RAD51 promotes non-homologous genetic rearrangements that are prevented by 53BP1

[1,2]Josée Guirouilh-Barbat*, [3]Wei Yu, [4]Loelia Babin, [4]Elisa Yaniz Galende, [5]Tatiana Popova, [1,2]Ayeong So, [3]Chloé Lescale, [5]Marc-Henri Stern, [4]Erika Brunet, [3]Ludovic Deriano, and [1,2]Bernard S. Lopez*.

[1]Gustave-Roussy cancer institute, CNRS UMR 8200, Université Paris-Saclay, Team labeled by "Ligue Nationale contre le cancer, Ligue 2017", 114 rue Edouard Vaillant, 94805 Villejuif, France.

[2]Institut Cochin, INSERM U1016, UMR 8104 CNRS, Université Paris-Descartes, 24 rue du Faubourg St Jacques, 75014 Paris, France.

[3]Genome Integrity, Immunity and Cancer Unit, Department of Immunology, Department of Genomes and Genetics, Institut Pasteur, 75015 Paris, France.

[4]Genome Dynamics in the Immune System Laboratory, Institut Imagine, INSERM, UMR 1163, Université Paris Descartes, Sorbonne Paris Cité, Equipe Labellisée Ligue Contre le Cancer, Paris, France.

[5]Institut Curie, PSL Research University, INSERM U830, DNA Repair and Uveal Melanoma (D.R.U.M.), Equipe labellisée par la Ligue Nationale contre le Cancer, Paris, 75248, France.

*Correspondence: bernard.lopez@inserm.fr; josee.guirouilh-barbat@inserm.fr

1

Abstract

Homologous recombination (HR), which requires long sequence homologies, is considered a high fidelity mechanism, preserving genome stability. In contrast, we show here that the central HR players RAD51 or BRCA2, promote genetic instability, fostering translocations and capture of ectopic chromosomal sequences when joining distant DNA breaks. Surprisingly, these events do not involve sequence homologies. Moreover, our data reveal that 53BP1 protects against RAD51-mediated non-homologous genetic rearrangements. Finally, analysis of a large panel of breast tumors revealed that BRCA2 proficiency is associated with increased frequency of capture of non-homologous sequences at junctions of structural variants (translocations, duplications, inversions, deletions). These data reveal that HR proteins (RAD51, BRCA2) possess the intrinsic capacity to generate genetic instability through sequence homology-independent processes, and that 53BP1 protects against it. We propose that BRCA2/RAD51-mediated genome instability occurs in the course of sequence homology search for HR.

Keywords:, Homologous recombination; RAD51; BRCA2; 53BP1; Double strand break repair; genomic instability; genome rearrangements; translocations; micro-homologies; human cells

Introduction

Genetic instability is associated with tumorigenesis, developmental syndromes and pre-mature ageing (Garinis et al., 2008; Hoeijmakers, 2009; Lans and Hoeijmakers, 2006; Negrini et al., 2010). Cells developed elaborate systems protecting genome stability, albeit allowing genetic diversity. While mechanisms involved in genome stability/diversity maintenance are largely documented, the molecular mechanisms driving genetic instability are far from being fully characterized, at a molecular level.

DNA double-strand break (DSB) is one of the most harmful lesions and a major source of genetic variability. Indeed, DSBs generate both genome rearrangements (e.g., translocations, inversions, duplications, deletions, and complex rearrangements) and mutagenesis at the sealed junctions themselves. Cells use two primary strategies to repair DSBs. The first strategy joins and ligates two DNA double-strand ends without requiring sequence homology and is referred to as non-homologous end-joining (NHEJ). The second strategy relies on the use of a DNA partner sharing long sequence homologies and is thus referred to as homologous recombination (HR).

HR, which requires hundreds of base pairs of homology between the two interacting DNA partners, is generally considered a high fidelity process essential for genome stability maintenance. HR is an evolutionarily highly conserved process that can be divided into sub-classes all initiated by common mechanisms (Haber, 2014): 1) Resection of DSBs, generating 3' single-stranded DNA tails (ssDNA); 2) loading of RAD51 on the ssDNA by BRCA2/PALB2, generating the actual active species of HR, i.e., the ssDNA/RAD51 filament; 3) search for sequence homologies and strand invasion of a homologous intact duplex DNA, representing thus the pivotal steps of HR performed by the ssDNA/RAD51 filament; and 4) DNA synthesis primed on the invading 3' end. Then, depending on the subsequent steps and resolution of intermediates, the process will

3

lead to gene conversion associated or not with crossovers, to break–induced replication (BIR) or to synthesis-dependent strand annealing (SDSA). Importantly, the length of homology between the two partners is a paramount parameter for HR. This point was extensively studied at the end of the last millennium leading to the definition of the MEPS (minimal effective treatment segment), ranging between 250 and 300 bp in mammalian cells (Bollag et al., 1989; Liskay et al., 1987; Lopez et al., 1992; Rubnitz and Subramani, 1984). Therefore, it is widely admitted that HR requires hundred(s) bp of sequence homology, in mammalian cells.

In addition to HR and NHEJ (in fact the KU-ligase IV-dependent canonical NHEJ, C-NHEJ), alternative processes mediated by microhomologies (2-4 bp) have been described: alternative end-joining (A-EJ), also called backup NHEJ (B-NHEJ) or microhomology-mediated end-joining (MMEJ); microhomology-mediated BIR (MMBIR); fork stalling and template switching (FoSTeS); and microhomology-mediated template switching (MMTS) (for review see (Deriano and Roth, 2013; So et al., 2017)). Because they are independent of KU-ligase 4 and they do not require long sequence homologies, but only few bp of microhomology (thus far from the MEPS), these events are distinct from both C-NHEJ and HR.

Using an intrachromosomal substrate (CD4-3200bp), we have shown that few kb of distance between two DNA double-stranded ends are sufficient to favor the capture of non-homologous ectopic chromosomal sequences (≥ 45 bp) at the sealed junction. Such kind of insertions jeopardizes genome integrity and is commonly observed in cancer cells. However, the underlying mechanisms remain to be characterized. The borders of the insertions are characterized by the absence of homologous sequences and, in contrast, the presence of microhomologies or unmappable patches of nucleotides; therefore, these events do not correspond to HR-related events (Guirouilh-Barbat et al.,
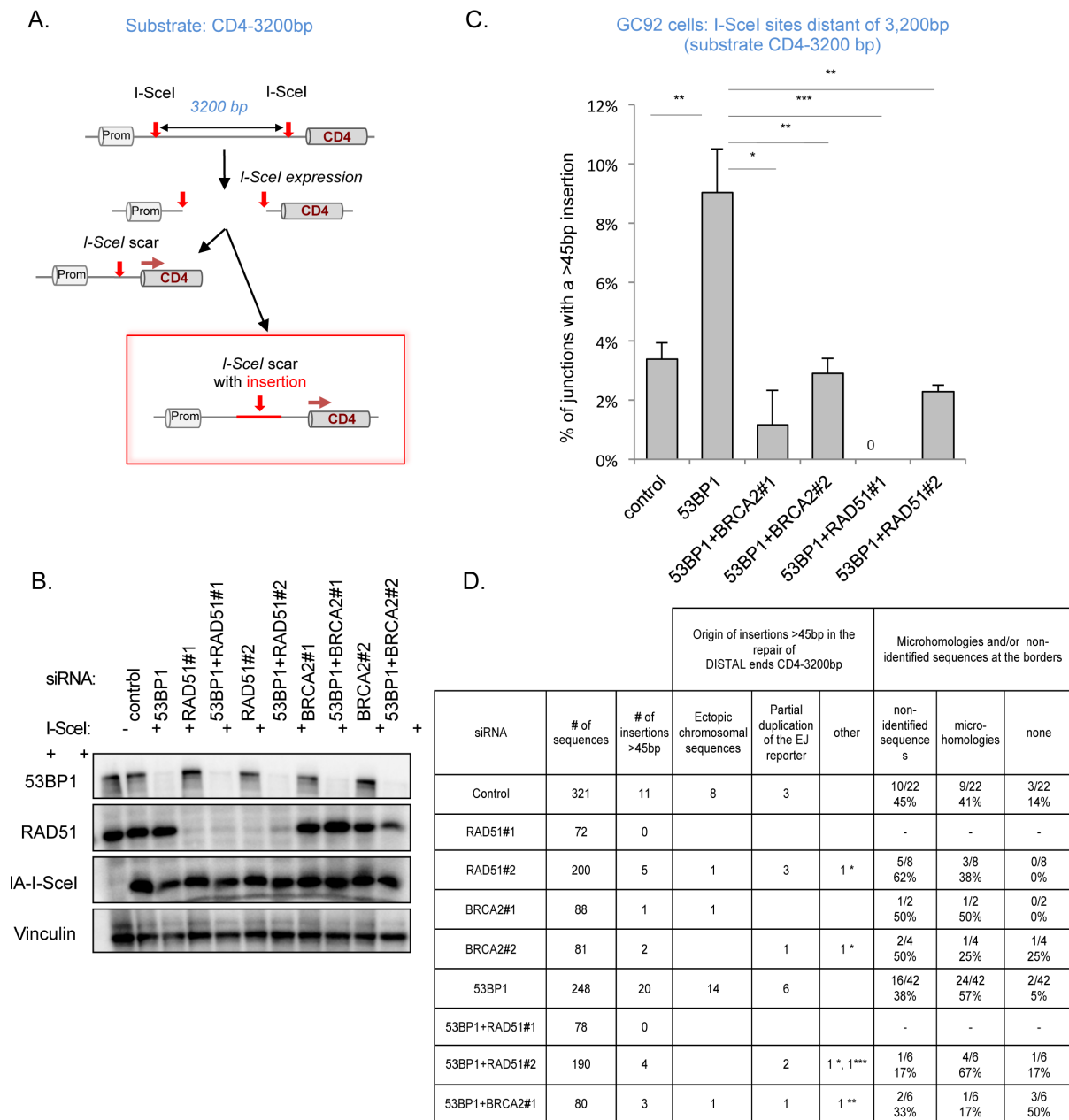
4

2016), but rather to microhomology-mediated insertions. We proposed that MMTS mediates the captures of ectopic sequences (Guirouilh-Barbat et al., 2016). The cohesion complex, which restrains the joining of distant ends by channeling DSB repair toward the use of sister chromatids, also limits the capture of ectopic sequences (Gelot et al., 2016). Because sister chromatid exchanges involve HR, we postulated that ablation of HR should also stimulate ectopic sequences capture. Surprisingly, silencing the HR proteins RAD51 or BRCA2 did not stimulate ectopic sequences capture but instead decreased it, although no homologous sequences were involved. The impact of RAD51 depletion was more particularly observed in cells depleted of 53BP1 in which ectopic sequences capture is favored. In addition, we found that RAD51 was also implicated in the generation of inter-chromosomal translocations themselves fostered by 53BP1 depletion, although no sequence homology was involved. Therefore the pivotal HR player RAD51 generates genetic instability in a process that does not require long sequence homology. This activity strongly contrasts with the canonical role of RAD51 in HR. In addition, our data reveal that 53BP1 protects against RAD51-mediated genetic instability. We propose a model in which BRCA2/RAD51 promote genome instability through non-homologous recombination (or micro-homologous recombination) when the RAD51/ssDNA filament scans the DNA in the course of homology search for HR. These data unravel a molecular mechanism driving genetic instability: mediated by the pivotal HR players RAD51 and counteracted by 53BP1. Therefore a balance between RAD51 and 53BP1 activities governs genome stability maintenance.

Results

53BP1 counteracts the capture of non-homologous ectopic sequences promoted by RAD51 and BRCA2, at junctions of distant DSBs. Using the intrachromosomal substrate CD4-3200bp (Fig. 1A), we previously reported the capture of ectopic sequences at the joining of 3200 bp distant ends, in a sequence homology-independent manner. Moreover, we show that ablation of 53BP1 strongly increased the frequency of ectopic non-homologous sequences capture (Guirouilh-Barbat et al., 2016). Here we show that silencing the pivotal HR protein RAD51 or BRCA2 completely abolished the stimulation of sequence capture resulting from 53BP1 silencing (Fig. 1B-D, Table 1 and Supplementary data S1). These results were unexpected because the CD4-3200bp substrate does not share sequence homology with the human genome, and thus ectopic sequence capture cannot be scored as HR events. In the presence of 53BP1, ablation of RAD51 or BRCA2 barely affected the frequency of ectopic sequences capture (Supplementary data S2).

Some captured sequences were partial duplications of the intrachromosomal reporter substrate, but the majority originated from other chromosomes (Fig. 1D and Supplementary data S1). Although both intra- and inter-chromosomal sequences captures are RAD51/BRCA2 dependent, no long stretches of sequence homology were present at the borders (Supplementary data S1); instead, 2-4 bp microhomologies or unmappable stretches of nucleotides were present. Few borders exhibited neither microhomologies nor unmappable sequences (Fig. 1D).

Therefore, in contrast with the canonical sequence homology-dependent role of RAD51/BRCA2 in HR and genome stability maintenance, RAD51 and BRCA2 promote the capture of non-homologous ectopic sequences, in human cells. In addition, these data reveal that 53BP1 protects against this RAD51- promoted genetic instability.
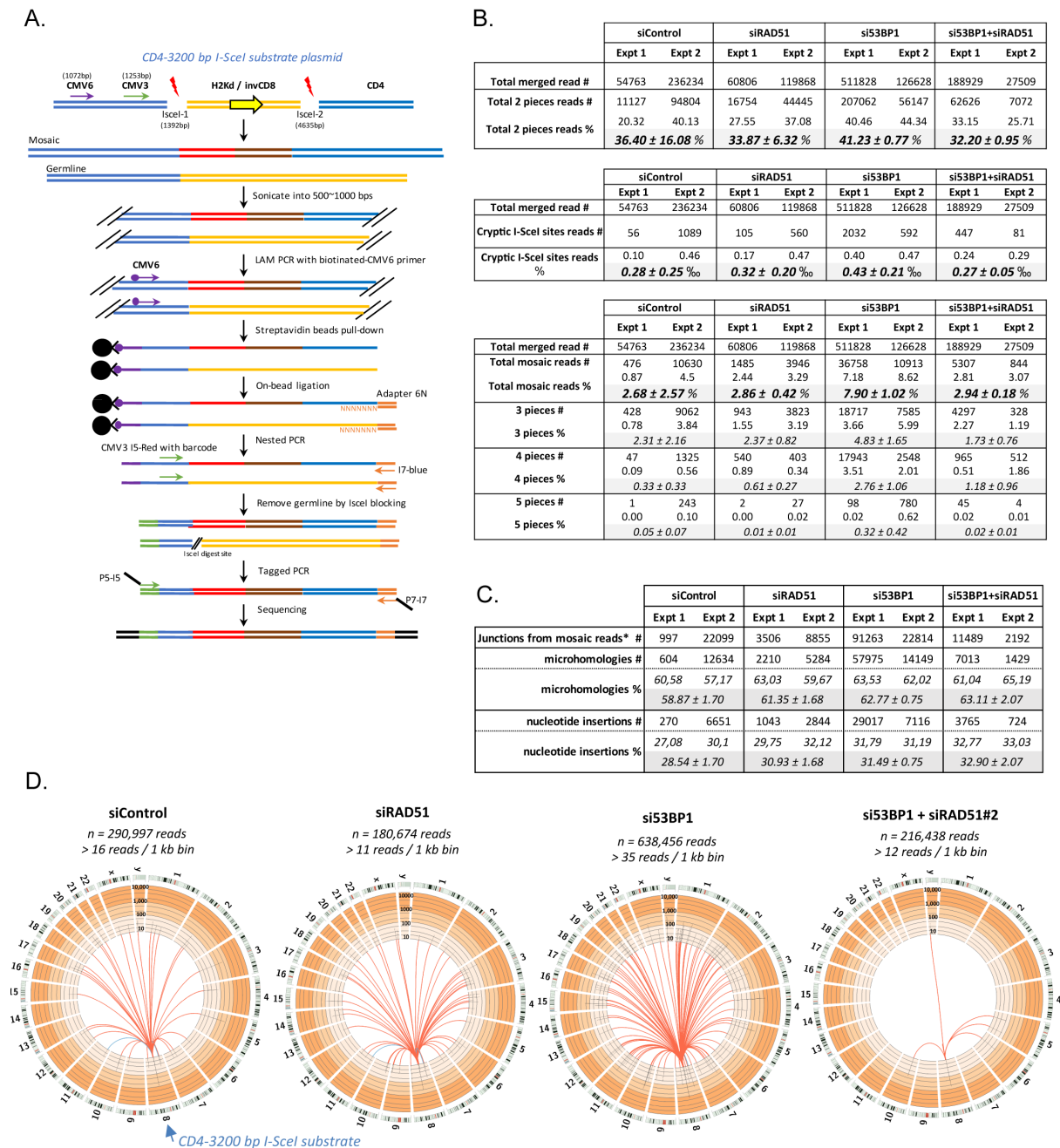
6

A. Substrate: CD4-3200bp

C. GC92 cells: I-SceI sites distant of 3,200bp (substrate CD4-3200 bp)

B.

D.

siRNA	# of sequences	# of insertions >45bp	Origin of insertions >45bp in the repair of DISTAL ends CD4-3200bp			Microhomologies and/or non-identified sequences at the borders		
			Ectopic chromosomal sequences	Partial duplication of the EJ reporter	other	non-identified sequences	micro-homologies	none
Control	321	11	8	3		10/22 45%	9/22 41%	3/22 14%
RAD51#1	72	0				-	-	-
RAD51#2	200	5	1	3	1 *	5/8 62%	3/8 38%	0/8 0%
BRCA2#1	88	1	1			1/2 50%	1/2 50%	0/2 0%
BRCA2#2	81	2		1	1 *	2/4 50%	1/4 25%	1/4 25%
53BP1	248	20	14	6		16/42 38%	24/42 57%	2/42 5%
53BP1+RAD51#1	78	0				-	-	-
53BP1+RAD51#2	190	4		2	1 *, 1***	1/6 17%	4/6 67%	1/6 17%
53BP1+BRCA2#1	80	3	1	1	1 **	2/6 33%	1/6 17%	3/6 50%

Fig. 1: BRCA2/RAD51 are required for ectopic sequences capture at the junction scars of distant ends. **A.** The CD4-3200bp substrate. **B.** Analysis of protein depletion and I-SceI expression in GC92 cells. **C.** Percentage of junctions (from I-SceI sites, 3200bp apart) containing ectopic sequences capture in the GC92 cells after 53BP1, RAD51 and/or BRCA2 depletion. Histograms represent the percentage of junctions with insertions >45 bp. Each value is the average of 2 to 7 independent experiments ± SEM. **D.** Origin of large insertions (>45bp) monitored in the repair of distal DSEs (CD4-3200bp) upon 53BP1, RAD51 and/or BRCA2 depletion, in the GC92 cell line. Values are calculated from 2 to 7 independent experiments and sequencing of 72 to 321 junction sequences (* I-SceI expression plasmid, ** bovine satellite, *** the whole insertion is non mappable).

7

Genome-wide analysis of chromosomal rearrangements at the I-SceI site. We then assessed the impact of RAD51 on genetic rearrangements at one DSB, at the genome wide level. We used linear amplification-mediated high-throughput genome-wide translocation sequencing (LAM-HTGTS) (Frock et al., 2015; Hu et al., 2016) to detect genome-wide chromosomal rearrangements at the CD4-3200bp chromosomal locus. As a bait for LAM-HTGTS, we chose the I-SceI DSB site close to the CD4 promoter of the CD4-3200bp chromosomal substrate and used two primers located 320 bp and 139 bp 5' upstream of the I-SceI site to amplify breakpoint junction sequences (Fig. 2A).

We analyzed genome-wide rearrangements at this I-SceI site upon I-SceI expression and silencing either RAD51, 53BP1 or both 53BP1 and RAD51. Analysis of the sequencing reads confirmed that a single copy of the CD4-3200bp substrate was located on chromosome 8 at approximately position 76,367 kb. We next searched for the presence of split reads among merged reads (Supplementary Fig. S3A). As expected, a majority of discordant reads corresponded to two-pieces split reads between the 5' promoter sequence of the I-SceI cutting site and sequences within the CD4-3200 bp substrate (Supplementary Fig. S3B). These events mainly correspond to end-joining events, as we have shown before (Grabarz et al., 2013; Guirouilh-Barbat et al., 2004, 2007; Rass et al., 2009). Other two-pieces split reads contained the bait promoter sequence and either unmappable sequences, the I-SceI donor plasmid sequence or sequences mapped to the human genome, including chromosomal translocations (Supplementary Fig. S3B). This latter category might also result from the direct end-joining of I-SceI-induced DSBs to prey DSBs occurring elsewhere in the genome.

A.

B.

	siControl		siRAD51		si53BP1		si53BP1+siRAD51	
	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
Total merged read #	54763	236234	60806	119868	511828	126628	188929	27509
Total 2 pieces reads #	11127	94804	16754	44445	207062	56147	62626	7072
Total 2 pieces reads %	20.32	40.13	27.55	37.08	40.46	44.34	33.15	25.71
	36.40 ± 16.08 %		*33.87 ± 6.32 %*		*41.23 ± 0.77 %*		*32.20 ± 0.95 %*	

	siControl		siRAD51		si53BP1		si53BP1+siRAD51	
	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
Total merged read #	54763	236234	60806	119868	511828	126628	188929	27509
Cryptic I-SceI sites reads #	56	1089	105	560	2032	592	447	81
Cryptic I-SceI sites reads %	0.10	0.46	0.17	0.47	0.40	0.47	0.24	0.29
	0.28 ± 0.25 ‰		*0.32 ± 0.20 ‰*		*0.43 ± 0.21 ‰*		*0.27 ± 0.05 ‰*	

	siControl		siRAD51		si53BP1		si53BP1+siRAD51	
	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
Total merged read #	54763	236234	60806	119868	511828	126628	188929	27509
Total mosaic reads #	476	10630	1485	3946	36758	10913	5307	844
Total mosaic reads %	0.87	4.5	2.44	3.29	7.18	8.62	2.81	3.07
	2.68 ± 2.57 %		*2.86 ± 0.42 %*		*7.90 ± 1.02 %*		*2.94 ± 0.18 %*	
3 pieces #	428	9062	943	3823	18717	7585	4297	328
3 pieces %	0.78	3.84	1.55	3.19	3.66	5.99	2.27	1.19
	2.31 ± 2.16		*2.37 ± 0.82*		*4.83 ± 1.65*		*1.73 ± 0.76*	
4 pieces #	47	1325	540	403	17943	2548	965	512
4 pieces %	0.09	0.56	0.89	0.34	3.51	2.01	0.51	1.86
	0.33 ± 0.33		*0.61 ± 0.27*		*2.76 ± 1.06*		*1.18 ± 0.96*	
5 pieces #	1	243	2	27	98	780	45	4
5 pieces %	0.00	0.10	0.00	0.02	0.02	0.62	0.02	0.01
	0.05 ± 0.07		*0.01 ± 0.01*		*0.32 ± 0.42*		*0.02 ± 0.01*	

C.

	siControl		siRAD51		si53BP1		si53BP1+siRAD51	
	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
Junctions from mosaic reads* #	997	22099	3506	8855	91263	22814	11489	2192
microhomologies #	604	12634	2210	5284	57975	14149	7013	1429
microhomologies %	60,58	57,17	63,03	59,67	63,53	62,02	61,04	65,19
	58.87 ± 1.70		*61.35 ± 1.68*		*62.77 ± 0.75*		*63.11 ± 2.07*	
nucleotide insertions #	270	6651	1043	2844	29017	7116	3765	724
nucleotide insertions %	*27,08*	*30,1*	*29,75*	*32,12*	*31,79*	*31,19*	*32,77*	*33,03*
	28.54 ± 1.70		*30.93 ± 1.68*		*31.49 ± 0.75*		*32.90 ± 2.07*	

D.

siControl
n = 290,997 reads
> 16 reads / 1 kb bin

siRAD51
n = 180,674 reads
> 11 reads / 1 kb bin

si53BP1
n = 638,456 reads
> 35 reads / 1 kb bin

si53BP1 + siRAD51#2
n = 216,438 reads
> 12 reads / 1 kb bin

CD4-3200 bp I-SceI substrate

Fig. 2: Linear amplification-mediated high-throughput genome-wide translocation sequencing (LAM-HTGTS) A. Linear amplification-mediated high-throughput genome-wide translocation sequencing (LAM-HTGTS) methodology outline. **B.** Numbers (#) of sequences containing 2 identified (upper table), I-SceI cryptic sites (middle table) or more than two identified pieces (lower table) and their percentages (%) relative to the total number of merged reads in control or 53BP1, RAD51 and 53BP1/RAD51-depleted GC92 cells. The results from two independent experiments are shown. **C.** Microhomology usage and presence of non-identified patches of nucleotides at junctions between identified pieces in mosaic reads containing 3 and 4 pieces. Numbers (#) and percentages (%) for each sample are indicated. The results from two independent experiments are shown. **D.** Circos plots representing interchromosomal junctions in

9

control, RAD51- or 53BP1-depleted and 53BP1/RAD51-depleted GC92 cells. The CD4-3200bp substrate is located on chromosome 8 (blue arrow). The number of total merged reads for two combined experiments and the cutoff of reads for 1 kb bins are indicated. The cutoff of reads is proportional to the total merged reads. The bins with a number of captured reads superior to the cutoff are linked by red or blue lines (red line: CD4-3200bp substrate-human genome joins; blue line: CD4-3200bp substrate-human genome with cryptic I-SceI site joins). The numbers of joins are plotted as black bars on a log scale with the indicated custom axis from 10 to 10000.

Consistent with this, we identified multiple sequences located at or near previously described cryptic I-SceI target sites (Frock et al., 2015), which likely represent direct end-joining events between two I-SceI-induced DSBs (Fig. 2B and Supplementary Fig. S3B). Therefore, two-pieces split reads correspond to rearrangements/translocations occurring by a mix of end-joining, and/or microhomology mediated mechanisms. These events were more frequent in 53BP1-depleted cells but this increase was abolished by co-depletion of RAD51 (Fig. 2B), showing that 1-RAD51 is implicated in some of microhomologies mediated rearrangements and 2- 53BP1 protects against these RAD51-mediated events.

In spite of the relatively short length of the reads (250-450bp) we also detected mosaic reads containing multiple (>2) mappable parts (Fig. 2B and Supplementary Fig. S3B). Most mosaics contained 3 mappable pieces, although four and even five pieces were occasionally recorded at lower frequencies (Fig. 2B and Supplementary Fig. S3B). Consistent with our previous analysis (see Fig. 1), silencing 53BP1 increased the occurrence of mosaics by approximately 3-fold compared to wild-type cells (Fig. 2B). Silencing RAD51 in 53BP1-silenced cells abolished the stimulation of such events (Fig. 2B).

Approximately 60% of the breakpoint junctions in mosaic reads containing 3 and 4 pieces exhibited microhomologies (Fig. 2C), with a median of 3 nucleotides in control

10

cells and up to 6 bp in 53BP1-silenced cells (Supplementary Fig. S4), and nearly 30% of joints contained non-identified nucleotide insertions (Fig. 2C), with a median of 9bp in control cells up to 20bp in 53BP1-silenced cells (Supplementary Fig. S4). These percentages are similar to those obtained in our previous and present experiments with the CD4-3200 bp substrate (Fig. 1, Supplementary data S1 and (Guirouilh-Barbat et al., 2016). Only few junctions contained neither microhomologies nor non-identified nucleotide insertions (Fig. 2C).

A large fraction of mosaic reads contained DNA sequences that mapped to the human genome (Fig. 2D and Supplementary Fig. S5), thus representing the capture of human ectopic chromosomal sequences or translocations. As the cryptic I-SceI sites were not over-represented in mosaic events, such events arise more likely through MMTS rather than direct DSB End-Joining. These mosaic events were stimulated in 53BP1-silenced cells but abolished in 53BP1- and RAD51-silenced cells (Fig. 2D and Supplementary Fig. S5).

These analyses show that RAD51 generates both genomic rearrangements themselves and capture of ectopic sequences at the junctions, in a sequence-homology-independent manner, and that 53BP1 counteracts RAD51-mediated non-homologous rearrangements.

53BP1 protects against RAD51-mediated t(2;5) translocation between non-homologous sequences. As LAM-HTGTS suggested that RAD51 fosters translocations independently of sequence homology, and that this is prevented by 53BP1, we directly addressed the impact of 53BP1 and RAD51 on the occurrence of a sequence homology-independent translocation targeted by CRISPR-Cas9, at another locus and in another cell line. The t(2;5) translocation between ALK and NPM1 sequences occurs in anaplasic

11

large cell lymphoma (ALCL). Such translocation can be reconstructed in cultured cells using specific CRISPR-Cas9 targeting (Piganeau et al., 2013) (Fig. 3A). Neither 53BP1 nor RAD51 silencing (Fig. 3B) significantly affected the efficiency of Cas9-GFP expression (Supplementary Fig. S6A) and the efficiency of cleavage by Cas9 monitored by TIDE analysis of INDELS at both sites (Supplementary Fig. S6B).

A.

B.

C.
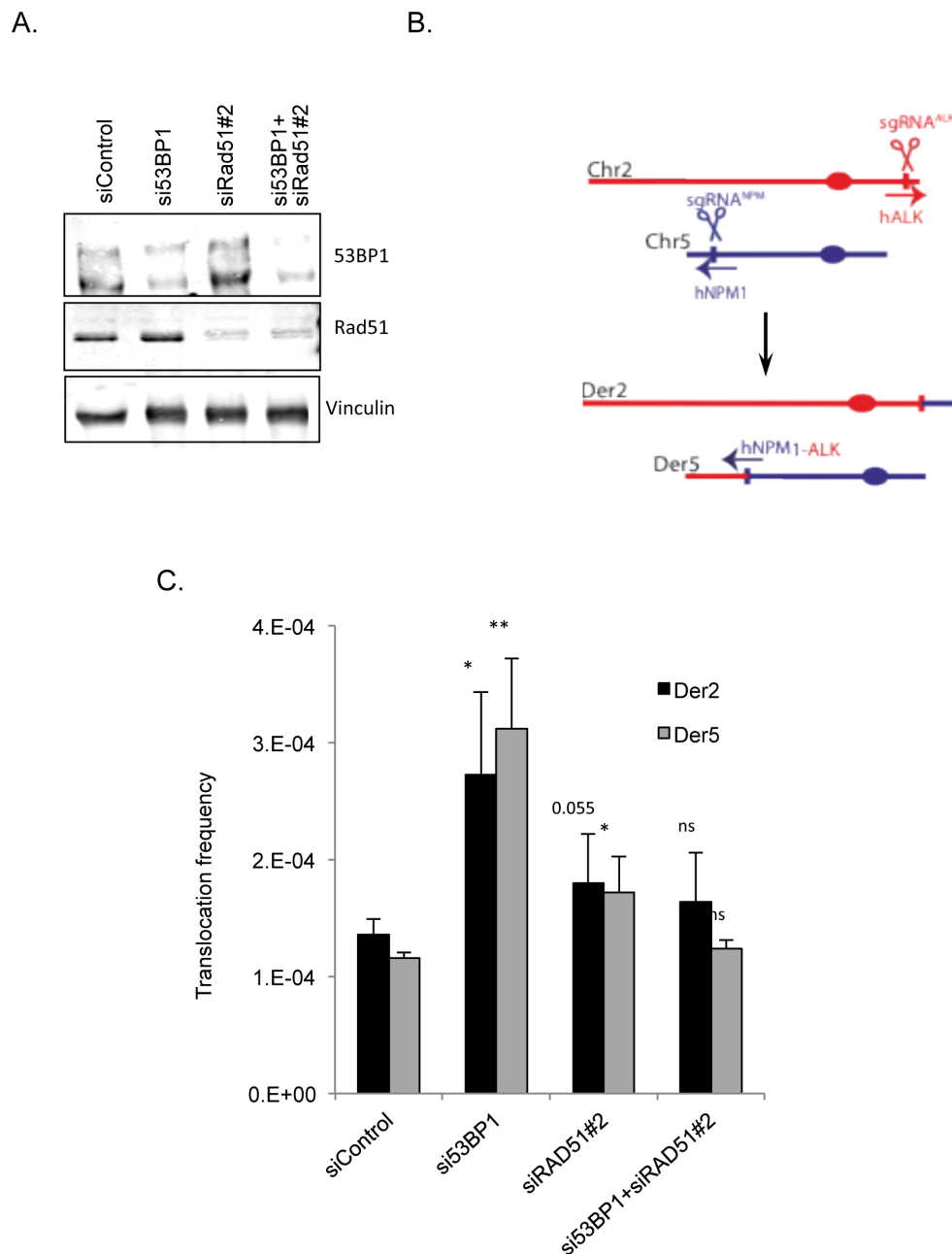
Fig. 3. The t(2;5) translocation between ALK and NPM1 induced by specific CRISPR-Cas9. A. Scheme of the reconstituted t(2;5) translocation with the 2 derivative chromosomes (der2 and der5). **B.** Analysis of 53BP1 and Rad51 expression. **C.**

Frequency of CRISPR-CAS9-induced t(2;5) hALK/hNPM1 translocation. The data correspond to the mean and SEM of 3 independent experiments.

In this novel set up, ablation of 53BP1 also resulted in a strong increase in the t(2;5) translocation frequency (Fig. 3C). Moreover, although no sequence homology was involved, ablation of RAD51 abolished the stimulation of translocation resulting from 53BP1 silencing (Fig. 3C). These data are fully consistent with that obtained with the LAM-HTGTS analysis described above, at a different locus, in a different cell line.

Collectively, our data point RAD51 as a driver of genetic instability, through events that do not involve sequence homology and 53BP1 as a protector against such events.

Capture of ectopic sequences occurs at SV junctions in human breast tumors. Then, we analyzed the junction scars of structural variants (SVs), *i.e.* translocations, duplications, deletions and inversions, in a physiopathological context. We considered the set of SV junctions from the whole genome sequencing (WGS) of 560 breast cancers described previously by S.Nik-Zainal et al (Nik-Zainal et al., 2016). For the basal analysis of junctions in breast tumors, we first excluded the cases with HR deficiency (HRD tumors) in which BRCA1/2 are inactivated or predicted to be inactivated (see M&Ms), in order to avoid any bias associated with HR deficient phenotype. This gave 419 cases corresponding to 47833 SVs, i.e., translocations, duplications, inversions, or deletions, therefore involving distant partners (>1000bp). 63% of SV junctions exhibited micro-homologies (Supplementary data Fig. S7), which is very similar to the frequency observed in the above HTGTS analysis (Fig. 2C). Scar junctions contained sequence insertions in 19% of SVs, including 3% of junctions with insertions of at least 45 bp that we have defined as the cut-off size of ectopic sequences capture (Guirouilh-Barbat et al., 2016) (Fig. 4A); this is in similar range than that observed with the CD4-3200 bp

13

substrate (compare Fig. 4A and 4B). Of note, each different type of SV (namely, duplications, inversions, deletions, and translocations) exhibited similar frequencies of ectopic sequences capture at the scar junctions (Fig. 4A and Supplementary data Fig. S6).

A. SVs in tumors genome, genome alteration >1000 pb
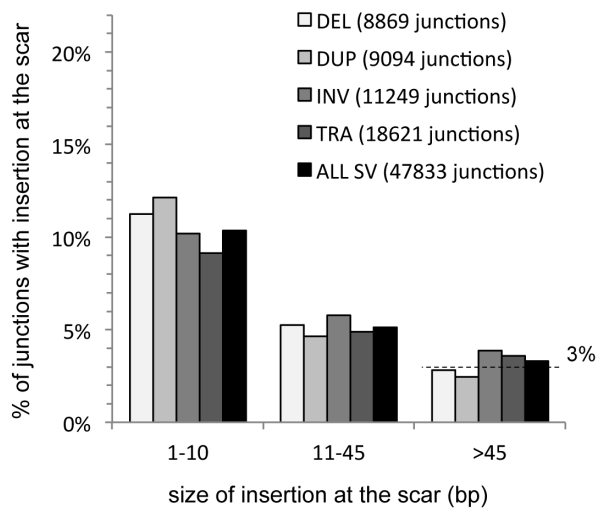
B. Junctions of two I-SceI sites distant of 3,200 bp in GC92 and GC49 cells (substrate CD4-3200bp)
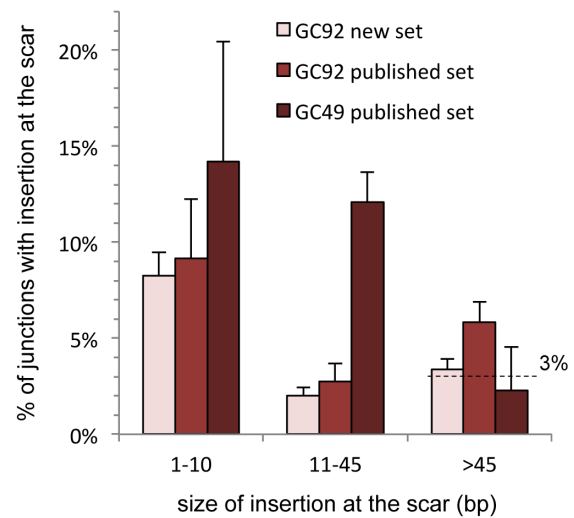
C.
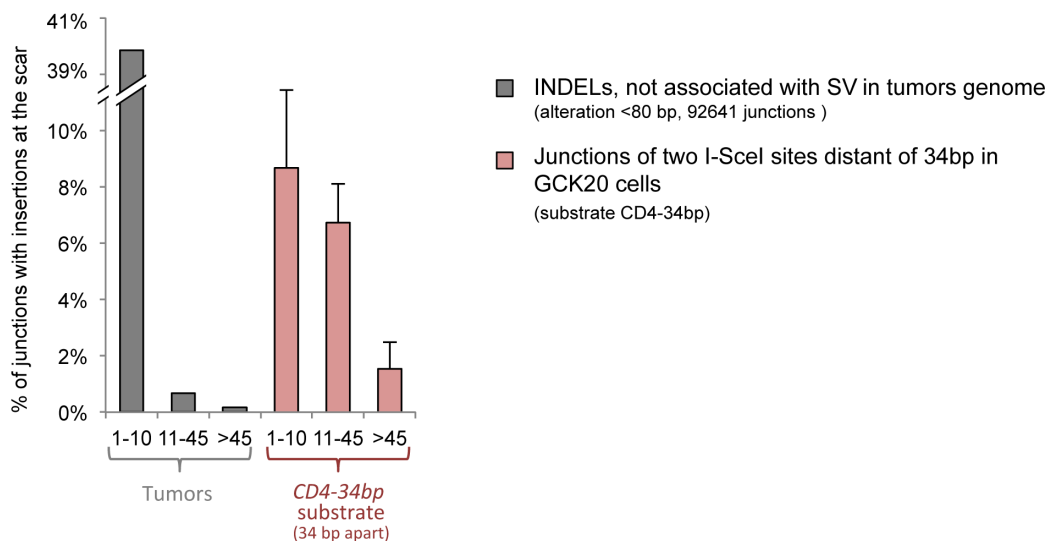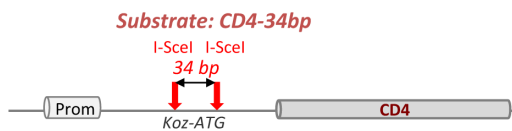
Fig. 4. Capture of ectopic sequences at the junction of SVs in breast tumors. A. Frequency of insertions at the junction scars (total # of analyzed junctions: 47833) of SVs (size of alteration > 1,000 bp) in a panel of 419 non HRD breast tumors. Approximately 20% of junctions contain insertions in all types of SVs. Histograms represent the fraction of scars with insertions of 1-10 bp, 11-45 bp or >45 bp; we propose that the latter arise through MMTS (Guirouilh-Barbat et al., 2016). **B.** Frequency of insertions at the junction scars of two I-SceI breakage sites 3,200 bp apart (substrate CD4-3200 bp, see Fig. 1B). GC92 and GC49 are two different clones bearing the CD4-3200 bp reporter (Guirouilh-Barbat et al., 2016). The new set of data represent the average of 5-7 independent experiments ± SEM. Published data were from (Guirouilh-Barbat et al., 2016). **C.** Gray bars: Frequency of insertions at the junction scars of INDELs not associated with SVs in tumors (size of alteration < 80 bp, total # of analyzed INDELs: 92641). Red bars: Frequency of insertions at the junction scars of two I-SceI breakage sites 34 bp apart in GCK20 cells. The substrate (CD4-34 bp) used is shown in the upper panel (Guirouilh-Barbat et al., 2016). The results represent the average of 5 independent experiments ± SEM that were used in a previous publication (Guirouilh-Barbat et al., 2016).

No long homologous sequences were present at the borders of the ectopic sequences capture events, suggesting that they occur through non-homologous-mediated process(es). Instead, the borders exhibit microhomologies and non-identified sequences, similar to those observed with the CD4-3200 bp intrachromosomal substrate.

We also analyzed the INDELs (small genome alteration < 80bp) not associated with SVs (92641 junctions), therefore that do not involve rearrangements of distant DNA partners. Insertions >45bp are only barely detectable (0.17% of all insertions) (Fig. 4C). This corresponds to the experimental assay with CD4-34 bp intra-chromosomal substrates (a derivative of the CD4-3200 bp where the two DNA ends are separated by only 34 bp) showing that the joining of close double-strand ends does not permit the capture of ectopic sequences (Fig. 4C and (Guirouilh-Barbat et al., 2016)). Taken together, the data confirm that capture of ectopic sequences is favored at junctions of distant partners, in non-HRD breast tumors.

Captures of non-homologous ectopic sequences at SV junctions are decreased in BRCA2-deficient tumors. The cohort of breast tumors also contains 28 cases with BRCA2 inactivation due to germline or somatic deleterious mutations and loss of the second allele (Nik-Zainal et al., 2016). Thus, we compared BRCA2 deficient tumors to the control group (high grade HER2-negative cases without other HR deficiency) for the frequency of non-homologous sequence capture at the SVs junctions (Fig. 5).
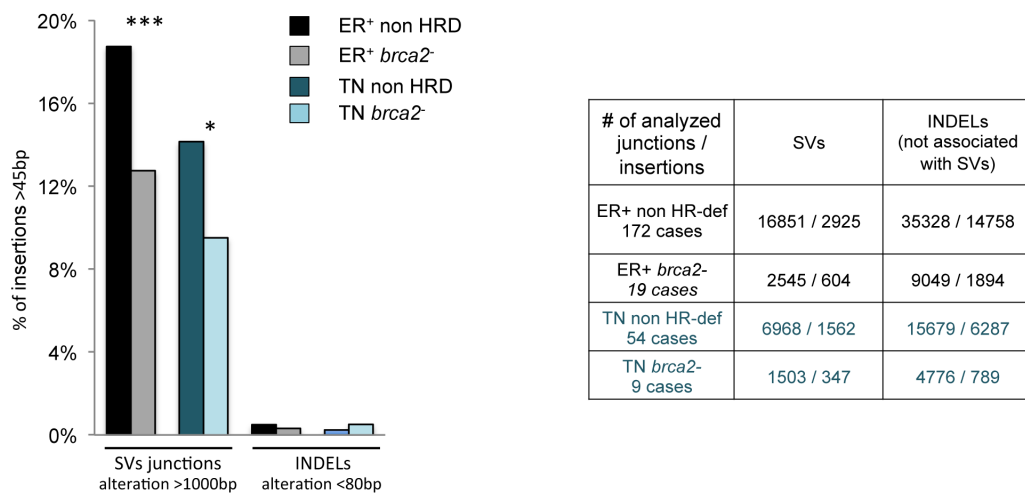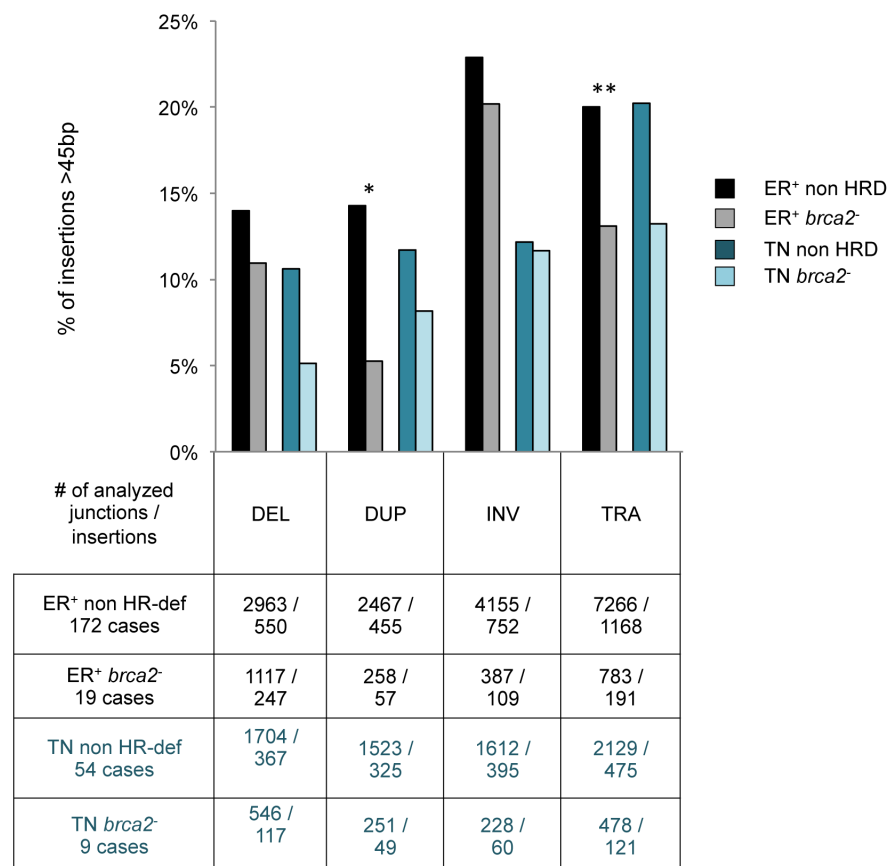
Fig. 5. Capture of ectopic sequences at SV scars in human non HR-deficient (non HR-def) *versus BRCA2***-deficient breast tumors. A.** Ectopic sequences capture in total SVs versus INDELs in non-SV events. Left four bars: SVs, involving distant DNA. . Right four bars: ectopic sequences capture at INDELs not associated with SVs, i.e. that do not involve distant DNA. **B.** Ectopic sequences capture in each kind of SVs. The bars represent the percentage of insertions that are >45 bp among all insertions. The number of events analyzed is shown in the tables. ER+: estrogen receptor positive tumors. TN: triple negative tumors.

17

Capture of ectopic sequences at all types of SV junctions is significantly decreased in *BRCA2*-deficient tumors, independently of their estrogen receptor or triple negative status (Fig. 5A and 5B). Moreover, consistent with the data above, the frequency of ectopic sequences capture in INDELs, which do not involve distant DNA, does not decrease in *BRCA2*-deficient cells (Fig. 5A and 5B). These data suggest that in tumors, capture of ectopic sequences associated with SVs (thus involving distant DNA) requires functional BRCA2, although this does not implicate sequence homologies.

The analysis in tumors confirms the implication of BRCA2 in the capture of ectopic sequences at junctions of SVs devoid of sequence homology. This finding fully supports the involvement of the HR players in genetic instability in a sequence homology-independent manner and in a physiopathological context.

Discussion

HR plays an essential role in genome stability maintenance through processes driven strictly by sequence homology (inspiring the process its name). In contrast to these widely admitted dogmas, the data herein highlight the double face of the pivotal HR actor(s) RAD51 and BRCA2 that intrinsically possess the capacity to foster genetic instability, such as translocations and ectopic sequences capture at SV junctions. In addition, we show here, that 53BP1 protects against RAD51-mediated non-homologous genetic instability. In *Saccharomyces cerevisiae*, HR has been shown to promote genome rearrangements and insertions, but via a process that is strictly dependent on sequence homology (Piazza et al., 2017). Complex chromosomes rearrangements generated by template switch with microhomologies at the junctions were described in yeast, after initiation of BIR and gene conversion, but were found to be independent of RAD51 (Anand et al., 2014; Tsaponina and Haber, 2014). Microhomology-mediated duplications in budding yeast were also found to be independent of HR and EJ machineries (Payen et al., 2008). Here we show that, in human cells, RAD51 can promote both ectopic sequences capture at SVs junctions and SVs themselves, although they do not involve sequence homology, or use only microhomologies far below the MEPS.

We observed these RAD51-mediated non-homologous events in different experimental setups and in the physiopathological context of Brest tumors. Therefore, RAD51 plays contradictory roles: on the one hand, RAD51 maintains genome stability, catalyzing HR but on the other hand, RAD51 can also generate genetic instability in a sequence homology-independent manner. Such genetic instability might account for the poor prognosis associated with RAD51 overexpression in breast, pancreatic, lung, head and neck cancers and colorectal carcinomas (Tennstedt et al., 2013) (and ref in).

19

Capture of ectopic sequences involves either microhomologies or non-identified patches of nucleotides. Microhomologies-mediated insertions could occur from different processes, i.e. alternative end-joining (A-EJ) or MMTS. Since ablation of HR stimulates A-EJ (So et al. submitted) but decreases microhomologies-mediated insertions (present data), this suggests that microhomologies-mediated insertions occur by MMTS rather than A-EJ.

Because rearrangements and insertions described here involve microhomologies far beyond the MEPS, they cannot be classified as HR events, although they are RAD51-dependent. Instead we propose a novel class of events: non-homologous recombination (nonHR) or micro-homologous recombination (μHR). To reconcile our data with the decades-old knowledge of HR, we propose the model shown in Fig. 6 where nonHR/μHR results from the different attempts of RAD51 to search for homology. Indeed it has been shown *in vitro* that the ssDNA/RAD51 filament scans the duplex DNA for homology through 8-bp blocks (Qi et al., 2015). As in our experiments the median length of microhomologies was from 3 to 6 bp, one can propose that, in living cells, the chromatin context might stabilize, at least transiently, pairing with microhomologies smaller than 8 bp. If a long homologous sequence reaching the MEPS is found, the intermediate is stabilized, triggering canonical HR. Below the MEPS, the intermediate should be rejected, and the ssDNA/RAD51 should engage another round of homology searches. However, as proposed for MMTS (Guirouilh-Barbat et al., 2016), DNA synthesis can be primed by the 3' invading strand before rejection; therefore, the rejected strand should contain a partial copy of the invaded molecule. Then, this ssDNA can anneal with the other double-strand ends of the recipient molecule, resulting in ectopic sequences capture. Several rounds of homology searches would account for both the observed mosaic sequence capture and the unidentified sequences at the junctions. One can

20

assume that canonical HR should be more stable biochemically and efficient than nonHR/μHR, but this should be largely compensated by the fact that few bp microhomologies are much more frequent than long homologous sequences. In addition, RAD51 increased chromosomal mobility after DNA damage in yeast (Smith et al., 2018). It is thus tempting to speculate that mammalian RAD51 could also do this, favoring inter chromosomal exchange that leads to SVs and capture of ectopic sequences.
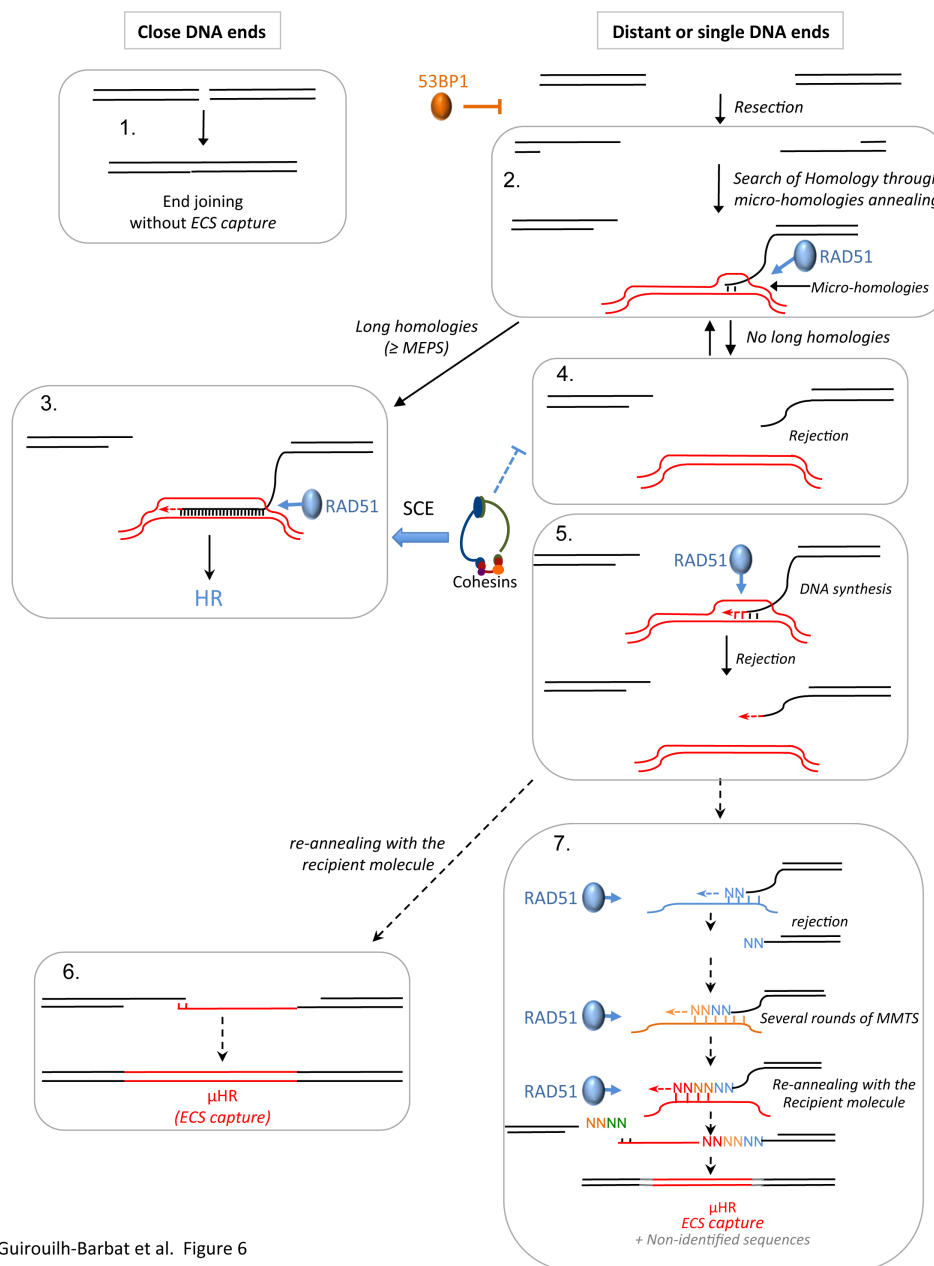
Guirouilh-Barbat et al. Figure 6

Fig. 6. Model for the role of the HR protein RAD51 in non-homologous genetic instability through MMTS. 1/ When two DNA double strand ends are close, insertions

(>45bp) were barely recorded (Guirouilh-Barbat et al., 2016). When DNA double-strand ends are distant (few kb), resection generates a 3' single strand end, which can be covered by RAD51 after loading by BRCA2. One can speculate that similar processes might occur on single-ended DSBs. **2/** The ssDNA/RAD51 filament then searches for homology, starts the scanning for microhomologies, which leads to ectopic matrix invasion. **3/** When a long homology is detected, the intermediate is stabilized, and synthesis occurs for canonical HR repair (left panel). **4/** If no long homology is found, the invading molecule is rejected and re-engaged in another round of homology search (2/). **5/** On some occasions, DNA synthesis can be initiated, as proposed for MMTS. However, the generated intermediate would be very unstable and thus rejected with the copied sequence. **6/** This intermediate re-anneals eventually with the recipient molecule through microhomologies. The resulting product is the insertion of a sequence bordered by microhomologies. **7/** When successive attempts to initiate synthesis by MMTS have occurred, the resulting product contains several sequences (mosaic) and/or the inserted sequences is bordered by unidentified patches of nucleotides. 53BP1 by impairing resection protects against these events. Moreover the cohesion complex represses the capture of ectopic sequences by i) favoring the interaction with the sister chromatids (which bears long identity) which favors sister chromatid exchanges (SCE); ii) by maintaining the two sister chromatids linked limiting the mobility of the broken ends and thus its capacity to interact with an ectopic sequence (Gelot et al., 2016).

One can expect that cells should have developed defense system(s) against RAD51-driven nonHR/μHR. First, we show here that 53BP1 counteracts nonHR/μHR. Recently, Ochs et al. showed that 53BP1 channels DSB repair toward RAD51 dependent HR by preventing hyper-resection (Ochs et al., 2016). Here we show that 53BP1 inhibits microhomology-mediated rearrangements by RAD51, which is an additional way of channeling RAD51 toward faithful HR. However, in breast tumor analysis, all tumors are wild type for 53BP1. This suggests that in a context of high genetic instability such as tumor cells, the protective effect of 53BP1 can be bypassed, at least for the capture of ectopic sequences at SV junctions that involve distant DNA partners. Second, we propose that the cohesion complex also protects against nonHR/μHR. Indeed, by maintaining the sister chromatids' embrace, the cohesion complex should i) limit the mobility of the DNA ends and thus rearrangements that involve distant sequences; ii) channel the search of

homology to the sister chromatid, which shares mega-bases of identity and thus favors sister chromatid exchange by HR. Remarkably, the ablation of the cohesion complex favors the capture of ectopic sequences and genome rearrangements (Gelot et al., 2016). Nevertheless, because nonHR/µHR does not need extended sequence homologies and requires only one DSB (in contrast to rearrangements promoted by end-joining), it might strongly jeopardize genome stability. In addition, because nonHR/µHR can generate mosaic events, it might be involved in complex rearrangements.

In the course of genome analyses, attempts are made to classify the mechanistic origins of the rearrangements in tumors based on the signature of the junctions (Weckselblatt and Rudd, 2015). These classifications should thus be re-evaluated considering the present data.

More generally, as a driver of genome plasticity, RAD51-mediated non-homologous recombination might play important role in and molecular evolution of the genome, in addition to fueling tumorigenesis.

Methods

Cells. CG92 and GC49 cell lines were derived from SV40-transformed GM639 human fibroblasts and contained the CD4-3200bp (pCOH-CD4) end-joining reporter (Rass et al., 2009). TERT-RPE-1 cells were cultured in DMEM-F12 (Gibco 10565-018) supplemented with 10% FBS (Gibco 10270106).

Transfection. The meganuclease I-SceI was expressed by transient transfection of the expression plasmid pCMV-HA-I-SceI (Liang et al., 1998) with Jet-PEI following the manufacturer's instructions (Polyplus Transfection). The expression of HA-tagged I-SceI was verified by Western blotting. For the silencing experiments, 50,000 cells were seeded 1 day before transfection, which was carried out using INTERFERin following the manufacturer's instructions (Polyplus Transfection) and 20 nmol siRNA: si53BP1 (onTarget plus SMARTpool for human TP53BP1) was purshased by Dharmacon (Chicago, IL, USA) and Control (5'- AUGAACGUGAAUUGCUCAA -3'), siRAD51#1 (cat# L003530-00-0010, Dharmacon), siRAD51#2 (3'UTR: 5'-GUGCUGCAGCCUAAUGAGA-3'), siBRCA2#1 (5'-GCUGAUCUUCAUGUCAUAA-3'), and siBRCA2#2 (5'-AACUGAGCAAGCCUCAGUCAA -3') were all synthesized by Eurofins. Forty-eight hours later, the cells were transfected with the pCMV-HA-I-SceI expression plasmid.

Analysis of DSB repair at the CD4-3200 bp intrachromosomal reporter. After transfection with the pCMV-HA-I-SceI plasmid and incubation for 72 h, the cells were pelleted and resuspended in PBS. After incubation for 1 h at 56°C with 2 mg/ml proteinase K, the proteinase K was inactivated at 95°C for 15 minutes, and genomic DNA

24

was used for the PCR amplification of the junction using the primers CMV-6 (5'-TGGTGATGCGGTTTTGGC-3') and CD4-int (5'-GCTGCCCCAGAATCTTCCTCT-3'). The predicted size of the PCR product was 732 nt. The PCR products were cloned with a TOPO PCR cloning kit (Invitrogen Life Technologies) and sequenced (Eurofins). For each sample, 2 to 7 experiments were analyzed. In each of these experiments, HA-I-SceI expression and silencing efficiency were verified by Western blot.

Western blot analysis. Cells were lysed in buffer containing 20 mM Tris-HCl (pH 7.5), 1 mM Na_2EDTA, 1 mM EGTA, 150 mM NaCl, 1% sodium deoxycholate, 1% (w/v) NP40, 2.5 sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM NA_3VO_4 and 1 μg/ml leupeptin supplemented with a complete mini protease inhibitor (Roche). Proteins (30–40 μg) were denatured, electrophoresed on 9% SDS-PAGE gels, transferred onto nitrocellulose membranes and probed with the following specific antibodies: anti-HA (MMS-101R, Covance), anti-53BP1 (#4937, Cell Signaling), anti-RAD51 (PC130, Millipore), and anti-vinculin (Abcam). BRCA2 depletion was checked on 4-12% Tris acetate gels according to the manufacturer's instructions and membranes were probed with anti-BRCA2 (#ab9143, Abcam). Immunoreactivity was visualized using an enhanced chemiluminescence detection kit (ECL, Pierce).

Linear amplification-mediated high-throughput genome-wide translocation sequencing (LAM-HTGTS). HTGTS was performed according to a published protocol (Hu et al., 2016) and is shown in Figure 2A. Briefly, DNA was purified and sonicated (Bioruptor, Diagenode) into 500–1,000 bp fragments. LAM-PCR was performed using bait primer coupled with biotin, 1U Phusion polymerase (Thermo Scientific) and the following cycles: 1x (98 °C for 120 s); 80x (95 °C for 30 s; 58 °C for 30 s; 72 °C for 90 s);

and 1x (72 °C for 120 s). Biotinylated PCR fragments were incubated with MyOne streptavidin C1 beads (Invitrogen) and rotated for 4 h in 1 M NaCl and 5 mM EDTA buffer at room temperature. After washes with B&W buffer (1 M NaCl; 5 mM Tris-HCl, 0.5 mM EDTA), on-bead ligation was performed using 2.5 mM bridge adaptor, 1 mM hexamine cobalt chloride (Sigma), 15 U T4 DNA ligase (Promega), 15% PEG-8000 (Sigma) and the following cycles: 25 °C for 1 h; 22°C for 2 h; and 16°C O/N. After washing three times with B&W buffer (1 M NaCl, 5 mM Tris-HCl (pH 7.4) and 0.5 mM EDTA (pH 8.0)), the on-bead ligated products were subjected to nested PCR using Phusion polymerase, locus-specific and adaptor primers and the following cycles: 1x (95 °C for 300 s); 15x (95 °C for 60 s; 60 °C for 30 s); and 1x (72 °C 600 s). Blocking digestion was then performed with *I-SceI* for 1 h to remove uncut germline DNA. After purification using a Qiagen column, recovered DNA was PCR-amplified using Illumina primers, Phusion polymerase and the following cycles: 1x (95°C for 180 s); 10x (95°C for 30 s; 60°C for 30 s; 72°C for 60 s); and 1x (72°C for 360 s). The tagged PCR products were size-selected for DNA fragments between 500–1000 bp on a 1% agarose gel and purified using a Qiagen column before loading onto an Illumina Miseq machine for paired-end 2X250 bp sequencing.

Biotin-CMV6	Biotin-TGG TGA TGC GGT TTT GGC
Adapter-upper-6N	GCG ACT ATA GGG CAC GCG TGG NNNNNN -NH2
Adapter-lower-NH2	/5Phos/CCA CGC GTG CCC TAT AGT CGC-NH2
P5-I5	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T
P7-I7	CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT C
I5-Red (CMV3)	TA CAC GAC GCT CTT CCG ATC T(6nts barcode)GT ACG GTG GGA GGT CTA TA
I7-Blue	CTC GGC ATT CCT GCT GAA CCG CTC TTC CGA TCT GAC TAT AGG GCA CGC GTG G

26

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals and Recombinant Proteins		
Dynabeads MyOne Streptavidin C1	Thermo Fisher Scientific	Cat#65001
Phusion High-Fidelity DNA Polymerase	NEB	Cat#M0530L
T4 DNA ligase	Promega	Cat#M1804
Proteinase K	Eurobio	Cat#GEXPRK01-B5
RNase A	Invitrogen	Cat#8003089
Taq polymerase	Invitrogen	Cat#18038042
MgCl2	Invitrogen	Cat#18038042
Hexammine cobalt (III) chloride	Sigma Life Sciences	Cat#H7891
PEG8000	Sigma Life Sciences	Cat#P2139
I-SceI	NEB	Cat#R0694S
Critical Commercial Assays		
MiSeq Reagent Kit v2 (500 cycle)	Illumina	Cat#MS-102-2003
QIAquick Gel Extraction Kit	QIAGEN	Cat#28704
Software and Algorithms		
ea-utils	Erik Aronesty., 2013	https://expressionanalysis.github.io/ea-utils/
BWA	Li H, et al., 2010	http://bio-bwa.sourceforge.net/
Circos	Krzywinski M et al., 2009	http://circos.ca/
FLASH	T .Magoc et al., 2011	https://ccb.jhu.edu/software/FLASH/

Data analysis for HTGTS. Sequencing reads in FASTQ files were parallel decoded and adaptor sequence trimmed out by US-utility tool (Aronesty, 2013). Paired-end reads with less than 20-nucleotide bait sequence at the 5' end of read 1 were filtered out. The remaining paired-end reads were merged by a _FLASH_ tool (Magoč and Salzberg, 2011) with a minimum of 10 bp overlap. The merged reads were mapped to the human hg38 reference genome combined with plasmid sequence using BWA (Li and Durbin, 2010) and the default setting. To identify different types of joins, including germlines, deletions, multiple copies and insertions, we first divided the plasmid sequence into

27

several parts, including a blue part, which started with the CMV3 primer and continued until the first I-SceI cut site. The H2Kd part was defined between the first and second I-SceI cut sites. The CD4 part was defined from the second I-SceI cut site to 500 bp downstream, and all rest of the plasmid was defined as the plasmid backbone. According to these definitions, we parsed the mapping of the merged reads. Reads without split mapping and with or without short indels in the first I-SceI cut site were defined as germline. If reads were mapped into two parts in which the 5' end was blue and CD4 was at the 3' end, these reads were defined as deletional joins. To identify multiple copies and insertions, the reads must be mapped separately to at least three parts. The reads include the blue part at the 5' end and the CD4 part at the 3' end. If one or more pieces of sequence originate from other defined parts but not from the close region (500 bp) of the H2Kd part around the I-SceI cut sites, then these reads are counted as multiple copies and insertions. If the reads included a blue part at the 5' end and 2 or more pieces of human chromosome sequences, they were also counted as a multiple copy and insertion. All analysis steps above were fulfilled by costumed Perl scripts developed in-house. Circos (Krzywinski et al., 2009) was used to visualize multiple copies and insertions with human chromosome sequences in each 1 kbp bin.

CRISPR-Cas9 translocation frequency. RPE-1 cells were transfected with 20 pmol siRNA by nucleofection (Amaxa Lonza 4D program EA-104 P3 solution) or Lipofectamine RNAiMAX transfection reagent (ThermoFisher) according to the manufacturer's instructions. Thirty-six hours after siRNA transfection, 10^6 cells were transfected a second time by nucleofection with pCas9-2A-GFP, sgRNA^{NPM1} and sgRNA^{ALK} (3 μg each) (Amaxa Lonza 4D program EA-104 P3 solution). siRNA inhibition was

analyzed by Western blot at the time of the second transfection. Transfection efficiency was assessed at 48 h by FACS. First-round PCR was performed on lysates with the primers used for translocation junction detection (listed below). Second-round PCR was performed with SYBR Green and real-time PCR conditions (Mx3005p, Agilent) (Brunet et al., 2009). PCR sequencing was performed on positive wells using forward primers and a BigDye™ Terminator v3.1 Cycle Sequencing Kit according to the manufacturer's instructions. The sequences were read on an Applied Biosystems™ 3500 Genetic Analyser. At D7, genomic DNA was also extracted from the pool of transfected cells using an E.Z.N.A Tissue DNA Kit (Omega Biotek). Translocation frequencies were estimated by serial dilution as previously described (Renouf et al., 2014).

sgRNA guide sequences:

sgRNA^{NPM1} TATATCCTCGAACTGCTACT

sgRNA^{ALK} GTCGGTCCATTGCATAGAGG

Primers sequences :

hNPM1-ALK (PCR1) hNA-Der5-F   ACCAAGTCCAGTTGCGTTTC

hNA-Der5-R   TTCCCTAACCACTGCCACTC

hNPM1-ALK (PCR2) hNA-Der5-NF   GAACTCCTGGCCTTAACGTG

hNA-Der5-NR   CCACCCTCTAGGGTTGTCAAT

hALK-NPM1 (PCR1) hAN-Der2-F   TCCTTCAGTGTCCATCACGA

hAN-Der2-R   GAACCTTGCTACCACCTCCA

hALK-NPM1 (PCR2) hAN-Der2-NF   GAGACATGCCCAGGACAGAT

hAN-Der2-NR   AGAGCACATGGGAAAAGGAA

Whole genome sequence analysis of breast tumors. We used the data from the Nik-Zainal et al. study (Nik-Zainal et al., 2016) supplementary tables describing 1) clinical annotation 2) structural variation and 3) SNVs and indels found in 560 whole genome sequenced breast

cancers. Triple negative, ER-positive HER2-negative high grade ductal carcinoma cases were selected according to clinical annotation. The cases were annotated as HR deficient according to the HRDetect signature (Davies et al., 2017). Cohort of HR proficient cases consisted of 429 cases. The cases were annotated as BRCA2 deficient if mutated in BRCA2 and classified as HR deficient (29 cases, including 19 ER positive and 9 triple negative cases). Non BRCA2 and HR deficient cases were excluded from the analysis. Control group of non-BRCA2 cases consisted of 110 ER positive and 54 triple negative subtypes.

For each category, we analyzed the length of insertions at the SV or INDEL scars.

Accession number

The raw sequencing data have been uploaded to GEO and the accession number is GSE126546

Acknowledgments

We are very grateful to Frederick Alt and colleagues for help and advices with setting up the LAM-HTGTS technique

Author contributions

Conceptualization: JG-B, BSL; Data curation: WY, TP; Funding acquisition: BSL, LD; Investigation: JG-B, WY, LB,EYG, AS, CL, EB; Methodology: JG-B, YW, TP, CL,MHS, EB, LD, BSL; Project administration: BSL; Resources: MHS, EB, LD, BSL; Supervision: JG-B, TP, EB, LD, BSL; Validation: JG-B, BSL; Visualization: JG-B;, WY, EB, LD, BSL; Writing – original draft: JG-B, BSL; Writing – review & editing: JG-B, WY, TP, CL, EB, LD, BSL.

Declaration of Interests.

The authors declare no competing interests

References

Anand, R.P., Tsaponina, O., Greenwell, P.W., Lee, C.-S., Du, W., Petes, T.D., and Haber, J.E. (2014). Chromosome rearrangements via template switching between diverged repeated sequences. Genes Dev. *28*, 2394–2406.

Aronesty, E. (2013). Comparison of Sequencing Utility Programs. Open Bioinforma. J. *7*, 1–8.

Bollag, R.J., Waldman, A.S., and Liskay, R.M. (1989). Homologous recombination in mammalian cells. Annu Rev Genet *23*, 199–225.

Davies, H., Glodzik, D., Morganella, S., Yates, L.R., Staaf, J., Zou, X., Ramakrishna, M., Martin, S., Boyault, S., Sieuwerts, A.M., et al. (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat. Med. *23*, 517–525.

Deriano, L., and Roth, D.B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu. Rev. Genet. *47*, 433–455.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.-J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. *33*, 179–186.

Garinis, G.A., van der Horst, G.T., Vijg, J., and Hoeijmakers, J.H. (2008). DNA damage and ageing: new-age ideas for an age-old problem. Nat Cell Biol *10*, 1241–1247.

Gelot, C., Guirouilh-Barbat, J., Le Guen, T., Dardillac, E., Chailleux, C., Canitrot, Y., and Lopez, B.S. (2016). The Cohesin Complex Prevents the End Joining of Distant DNA Double-Strand Ends. Mol. Cell *61*, 15–26.

Grabarz, A., Guirouilh-Barbat, J., Barascu, A., Pennarun, G., Genet, D., Rass, E., Germann, S.M., Bertrand, P., Hickson, I.D., and Lopez, B.S. (2013). A Role for BLM in Double-Strand Break Repair Pathway Choice: Prevention of CtIP/Mre11-Mediated Alternative Nonhomologous End-Joining. Cell Rep. *5*.

Guirouilh-Barbat, J., Huck, S., Bertrand, P., Pirzio, L., Desmaze, C., Sabatier, L., and Lopez, B.S. (2004). Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. Mol. Cell *14*, 611–623.

Guirouilh-Barbat, J., Rass, E., Plo, I., Bertrand, P., and Lopez, B.S. (2007). Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. Proc. Natl. Acad. Sci. U. S. A. *104*.

Guirouilh-Barbat, J., Gelot, C., Xie, A., Dardillac, E., Scully, R., and Lopez, B.S. (2016). 53BP1 Protects against CtIP-Dependent Capture of Ectopic Chromosomal Sequences at the Junction of Distant Double-Strand Breaks. PLOS Genet. *12*, e1006230.

Haber, J.E. (2014). Genome stability. DNA repair and recombination.

Hoeijmakers, J.H. (2009). DNA damage, aging, and cancer. N Engl J Med *361*, 1475–1485.

Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., and Frock, R.L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-

mediated high-throughput genome-wide translocation sequencing. Nat. Protoc. *11*, 853–871.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Lans, H., and Hoeijmakers, J.H. (2006). Cell biology: ageing nucleus gets out of shape. Nature *440*, 32–34.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Liang, F., Han, M., Romanienko, P.J., and Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci U S A *95*, 5172–5177.

Liskay, R.M., Letsou, A., and Stachelek, J.L. (1987). Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics *115*, 161–167.

Lopez, B.S., Corteggiani, E., Bertrand-Mercat, P., and Coppey, J. (1992). Directional recombination is initiated at a double strand break in human nuclear extracts. Nucleic Acids Res *20*, 501–506.

Magoč, T., and Salzberg, S.L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability--an evolving hallmark of cancer. Nat Rev Mol Cell Biol *11*, 220–228.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

Ochs, F., Somyajit, K., Altmeyer, M., Rask, M.B., Lukas, J., and Lukas, C. (2016). 53BP1 fosters fidelity of homology-directed DNA repair. Nat. Struct. Mol. Biol. *23*, 714–721.

Payen, C., Koszul, R., Dujon, B., and Fischer, G. (2008). Segmental Duplications Arise from Pol32-Dependent Repair of Broken Forks through Two Alternative Replication-Based Mechanisms. PLoS Genet. *4*, e1000175.

Piazza, A., Wright, W.D., and Heyer, W.D. (2017). Multi-invasions Are Recombination Byproducts that Induce Chromosomal Rearrangements. Cell *170*, 760–773.e15.

Piganeau, M., Ghezraoui, H., De Cian, A., Guittat, L., Tomishima, M., Perrouault, L., René, O., Katibah, G.E., Zhang, L., Holmes, M.C., et al. (2013). Cancer translocations in human cells induced by zinc finger and TALE nucleases. Genome Res. *23*, 1182–1193.

Qi, Z., Redding, S., Lee, J.Y., Gibb, B., Kwon, Y., Niu, H., Gaines, W.A., Sung, P., and Greene, E.C. (2015). DNA sequence alignment by microhomology sampling during homologous recombination. Cell *160*, 856–869.

Rass, E., Grabarz, A., Plo, I., Gautier, J., Bertrand, P., and Lopez, B.S. (2009). Role of Mre11

in chromosomal nonhomologous end joining in mammalian cells. Nat. Struct. Mol. Biol. *16*, 819–824.

Rubnitz, J., and Subramani, S. (1984). The minimum amount of homology required for homologous recombination in mammalian cells. Mol Cell Biol *4*, 2253–2258.

Smith, M.J., Bryant, E.E., and Rothstein, R. (2018). Increased chromosomal mobility after DNA damage is controlled by interactions between the recombination machinery and the checkpoint. Genes Dev. *32*, 1242–1251.

So, A., Le Guen, T., Lopez, B.S., and Guirouilh-Barbat, J. (2017). Genomic rearrangements induced by unscheduled DNA double strand breaks in somatic mammalian cells. FEBS J.

Tennstedt, P., Fresow, R., Simon, R., Marx, A., Terracciano, L., Petersen, C., Sauter, G., Dikomey, E., and Borgmann, K. (2013). RAD51 overexpression is a negative prognostic marker for colorectal adenocarcinoma. Int. J. Cancer *132*, 2118–2126.

Tsaponina, O., and Haber, J.E. (2014). Frequent Interchromosomal Template Switches during Gene Conversion in S. cerevisiae. Mol. Cell *55*, 615–625.

Weckselblatt, B., and Rudd, M.K. (2015). Human Structural Variation: Mechanisms of Chromosome Rearrangements. Trends Genet. *31*, 587–599.

Supplemental data

35

RAD51 promotes non-homologous genetic rearrangements that are prevented by 53BP1

Josée Guirouilh-Barbat, Wei Yu, Loelia Babin, Elisa Yaniz Galende, Tatiana Popova,

Ayeong So, Chloé Lescale, Marc-Henri Stern, Erika Brunet, Ludovic Deriano, and Bernard

S. Lopez.

Supplemental data S1

Fig. S1. - Sequence of the junction scars at the I-SceI site of the CD4-3200bp substrates.

S1A. Sequences of captured ECS at junction scars of the two I-SceI sites in GC92 cells bearing the CD4-3200bp reporter after 53BP1, RAD51, and/or BRCA2 siRNA transfection. Insertions were blasted using the BLAST program of the National Centre of Biotechnology Information, National Institutes of Health, Bethesda MD, USA). Insertions were blasted to the End Joining reporter, the I-SceI expression plasmid, the mitochondrion genome, the human genome (*Homo sapiens*, taxid: 9606), Human ALU repeat elements and the Nucleotide collection, using megablast and discontinuous megablast. Sequences identified as ''unmappable stretches of nucleotides'' were identified by neither of these searches.

In red : microhomologies ; in yellow: unmappable stretches of nucleotides ; lower case : sequence of the CD4-3200bp reporter.

S1A. Sequences of insertions

in red : microhomologies

in yellow: non template

lower case : sequence of the CD4-3200bp reporter

siCTRL

155nt

1-154: Homo sapiens chromosome 16, alternate assembly CHM1_1.1 3221396 to 3221549

tttggcCCAACAGGATTTGACCCTGAGGCCCACTCTCACCCTAATCATAACCGCAAAACCACCAG CGCCTGGAGAGAGAGTGAGAGAGAAACAGAAACGGAGCGAGTGTGTGTCGCTGTGATGCCCT TTGCCGTCGCTGCTCATCCCCAGTGACCTCCTGA

395 nt

1-395: Sequence répétée non Alu Homo sapiens chromosome 12, alternate assembly CHM1_1.1 48768696 to 48769089

tggtgatgcggttttggcGTGGATGATCTTTTTGTTGACGTTGATGCTATTCCTGTTTGTTGGTTTTC CTTCTAACAGTCAGGCCCCTCAGCTGCAGGTCTGTTGGAGTTTGCTGGAAGTCCACTCCAGAC TCTACTTGCCTGGGTATCACCAACGGAGGCTGCAGAACACCAAATATTGCTGCCTGATCCTTC

CTTCCTCTAGAAGCTTCATCACAGAGGGGCACCCACCTGTTTGAGGTGTCTGTCGGCCCCTAC
TGGGAGGTGTCTCCCAGTCAGGCTACACAGGGGTCAAGGACCCGCTTGAGGAGGCAGTCTGTC
CATTTTCGGATCTTGAACGCCATGCTGAGAAACCACTGCTCTCTTCAGAGCTGTCAGATAGG
GACGTTTTAAGTCTGCAGAAGCTGTCTGCTgc

117nt

3-70 CD4-3200bp reporter ; 71-117 CD4-3200bp reporter (different location)

TATCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCA
AAAAGCTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATtatccct

298 nt

16-290: Homo sapiens chromosome 3, alternate assembly CHM1_1.1 12139590 to
12139864

TCACTGCCACCAGAATCAAGCGATTCTCCCGCCTCAGCCTCCTGAATAGCTGGGATTACAGGC
GCCTGCCACCATGCTTGGCTAATTTTTTGTATTTTTAGTAGAGACAGCGTTTCACTATGTTGA
CCAGGCTGGTCTTGAACGCCTGACCTCGTGATCCACCTGCCTCAGCCTCCCAAAGTGCTGGGA
TTACAGGCGTGAGCCACCACGTCTGGCCACAAATTTCTTTAAGAGAACCCATATACTAGGAGC
CTGTAACAACCAGGCTGTCAAAGTGGGGCAGCGGTGCTGATGTTAG

220nt

1-223: Homo sapiens chromosome 7, alternate assembly CHM1_1.1 72739670 to
72739855 ; 1-186:Alu sequence

tggtgatgcggttttggcTCACTGCAACCTCCGTCTCCCAGGTTCAAGCGATTCTCATGCCTCAGACTC
CCGAGTAGCTGAGATTACAGGCGTGCGCCACCATGCCTGGCTAATTTTTGTATTTTTAGTAGA
GACGGTGATTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGATGATCTGCCAAC
CTCGCCCTCATAGGCCTGAGATTTTAAAGCATGCGTGGGAATATATAGTTTAG

164nt

1-160: Homo sapiens chromosome 16, alternate assembly CHM1_1.1 3221396 to
3221556

tttggcCCAACAGGATTTGACCCTGAGGCCCACTCTCACCCTAATCATAACCGCAAAACCACCAG
CGCCTGGAGAGAGAGTGAGAGAGAAACAGAAACGGAGCGAGTGTGTGTCGCTGTGATGCCCT
TTGCCGTCGCTGCTCATCCCCAGTGACCTCCTGAACTTTCAGG

154nt

1-153: Homo sapiens chromosome 16, alternate assembly CHM1_1.1 3221396 to 3221548

tttggcCCAACAGGATTTGACCCTGAGGCCCACTCTCACCCTAATCATAACCGCAAAACCACCAG
CGCCTGGAGAGAGAGTGAGAGAGAAACAGAAACGGAGCGAGTGTGTGTCGCTGTGATGCCCT
TTGCCGTCGCTGCTCATCCCCAGTGACCTCCT<mark>A</mark>

317 nt

4-317 CD4-3200bp reporter

<mark>AAC</mark>TGCTGATGCCCGTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGAT
TTTCAAGTCTCCACCCCATAGACGTCAATGGGAATTTGTTTTGGCACCAAAATCAACGGGACT
TTCCAAAATCTAGTAACAACTCCGCCCCATTGACCCAAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCTTTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAA
TTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCTCTAGAGCAACACGGAAGGAA
TTAatgtgccgag

157 nt

1-157: CD4-3200bp reporter

ggaaggaCAGAAAGCGGGTCAACGGTGAGGCCATGGTGTGCTCTCCCCTAGAGCCCTACGGAAT
TCGAGCTCGCCCGGGGATCTCGAGGTCACCCTGCGGTGTCGTCCATCACAGTTTGCCAGTGAT
ACACATGGGGATCAGCAATCGCGCATAATCGATATTAccctatcc

113 nt

1-110: Homo sapiens chromosome 8, GRCh38.p12 Primary Assembly 136408078 to 136408187

ggaattCCTAAGAGCAGAACTATAGTTTGTCACCTCCACAACTCCAGGGACTGGCACTAGATTG
GCATATTTTTCAGTAAATGTTGAATAAAGGAATGAATCTTACTTCAAAGAGG<mark>TCC</mark>

107 nt

1-99: seq répétée

ctgttatGAAGGCCACAAAGTGGTCCAAATATCCACTTGCAGATTCTACAAAAAGAGTGTTTGAA
AGCTGAACTATGAAAGCAAGGTTCAACTCTGTGAGTAATTC<mark>CTTCCGTG</mark>

39

si53BP1

130nt

1-61: CD4-3200bp reporter; 62-129: CD4-3200bp reporter (different location)

cctgttatTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACA
AGTTAGCTTTTTGCAAAAGCCTAGGCCTCCAAAAAAGCCTCCTCACTACTTCTGGAATAGCTC
AGAGGCCGATccctatctag

130nt

1-61: CD4-3200bp reporter ; 62-129: CD4-3200bp reporter (different location)

ccctgttatTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAAC
AAGTTAGCTTTTTGCAAAAGCCTAGGCCTCCAAAAAAGCCTCCTCACTACTTCTGGAATAGCT
CAGAGGCCGATccctatctag

130nt

1-61: CD4-3200bp reporter ; 62-129: CD4-3200bp reporter (different location)

ctgttatTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAA
GTTAGCTTTTTGCAAAAGCCTAGGCCTCCAAAAAAGCCTCCTCACTACTTCTGGAATAGCTCA
GAGGCCGATccctatctag

254nt

1-222: Homo sapiens chromosome 8, alternate assembly CHM1_1.1 94640757 to 94640978

cacggaaggGAAGAGGGAAAGGACATGGAATGTTTGAGAAGCAAACACACACTTCTCCCCAGAG
TGATTCTGTGCATTAACCAGACACGTCTCCATTCTGCTCCCATGGACCTACTGCTGGTGATTC
CCATTTGCAGTCGCCTAACCTACTACCCTTACGCTGTTCTCTGCTGTGTTCCTCAACCATCAGC
ATTTGGGGCAATTCTTCACAGCACAGGACTGCCCTCATATTTTAAATAAGAAATGTGTGATA
ACATCAAAAG

154nt

1-153: Homo sapiens chromosome 16, alternate assembly CHM1_1.1 3221396 to 3221549

tttggcCCAACAGGATTTGACCCTGAGGCCCACTCTCACCCTAATCATAACCGCAAAACCACAGC
GCCTGGAGAGAGAGTGAGAGAGAAACAGAAACGGAGCGAGTGTGTGTCGCTGTGATGCCCTT
TGCCGTCGCTGCTCATCCCCAGTGACCTCCTG<mark>A</mark>

237 nt

1-140 Homo sapiens chromosome 19, alternate assembly CHM1_1.1 5640605 to
5640746 ; 1-41: Alu sequence; 40-140: Alu sequence; 140-200: Alu sequence ; 40-215:
Homo sapiens chromosome 19, alternate assembly CHM1_1.1 1832221 to 1832391

tttggcTCACTGCAACCTCCACCTCCTGGGTTCAAGCGATTCTCCTGTATTTTTAGTAGAGATGG
GGTTTCACCATGCTGGCCAGGCTGGTCTTGAACTCCTGACCTGGTGATCTGCCCGCCTCGGCC
TCCCAAAGTGCTGGGATTGCCTCAGCCTCCCAAGCAGCTGAGATTACAGGCATGCACCACCAT
GCTCGGCTAATTTTTACAGGCGTGAGCCAC<mark>TGCACCCGGACCACACCTAGCA</mark>

238 nt

1-224 Alu sequence

tttggcTCACTGCAACCTCCACCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGCAGC
TGAGATTACAGGCATGCACCACCATGCTCGGCTAATTTTTGTATTTTTAGTAGAGATGGGGT
TTCACCATGCTGGCCAGGCTGGTCTTGAACTCCTGACCTGGTGATCTGCCCGCCTCGGCCTCCC
AAAGTGCTGGGATTACAGGCGTGAGCCACTGCACCCGG<mark>ACCACACCTATGCA</mark>

241nt

1-224 Alu sequence

tttggcTCACTGCAACCTCCACCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGCAGC
TGAGATTACAGGCATGCACCACCATGCTCGGCTAATTTTTGTATTTTTAGTAGAGATGGGGT
TTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTGGTGATCTGCCCGCCTCGGCCTCCC
AAAGTGCTGGGATTACAGGCGTGAGCCACTGCACCCGG<mark>ACCACACCTATGCATCG</mark>

386nt

1-384 Homo sapiens chromosome 8 alternate assembly CHM1_1.1 29167094 to
29167478

gcggttttggcTCACTGCAGGCTCCGCCCCCCGGGGTTCACACCATTCTCCTGCCTTAGCCTCCCGA
GTAGCTGGGACTACAGGTGCCCGCCACCTCGCCTGACTAGTTTTTTGTATTTTTAGTAGAGAC
GGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCTCCTTGG
CCTCCCAAAGTGCTGGGATTACATACGTGAGCCACTGCGCCCAGCCTGAAGTTTCTTTTCTGA

41

CAGAAATAGGAAGGAAATTTAGACAGAAGTACGAGCTTTAGCAGAAAAATTGTTTTTCAACC
CCAGTAGTCAGAGTCTAATAAACCTTATTTTAATTGGCAAATTAAAAGAAATGAACTCTAAA
ACACTTCAAGTAAGT<mark>CA</mark>

69nt

1-56: Homo sapiens chromosome 8 GRCh38.p1285269931 to 85269986

gtcgtaTTATATTGTAACCCACAATATCGTTTACAGGCAAAATAATGGTATATCATGAGACA<mark>GA
TTGTTTGGCCG</mark>

185nt

1-184: Alu sequence

gttttggcTCACTGCAACCTCCACCTCCCAGGTTCAACAATTCTCGTGTCTCAGCCTCCCAAGTAG
CTGGGACTACAGGCGCCTGCCGCCACACCTGACTAATTTTTGTATTTTTGGTAGAGACGGAGT
TTCCCTATGTTGCCCAATCCGGTCTCGAACTCCTGGGCTCAAGTGATGAGCCCACCTCGGCCT
<mark>A</mark>

94nt

33-94 : Homo sapiens chromosome 1, GRCh38.p12 Primary Assembly 86576654 to 86576715

ggtctatata<mark>CACAGAGCTCTCGGCTCTTGGTGATGCGGTTTT</mark>GGCAGGAAAAGAAAGCAGGAGT
AATCCTCGACAGCTGGAACAAAGTATAAATATAGTCCTGagaactg

114 nt

1-108 repeated sequence

ATTGATTATTCCATTACATTCCATTCGATGATTCCATTTGAGTCCATTCAATGATTCTATTCG
ATTCCATTCAATAATTCCATTCGATTCCATTTGATGATGATCCAT<mark>CTCACG</mark>

167 nt

1-159: Homo sapiens chromosome 16, GRCh38.p12 Primary Assembly 3171420 to 3171578

gatgcggttttggcCCAACAGGATTTGACCCTGAGGCCCACTCTCACCCTAATCATAACCGCAAAAC
CACCAGCGCCTGGAGAGAGAGTGAGAGAGAAACAGAAACGGAGCGAGTGTGTGTCGCTGTGA
TGCCCTTTGCCGTCGCTGCTCATCCCCAGTGACCTCCTGAACTT<mark>CAGAGCTC</mark>c

42

266 nt

1-99: : Homo sapiens chromosome 16, GRCh38.p12 Primary Assembly 3681336 to 3681434 ; 96-255: Homo sapiens chromosome 16, GRCh38.p12 Primary Assembly

ggtgatgcggttttggcaCCCATTCTGCCTGTGCTGCCTACAATGGGAAGGCTGACAGCTGGCCAGTC
TCCCTGGCCTGCTGCTTGCTGTGGATTATTTGCTGAGTGGCACTGCCCATTCTGCCTGTGCTG
CCTACAATGGGAAGGCTGACAGCTGGCCAGTCTCCCTGGCCTGCTGCTTGCTGTGGATTATTT
GCTGAGTGGCACTGACCACCCTTGCTTTGCCAATGGAGACAGCACTGGCGCCTGGTGCTAGTG
GGCTCCCAAGGCTCCTTTCCGGGTTCT

197 nt

1-180: CD4-3200bp reporter

ggagacccaagctgTGGTGATACGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTC
ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAATTTGTTTTGGCACCAAAATCA
ACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGT
ACTGGCTTATCGAAAACAC

99 nt

1-99 : pCOH 45084606

gaaggaattCGAGCTCGCCCGGGGATCTCGAGGTCACCCTGACGGTGTCGTCCATCACAGTTTGC
CAGTGATACACATGGGGATCAGCAATCGCGCATAATCGATATTaccctgtta

99 nt

1-99 : pCOH 45084606

gaaggaattCGAGCTCGCCCGGGGATCTCGAGGTCACCCTGACGGTGTCGTCCATCACAGTTTGC
CAGTGATACACATGGGGATCAGCAATCGCGCATAATCGATATTaccctgttat

354 nt

1-354 : Homo sapiens chromosome 14, GRCh38.p12 Primary Assembly 75197885 to 75198238, 1-95 repeated sequence

ggttttggcaTGTTGGCCAGACTGGTCTCGAATTCCTGACCTCAGGTGATCCGCCCGCCTCGGCCT
CCCAAAATGCTGGAATTATAGGCGTCAGCCACTGCGCCCTGCTAATGGACAACCTTTAAACCT
ACATGTTTGGCTATTTCTTAGTATTTTAGGGTTGCTTCGTGCTCTGGCCTGGGTTTGCAGGAA
AAATGGGAGAGAAGAGGAGAAAAGGGCTGGAAGAGGCTTGGCCATGGTAGAAGGCATGGGTG

GGTGTGAGGAGAAGGAAACAGAGTGATGGGCCAGCAAGGAGTCACAGGGTTTCGGGATCTGG
GCTAGCGACCCTCCTGGCCTCCAATGGGCTAACACGGAAAAGAGGGATc

129 nt

1-129: Homo sapiens chromosome 9, GRCh38.p12 Primary Assembly 99847500 to 99847628

ggttttggcaAGGTGGTGCTTACCTGTGATCATGCTCAGTGCTGACAGGCAGGCTAAGGCTTGGA
TATCAAGGTTCAGGCTCTGCAAATTTAAGGAAAAGTCTTTAATAGAGTCGAACCACTCCCCA
AATCCACGAAGG

siRAD51 #1

siRAD51#2

49nt

1-49 pBACSe

ccctgttCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGGCTGCAGGtccctatc

1-130 Homo sapiens chromosome 5, GRCh38.p12 Primary Assembly 172716876 to 172717005

ttaccctCTTGGGAGCCGACAAGGGGGTCTCTAGCGCCACCTTCCGGGCATCCGGGAAATGTCCG
TACAGTCCCACCAAGGCCCACCTGCCGCCTCGACATTCATAAATAAACGCCAACAATGACAGC
GACAGCAACTGGTTATTACCCTCAG

1-140 pCOH CD4 190-329

cctgttatcGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAA
AAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC
ATACTCTTCCTTTTTCAATATC

177 nt

72-123: Homo sapiens chromosome 9, GRCh38.p12 Primary Assembly 16273586 to 16273637

45

GAAAGAGTTCATTGTTTCTGATTGTAAAAATAAAACTTAATATATATGCTCTTCCAAGAAGA
ATCACTGACTACATAAATGGATAAAGAAGAAAAGGCAATTCATCCATAGTCCTGCCACACAG
AGAAAGCCATTCTTCGTATTTTGGCATATGTACTTTTTAATAAAAGTATACTT

156 nt

1-154 : Homo sapiens chromosome 16, GRCh38.p12 Primary Assembly 3681281 to 3681434

ttttggcaCCCATTCTGCCTGTGCTGCCTACAATGGGAAGGCTGACAGCTGGCCAGTCTCCCTGGC
CTGCTGCTTGCTGTGGATTATTTGCTGAGTGGCACTGACCACCCTTGCTTTGCCAATGGAGAC
AGCACTGGCGCCTGGTGCTAGTGGGCTCCCATGCT

si53BP1+ siRAD51#1

si53BP1+ siRAD51#2

214 nt

1-210: CD4-3200bp reporter

cgctctagagcaacTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAG
GAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTC
TTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTT
GAATGTATTTAGAAAAATAAA CAAATAGGGGTTGGAAcctatctagatat

48 nt

1-48: CD4-3200bp reporter

cgactcactatagATATGAAATCACGCCATGTAGTGTATTGACCGATTCCTTGCGGTCCGAtctagat
atgaa

39 nt

39 nt : non template

accctgttatATTATATCCCTATCTAGGAATTACCCTGTTATATTATATccctatctag

124 nt

1-124: pBSCe ISceI

47

accctgttaCCATCATCCATGAACCAGTATGCCAGAGACATCGGGGTCAGGTAGTTTTCAACCAG
GTTGTTCGGGATGGTTTTTTTGTTGTTAACGATGAACAGGTTAGCCAGTTTGTTGAAAGCTT
GGTGTTtccctatc

siBRCA2 #1

153 nt

1-154 Homo sapiens chromosome 16, alternate assembly CHM1_1.1 3221396 to 3221549

atgcggtttttggcCCAACAGGATTTGACCCTGAGGCCCACTCTCACCCTAATCATAACCGCAAAACC
ACAGCGCCTGGAGAGAGAGTGAGAGAGAAACAGAAACGGAGCGAGTGTGTGTCGCTGTGATG
CCCTTTGCCGTCGCTGCTCATCCCCAGTGACCTCCTGAA

siBRCA2 #2

111 nt

9-111: pBASCe

GACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGT
GCGGGGGGAGCGGCTCGGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAtatctagat

164 nt

3-164: pCOH 1362-1201

CACGTGTTGCTCTAGAGCGCTAGCCAGCTTGGGTCTCCCTATAGTGAGTCGTATTAATTTCGA
TAAGCCAGTAAGCAGTGGGTTCTCTAGTTAGCCAGAGAGCTCTGCTTATATAGACCTCCCACC
GTACACGCCTACCGCCCATTTGCGTCAATGGGGCGGAG ttcgaattcgag

49

si53BP1+ siBRCA2 #1

132 nt

17-132: Bovine satellite DNA fragment

TATTGATCACGTGGCTGATCATGCACTGATCACGTGACTATCATGCACTGCTCACGTGGCTGA
TCATGCGATGATCACGCAGCTACCATGTACTGGTCACATGATTGATCATGAACTGATCACGTG
ACTGATtatccctatc

120 nt

1-120: Homo sapiens chromosome 1, alternate assembly CHM1_1.1 86165308 to 86165427

acggaaggaattTCCCAATACCGTTCCTAAAAATTAGAACTACTTAGAAATTTACCCTTGGCTGTT
ATGAGGAAAAATCTGAGTCTATAGGAGTCAGTTTGTTTTCACTAAATGAGCTTGAAGCATGA
CTTGtgttatcccta

66 nt

4-67: pCOH 1300-1363

AGATATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCTCTAGAGCAA
CACGGtcacccattcgaatt

si53BP1 + siBRCA2 #2

91 nt

6-85: Homo sapiens chromosome 5, alternate assembly CHM1_1.1 35891395 to 35891481

<mark>CTTGG</mark>TATGTATGTATGTATGCATACATATGTATGCATACATGTATACATATGGTATACATA
TGGTATGCATGTATGTAGGCATA<mark>CCAATA</mark>

156 nt

1-154: Homo sapiens chromosome 16, alternate assembly CHM1_1.1 3221396 to 3221549

ggttttggcCCAACAGGATTTGACCCTGAGGCCCACTCTCACCCTAATCATAACCGCAAAACCACC
AGCGCCTGGAGAGAGAGTGAGAGAGAAACAGAAACGGAGCGAGTGTGTGTCGCTGTGATGCC
CTTTGCCGTCGCTGCTCATCCCCAGTGACCTCCTG<mark>AA</mark>

S1B. Sequences of all junctions of the two I-SceI sites in GC92 cells bearing the CD4-3200bp reporter after 53BP1, RAD51, and/or BRCA2 siRNA transfection (see excel document).
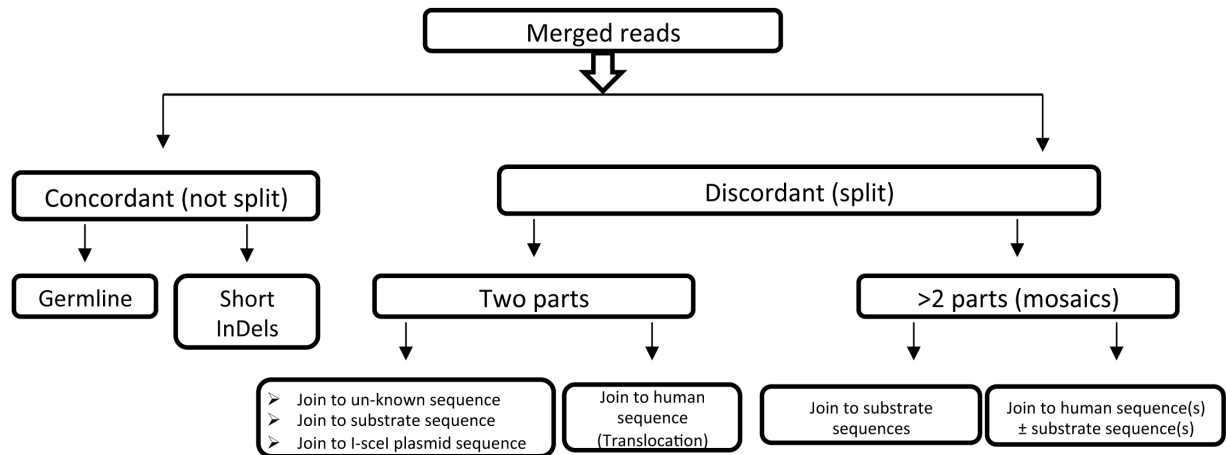
Supplemental data S2
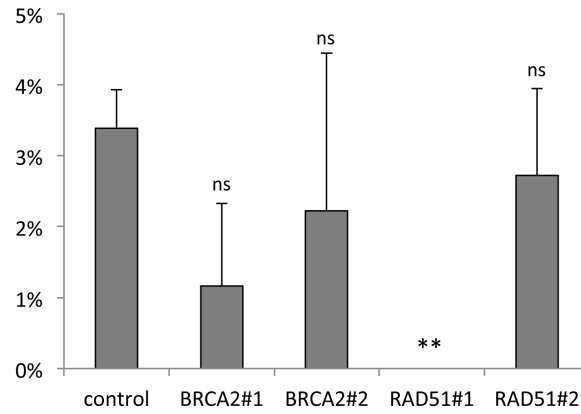
Fig. S2. ECS capture in the presence of 53BP1 after BRCA2 or RAD51 depletion in GC92 cells. Bars represent the average of 2-8 experiments (mean ±SEM).
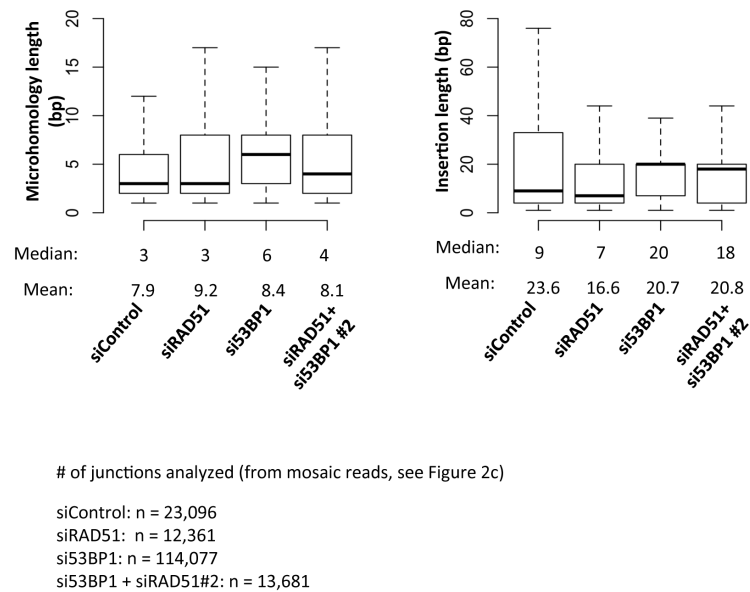
Supplemental data S3

A.

B.

			Expt 1				Expt 2			
			siControl	siRAD51	si53BP1	si53BP1+siRAD51	siControl	siRAD51	si53BP1	si53BP1+siRAD51
		Total merged read #	54763	60806	511828	188929	236234	119868	126628	27509
		Concordant (Germline + InDels) read # (%)	43160 (78.8)	42567 (70)	268008 (52.4)	120996 (64.04)	130800 (55.4)	71477 (59.6)	59568 (47)	19 593 (71.2)
Discordant/split read	2 parts	Total read # (%)	11127 (20.32)	16754 (27.55)	207062 (40.45)	62626 (33.15)	94804 (40.13)	44445 (37.07)	56147 (44.34)	7072 (25.70)
		contain unknown sequence	952 (1.74)	604 (0.99)	23163 (4.53)	1027 (0.54)	2375 (1.01)	930 (0.78)	1129 (0.89)	332 (1.21)
		contain substrate sequence	8212 (14.99)	10860 (17.86)	125570 (24.53)	30065 (15.91)	65658 (27.79)	30968 (25.83)	33042 (26.09)	3894 (14.15)
		contain I-scel donor plasmid sequence	699 (1.28)	905 (1.49)	14140 (2.76)	2390 (1.27)	7964 (3.37)	2307 (1.92)	3584 (2.83)	340 (1.24)
		contain human genome sequence (Translocations)*	1264 (2.31)	4385 (7.21)	44189 (8.63)	29144 (15.43)	18807 (7.89)	10240 (8.54)	18392 (14.52)	2506 (9.11)
	>2 parts (mosaics)	Total read # (%)	476 (0.87)	1485 (2.44)	36758 (7.18)	5307 (2.81)	10630 (4.5)	3946 (3.29)	10913 (8.62)	844 (3.07)
		contain substrate sequence only	241(0.44)	851 (1.40)	25670 (5.01)	1925 (1.02)	3790 (1.60)	1068 (0.89)	2358 (1.86)	640 (2.32)
		contain human genome sequence	235 (0.43)	634 (1.04)	11088 (2.16)	3382 (1.79)	6840 (2.89)	2878 (2.4)	8555 (6.75)	204 (0.74)

Figure S3. A. Scheme of LAM-HTGTS analysis pipeline. B. Numbers of LAM-HTGTS-sequenced and analyzed reads.
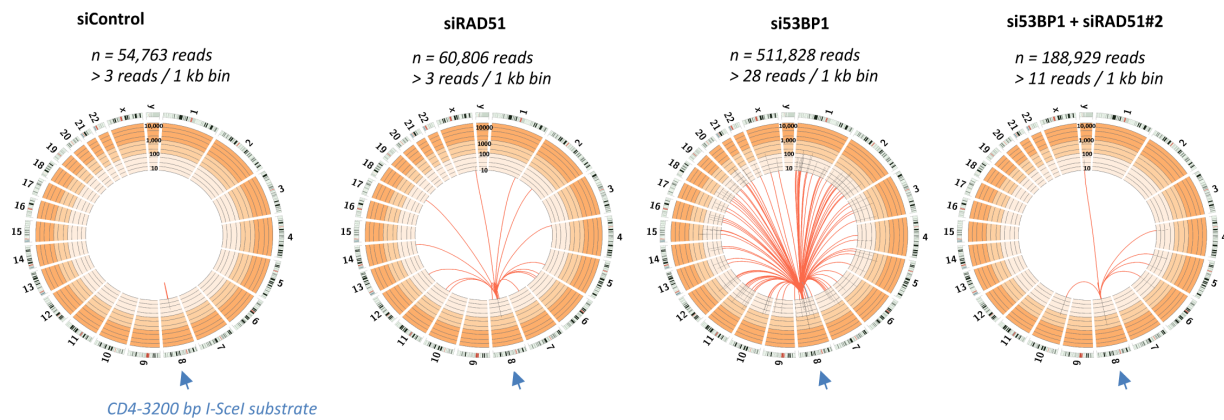
Supplemental data S4

# of junctions analyzed (from mosaic reads, see Figure 2c)

siControl: n = 23,096
siRAD51: n = 12,361
si53BP1: n = 114,077
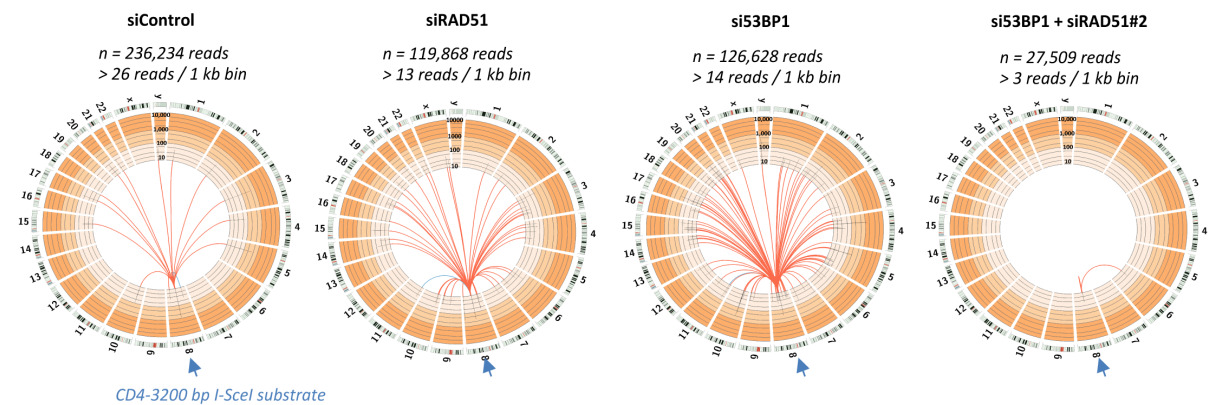si53BP1 + siRAD51#2: n = 13,681

Fig. S4- Microhomology (left panel) and insertion length (right panel) in the junction sequences of mosaic reads. The thick horizontal line indicates the median, and the upper and lower lines correspond to the 25th and 75th percentiles.
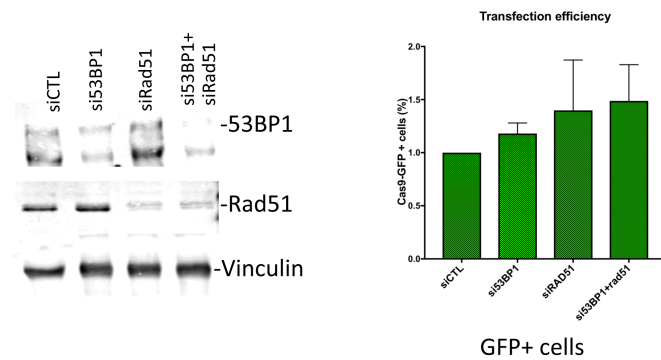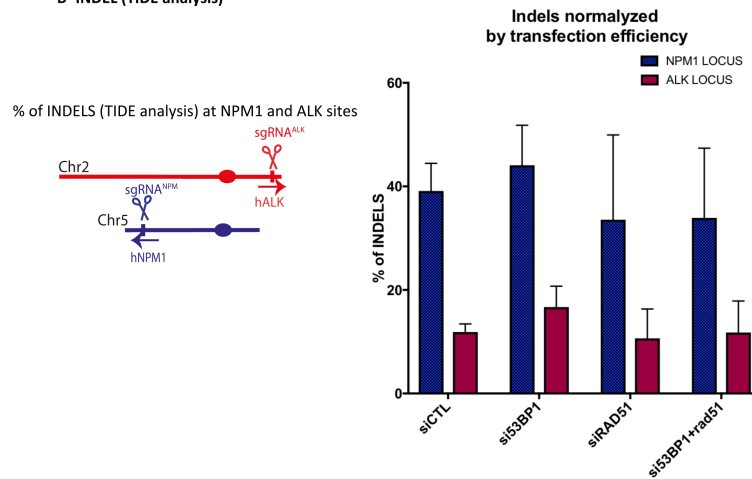
Supplemental data S5.

EXPERIMENT 1

Fig. S5- Biological duplicates of the LAM-HTGTS experiments. Circos plots represent interchromosomal junctions.

Supplemental data S6

A- Expression of 53BP1 and Rad51 (left panel) and Cas9-GFPtransfection efficiency (rigth panel)

B- INDEL (TIDE analysis)

Fig. S6- A) Analysis of 53BP1 and Rad51 expression (A-left panel) and Cas9-GFP transfection efficiency (A-right panel) and (B) % of INDELs corresponding to CRISPR /Cas9 cleavage efficiency upon silencing of RAD51 and/or 53BP1.

Supplemental data S7

dataset : 419 non HRD cases

SV	#					%				
	DEL	DUP	INV	TRA	ALL	DEL	DUP	INV	TRA	ALL
microHom	5528	5958	6953	11448	29887	62%	66%	62%	62%	63%
ins_1 to 10	1623	1380	2067	3892	8962	11.3%	12.1%	10.2%	9.2%	10.4%
ins_11 to 45	997	1108	1146	1707	4958	5.3%	4.6%	5.8%	4.9%	5.1%
ins_more 45	467	422	649	910	2448	2.9%%	2.5%	3.9%	3.6%	3.3%
no microhom /no insert	254	226	434	664	1578	18%	15%	18%	21%	19%
total junctions #	8869	9094	11249	18621	47833					

Simple Insertions in INDELs	#	%
Deletion	54936	59.30%
ins_1 to 10	36926	39.86%
ins_11 to 45	622	0.67%
ins_more 45	157	0.17%
total junctions #	92641	

Fig. S7- Analysis of SVs and INDELs junctions in non-HRD breast tumors.