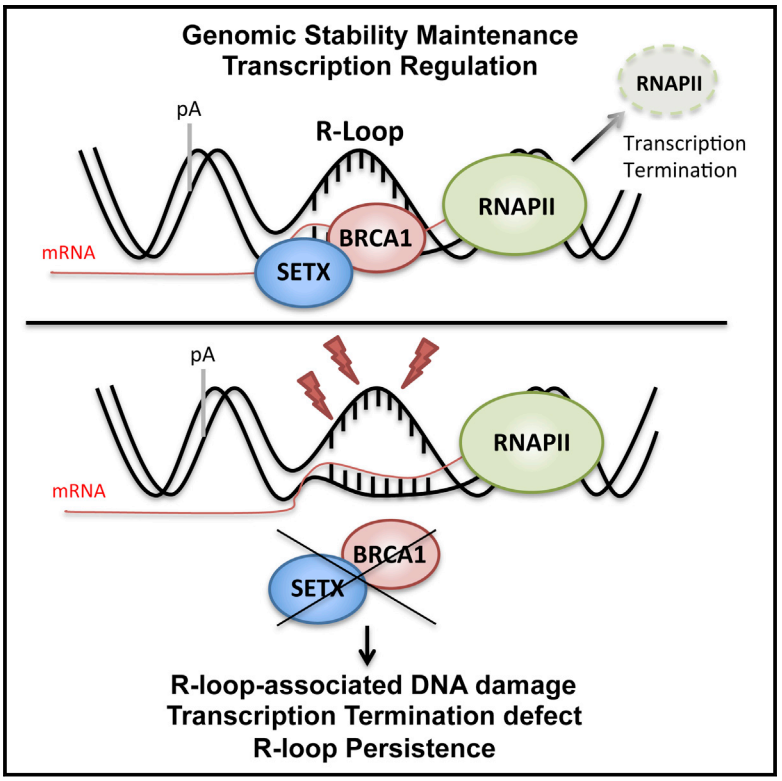
Article

Molecular Cell

BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair

Graphical Abstract



Highlights

- Endogenous BRCA1 and senataxin (SETX) interact in a BRCA1-driven process
- BRCA1/SETX complexes are recruited to R-loop-associated termination regions (TRs)
- BRCA1/SETX complexes suppress transcriptional DNA damage arising at nearby R-loops
- BRCA1 breast cancers reveal indel mutations near BRCA1 TR binding regions

Authors

Elodie Hatchi,
Konstantina Skourti-Stathaki, ...,
Nicholas J. Proudfoot,
David M. Livingston

Correspondence

elodiey_hatchi@dfci.harvard.edu (E.H.),
david_livingston@dfci.harvard.edu
(D.M.L.)

In Brief

Transcriptional R-loops represent a potential threat to genome integrity. Hatchi et al. show that BRCA1, in partnership with SETX, is engaged in a DNA repair mechanism that deals with R-loop-associated genomic instability at transcriptional termination pause sites.

Hatchi et al., 2015, Molecular Cell *57*, 636–647
February 19, 2015
http://dx.doi.org/10.1016/j.molcel.2015.01.011

CellPress

CellPress

Molecular Cell
Article

BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair

Elodie Hatchi,[1,2,*] Konstantina Skourti-Stathaki,[3] Steffen Ventz,[4,5] Luca Pinello,[4,5] Angela Yen,[6,7] Kinga Kamieniarz-Gdula,[3] Stoil Dimitrov,[1,2] Shailja Pathania,[1,2] Kristine M. McKinney,[1,2] Matthew L. Eaton,[6,7] Manolis Kellis,[6,7] Sarah J. Hill,[1,2] Giovanni Parmigiani,[4,5] Nicholas J. Proudfoot,[3] and David M. Livingston[1,2,*]

[1]Department of Genetics, Harvard Medical School, Boston, MA 02215, USA
[2]Department of Cancer Biology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA
[3]Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK
[4]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA
[5]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA
[6]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[7]Computer Science and Artificial Intelligence Laboratory (CSAIL), MIT, Cambridge, MA 02139, USA
*Correspondence: elodiey_hatchi@dfci.harvard.edu (E.H.), david_livingston@dfci.harvard.edu (D.M.L.)
http://dx.doi.org/10.1016/j.molcel.2015.01.011

SUMMARY

The mechanisms contributing to transcription-associated genomic instability are both complex and incompletely understood. Although R-loops are normal transcriptional intermediates, they are also associated with genomic instability. Here, we show that BRCA1 is recruited to R-loops that form normally over a subset of transcription termination regions. There it mediates the recruitment of a specific, physiological binding partner, senataxin (SETX). Disruption of this complex led to R-loop-driven DNA damage at those loci as reflected by adjacent γ-H2AX accumulation and ssDNA breaks within the untranscribed strand of relevant R-loop structures. Genome-wide analysis revealed widespread BRCA1 binding enrichment at R-loop-rich termination regions (TRs) of actively transcribed genes. Strikingly, within some of these genes in BRCA1 null breast tumors, there are specific insertion/deletion mutations located close to R-loop-mediated BRCA1 binding sites within TRs. Thus, BRCA1/SETX complexes support a DNA repair mechanism that addresses R-loop-based DNA damage at transcriptional pause sites.

INTRODUCTION

Germline mutations of the breast cancer susceptibility gene 1 (*BRCA1*) significantly increase the risk of developing breast and ovarian cancers (Tutt and Ashworth, 2002). Since its discovery, researchers have intensively investigated the mechanisms by which full-length BRCA1 (p220) functions as a tumor suppressor.

BRCA1 is engaged in numerous direct and indirect physical interactions with specific partner proteins (Huen et al., 2010). As part of several complexes, BRCA1 contributes to double-strand DNA break (DSB) repair by homologous recombination (HR), stalled fork repair, cell-cycle checkpoint activation, transcription regulation, heterochromatin maintenance, mitotic spindle formation, RNA splicing control, and estrogen metabolism (Gardini et al., 2014; Gorski et al., 2011; Mullan et al., 2006; Pathania et al., 2011; Savage et al., 2014a, 2014b; Venkitaraman, 2014; Zhu et al., 2011). Many of these BRCA1 properties and, in particular, those that protect genome integrity, probably contribute to its tumor suppressing function (Silver and Livingston, 2012; Tutt and Ashworth, 2002). However, in the absence of a coherent underlying mechanism, there is still no definitive evidence of which specific BRCA1 functions are required for breast and ovarian cancer suppression (Huen et al., 2010; Venkitaraman, 2014). Therefore, identifying new BRCA1 binding partners and their associated functions may yield valuable insights.

Our laboratory has identified senataxin (SETX) as a putative BRCA1 binding partner in a yeast two hybrid and several, independent TAP-MS-based genome-wide BRCA1/protein interaction screens (Hill et al., 2014; and D.M.L., unpublished data). Disruption of the *Setx* gene in mice leads to a defect in spermatogenesis, caused by failure of meiotic recombination (Becherel et al., 2013). *Sen1*, the *SETX* yeast homolog, was shown to contribute to the processing of various RNA species and to the distribution of RNA polymerase II (RNAPII) across the genome (Mischo et al., 2011; Steinmetz et al., 2006; Ursic et al., 1997). This probably occurs via its direct interaction with RNAPII and certain RNA processing factors (Suraweera et al., 2009).

While transcription is an essential cellular process, it also represents a potential threat to genome integrity (Kim and Jinks-Robertson, 2012). Several studies indicate that highly transcribed genes exhibit increased rates of mutation and illegitimate recombination (Gaillard et al., 2013). Moreover, a large body of evidence indicates that mutations in certain factors involved at the interface of transcription and RNA processing are associated with genomic instability (Bhatia et al., 2014; Chan et al., 2014; Kleiman and Manley, 1999; Kleiman et al., 2005; Li and Manley, 2006; Stirling et al., 2012). An emerging view is that these mutants contribute to the above-noted phenomena through a common mechanism, which induces the abnormal persistence of co-transcriptional R-loops

(three-stranded structures, each consisting of an RNA:DNA hybrid plus the coding strand DNA). Although R-loops are a naturally occurring consequence of transcription and are essential for diverse cellular events (Skourti-Stathaki and Proudfoot, 2014), they can be potentially deleterious to some cellular functions and compromise genome integrity (Aguilera and García-Muse, 2012; Hamperl and Cimprich, 2014). Indeed, unresolved R-loop structures can expose the displaced, coding ssDNA to nicking and/or other forms of damage (Daniel and Nussenzweig, 2013; Wimberly et al., 2013), as well as impair transcription (Aguilera, 2002; Huertas and Aguilera, 2003) and DNA replication fork progression (Gan et al., 2011; Helmrich et al., 2011).

Interestingly, *SETX* is involved in RNAPII transcription termination and resolves R-loops that form at G-rich transcription pause sites (Skourti-Stathaki et al., 2011). It also associates with processing replication forks and facilitates their progression through RNAPII transcribed genes by displacing R-loops (Alzu et al., 2012). In part through its genetic interaction with DNA repair genes involved in HR, senataxin also protects the genome from transcription-associated instability (Mischo et al., 2011; Ursic et al., 2004). Similarly, SETX, by resolving R-loops at sites of transcription and replication collision, is engaged at the interface of replication stress, transcription, and DNA damage (Yüce and West, 2013).

Interestingly, BRCA1-containing complexes restrict DNA damage induced by aberrant transcription or RNA processing via proposed interactions with multiple transcription and RNA processing factors, including RNAPII (Anderson et al., 1998; Bennett et al., 2008; Kawai and Amano, 2012; Kleiman and Manley, 1999; Kleiman et al., 2005; Savage et al., 2014a; Scully et al., 1997).

In view of these associations, we have asked whether BRCA1 plays a significant role in the repair of R-loop-associated DNA damage arising at termination sites. We find that BRCA1 and SETX form a physiological complex, recruited in a BRCA1-dependent manner to a subset of transcription termination pause sites of highly transcribed genes. There they act to suppress co-transcriptional R-loop-associated DNA damage. Unexpectedly, in breast tumor tissues carrying inherited BRCA1 mutations, insertion/deletion somatic mutations were found in the vicinity of BRCA1-bound gene termination sites where BRCA1 normally engages in the repair of R-loop-associated DNA damage.

RESULTS

Identification of BRCA1 as a Scaffolding Partner for SETX at the β-actin Transcription Termination Site

To investigate whether BRCA1 is involved in R-loop-driven DNA damage responses, we first assessed the physiological relevance of a BRCA1 and SETX interaction, recently identified in our proteomic screens (Hill et al., 2014) and suggested by others (Becherel et al., 2013). Endogenous BRCA1 and SETX co-immunoprecipitation (co-IP) was carried out in HeLa cell extracts, using two, different mono-specific antibodies (Abs) (Figures S1A and S1B). IP of each protein revealed significant co-IP of the other. Irrelevant IgG gave negative results. These results imply the existence of endogenous BRCA1/SETX-containing complexes in these cells (Figure 1A). Weak or undetectable co-IP followed RNAi-driven BRCA1 and SETX depletion, thereby validating Ab specificity (Figure S1C). Two-way BRCA1/SETX co-IP was also apparent in primary, diploid human BJ-hTERT fibroblast extracts (Figure 1A, bottom), suggesting the existence of a physiological interaction between BRCA1 and SETX.

In an effort to map BRCA1 and SETX domains that participate in their interaction, we tested multiple, GST-tagged recombinant fragments that contain various BRCA1 and SETX sequences in glutathione S-transferase (GST) binding assays (Scully et al., 1997; Suraweera et al., 2009) (Figure S1D). The results showed that SETX interacts with F6C, which contains a BRCT (BRCA1 C-terminal) motif (Figure 1B). Intriguingly, F6, which also includes F6C, did not interact with SETX, although F6 was sufficiently folded to interact with another BRCT domain binding protein, BACH1 (Figure S1E) (Cantor et al., 2001). Reciprocal SETX mapping revealed that fragment S8 interacted with BRCA1 (Figure 1C). These results imply that specific structural units of both proteins participate in their interaction. A more extensive analysis will be required to obtain a full picture of this process.

In light of the existence of BRCA1/SETX binding, we tested for BRCA1 co-recruitment at transcription termination sites associated with R-loops by chromatin immunoprecipitation (ChIP) performed on the endogenous human β-actin gene (Figure 1D) (Skourti-Stathaki et al., 2011). Cross-linked HeLa cell chromatin was harvested before and after siRNA depletion of either BRCA1 or SETX. Depletion efficiency was verified at the protein level (Figure 1E).

Specifically, we found that in mock siRNA-transfected cells BRCA1 binding was significantly enriched (3- to 6-fold) at the transcription termination site (5′ pause and pause site probes) when compared to the signals obtained following BRCA1 depletion (Figure 1F). Notably, while BRCA1 depletion suppressed these binding signals, SETX depletion did not. We also confirmed the existence of SETX accumulation at the β-actin transcription termination pause site (Skourti-Stathaki et al., 2011) (Figure 1G). BRCA1 depletion impaired SETX binding to these sites, implying that its recruitment is BRCA1 dependent at this locus. Similar results were obtained when BRCA1 and SETX hairpins were substituted for cognate siRNA species (Figures S2A and S2B). These findings demonstrate a physiological BRCA1/SETX interaction at the β-actin transcription termination site and suggest that BRCA1 acts as an anchor for SETX therein.

R-Loop Formation at the β-actin Gene Transcription Termination Site Triggers BRCA1/SETX Complex Recruitment

To search for a connection between the formation of R-loops and the recruitment of BRCA1/SETX at the β-actin termination site, we first asked whether RNA:DNA hybrids forming within this gene are affected by BRCA1 and SETX depletion. We measured R-loop formation over the β-actin gene, employing DNA:RNA hybrid immunoprecipitation (DIP) of uncross-linked DNA using a monoclonal Ab specific for RNA:DNA hybrids (S9.6) (Skourti-Stathaki et al., 2011). As shown in Figure 2A, a modest, but reproducible, increase in R-loop abundance was detected within the β-actin gene pause site, following BRCA1 or SETX depletion. While this result suggests that BRCA1 and SETX affect R-loop formation, it is important to note that BRCA1 depletion also elicited a phenotype suggestive of a

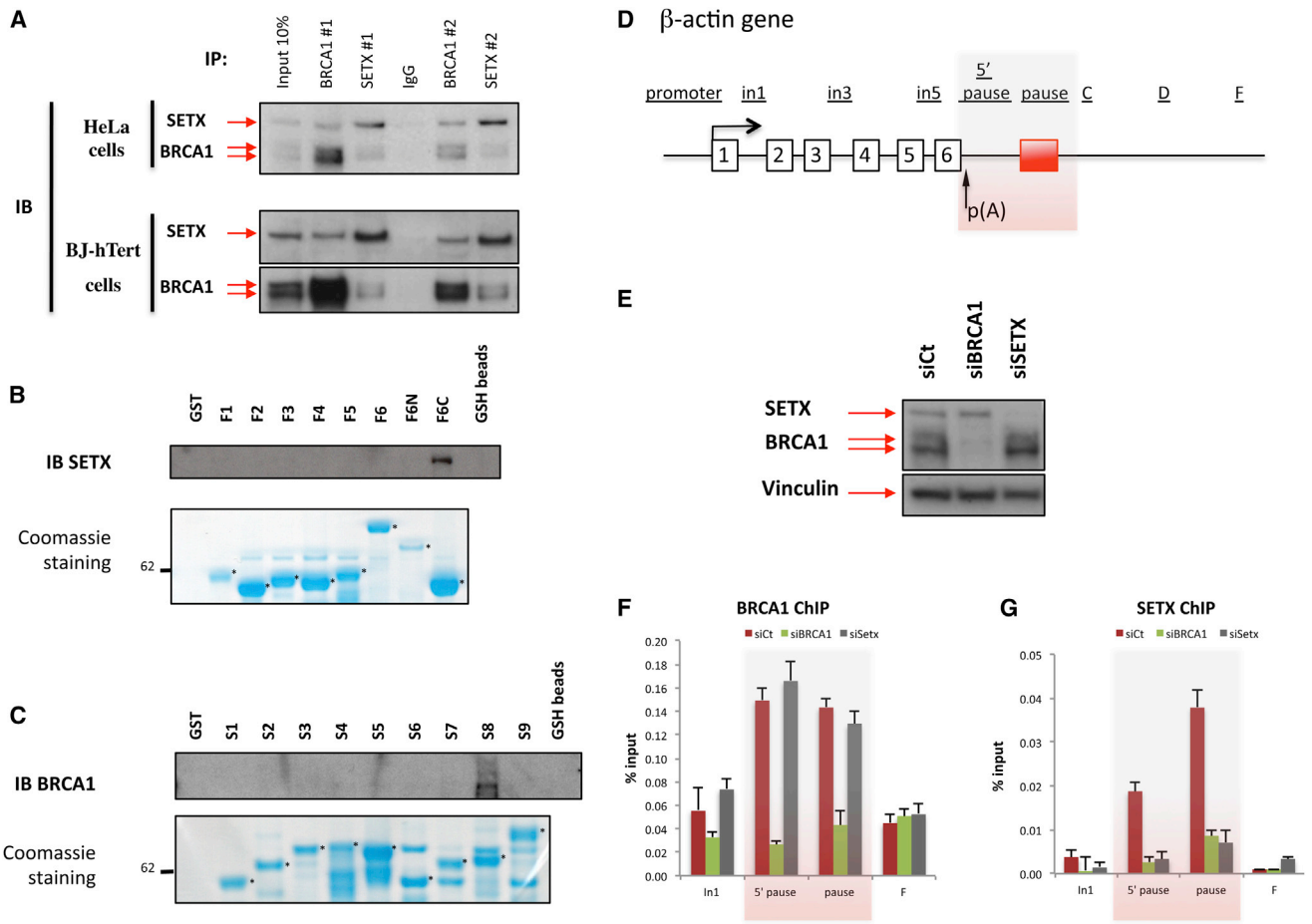


Figure 1. BRCA1 Interacts with SETX and Is Required for SETX Recruitment to the R-Loop-Associated Termination Pause Region of the Human β-actin Gene

(A) Co-immunoprecipitation (IP) of endogenous BRCA1 and SETX in HeLa cell (top) and in BJ-hTert human fibroblast extracts (bottom), using antibodies against BRCA1 (BRCA1 #1/#2) or SETX (SETX #1/#2). IgG, negative control.

(B) In vitro interaction assay using recombinant GST-BRCA1 fragments performed in HeLa cells. GSH bead-bound proteins were eluted and analyzed by immunoblot. Bottom: relative abundance of each recombinant affinity purified fusion protein (marked with a star) visualized by SDS-PAGE after Coomassie staining. Additional bands are readthrough and/or degradation products of recombinant proteins.

(C) Same as in (B) but with recombinant GST-SETX fragments.

(D) Schematic representation of the human β-actin gene: exons are numbered, and the red box reflects the existence of a transcription pause element. Location of primers used for the ChIP experiments are depicted above the diagram.

(E) Immunoblot that reflects the efficiency of the siRNA-mediated depletion of BRCA1 and SETX. Top: SETX and BRCA1; bottom: Vinculin used as a loading control.

(F) ChIP analyses performed on the β-actin gene from mock-treated (red bars), BRCA1- (green bars), or SETX-depleted cells (gray bars) using BRCA1 Ab #3. Average ChIP values ± SD from three independent experiments are shown.

(G) Same ChIP experiments as in (F) but with SETX Ab #1.

See also Figures S1 and S2.

transcription termination defect that was previously reported upon SETX depletion. It was manifest by RNAPII accumulation at the pause site (Skourti-Stathaki et al., 2011) (Figure S2C). The recruitment of other RNA:DNA helicases in the absence of SETX to resolve R-loops over these regions, as well as potential differences in transcription levels after BRCA1 and SETX depletion, might explain the lack of a more substantial increase in R-loops levels.

We then tested BRCA1 and SETX chromatin occupancy over the β-actin gene following R-loop suppression. ChIP experiments were performed in HeLa cells before and after overexpression of GFP-tagged RNaseH1 (Figure S3A), which significantly reduced R-loop formation over the β-actin gene (Skourti-Stathaki et al., 2014). In RNaseH1-overexpressing cells, we observed a significant suppression of BRCA1 and SETX binding at the transcription termination site, as compared with control cells (Figure 2B). Moreover, as shown in Figure 2C, RNaseH1 overexpression led to a reduction in the co-IP signal, implying that BRCA1/SETX complex formation at these genomic loci is mediated by R-loops.

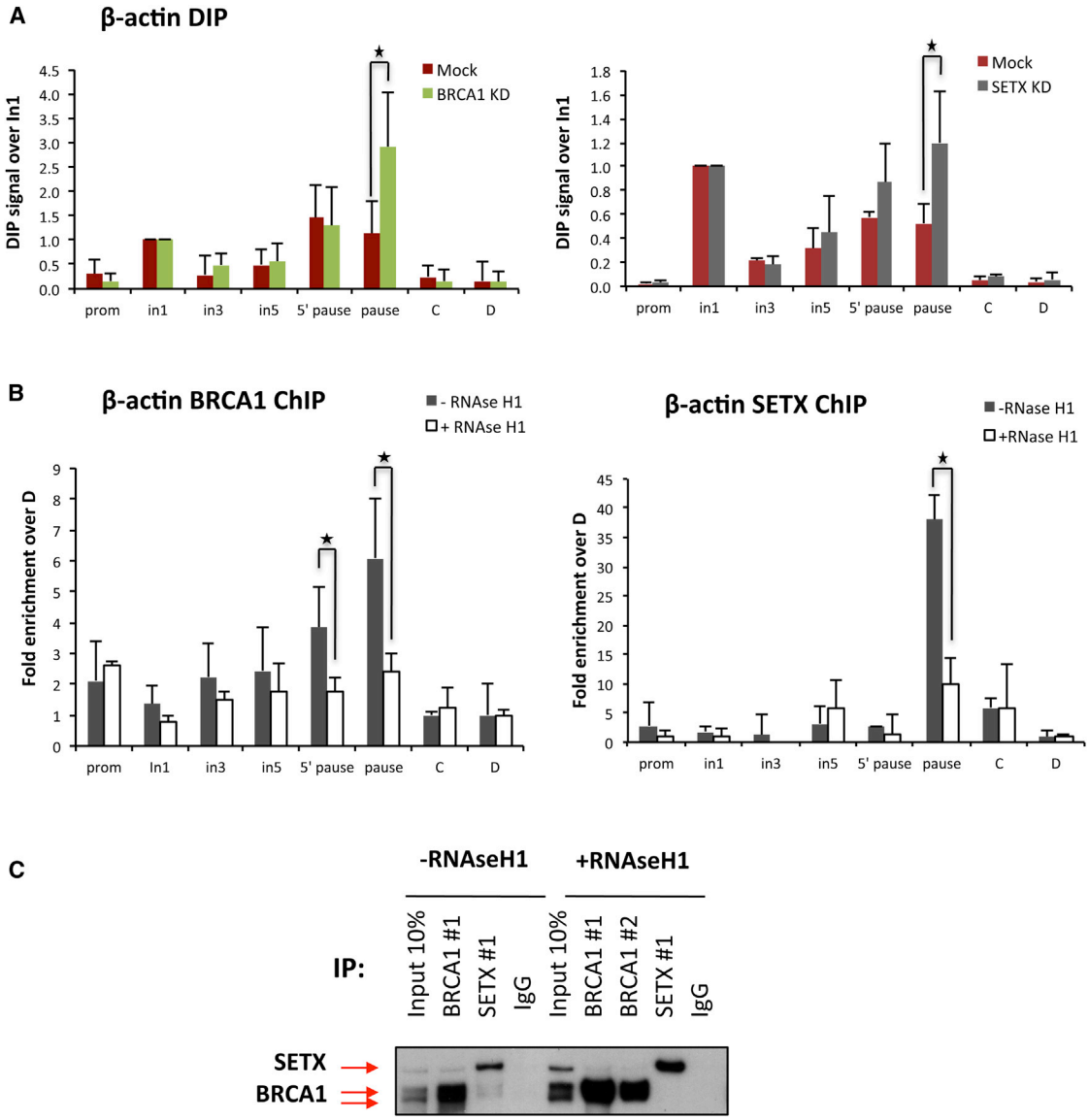


Figure 2. BRCA1 Recruitment to the β-actin Termination Pause Site Is Mediated by R-Loops

(A) RNA:DNA immunoprecipitation (DIP) analyses performed in HeLa from mock-treated or BRCA1-depleted cells (BRCA1 KD) (left) and mock-treated or SETX-depleted cells (SETX KD) (right).

(B) BRCA1 and SETX ChIP experiment performed in control conditions (−RNaseH1) or with RNaseH1 overexpression (+RNaseH1). Average DIP and ChIP values ± SD from three, independent experiments are shown. *$p < 0.05$.

(C) BRCA1 and SETX co-IP experiments performed in HeLa cells ± RNaseH1 expression. IgG, negative control.

Thus, our results suggest that BRCA1/SETX complexes bind at termination region (TR) sites in response to local R-loop formation and that they also suppress R-loop abundance, in part, by participating in the regulation of transcription and of transcription termination.

BRCA1/SETX Complexes Protect DNA at R-Loop-Associated Termination Pause Sites from the Development of ssDNA Breaks

Since R-loop accumulation can increase the risk of genomic instability by promoting DNA damage (Hamperl and Cimprich, 2014; Skourti-Stathaki and Proudfoot, 2014), we tested whether BRCA1 and/or SETX depletion triggers the generation of DNA damage in the vicinity of persistent R-loops. Specifically, γH2AX ChIP assays were performed over the β-actin gene. After depletion of either BRCA1 or SETX, we observed significantly increased signals of γH2AX largely restricted to the termination region (Figure 3A). This implied that loss of BRCA1/SETX complexes over the R-loop-associated pause site contributed to the accumulation of local damage.

Since several studies suggest that the ssDNA of an R-loop is prone to transcription-associated mutations or DNA breaks

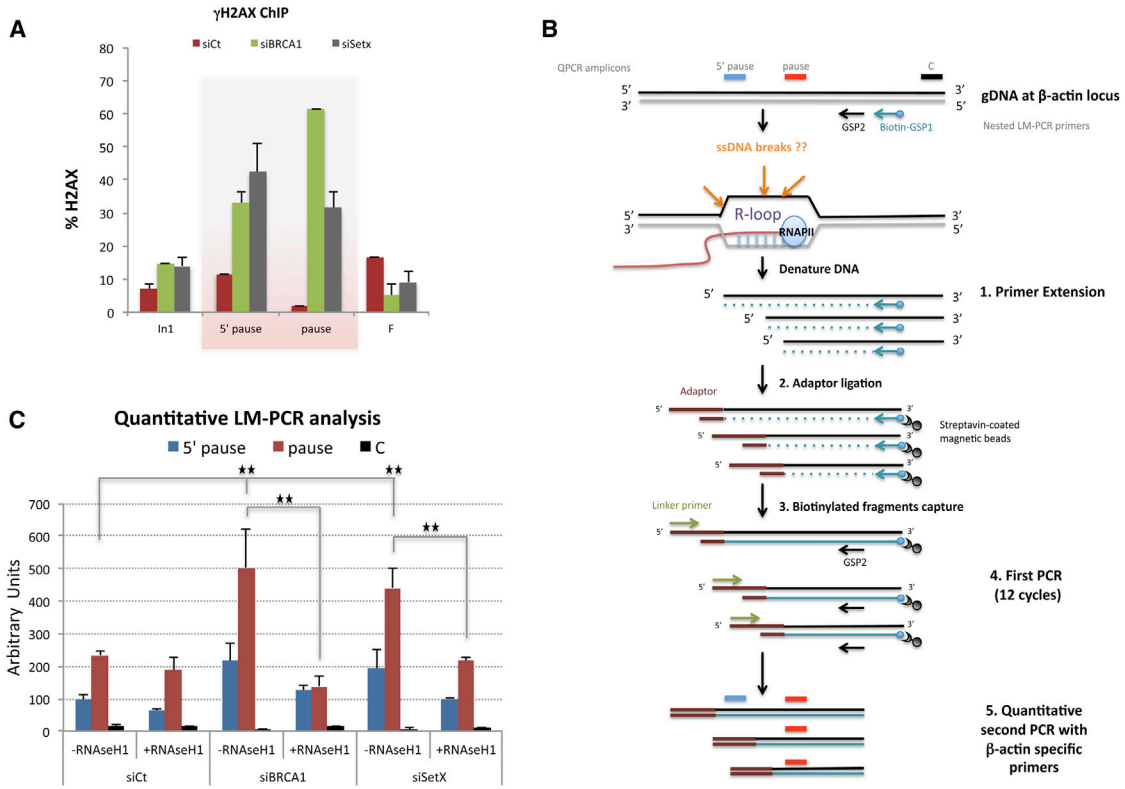


Figure 3. BRCA1/SETX Complex at the β-actin Pause Site Protects Cells from R-Loop-Driven ssDNA Breaks

(A) ChIP analysis performed on the β-actin gene as in Figure 1 using γH2AX and total H2AX antibodies. Histograms represent the proportion of total H2AX phosphorylated on Ser139 (i.e., γH2AX). Average ChIP values ± SD from three independent experiments are shown.

(B) LM-PCR strategy used to identify R-loop-associated ssDNA breaks on the coding strand. See Supplemental Experimental Procedures.

(C) Quantitative detection of ssDNA after LM-PCR performed on the β-actin gene before and after BRCA1 or SETX knockdown and with or without ectopic RNaseH1 expression. QPCR values are average ± SD from three independent experiments. *$p < 0.05$, **$p < 0.007$ by one-tailed Student's t test.

See also Figure S3.

(Aguilera and García-Muse, 2012; Wimberly et al., 2013), we adapted a ligation-mediated quantitative PCR (LM-qPCR) approach to test whether ssDNA breaks had occurred on the non-template strand following BRCA1 or SETX depletion (Figure 3B). Briefly, we ligated an adaptor to putative ssDNA ends after performing a primer extension reaction over the far 3′ end of the β-actin gene. The goal was to amplify any post-ligation DNA fragments that harbor both the adaptor sequence and a segment of the 3′ end of the β-actin gene.

In the absence of BRCA1 or SETX, significant signals were observed when using 5′ pause and pause site primers, thereby reflecting the existence of one or more ssDNA breaks located near the termination region (Figures 3C and S3B). The 3′ R-loop-free C region (Figure 2A) was used as a negative control to validate the specificity of the above-noted results. We also tested the anti-sense strand and failed to observe LM-qPCR signals above background following BRCA1 or SETX depletion (Figure S3C). This suggested that the observed breaks affected only the non-template strand.

We then asked whether RNaseH1 overexpression suppressed the appearance of ssDNA breaks in the β-actin gene in BRCA1- or SETX-depleted cells. The background signals in control region C remained barely detectable, while RNaseH1-expressing BRCA1- or SETX-depleted cells revealed a significant signal decrease within the 5′ pause and pause regions (Figures 3C and S3B). Taken together, these data argue that, in the β-actin gene, BRCA1 and SETX are involved in the prevention/repair of ssDNA breaks arising within the coding strand R-loop near their 3′ end binding site(s).

BRCA1/SETX Complexes Form at Several R-Loop-Associated Genomic Loci and Protect Cells from the Development of ssDNA Breaks

To search for a global effect of BRCA1/SETX function, we first searched for evidence within two gene sets. On one hand, we analyzed two genes (*ENSA*, *Gemin7*) in which transcription termination is also dependent on R-loops (Skourti-Stathaki et al., 2014). In both genes, BRCA1 and SETX binding was enriched at the relevant transcription termination regions as compared to an irrelevant intronic segment (Figures 4A, 4B, S4A, and S4B). Although the BRCA1 antibody used in ChIP experiments displayed some background signal, its specificity was confirmed by its sensitivity to BRCA1 depletion. In parallel, we studied the *Akirin1* and *cyclinB1* genes in which transcription

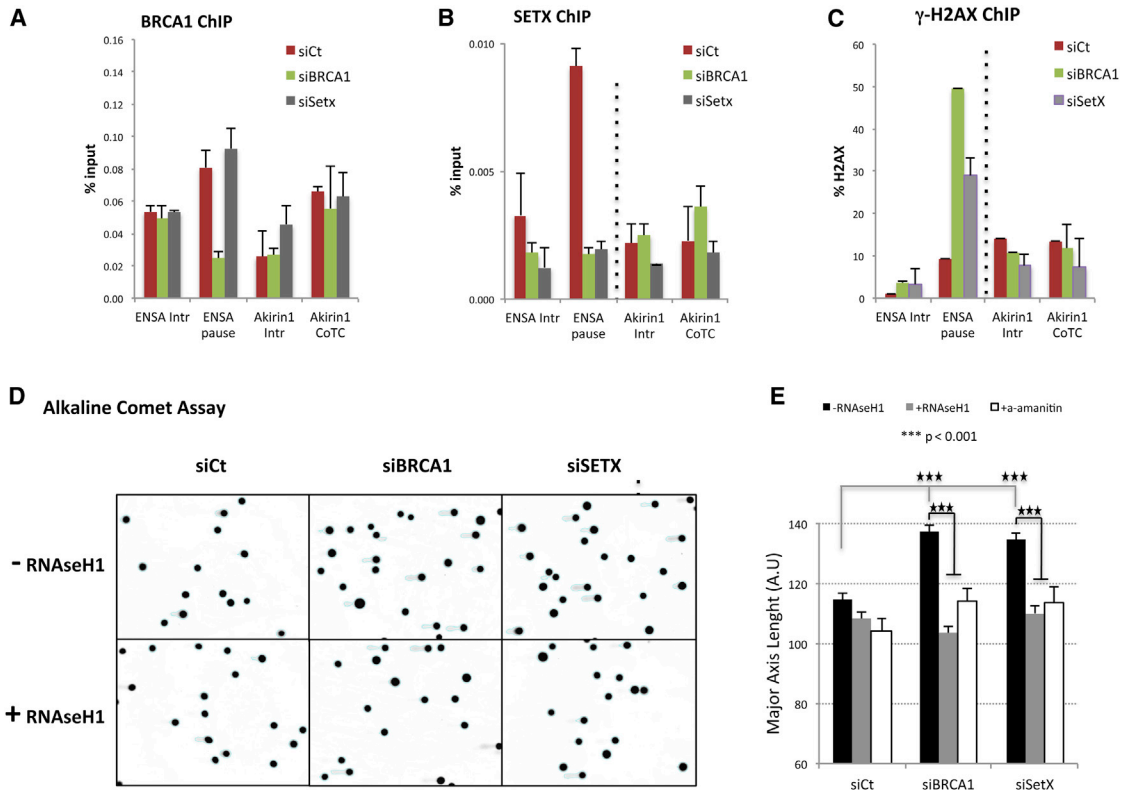


Figure 4. DNA Damage Arising in Absence of BRCA1/SETX Complexes at Termination Pause Sites Is R-Loop Dependent

(A and B) BRCA1 (A) and SETX (B) ChIP analyses performed on the ENSA and Akirin1 genes in HeLa cells transfected with siCt, siBRCA1, or siSETX. ENSA and Akirin1 transcription termination is regulated by R-loops and CoTC sequences, respectively. Intronic regions (Intr) were studied as controls. Average ChIP values ± SD from three independent experiments are shown.

(C) γH2AX ChIP experiments performed as in (A) and analyzed as in Figure 3A. Average ChIP values ± SD from three independent experiments are shown.

(D) Representative pictures of comet assays performed under alkaline conditions in HeLa cells transfected with siCt, siBRCA1, or siSETX in the absence (−RNaseH1) and presence (+RNaseH1) of ectopic RNaseH1 expression.

(E) Quantitative analysis of comet tail lengths for each condition showed in (D). Average tail lengths ±SEM from three independent experiments are shown. ***p < 0.001 by two-tailed Student's t test.

See also Figure S4.

termination is regulated by an R-loop-free co-transcriptional cleavage (CoTC) mechanism (Nojima et al., 2013; Skourti-Stathaki et al., 2014; White et al., 2013). BRCA1 and SETX ChIP analysis revealed no specific signals in either intronic region or the CoTC regions of *Akirin1* and *cyclinB1*, reinforcing the hypothesis that BRCA1/SETX recruitment requires R-loop formation (Figures 4A, 4B, S4A, and S4B).

We also examined the DNA damage signaling response in these two categories of genes by searching for γH2AX signals over the termination regions using ChIP qPCR. Similar to *β-actin*, increased γH2AX signals were clearly visible over the pause element of both *ENSA* and *Gemin7* following BRCA1 or SETX depletion, but not over the control intronic region (Figures 4C and S4C). By contrast, no γH2AX enrichment was detected at "R-loop-free" CoTC termination regions within *Akirin1* or *cyclinB1*. These results are consistent with our β-actin ChIP data (Figure 3A) and with the presence of local γH2AX signals in the absence of BRCA1/SETX complexes over *ENSA* and *Gemin7*. Thus, they can be viewed as a reflection of local R-loop-associated DNA damage. These findings imply that BRCA1/SETX complexes participate in the repair or prevention of this form of damage in these two genes bearing R-loop-associated transcription termination regions.

We then asked whether BRCA1/SETX function at R-loop-forming termination sites is a genome-wide phenomenon. First, comet assays were performed to search for widespread DNA breaks as a consequence of BRCA1 or SETX depletion. As predicted, under alkaline conditions, which detect both ssDNA breaks and DSBs, we observed a significant increase in comet tail lengths in both BRCA1- and SETX-depleted cells. Moreover, RNaseH1 overexpression in both settings decreased the abundance of cells with comets as well as their length (Figures 4D and 4E). Similar results were observed when the cells were treated with a low dose of α-amanitin that acts to inhibit RNAPII elongation. These results confirm that R-loop structures, as by-products of transcription, create numerous fragile genomic sequences prone to BRCA1/SETX depletion-associated damage. Although we also observed an increase in comet tail lengths in BRCA1- and SETX-depleted cells assayed in neutral buffer conditions (probably reflecting DSB), these DNA damage phenotypes did not revert after

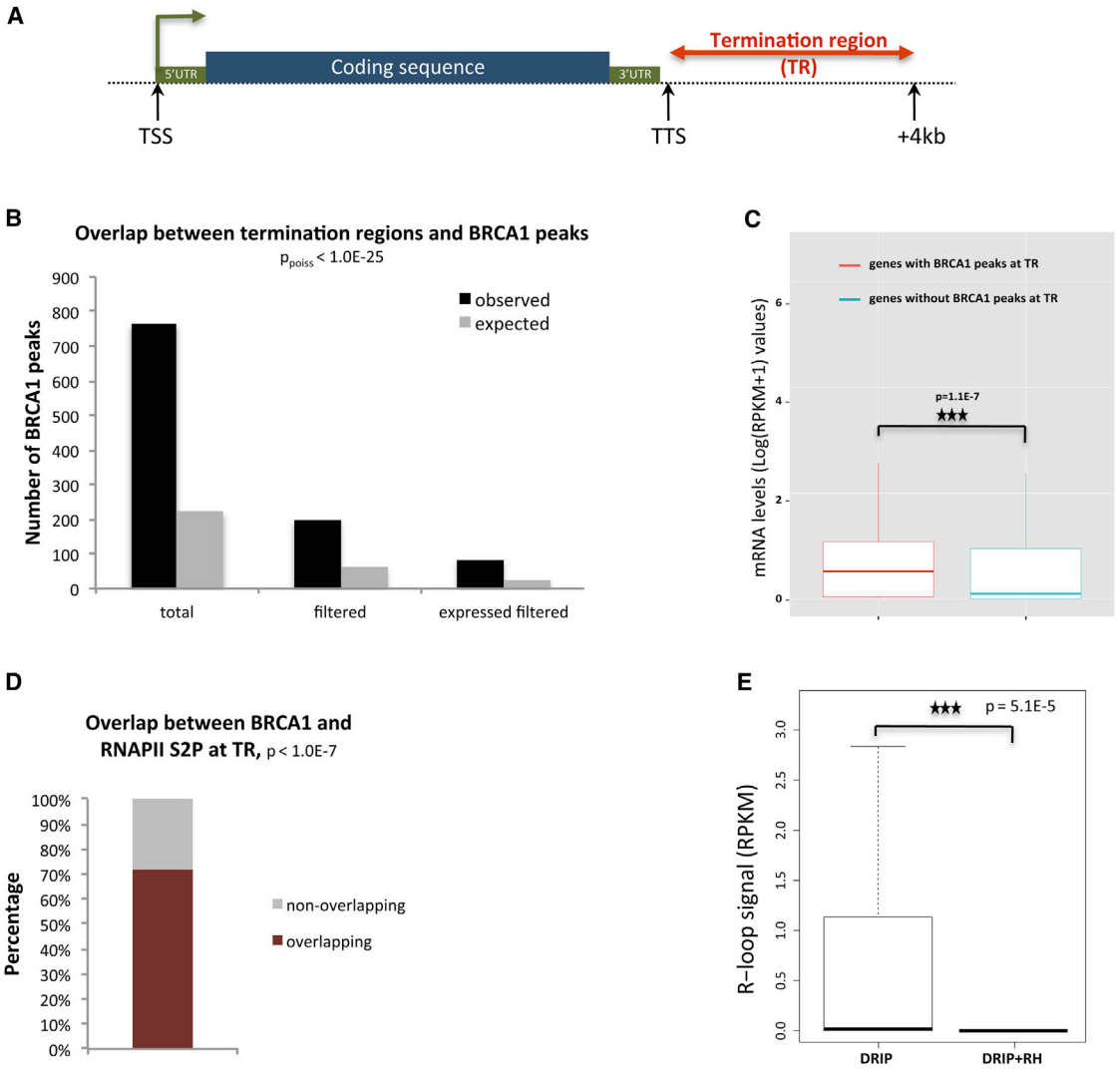


Figure 5. BRCA1 Binds the Transcription Termination Region of a Substantial Subset of Actively Expressed Mammalian Genes

(A) Diagram of candidate BRCA1 TR binding regions. Putative termination regions (TRs) were defined as segments extending from the TTS to TTS + 4 kb.

(B) Total number of observed and expected overlaps between TR and BRCA1 peaks (p_{poiss} = 2.2E-177). "Filtered": BRCA1 ChIP-seq peaks divested of those overlapping promoter regions and transcripts (p_{poiss} = 1.2E-42). "Expressed filtered": BRCA1 peaks present in TR region of expressed genes (p_{poiss} = 5E-24).

(C) Gene expression comparison between genes with (red) and without (blue) BRCA1 bound to relevant TR. Boxplots reflect the median (50th percentile) of mRNA expression. ***p = 1.1E-7, Mann-Whitney test. Outliers have been omitted from the plot.

(D) Overlap between BRCA1 TR and RNAPIISer2P peaks, overlap enrichment over random (***p < 1.0E-7), see Experimental Procedures.

(E) Boxplots showing DRIP-seq signals (RPKM) of DRIP samples compared with DRIP+RH controls (treated with RNaseH1) in BRCA1 TR candidate regions, ***p = 5.101E-5, paired Wilcoxon test.

See also Figure S5 and Table S1.

overexpression of RNaseH1 or exposure to α-amanitin (Figure S4D). Thus, R-loop-dependent DNA damage following BRCA1/SETX depletion is mainly composed of ssDNA breaks.

BRCA1 Is Globally Associated with Transcription Termination Sites of Highly Transcribed Genes

In light of the possibility that R-loop-associated termination regions are characteristics of larger numbers of genes, we performed a meta-analysis from several deep-sequencing datasets to search for genomic co-localization of BRCA1 peaks at transcription termination regions. We used the ENCODE BRCA1 ChIP-seq dataset to identify candidate genes regulated by this mechanism (Consortium, 2012). For each gene, a candidate transcription TR was defined as a 4 kb segment downstream of the transcription termination site (TTS) (Figure 5A).

Initially, 764 distinct BRCA1 peaks significantly overlapping a genomic TR were identified. Since BRCA1 is also associated with a subset of promoters (K.M.M. et al., unpublished data) (Gardini et al., 2014), we further filtered these data to exclude any peaks overlapping promoter regions and their associated

transcripts. In this filtered gene set, 196 BRCA1 peaks, each associated with a TR, corresponded to 184 genes that are referred to here as BRCA1 TR genes (Figure 5B and Table S1).

Since the DNA damage that developed over R-loop-associated TRs upon BRCA1/SETX depletion is proposed to be R-loop and, therefore, transcription dependent, we analyzed the expression levels of these BRCA1 TR genes. RNA-seq data were used to analyze HeLa cell transcriptomes available from ENCODE. The expression profiles of genes whose filtered TRs did or did not overlap BRCA1 ChIP-seq peaks were compared. The data showed that BRCA1 TR genes clearly displayed higher mRNA levels ($p = 1.1E\text{-}7$), when compared to genes whose termination sites are not associated with a BRCA1 binding site (Figures 5C and S5 and Table S1). These results indicate that BRCA1 TR genes are more highly expressed than those lacking BRCA1 TR binding peaks.

To explore further whether BRCA1 binding at TR is associated with transcription termination, we searched for evidence of the co-localization at termination sites of BRCA1 peaks and RNAPII paused at the 3′ end of the relevant genes (RNAPII S2P) (Davidson et al., 2014). Here we used ENCODE RNAPII CTD Ser2-P ChIP-seq data (Consortium, 2012). This analysis showed that more than 70% of BRCA1 peaks within TRs overlap paused RNAPII (Figure 5D and Table S1), suggesting that BRCA1 is actively engaged in transcription-associated events at termination sites. To investigate whether BRCA1 TR are associated with R-loops, we compared them with the RNA:DNA IP database (DRIP-seq) generated with cells before and after RNaseH1 treatment (DRIP+RH) (Ginno et al., 2012). We found that BRCA1 TR peaks show a significant enrichment of DRIP signal as compared to DRIP+RH (Figure 5E), implying the existence of R-loop formation at the BRCA1-associated TR loci.

These results suggest that in mammalian cells BRCA1 binds to TR (associated with paused RNAPII) in a substantial subset of highly transcribed genes, whose pause sites reflect a strong tendency to form R-loops. Taken together, our findings indicate that BRCA1 and SETX participate in the prevention/repair of ssDNA damage occurring at specific regions (TRs) in response to the transcription-associated formation of R-loops in physiological conditions. Conceivably, these genes require BRCA1/SETX-dependent DNA damage surveillance to ensure regulation of transcription (Huppert et al., 2008; Skourti-Stathaki and Proudfoot, 2014).

BRCA1 Mutant Breast Cancers Reveal Genomic Alterations at BRCA1-Associated Termination Sites

We next asked whether a defect in the newly detected BRCA1/SETX function could participate in the pathogenesis of BRCA1-deficient breast cancer tumors. Recent large-scale DNA sequencing screens performed in various human cancers have shed light on the nature and location of associated somatic mutations (Alexandrov et al., 2013; Pleasance et al., 2010). We performed a mutation analysis in the 184 BRCA1 TR genes, using the complete catalog of somatic mutations obtained from the whole-genome sequencing of 21 breast cancers, a subset of which ($n = 5$) were BRCA1 mutant (Nik-Zainal et al., 2012). This catalog includes single base mutations (SBMs), insertions and deletions (indels), and chromosomal rearrangements. For each patient, the genome of healthy/normal mammary tissue was also sequenced, and any mutations therein were deleted from the patient mutation analysis. Since R-loops can be long, we first included promoter regions in this search. More specifically, we first defined for each BRCA1 TR gene a whole gene region of special interest from TSS − 1,250 bp to TTS + 5 kb and searched for BRCA1-specific mutations.

We divided the 21 patient datasets into 3 breast cancer types: 12 sporadic (i.e., tumors WT/WT for both BRCA1 and BRCA2), 5 BRCA1, and 4 BRCA2 mutant tumors. Figure 6A compares each of the mutant groups to WT, in search of a quantitative measure of any difference in mutation rate (i.e., effect size) between any 2 groups. The effect size was computed as a standardized difference in mutation counts when all selected genes are considered together. It was obtained separately for each type of mutation analyzed. The results clearly indicate that, within the BRCA1 TR gene regions of interest, there was significant enrichment for indels only in the mutant BRCA1 tumors ($z = 2.875$, $p = 0.009$), despite the fact that more indels were detected throughout the entire genomes of BRCA2 than BRCA1 breast cancers (Nik-Zainal et al., 2012). By contrast, these genomic loci were enriched for SBM in the BRCA2 tumors ($z = 2.144$, $p = 0.021$). Of note, no significant enrichment for rearrangements was observed. These results indicate that, within BRCA1 TR genes, distinct mutational patterns exist that separate BRCA1 and BRCA2 tumors. This suggests that any biochemical defects that the absence of these proteins elicits may, at least in part, be different.

In parallel, we also performed a global mutation analysis among the "R-loop free" CoTC-regulated genes ("negative" genes) (Nojima et al., 2013) where BRCA1 is not recruited (Figure 4A). We failed to detect any significant change in the species of mutations detected in either BRCA1 or BRCA2 null tumors (Figure S6).

To further investigate the mutational profile at the 3′ ends of the BRCA1 TR genes, we focused on a narrower region defined as: TTS ± 4 kb. Strikingly, the only significant enrichment observed was for indels in BRCA1 mutant tumors ($z = 3.464$, $p = 0.002$) (Figure 6B). Precise mapping of the indels in the BRCA1 TR genes in the BRCA1 null tumors indicated that 3 out of 6 were located ~300–400 bp from the BRCA1 TR peak; 2 were located further away (last intron and ~1.8 kb downstream of the BRCA1 peak); and another was close to the promoter region (Figures 6C–6E and Table S2). The absence of SBM enrichment in the 3′ end regions of the TR genes in BRCA2 mutant carriers suggests that most of it accumulates at 5′ ends and within the gene body. Of note, none of the few mutations detected in the CoTC genes was located in 3′ end regions.

Overall, these results show that BRCA1-deficient tumors are significantly enriched for indels in BRCA1 TR genes compared to BRCA2 and sporadic breast cancers. A significant fraction of the BRCA1 tumor-associated indels lie in the vicinity of BRCA1 binding TR loci, supporting the view that BRCA1 plays a critical role in preventing/repairing R-loop-mediated damage in the vicinity of R-loop-associated TR.

DISCUSSION

In this study, a combination of biochemical, molecular, and genetic data provides evidence that a newly identified BRCA1/SETX

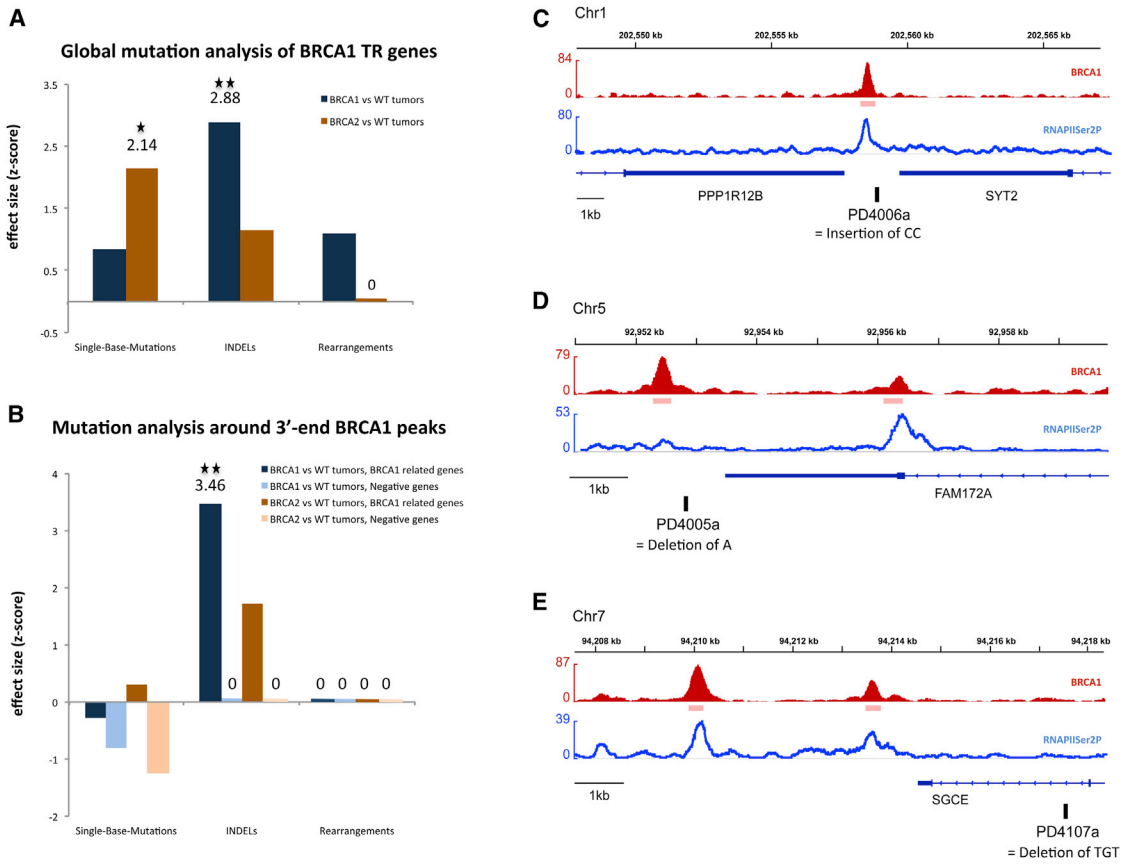


Figure 6. BRCA1-Deficient Breast Cancers Reveal Genomic Abnormalities at and near BRCA1-Associated Termination Sites

(A and B) Global mutational analysis carried out in the 184 BRCA1 TR genes using the complete whole-genome catalog of somatic mutations from 21 breast cancers (Nik-Zainal et al., 2012). Effect size comparison (one-tailed CMH *Z* score) between the different tumor subgroups when testing the region from TSS − 1,250 bp to TTS + 5 kb (A) or ± 4 kb from TTS (B). WT tumors = non-BRCA1/BRCA2, and negative genes = CoTC genes. Statistical significant: *$p < 0.05$ and **$p < 0.01$.

(C–E) ChIP-seq profiles of BRCA1 (red) and RNAPIISer2P (blue) in BRCA1 TR genes and location of indels (black boxes). Chr, chromosome.

See also Figure S6 and Table S2.

complex is required to restrain the development of R-loop-mediated DNA damage at specific genomic regions. In particular, we have delineated a sequence of events that is predicted to occur over a substantial subset of R-loop-associated transcriptional pause sites and, when impaired, results in the BRCA1 TR-associated indel mutations observed in BRCA1$^{-/-}$ breast cancer genomes.

Under physiological conditions, R-loops play a role in transcription termination at G-rich pause sites, like those in *β-actin*, *Gemin7*, and *ENSA* (Skourti-Stathaki et al., 2014). They are also present at promoter regions where there is a region of GC-skew that ensures the protection of vicinal CpG island sequences against DNA methylation (Ginno et al., 2012). Though these structures represent R-loops associated with distinct biological functions, they are, nonetheless, still predicted to induce DNA damage (for review see Skourti-Stathaki and Proudfoot, 2014).

How BRCA1 recognizes R-loop structures remains to be determined. Current knowledge focuses mainly upon the factors that regulate their formation (Chan et al., 2014; Hamperl and Cimprich, 2014; Skourti-Stathaki and Proudfoot, 2014). However, little is known about potential direct interaction between specific proteins and R-loops. Several possibilities exist. For example, there could be direct recognition of an RNA:DNA hybrid, e.g., as occurs with the hybrid-binding domain (HBD) of RNaseH1 (Cerritelli and Crouch, 2009), and/or of the non-template ssDNA. For BRCA1, it is also possible that the branched structure of an R-loop resembles the DNA flaps, branched DNA, and/or four-way junctions for which BRCA1 exhibits an intrinsic affinity (Paull et al., 2001). Alternatively, BRCA1 recruitment to R-loops could be mediated by other protein complexes that normally interact with these special DNA structures. These could be components of the RNAPII holoenzyme, splicing machinery, and/or specific chromatin remodeling complexes that are associated with R-loops (Bochar et al., 2000; Boulé and Zakian, 2007; Savage et al., 2014a; Scully et al., 1997; Yarden and Brody, 1999).

Results presented here show that a deficiency in BRCA1/SETX function results in unrepaired ssDNA breaks on the non-template strand at certain R-loop-associated TRs. γH2AX

accumulation was observed by ChIP, although the extent of its genomic presence appears to be more restricted than observed at sites of DSB (Paull et al., 2000). It is tempting to speculate that the restricted presence of γH2AX signals is a reflection of the presence of ssDNA breaks as the source of DNA damage and/or that γH2AX spreading is antagonized by high levels of transcription (Iacovoni et al., 2010).

The above considerations aside, how DNA damage occurs at such sites remains unknown. Multiple molecular mechanisms may foster or contribute to the DNA damage associated with R-loops. Superhelical stress and G4 and/or flap endonucleases could generate local ssDNA breaks, some of which could devolve into DSB (Hamperl and Cimprich, 2014). In addition, genome-wide sequencing of human cancers suggests that non-random, clustered mutations may be concentrated in long ssDNA regions, some of which might form R-loop structures (Roberts et al., 2012). Indeed, during transcription, possibly due to the relative chemical susceptibility to damage of ssDNA, there is an increase in the mutation rate associated with the activity of editing enzymes like activation-induced cytosine deaminase (AID) or apolipoprotein B mRNA-editing catalytic polypeptide proteins (APOBEC) (Alexandrov et al., 2013; Beale et al., 2004; Chan et al., 2012). Recently, the latter have been suggested to play a role in the mutational processes that affect breast cancer genomes (Burns et al., 2013; Nik-Zainal et al., 2012, 2014).

Taken together, our findings show that BRCA1 contributes to the control and/or repair of R-loop-mediated DNA damage at specific sites, which is, potentially, a significant contributor to the maintenance of genomic stability. Similarly, the implications of a proposed role for BRCA2 in transcription-associated recombination (TAR) and in the processing of R-loops in partnership with RNA processing factors suggest that these structures are a source of cancer-related instability (Bhatia et al., 2014; Gallardo et al., 2003; Huertas and Aguilera, 2003; Savolainen and Helleday, 2009).

BRCA1- and *BRCA2*-mutant cancers exhibit differences in histopathology, gene expression profiles, and clinical course, even though they share similarities in their marked levels of genomic instability (Nik-Zainal et al., 2012). By contrast, our observations imply that, at certain transcription termination regions, the mutational signatures of BRCA1 and BRCA2 null tumors are different. Conceivably, these differences arise from the possibility that BRCA1 and BRCA2 respond to/interact with R-loops differently and even in different contexts. Additional whole-genome sequencing from greater numbers of BRCA mutant cancers would be required to address these possibilities in the future. Further studies will also be needed to determine whether such BRCA1/BRCA2 mutational differences contribute directly or indirectly to the biological differences between BRCA1 and BRCA2 breast cancers.

EXPERIMENTAL PROCEDURES

For detailed experimental procedures, see Supplemental Experimental Procedures.

Co-Immunoprecipitation Analysis

Whole-cell extracts (WCE) were prepared as previously described (Tardat et al., 2010), except that the lysis mixture was sonicated with a Sonic Dismembrator for 15 s at an amplitude of 20% (Fisher Scientific, Model 120). 200 μg of WCE were incubated for 2 hr at 4°C with anti-BRCA1#1 (SG11, mouse monoclonal), anti-BRCA1#2 (MS110, mouse monoclonal), anti-SETX#1 (A301-105A, Bethyl), anti-SETX#2 (A301-104A, Bethyl), or control IgG. Immune complexes were collected with protein A/G magnetic beads (Dynabeads, Invitrogen) for 15 min at 4°C and washed with IP buffer (with 1 mM DTT) and increasing concentrations of KCl (50/100/150 mM KCl). Bound proteins were eluted in LDS sample buffer 1× (Invitrogen) and analyzed by immunoblotting.

qPCR ChIP and DIP Experiments

ChIP experiments on BRCA1, SETX, and γH2AX were performed using a modified version of N.J.P. laboratory's protocol (Skourti-Stathaki et al., 2011), as detailed in the Supplemental Experimental Procedures. DNA:RNA hybrid precipitation (DIP) analysis was performed with the specific monoclonal RNA:DNA hybrid Ab (S9.6) as described in Skourti-Stathaki et al. (2011). Sequences of the DNA primers are listed in Supplemental Experimental Procedures.

Mutational Analyses of Breast Cancers

Detailed description of the computational and statistical analyses is available in Supplemental Experimental Procedures.

The integrative Genomics Viewer (IGV) genome browser was used to visualize the indels and the BRCA1 and RNAPIISer2P ChIP-seq profiles across the termination regions of the BRCA1 TR genes (Thorvaldsdóttir et al., 2013).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2015.01.011.

AUTHOR CONTRIBUTIONS

S.J.H. and D.M.L. originally suggested and first conceived the rationale for this project. E.H. designed the overall project with help from K.S.-S., K.M.M., D.M.L., and N.J.P. E.H. performed the experiments and analyzed the experimental data unless otherwise stated. E.H. received advice, reagents, and help for the ChIP experiments and the co-immunoprecipitation from K.S.-S., K.K.M., S.P., and S.D. K.S.-S. performed and analyzed the BRCA1 and SETX DIP and part of BRCA1 and SETX ChIP experiments. A.Y., M.L.E., and M.K. performed the genomic analysis of BRCA1 ChIP-seq data and K.K.-G. the meta-analyses correlating RNAPII and DRIP datasets. S.V., L.P., and G.P. performed the global mutational analyses of the breast tumors. E.H. and D.M.L. wrote the manuscript. All authors contributed to the editing of the manuscript.

ACKNOWLEDGMENTS

We thank all Livingston laboratory members for support, technical advice, and helpful discussions. E.H., D.M.L., and other Livingston laboratory members were supported by grants from the Breast Cancer Research Foundation, the Susan G. Komen Foundation for the Cure (SAC140022), and from the National Cancer Institute (NCI) - Mechanisms of Breast Development and Carcinogenesis (2PO1CA80111-16) and BRCA1 Function in Post Damage Foci (5R01CA136512-05). K.K.-G. was supported by Marie Curie IEF. K.S.-S. and N.J.P. were supported by a Programme grant from the Wellcome Trust and G.P. by NCI 5P30 CA006516-46.

Received: August 25, 2014
Revised: November 21, 2014
Accepted: January 5, 2015
Published: February 19, 2015

REFERENCES

Aguilera, A. (2002). The connection between transcription and genomic instability. EMBO J. *21*, 195–201.

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol. Cell *46*, 115–124.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Alzu, A., Bermejo, R., Begnis, M., Lucca, C., Piccini, D., Carotenuto, W., Saponaro, M., Brambati, A., Cocito, A., Foiani, M., and Liberi, G. (2012). Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. Cell *151*, 835–846.

Anderson, S.F., Schlegel, B.P., Nakajima, T., Wolpin, E.S., and Parvin, J.D. (1998). BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. Nat. Genet. *19*, 254–256.

Beale, R.C.L., Petersen-Mahrt, S.K., Watt, I.N., Harris, R.S., Rada, C., and Neuberger, M.S. (2004). Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. J. Mol. Biol. *337*, 585–596.

Becherel, O.J., Yeo, A.J., Stellati, A., Heng, E.Y.H., Luff, J., Suraweera, A.M., Woods, R., Fleming, J., Carrie, D., McKinney, K., et al. (2013). Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. PLoS Genet. *9*, e1003435.

Bennett, C.B., Westmoreland, T.J., Verrier, C.S., Blanchette, C.A.B., Sabin, T.L., Phatnani, H.P., Mishina, Y.V., Huper, G., Selim, A.L., Madison, E.R., et al. (2008). Yeast screens identify the RNA polymerase II CTD and SPT5 as relevant targets of BRCA1 interaction. PLoS ONE *3*, e1448.

Bhatia, V., Barroso, S.I., García-Rubio, M.L., Tumini, E., Herrera-Moyano, E., and Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature *511*, 362–365.

Bochar, D.A., Wang, L., Beniya, H., Kinev, A., Xue, Y., Lane, W.S., Wang, W., Kashanchi, F., and Shiekhattar, R. (2000). BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. Cell *102*, 257–265.

Boulé, J.-B., and Zakian, V.A. (2007). The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. Nucleic Acids Res. *35*, 5809–5818.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature *494*, 366–370.

Cantor, S.B., Bell, D.W., Ganesan, S., Kass, E.M., Drapkin, R., Grossman, S., Wahrer, D.C., Sgroi, D.C., Lane, W.S., Haber, D.A., and Livingston, D.M. (2001). BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell *105*, 149–160.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. FEBS J. *276*, 1494–1505.

Chan, K., Sterling, J.F., Roberts, S.A., Bhagwat, A.S., Resnick, M.A., and Gordenin, D.A. (2012). Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. PLoS Genet. *8*, e1003149.

Chan, Y.A., Hieter, P., and Stirling, P.C. (2014). Mechanisms of genome instability induced by RNA-processing defects. Trends Genet. *30*, 245–253.

Consortium, T.E.P.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Daniel, J.A., and Nussenzweig, A. (2013). The AID-induced DNA damage response in chromatin. Mol. Cell *50*, 309–321.

Davidson, L., Muniz, L., and West, S. (2014). 3′ end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. Genes Dev. *28*, 342–356.

Gaillard, H., Herrera-Moyano, E., and Aguilera, A. (2013). Transcription-associated genome instability. Chem. Rev. *113*, 8638–8661.

Gallardo, M., Luna, R., Erdjument-Bromage, H., Tempst, P., and Aguilera, A. (2003). Nab2p and the Thp1p-Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. J. Biol. Chem. *278*, 24225–24232.

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev. *25*, 2041–2056.

Gardini, A., Baillat, D., Cesaroni, M., and Shiekhattar, R. (2014). Genome-wide analysis reveals a role for BRCA1 and PALB2 in transcriptional co-activation. EMBO J. *33*, 890–905.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell *45*, 814–825.

Gorski, J.J., Savage, K.I., Mulligan, J.M., McDade, S.S., Blayney, J.K., Ge, Z., and Harkin, D.P. (2011). Profiling of the BRCA1 transcriptome through microarray and ChIP-chip analysis. Nucleic Acids Res. *39*, 9536–9548.

Hamperl, S., and Cimprich, K.A. (2014). The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. DNA Repair (Amst.) *19*, 84–94.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol. Cell *44*, 966–977.

Hill, S.J., Rolland, T., Adelmant, G., Xia, X., Owen, M.S., Dricot, A., Zack, T.I., Sahni, N., Jacob, Y., Hao, T., et al. (2014). Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. Genes Dev. *28*, 1957–1975.

Huen, M.S.Y., Sy, S.M.H., and Chen, J. (2010). BRCA1 and its toolbox for the maintenance of genome integrity. Nat. Rev. Mol. Cell Biol. *11*, 138–148.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol. Cell *12*, 711–721.

Huppert, J.L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008). G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res. *36*, 6260–6268.

Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. EMBO J. *29*, 1446–1457.

Kawai, S., and Amano, A. (2012). BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. J. Cell Biol. *197*, 201–208.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. Nat. Rev. Genet. *13*, 204–214.

Kleiman, F.E., and Manley, J.L. (1999). Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. Science *285*, 1576–1579.

Kleiman, F.E., Wu-Baer, F., Fonseca, D., Kaneko, S., Baer, R., and Manley, J.L. (2005). BRCA1/BARD1 inhibition of mRNA 3′ processing involves targeted degradation of RNA polymerase II. Genes Dev. *19*, 1227–1237.

Li, X., and Manley, J.L. (2006). Cotranscriptional processes and their influence on genome stability. Genes Dev. *20*, 1838–1847.

Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Mol. Cell *41*, 21–32.

Mullan, P.B., Quinn, J.E., and Harkin, D.P. (2006). The role of BRCA1 in transcriptional regulation and cell cycle control. Oncogene *25*, 5854–5863.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

Nik-Zainal, S., Wedge, D.C., Alexandrov, L.B., Petljak, M., Butler, A.P., Bolli, N., Davies, H.R., Knappskog, S., Martin, S., Papaemmanuil, E., et al. (2014). Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat. Genet. *46*, 487–491.

Nojima, T., Dienstbier, M., Murphy, S., Proudfoot, N.J., and Dye, M.J. (2013). Definition of RNA polymerase II CoTC terminator elements in the human genome. Cell Rep. *3*, 1080–1092.

Pathania, S., Nguyen, J., Hill, S.J., Scully, R., Adelmant, G.O., Marto, J.A., Feunteun, J., and Livingston, D.M. (2011). BRCA1 is required for postreplication repair after UV-induced DNA damage. Mol. Cell *44*, 235–251.

Paull, T.T., Rogakou, E.P., Yamazaki, V., Kirchgessner, C.U., Gellert, M., and Bonner, W.M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. Curr. Biol. *10*, 886–895.

Paull, T.T., Cortez, D., Bowers, B., Elledge, S.J., and Gellert, M. (2001). Direct DNA binding by Brca1. Proc. Natl. Acad. Sci. USA *98*, 6086–6091.

Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.-L., Ordóñez, G.R., Bignell, G.R., et al. (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. Nature *463*, 191–196.

Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., Mieczkowski, P.A., et al. (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol. Cell *46*, 424–435.

Savage, K.I., Gorski, J.J., Barros, E.M., Irwin, G.W., Manti, L., Powell, A.J., Pellagatti, A., Lukashchuk, N., McCance, D.J., McCluggage, W.G., et al. (2014a). Identification of a BRCA1-mRNA splicing complex required for efficient DNA repair and maintenance of genomic stability. Mol. Cell *54*, 445–459.

Savage, K.I., Matchett, K.B., Barros, E.M., Cooper, K.M., Irwin, G.W., Gorski, J.J., Orr, K.S., Vohhodina, J., Kavanagh, J.N., Madden, A.F., et al. (2014b). BRCA1 deficiency exacerbates estrogen-induced DNA damage and genomic instability. Cancer Res. *74*, 2773–2784.

Savolainen, L., and Helleday, T. (2009). Transcription-associated recombination is independent of XRCC2 and mechanistically separate from homology-directed DNA double-strand break repair. Nucleic Acids Res. *37*, 405–412.

Scully, R., Anderson, S.F., Chao, D.M., Wei, W., Ye, L., Young, R.A., Livingston, D.M., and Parvin, J.D. (1997). BRCA1 is a component of the RNA polymerase II holoenzyme. Proc. Natl. Acad. Sci. USA *94*, 5605–5610.

Silver, D.P., and Livingston, D.M. (2012). Mechanisms of BRCA1 tumor suppression. Cancer Discov *2*, 679–684.

Skourti-Stathaki, K., and Proudfoot, N.J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. *28*, 1384–1396.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol. Cell *42*, 794–805.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol. Cell *24*, 735–746.

Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev. *26*, 163–175.

Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., and Lavin, M.F. (2009). Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. Hum. Mol. Genet. *18*, 3384–3396.

Tardat, M., Brustel, J., Kirsh, O., Lefevbre, C., Callanan, M., Sardet, C., and Julien, E. (2010). The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. Nat. Cell Biol. *12*, 1086–1093.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Tutt, A., and Ashworth, A. (2002). The relationship between the roles of BRCA genes in DNA repair and cancer predisposition. Trends Mol. Med. *8*, 571–576.

Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F., and Culbertson, M.R. (1997). The yeast SEN1 gene is required for the processing of diverse RNA classes. Nucleic Acids Res. *25*, 4778–4785.

Ursic, D., Chinchilla, K., Finkel, J.S., and Culbertson, M.R. (2004). Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. Nucleic Acids Res. *32*, 2441–2452.

Venkitaraman, A.R. (2014). Cancer suppression by the chromosome custodians, BRCA1 and BRCA2. Science *343*, 1470–1475.

White, E., Kamieniarz-Gdula, K., Dye, M.J., and Proudfoot, N.J. (2013). AT-rich sequence elements promote nascent transcript cleavage leading to RNA polymerase II termination. Nucleic Acids Res. *41*, 1797–1806.

Wimberly, H., Shee, C., Thornton, P.C., Sivaramakrishnan, P., Rosenberg, S.M., and Hastings, P.J. (2013). R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. Nat. Commun. *4*, 2115.

Yarden, R.I., and Brody, L.C. (1999). BRCA1 interacts with components of the histone deacetylase complex. Proc. Natl. Acad. Sci. USA *96*, 4983–4988.

Yüce, Ö., and West, S.C. (2013). Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. Mol. Cell. Biol. *33*, 406–417.

Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature *477*, 179–184.

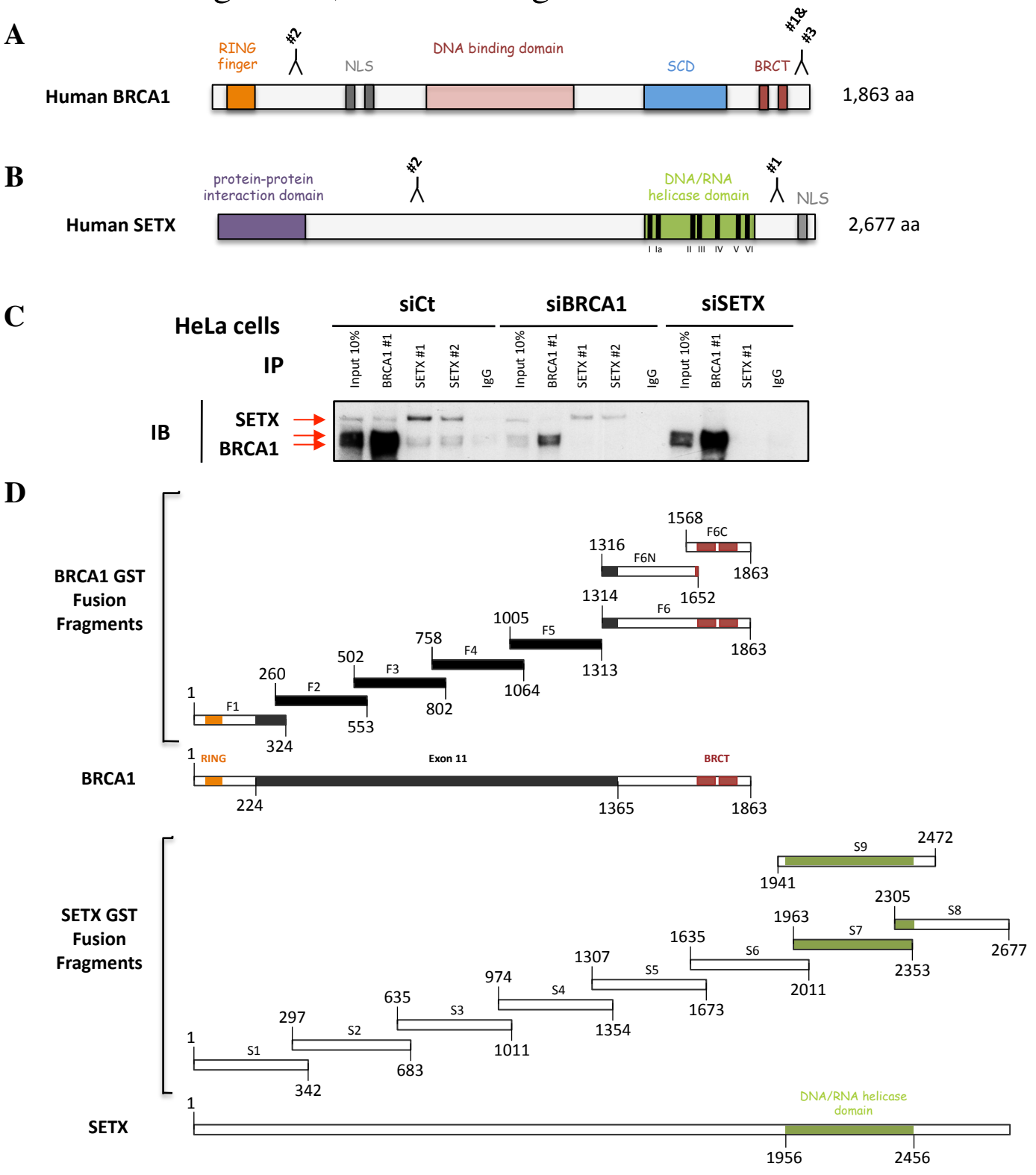
Molecular Cell

Supplemental Information

BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair

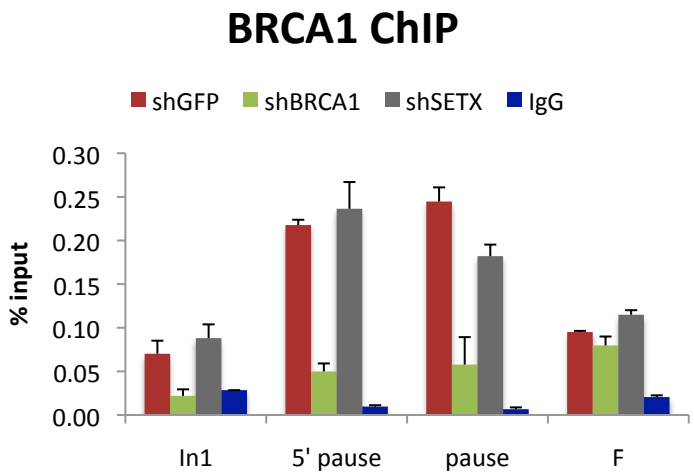
Elodie Hatchi, Konstantina Skourti-Stathaki, Steffen Ventz, Luca Pinello, Angela Yen, Kinga Kamieniarz-Gdula, Stoil Dimitrov, Shailja Pathania, Kristine McKinney, Matthew L. Eaton, Manolis Kellis, Sarah J. Hill, Giovanni Parmigiani, Nicholas J. Proudfoot, and David M. Livingston

Hatchi et al. Figure S1, related to Figure 1

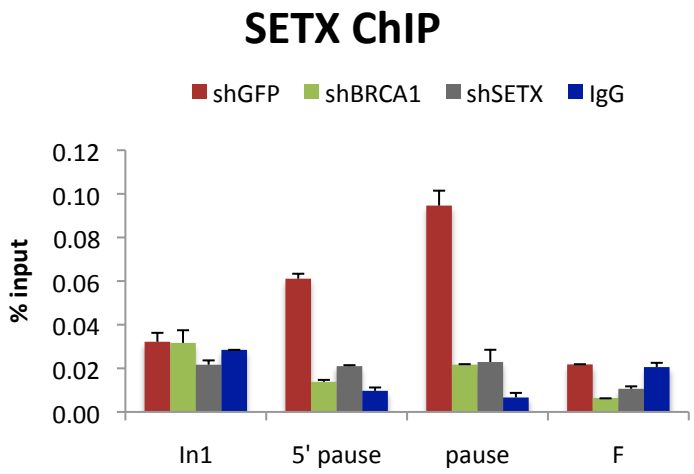


Hatchi et al. Figure S2, related to Figure 1

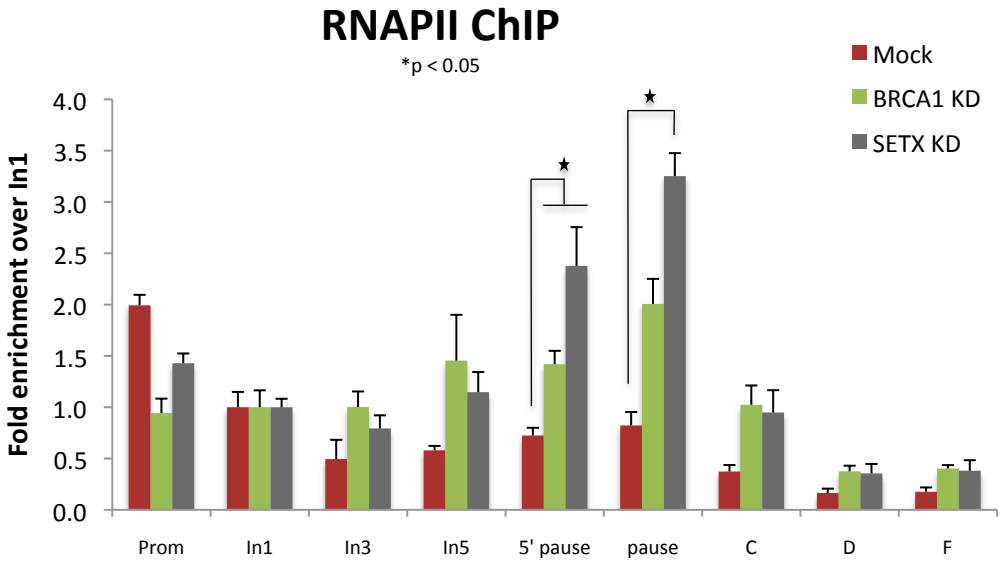
A



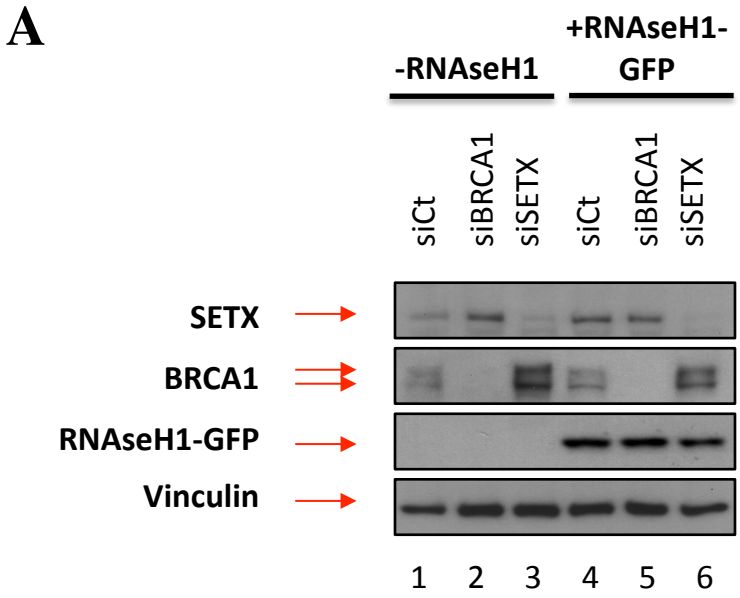
B

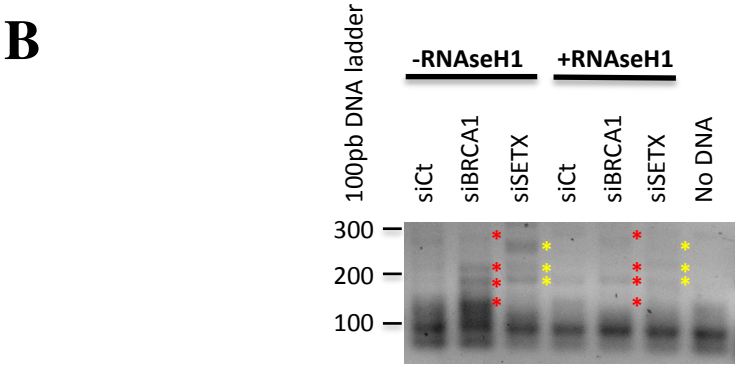


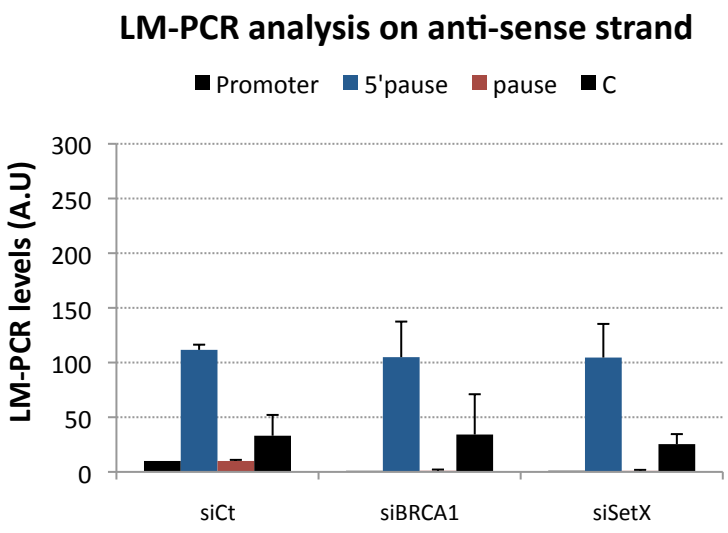
C



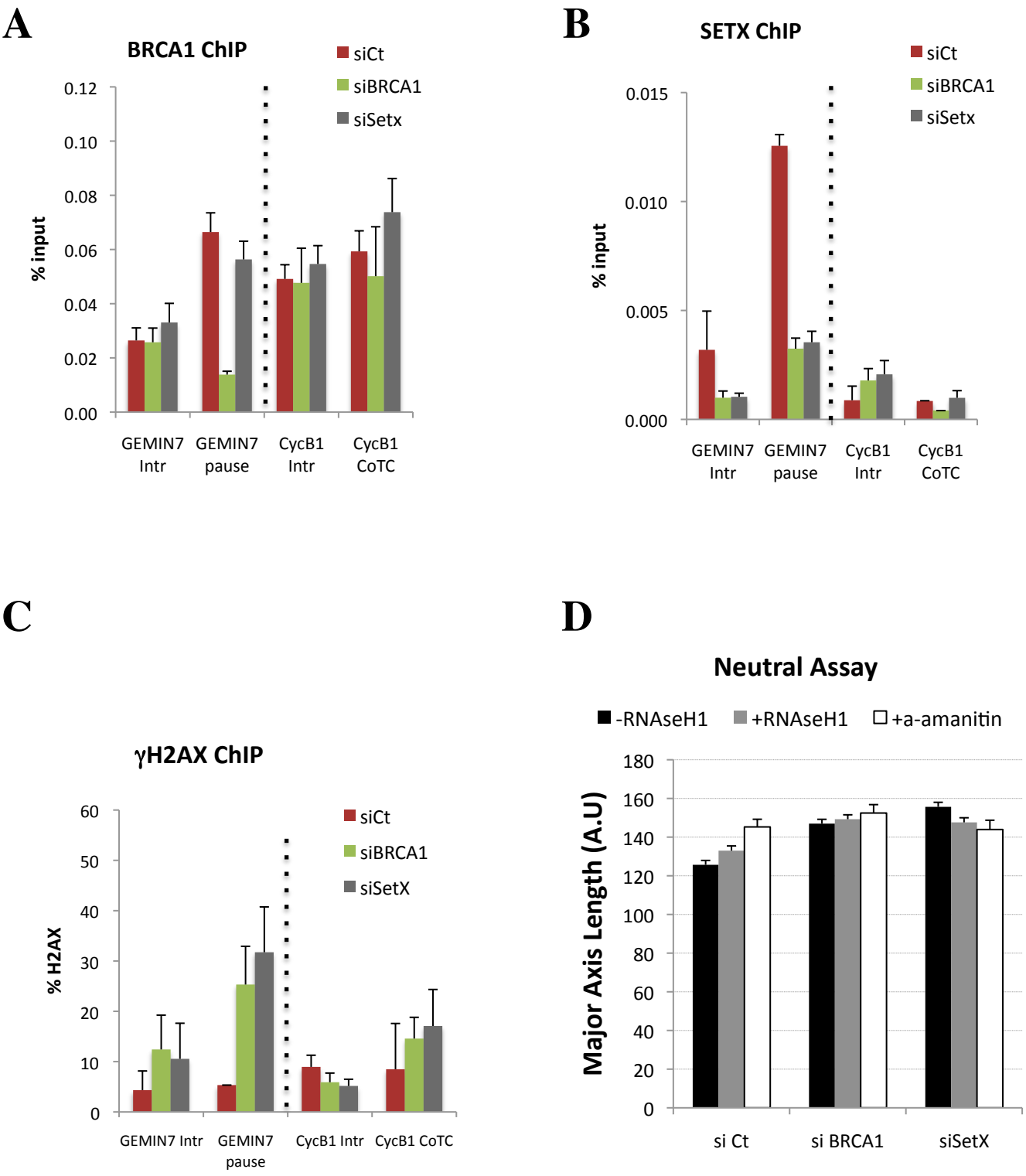
Hatchi et al. Figure S3, related to Figures 2 and 3



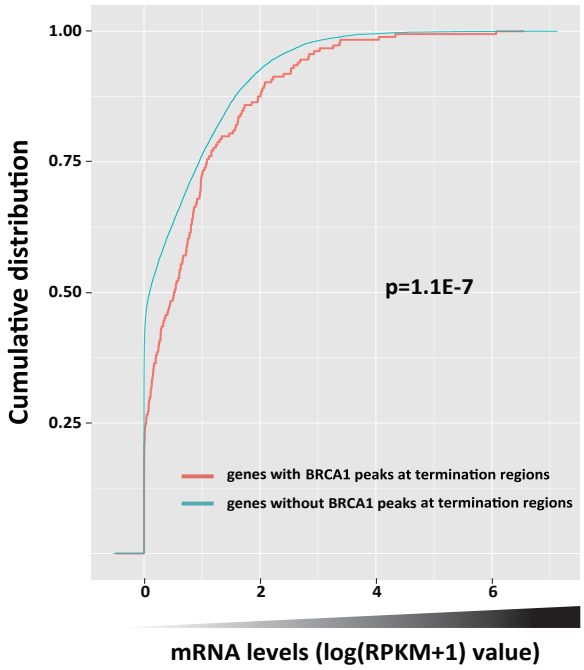




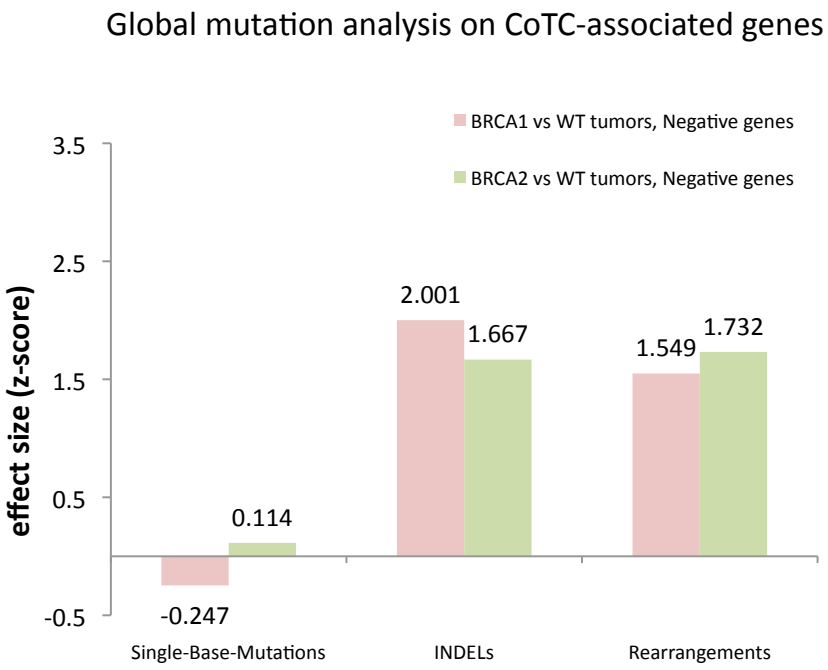
Hatchi et al. Figure S4, related to Figure 4



Hatchi et al. Figure S5, related to Figure 5



Hatchi et al. Figure S6, related to Figure 6



Supplemental Figure Legends

Figure S1. BRCA1 and SETX antibodies, epitope validation, and GST binding results. Related to Figure 1.

(A) Human BRCA1 p220 (referred to as BRCA1) is a multifunctional protein bearing a RING domain and paired BRCT motifs. The location of three BRCA1 monoclonal antibody epitopes is indicated. BRCA1#3 is A300-000A from Bethyl Laboratories and was used for BRCA1 ChIP assays. NLS, nuclear localization site; SCD, super-coiled domain; BRCTs, *BRCA1* C-Terminal motifs.

(B) Human SETX map. The N-terminal domain shared by human SETX and the yeast homolog Sen1 proteins is indicated in purple. The helicase domain is indicated in green. The position of the 7 helicase motifs (I–VI) is noted. The location of SETX antibody epitopes is indicated.

(C) Co-immunoprecipitation (IP) showing that the interaction between endogenous BRCA1 and endogenous SETX in HeLa cells is lost after depletion of either protein. These results confirm the specificity of the antibodies targeting BRCA1 and SETX. Input= 10% of protein used for co-IP. IgG, negative control.

(D) Schematic representation of the BRCA1 and SETX fragments produced in *E.coli* as GST fusion proteins and respectively tested for binding to SETX and BRCA1 by GST binding assays (Figure 1B).

For BRCA1, the RING domain is shown by the orange N-terminal box; exon 11 is indicated in black; and the 2 BRCT motifs, located near the C-terminus are in dark red.

The numbers represent BRCA1 amino acid residues. For SETX, the helicase domain is shown is green.

(E) BRCA1 fragment/BACH1 binding assays performed with GST-BRCA1 fragments and analyzed by immunoblotting. Blots generated with eluates of each GST-fused BRCA1 fragment were probed with anti-BACH1 antibody. Bottom panel: Relative abundance of each recombinant affinity purified fusion protein (marked with a star) visualized by SDS-PAGE after Coomassie staining. Additional bands could indicate abnormally translated or degradation products of the relevant recombinant proteins.

Figure S2. Validation of BRCA1 and SETX recruitment to β-actin termination sites. Related to Figure 1.

(A-C) Quantitative ChIP analyses performed with BRCA1 #3 (A), SETX #1 (B) and total RNAPII (C) antibodies or negative control IgG (blue bars) on the human β-actin gene in cells infected with lentiviruses encoding shGFP (red bars), shBRCA1 (green bars) or shSETX (grey bars). Mean values (± SD) of ChIP results derived from two independent experiments are shown. $p^* < 0.05$.

Figure S3. Identification of R-loop-associated ssDNA breaks by LM-PCR. Related to Figures 2 and 3.

(A) Immunoblot showing the extent of siRNA-mediated depletion of BRCA1 and SETX and the level of RNAseH1 expression in HeLa cells used for comet assays and ssDNA

break analysis. Immunoblots were performed with SETX #1 (top panel), BRCA1 SD118 (middle panel) and GFP (bottom panel) antibodies. Vinculin was used as a loading control.

(B) Detection of multiple LM-PCR products was undertaken after capture of the primer-extended and ligated fragments. This was followed by a regular PCR reaction with linker and GSP2 primers, respectively annealing to the adaptor and the far 3'-end of the β-actin gene. Analysis was performed by ethidium bromide-stained agarose gel electrophoresis. Red and yellow stars indicate the relevant PCR products that were selectively amplified in BRCA1- or SETX-depleted cells.

(C) Quantitative detection of ssDNA after LM-PCR analyses performed across several locations on the anti-sense strand of the β-actin gene in control and BRCA1 or SETX knockdown conditions. C, a 3'-end segment of the β-actin gene that lies downstream of its termination pause site.

Figure S4. ChIP analyses at the *Gemin7* and *CyclinB1* genes and Neutral comet Assay. Related to Figure 4.

(A-C) Quantitative ChIP analyses using antibodies directed against BRCA1 (A), SETX (B) or γH2AX/H2AX (C) performed on *Gemin7* and *CyclinB1* (CycB1) genes. Transcriptional termination of *Gemin7* and *CyclinB1* is mediated by R-loops (pause) and CoTC sequences, respectively. Intronic regions (Intr) of the two genes were used as

controls. ChIP profiles based on three independent experiments show mean values (± SD).

(D) Comet assay analysis performed in neutral buffer conditions. Histograms depict the quantitative analysis of comet tail lengths for each condition. Data are represented as the mean ± SEM from three independent experiments.

Figure S5. Expression analysis of BRCA1 TR genes. Related to Figure 5.

(A) Cumulative distribution plot comparing the expression of genes with (red curve) and without (blue curve) BRCA1 bound at termination regions. The level of mRNA expression increases with the value of log(RPKM+1); the gradient bar reflects the amplitude of expression (grey: low to black: high). Differences in expression levels between the red and blue groups were validated by the results of a Mann-Whitney-Wilcoxon statistical test (p=1.1E-7).

Figure S6. Mutational analysis of CoTC-associated genes. Related to Figure 6.

Global mutational analysis carried out in "R-loop free" CoTC genes using the complete whole-genome catalog of somatic mutations from 21 breast cancers (Nik-Zainal et al., 2012). The histogram illustrates the effect size comparison (one tailed CMH z-score) between the different tumor subgroups: BRCA1 mutant or BRCA2 mutant vs WT (non BRCA1/BRCA2 mutated) when testing the region ±4kb from a TTS.

Supplemental Tables

Table S1: List of BRCA1 peaks at termination regions, Related to Figure 5. List of BRCA1 ChIP-seq peaks identified within Termination Regions (TR) showing genomic information: chromosomes, coordinates, strand, gene name associated, RNA-seq RPKM values and overlap with RNAPIISer2P ChIP-seq peaks.

Table S2: List of the indels identified within the BRCA1 TR and CoTC genes, Related to Figure 6. BRCA1 TR and CoTC genes were tested for mutations within 21 breast cancers containing BRCA1 or BRCA2 germline mutations. Genomic information of the indels identified are shown.

Supplemental Experimental Procedures

Cell lines and culture conditions. HeLa cells and immortalized human fibroblasts (BJ-hTert) were grown at 37°C in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (GIBCO) in a 10% CO_2-containing atmosphere. To secure the inhibition of RNA polymerase II, we incubated the cells overnight with 10uM α-amanitin (Sigma) prior to harvesting the cells.

Plasmids and knockdown experiments: siRNA transfection and lentivirus infection. Transfection of GFP-RNAseH1 vectors (Skourti-Stathaki et al., 2011) was performed with Fugene6 transfection reagent (Promega) according to the manufacturer's instructions. The following siRNA reagents were used in this study: siBRCA1,

CAGCUACCCUUCCAUCAUA; siSETX, GCCAGAUCGUAUACAAUUA and siCt, GCGCGCUUUGUAGGAUUCG. All siRNAs were transfected into cells at 30nM (final concentration) in the presence of Lipofectamine RNAiMAX (Life Technologies). Cells were analyzed 72h post-transfection. These shRNAs were purchased from Sigma: shGFP, SHC005; shBRCA1, TRCN0000010305 and shSETX, TRCN0000051516.

GST binding assays. GST fragments (F1-F6 for BRCA1 and S1-S9 for SETX) have been described (Scully et al., 1997; Suraweera et al., 2009). C-terminal fragments of BRCA1 (F6N and F6C) were PCR amplified and cloned into the GST bacterial expression vector, pET42a. GST fusion proteins were generated in BL21(DE3)pLysS *E. coli* strains for 3h at 37°C after addition of 0.5 mM IPTG (isopropyl-β-D-thio-galactoside; Invitrogen). Bacteria were pelleted, resuspended in sonication buffer [50 mM Tris-HCl pH 7.6, 150 mM NaCl, 10% (v/v) glycerol, 0.5% (v/v) NP40, 1mM DTT and protease inhibitors (Roche)] and sonicated on ice. GST-tagged proteins were purified from bacterial lysates using Glutathione Sepharose 4B (GE Healthcare) and eluted from the beads with Elution Buffer (50 mM Tris pH 8, 150 mM NaCl, 10% (v/v) glycerol, 0.1% (v/v) NP40, 1mM DTT, 10mM reduced glutathione (Sigma) and protease inhibitors. Purified recombinant GST fusion proteins (2 μg of each) were bound to the Glutathione Sepharose 4B and incubated with a relevant pre-cleared WCE for 2h at 4°C with rotation. Beads were washed with immunoprecipitation buffer (detailed in Experimental Procedures supplemented with 1mM DTT and increasing concentrations of KCl (50, 100 and 150mM). Bound proteins were eluted in LDS sample buffer 1X (Invitrogen) and analyzed by immunoblotting.

Immunoblotting. Cells were lysed for 30 min on ice with NETN-420 [420 mM NaCl, 0.5 mM EDTA, 20 mM Tris-HCl (pH 8.0), 0.5% (v/v) Nonidet P-40 (NP-40)] and supplemented with protease inhibitors (Roche). Whole cell extracts (WCE) were centrifuged at 14,000rpm for 10min at 4ºC. Protein concentrations were measured using the BCA protein Assay Kit (Pierce). From each sample, 15ug of WCE were separated in NuPAGE® 3-8% Tris-Acetate precast gels (or 4-12% Bis-Tris for GST binding assays) and blotted onto 0.2μm nitrocellulose membranes. The following antibodies were used: mouse monoclonal anti-BRCA1 (SD118, Millipore), rabbit polyclonal anti-SETX (#1: A301-105A, Bethyl), rabbit polyclonal anti-BACH1 (B1310, Sigma) and mouse monoclonal anti-Vinculin (G-11, Santa Cruz).

Chromatin immunoprecipitation (ChIP). Cells were cross- linked in DMEM containing 1% formaldehyde at room temperature for 10 min with rotation. Cross-linking was stopped by the addition of glycine to a final concentration of 0.125 M. Cells were harvested and resuspended in cell lysis buffer (5mM PIPES, 85 mM KCl, 0.5% (v/v) NP40 and protease inhibitors) for 10 min on ice. Nuclei were then pelleted by centrifugation and resuspended in nuclear lysis buffer (25mM Tris-HCl pH 8, 5 mM EDTA, 1% (v/v) SDS, and protease inhibitors: Roche) for 10 min on ice. Chromatin was sheared using a cooled Bioruptor bath sonicator (Diagenode) for 20 min with 30 s pulses and 30 s pauses in order to obtain chromatin fragments with an average size of 500 bp. Quality and size of chromatin fragments was monitored by ethidium-bromide stained agarose gel electrophoresis after DNA purification. Chromatin was diluted at least 10 times into ChIP dilution buffer (1mM EDTA pH 8, 15mM Tris pH 8, 150mM NaCl

complemented with protease inhibitors). An aliquot representing 5% of the total amount of chromatin used for each IP was stored for further analysis. Chromatin (15 µg per condition) was then incubated overnight with protein A/G magnetic beads (Dynabeads, Invitrogen) coupled to the appropriate antibody: 5 µg of anti-BRCA1 (#3: A300-000A, Bethyl), 5 µg of anti-SETX (#1: A301-105A, Bethyl), 1 µg of anti–γH2AX (#05-636, Millipore), 1 µg of anti-histone H2A.X (#07-627, Millipore) and 5ug of total RNAPII (sc-9001, Santa Cruz). Immunoprecipitates were exposed to 8 serial washes for 5-10 min each on a rotating wheel at 4°C in the following buffers, complemented with protease inhibitors: 1X: 20mM Tris-HCl pH 8, 2mM EDTA, 0.1% SDS, 1% (v/v) Triton X-100 and 165mM NaCl; 1X: 20mM Tris-HCl pH 8, 2mM EDTA, 0.1% SDS, 1% (v/v) Triton X-100 and 500mM NaCl; 1X: 10mM Tris-HCl pH 8, 1mM EDTA, 1% (v/v) NP-40, 1% Na-deoxycholate and 250mM LiCl; 2X: 50mM HEPES pH 7.6, 1mM EDTA, 1% (v/v) NP-40, 0.7% Na-deoxycholate and 500mM LiCl and 3X with TE (10mM Tris.HCl pH7.5, 1mM EDTA). DNA was eluted from the magnetic beads with 0.1 M sodium carbonate and 1% (w/v) SDS. Samples were then brought to 300 mM NaCl and RNAse A (10 µg/ mL) and incubated at 65°C overnight to reverse crosslinks. DNA was purified with a PCR purification kit (Qiagen) according to the manufacturer's instructions. ChIP samples were quantified by quantitative real-time PCR using SYBR green mix (iTaq Universal, Biorad) and the primers of interest (see Table 1 for primers sequences). We analyzed the data using ABI PRISM 7900 Sequence Detection system software (Applied Biosystems). The results were calculated as %Input, representing the proportion of 100% Input (total amount of chromatin used for each IP) or as fold enrichment over a given region.

Comet Assay. Alkaline and neutral comet assays were performed on HeLa cells using a Single Cell Gel Electrophoresis Assay Kit (Trevigen) according to the manufacturer's instructions. 1500 cells were spotted onto each sample area. The CellProfiler software (Carpenter et al., 2006) was used for analysis and quantification of the results.

LM-PCR and quantification of ssDNA breaks. HeLa cells were harvested 72h after siRNA-mediated depletion of BRCA1 or SETX, and genomic DNA (gDNA) was routinely purified using a DNeasy® kit (Qiagen). The LM-1 and LM-2 oligonucleotides were annealed as described (Schlissel et al., 1993), and 300 pmoles of these adapters were phosphorylated with T4 Polynucleotide Kinase (New England Biolabs) and purified as described (Sambrook and Russell, 2006). Antisense, gene- specific primers (GSPs) for the β-actin gene were used to detect ssDNA breaks on the sense strand of the gDNA. The assay was performed using 1 μg of gDNA per sample. DNA was denatured and first annealed with biotinylated-GSP1 primers. Primer extension from the 3'-end of the β-actin gene was accomplished with Deep Vent® DNA polymerase (New England Biolabs). Then, the ds blunt end DNA was ligated to the phosphorylated asymmetric/5'-overhanged double-stranded adaptor for 2 hrs at 22°C using 2.5 U of T4 DNA ligase (New England Biolabs) and a final adaptor concentration of 1nM. Dynal M280 streptavidin-coated magnetic beads (Invitrogen) were used to capture the ligated and extended β-actin DNA fragments, according to manufacturer's instructions. The beads were resuspended in 0.1X TE, and the equivalent of 50-100 ng of purified and ligated

product were used as a template for a first round PCR reaction, using a nested gene specific primer (GSP2) and a linker primer. Samples were amplified for 12 cycles in a 3 step-protocol with Phusion High-Fidelity DNA polymerase in GC buffer (Thermo Scientific). PCR products were diluted (1/50) and processed by qPCR to amplify specific regions of interest along the β-actin gene, the 5'pause region and the pause region, itself, using the same primers that were used for ChIP experiments. We also included a negative control region, which was interrogated with primer C. It is located further downstream and outside of the primer- extended β-actin gene. As for ChIP experiments, the quantification of PCR products was analyzed with SYBR green (iTaq Universal, Biorad) using ABI PRISM 7900 Sequence Detection system software (Applied Biosystems). See Table 1 for primers sequences.

Genomic Analyses. Protein-coding transcripts were identified from an hg19 build based on GENCODE v10 annotations (Consortium, 2012; Harrow et al., 2012). Putative transcription termination regions (TR) were defined as a 4 kb window downstream from the transcription termination site (TTS, end of the 3' UTR). When there were multiple transcripts sharing the same gene ID, the most downstream TTS was used. Termination regions were filtered out when they overlapped promoter regions.

Using the same protein-coding transcripts as above, we defined promoter regions as being located within a sequence 1250 bp upstream and 250 bp downstream of the transcription start site (TSS=beginning of the 5' UTR). When there were multiple transcripts sharing the same gene ID, the most upstream TSS was used. To define BRCA1 peaks, we used the officially processed, HeLa-S3 BRCA1 peaks as called by

ENCODE (Consortium, 2012; Quinlan and Hall, 2010) with the SPP algorithm. The file is accessible at: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/wgEncodeAwgTfbsSydhHelas3Brca1a300IggrabUniPk.narrowPeak.gz. When investigating the overlap between termination regions and BRCA1 peaks, we filtered out any BRCA1 peaks that overlapped promoters and/or transcripts.

We defined a BRCA1 peak/TR overlap as that of a termination region overlapping a BRCA1 peak, with the use of Bedtools (Consortium, 2012; Djebali et al., 2012; Quinlan and Hall, 2010). We calculated the number of instances observed between any two groups, as well as the expected instances of overlap based on the hypothetical uniform distribution of the terminations regions (TR) and BRCA1 peaks. Specifically, let n_1 and n_2 be the respective number of regions of TR and BRCA1 peaks, while l_1 and l_2 is the respective total length of regions corresponding to TR and BRCA1 peaks. Furthermore, let x be the total size of the genome. Then, the expected instances of overlaps would be approximately $E[Overlaps] = (n_1*l_2+n_2*l_1)/x$, assuming uniform distributions of TR and BRCA1 peaks over the entire genome. When using all end sites (TR) and BRCA1 peaks, our search space was the size of the genome, while the use of filtered end sites and BRCA1 peaks allowed for the reduced search space of the size of the genome without promoters.

Using the expected and observed number of overlaps, we identified statistically significant enrichments for BRCA1 peaks at putative pause sites. To test the significance of the results, we used the upper tail of the cumulative Poisson distribution, with the expected number of overlaps as the mean.

For gene expression data, RNA-Seq quantifications were downloaded from the ENCODE RNA-Seq dashboard (Consortium, 2012; Djebali et al., 2012). The RPKM scores were exported from the CSHL HeLa-S3, Long PolyA+, whole-cell extract RNA-Seq data with GENCODE v10 annotations. The quantifications were generated using the GeneGencV10IAcuff quantification view. The file is accessible at http://genome.crg.es/~jlagarde//encode/pre-DCC/wgEncodeCshlLongRnaSeq/20120220_long_quantifications_gencodev10_cufflinks_cshl_NOT_SUBMITTED/LID16633-LID16634_GeneGencV10IAcuff.gff.

With the gene expression data, we were able to compare the expression of genes whose filtered end sites did or did not overlap the filtered BRCA1 peaks. We transformed the RPKM scores to log(RPKM+1), and generated a geometric box plot (Figure 5C) and a cumulative distribution plot (Figure S4) using ggplot2 (Harrow et al., 2012; Wickham, 2009). Any gene whose putative pause site was filtered out in the previous steps was not included. The statistical significance of the variation between the two groups was tested using the non-parametric two-sided Mann-Whitney test.

To evaluate the overlap between BRCA1 peaks at end sites and RNAPIISer2P peaks, we used the RNAPIISer2P ChIP-seq data available from ENCODE at: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhTfbs/wgEncodeSydhTfbsHelas3Pol2s2IggrabPk.narrowPeak.gz. To compute the p-value of BRCA1 TR and RNAPIISer2P peak overlap enrichment, we generated 10E7 random datasets having peak lengths and chromosomal distributions matching the experimental dataset. None of them had the same or a larger number of overlaps than the original file, hence $p<1.0E-7$. For the meta-analysis of co-incidence between BRCA1 peaks at ends sites and

R-loops, we used the DRIP-seq data (Ginno et al., 2012). The statistical significance of the variation between the DRIP and DRIP+RH samples was tested using the non-parametric two-sided paired Wilxocon test, as the data did not conform to normal distribution.

Functional enrichment analyses. Gene IDs of BRCA1 TR genes corresponding to genes with at least one BRCA1 binding site in the termination region were uploaded into Ingenuity Pathway Analysis (IPA) to search for biological and molecular functions as well as disease associations that were significantly enriched in the data set. Each p-value was determined by the Fisher's exact test and corresponds to the probability that, in a given category, we could reach the observed enrichment, or a larger one, by chance alone. Select enrichments for the BRCA1 TR gene set are shown in Figure 5F, and the full list is accessible in Table S2.

Mutational analyses of breast cancers. The BRCA1 TR genes as well as the CoTC "negative" genes (Nojima et al., 2013) were annotated in the hg19 build based on ENSEMBL format version 74 and scanned for each type of mutation. For each type of mutation, we computed a standardized mutation score (z-score) as follows. For each gene, we compared the observed number of mutations in BRCA1 or BRCA2 breast tumors to the number of mutations we would expect under the null hypothesis that the mutation rate in BRCA1 or BRCA2 breast tumor subgroup is equal to the mutation rate in the sporadic tumors (WT).

Then, for each subgroup of tumors, we computed the sum of the differences between observed and expected mutations across all genes; finally we standardized this sum as: z-score = (Total Observed – Total Expected) / Standard Deviation of Total Observed. This analysis produces a global standardized mutation score, where positive (or negative) values indicated a higher (or lower) number of mutations than expected under the null hypothesis. Since mutational events are rare, we computed the p-values by Monte-Carlo simulation. First, we drew 10,000 global standardized mutation scores from the exact null distribution. Under the null hypothesis, the mutation score is equal to a sum of transformed hypergeometric random variables, one for each gene. Second, we computed the p-value as the proportion of simulated mutation scores, which are greater than, or equal to, the observed mutation score.

Sequence of DNA primers

Name	**Sequence (5' →3')**
Human β-actin gene	
Prom (F)	GAGGGGAGAGGGGGTAAA
Prom (R)	AGCCATAAAAGGCAACTTTCG
In1 (F)	CGGGGTCTTTGTCTGAGC
In1 (R)	CAGTTAGCGCCCAAAGGAC
In3 (F)	TAACACTGGCTCGTGTGACAA
In3 (R)	AAGTGCAAAGAACACGGCTAA
In5 (F)	GGAGCTGTCACATCCAGGGTC
In5 (R)	TGCTGATCCACATCTGCTGG
5'pause (F)	TTACCCAGAGTGCAGGTGTG
5'pause (R)	CCCCAATAAGCAGGAACAGA
pause (F)	GGGACTATTTGGGGGTGTCT
pause (R)	TCCCATAGGTGAAGGCAAAG
C (F)	TGGGCCACTTAATCATTCAAC
C (R)	CCTCACTTCCAGACTGACAGC
D (F)	CAGTGGTGTGGTGTGATCTTG
D (R)	GGCAAAACCCTGTATCTGTGA
F (R)	CCATCACGTCCAGCCTATTT
F (F)	TGTGTGAGTCCAGGAGTTGG
Human ENSA gene	
Intr (F)	GCAACATGCTGGAAGAGAGAG
Intr (R)	AAACTACAGTGCCCCCTTAGC
3' end (F)	TCTGCCTGTATTTGTGTGCTG

3’ end (R)	GCAGCCCTCGTCTGTATAATG
Human Gemin7 gene	
Intr (F)	GATTCTATTTGGGCCACCTATG
Intr (R)	GGGAGGCATCTAAACCTCATC
3’ end (F)	AGCTCACGCTGGTTCTTTCTT
3’ end (R)	GAAATTCCAAAGGCGAGAGAC
Human Akirin1 gene	
Intr (F)	GGCATTACAGGAGTGCTACACA
Intr (R)	AATGGCCTACTCAATGCCTTC
3’ end (F)	TGTCTGTGGCAGTTTTCACAC
3’ end (R)	GGCCAGCACACTTTTCTGTAA
Human Cyclin B1 gene	
Intr (F)	CTTGGAAGGCCCATACTGATT
Intr (R)	GGAGTTGATGCTTGAGCTGTT
3’ end (F)	GCAGGAGGGTATGCCATCTA
3’ end (R)	TGCTTAAGGACTGGGATATGGT
LM-PCR	
Long Adaptor oligonucleotides (LM1)	GCATGGATGTTTTCCCAGTCACGACGTTGTGGGGACAAGTTTGTA CAAAAAAGCAGGCTTCCAAC
Short Adaptor oligonucleotides (LM2)	GTTGGAAGCCTGCTTTTTTGTACAAACTTGTCCCC
Linker primer	GCATGGATGTTTTCCCAGTCACGACGTTGT
GSP1 (5’-biotin)	GCAGAATCCAGACCTCAGCCCATAGCTAACCAGA
GSP2	TGTCTGGATGAACAGGTAGGAAT
SGSP1 (5’biotin)	TAACACTGGCTCGTGTGACAA
SGSP2	GAGCTGTCACATCCAGGGTC

Supplemental references

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biology *7*, R100.

Consortium, T.E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. Molecular Cell *45*, 814–825.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Research *22*, 1760–1774.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell *149*, 979–993.

Nojima, T., Dienstbier, M., Murphy, S., Proudfoot, N.J., and Dye, M.J. (2013). Definition of RNA Polymerase II CoTC Terminator Elements in the Human Genome. Cell Reports *3*, 1080–1092.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Sambrook, J., and Russell, D.W. (2006) Purification of Radiolabeled Oligonucleotides by Precipitation with Ethanol. Molecular Cloning. Cshprotocols.Cshlp.org.Ezp-Prod1.Hul.Harvard.Edu.

Schlissel, M., Constantinescu, A., Morrow, T., Baxter, M., and Peng, A. (1993). Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. Genes Dev. *7*, 2520–2532.

Scully, R., Chen, J., Plug, A., Xiao, Y., Weaver, D., Feunteun, J., Ashley, T., and Livingston, D.M. (1997). Association of BRCA1 with Rad51 in mitotic and meiotic cells. Cell *88*, 265–275.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Molecular Cell *42*, 794–805.

Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., and Lavin, M.F. (2009). Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. Human Molecular Genetics *18*, 3384–3396.

Wickham, H. (2009). ggplot2 (Springer).