Mechanism of tandem duplication formation in BRCA1-mutant cells

Nicholas A. Willis1, Richard L. Frock2, Francesca Menghi3, Erin E. Duffey1, Arvind Panday1, Virginia Camacho4, E. Paul Hasty5, Edision T. Liu1,6, Frederick W. Alt7 & Ralph Scully8

Small, approximately 10-kilobase microhomology-mediated tandem duplications are abundant in the genomes of BRCA1-linked but not BRCA2-linked breast cancer. Here we define the mechanism underlying this rearrangement signature. We show that, in primary mammalian cells, BRCA1, but not BRCA2, suppresses the formation of tandem duplications at a site-specific chromosomal replication fork barrier imposed by the binding of Tus proteins to an array of Ter sites. BRCA1 has no equivalent role at chromosomal double-stranded DNA breaks, indicating that tandem duplications form specifically at stalled forks. Tandem duplications in BRCA1 mutant cells arise by a replication restart–bypass mechanism terminated by end joining or by microhomology-mediated template switching, the latter forming complex tandem duplication breakpoints. Solitary DNA ends form directly at Tus–Ter, implicating misrepair of these lesions in tandem duplication formation. Furthermore, BRCA1 inactivation is strongly associated with ~10 kilobase tandem duplications in ovarian cancer. This tandem duplicator phenotype may be a general signature of BRCA1-deficient cancer.

BRCA1 suppresses Tus–Ter–induced TDs

We previously described a Rosa26-targeted 6×Ter–HR reporter for simultaneous measurement of STGC and LTGC in mammalian cells, in response to a Tus–Ter-mediated RFB or a chromosomal DSB induced by the rare-cutting I-SceI homing endonuclease10,26. In the reporter shown in Fig. 1a, STGC converts the cell to GFP+RFP+, while LTGC converts it to GFP−RFP+, by replicative duplication of an RFP expression cassette. In response to a Tus–Ter block (after transient Tus expression), we observed extremely low levels of a novel GFP−RFP+ repair product (Fig. 1b). We compared Tus–Ter–induced GFP−RFP+ products in cells expressing wild-type Brca1 (Brca1flox/flox1) versus isogenic cells that express hypomorphic Brca1 alleles lacking the in-frame exon 11 (Brca1flox exon11). Loss of wild-type Brca1 reduced Tus–Ter–induced STGC and increased Tus–Ter–induced LTGC as previously described10 (Fig. 1b and Extended Data Fig. 1). Brca1flox exon11 cells revealed an approximately 10-fold increase in Tus–Ter–induced GFP RFP+ products in comparison to Brca1flox/flox1 cells. GFP RFP+ products were further increased by short interfering RNA (siRNA)-mediated depletion of the residual Brca1 hypomorphic gene product and were not suppressed by RAD51 depletion (Fig. 1b and Extended Data Fig. 1). In parallel, I-SceI induced low levels of GFP RFP+ products that were only marginally increased by the loss of BRCA1 (Extended Data Fig. 1). Thus, BRCA1 suppresses a novel RAD51-independent GFP RFP+ outcome primarily during the stalled fork response.

Replication fork stalling at abnormal DNA structures or after collision with transcription complexes is a source of genomic instability in cancer and in developmental disorders1–5. Homologous recombination (HR) at stalled or collapsed forks can either suppress or promote genomic instability6,7. To study repair at stalled mammalian replication forks, we previously adapted the Escherichia coli Tus–Ter replication fork barrier (RFB)8,9 to trigger locus-specific fork stalling and HR on a mammalian chromosome10. We uncovered functions for BRCA1, BRCA2 and RAD51 in suppressing aberrant replicative HR responses at stalled forks. In wild-type cells, conservative short tract gene conversion (STGC) is the major HR product at Tus–Ter. In cells that lack BRCA1 and RAD51, approximately 85% of all Tus–Ter–induced HR events are resolved by aberrant long tract gene conversion (LTGC)10—a replicative response to fork stalling that is potentially analogous to break-induced replication in yeast11–13. BRCA1, BRCA2, RAD51 and the Fanconi anaemia (FA) genes have additional non-HR functions at stalled forks, where they protect DNA from degradation by the MRE11 nuclease14. BRCA1, together with its heterodimeric partner BARD1, has also been implicated in the removal of the CMG replicative helicase from the stalled fork15. BRCA1, BARD1 and the BRCA1-interacting protein CtIP have BRCA2-independent functions in DNA end processing16–18. BRCA1–BARD1 interacts with RAD51 directly and also indirectly via PALB2–BRCA219,20. Thus, BRCA1 performs several functions at the stalled fork and in double-strand break (DSB) repair, only some of which are shared with BRCA2.

Recently, a rearrangement signature specifically associated with BRCA1 loss was identified in the human breast cancer genome—the presence of abundant small (~10 kb) tandem duplications (TDs) with microhomologous breakpoints21,22. This chromotype, which differs from the larger (>100 kb) TDs noted previously in the cancer genome23,24, was termed ‘rearrangement signature 3’ or ‘group 1 TD phenotype (TDP)22,25. We will use the latter term here. Group 1 TDs are strongly associated with the loss of BRCA1 but not with the loss of BRCA2, and are enriched at loci that disrupt tumour suppressor genes, suggesting that group 1 TDs promote tumorigenesis in BRCA1-linked breast cancer21,22. However, the mechanisms that connect BRCA1 loss with TD formation remain undefined. Similarly, it is unclear whether the suppression of group 1 TDs is an intrinsic BRCA1 function that operates in primary cells. In this study, we address these questions by analysing 2–6-kb microhomology-mediated TDs that arise at a Tus–Ter site-specific chromosomal RFB in primary mouse embryonic stem (ES) cells.
To determine the rearrangement underlying the GFP RFP+ outcome, we analysed 6×Ter-HR reporter structure in Tus−Ter-induced GFP RFP+ clones (Fig. 2). We used fluorescence-activated cell sorting (FACS) to isolate Tus−Ter-induced GFP RFP+ clones from Brca1fl/exon11 Brca1−/uni2032 Brca1fl/exon11 6×Ter-HR reporter cells, in parallel with Tus−Ter-induced GFPRFP+ LTGC and STGC products. Representative primary FACS data for Brca1[Δexon11] and Brca1[Δexon11] Brca1−/uni2032 6×Ter-HR reporter cells co-transfected with wild-type Tus and short interfering RNA (siRNA) against Luc or Brca1. FACS plots produced from pooled data of technical duplicate samples from three independent experiments. Numbers represent percentages. See Extended Data Fig. 1 for additional primary data, quantification and Brca1 mRNA depletion. Red arrowhead denotes GFP−RFP+ repair products in Brca1[Δexon11] cells depleted of BRCA1.

Figure 2 | Tus−Ter-induced GFP−RFP+ repair products are microhomology-mediated tandem duplications. a, STGC, LTGC and GFP−RFP+ products. Elements as in Fig. 1a. Red half-arrows denote primers for breakpoint PCR. B, BglII site. Southern blotting with GFP probe fragment sizes indicated. Grey hatched box denotes the breakpoint of the GFP−RFP+ product. b, Analysis of GFP−RFP+ repair products. Top, Southern blots of Tus−Ter-induced STGC, LTGC and GFP RFP+ products in Brca1[Δexon11] 6×Ter-HR reporter cells. M, molecular mass marker lane. Red asterisk denotes example of class 1 GFP−RFP+ repair product. Blue asterisk denotes example of class 2 product. See Extended Data Fig. 2 for sequence analysis of these two clones. Bottom, breakpoint PCR products. c, GFP−RFP+ products are microhomology-mediated TDs. Cartoons show typical class 1 and 2 TDs. Elements as in a and Fig. 1a. Green line denotes TD breakpoint. For gel source data, see Supplementary Fig. 1.

Specificity of TD suppression by BRCA1
To determine whether TD suppression at Tus−Ter-stalled forks is specific to BRCA1, we studied the contribution of additional stalled fork metabolism/repair proteins to Tus−Ter-induced repair. We compared, in parallel, the effect of siRNA-mediated depletion of candidates on Tus−Ter-induced versus I-SceI-induced repair in Brca1[Δexon11] cells versus Brca1[Δexon11] cells, using siRNA against luciferase as control. In Brca1[Δexon11] cells (that is, expressing wild-type Tus), we identified BRCA1, BARD1 and CtIP as major suppressors of Tus−Ter-induced TDs (Extended Data Fig. 3a). CtIP acts largely independently of BRCA1 as a TD suppressor, as it does in certain other repair functions26,27. By contrast, BRCA2, RAD51, FANCA, FANC2 or SLX4 (also known as FANCP) suppressed TDs modestly or not at all, despite evidence that these proteins support Tus−Ter-induced HR, as expected from previous studies10,28 (Extended Data Fig. 3b). I-SceI-induced GFP RFP+ products were not regulated by the aforementioned proteins (Extended Data Fig. 3a). Depletion of the FANC translocase29 or the Bloom’s syndrome helicase (BLM)30 did not induce TDs in Brca1[Δexon11] cells but unexpectedly increased Tus−Ter-induced TDs around 15-fold in Brca1[Δexon11] cells (Extended Data Fig. 3a, c). FANC and BLM can each disassemble late recombination intermediates but are also implicated in stalled fork metabolism31. Loss of FANC or BLM affected Tus−Ter-induced TDs quantitatively but not qualitatively (Extended Data Fig. 4a; see also TD breakpoint analysis, below). Co-depletion of FANC and BLM in Brca1[Δexon11] cells produced additive effects on TD formation (Extended Data Fig. 4b), suggesting that the two proteins act independently to suppress Tus−Ter-induced TDs. As a further test of the relative contributions of BRCA1 and BRCA2 to TD suppression, we depleted BRCA1, BARD1, BRCA2 or RAD51 in combination with FANCM or BLM in Brca1[Δexon11] 6×Ter-HR reporter cells. Consistent with the above findings, co-depletion of BRCA1, BARD1 or CtIP with FANCM or BLM induced Tus−Ter-induced TDs, whereas co-depletion of BRCA2 or RAD51 with FANCM or BLM had a minimal effect.

ArticolereSearcH
Figure 3 | Candidate mechanisms of Tus–Ter-induced TDs. a, Breakage-fusion model. GFP elements not shown. Black lines denote parental DNA. Blue lines denote nascent strands of conventional replication. Half arrows denote nascent strand 3' ends. Scissors indicate sites of fork breakage. Pink dashed arrow denotes the fusion of broken sisters by end joining. b, MMBIR model. Red half arrow denotes repair synthesis during bubble migration. Other symbols are as in a. c, Replication restart-bypass model. The leftward fork undergoes aberrant replication restart—for example, engaging a bubble migration mechanism, as shown. The rightward fork bypasses the leftward nascent strand and stalls at Tus–Ter. End joining (pink dashed arrow) completes the TD. Symbols as in a and b. d, Summary of predictions made by the three TD models.

Mechanism of TD formation

Three different mechanisms could mediate TD formation at stalled forks. The first invokes breakage of both sister chromatids and their fusion by end joining (breakage–fusion; Fig. 3a). The 'partner' sister chromatid (the sister that does not acquire a TD) would be broken and rearranged during this process. A second model invokes TD initiation by microhomology-mediated synopsis of a free DNA end generated at the stalled/collapsed rightward fork (Fig. 3b), priming TD formation by microhomology-mediated break-induced replication (MMBIR)33,34. A third mechanism entails aberrant 'replication restart' of the stalled/collapsed leftward fork (Fig. 3c). By analogy with previously described RAD51-independent replication restart mechanisms35–38, processing of the collapsed leftward fork primes extension of the stalled leading strand by a migrating bubble mechanism that resembles break-induced replication13 (Fig. 3c). The approaching conventional rightward fork bypasses the restarted leftward nascent strand and re-copies the TD tract before stalling at Tus–Ter (replication restart-bypass; Fig. 3c). According to this model, the 'upstream' site of the TD breakpoint (defined in Extended Data Fig. 2a) marks the site of displacement of the leftward nascent strand, and the 'Ter-proximal' site (Extended Data Fig. 2a) is derived from a free DNA end formed at the Tus–Ter-stalled rightward fork. Note that the fork breakage shown in Fig. 3b, c is not a requirement of these models, since a free DNA end could alternatively be generated at Tus–Ter by fork regression39. Indeed, high frequency rearrangements observed at a site-specific RFB in Schizosaccharomyces pombe are not accompanied by evidence of fork breakage37.

Figure 4 | A replicative mechanism involving classical non-homologous end joining mediates Tus–Ter-induced TDs. a, Aneuploidy induced by 30 μM cytochalasin B (CB) during TD induction. b, Analysis of Tus–Ter-induced TDs in Brca1loxPlox ES cells. c, Breakage–fusion by end joining (breakage–fusion; Fig. 3a). The 'partner' sister chromatid (the sister that does not acquire a TD) would be broken and rearranged during this process. A second model invokes TD initiation by microhomology-mediated synopsis of a free DNA end generated at the stalled/collapsed rightward fork (Fig. 3b), priming TD formation by microhomology-mediated break-induced replication (MMBIR)33,34. A third mechanism entails aberrant 'replication restart' of the stalled/collapsed leftward fork (Fig. 3c). By analogy with previously described RAD51-independent replication restart mechanisms35–38, processing of the collapsed leftward fork primes extension of the stalled leading strand by a migrating bubble mechanism that resembles break-induced replication13 (Fig. 3c). The approaching conventional rightward fork bypasses the restarted leftward nascent strand and re-copies the TD tract before stalling at Tus–Ter (replication restart-bypass; Fig. 3c). According to this model, the 'upstream' site of the TD breakpoint (defined in Extended Data Fig. 2a) marks the site of displacement of the leftward nascent strand, and the 'Ter-proximal' site (Extended Data Fig. 2a) is derived from a free DNA end formed at the Tus–Ter-stalled rightward fork. Note that the fork breakage shown in Fig. 3b, c is not a requirement of these models, since a free DNA end could alternatively be generated at Tus–Ter by fork regression39. Indeed, high frequency rearrangements observed at a site-specific RFB in Schizosaccharomyces pombe are not accompanied by evidence of fork breakage37.

As summarized in Fig. 3d, the breakage–fusion, MMBIR and replication restart-bypass models predict different fates of the partner sister chromatid during TD formation and/or different dependencies on end joining. To retrieve the partner sister chromatid, we induced mitotic non-disjunction during the cell cycle in which the TD formed, by treating Tus-transfected FANC–depleted Brca1loxPlox cells with 30 μM cytochalasin B for 24 h immediately before FACS cloning of GFP+RFP− cells (Fig. 4a and Methods). Southern analysis of 60 independent GFP+RFP− aneuploid clones revealed no off-size GFP-hybridizing bands other than the TD itself in any clone, providing no support for the
breakage–fusion model (Fig. 4a and Supplementary Fig. 1). By contrast, 11 out of 60 GFP RFP+ clones contained two copies of the 6 Ter-HR reporter: one that had undergone a TD and one that retained the parental structure (Fig. 4b). In eight of these clones, re-cloning failed to separate the two reporter copies, confirming that the TD and unrearranged reporter were present in the same cell (Extended Data Fig. 6a, b). We obtained direct TD breakpoint sequences for 6 of these 8 clones. One TD breakpoint was blunt, one entailed insertion of one nucleotide and four revealed microhomology. The spontaneous non-disjunction rate for this cell line is around 1 in 1,000. The fact that 8 out of 60 CB-induced ‘non-disjunction’ TD clones revealed an unaltered partner sister chromatid indicates that TDs form at Tus–Ter primarily via a replicative mechanism, not by breakage–fusion. Interestingly, solitary (+) DNA ends are more common than (–) ends, indicating the presence of solitary DNA ends. Note virtual absence of signal in empty vector (EV) controls. Maps represent pooled data from two (I-SceI), three (Tus), or two (EV) independent replicates.

**TD breakpoint analysis**

To understand the mechanisms underlying TD formation better, we analysed in detail the sequence of Tus–Ter-induced TDs from Brca1△exon11 cells depleted of FANC, BRCA1 or BLM. TD spans varied from around 2 kb to 6 kb, which represent the technical boundaries of TD detection using this reporter (Extended Data Fig. 7a). We confirmed that BRCA1 and BARD1 are the dominant TD suppressors in Brca1△exon11 groups and for Brca1△exon11 cells depleted of FANC, data pooled from three independent biological replicates. All other Brca1△exon11 groups and for Brca1△exon11 cells depleted of FANC, data pooled from three independent biological replicates.

**Figure 5** | Solitary DNA ends form at Tus–Ter-stalled forks. CRISPR–Cas9 induces ‘bait’ DSBs approximately 30 kb from 6 Ter array + I-SceI site at Rosa26. a, I-SceI-induced two-ended DSB produces balanced (+) and (–) ends in HTGTS. Half-arrow denotes HTGTS sequencing primer. Weighted black line denotes duplex DNA. b, Focused rightward fork breakage produces (+) orientation DNA ends in HTGTS. c, Alternatively, solitary (+) DNA end forms via regression of rightward fork. Thus, stalled rightward forks generate (+) ends, irrespective of mechanism. Stalled leftward forks (not shown) generate (–) ends. d, HTGTS breakpoints in FANC-depleted Brca1△exon11 cells containing a single Rosa26-targeted 6 Ter-I-SceI-GFP cassette. Grey area/orange triangles: 6× Ter-array.

I-SceI-induced DSBs produce expected symmetrical pattern in HTGTS. Tus–Ter induces asymmetrical pattern with approximately sevenfold more (+) ends than (–) ends, indicating the presence of solitary DNA ends. Note virtual absence of signal in empty vector (EV) controls. Maps represent pooled data from two (I-SceI), three (Tus), or two (EV) independent replicates. e, Tus–Ter-induced HTGTS in Brca1△exon11 or Brca1△exon11 cells receiving siRNAs shown. For all Brca1△exon11 groups and for Brca1△exon11 cells depleted of FANC, data pooled from three independent biological replicates. All other Brca1△exon11 groups and for Brca1△exon11 cells depleted of FANC, data pooled from three independent biological replicates.
breakpoints were homoeologous, containing 1–2 bp internal mismatches within longer microhomology tracts of 4–10 bp, with no consistent strand preference of mismatch correction (Extended Data Fig. 7c). Notably, 6 out of 231 (2.6%) TDs contained complex breakpoints (Extended Data Fig. 7d), suggestive of microhomology-mediated template switching52. Template switching is associated with TD formation in E. coli, break-induced replication in S. cerevisiae and alternative end joining in mammalian cells.40,43–47 It has been invoked to explain complex breakpoints associated with replication stress in the cancer genome33,48. Our findings provide direct evidence of microhomology-mediated template switching at stalled mammalian replication forks.

**Solitary DNA ends form at Tus–Ter**

The Ter-proximal site of the TD represents the product of rightward fork stalling at Tus–Ter (Fig. 3c). Indeed, Ter-proximal sites were clustered near the first Ter elements encountered by the rightward fork, a minority being distributed upstream (Extended Data Fig. 7e). By contrast, upstream TD sites were more widely distributed (Extended Data Fig. 7f). To determine whether Ter-proximal TD sites correspond to detectable DNA lesions at Tus–Ter, we used high-throughput genome-wide translocation sequencing (HTGTS)49,50 to map translocation–competent DNA ends at Tus–Ter. As bait for HTGTS, we induced a Cas9–CRISPR–mediated DSB approximately 30 kb from the 6×Ter array at Rosa26 (Fig. 5). (The translocations studied here are strictly, intrachromosomal rearrangements.) Control I-SceI-induced two-ended DSBs should produce equal representation of (+) and (−) DNA ends in HTGTS mapping (Fig. 5a). By contrast, rightward forks arriving at Tus–Ter (Fig. 5b) might generate predominantly (+) DNA ends, whereas leftward forks (not shown) would generate (−) DNA ends. This polarity is expected whether the DNA end is generated directly by breakage at the branch-point of the stalled fork (Fig. 5b) or indirectly via fork regression (Fig. 5c). Notably, if either sister chromatid were broken anywhere other than at the branch-point of the stalled fork, this would generate a conventional two-ended DSB with equal representation of (+) and (−) DNA ends.

As expected, FANC-mutant Brca1Δcom12 cells co-transported with control I-SceI and the CRISPR–Cas9 bait vectors revealed symmetrical HTGTS distributions of (+) and (−) DNA ends that mapped to the I-SceI site adjacent to the Ter array49 (Fig. 5d). Conversely, translocations into Tus–Ter in FANC-mutant Brca1Δcom12 cells were highly asymmetric. We noted an approximately sevenfold excess of (+) ends compared with (−) ends (Fig. 5d), indicating that solitary DNA ends predominate at Tus–Ter-stalled forks. Tus–Ter HTGTS breakpoints were tightly focused on the Ter array and were microhomology biased in comparison to I-SceI HTGTS breakpoints, revealing a 1–2 bp microhomology preference reminiscent of Tus–Ter-induced TD breakpoints (Extended Data Fig. 8a; compare with Extended Data Fig. 7b). Furthermore, translocations at Tus–Ter were more abundant into the Ter sites first encountered by the approaching replication fork (Fig. 5d).

In all treatment groups, including cells containing wild-type Brca1, the distributions of Tus–Ter HTGTS breakpoints were similar (Fig. 5e and Extended Data Fig. 8b). However, a quantitative effect of BRCA1 on the formation of DNA ends at Tus–Ter is not excluded. In all treatment groups, the distribution of Ter-proximal TD sites (products of rightward fork stalling) was significantly shifted in comparison to the distribution of Tus–Ter HTGTS (+) ends (also products of rightward fork stalling; Extended Data Fig. 8b). Taken together, these findings suggest that the Ter-proximal site of the TD breakpoint arises from a solitary DNA end generated at the Tus–Ter RFB, which is further processed before being misrepaired in Brca1 mutants to form a TD.

**TD phenotype in BRCA1 mutant cancer**

Our data suggest that TD suppression at stalled replication forks is an intrinsic function of BRCA1. If so, the TD phenotype might be a general feature of BRCA1 loss in cancer. To test this idea, we analysed TDs occurring in 92 cancers from the Australian Ovarian Cancer Study (AOCS; http://www.aocstudy.org/), for which whole-genome sequence, BRCA1 promoter methylation status and transcriptome data are available. We noted a strong association between loss of BRCA1 by mutation or promoter methylation and group 1 TDP (Extended Data Fig. 9). Re-analysis of the Sanger Institute dataset22 using our TD algorithm confirmed that TDP group 1 is strongly associated with BRCA1 loss but not with BRCA2 loss. Indeed, in the Sanger dataset filtered to include only triple-negative breast cancers (which included almost all the group 1 TDP breast cancers), BRCA2 inactivation was negatively associated with group 1 TDP (Extended Data Fig. 9b). In the AOCS and Sanger TNBC datasets, we observed no association between group 1 TDP and either mutation or aberrant expression of FANCm or BLM (Extended Data Fig. 10). Whether these genes function as TD suppressors in human tumorigenesis therefore remains to be determined.

**Conclusion**

We demonstrate that BRCA1 but not BRCA2 is a major suppressor of TDs at a Tus–Ter RFB in primary mammalian cells. These findings recapitulate the group 1 TD phenotype of BRCA1-mutant breast cancers21,22,25. We therefore propose that group 1 TDs in BRCA1-linked cancer arise by defective processing of stalled replication forks. Extending these observations across tumour types, we observed a strong group 1 TDP in ovarian cancers that lack BRCA1. The group 1 TDP may therefore serve as a useful biomarker of BRCA1 loss in other cancer types51. Furthermore, our findings suggest that inactivation of BARD1 or RBBP8 (which encodes CtIP) may also be associated with group 1 TDP cancers.

Our analysis suggests that Tus–Ter-induced TDs in BRCA1 mutant cells arise by an aberrant replication restart-bypass mechanism terminated by end joining. Certain key elements of this mechanism are conserved in yeast7,13,37,38. However, it remains to be determined precisely how BRCA1 and BARD1, of which there are no yeast orthologues, suppress these aberrant stalled fork responses. BRCA1 and BARD1 have BRCA2-independent roles in DNA end-processing and in CMG helicase unloading at the stalled fork15–18. Thus, several distinct BRCA1-mediated functions might suppress TD formation at stalled forks. A notable aspect of this study is the finding that solitary DNA ends predominate at Tus–Ter-stalled forks. We detected these lesions in both TD-prone and control cells, suggesting that the production of solitary DNA ends is a generalized, perhaps physiological, response to fork stalling59. The element of TD formation that is specific to BRCA1 loss therefore appears to be the ‘licensing’ of an aberrant replication restart mechanism at stalled forks.

**Online Content**

Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 August 2016; accepted 9 October 2017.
Published online 22 November 2017.

1. Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. Nat. Cell Biol. 16, 2–9 (2014).
2. Duxin, J. P. & Walter, J. C. What is the DNA repair defect underlying Fanconi anemia? Curr. Opin. Cell Biol. 37, 49–60 (2015).
3. D’Andrea, A. D. Susceptibility pathways in Fanconi anemia’s and breast cancer. N. Engl. J. Med. 362, 1909–1919 (2010).
4. Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. Mol. Cell. 40, 179–204 (2010).
5. Garcia-Muse, T. & Aguillera, A. Transcription–replication conflicts: how they occur and how they are resolved. Nat. Rev. Mol. Cell Biol. 17, 553–563 (2016).
6. Carr, A. M. & Lambert, S. Replication stress-induced genome instability: the dark side of replication maintenance by homologous recombination. J. Mol. Biol. 425, 4733–4744 (2013).
7. Mayle, R. et al. Mus81 and converging forks limit the mutagenicity of replication fork breakage. Science 349, 742–747 (2015).
8. Mulcair, M. D. et al. A molecular mousetrap determines polarity of termination of DNA replication in E. coli. Cell 125, 1309–1319 (2006).
9. Larsen, N. B., Sass, E., Suski, C., Mankouri, H. W. & Hickson, I. D. The Escherichia coli Tus–Ter replication fork barrier causes site-specific DNA replication perturbation in yeast. Nat. Commun. 5, 3574 (2014).
10. Willis, N. A. et al. BRCA1 controls homologous recombination at Tus/Ter-stalled mammalian replication forks. Nature 510, 556–599 (2014).
11. Llorente, B., Smith, C. E. & Symington, L. S. Break-induced replication: what is it and what is it for? Cell Cycle 7, 839–844 (2008).
12. Anand, R. P., Lovett, S. T. & Haber, J. E. Break-induced DNA replication. Cold Spring Harb. Perspect. Biol. 5, a013973 (2013).
13. Saini, N. et al. Migrating bubble during break-induced replication drives compensative DNA synthesis. Nat. Cell Biol. 15, 739–733 (2006).
14. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. Cancer Cell 22, 106–116 (2012).
15. Long, D. T., Joukov, V., Budzowska, M. & Walter, J. C. BRCA1 promotes nucleotide excision repair at sites of ultraviolet damage in mammalian homologs with distinct mutagenic consequences. Mol. Cell. Biol. 24, 9305–9316 (2004).
16. Bunting, S. F. et al. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell 141, 243–254 (2010).
17. Smith, C. E., Llorente, B. & Symington, L. S. Template switching after UV-induced DNA damage. Mol. Cell 44, 235–251 (2011).
18. Sy, S. M., Huen, M. S. & Chen, J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proc. Natl Acad. Sci. USA 113, 7155–7160 (2016).
19. Zhao, W. et al. BRCA1–BARD1 promotes RAD51-mediated homologous DNA pairing. Nature 550, 360–365 (2017).
20. Menghi, F. & Liu, E. T. Reply to Watkins et al.: Whole-genome sequencing-based identification of diverse tandem duplication phenotypes in human cancers. Proc. Natl Acad. Sci. USA 113, E5259–E5260 (2016).
21. Chandramouly, G. et al. BRCA1 and CIP5 suppress long-tract gene conversion between sister chromatids. Nat. Commun. 4, 2404 (2013).
22. Polato, F. et al. CIP1 recombination is essential for viability and can operate independently of BRCA1. J. Exp. Med. 211, 1027–1036 (2014).
23. Popova, T. et al. Ovarian cancers harboring inactivating mutations in CDK12 display a distinct genomic instability pattern characterized by large tandem duplications. Cancer Res. 76, 1882–1891 (2016).
24. Watkins, J., Tutt, A. & Grigoriadis, A. Tandem duplications contribute to not one but two distinct phenotypes. Proc. Natl Acad. Sci. USA 113, E5257–E5258 (2016).
25. Menghi, F. & Liu, E. T. Reply to Watkins et al.: Whole-genome sequencing-based identification of diverse tandem duplication phenotypes in human cancers. Proc. Natl Acad. Sci. USA 113, E5259–E5260 (2016).
26. Chandramouly, G. et al. BRCA1 and CIP5 suppress long-tract gene conversion between sister chromatids. Nat. Commun. 4, 2404 (2013).
27. Polato, F. et al. CIP1 recombination is essential for viability and can operate independently of BRCA1. J. Exp. Med. 211, 1027–1036 (2014).
28. Nakanishi, K. et al. Homology-directed Fanconi anemia pathway crosslink repair is dependent on DNA replication. Nat. Struct. Mol. Biol. 18, 500–503 (2011).
29. Xue, X., Sung, P. & Zhao, X. Functions and regulation of the multitasking FANC family of DNA motor proteins. Genes Dev. 29, 1777–1788 (2015).
30. Lassar, N. J. & Hickson, I. D. ResQ helicases: conserved guardians of genomic integrity. Adv. Exp. Med. Biol. 767, 161–184 (2013).
31. Deans, A. J. & West, S. C. DNA interstrand crosslink repair and cancer. Nat. Rev. 11, 467–480 (2011).
32. Morimitsu, Y.-L., Zhao, G. & Hasty, P. Cells deleted for Brca2 COOH terminus exhibit hypersensitivity to gamma-radiation and premature senescence. Cancer Res. 58, 3441–3447 (1998).
33. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediates break-induced replication model for the origin of human copy number variation. PLoS Genet. 5, e1000327 (2009).
34. Payen, C., Koszul, R., Dujon, B. & Fischer, G. Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-dependent mechanisms. PLoS Genet. 4, e1000175 (2008).
35. Bhowmick, R., Minohcheromy, S. & Hickson, I. D. RAD52 facilitates mitotic DNA synthesis following replication stress. Mol. Cell 64, 1117–1126 (2016).
36. Doe, C. L., Osman, F., Dixon, J. & Whitby, M. C. DNA repair by a Rad22-Mus81-dependent pathway that is independent of Rhp51. Nucleic Acids Res. 32, 5570–5581 (2004).
37. Lambert, S. et al. Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. Mol. Cell 39, 346–359 (2010).
38. Nguyen, M. G., Jahan, M., Morrow, C. A., Osman, F. & Whitby, M. C. Recombination occurs within minutes of replication blockage by RTS1 producing restarted forks that are prone to collapse. eLife 4, e04539 (2015).
Molecular biology and siRNAs. The vectors for Ter HR reporters described were constructed by conventional cloning methods using a previously described 6×Ter-HR and RFP-SCR reporters16,26. PHV-NAT-CD52 vectors were derived from pHV-ZsGreen, a gift from B. Welsh and Z. Wedberg (Addgene plasmid 18121)27. Ter-containing plasmids were amplified in JFC33 (Tur+) strains of E. coli. siRNA SMARTpools were purchased from Dharmacon. All plasmids used for transfection were prepared by endotoxin-free maxiprep (QIAGEN Sciences).

Mouse cell culture. Mouse ES cells were authenticated as described in the text and were periodically tested for mycoplasma contamination. Only mycoplasma-free cells were used in experiments described here. Cells were thawed on mouse embryonic fibroblast (MEF) feeders and maintained in ES medium on gelatinized plates. 10 μg of Ter/HR reporter Rosa26-targeting plasmids per 1 × 10⁶ cells were linearized using PspI and introduced by electroporation. ES cells were seeded onto 6-cm dishes containing puromycin-resistant feeders and plates supplemented with 4 μg/ml puromycin for 24 h. Individual colonies were picked 7–10 days later. Rosa26-targeted lines were screened for by PCR. Reporter cassette integration and overall structure were verified for targeted lines by Southern blotting. Multiple Brca1-deficient or Xrcc4-deficient ES clones were generated by transient adenovirus-mediated Cre expression. Rosa26 genotyping primers: Rosa26 sense 5′-CATCAAGGAAAACCTGGACTACTG-3′; Ter-HR reporter antisense 5′-CCTCGGCTAGTTAGGGGATC-3′. Brca1 exon 11 status was determined by PCR: Brca1 5′-sense 5′-CTGGTATGTTGAGGATTCC-3′; Brca1 exon 11 antisense 5′-CAATAAATTCTGCTTGGACCC-3′; Brca1 exon 11 sense 5′-GGAAATGGCAACTTGCCTAG-3′; Brca1 3′-antisense 5′-CTCCGGAGCAGCTGCTTGAAG-3′. Xrcc4 status was determined by PCR: Xrcc4 5′-sense 5′-TTACGCTTACACGCTATAAG-3′; floxed allele, Xrcc4 5′-antisense 5′-GCACCTTTGCCTACTAAGCCATCTCAC-3′; Xrcc4 exon 3 deleted allele, Xrcc4 5′-antisense 5′-TGAAAATGGCCGACATTACCC-3′. Brca2 exons 26 and 27 status was determined by PCR: Brca2 intron 25′-sense 5′-TTCTACAAGGGCAGCTAATAAAG-3′; Brca2 exon 27′-antisense 5′-CGTCTTCCCTTCACTCAAG-3′.

Recombination assays. Approximately 1.6 × 10⁶ cells were co-transfected in suspension with 0.35 μg empty vector, pcDNA3-myc NLS-Tus19 or pcDNA3-myc NLS-1-Scel20, and 20 pmol ONTargetPlus-smartpool using Lipofectamine 2000 (Invitrogen). GFP ‘+’, GFP ‘−’ and GFP ‘F’ frequencies were scored 72 h after transfection by flow cytometry using a Becton Dickinson 5 Laser LSRIII in duplicate. For each duplicate sample, 3 × 10⁶–6 × 10⁶ total events were scored. Repair frequencies presented are corrected for background events and for transfection efficiency (50–85%). Transfection efficiency was measured by parallel transfection with 0.05 μg wild-type GFP expression vector, 0.30 μg control vector and 20 pmol siRNA. For depletion of two gene targets, 10 pmol of each siRNA was used, while single depletion controls received 10 pmol of the target siRNA and 10 pmol of control luciferase siRNA. Data presented represent mean and error bars represent the s.e.m. of between 5 (n = 5) and 11 (n = 11) independent experiments (n values given in figure legends).

Statistical methods. Figure legends specify the sample number in terms of the number of replicates within each individual experiment (typically two) and the number of independent experiments (n) that were performed to generate the data presented. For repair frequency statistical analysis, the arithmetic mean of samples collected for independent experiments was calculated and data points for each independent experiment were used to calculate a s.d./n, in which n indicates the number of independent experiments. Differences between sample pairs repair frequencies were analysed by Student’s two-tailed unpaired t-test, assuming unequal variance using GraphPad Prism v6.0d software. P values are indicated in each figure legend. Additional statistical analyses are as described in figure legends. No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

RT–qPCR analysis. RNA from transfected cells was extracted using QIAGEN RNAasy Mini Kit (QIAGEN Sciences) 48 h after transfection. First-strand cDNA synthesis was performed on an ABI 7300 Real Time PCR System using PrimeScript SYBR Green RNA-to cDNA 1-Step Kit (Applied Biosystems). SYBR green RT–qPCR assays of GAPDH and the siRNA-targeted gene were performed. We used the NIH NCBI Nucleotide utility to generate gene-specific primer sequences for mouse Brca1, Brca2, Rad51, Bard1, Ctip, Slx4, Fanca, Fancc and Gapdh. Primers for RT–PCR: Brca1 sense 5′-ATAGACCTGGAGAGGATGCTG-3′; Brca1 antisense 5′-CTGGGCGATTGTGCTTCTCT-3′; Brca2 sense 5′-TCTGCACACTGTAAGATGTC-3′; Brca2 antisense 5′-TCAAAGCTGGGCAGGAATTGTTAAATG-3′; Slx4 sense 5′-GTCGTTATCCTCGCTGAAGG-3′; Slx4 antisense 5′-GCACCTTTGGTGTCTCTGG-3′; Ctip sense 5′-AGGAGAGGAGGGAGCTG-3′; Ctip antisense 5′-TGAATAACCTCCTGGGCTG-3′; Fanca sense 5′-GGGAGCCTTGTACAACTGAT-3′; Fanca antisense 5′-GCCCGAGCGTCGTCATGTT-3′; Fanpds sense 5′-GCGGAGCTGAAGGACTGATGTCG-3′; Fanpds antisense 5′-ACTACATGCAAGACAGATG-3′; Gapdh sense 5′-GAAGCTAGGGTGTTGAAGACCT-3′; Gapdh antisense 5′-ATTCACACTGCCCTAAGTAC-3′; AACTC-3′. We used the Roche ProbeFinder utility based on Primer 3 software (Whitehead Institute, MIT) to generate gene-specific primer sequences for mouse Fancm and Bmnl. Fancm sense 5′-GTCCTTATCTGCGTGTAAGG-3′; Fancm antisense 5′-TTTTGTTAGCTAGCTTATAGTATG-3′; Bmnl sense 5′-GCGGAGCTGAAGGACTGATGTCG-3′; Bmnl antisense 5′-GCCACACTGCCCTAAGTAC-3′. Error bars represent the s.d. of the ΔCt value (s.d. = [ΔCt(mut) – ΔCt(wt)]/N).

Western blotting. Cells were lysed using RIPA buffer (50 mM Tris–HCl, pH 8.0, 250 μM NaCl, 0.1% sodium dodecyl sulphate, 1% NP-40 containing the protease inhibitors, PMSF, and Roche complete protease inhibitor tablet) and resolved by 10% bis-Tris SDS–PAGE (Invitrogen). Protein expression was analysed by immuno blotting using the following antibodies: β-tubulin (Abcam ab6064, 1:4,000), human RAD51 (aliquote B32, 1:500, mouse XRCC4 (Abcam ab97351, 1:3,000).

Southern blotting. Southern blotting of BglII- or Asel-digested genomic DNA was performed using a GPD cDNA probe by methods described previously16,26,27.

For all experiments, mouse ES cells containing a single, intact copy of the reporter integrated at the Rosa26 locus carrying 6-well plates (5 × 10⁶–10 × 10⁶ cells) using a Puregene DNA Isolation Kit (QIAGEN Sciences).

Individual repair clone capture and molecular analysis. Individual GFP ‘F’, GFP ‘+’, or GFP ‘F’ cells were captured by FACS 72 h after transfection using a FACSAria II SORP running FACSDiva software v.6.1.3. To capture aneuploid ‘non-disjunction’ clones, individual GFP ‘F’ or GFP ‘+’ cells were FACS-sorted 48 h after transfection. Cytochalasin B induced mitotic arrest and nondisjunction, 24 h after transfection cells were incubated for 22 h and FACS-sorted for 2 h in 20 μM dihydrocytochalasin B (Sigma Aldrich, D1641). Isolated colonies from single cells were picked from 6 cm dishes containing feeding MFs and individual repair clones expanded onto 24-well plates also containing feeding MFs. Genomic DNA was extracted from ES clones subsequently expanded and grown to confluence on gelatinized 6-well plates (~5 × 10⁶–10 × 10⁶ cells) using a Puregene DNA Isolation Kit (QIAGEN Sciences).

LTGC and TD breakpoint junction PCR was performed using Taq DNA Polymerase (QIAGEN Sciences) according to manufacturer’s instructions using primers unique to HR cassette synthetic RIPF exons: RIPF exon A sense 5′-ATGTCAGGCATCAAGCTCAG-3′, RIPF exon A antisense 5′-TTGACCTGACGTTCGAGCCTTTCCTC-3′; RIPF exon B sense 5′-GGGAGCCTTGTACAACTGAT-3′, RIPF exon B antisense 5′-GCCCGAGCGTCGTCATGTT-3′; RIPF exon C sense 5′-GGCAGGAGTAACTTGC-3′, RIPF exon C antisense 5′-TGACCTGAGTAACGAGCAGCCG-3′; Cas9 sense 5′-AAATCTCCTTCAAGGCAAGCAGGATGCTG-3′; Cas9 antisense 5′-CTGGGTCTGTATGAGAAGG-3′.

Unpaired PCR product sequencing was performed by Eton Bioscience using the Cas9 expression plasmid co-transfected with the bait plasmid and transgenic mice 6 weeks were used. LAM–HTGTS libraries were prepared and analysed as outlined previously16,28.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
bridge adaptor-sense-GCGACTATAGGCAAGCGTGGNNNNN-NH$_2$; bridge-adaptor-antisense 5-PhosCCACGCGTGCCTATAGTGCN-5' nested-PCR, 15-nested-ACACCTCTTTCCCTACACGAGCTCTTCCGATCT-5nt barcode-nested primer; CRISPR-Cas9-nested primer-CATGGCGGAAAGTAGATCCGGCTTCCCGATCTGACTATAGGCAAGCGTGG; tagged-PCR, P5-I5-AATGATACGGCGACCACGACGCTCTTCCGATCT, P7-I7-CAAGCAGGAACCGCTCTTCCGATCTGACTATAGGCAAGCGTGG; tagged-PCR, P5-I5-AATGATACGGCGACCACGACGCTCTTCCGATCT, P7-I7-CAAGCAGGAACCGCTCTTCCGATCTGACTATAGGCAAGCGTGG.

**Data availability.** The datasets generated during and/or analysed during the current study (for example, the recombination/repair assays analysed throughout the paper, with quantitation by FACS) are available from the corresponding author on reasonable request. Source data for the figures are available in Supplementary Fig. 1. The HTGTS datasets (10 datasets, corresponding to a total of 25 independent HTGTS experiments) are deposited in the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi), accession number GSE103624.

52. Welm, B. E., Dijkgraaf, G. J., Bledau, A. S., Welm, A. L. & Werb, Z. Lentiviral transduction of mammary stem cells for analysis of gene function during development and cancer. *Cell Stem Cell* **2**, 90–102 (2008).

53. Puget, N., Knowlton, M. & Scully, R. Molecular analysis of sister chromatid recombination in mammalian cells. *DNA Repair (Amst.)* **4**, 149–161 (2005).
Extended Data Figure 1 | BRCA1 suppresses RAD51-independent Tus–Ter-induced GFP RFP+ repair outcomes. a, Repair frequencies in Brca1fl/exon11 and Brca1Δexon11 6×Ter-HR reporter cells transfected with Tus or I-SceI and with either control Luciferase siRNA (siLUC) or Brca1 SMARTpool (siBRCA1). Columns represent mean of technical duplicate samples from ten independent experiments (that is, \( n = 10 \)). Error bars denote s.e.m. Tus-induced HR, Brca1fl/exon11 cells, \( t \)-test siBRCA1 versus siLUC: all measurements \( P < 0.01 \); Brca1Δexon11 cells, siBRCA1 versus siLUC: total HR: \( P = 0.0470 \); STGC: \( P = 0.0003 \); LTGC: not significant; LTGC/total HR: \( P < 0.0001 \); GFP RFP+: \( P = 0.0010 \). I-SceI-induced HR, Brca1fl/exon11 cells, \( t \)-test siBRCA1 versus siLUC: all measurements \( P < 0.05 \); Brca1Δexon11 cells, \( t \)-test siBRCA1 versus siLUC: all measurements \( P < 0.05 \); Brca1Δexon11 cells, \( t \)-test siBRCA1 versus siLUC: all measurements \( P < 0.02 \). b, Representative primary FACS data for Brca1fl/exon11 and Brca1Δexon11 6×Ter-HR reporter cells transfected with empty vector, Tus or I-SceI and with siLUC or siBRCA1. Tus-transfected samples reproduced from Fig. 1b. FACS plots produced from pooled data of technical duplicate samples from three independent experiments. Numbers represent percentages. c, RT–qPCR analysis of Brca1 mRNA in siRNA-treated cells. Data normalized to Gapdh and expressed as fold difference from siLUC sample from the same experiment (\( \Delta \Delta \text{Ct} = [\text{Ct}_{\text{target}} - \text{Ct}_{\text{Gapdh}}] - [\text{Ct}_{\text{siLUC}} - \text{Ct}_{\text{siGAPDH}}] \)). Error bars denote s.d. of \( \Delta \Delta \text{Ct} \) value (s.d. = \( \sqrt{\text{s.d.}_{\text{target}}^2 + \text{s.d.}_{\text{Gapdh}}^2} \)). d, Frequencies of GFP RFP+ events in Brca1fl/exon11 and Brca1Δexon11 6×Ter-HR reporter cells transfected with Tus or I-SceI and with siLUC, siBRCA1, or Rad51 SMARTpool (siRAD51). Columns represent mean of technical duplicate samples, \( n = 5 \). Error bars denote s.e.m. Tus-induced GFP RFP+, Brca1fl/exon11 cells, \( t \)-test: all comparisons \( P < 0.05 \). Tus-induced GFP RFP+, Brca1Δexon11 cells, \( t \)-test: all comparisons \( P < 0.03 \). Abundance of RAD51 protein in siRNA-treated Brca1fl/exon11 and Brca1Δexon11 6×Ter-HR reporter ES cells. For gel source data, see Supplementary Fig. 1.
Extended Data Figure 2 | Examples of breakpoint sequence analysis of Tus–Ter-induced GFP RFP products. Class 1 and class 2 rearrangements are microhomology-mediated TDs. a, Structure of the GFP RFP class 1 rearrangement marked with red asterisk in Fig. 2. Cartoon elements as in Figs 1 and 2; orange triangle represents 6 × Ter array. Right cartoon denote schematic of TD breakpoint. Grey number denote site of Ter-proximal breakpoint relative to Ter array. In this TD clone, this breakpoint is located 333 bp upstream of the first nucleotide of the first Ter site encountered by the rightward replication fork (that is, position +333). Black number denotes number of base pairs of microhomology at the breakpoint (in this clone, microhomology = 2).

Grey arrows identify the orientation of the segments of the TD, relative to the reporter. Top text box, the direct sequence of the TD breakpoint. Green bold text denotes fragments of GFP open reading frame (ORF). Red bold letters denote 2-bp microhomology breakpoint. Black text denotes other reporter sequences. Bottom text box, overlay of TD breakpoint ends (green for GFP sequences and red for 2-bp microhomology breakpoint) on full-length wild-type GFP (grey). © 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

b, Structure of the GFP–RFP class 2 rearrangement marked with blue asterisk in Fig. 2. Blue letter ‘B’ indicates a BglII site retained within the TD breakpoint. Right cartoon, schematic of TD breakpoint, elements as in a. In this TD clone, the Ter-proximal TD breakpoint is located 8 bp downstream of the first nucleotide of the first Ter site encountered by the rightward replication fork (that is, position +8).

Text box, direct sequence of TD breakpoint. Green bold text denotes fragments of GFP ORF. Orange highlighting: 8-bp fragment of first Ter element retained within the TD breakpoint. Red bold letter denotes 1-bp microhomology breakpoint. Blue highlighting denotes BglIII site retained within the TD breakpoint. Black text denotes other reporter sequences. © 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 3 | Specificity of BRCA1 loss on Tus–Ter-induced TDs. a, Tus–Ter-induced and I-SceI-induced TD (GFP−RFP+) products in Brca1fl/exon11 or Brca1Δ/exon11 6× Ter-HR cells depleted of indicated repair proteins. Induction of repair products was calculated relative to siLUC controls (which therefore score as 1). Data represents mean of between eight and ten independent experiments, each experimental data point collected as technical duplicates (replicates: siBRCA1, n = 10; siBARD1, n = 9; siCTIP, n = 9; siBLM, n = 8; siFANCd, n = 9; siBRCA2, n = 8; siFANCA, n = 9; siFANCD2, n = 10; siRAD51, n = 9; siSLX4, n = 9). Error bars denote s.e.m. b, Tus-induced and I-SceI-induced STGC (GFP+RFP−) products in Brca1fl/exon11 or Brca1Δ/exon11 6× Ter-HR cells depleted of repair proteins indicated. Replicates and error bars as in a. c, Representative primary FACS data for Brca1Δ/exon11 6× Ter-HR reporter cells co-transfected with empty vector (EV), Tus or I-SceI expression vectors (as shown) and siRNAs as shown. FACS plots pooled from technical duplicate samples of four independent experiments. Numbers represent percentages. d, RT–qPCR analysis of Blm, Fancm, Brca2, Fanca, Slx4, Ctip and Bard1 mRNA. Data normalized to Gapdh and expressed as a fold difference from siLUC-treated sample from the same experiment (x = −2^ΔΔCt, with ΔΔCt = [Ctarget − CGapdh] − [CsilUC − CGapdh]). Error bars represent the s.d. of the ΔCt value (s.d. = √[s.d. target^2 + s.d. Gapdh^2]).
Extended Data Figure 4 | Tus−Ter-induced TDs in FANCM- or BLM-depleted Brca1Δ/exon11 6 × Ter-HR reporter cells. a, Southern blot analysis of Tus−Ter-induced LTGC and GFP−RFP+ TD products in FANCM or BLM-depleted Brca1Δ/exon11 6 × Ter-HR reporter cells (BglII digest, GFP probe). M, molecular mass marker lane. TD breakpoints were identified by PCR product sequencing. b, Repair frequencies in Brca1fl/exon11 and Brca1Δ/exon11 6 × Ter-HR reporter cells transfected with siLUC, siFANCM, siBLM or siFANCM plus siBLM in combination. Columns represent mean of technical duplicate samples, n = 7. Error bars denote s.e.m. Tus−Ter-induced total HR, Brca1Δ/exon11 cells, t-test: all siFANCM samples versus those with no siFANCM: P < 0.001; Brca1Δ/exon11 cells, t-test: all samples versus siLUC: P < 0.002; siFANCM versus siFANCM plus siBLM: P = 0.0240; siBLM versus siFANCM plus siBLM: P = 0.0294. Tus−Ter-induced TD, Brca1Δ/exon11 cells, t-test: siFANCM or siBLM versus siLUC: P < 0.002; siFANCM versus siFANCM plus siBLM: not significant; siFANCM plus siBLM versus all others: P < 0.0001, I-SceI-induced total HR, Brca1Δ/exon11 cells, t-test: siFANCM versus siBLM: P = 0.0265, I-SceI-induced STGC, Brca1Δ/exon11 cells, t-test: siFANCM versus siLUC or siBLM: P < 0.05; siBLM versus siFANCM plus siBLM: P = 0.0445, I-SceI-induced LTGC: not significant. I-SceI-induced ratio LTGC:total HR, Brca1Δ/exon11 cells, t-test: all samples versus siLUC: P < 0.05; siFANCM versus siBLM: P = 0.0245, I-SceI-induced TD, Brca1Δ/exon11 cells, t-test: all samples versus siLUC: P < 0.02. For gel source data, see Supplementary Fig. 1.
**Extended Data Figure 5 | BRCA2 is not a major suppressor of Tus–Ter-induced TDs.**

**a,** GFP–RFP+ products in Brca16×Ter-HR cells transfected with siFANCM or siBLM alone or together with siBRCA1, siBARD1, siBRCA2 or siRAD51. Columns represent mean of technical duplicate samples, n = 5. Error bars denote s.e.m. Tus-induced TDs, t-test: siFANCM plus siBRCA1 or siBARD1 versus all other samples: P < 0.01. siBLM plus siBRCA1 or siBARD1 versus all other samples: P < 0.03. I-SceI-induced TDs, t-test: all comparisons not significant.

**b,** GFP–RFP+ products in Brca16×Ter-HR cells after depletion of CtIP. Columns represent mean of technical duplicate samples, n = 11. Error bars denote s.e.m. Tus-induced TD t-test: all samples versus siLUC: P < 0.01; siFANCM plus siCtIP versus siCtIP or siFANCM: P < 0.001; siFANCM plus siBRCA1 versus all other siFANCM samples: P < 0.0001; siBLM plus siCtIP versus siBLM: P < 0.0001; siBLM plus siBRCA1 versus all other siBLM samples: P < 0.00001. I-SceI-induced TD t-test: all samples versus siLUC: P < 0.05; siFANCM plus siCtIP versus siCtIP: P = 0.0311; siFANCM plus siBRCA1 versus all other siFANCM samples: P < 0.01; siFANCM plus siCtIP versus siFANCM: not significant; siBLM plus siBRCA1 versus all other siBLM samples: P < 0.01; siBLM plus siCtIP versus siBLM: not significant.

**c,** GFP–RFP+ products in two independently derived Brca2lex1/lex2 single-copy 6×Ter-HR reporter clones transfected with siRNAs as shown. Columns represent mean of technical duplicate samples, n = 8. Error bars denote s.e.m. Clone #3 Tus-induced TD t-test: siFANCM plus siBRCA1 versus all other samples: P < 0.01; siLUC versus siFANCM plus siBRCA2: P = 0.0131; siFANCM versus siFANCM plus siBRCA2: not significant. Clone #56 Tus-induced TD t-test: siFANCM plus siBRCA1 versus all other samples: P < 0.003; siFANCM versus siFANCM plus siBRCA2: not significant. Clone #3 and clone #56 I-SceI-induced TD: not significant.

**d,** RT–qPCR analysis of siRNA-treated Brca2lex1/lex2 6×Ter-HR cells. Data normalized to Gapdh and expressed as a fold difference from siLUC sample (x = 2ΔΔCt, with ΔΔCt = [Ctarget - CsiLUC] - [CsiLUC - CsiGAPDH]). Error bars denote s.d. of the ΔCt value (s.d. = [±sdtarget + sdSIUC] - [sdSIUC + sdGAPDH]). **e,** Brca2 gene structure in Brca2lex1/lex2 reporter cells. Grey boxes denote Brca2 exons. PCR primers a, b, and c indicated by arrows. neo denotes neomycin-resistance gene. Asterisk denotes partial deletion of exon 26. For gel source data, see Supplementary Fig. 1.
Extended Data Figure 6 | Tus–Ter–induced TDs arise by a replicative mechanism involving canonical end-joining. a, Southern blot analysis of aneuploid TD clones (AseI digest of gDNA, full-length GFP probe). Same data as Fig. 4b. Parental Ter–HR reporter (P) marks size of unaltered reporter. b, Southern blot analysis of 19 reclones of aneuploid TD clones (AseI digest of gDNA, full-length GFP probe) that contained a second reporter copy. M, molecular mass; R, original aneuploid clone. Lanes 3–20, 19 independent re-clones. For parental and TD structure, see Fig. 4b.

c, Tus–Ter–induced TDs in FANCM-depleted Xrcc4fl/fl (#8) and Xrcc4Δ/Δ (#11) cells co-transfected with siRNAs shown. Mean of technical duplicates, n = 5. Error bars denote s.e.m. P values from a Student’s t-test apply to both #8 and #11 data unless otherwise stated. siFANCM plus siBRCA1 or siFANCM plus siBARD1 versus all other samples in clone #8: P < 0.02, except for clone #11; siFANCM plus siBRCA1 versus siFANCM plus siRAD51: not significant; siFANCM plus siBRCA1 versus siFANCM plus siBARD1: not significant; siFANCM plus siBRCA2 or siFANCM plus siRAD51 versus siLUC or siFANC: not significant. d, Tus–Ter–induced TDs in BLM-depleted Xrcc4fl/fl (#8) and Xrcc4Δ/Δ (#11) cells co-transfected with siRNAs shown. Mean of technical duplicates, n = 5. Error bars denote s.e.m. P values from a Student’s t-test apply to both #8 and #11 data unless otherwise stated. siBLM plus siBRCA1 or siBLM plus siBARD1 versus all other samples in clone #8: P < 0.05. In clone #11, siBLM plus siBRCA1 or siBLM plus siBARD1 versus siBLM plus siRAD51 or siBLM plus siBRCA2: not significant; siBLM plus siBRCA1 versus siBLM plus siBARD1: not significant. siBLM plus siBRCA2 or siBLM plus siRAD51 versus: siLUC or siBLM: not significant. e, RAD51 western blot in siRNA-treated #8 and #11 cells.
f, RT–qPCR analysis of Fancm, Brca1, Bard1, Blm, and Brca2 mRNA in siRNA-treated #8 and #11 cells. Data normalized to Gapdh and expressed as fold difference from siLUC sample (ΔΔCt, with ΔΔCt = [Ct,target] − [Ct,Gapdh] − [Ct,silUC] − [Ct,silGAPDH]). Error denote bars s.d. of ΔΔCt value (s.d. = s.d.target + s.d.Gapdh). g, RT–qPCR analysis of Brca1, Fancm and Blm mRNA in siRNA-treated Xrcc4Δ/Δ (#11) cells lentivirally transduced with pHIV-EV or pHIV-mXRCC4 (X4). See f for normalization and error bar details. For gel source data, see Supplementary Fig. 1.
Extended Data Figure 7 | Breakpoint analysis of Tus–Ter-induced TDs. a, Span of TDs in Brca1Δexon11 6×Ter-HR reporter siFANCM (121 independent TDs), siBRCA1 (44 independent TDs), or siBLM (66 independent TDs) treatment groups. b, Microhomology usage at breakpoint of Tus–Ter-induced TDs for Brca1Δexon11 cells depleted of FANCM, BRCA1 or BLM. Numbers denote total number of breakpoints with microhomology \( \leq 5 \), excluding untemplated insertions. Grey line denotes expected microhomology usage by chance alone. c, Strand preference of mismatch correction in 14 homeologous breakpoints (that is, microhomology with internal mismatches) of Tus–Ter-induced TDs from Brca1Δexon11 cells transfected with siRNAs shown. ‘C/T’ indicates C–T mismatch. A TD site (that is, Ter-proximal or upstream) that underwent mismatch correction is noted. d, Template switches associated with six TD breakpoints. Cartoon format as in Extended Data Fig. 2a. Light grey arrows identify orientation of TD segments relative to the parental reporter. Grey numbers denote position of Ter-proximal sites relative to first Ter site encountered by rightward fork. Black numbers denote breakpoint microhomology use (bp). Template switch insertions as shown. e, Distribution of Ter-proximal sites of TD breakpoints in Brca1Δexon11 cells for each treatment group, relative to first Ter site encountered by rightward fork. 10-bp binned data. Grey area/orange triangles denote 6×Ter array. Bottom, distribution of Ter-proximal TD sites in Brca1Δexon11 6×Ter-HR reporter cells transfected with siFANCM, siBRCA1 or siBLM. The source data are identical to that used for histograms in the top panels, but has been re-presented as ‘survival’ curves, scoring the probability that a Ter-proximal TD site will be positioned to the right of the nucleotide in question. Hence, all groups at nucleotide position −800 are at 100% and all reach 0% by position +300. Mantel–Cox log-rank statistical tests between all pairs are not significant. f, Distribution of ‘upstream’ sites of TD breakpoints in Brca1Δexon11 cells for each treatment group, relative to splice acceptor of RFP exon B. 100-bp binned data.
Extended Data Figure 8 | Analysis of TD and HTGTS breakpoints.

a, Microhomology usage in HTGTS (+) end breakpoints for Tus–Ter-induced translocations from Brca1Δ/exon11 cells treated with siLUC (655), siFANCM (612), siBRCA1 (548) or siBLM (633) or Brca1fl/exon11 cells treated with siLUC control (636) siFANCM (658), siBRCA1 (403) or siBLM (405) or I-SceI-induced HTGTS breakpoints for Brca1Δ/exon11 cells treated with siFANCM (all: 954; +: 506; −: 403). Breakpoints with insertions or with microhomology use >6 were not included in this analysis. Note that HTGTS breakpoints at Tus–Ter are microhomology skewed in comparison to HTGTS breakpoints at I-SceI.

b, Comparison of distributions of Ter-proximal TD sites and HTGTS (+) end breakpoint distribution for Brca1Δ/exon11 6×Ter cells treated with siFANCM (679), siBRCA1 (630), or siBLM (724). Mantel–Cox log-rank test for TD versus HTGTS: siLUC versus siFANCM, P = 0.0171; siLUC versus siBRCA1, P = 0.0003; siLUC versus siBLM, P < 0.0001; siFANCM versus siBRCA1, P = 0.1528; siFANCM versus siBLM, P = 0.0017; siBLM versus siBRCA1, P = 0.1213. Gehan–Breslow–Wilcoxon log-rank test for HTGTS: siLUC versus siFANCM, P = 0.3108; siLUC versus siBRCA1, P = 0.0099; siLUC versus siBLM, P < 0.0001; siFANCM versus siBRCA1, P = 0.0166; siFANCM versus siBLM, P < 0.0001; siBLM versus siBRCA1, P = 0.0751. 6×Ter array indicated by the grey-shaded region. Orange triangles denote individual Ter sites within the 6×Ter array. Nucleotide position 0 represents first nucleotide of first Ter site encountered by the rightward fork. For all Brca1Δ/exon11 treatment groups and Brca1fl/exon11 cells depleted of FANCM, each sample group represents pooled data from three independent biological replicates. For all other Brca1fl/exon11 treatment groups, data shown are from two pooled biological replicates.
Extended Data Figure 9 | BRCA1 loss in ovarian and breast carcinomas is associated with widespread TDs of approximately 10 kb (group 1 TDs). a, Analysis of 92 human ovarian carcinoma genomes (available through the Australian Ovarian Cancer Study (AOCS); http://www.aocstudy.org) and 560 breast carcinoma (BC) genomes (available through the Wellcome Trust Sanger Institute; http://cancer.sanger.ac.uk/cosmic), including 163 triple-negative breast cancer (TNBC) genomes. For each dataset, samples are sorted on the x axis based on increasing number of somatic TDs. y axis: log10 of TD span (in kilobases) within each cancer genome, with median marked with circle. Samples featuring a TDP group 1 profile are indicated in orange. Abrogation of BRCA1 and BRCA2 (by germline mutation, somatic mutation or promoter methylation), and of CDK12 (by somatic mutation) is noted according to key. b, Top, exact numbers of samples analysed for each dataset and each genetic/genomic subgroup indicated in boxes, with digits colour-coded according to key in a. Orange boxes denote group 1 TDP; white boxes denote not group 1 TDP. The numbers comprise only samples for which the relevant genetic annotation is available. Bar charts show percentages of cancer samples with abrogation of BRCA1 (red) or BRCA2 (blue) among the two cancer subsets with or without a TDP group 1 profile; P values calculated by Fisher's exact test. c, Percentages of cancer samples with (orange) or without (grey) a TDP group 1 profile among the entire datasets and the subsets of samples showing abrogation of BRCA1 (B1m) or BRCA2 (B2m); P values calculated by probability mass function.
Downregulation of BRCA1 expression is the most prominent and consistent transcriptional feature of ovarian and breast carcinomas associated with TDP group 1 profile. Box plots comparing expression levels between cancer samples with (orange) or without (grey) a TDP group 1 profile, relative to nine DNA replication/repair genes, for which a role as potential contributors to the widespread TD formation in cancer has been investigated or suggested. Numbers under each dataset represent number of cancers for which expression data are available. P values calculated by Student’s t-test.
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size
   
   Describe how sample size was determined.
   
   Each experiment contributed one data point to a given value. A sample size of n=6 means that the data shown represents (for example) the mean value of 6 independent experiments. We used the number of experiments to calculate the P value for a given experiment. Statistical methods are reported in the "Methods" section.

2. Data exclusions
   
   Describe any data exclusions.
   
   We excluded experiments in which the transfection efficiency was less than 40%. These were pre-established criteria, based on our previous experience.

3. Replication
   
   Describe whether the experimental findings were reliably reproduced.
   
   Yes, on numerous independent experiments and over an extended period of time, using several distinct cell types (as detailed in the m/s).

4. Randomization
   
   Describe how samples/organisms/participants were allocated into experimental groups.
   
   Most experiments shown entailed parallel treatments of the same cell line. In these cases, randomization is not applicable.

5. Blinding
   
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   
   Not blinded.

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   n/a Confirmed

   - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   - A statement indicating how many times each experiment was replicated
   - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
   - A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   - The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
   - A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   - Clearly defined error bars

   See the web collection on statistics for biologists for further resources and guidance.
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

- Tandem duplication score software is published in Menghi et al. Proc Natl Acad Sci U S A. 2016 Apr 26;113(17):E2373-82. doi: 10.1073/pnas.1520010113. PMID: 27071093.
- Software for HTGTS analysis was published in Hu et al. Nat Protoc. 2016 May;11(5):853-71. doi: 10.1038/nprot.2016.043. Epub 2016 Mar 31. PMID:27031497

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

- No restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

- We used anti-mouse XRCC4 Ab (Abcam) and validated it by comparing wild type cells vs. XRCC4 null cells.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

- We used mouse ES cells described previously:
  1. BRCA1 exon 11 conditional cells as described in Willis et al. Nature. 2014 Jun 26;510(7506):556-9. doi: 10.1038/nature13295. Epub 2014 Apr 28. PMID: 24776801
  2. XRCC4 conditional mouse ES cells, as described in Xie et al Nat Struct Mol Biol. 2009 Aug;16(8):814-8. doi: 10.1038/nsmb.1640. Epub 2009 Jul 26. PMID:19633669

b. Describe the method of cell line authentication used.

- Southern blot, PCR using specific primers and/or western blotting. In the cell lines used, we also reproduced phenotypes that have previously been associated with these cells.

c. Report whether the cell lines were tested for mycoplasma contamination.

- Yes, we use continuous prophylactic treatment in the tissue culture medium and we periodically test for mycoplasma contamination. We discard any cell lines that are shown to be mycoplasma positive. Any mycoplasma contaminations typically cause cell death in the cell culture systems we use.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

- N/A

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

- N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

- N/A
Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

- **Data presentation**

  For all flow cytometry data, confirm that:

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

- **Methodological details**

  5. Describe the sample preparation. All data represents mouse embryonic stem cell lines, readouts of repair reporter assays with measurement of GFP and RFP fluorescence.
  6. Identify the instrument used for data collection. Becton Dickinson 5 Laser LSRII
  7. Describe the software used to collect and analyze the flow cytometry data. The FACS data was analyzed using FlowJo software.
  8. Describe the abundance of the relevant cell populations within post-sort fractions. The percentages of the populations are shown.
  9. Describe the gating strategy used. Gating was performed as shown in the figures, centered on the relevant population. Gating strategy is shown directly in Fig. 1b and Extended Data Fig. 3c of the m/s.

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