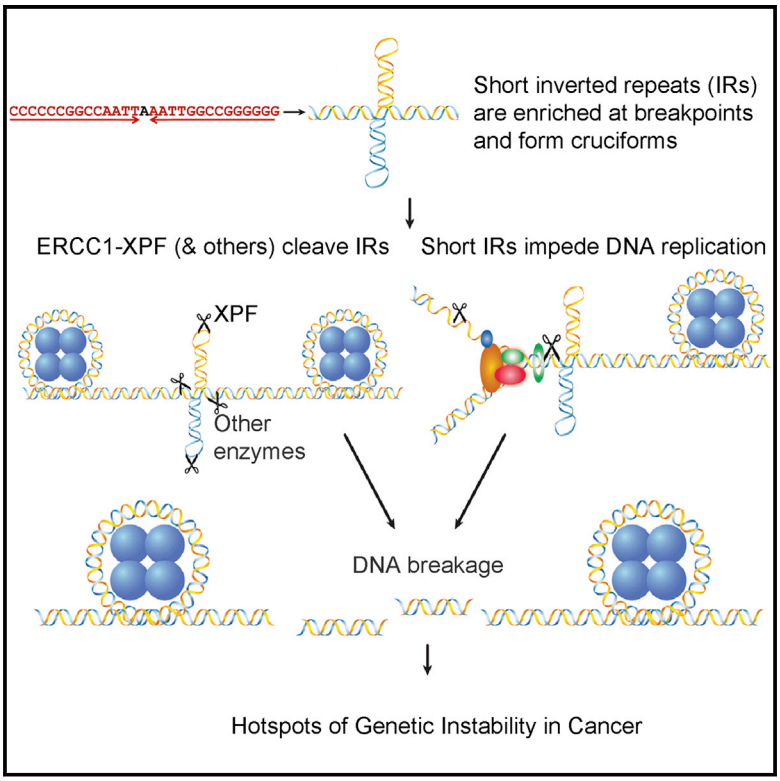
Report

Cell Reports

Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes

Graphical Abstract



Highlights

- Short inverted repeat (IR) sequences are enriched at human cancer breakpoints
- Short IRs stimulate DNA double-strand breaks and deletions in mammalian cells
- Short IRs impede DNA replication forks in mammalian cells
- ERCC1-XPF cleaves IRs and is required for IR-induced chromosome breakage

Authors

Steve Lu, Guliang Wang, ..., Scott Spitser, Karen M. Vasquez

Correspondence

karen.vasquez@austin.utexas.edu

In Brief

The significance of short inverted repeats (IRs) in the human genome is unclear. Lu et al. find that short IRs are enriched at translocation breakpoints and provide evidence for "replication-related" and "DNA-structure-specific cleavage" models of IR-induced DNA breaks and mutations in mammalian cells, suggesting a role in cancer-associated genomic instability.

Lu et al., 2015, Cell Reports *10*, 1674–1680
March 17, 2015
http://dx.doi.org/10.1016/j.celrep.2015.02.039

OPEN ACCESS
CellPress

Cell Reports
Report

Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes

Steve Lu,[1,2] Guliang Wang,[1,2] Albino Bacolla,[1] Junhua Zhao,[1] Scott Spitser,[1] and Karen M. Vasquez[1,*]
[1]Division of Pharmacology and Toxicology, College of Pharmacy, The University of Texas at Austin – Dell Pediatric Research Institute, 1400 Barbara Jordan Boulevard R1800, Austin, TX 78723, USA
[2]Co-first author
*Correspondence: karen.vasquez@austin.utexas.edu
http://dx.doi.org/10.1016/j.celrep.2015.02.039

SUMMARY

Analyses of chromosomal aberrations in human genetic disorders have revealed that inverted repeat sequences (IRs) often co-localize with endogenous chromosomal instability and breakage hotspots. Approximately 80% of all IRs in the human genome are short (<100 bp), yet the mutagenic potential of such short cruciform-forming sequences has not been characterized. Here, we find that short IRs are enriched at translocation breakpoints in human cancer and stimulate the formation of DNA double-strand breaks (DSBs) and deletions in mammalian and yeast cells. We provide evidence for replication-related mechanisms of IR-induced genetic instability and a novel XPF cleavage-based mechanism independent of DNA replication. These discoveries implicate short IRs as endogenous sources of DNA breakage involved in disease etiology and suggest that these repeats represent a feature of genome plasticity that may contribute to the evolution of the human genome by providing a means for diversity within the population.

INTRODUCTION

Genetic analyses of cancer-related genetic instability events have detected regions of the human genome that are hypersusceptible to breakage, which can lead to the deregulation of oncogenes and/or inactivation of tumor suppressors (Popescu, 2003). Interestingly, many such regions contain sequences that can adopt alternative conformations (i.e., non-B DNA), and several of these conformations have been shown to be sources of genetic instability (Kurahashi et al., 2004; Nasar et al., 2000; Wang and Vasquez, 2006), yet the underlying mechanisms are not clear.

Cruciform DNA structures can form at inverted repeat (IR) sequences, where the two symmetric sequences undergo intrastrand base-pairing. Long IRs (>500 bp), which are rare in the human genome, have been found at sites of gross chromosomal rearrangements (Kurahashi and Emanuel, 2001) and can cause DSBs, stimulating recurrent constitutional translocations in sperm leading to chromosomally unbalanced offspring (Ho et al., 2012; Kurahashi and Emanuel, 2001; Kurahashi et al., 2004; Tanaka et al., 2006). Long IRs can also contribute to deletions, recombination, and gene amplifications (Akgün et al., 1997; Cunningham et al., 2003; Gordenin et al., 1993; Losch et al., 2007; Mizuno et al., 2009; Nag and Kurst, 1997; Tanaka et al., 2002; VanHulle et al., 2007; Zhou et al., 2001). In contrast, short perfect IRs are abundant in the human genome; ~80% of all IRs are under 100 bp (Wang and Leung, 2006). However, the mutagenic potential of short IRs has not been well characterized.

This study fills a gap in our understanding of the role of short IRs in genomic instability in mammals by providing evidence that cruciforms formed at short IRs (≤30 bp) can stimulate DSBs by stalling DNA replication forks and/or by activating enzymes (i.e., ERCC1-XPF) that cleave the structures, causing deletions. These findings provide a mechanistic explanation for the co-localization between short IRs and human cancer breakpoints and support the hypothesis that non-B DNA is involved in genetic instability, disease etiology, and evolution.

RESULTS

Short IRs Adopt Cruciform Structures and Induce Genetic Instability

To determine the mutagenic potential of short IRs in mammalian cells, we inserted a 29-bp cruciform-forming IR or a 29-bp control sequence into a *lacZ′* mutation-reporter gene on the vector pUCNIM (Figure 1A). Cruciform formation on the plasmid (pU^+) was confirmed by T7 endonuclease I cleavage (Figure S1). pU^+ and the control pUCON were introduced into mammalian COS-7 cells and screened for mutations 48 hr post-transfection. pU^+ stimulated mutations ~3-fold above that of pUCON (9.2×10^{-3} versus 2.9×10^{-3}: $p < 0.003$; chi-square test, Figure 1B), demonstrating that short IRs are mutagenic in mammalian cells. Restriction digestion and sequencing analyses revealed that >90% of IR-induced mutants consisted of large deletions (>200 bp) ablating the IR sequence (Figures 1C and S2), with some mutants suffering deletions of >2,000 bps (Figure S2A).

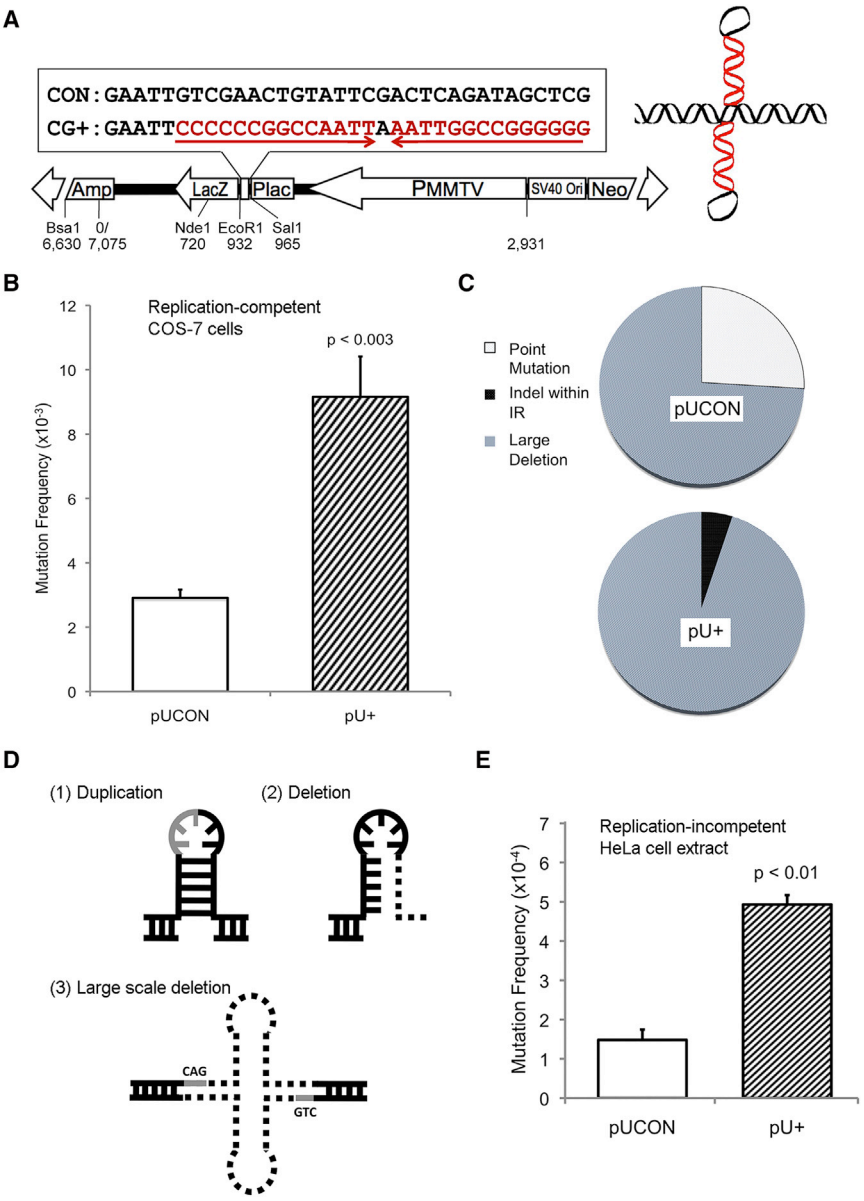


Figure 1. Short IRs Are Mutagenic

(A) Schematic of the reporter plasmids and IR sequence (red); cloning and restriction sites are indicated.

(B) IR-induced mutations on plasmids replicated in COS-7 cells for 48 hr. Error bars indicate SEM of four to five replicates.

(C) Spectrum of IR-induced mutations in mammalian cells.

(D) Schematic diagram of IR-induced mutants.

(E) IR-induced mutagenesis after 6 hr in replication-incompetent HeLa extracts. Error bars represent the SEM of three experiments.

Most deletion mutants contained 1–6 bp of microhomologies at the breakpoint junctions, suggesting that they were generated from DSBs processed by an error-prone microhomology-mediated end joining (MMEJ) mechanism (Figure S2B). The remaining ~10% of IR-induced mutants contained small insertions or deletions (indels) within the IR sequence (Figures 1C and 1D), suggesting a structure-specific cleavage model where the single-stranded cruciform tips represent endonucleolytic targets for a "center-break mechanism" (Lobachev et al., 2007). No point mutations were identified. In contrast, the control (pUCON) mutants contained >20% single base substitutions, with the control insert remaining intact in 74% of the deletion clones (Figure 1C).

Similar results were obtained in replication-incompetent HeLa cell extracts. Here, IRs induced mutations >4-fold above background (4.3×10^{-4} for pU^+ versus 0.9×10^{4} for pUCON: $p < 0.01$; chi-square test, Figure 1E), similar to the levels obtained when DNA replication was enabled by supplementing the extracts with large T antigen (Figure S3). About 60% of the mutations were large deletions (>200 bp) with microhomologies at the breakpoint junctions, suggesting that they were largely products of error-prone DSB repair.

Short IRs Stimulate DSBs in Replication-Competent and -Incompetent Systems

To provide direct evidence that the IRs stimulated DSBs in mammalian cells, ligation-mediated PCR (LM-PCR) was performed (Wang et al., 2006; Wang and Vasquez, 2004) using an upstream primer (~200 bp from the IR) and a primer within the linker. The results revealed a breakpoint hotspot (BH1) at the IR, and another (BH2) 60 bp upstream of the IR (Figure 2A). Sequencing of the PCR products mapped the locations of DSBs near the base of the predicted cruciform stem (1–10 bp upstream of the IR) and ~60 bp upstream of the IR (Figure 2C). In both regions breakpoints were clustered within a small area, consistent with DNA-structure-induced DSBs. By contrast, no distinct DSB hotspots were identified on the control plasmid.

Interestingly, the locations of the IR-specific DSBs in replication-incompetent HeLa extracts were similar to those generated in replication-competent COS-7 cells; both BH1 and BH2 were clearly present, while BH2 was diminished in the replication-competent in vitro system (Figure 2B). These results implicate both replication-independent and replication-related cleavage mechanisms for IR-induced DSBs.

Short IRs Stall DNA Replication Forks In Vivo

To explore plausible mechanisms of IR-induced genetic instability, we examined the ability of short IRs to impede DNA polymerase, as long IRs can stall DNA replication forks in yeast

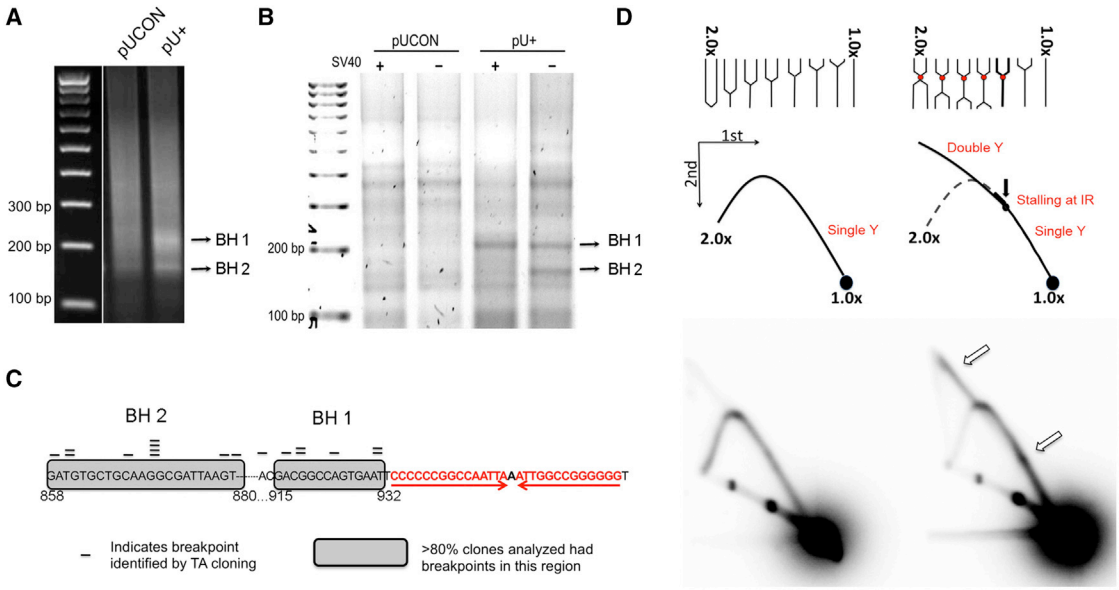


Figure 2. Replication-Related and -Independent Mechanisms of IR-Induced DSBs

(A) Mapping of IR-induced DSBs in replication competent COS-7 cells. LM-PCR amplified regions between the upstream specific primer and the linker (IR-induced breakpoints) were separated on 1.8% agarose gels. Breakpoint hotspots 1 and 2 (BH 1&2) are indicated.

(B) DSBs identified by LM-PCR on plasmids recovered from replication-proficient ($SV40^+$) or -deficient ($SV40^-$) extracts.

(C) Mapping of breakpoints by sequencing of cloned PCR products. Gray lines represent the incidence of breaks at this position.

(D) IRs stall replication forks in COS-7 cells. Upper panel: schematic of a smooth Y-arc from 2D gel electrophoresis for the control plasmid (left), and the bulge on the Y-arc, as well as the double-Y-shaped replication intermediates caused by a replication barrier from the IR-containing plasmid. Lower panel: short IRs stall replication. The arrows designate the bulge on the Y-shaped arc, indicative of stalled replication intermediates, and the double Y-shaped replication intermediates containing IR-stalled replication forks colliding with forks progressing from the opposite direction. A representative image of four independent analyses is shown.

(Voineagu et al., 2008). We performed two-dimensional (2D) gel electrophoresis to separate replication intermediates recovered from mammalian COS-7 cells. After removing unreplicated DNA by DpnI digestion, replication intermediates consisting of the NdeI and BsaI fragments (containing the inserts) of control (pCON) and IR-containing (pS^+) plasmids were analyzed. We detected the expected smooth Y-shaped replication arc on the control plasmid. In contrast, the arc generated from the IR-containing plasmid contained a bulge on the right arm of the Y-arc mapping to the IR site, and a corresponding weaker signal on the left arm, indicating replication fork stalling at the IR sequence (Figure 2D). In addition, the IR-containing samples (pS^+) contained substantially more double Y-arc replication intermediates than the control samples, indicative of stalled replication forks at the IR colliding with replication forks progressing from the opposite direction. This result supports replication stalling-related breakage as a likely mechanism for instability in vivo, although stimulation of recombination by stalled replication could also lead to rearrangements without DSBs (Lambert et al., 2005).

The DNA Repair Complex ERCC1-XPF Cleaves Cruciforms and Is Involved in IR-Induced Genetic Instability

To identify gene products responsible for cleaving cruciforms in vivo, we inserted the IR or control sequences downstream of the selectable *URA3* gene in a yeast artificial chromosome (YAC) and screened for loss of *URA3* function in wild-type BY4742 and a number of gene-deficient yeast strains (Wang et al., 2009). The IR-induced loss of *URA3* function was substantially diminished in the absence of *RAD1*, the yeast homolog of human *XPF* (6.5×10^{-5} versus 2.1×10^{-5}, $p = 0.0018$), suggesting that Rad1p/XPF was required for cruciform-induced DSBs and subsequent genetic instability (Figure 3A).

These results were confirmed in human GM08437B (SV40-transformed XP2YO) isogenic XPF-proficient (XPF^+) and XPF-deficient (XPF^-) cells (Wu et al., 2007), where pU^+ induced mutations 2.7-fold over pUCON (32.6×10^{-4} versus 11.8×10^{-4}) in XPF^+ cells, while induction was reduced to 1.6-fold in XPF^- cells (20.0×10^{-4} versus 12.8×10^{-4}) (Figure 3B). The IR-induced mutation spectra were similar in the presence or absence of XPF, with >90% of mutants undergoing large deletions (data not shown).

Moreover, purified human recombinant ERCC1-XPF cleaved the cruciform structures in vitro, to yield an ~35 nt fragment (Figure 3C), supporting the genetic data. Interestingly, incision by ERCC1-XPF occurred on the 3′ side of the cruciform loop, rather than 5′ of the single-strand/double-strand junction, as predicted based on its cleavage pattern on a stem-loop, a well-established substrate (Figure 3D).

Short IRs Are Enriched at Translocation Breakpoints in Human Cancer Genomes

To support the biological significance of our findings, we investigated whether short IRs promote genetic instability in cancer.

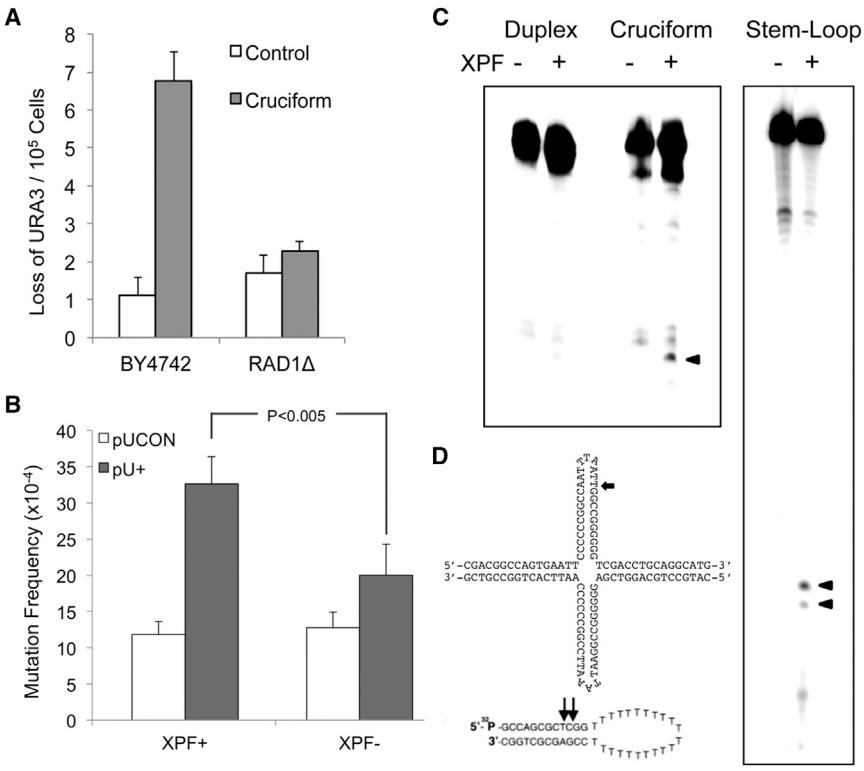


Figure 3. RAD1/XPF Is Involved in IR-Induced Genetic Instability in Eukaryotes

(A) IR-induced loss of *URA3* function is *RAD1* dependent. BY4742 is the wild-type yeast strain. Error bars show the SD of three to four experiments.

(B) IR-induced mutagenesis is reduced in human XPF-deficient (XPF^{-}) cells relative to XPF-proficient (XPF^{+}) cells. Error bars indicate SEM of three replicates. $p < 0.005$ in a chi-square test.

(C) Purified human recombinant ERCC1-XPF cleaves cruciforms (left). A stem-loop (right) served as a positive control (Kuraoka et al., 2000), and a duplex substrate served as a negative control. Cleavage products are indicated by arrows.

(D) Sequences and structures of the substrates used in (C).

We developed a computer program to map such repeats within ±100 bp of 19,956 translocation breakpoints from sequenced cancer genomes (COSMIC at http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/). The average number of 7–30 bp IRs in the ±100 bp regions surrounding the breakpoints was significantly greater than in 20,000 random control sequences extracted from the reference human genome (mean ± SD 28.5 ± 4.7 versus 21.4 ± 5.0; $p = 1.14 \times 10^{-31}$; two-tailed t test), and their location peaked at the cancer breakpoint sites (Figure 4A). Moreover, there were up to five times more IR sequences containing stem lengths of 10–30 bp in cancer translocations than in the control sequences (Figure 4B). Thus, we conclude that short IRs induce chromosomal translocations in human cancer, likely through the ability to extrude into cruciforms.

DISCUSSION

Sequences with the capacity to adopt cruciforms are abundant in the human genome, with IRs ≥8 bp found in ~1 in every 5,600 bp in the human genome (Schroth and Ho, 1995). Herein, we found that short IRs (stem length 10–30 bp) are significantly enriched within 200 bps surrounding 19,956 translocation breakpoints in human cancer genomes. Toward a mechanistic understanding of this co-localization, we provide direct evidence that short IRs can induce genetic instability in mammalian cells. Short IRs stimulated DSBs, resulting in large deletions containing microhomologies at the breakpoint junctions in >90% of the mutants (Figure S2). These findings suggest that IR-induced DSBs are processed by MMEJ, a pathway implicated in chromosomal translocations in cancer (McVey and Lee, 2008). We have identified at least two plausible mechanisms by which IRs can induce DSBs and genetic instability: replication fork stalling and collapse, resulting in DNA replication-dependent mutagenesis; and replication-independent, structure-specific cleavage by ERCC1-XPF and, perhaps, additional nucleases.

It has been shown that long IRs consisting of two inverted Alu sequences (~300 bp) slowed the progression of replication forks in bacteria, yeast, and mammals, promoting DNA strand breaks (Pearson et al., 1996; Voineagu et al., 2008). Here, we found that much shorter IRs of 14 bps can stall replication forks (Figure 2D), which can lead to DNA strand breaks. This provides direct evidence for a replication-related mechanism of short IR-induced genetic instability in vivo.

We also found that short IRs stimulated mutations in the absence of DNA replication, implicating a replication-independent mechanism of IR-induced mutagenesis. Moreover, IR-induced DSBs in replication-incompetent HeLa extracts differed from those on replicating templates (Figures 2A and 2B). These findings strengthen our earlier hypothesis (Wang et al., 2006) that, in addition to replication-related mechanisms, there are replication-independent mechanisms of DNA-structure-induced genetic instability that rely on structure-specific endonuclease cleavage. The identification of mutants containing duplications of the spacer region between the IRs or deletions of half of the IR sequence (Figure 1D) provides further support for structure-specific cleavage of non-B DNA structures. These mutations may have arisen from a "center-break mechanism," where the cruciform tips were targets for endonucleases (Lobachev et al., 2007). These deletion events, together with the results from LM-PCR, suggest that repair nucleases may contribute to IR-induced mutagenesis by structure-specific cleavage in mammalian cells.

Several proteins have been implicated in the mutagenesis induced by long IRs or hairpins formed by CNG repeats, including Holliday junction resolvases (Inagaki et al., 2013; Lewis

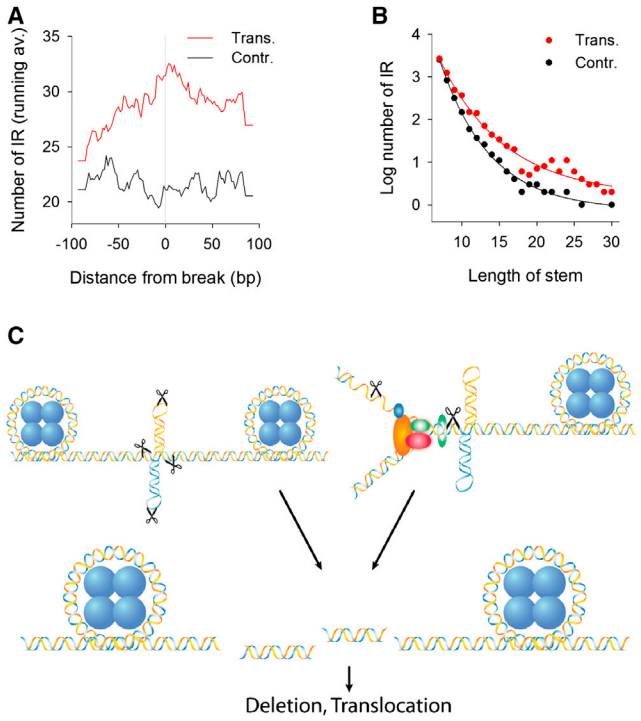


Figure 4. Short IRs Are Associated with Translocations in Cancer Genomes

(A) Running average of IR motifs in the cancer (red) and control (black) data sets whose center loop positions are located at each base along the ±100 bp flanking cancer translocation breakpoints or the control sites.

(B) Distribution of stem lengths for the IRs found in the cancer and control data sets.

(C) Model of replication-related genetic instability where a cruciform impedes DNA replication (right) and replication-independent mechanisms of structure-specific cleavage (e.g., XPF) (left).

and Coté, 2006; Lobachev et al., 2007; Mankouri et al., 2011), Topoisomerase II (Froelich-Ammon et al., 1994), SLX1 and Mus81-Mms4 endonucleases (Agostinho et al., 2013; Ashton et al., 2011; Castor et al., 2013), the MRE11-RAD50-NBS1 complex (Trujillo and Sung, 2001), the MSH2-MSH3 complex (Owen et al., 2005), and nucleotide excision repair factors (Butler et al., 2002). More details can be found in our recent review (Vasquez and Wang, 2013). Whether these proteins and/or other factors, such as transcription, act on short perfect IRs in a fashion similar to that of long IRs warrants further investigation.

We found that lack of XPF in human cells or *Rad1* in yeast significantly reduced IR-induced instability. Although ERCC1-XPF has been implicated in Holliday junction resolution (Agostinho et al., 2013; Oh et al., 2008) and can cleave at the 5′ end of a four-way structure (Agostinho et al., 2013) or 5′ to a stem-loop junction (Figures 3C and 3D), we found that cleavage occurred at the 3′ side of the cruciform tip (Figures 3C and 3D). Whether this behavior is due to local structural features or a property common to short cruciforms remains to be determined. In any case, ERCC1-XPF cleavage on both tips of a cruciform will lead to DSBs. The differences in the predominant cleavage sites identified by cleavage assays (Figure 3C) and LM-PCR analysis (Figures 2A and 2B) may have occurred because LM-PCR identifies the most upstream breakpoints after DSB processing.

Most IRs present in eukaryotic genomes are >60% A+T rich (Schroth and Ho, 1995). We also found that the cancer-associated IRs were more A+T rich than those of the control data set (0.79 ± 0.04 versus 0.72 ± 0.05; $p = 2.39 \times 10^{-45}$) (Figure S4). We performed similar experiments with a plasmid containing the same repeat length (29 bp), but A+T rich, and observed similar results to those obtained with the G+C-rich IR. Moreover, we found that a very short IR (5 bp in each arm) was mutagenic (data not shown). Thus, the effect of IRs on replication and mutagenesis was for the most part dependent on the cruciform structure itself.

Collectively, our data reveal both replication-related and -independent mechanisms for IR-induced mutagenesis. During replication, the duplex is separated into single strands and negative supercoiling is generated, which facilitates cruciform extrusion (Lilley, 1980). Our data support a model in which cruciforms can stall replication forks and/or can be recognized and processed in a structure-dependent fashion, thereby leading to DSBs (Figure 4C). The resulting DSBs are then repaired in an error-generating fashion by MMEJ. Our findings support the conclusion that IR-mediated rearrangements contribute to human disease; thus, a better understanding of these processes may lead to novel strategies to treat or prevent such diseases caused by genetic instability.

EXPERIMENTAL PROCEDURES

Plasmids

A 29-bp IR and a control sequence unable to adopt non-B DNA were cloned into the *lacZ′* mutation-reporter vector pUCNIM at the EcoRI-SalI cassette (Figure 1A). The plasmids were named pU^+ and pUCON, respectively. The four-way junction-specific T7 endonuclease I was used to determine cruciform formation.

Mutagenesis Assays

Plasmid DNA was transfected into COS-7, XPF-proficient or -deficient cells using AMAXA Nucleofector Kit V according to recommended protocols. After 48 hr, plasmids were recovered using a QIAGEN Miniprep kit and treated with DpnI to remove unreplicated plasmids. Mutants were identified through a DH5α blue-white screen, as described (Wang et al., 2006). Mutation frequencies in mammalian cells were adjusted by subtracting the bacterial mutation frequencies. Student's t tests were used to determine statistical significance if not otherwise described.

LM-PCR

LM-PCR was carried out as described (Wang et al., 2006; Wang and Vasquez, 2004). The region between the upstream primer and IR was ~200 bp. Amplified PCR products were separated on 1.8% agarose gels, bands purified using Promega Wizard SV Gel and PCR Clean-Up System, and cloned into the Promega pGEM-T vector for sequencing.

In Vitro Mutagenesis Assays

HeLa cell-free extracts from a CHIMERx DNA replication assay kit were used as a replication-incompetent system. Purified large T antigen was added to allow for replication in extracts. Fifty nanograms of plasmid DNA was incubated for 6 hr at 37°C and then cleaved with DpnI (from replicating extracts). IR-induced mutants were determined by blue-white screening and analyzed by restriction digestion and DNA sequencing.

Two-Dimensional Gel Electrophoresis

The control and IR sequences, identical to those in the mutation reporter plasmids, were inserted into a different backbone and named pCON and pS^+, according to their inserts. Plasmids were transfected into COS-7 cells, and after 24 hr replication intermediates were isolated and separated by 2D gel electrophoresis, as described (Krasilnikova and Mirkin, 2004). DNA was transferred to nitrocellulose membranes, and replication intermediates were identified with radiolabeled probes for the NdeI-BsaI inserts.

IR-induced Loss of *URA3* in Yeast

Loss of *URA3* function in yeast was assessed as described (Wang et al., 2009). Briefly, YACs containing the IR or control sequences were introduced into wild-type yeast strain BY4742, or mutant yeast strains. Yeast cells were plated on 5-fluoroorotic acid (FOA) selective media to screen for loss of *URA3*.

ERCC1-XPF Cleavage Assays

Sixty nanograms of purified human recombinant ERCC1-XPF or BSA was incubated with 4×10^{-8} M DNA substrates (Figure 3D; Supplemental Information) at 30°C for 1 hr in a buffer containing 50 mM Tris (pH 8.0), 750 μM $MnCl_2$, 500 μM DTT, and 20 mM NaCl. The cleavage products were resolved by electrophoresis on 12% denaturing polyacrylamide gels and visualized by a Typhoon PhosphorImager.

Bioinformatic Analyses

The cancer genome translocation breakpoint data set was obtained from COSMIC at http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/. Perfect IR motifs, stem length 7–30 and loop size 0–7 bases, were retrieved using custom scripts; for overlapping IRs, only the longest match was output. The bedtools random utility was used to generate the set of 20,000 non-gap-matching 200-bp sequences.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.02.039.

AUTHOR CONTRIBUTIONS

S.L., G.W., J.Z., S.S., and A.B. performed experiments; S.L., G.W., J.Z., A.B., and K.M.V. analyzed data; K.M.V. and G.W. conceived the project; and S.L., G.W., A.B., and K.M.V. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Drs. Richard Wood for providing purified ERCC1-XPF protein and Rick Finch for critical comments. This work was supported by NIH/NCI grants CA093729 and CA187854 to K.M.V. and NSF grant ACI-1134872 to the Texas Advanced Computing Center (TACC).

Received: August 19, 2014
Revised: January 26, 2015
Accepted: February 16, 2015
Published: March 12, 2015

REFERENCES

Agostinho, A., Meier, B., Sonneville, R., Jagut, M., Woglar, A., Blow, J., Jantsch, V., and Gartner, A. (2013). Combinatorial regulation of meiotic holliday junction resolution in C. elegans by HIM-6 (BLM) helicase, SLX-4, and the SLX-1, MUS-81 and XPF-1 nucleases. PLoS Genet. *9*, e1003591.

Akgün, E., Zahn, J., Baumes, S., Brown, G., Liang, F., Romanienko, P.J., Lewis, S., and Jasin, M. (1997). Palindrome resolution and recombination in the mammalian germ line. Mol. Cell. Biol. *17*, 5559–5570.

Ashton, T.M., Mankouri, H.W., Heidenblut, A., McHugh, P.J., and Hickson, I.D. (2011). Pathways for Holliday junction processing during homologous recombination in Saccharomyces cerevisiae. Mol. Cell. Biol. *31*, 1921–1933.

Butler, D.K., Gillespie, D., and Steele, B. (2002). Formation of large palindromic DNA by homologous recombination of short inverted repeat sequences in Saccharomyces cerevisiae. Genetics *161*, 1065–1075.

Castor, D., Nair, N., Déclais, A.C., Lachaud, C., Toth, R., Macartney, T.J., Lilley, D.M., Arthur, J.S., and Rouse, J. (2013). Cooperative control of holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases. Mol. Cell *52*, 221–233.

Cunningham, L.A., Coté, A.G., Cam-Ozdemir, C., and Lewis, S.M. (2003). Rapid, stabilizing palindrome rearrangements in somatic cells by the center-break mechanism. Mol. Cell. Biol. *23*, 8740–8750.

Froelich-Ammon, S.J., Gale, K.C., and Osheroff, N. (1994). Site-specific cleavage of a DNA hairpin by topoisomerase II. DNA secondary structure as a determinant of enzyme recognition/cleavage. J. Biol. Chem. *269*, 7719–7725.

Gordenin, D.A., Lobachev, K.S., Degtyareva, N.P., Malkova, A.L., Perkins, E., and Resnick, M.A. (1993). Inverted DNA repeats: a source of eukaryotic genomic instability. Mol. Cell. Biol. *13*, 5315–5322.

Ho, B., Baker, P.M., Singh, S., Shih, S.J., and Vaughan, A.T. (2012). Localized DNA cleavage secondary to genotoxic exposure adjacent to an Alu inverted repeat. Genes Chromosomes Cancer *51*, 501–509.

Inagaki, H., Ohye, T., Kogo, H., Tsutsumi, M., Kato, T., Tong, M., Emanuel, B.S., and Kurahashi, H. (2013). Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations. Nat. Commun. *4*, 1592.

Krasilnikova, M.M., and Mirkin, S.M. (2004). Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Mol. Cell. Biol. *24*, 2286–2295.

Kurahashi, H., and Emanuel, B.S. (2001). Long AT-rich palindromes and the constitutional t(11;22) breakpoint. Hum. Mol. Genet. *10*, 2605–2617.

Kurahashi, H., Inagaki, H., Yamada, K., Ohye, T., Taniguchi, M., Emanuel, B.S., and Toda, T. (2004). Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations. J. Biol. Chem. *279*, 35377–35383.

Kuraoka, I., Kobertz, W.R., Ariza, R.R., Biggerstaff, M., Essigmann, J.M., and Wood, R.D. (2000). Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. J. Biol. Chem. *275*, 26632–26636.

Lambert, S., Watson, A., Sheedy, D.M., Martin, B., and Carr, A.M. (2005). Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. Cell *121*, 689–702.

Lewis, S.M., and Coté, A.G. (2006). Palindromes and genomic stress fractures: bracing and repairing the damage. DNA Repair (Amst.) *5*, 1146–1160.

Lilley, D.M. (1980). The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. Proc. Natl. Acad. Sci. USA *77*, 6468–6472.

Lobachev, K.S., Rattray, A., and Narayanan, V. (2007). Hairpin- and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. Front. Biosci. *12*, 4208–4220.

Losch, F.O., Bredenbeck, A., Hollstein, V.M., Walden, P., and Wrede, P. (2007). Evidence for a large double-cruciform DNA structure on the X chromosome of human and chimpanzee. Hum. Genet. *122*, 337–343.

Mankouri, H.W., Ashton, T.M., and Hickson, I.D. (2011). Holliday junction-containing DNA structures persist in cells lacking Sgs1 or Top3 following exposure to DNA damage. Proc. Natl. Acad. Sci. USA *108*, 4944–4949.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. *24*, 529–538.

Mizuno, K., Lambert, S., Baldacci, G., Murray, J.M., and Carr, A.M. (2009). Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. Genes Dev. *23*, 2876–2886.

Nag, D.K., and Kurst, A. (1997). A 140-bp-long palindromic sequence induces double-strand breaks during meiosis in the yeast Saccharomyces cerevisiae. Genetics *146*, 835–847.

Nasar, F., Jankowski, C., and Nag, D.K. (2000). Long palindromic sequences induce double-strand breaks during meiosis in yeast. Mol. Cell. Biol. *20*, 3449–3458.

Oh, S.D., Lao, J.P., Taylor, A.F., Smith, G.R., and Hunter, N. (2008). RecQ helicase, Sgs1, and XPF family endonuclease, Mus81-Mms4, resolve aberrant joint molecules during meiotic recombination. Mol. Cell *31*, 324–336.

Owen, B.A., Yang, Z., Lai, M., Gajec, M., Badger, J.D., 2nd, Hayes, J.J., Edelmann, W., Kucherlapati, R., Wilson, T.M., and McMurray, C.T. (2005). (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. Nat. Struct. Mol. Biol. *12*, 663–670, Corrected in Nat. Struct. Mol. Biol. 2005;12:824 (Gajek, Maciez [corrected to Gajec, Maciej]).

Pearson, C.E., Zorbas, H., Price, G.B., and Zannis-Hadjopoulos, M. (1996). Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. J. Cell. Biochem. *63*, 1–22.

Popescu, N.C. (2003). Genetic alterations in cancer as a result of breakage at fragile sites. Cancer Lett. *192*, 1–17.

Schroth, G.P., and Ho, P.S. (1995). Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. Nucleic Acids Res. *23*, 1977–1983.

Tanaka, H., Tapscott, S.J., Trask, B.J., and Yao, M.C. (2002). Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. Proc. Natl. Acad. Sci. USA *99*, 8772–8777.

Tanaka, H., Bergstrom, D.A., Yao, M.C., and Tapscott, S.J. (2006). Large DNA palindromes as a common form of structural chromosome aberrations in human cancers. Hum. Cell *19*, 17–23.

Trujillo, K.M., and Sung, P. (2001). DNA structure-specific nuclease activities in the Saccharomyces cerevisiae Rad50*Mre11 complex. J. Biol. Chem. *276*, 35458–35464.

VanHulle, K., Lemoine, F.J., Narayanan, V., Downing, B., Hull, K., McCullough, C., Bellinger, M., Lobachev, K., Petes, T.D., and Malkova, A. (2007). Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. Mol. Cell. Biol. *27*, 2601–2614.

Vasquez, K.M., and Wang, G. (2013). The yin and yang of repair mechanisms in DNA structure-induced genetic instability. Mutat. Res. *743-744*, 118–131.

Voineagu, I., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2008). Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. Proc. Natl. Acad. Sci. USA *105*, 9936–9941.

Wang, Y., and Leung, F.C. (2006). Long inverted repeats in eukaryotic genomes: recombinogenic motifs determine genomic plasticity. FEBS Lett. *580*, 1277–1284.

Wang, G., and Vasquez, K.M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proc. Natl. Acad. Sci. USA *101*, 13448–13453.

Wang, G., and Vasquez, K.M. (2006). Non-B DNA structure-induced genetic instability. Mutat. Res. *598*, 103–119.

Wang, G., Christensen, L.A., and Vasquez, K.M. (2006). Z-DNA-forming sequences generate large-scale deletions in mammalian cells. Proc. Natl. Acad. Sci. USA *103*, 2677–2682.

Wang, G., Zhao, J., and Vasquez, K.M. (2009). Methods to determine DNA structural alterations and genetic instability. Methods *48*, 54–62.

Wu, Y., Zacal, N.J., Rainbow, A.J., and Zhu, X.D. (2007). XPF with mutations in its conserved nuclease domain is defective in DNA repair but functions in TRF2-mediated telomere shortening. DNA Repair (Amst.) *6*, 157–166.

Zhou, Z.H., Akgün, E., and Jasin, M. (2001). Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences. Proc. Natl. Acad. Sci. USA *98*, 8326–8333.

Cell Reports

Supplemental Information

Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes

Steve Lu, Guliang Wang, Albino Bacolla, Junhua Zhao, Scott Spitser, and Karen M. Vasquez

SUPPLEMENTAL FIGURES

Figure S1 related to Figure 1

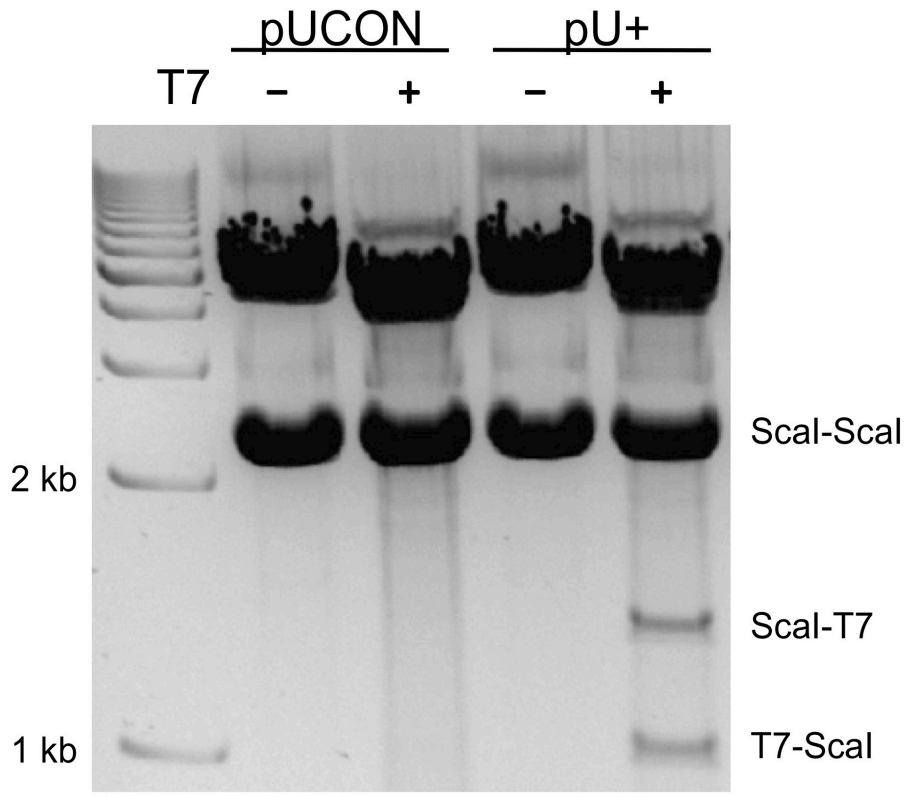
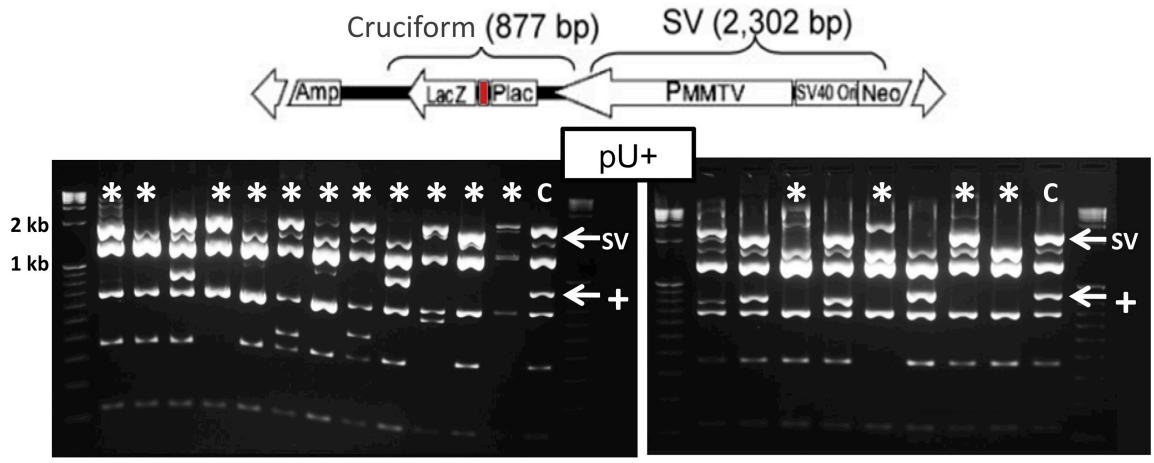


Figure S2 related to Figure 1

A



B

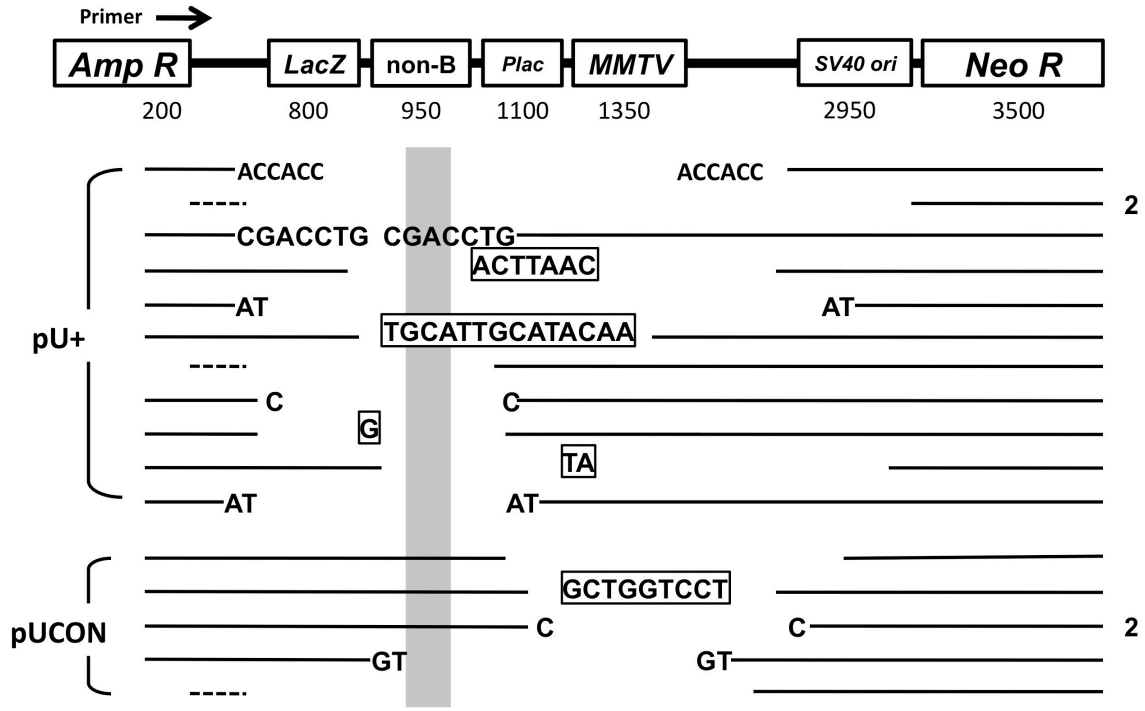


Figure S3 related to Figure 2

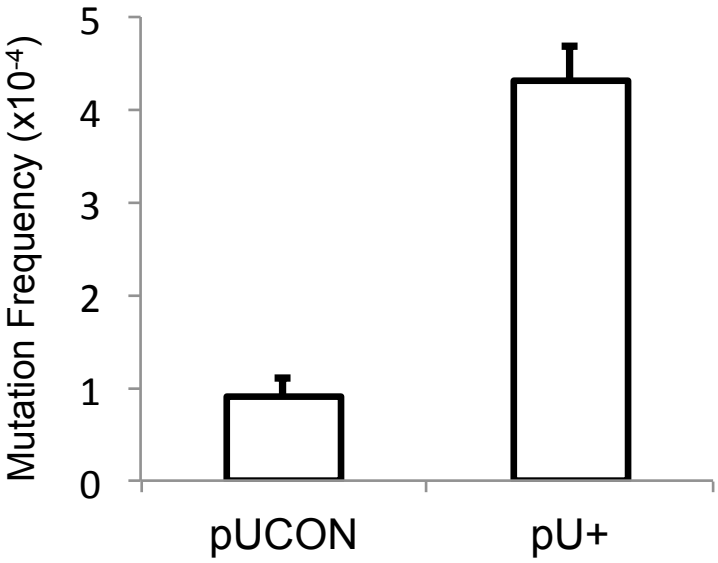
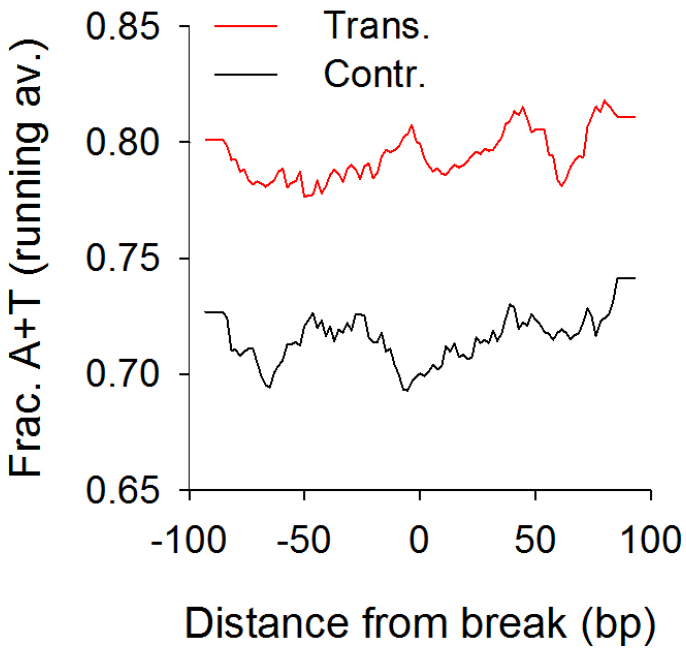


Figure S4 related to Figure 4



SUPPLEMENTAL INFORMATION

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Oligonucleotides Used for Cruciform and Stem-loop Structures in ERCC1-XPF Cleavage Assays. Two 61-nt oligonucleotides, CFO1 and CFO2, used to assemble a cruciform structure contain 29-nt inverted repeats (IRs) in the center that can form a hairpin structure but are not able to anneal to each other, flanked by 16-nt regions on either side that are complementary to each other. CFO1 (5'-CGACG GCCAG TGAAT TCCCC CCGGC CAATT TAATT GGCCG GGGGG TCGAC CTGCA GGCAT G) and CFO2 (5'-C ATGCC TGCAG GTCGA GGGGG GCCGG AATTA AATTC CGGCC CCCCA ATTCA CTGGC CGTCG) were annealed at 95°C followed by slow cooling in DNA annealing buffer (10 mM Tris, pH 8.0, 50 mM NaCl, and 1 mM EDTA). The sequence of the stem-loop structure formed on DSFO1 is: (5'-C ATGCC TGCAG GTCGA CCCCC CGGCC AATTA AATTG GCCGG GGGGA ATTCA CTGGC CGTCG).

SUPLEMENTAL FIGURE LEGENDS

Figure S1: Detection of Secondary DNA Structures by T7 Endonuclease I Sensitivity Assay, Related to Figure 1.

In order to confirm cruciform structure formation at the short IR sequence, plasmids pU+ and pUCON were first treated with T7 endonuclease I, which cleaves at the four-way junction between B-DNA and the cruciform stems, and then digested with ScaI to release ScaI-ScaI fragments containing the IR insert or the control sequence. The 1.2 kb ScaI-T7 and 1 kb T7-ScaI fragments indicate cruciform extrusion at the IR sequence in a subpopulation of plasmids. Thus, our data indicate that the short IR sequence can form a cruciform structure *in vitro*.

Figure S2: Cruciforms Induce Large-Scale Deletions in Mammalian Cells, Related to Figure 1.

(A) Restriction digestion analysis of cruciform-induced *lacZ´* mutants in mammalian COS-7 cells. Twenty randomly picked mutants were subjected to restriction digestion analysis and/or sequencing to reveal the mutation spectra. Plasmids were digested into

7 fragments after incubation with Eag1 and BssS1. The 877-bp fragment (labeled with a '+') contains the cruciform-forming sequence, and the 2,302-bp fragment (labeled 'SV') contains the SV40 origin. The lanes marked "C" contain DNA isolated from wild-type colonies and are used as controls.

(B) Schematic diagram of deletion mutants. The majority of cruciform-induced mutants contained deletions that included the IR sequences. Letters that are boxed indicate insertions within the deleted regions. Numbers on the right designate the number of identical mutants. Dotted lines represent deletions that occurred near the primer sequence. Unboxed letters represent microhomologies. The grey rectangle shows the location of the IR.

Figure S3: Cruciform-Induced Mutagenesis in Replication-Proficient HeLa Cell Extracts, Related to Figure 2.

Plasmids pU+ and pUCON were incubated for 6 h in HeLa cell extracts supplemented with SV40 large T antigen to support replication. The plasmids were then purified and subjected to blue-white screening to detect mutations that inactivated the *lacZ´* reporter gene. The results indicate that pU+ induced mutations ~3-fold above that of the control plasmid, pUCON (4.9×10^{-4} vs. 1.5×10^{-4}: $P < 0.02$).

Figure S4: IRs at Translocation Breakpoints in Cancer Genomes are A+T-Rich, Related to Figure 4.

Running average of A+T fractions for the IRs along ±100-bp flanking cancer translocation breakpoints (red) or the randomly picked control sites (black). The results show that cancer-associated IRs were more A+T-rich than those of the control set (mean ± SD, 0.79 ± 0.04 versus 0.72 ± 0.05; $P = 2.39 \times 10^{-45}$).