphisms (SNPs) detected after 4 passages in -RAP and +RAP cells; data are represented as violin plots, where the shape indicates data distribution and horizontal dotted white lines indicate median; differences were tested with a Kruskal–Wallis test (*** denotes $P< 0.0001$). **(C)** Heatmaps showing density of new SNPs (SNPs/Kb) detected after 4 passages in each chromosome (01-36). **(D)** Metaplots of normalized density of new SNPs (SNPs/Kb) after 4 passages; data is shown for 60 kb around MFAseq-mapped replication origins (SSR^{ORI}, n= 36), and all other strand switch regions that do not show origin activity ($SSR^{non\text{-}ORI}$, n= 95).

Figure 6. Genome-wide mapping reveals impaired initiation of DNA replication upon induced knockout of RAD51. (A) Graphs show the distribution of sites of DNA synthesis initiation across two complete chromosomes in the indicated cell lines, in each case grown in the absence (-RAP) or the presence (+RAP) of rapamycin; MFA-seq signal is represented by Z-scores across the chromosomes, calculated by comparing

read depth coverage of DNA from exponentially growing cells relative to stationary cells; the bottom track for each chromosome displays coding sequences, with genes transcribed from right to left in red, and from left to right in blue. **(B)** Metaplots of MFA-seq signal found in every chromosome, centred on the previously mapped constitutive DNA replication origin (SSR^{ORI}) ±0.15 Mb, in -RAP and +RAP cells. **(C)** Metaplots of MFA-seq signal across 0.15 Mb of sequence from all chromosomes ends in -RAP and +RAP cells. In B and C, *p* values were determined using Wilcoxon test.

Acknowledgements. We thank all current and previous members of the McCulloch and Tosi labs for input, and S.Duncan and J.Mottram for discussions regarding CRISPR/Cas9 and DiCre. This work was supported by the BBSRC [BB/N016165/1, BB/R017166/1] and by the EU [Marie Sklodowska-Curie Individual Fellowship, RECREPEMLE]. The Wellcome Centre for Integrative Parasitology is supported by core funding from the Wellcome Trust [104111].

Competing interests. The authors declare no competing interests

References

1. Scully R, Panday A, Elango R, & Willis NA (2019) DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature reviews. Molecular cell biology*.
2. Bell JC & Kowalczykowski SC (2016) RecA: Regulation and Mechanism of a Molecular Search Engine. *Trends in biochemical sciences* 41(6):491-507.
3. Sullivan MR & Bernstein KA (2018) RAD-ical New Insights into RAD51 Regulation. *Genes (Basel)* 9(12).
4. Zhang S*, et al.* (2017) Structural basis for the functional role of the Shu complex in homologous recombination. *Nucleic Acids Res* 45(22):13068-13079.
5. Gaines WA*, et al.* (2015) Promotion of presynaptic filament assembly by the ensemble of S. cerevisiae Rad51 paralogues with Rad52. *Nature communications* 6:7834.
6. Taylor MRG*, et al.* (2015) Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination. *Cell* 162(2):271-286.
7. Rosenbaum JC*, et al.* (2019) The Rad51 paralogs facilitate a novel DNA strand specific damage tolerance pathway. *Nature communications* 10(1):3515.
8. Saxena S, Somyajit K, & Nagaraju G (2018) XRCC2 Regulates Replication Fork Progression during dNTP Alterations. *Cell reports* 25(12):3273-3282 e3276.
9. Pond KW, de Renty C, Yagle MK, & Ellis NA (2019) Rescue of collapsed replication forks is dependent on NSMCE2 to prevent mitotic DNA damage. *PLoS genetics* 15(2):e1007942.
10. Malacaria E*, et al.* (2019) Rad52 prevents excessive replication fork reversal and protects from nascent strand degradation. *Nature communications* 10(1):1412.
11. Bhat KP*, et al.* (2018) RADX Modulates RAD51 Activity to Control Replication Fork Protection. *Cell reports* 24(3):538-545.
12. Petermann E, Orta ML, Issaeva N, Schultz N, & Helleday T (2010) Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. *Mol.Cell* 37(4):492-502.
13. Somyajit K, Saxena S, Babu S, Mishra A, & Nagaraju G (2015) Mammalian RAD51 paralogs protect nascent DNA at stalled forks and mediate replication restart. *Nucleic Acids Res* 43(20):9835-9855.

14. Kogoma T & von Meyenburg K (1983) The origin of replication, oriC, and the dnaA protein are dispensable in stable DNA replication (sdrA) mutants of Escherichia coli K-12. *The EMBO journal* 2(3):463-468.
15. Asai T, Sommer S, Bailone A, & Kogoma T (1993) Homologous recombination-dependent initiation of DNA replication from DNA damage-inducible origins in Escherichia coli. *The EMBO journal* 12(8):3287-3295.
16. Asai T & Kogoma T (1994) D-loops and R-loops: alternative mechanisms for the initiation of chromosome replication in Escherichia coli. *Journal of bacteriology* 176(7):1807-1812.
17. Kogoma T (1997) Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiology and molecular biology reviews : MMBR* 61(2):212-238.
18. Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, & Allers T (2013) Accelerated growth in the absence of DNA replication origins. *Nature* 503(7477):544-547.
19. Piazza A & Heyer WD (2019) Homologous Recombination and the Formation of Complex Genomic Rearrangements. *Trends in cell biology* 29(2):135-149.
20. Lee CS & Haber JE (2015) Mating-type Gene Switching in Saccharomyces cerevisiae. *Microbiology spectrum* 3(2):MDNA3-0013-2014.
21. McCulloch R, Morrison LJ, & Hall JP (2015) DNA Recombination Strategies During Antigenic Variation in the African Trypanosome. *Microbiology spectrum* 3(2):MDNA3-0016-2014.
22. Trenaman A*, et al.* (2013) Trypanosoma brucei BRCA2 acts in a life cycle-specific genome stability process and dictates BRC repeat number-dependent RAD51 subnuclear dynamics. *Nucleic Acids Res* 41(2):943-960.
23. Hartley CL & McCulloch R (2008) Trypanosoma brucei BRCA2 acts in antigenic variation and has undergone a recent expansion in BRC repeat number that is important during homologous recombination. *Mol.Microbiol.* 68(5):1237-1251.
24. Moraes Barros RR*, et al.* (2012) Anatomy and evolution of telomeric and subtelomeric regions in the human protozoan parasite Trypanosoma cruzi. *BMC Genomics* 13:229.
25. Weatherly DB, Peng D, & Tarleton RL (2016) Recombination-driven generation of the largest pathogen repository of antigen variants in the protozoan Trypanosoma cruzi. *BMC Genomics* 17(1):729.
26. Alves CL*, et al.* (2018) The recombinase Rad51 plays a key role in events of genetic exchange in Trypanosoma cruzi. *Scientific reports* 8(1):13335.
27. Laffitte MC*, et al.* (2016) Chromosomal Translocations in the Parasite Leishmania by a MRE11/RAD50-Independent Microhomology-Mediated End Joining Mechanism. *PLoS genetics* 12(6):e1006117.
28. Laffitte MN, Leprohon P, Papadopoulou B, & Ouellette M (2016) Plasticity of the Leishmania genome leading to gene copy number variations and drug resistance. *F1000Res* 5:2350.
29. Genois MM*, et al.* (2015) Roles of Rad51 paralogs for promoting homologous recombination in Leishmania infantum. *Nucleic Acids Res* 43(5):2701-2715.
30. Ubeda JM*, et al.* (2014) Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite Leishmania. *PLoS biology* 12(5):e1001868.
31. Laffitte MC*, et al.* (2014) Formation of linear amplicons with inverted duplications in Leishmania requires the MRE11 nuclease. *PLoS genetics* 10(12):e1004805.
32. Genois MM*, et al.* (2014) DNA repair pathways in trypanosomatids: from DNA repair to drug resistance. *Microbiology and molecular biology reviews : MMBR* 78(1):40-73.
33. Lachaud L*, et al.* (2014) Constitutive mosaic aneuploidy is a unique genetic feature widespread in the Leishmania genus. *Microbes and infection / Institut Pasteur* 16(1):61-66.
34. Bussotti G*, et al.* (2018) Leishmania Genome Dynamics during Environmental Adaptation Reveal Strain-Specific Differences in Gene Copy Number Variation, Karyotype Instability, and Telomeric Amplification. *mBio* 9(6).
35. Dumetz F*, et al.* (2017) Modulation of Aneuploidy in Leishmania donovani during Adaptation to Different In Vitro and In Vivo Environments and Its Impact on Gene Expression. *mBio* 8(3).

36. Prieto Barja P*, et al.* (2017) Haplotype selection as an adaptive mechanism in the protozoan pathogen Leishmania donovani. *Nat Ecol Evol* 1(12):1961-1969.
37. Zackay A*, et al.* (2018) Genome wide comparison of Ethiopian Leishmania donovani strains reveals differences potentially related to parasite survival. *PLoS genetics* 14(1):e1007133.
38. Ubeda JM*, et al.* (2008) Modulation of gene expression in drug resistant Leishmania is associated with gene amplification, gene deletion and chromosome aneuploidy. *Genome biology* 9(7):R115.
39. Marques CA, Dickens NJ, Paape D, Campbell SJ, & McCulloch R (2015) Genome-wide mapping reveals single-origin chromosome replication in Leishmania, a eukaryotic microbe. *Genome biology* 16(1):230.
40. Lombrana R*, et al.* (2016) Transcriptionally Driven DNA Replication Program of the Human Parasite Leishmania major. *Cell reports* 16(6):1774-1786.
41. Marques CA & McCulloch R (2018) Conservation and Variation in Strategies for DNA Replication of Kinetoplastid Nuclear Genomes. *Current genomics* 19(2):98-109.
42. McKean PG, Keen JK, Smith DF, & Benson FE (2001) Identification and characterisation of a RAD51 gene from Leishmania major. *Mol.Biochem.Parasitol.* 115(2):209-216.
43. Genois MM*, et al.* (2012) Interactions between BRCA2 and RAD51 for promoting homologous recombination in Leishmania infantum. *Nucleic Acids Res* 40(14):6570-6584.
44. Proudfoot C & McCulloch R (2005) Distinct roles for two RAD51-related genes in Trypanosoma brucei antigenic variation. *Nucleic Acids Res.* 33(21):6906-6919.
45. Proudfoot C & McCulloch R (2006) Trypanosoma brucei DMC1 does not act in DNA recombination, repair or antigenic variation in bloodstream stage cells. *Mol.Biochem.Parasitol.* 145(2):245-253.
46. Dobson R, Stockdale C, Lapsley C, Wilkes J, & McCulloch R (2011) Interactions among Trypanosoma brucei RAD51 paralogues in DNA repair and antigenic variation. *Mol.Microbiol.* 81(2):434-456.
47. Peacock L*, et al.* (2011) Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly. *Proceedings of the National Academy of Sciences of the United States of America* 108(9):3671-3676.
48. Zhang WW, Lypaczewski P, & Matlashewski G (2017) Optimized CRISPR-Cas9 Genome Editing for Leishmania and Its Use To Target a Multigene Family, Induce Chromosomal Translocation, and Study DNA Break Repair Mechanisms. *mSphere* 2(1).
49. McCulloch R & Barry JD (1999) A role for RAD51 and homologous recombination in Trypanosoma brucei antigenic variation. *Genes & development* 13(21):2875-2888.
50. Liu B*, et al.* (2009) Trypanosomes have six mitochondrial DNA helicases with one controlling kinetoplast maxicircle replication. *Mol.Cell* 35(4):490-501.
51. Byrd AK & Raney KD (2017) Structure and function of Pif1 helicase. *Biochemical Society transactions* 45(5):1159-1171.
52. Deegan TD, Baxter J, Ortiz Bazan MA, Yeeles JTP, & Labib KPM (2019) Pif1-Family Helicases Support Fork Convergence during DNA Replication Termination in Eukaryotes. *Molecular cell* 74(2):231-244 e239.
53. Glover L & Horn D (2012) Trypanosomal histone gammaH2A and the DNA damage response. *Mol.Biochem.Parasitol.* 183(1):78-83.
54. Tiengwe C*, et al.* (2012) Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of Trypanosoma brucei. *Cell reports* 2(1):185-197.
55. Muller CA & Nieduszynski CA (2012) Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome research* 22(10):1953-1962.
56. Conway C, Proudfoot C, Burton P, Barry JD, & McCulloch R (2002) Two pathways of homologous recombination in Trypanosoma brucei. *Molecular microbiology* 45(6):1687-1700.
57. Glover L, Jun J, & Horn D (2011) Microhomology-mediated deletion and gene conversion in African trypanosomes. *Nucleic Acids Res.* 39(4):1372-1380.
58. Glover L, McCulloch R, & Horn D (2008) Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. *Nucleic Acids Res* 36(8):2608-2618.

59. Zhang WW & Matlashewski G (2015) CRISPR-Cas9-Mediated Genome Editing in Leishmania donovani. *mBio* 6(4):e00861.
60. Sollelis L*, et al.* (2015) First efficient CRISPR-Cas9-mediated genome editing in Leishmania parasites. *Cellular microbiology* 17(10):1405-1412.
61. Lander N, Li ZH, Niyogi S, & Docampo R (2015) CRISPR/Cas9-Induced Disruption of Paraflagellar Rod Protein 1 and 2 Genes in Trypanosoma cruzi Reveals Their Role in Flagellar Attachment. *mBio* 6(4):e01012.
62. Peng D, Kurup SP, Yao PY, Minning TA, & Tarleton RL (2015) CRISPR-Cas9-mediated single-gene and gene family disruption in Trypanosoma cruzi. *mBio* 6(1):e02097-02014.
63. Iantorno SA*, et al.* (2017) Gene Expression in Leishmania Is Regulated Predominantly by Gene Dosage. *mBio* 8(5).
64. Costantino L & Koshland D (2018) Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. *Molecular cell* 71(4):487-497 e483.
65. Briggs E, Hamilton G, Crouch K, Lapsley C, & McCulloch R (2018) Genome-wide mapping reveals conserved and diverged R-loop activities in the unusual genetic landscape of the African trypanosome genome. *Nucleic Acids Res* 46(22):11789-11805.
66. Kim HS (2019) Genome-wide function of MCM-BP in Trypanosoma brucei DNA replication and transcription. *Nucleic Acids Res* 47(2):634-647.
67. Santosa V, Martha S, Hirose N, & Tanaka K (2013) The fission yeast minichromosome maintenance (MCM)-binding protein (MCM-BP), Mcb1, regulates MCM function during prereplicative complex formation in DNA replication. *The Journal of biological chemistry* 288(10):6864-6880.
68. Nishiyama A, Frappier L, & Mechali M (2011) MCM-BP regulates unloading of the MCM2-7 helicase in late S phase. *Genes & development* 25(2):165-175.
69. Marques CA*, et al.* (2016) Diverged composition and regulation of the Trypanosoma brucei origin recognition complex that mediates DNA replication initiation. *Nucleic Acids Res* 44(10):4763-4784.
70. Mosig G (1998) Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annual review of genetics* 32:379-413.
71. Mosig G, Gewin J, Luder A, Colowick N, & Vo D (2001) Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *Proceedings of the National Academy of Sciences of the United States of America* 98(15):8306-8311.
72. Wilkinson DE & Weller SK (2003) The role of DNA recombination in herpes simplex virus DNA replication. *IUBMB Life* 55(8):451-458.
73. Lee PH, Meng X, & Kapler GM (2015) Developmental Regulation of the Tetrahymena thermophila Origin Recognition Complex. *PLoS genetics* 11(1):e1004875.
74. Natsume T*, et al.* (2017) Acute inactivation of the replicative helicase in human cells triggers MCM8-9-dependent DNA synthesis. *Genes & development* 31(8):816-829.
75. Akiyoshi B & Gull K (2014) Discovery of unconventional kinetochores in kinetoplastids. *Cell* 156(6):1247-1258.
76. Garcia-Silva MR*, et al.* (2017) Identification of the centromeres of Leishmania major: revealing the hidden pieces. *EMBO reports* 18(11):1968-1977.
77. Forsburg SL & Shen KF (2017) Centromere Stability: The Replication Connection. *Genes (Basel)* 8(1).
78. Onaka AT*, et al.* (2016) Rad51 and Rad54 promote noncrossover recombination between centromere repeats on the same chromatid to prevent isochromosome formation. *Nucleic Acids Res* 44(22):10744-10757.
79. McFarlane RJ & Humphrey TC (2010) A role for recombination in centromere function. *Trends in genetics : TIG* 26(5):209-213.
80. Santos R*, et al.* (2017) A DiCre recombinase-based system for inducible expression in Leishmania major. *Molecular and biochemical parasitology* 216:45-48.
81. Beneke T*, et al.* (2017) A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *R Soc Open Sci* 4(5):170095.

82. Damasceno JD*, et al.* (2018) Conditional genome engineering reveals canonical and divergent roles for the Hus1 component of the 9-1-1 complex in the maintenance of the plastic genome of Leishmania. *Nucleic Acids Res* 46(22):11835-11846.
83. Duncan SM*, et al.* (2016) Conditional gene deletion with DiCre demonstrates an essential role for CRK3 in Leishmania mexicana cell cycle regulation. *Molecular microbiology* 100(6):931-944.
84. Stortz JA*, et al.* (2017) Genome-wide and protein kinase-focused RNAi screens reveal conserved and novel damage response pathways in Trypanosoma brucei. *PLoS pathogens* 13(7):e1006477.
85. Afgan E*, et al.* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46(W1):W537-W544.
86. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
87. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
88. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
89. McKenna A*, et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297-1303.
90. Garrison E & Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv e-prints.*
91. Ramirez F*, et al.* (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44(W1):W160-165.
92. Hahne F & Ivanek R (2016) Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol* 1418:335-351.

Figure 1

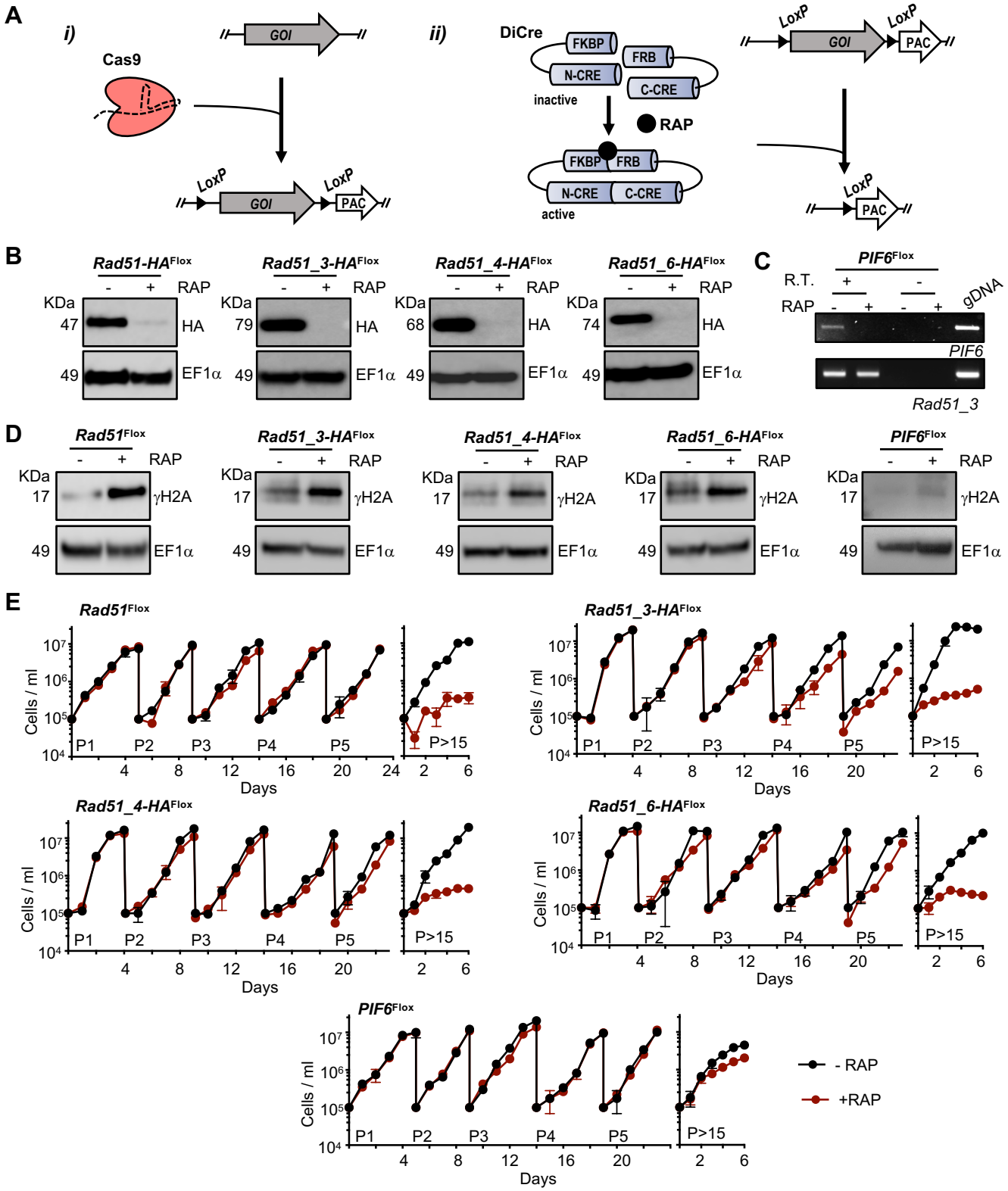
A
i)
GOI
Cas9
LoxP
GOI
PAC
ii)
DiCre
FKBP
FRB
N-CRE
C-CRE
inactive
RAP
active
B
Rad51-HAFlox
Rad51_3-HAFlox
Rad51_4-HAFlox
Rad51_6-HAFlox
KDa
HA
EF1α
C
PIF6Flox
R.T.
RAP
gDNA
PIF6
Rad51_3
D
Rad51Flox
γH2A
E
Cells / ml
Days
P1
P2
P3
P4
P5
P>15
- RAP
+RAP

Figure 2

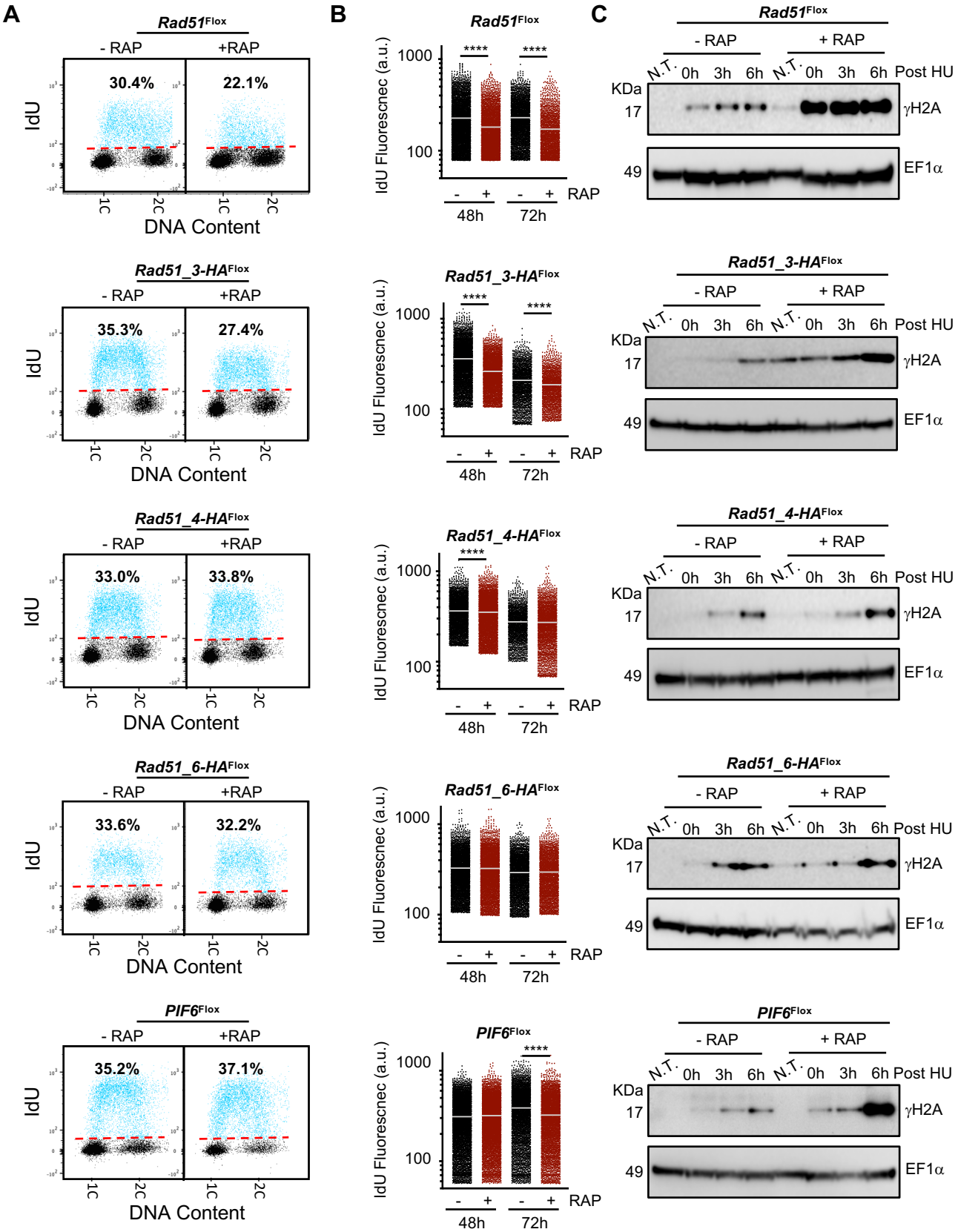
A
B
C
Rad51Flox
- RAP
+RAP
30.4%
22.1%
IdU
1C
2C
DNA Content
IdU Fluorescnec (a.u.)
1000
100
****
- + - + RAP
48h
72h
N.T.
0h
3h
6h
+ RAP
Post HU
KDa
17
γH2A
49
EF1α
Rad51_3-HAFlox
35.3%
27.4%
Rad51_4-HAFlox
33.0%
33.8%
Rad51_6-HAFlox
33.6%
32.2%
PIF6Flox
35.2%
37.1%

Figure 3

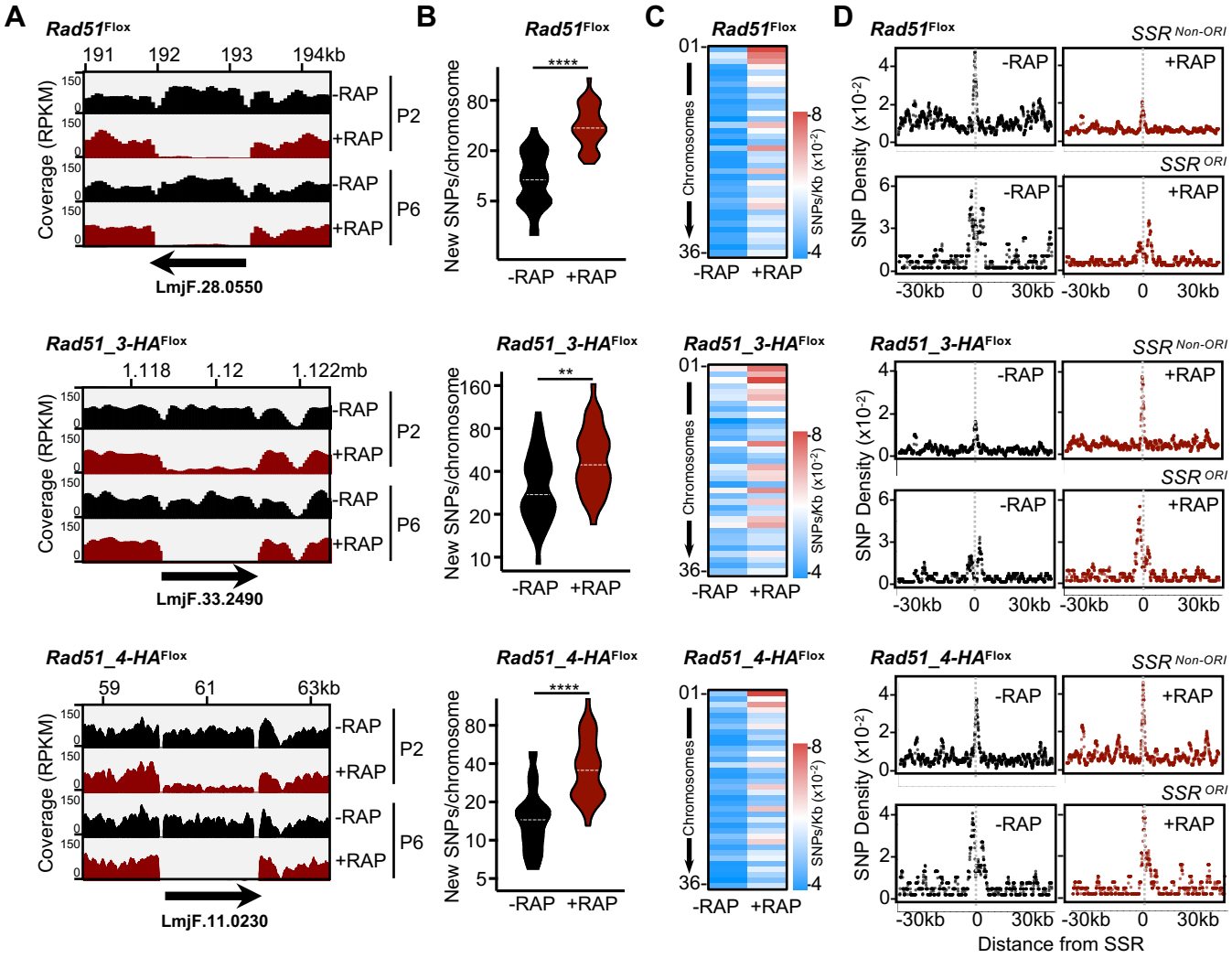
A
Rad51Flox
191
192
193
194kb
Coverage (RPKM)
-RAP
+RAP
P2
P6
LmjF.28.0550
Rad51_3-HAFlox
1.118
1.12
1.122mb
LmjF.33.2490
Rad51_4-HAFlox
59
61
63kb
LmjF.11.0230
B
New SNPs/chromosome
****
**
-RAP
+RAP
C
Chromosomes
01
36
SNPs/Kb (x10-2)
D
SSR Non-ORI
SSR ORI
SNP Density (x10-2)
-30kb
0
30kb
Distance from SSR

Figure 4

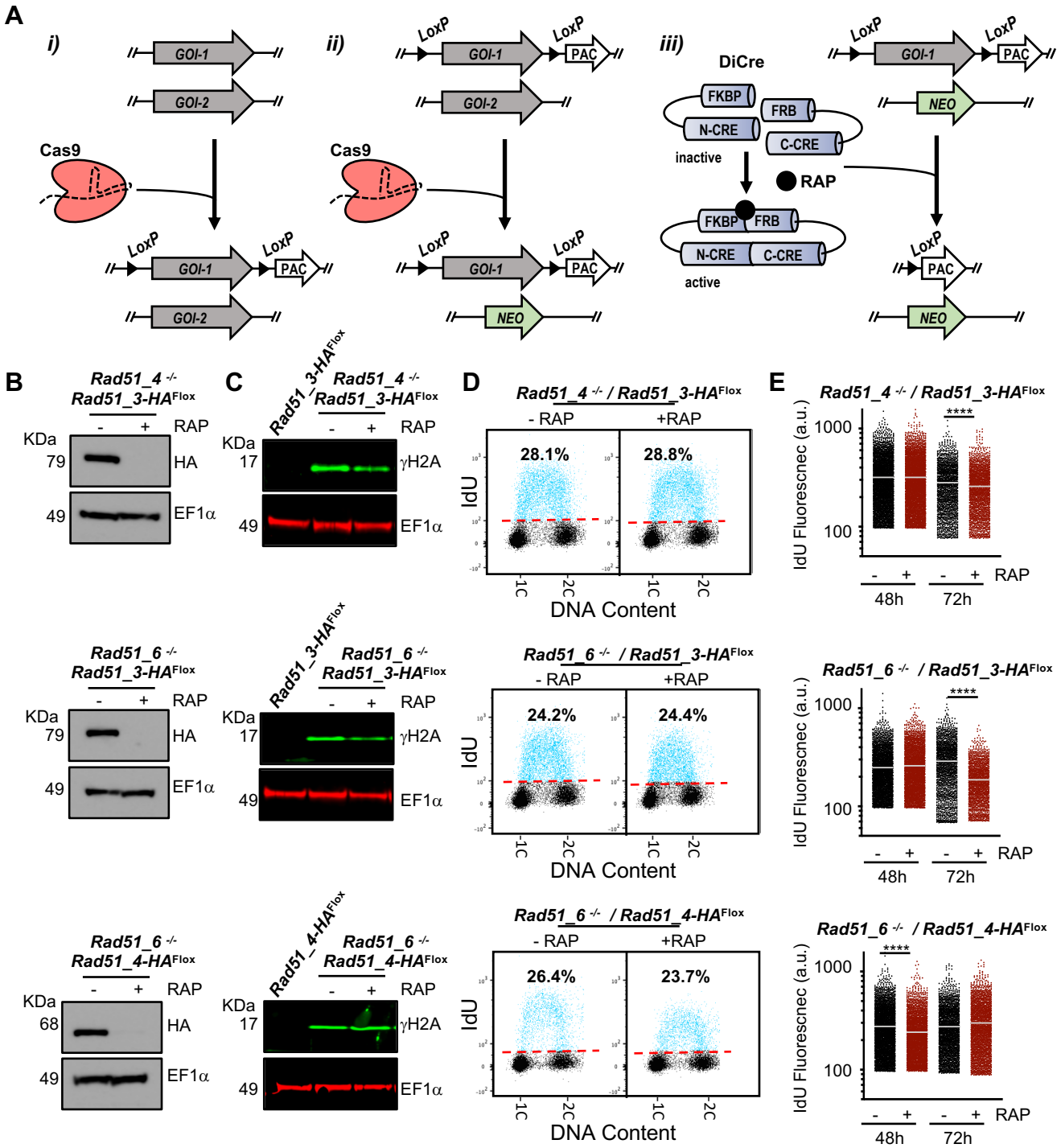
A
i)
GOI-1
GOI-2
Cas9
LoxP
PAC
ii)
iii)
DiCre
FKBP
FRB
N-CRE
C-CRE
inactive
RAP
active
NEO
B
Rad51_4 -/-
Rad51_3-HAFlox
- + RAP
KDa
79
HA
49
EF1α
C
Rad51_3-HAFlox
Rad51_4 -/-
Rad51_3-HAFlox
- + RAP
KDa
17
γH2A
49
EF1α
D
Rad51_4 -/- / Rad51_3-HAFlox
- RAP
+RAP
28.1%
28.8%
IdU
1C
2C
DNA Content
E
Rad51_4 -/- / Rad51_3-HAFlox
IdU Fluorescnec (a.u.)
1000
100
****
- + - + RAP
48h
72h
Rad51_6 -/-
Rad51_3-HAFlox
Rad51_6 -/- / Rad51_3-HAFlox
24.2%
24.4%
Rad51_6 -/-
Rad51_4-HAFlox
Rad51_4-HAFlox
Rad51_6 -/- / Rad51_4-HAFlox
68
26.4%
23.7%

Figure 5

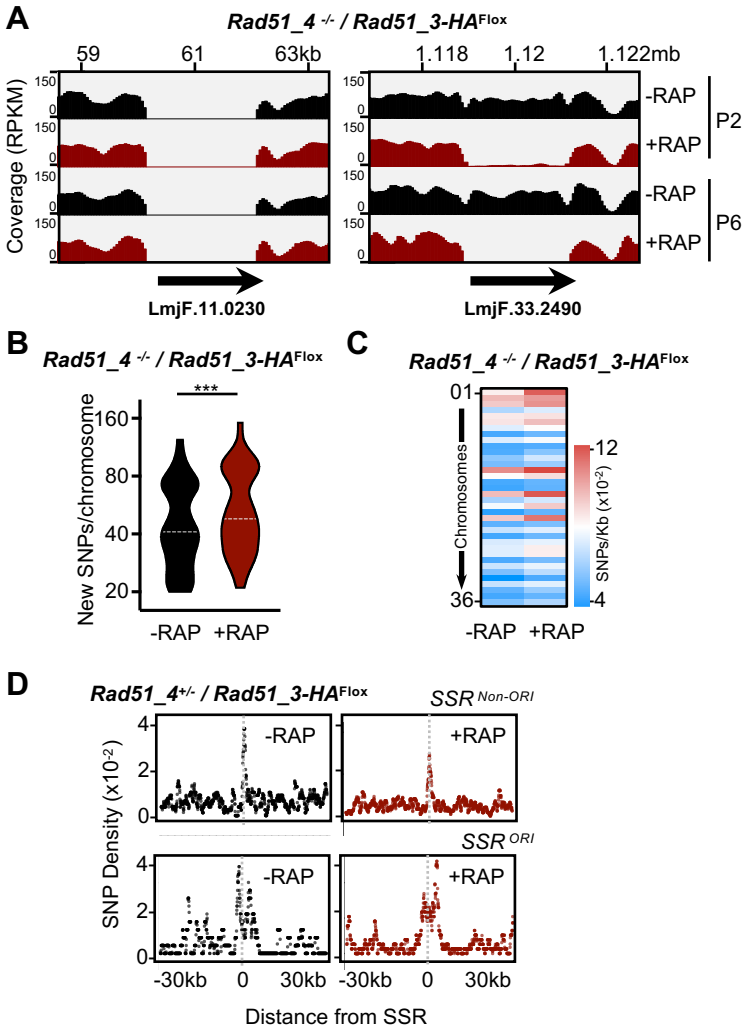
A
Rad51_4 -/- / Rad51_3-HAFlox
59
61
63kb
1.118
1.12
1.122mb
Coverage (RPKM)
150
0
-RAP
+RAP
P2
P6
LmjF.11.0230
LmjF.33.2490
B
Rad51_4 -/- / Rad51_3-HAFlox
***
New SNPs/chromosome
160
80
40
20
-RAP
+RAP
C
Rad51_4 -/- / Rad51_3-HAFlox
01
36
Chromosomes
12
4
SNPs/Kb (x10-2)
-RAP
+RAP
D
Rad51_4+/- / Rad51_3-HAFlox
SSR Non-ORI
SSR ORI
SNP Density (x10-2)
4
2
0
-RAP
+RAP
-30kb
0
30kb
Distance from SSR

Figure 6

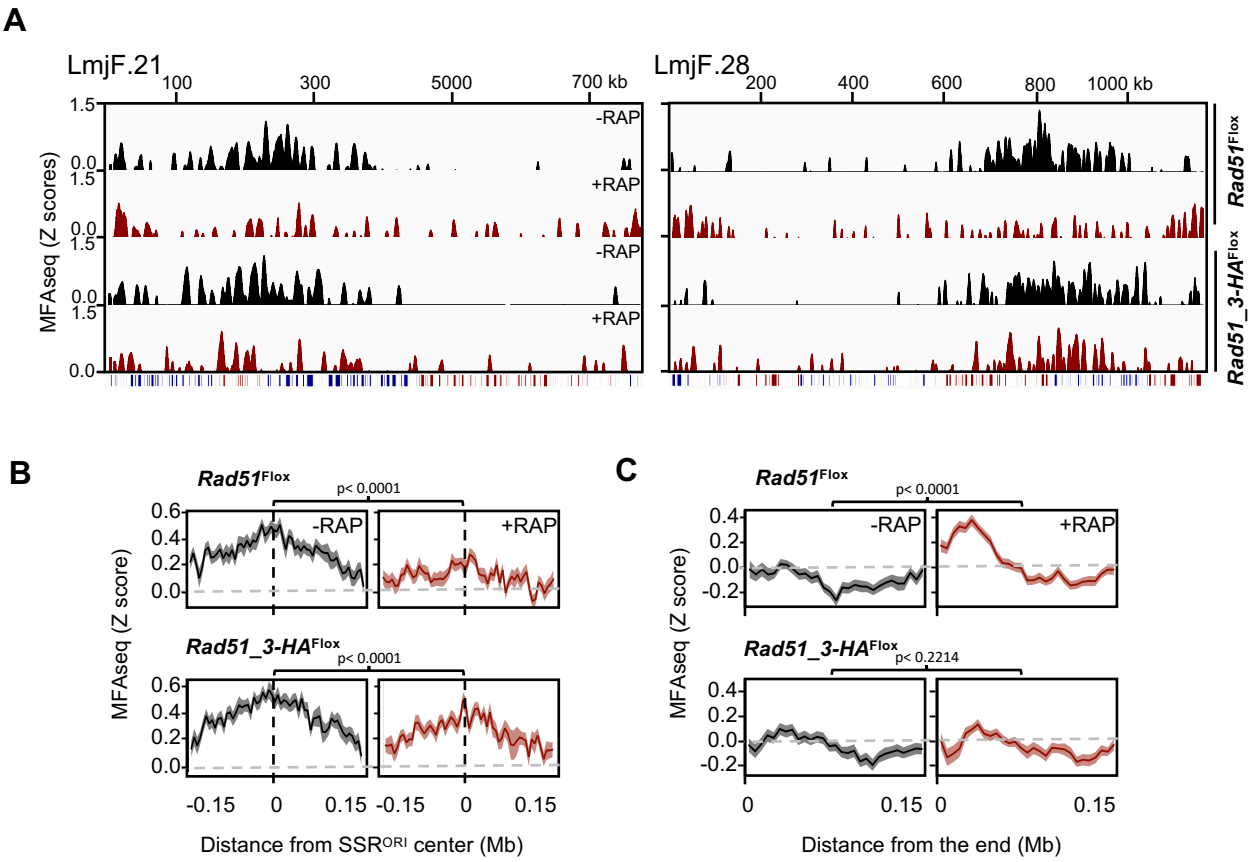
A
LmjF.21
100
300
5000
700 kb
LmjF.28
200
400
600
800
1000 kb
MFAseq (Z scores)
-RAP
+RAP
-RAP
+RAP
Rad51Flox
Rad51_3-HAFlox
B
Rad51Flox
p< 0.0001
-RAP
+RAP
Rad51_3-HAFlox
p< 0.0001
MFAseq (Z score)
-0.15
0
0.15
Distance from SSRORI center (Mb)
C
Rad51Flox
p< 0.0001
-RAP
+RAP
Rad51_3-HAFlox
p< 0.2214
MFAseq (Z score)
0
0.15
Distance from the end (Mb)

Table S1

Primers used in this study

Name	Orientation	Sequence (5'- 3')	Target	Purpose
a-1	Forward	GCGTTGGTGATGCTCGGGAT	Rad51 ORF	Detect integration of GOI^{Flox}
a-2		GTGGCAGAGAACCGCACAAA	Rad51_3 ORF	
a-3		TGCCATCCTTCTTGTATCTA	Rad51_4 ORF	
a-4		AGCACGGCATCGAGGATGCA	Rad51_6 ORF	
a-5		GGGCGAAAACAGAGGATGGG	PIF6 ORF	
b-1	Reverse	GAAAGAAAAGGCGTGAAGC	Rad51 3'UTR	
b-2		TGACGTAGCAACTCGGCG	Rad51_3 3'UTR	
b-3		GCACGACATCATGCTCG	Rad51_4 3'UTR	
b-4		GTGGAGAGGAGGGCAGAG	Rad51_6 3'UTR	
b-5		CGCGTATATATGGCGAGG	PIF6 3'UTR	
c	Forward	CGTTGTAAAACGACGGCCAG	Upstream of *LoxP*	Detect excision of GOI^{Flox}
d	Reverse	CGTGGGCTTGTACTCGGT	PAC	
Rad51	Forward	TAC **CAT ATG**CAGACCCGTTCCAAGGCCAA	Rad51 ORF	PCR amplification of GOI's ORFs
	Reverse	TAC **ACTAGT** ATCCCGAGCATCACCAACGC		
Rad51_3	Forward	TAC **ACTAGT** ATGGCACTGATCGAGTGCGC	Rad51_3 ORF	
	Reverse	TAC **ACTAGT** TTTGTGCGGTTCTCTGCCAC		
Rad51_4	Forward	TAC **CAT ATG**AACAAGTCGCAGTTCTACGC	Rad51_4 ORF	
	Reverse	TAC **ACTAGT** TAGATACAAGAAGGATGGCA		
Rad51_6	Forward	TAC **CAT ATG**AGCGCCATAGTGCCTGCTGC	Rad51_6 ORF	
	Reverse	TAC **ACTAGT** TGCATCCTCGATGCCGTGCT		
PIF6	Forward	TAC **CAT ATG**GCGGCTGTTGCGATACG	PIF6 ORF	
	Reverse	TAC **ACTAGT** CCCATCCTCTGTTTTCGCCC		
G00	Reverse	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATA ACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC		Generate sgRNA
JD102	Forward	**N_{24}** GAGCAGTCGAATGCCGAGCG **N_{20}**	Rad51 5'UTR	
JD103		**N_{24}** TCGTGGATGTCGCGGCAAGC **N_{20}**	Rad51 3'UTR	
JD018		**N_{24}** TTCTCTTCTCTCTCTTCTCT **N_{20}**	Rad51_3 5'UTR	
JD019		**N_{24}** ATTCGAGCAAGTGAAGGGTG **N_{20}**	Rad51_3 3'UTR	
JD020		**N_{24}** TGTCTTGGTGACACAGTAAG **N_{20}**	Rad51_4 5'UTR	
JD021		**N_{24}** CCCTTGGCATCCTTATCGTC **N_{20}**	Rad51_4 3'UTR	
JD022		**N_{24}** ACGAGGACAGGAGGGGAGAC **N_{20}**	Rad51_6 5'UTR	
JD023		**N_{24}** GTGCGTGGCACTGAGAAGTA **N_{20}**	Rad51_6 3'UTR	
JD104		**N_{24}** GGTGGGGGTGGGGTCGAGAG **N_{20}**	PIF6 5'UTR	
JD105		**N_{24}** GTGGTACACGTGTGTGTGTG **N_{20}**	PIF6 3'UTR	
JD031	Forward	ACTGCACTGGCGTCTCCTCCGCCCTTTTTC **NN_{10}**	pGL2314RAD51^{Flox}	Generate Donor fragment
JD032	Reverse	GCCGAGAAGGGCGTGCGATACTTTTCGATG **NN_{11}**		
JD033	Forward	GTGCCCGGCTTGTGCGGAGCGCCAGTCGCC **NN_{10}**	pGL2314RAD51_3^{Flox}	
JD034	Reverse	GGCAGCGAGCGCAAACTTGTGTGTGTGTGT **NN_{11}**		
JD035	Forward	CGTGTGTCTCCCCTCAGTCCTCGCCGCCCC **NN_{10}**	pGL2314RAD51_4^{Flox}	
JD036	Reverse	AGTGTAGGAGAGTGAGATCGCGCGCGGCAA **NN_{11}**		
JD037	Forward	TGTGTGTCTACATCTGTGCGTACGCCGAAG **NN_{10}**	pGL2314RAD51_6^{Flox}	
JD038	Reverse	GCCCTCGCATTCGGCGAGACAGACAGAGAG **NN_{11}**		
JD039	Forward	ACAGTCAAACATCGTGTCTTTCTGTTCTGT **NN_{10}**	pGL2314PIF6^{Flox}	
JD040	Reverse	CGCCACATTGCCGAGCCGCAGCTTCGGGGC **NN_{11}**		

CAT ATG = *Nde*I cut site; **ACTAGT** = *Spe*I cut site; **N_{24}** = GAAATTAATACGACTCACTATAGG; **N_{20}** = GTTTTAGAGCTAGAAATAGC; **NN_{10}** = CGTTGTAAAACGACGGCCAG; **NN_{11}** = CTCCATAGATCCATCCGCAC

Figure S1

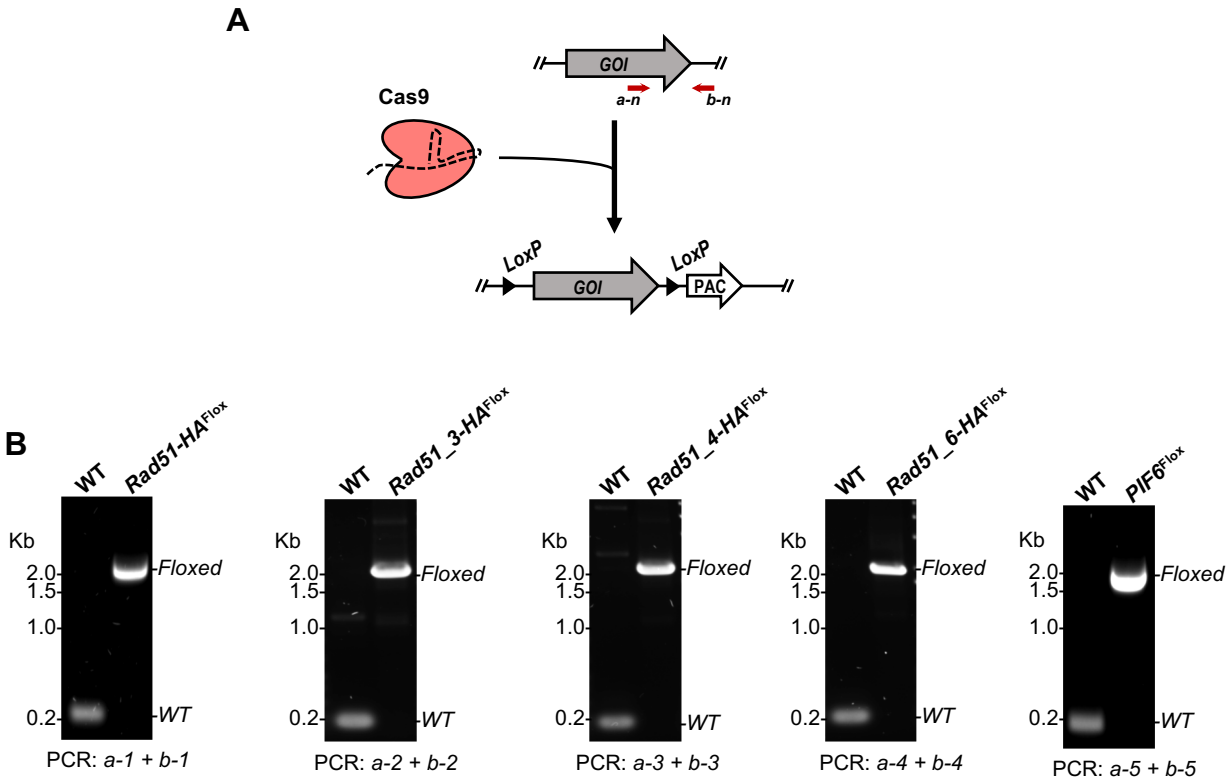

Figure S1: Combining CRISPR/Cas9 and DiCre to rapidly generate cell lines for inducible knockout. (A) Cas9 was used to replace all copies of a gene of interest (*GOI*) by a version of the same *GOI* flanked by *LoxP* sites (GOI^{Flox}); **(B)** PCR analysis of genomic DNA extracted from the indicated cell lines; approximated annealing positions for primers *a-n* and *b-n* (where n varies from 1 to 5, indicating a distinct sequence for the targeted *GOI* in each cell line) are shown in (A).

Figure S2

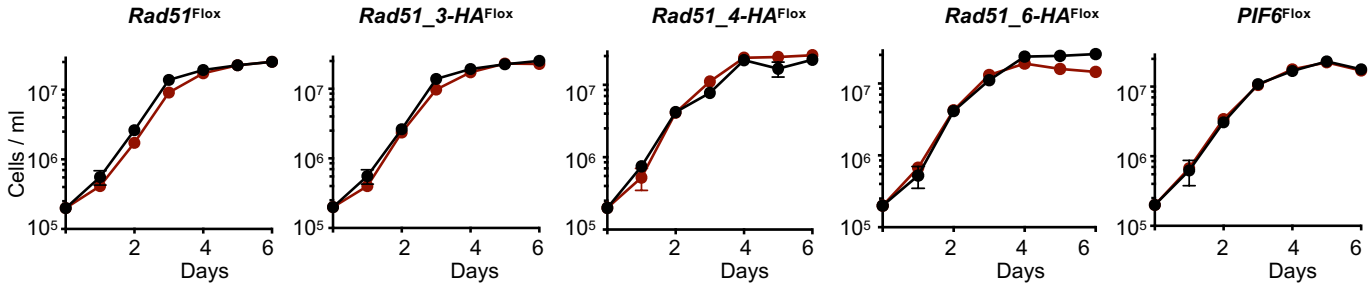

Figure S2: Growth profile analysis. Representative growth curves of the indicated cell lines (red lines) as compared to the background cell line expressing Cas9 and Dicre (black line); growth curves were started with $2x10^5$ cells/ml; cell density was assessed every 24 h and error bars depict standard error of the mean (S.E.M.).

Figure S3

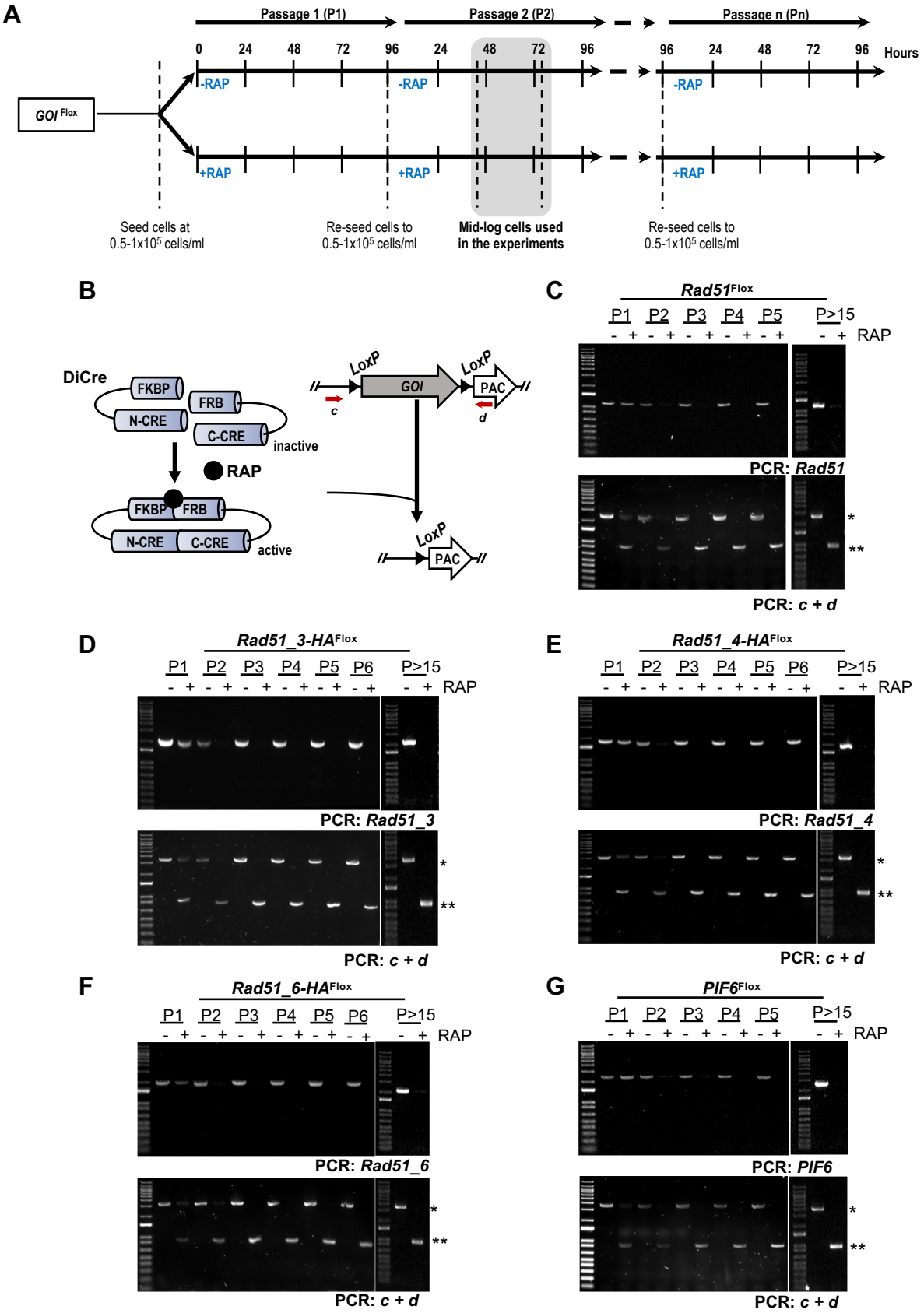
A
Passage 1 (P1)
Passage 2 (P2)
Passage n (Pn)
0 24 48 72 96 24 48 72 96 96 24 48 72 96 Hours
GOI Flox
-RAP
+RAP
Seed cells at 0.5-1x10⁵ cells/ml
Re-seed cells to 0.5-1x10⁵ cells/ml
Mid-log cells used in the experiments
Re-seed cells to 0.5-1x10⁵ cells/ml
B
DiCre
FKBP
FRB
N-CRE
C-CRE
inactive
RAP
active
LoxP
GOI
PAC
c
d
C
Rad51Flox
P1 P2 P3 P4 P5 P>15
RAP
PCR: Rad51
PCR: c + d
D
Rad51_3-HAFlox
P1 P2 P3 P4 P5 P6 P>15
PCR: Rad51_3
PCR: c + d
E
Rad51_4-HAFlox
P1 P2 P3 P4 P5 P6 P>15
PCR: Rad51_4
PCR: c + d
F
Rad51_6-HAFlox
P1 P2 P3 P4 P5 P6 P>15
PCR: Rad51_6
PCR: c + d
G
PIF6Flox
P1 P2 P3 P4 P5 P>15
PCR: PIF6
PCR: c + d

Figure S3: Dynamics of KO induction. (A) Illustration of KO induction scheme; cells were seeded in medium with (+RAP) or without (-RAP) rapamycin; after 4 days (~96 hours) of cultivation, cells were re-seeded, cultivated further and then diluted again; all the experiments reported here were performed in cells subjected to this induction protocol; times points indicated in the figures refer to the second passage (P2) .**(B)** Illustration of GOI^{Flox} excision catalyzed by DiCre, as induced by rapamycin. **(C)-(G)** PCR analysis of genomic DNA from the indicated cell lines throughout the indicated passages; DNA was extracted from cells ~72h of each passage; approximated annealing positions for primers *c* and *d* are shown in (A); (*) and (**), GOI^{Flox} and GOI^{Flox} after excision, respectively.

Figure S4

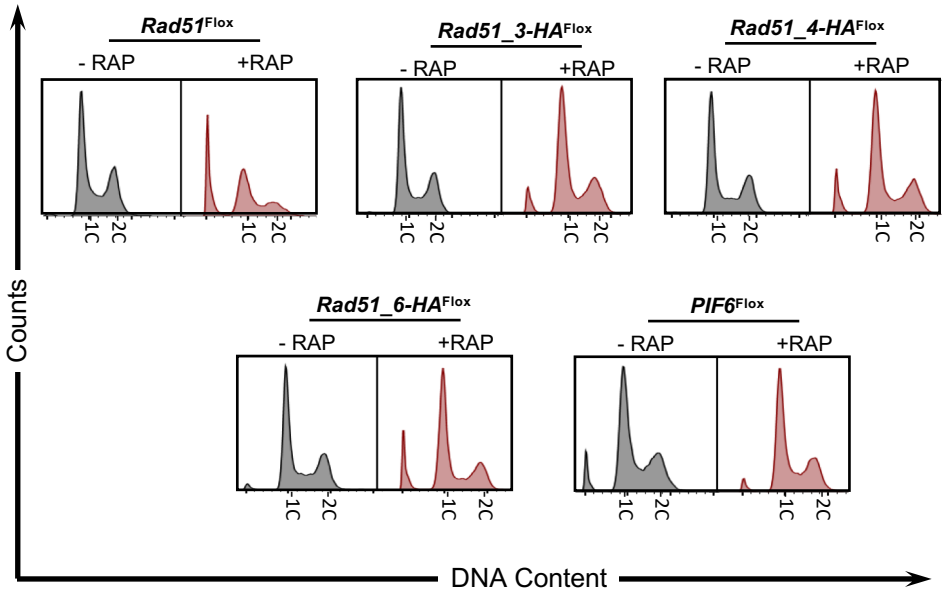

Figure S4: Analysis of DNA content profile upon prolonged cultivation after KO induction of homologous recombination factors. Representative histograms from FACS analysis to determine the distribution of cell population according to DNA content in cells kept in culture for more than 15 passages; 30,000 cells were analysed per sample; 1C and 2C indicate one DNA content (G1) and double DNA content (G2/M), respectively.

Figure S5

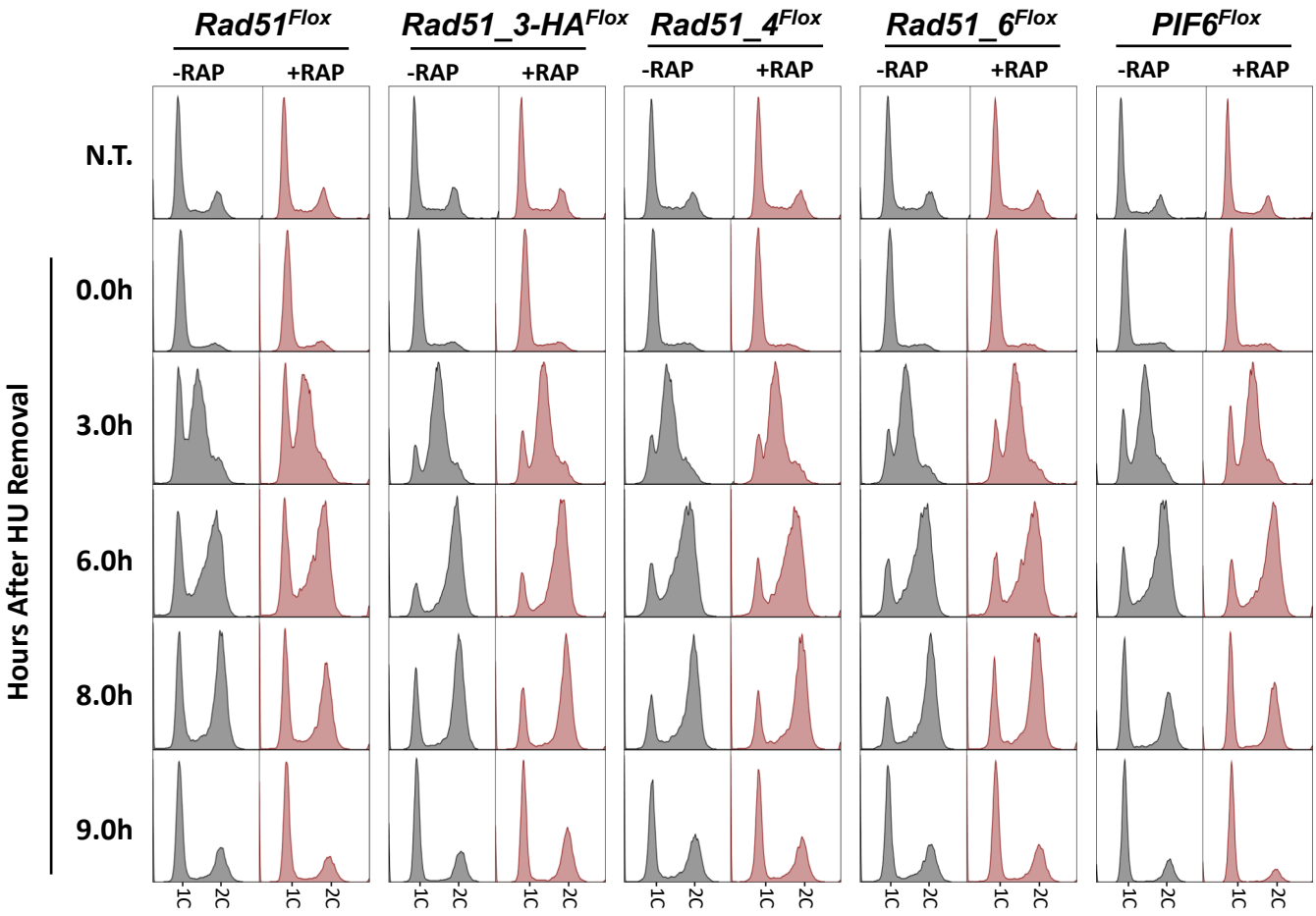

Figure S5: Cell cycle progression analysis after replication stress upon KO induction of homologous recombination factors. The indicated cell lines were left untreated (N.T.) or treated ~8 hours with 5mM HU and then re-seeded in HU-free medium; cells were collected at the indicated time points after HU removal, fixed, stained with Propidium Iodide and analysed by FACS; 1C and 2C indicate one DNA content (G1) and double DNA content (G2/M), respectively.

Figure S6

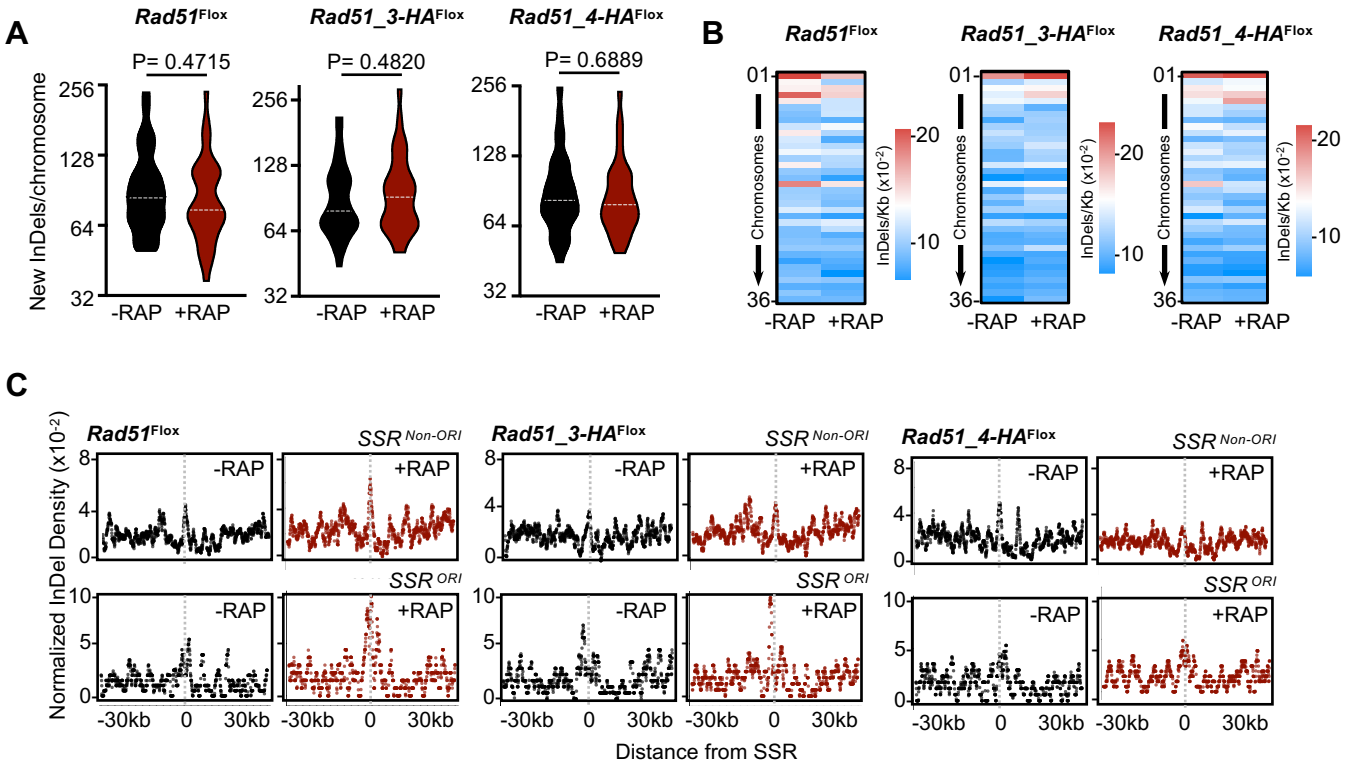

Figure S6: Whole genome analysis of InDels accumulation patterns upon KO induction of single or combined homologous recombination factors. (A) Quantification of the number of new InDels detected after 4 passages(between P6 and P2); data are represented as violin plots, where shape indicates the distribution of pooled data and horizontal doted white lines indicate the median; differences were tested with Mann-Whitney test and P values are shown above each pair of data sets **(B)** Heatmap representing density of new InDels (InDels/Kb) detected after 4 passages in each chromosome. **(C)** Metaplots of normalized density of new InDels (InDels/Kb) after 4 passages is plotted +/- 30 Kb around the centre of either SSR^{ORI} (n = 36) or $SSR^{non\text{-}ORI}$ (n = 95) for the indicated cell lines.

Figure S7

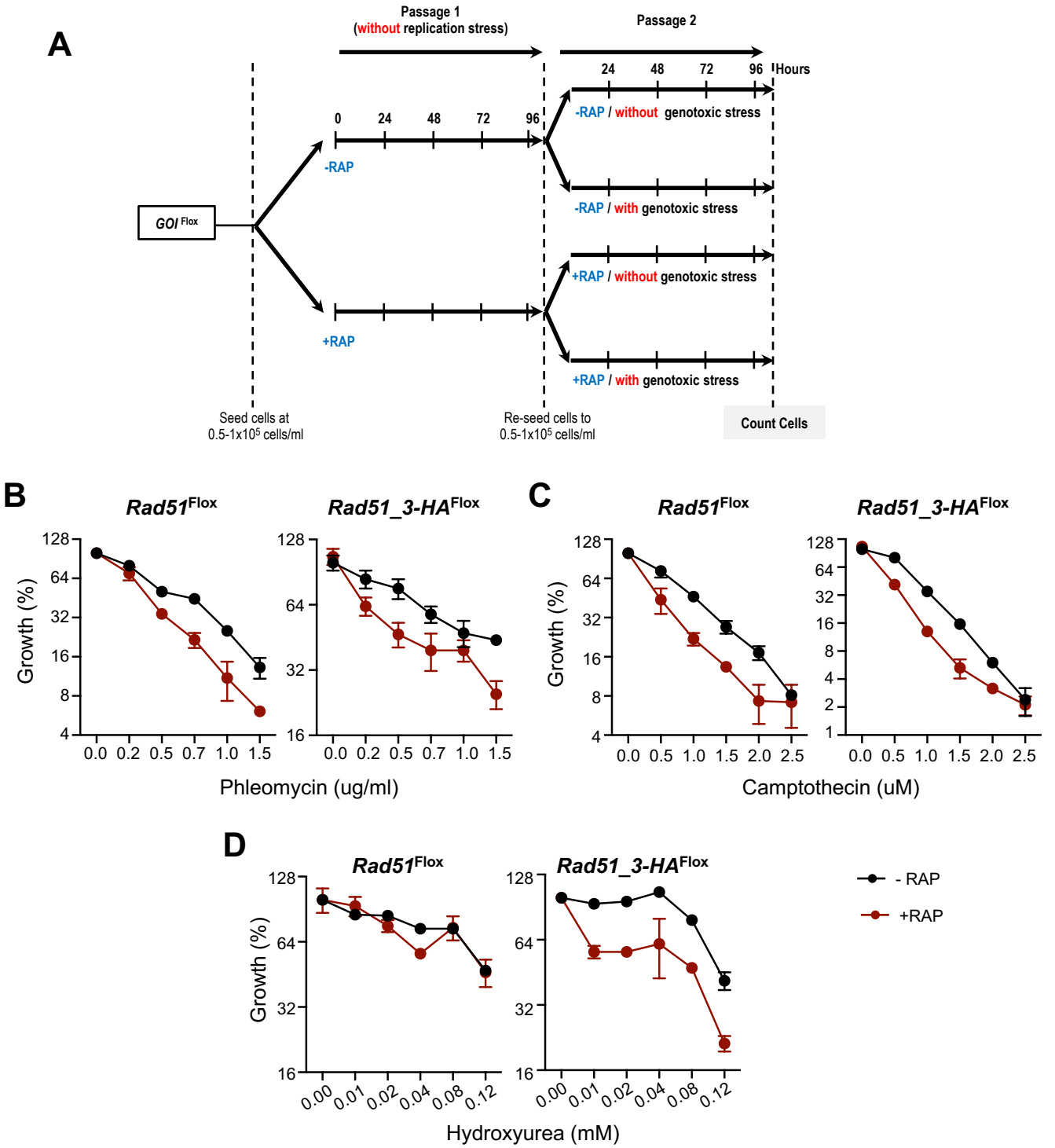

Figure S7: Genotoxic stress resistance profile upon KO induction of Rad51 and Rad51_3 . (A) Experimental design to evaluate resistance to genotoxic agents as shown in (B-D); cells were seeded in medium with (+RAP) or without (-RAP) rapamycin, in the absence of any genotoxic drug; after ~96 hours of the first passage, cells were re-seeded in medium with or without genotoxic agents at various concentration; after ~96 hours of the second passage, cell density in each condition was determined. **(B – D)** Relative growth of cells incubated with the indicated concentration of the indicated genotoxic agents, during passage 2; growth in each concentration is expressed as percentage of proliferation relative to cells cultivated without the genotoxic drugs; error bars depict SD.

Figure S8

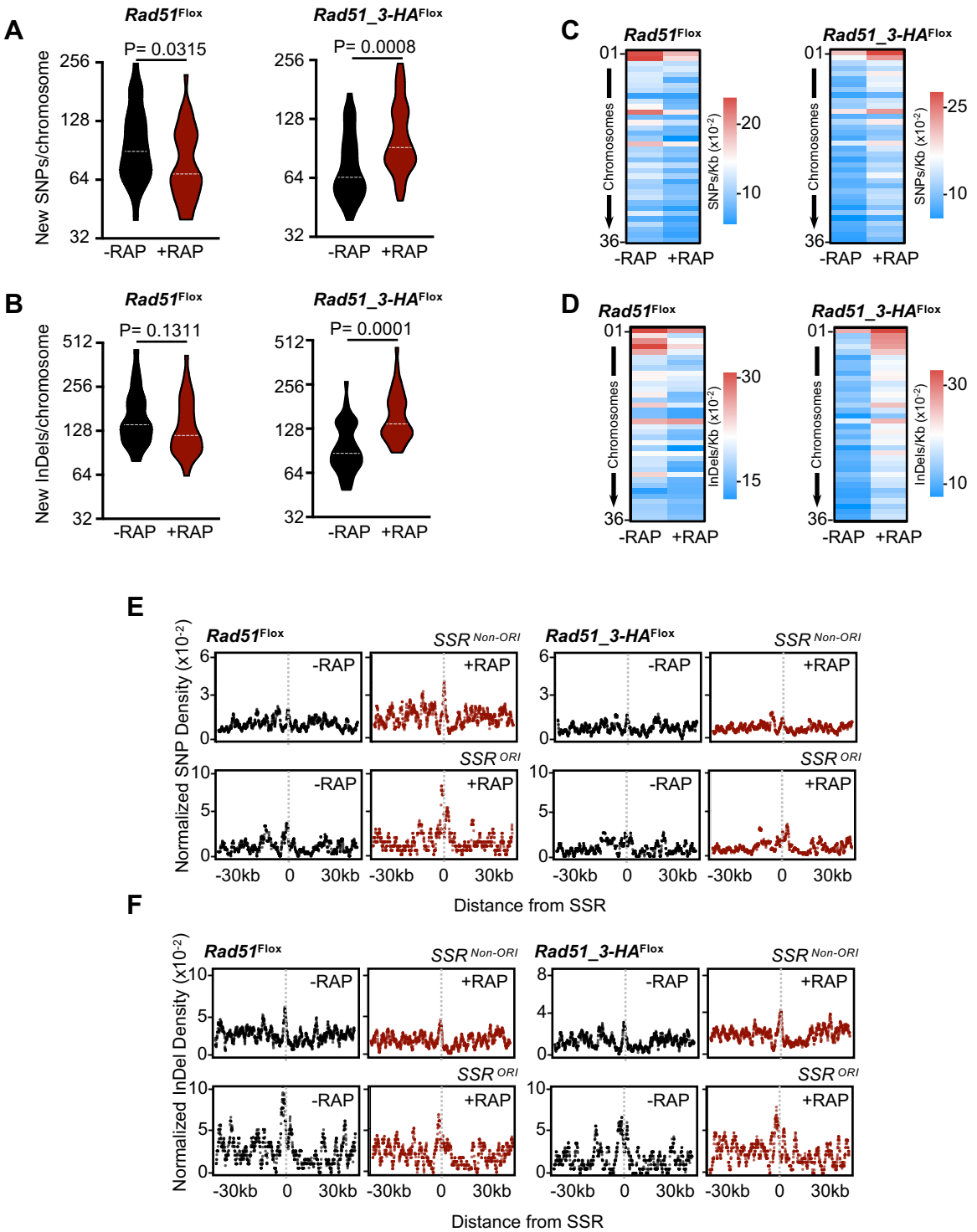

Figure S8: Whole genome analysis of SNPs and InDels accumulation patterns after replication stress upon KO induction of Rad51 and Rad51_3. (A) and **(B)** Quantification of the number of new SNPs and new InDels, respectively, detected 96 hours after cells were released from HU treatment; data are represented as violin plots, where shape indicates the distribution of pooled data and horizontal doted white lines indicate the median; differences were tested with Mann-Whitney test and P values are shown above each pair of data sets. **(C)** and **(D)** Heatmap representing density of new SNPs (SNPs/Kb) and new InDels (InDels/Kb), respectively, detected 96 hours after cells were released from HU treatment. **(E)** and **(F)** Metaplots of normalized density of new SNPs (SNPs/Kb) and new InDels (InDels/Kb), respectively, detected 96 hours after cells were released from HU treatment is plotted +/- 30 Kb around the centre of either SSR^{ORI} (n = 36) or $SSR^{non\text{-}ORI}$ (n = 95) for the indicated cell lines.

Figure S9

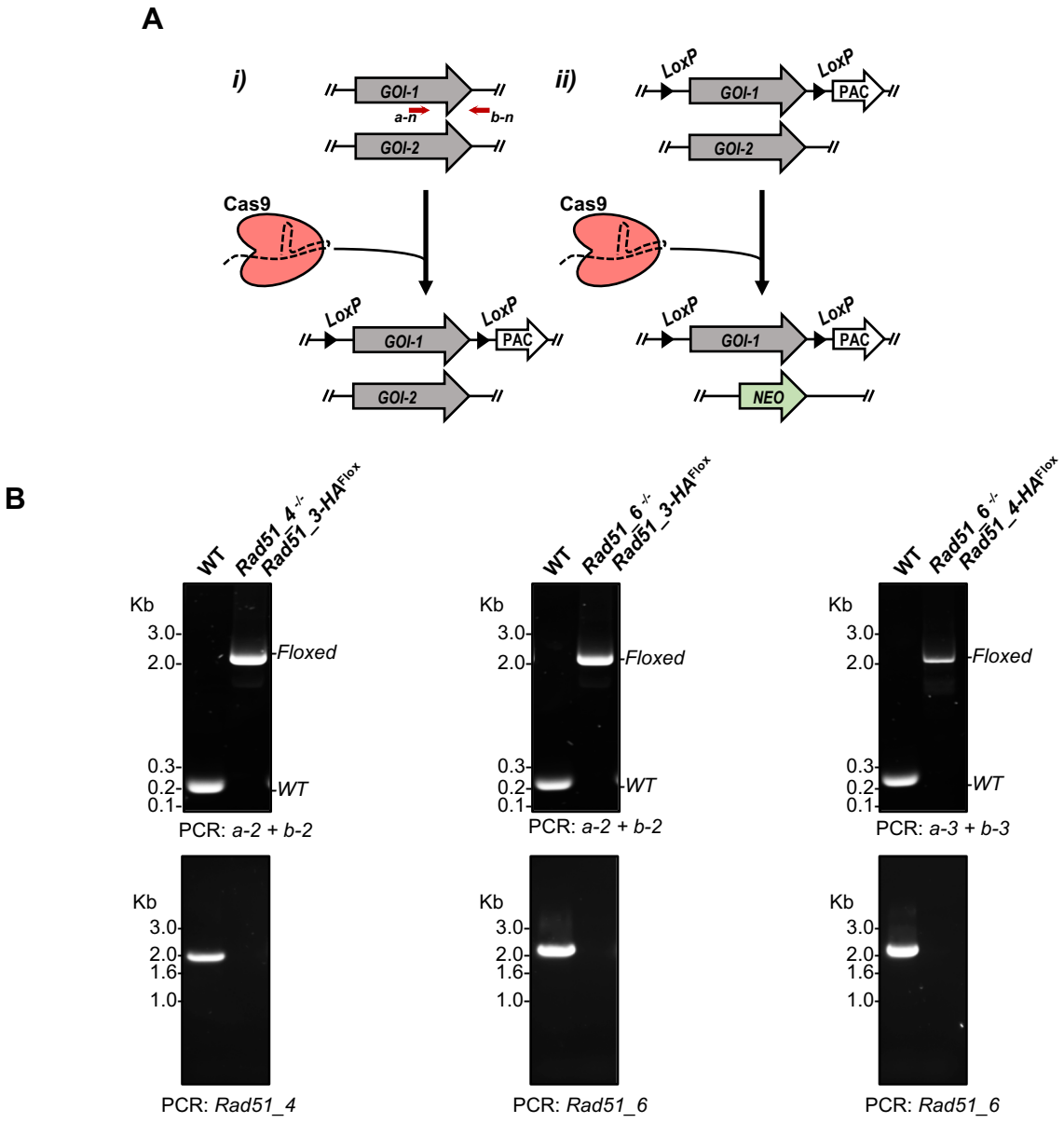

Figure S9: Combining CRISPR/Cas9 and DiCre to rapidly generate double KO cell lines. (A) *i)* Cas9 was used to replace all copies of a gene of interest (*GOI-1*) by a version of the same *GOI* flanked by *LoxP* sites (*GOI-1*^{Flox}); *ii)* in the same cell line, Cas9 was used to replace all copies of another gene of interest (*GOI-2)* by a Neomycin resistance gene (NEO); **(B)** PCR analysis of genomic DNA extracted from the indicated cell lines; approximated annealing positions for primers *a-n* and *b-n* (where n varies from 1 to 5, indicating a distinct sequence for the targeted *GOI* in each cell line) are shown in (A).

Figure S10

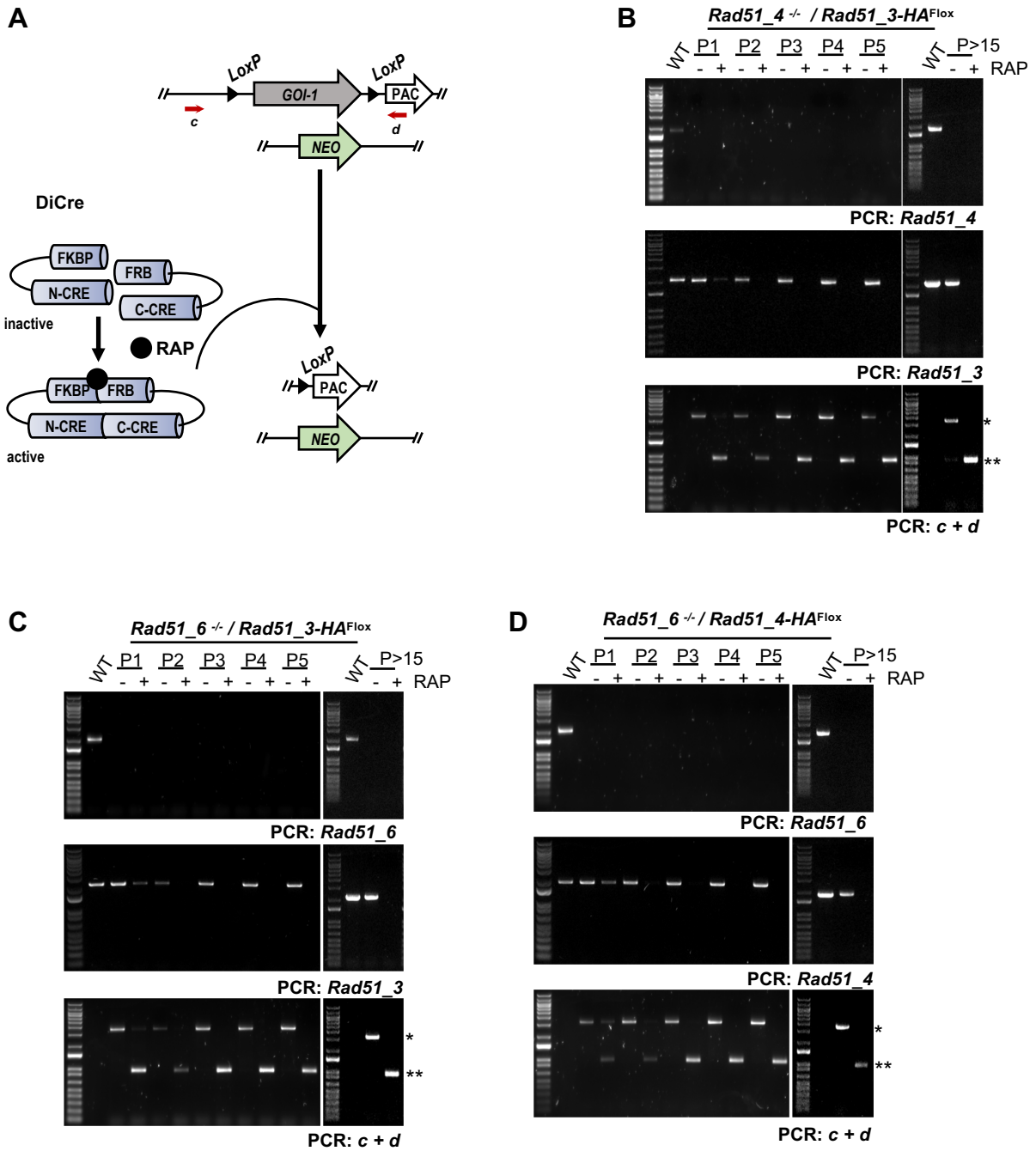

Figure S10: Dynamics of KO induction. Induction was performed as depicted in Figure S3A. **(A)** Illustration of GOI^{Flox} excision catalyzed by DiCre, as induced by rapamycin, to render double KO cells. **(B) - (D)** PCR analysis of genomic DNA from the indicated cell lines throughout the indicated passages; approximated annealing positions for primers *c* and *d* are shown in (A); (*) and (**), *GOI^{Flox}* and *GOI^{Flox}* after excision, respectively.

Figure S11

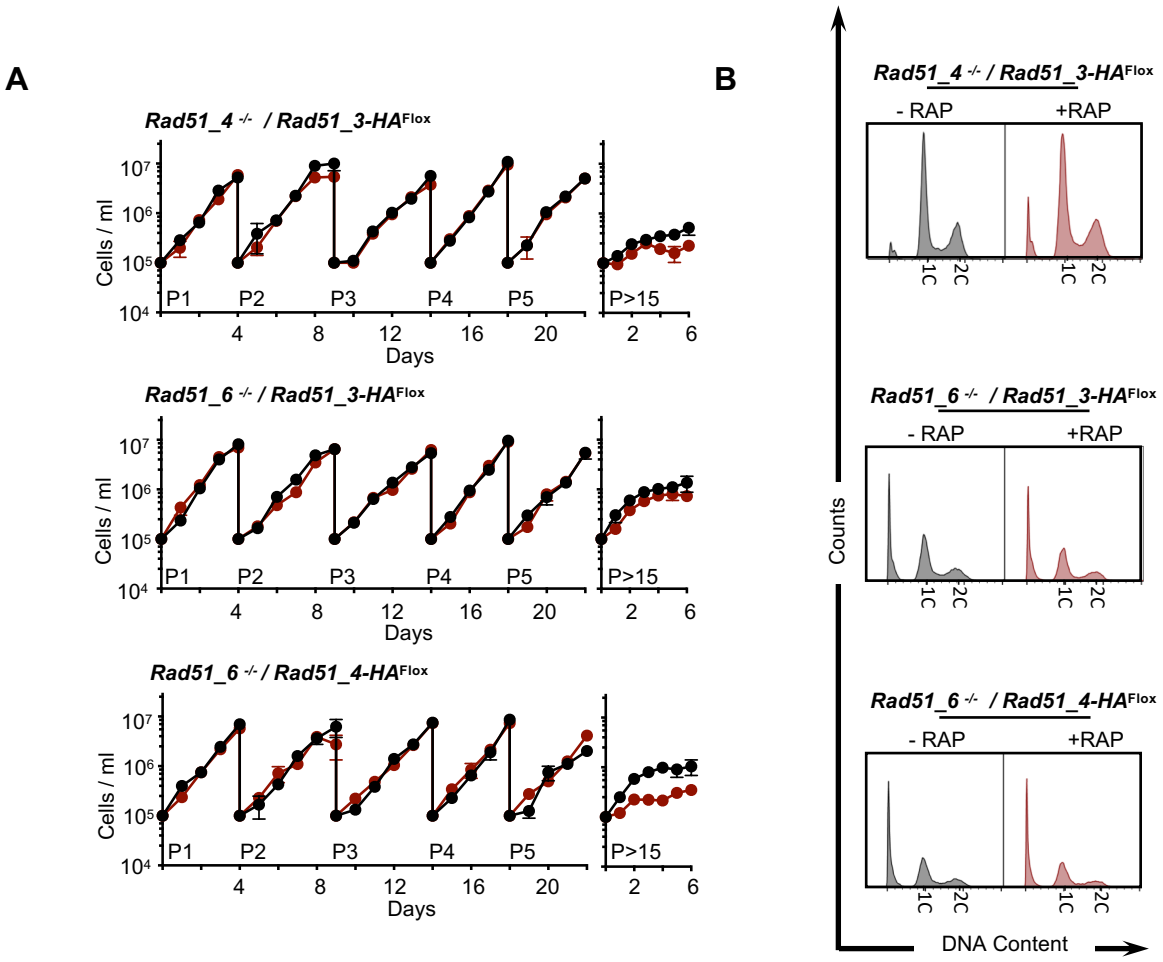

Figure S11: Effects of double KO of Rad51 paralogues. (A) Representative growth curves of the indicated cell lines in the presence or absence of RAP; cells were seeded at 10^5 cells/ml in day 0 and re-seeded every 4 - 5 days to complete five passages (P1 to P5); growth profile was also evaluated after cells were kept in culture for more than 15 passages (>P15); cell density was assessed every 24 h and error bars depict standard error of the mean (S.E.M.). **(B)** Representative histograms from FACS analysis to determine the distribution of cell population according to DNA content in cells kept in culture for more than 15 passages; 30,000 cells were analysed per sample; 1C and 2C indicate one DNA content (G1) and double DNA content (G2/M), respectively.

Figure S12

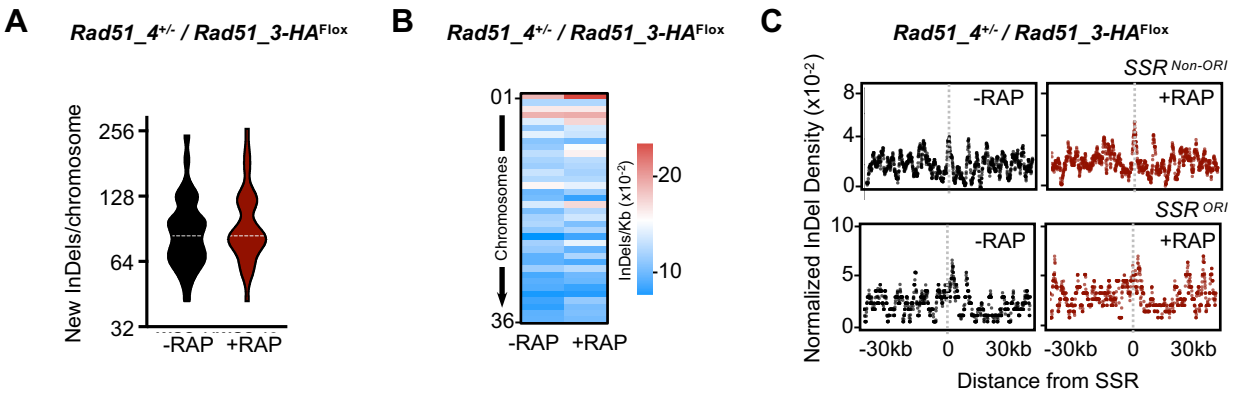

Figure S12: Whole genome analysis of InDels accumulation patterns upon double KO of Rad51-3 and Rad51-4. (A) Quantification of the number of new InDels detected after 4 passages(between P6 and P2); data are represented as violin plots, where shape indicates the distribution of pooled data and horizontal doted white lines indicate the median. **(B)** Heatmap representing density of new InDels (InDels/Kb) detected after 4 passages in each chromosome. **(C)** Metaplots of normalized density of new InDels (InDels/Kb) after 4 passages is plotted +/- 30 Kb around the centre of either SSR^{ORI} (n = 36) or $SSR^{non\text{-}ORI}$ (n = 95) for the indicated cell lines.

Figure S13

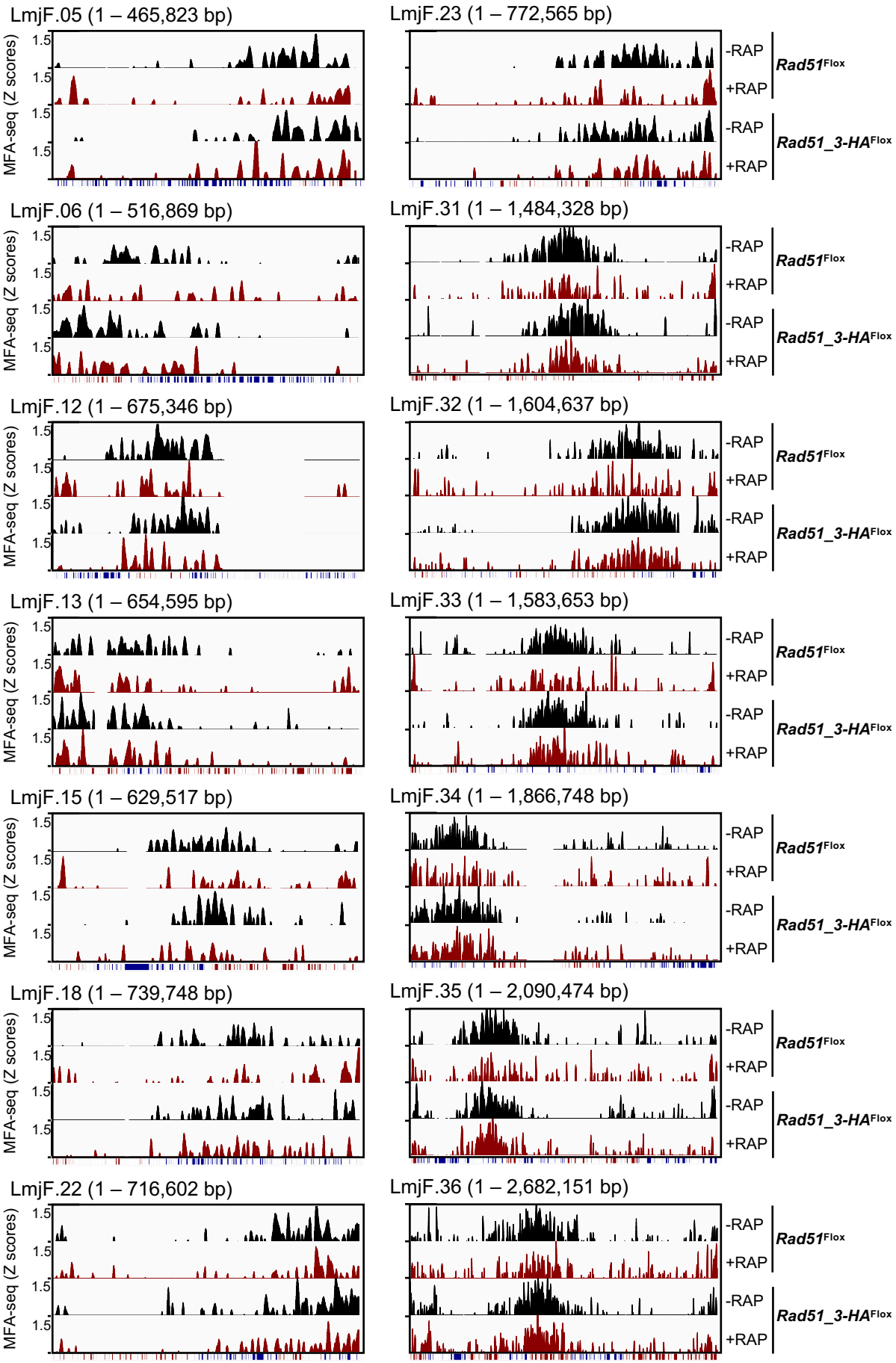
LmjF.05 (1 – 465,823 bp)
LmjF.23 (1 – 772,565 bp)
LmjF.06 (1 – 516,869 bp)
LmjF.31 (1 – 1,484,328 bp)
LmjF.12 (1 – 675,346 bp)
LmjF.32 (1 – 1,604,637 bp)
LmjF.13 (1 – 654,595 bp)
LmjF.33 (1 – 1,583,653 bp)
LmjF.15 (1 – 629,517 bp)
LmjF.34 (1 – 1,866,748 bp)
LmjF.18 (1 – 739,748 bp)
LmjF.35 (1 – 2,090,474 bp)
LmjF.22 (1 – 716,602 bp)
LmjF.36 (1 – 2,682,151 bp)
MFA-seq (Z scores)
1.5
-RAP
+RAP
Rad51Flox
Rad51_3-HAFlox

Figure S13: Genome-wide mapping of replication initiation upon RAD51 and RAD51-3 KO. Graphs show the distribution of sites of DNA synthesis initiation across the indicated chromosomes in the indicated cell lines, in each case grown in the absence (-RAP) or the presence (+RAP) of rapamycin. MFA-seq is represented by Z-scores across the chromosomes, calculated by comparing read depth coverage of DNA from exponentially growing cells relative to stationary cells; the bottom track for each chromosome displays coding sequences, with genes transcribed from right to left in red, and from left to right in blue .