Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes

Anne Helmrich,^{1,*} Monica Ballarino,¹ and Laszlo Tora^{1,*}

¹Department of Functional Genomics and Cancer, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U 964, Université de Strasbourg, BP 10142-67404 ILLKIRCH Cedex, France *Correspondence: anneh@igbmc.fr (A.H.), laszlo@igbmc.fr (L.T.)

DOI 10.1016/j.molcel.2011.10.013

SUMMARY

We show that the time required to transcribe human genes larger than 800 kb spans more than one complete cell cycle, while their transcription speed equals that of smaller genes. Independently of their expression status, we find the long genes to replicate late. Regions of concomitant transcription and replication in late S phase exhibit DNA break hot spots known as common fragile sites (CFSs). This CFS instability depends on the expression of the underlying long genes. We show that RNA:DNA hybrids (R-loops) form at sites of transcription/replication collisions and that RNase H1 functions to suppress CFS instability. In summary, our results show that, on the longest human genes, collisions of the transcription machinery with a replication fork are inevitable, creating R-loops and consequent CFS formation. Functional replication machinery needs to be involved in the resolution of conflicts between transcription and replication machineries to ensure genomic stability.

INTRODUCTION

Gene transcription and DNA replication are fundamental genetic processes that are essential for cell growth and division. They are carried out by large protein complexes progressing at high speed and for long distances along the chromosomes.

In mammalian cells, the 0.5 MDa twelve subunit DNA-dependent RNA polymerase II (Pol II) enzyme copies 18–72 nucleotides of DNA per second into RNA (Darzacq et al., 2007; Pérez-Ortín et al., 2007; Singh and Padgett, 2009). This Pol II transcription rate is independent of gene length and intron density (Singh and Padgett, 2009). The longest human gene, *CNTNAP2*, spans 2.3 Mb on genomic DNA. Depending on which published Pol II kinetic parameters one uses, this gene would be transcribed for a period of 9 to 35 hr. Indeed, another very long gene, the 2.2 Mb *DMD* locus, was shown to be transcribed over a 16 hr time course (Tennyson et al., 1995). Since fast-cycling mammalian cells have a cell-cycle time of about ten hours, such longterm transcription cycles would be expected to interfere with replication in the S phase of the cell cycle.

Prokaryotes have multiple mechanisms to minimize interference between replication and transcription (Brewer, 1988; Mirkin and Mirkin, 2007; Pomerantz and O'Donnell, 2008). In higher eukaryotes, replication and transcription are coordinated processes and, unlike in bacteria, they occur within spatially and temporally separated domains (Wei et al., 1998). Active transcription usually occurs in the G1 phase. When transcription occurs in S phase, it was suggested to be spatially separated from replication sites, as shown in human and mouse erythroid cells for the highly expressed beta-globin genes (Vieira et al., 2004).

Even though it has been established that transcription increases genomic instability, the molecular mechanisms involved remain poorly understood. In E. coli and Saccharomyces cerevisiae, transcription was shown to impede replication, thereby resulting in arrested replication forks (Mirkin and Mirkin, 2005; Prado and Aguilera, 2005). Transcription in budding yeast is known to induce genomic instability by blocking replication fork progression (Prado and Aguilera, 2005). The induction of gene expression in Chinese hamster cells provoked S-phase-dependent recombination events within the transcription unit, suggesting that collisions of the transcription and replication machineries also cause genomic instability in mammals (Gottipati et al., 2008). Furthermore, the perturbation of posttranscription processes like RNA splicing and export lead to the formation of DNA breaks, as has been described in yeast and mammalian cells depleted for Topoisomerase 1 (TOP1), THO, or ASF/SF2 (Huertas and Aguilera, 2003; Li and Manley, 2005; Gómez-González et al., 2009; Tuduri et al., 2009). These studies also showed a strong link between hyperrecombination phenotypes and the presence of slowed or stalled replication forks and R-loops. R-loops are RNA:DNA hybrids between nascent transcripts and the DNA template strand, while the nontemplate strand remains as single-stranded DNA (Reaban et al., 1994; Yu et al., 2003). The hydrolysis of RNA from such RNA:DNA hybrids by the RNase H1 enzyme suppresses both replication stress and genomic instability (Gómez-González et al., 2009; Tuduri et al., 2009). DNA breaks at the sites of RNA:DNA duplexes are thought to be caused either by hindered replication fork progression, or by exposure of the nontemplate DNA strand to cleavage and recombination (Aguilera and Gómez-González, 2008, and references therein). Besides their implication in genomic instability



Figure 1. Transcription throughout Long CFS-Associated Genes Takes More Than One Cell Cycle

(A–C) B-lymphoblasts were separated in four cell-cycle fractions (as indicated) and RNA was isolated. In (A), RT-qPCR quantification of *FHIT*, *WWOX* and *IMMP2L* pre-mRNAs. Transcript formation was monitored with different primer pairs (hybridizing to introns) along the CFS-associated genes. Colored bars (as indicated) show transcript quantification (±standard deviations) in the cell cycle fractions as % of total transcript quantification. In (B), RT-qPCR quantification of pre-mRNAs of short nonfragile genes from the four cell cycle fractions was carried out as control. Transcript quantification was calculated as in panel A. In (C), Schematic view of the transcription timing through the long genes *FHIT*, *WWOX* and *IMMP2L*. Arrows in the bottom indicate the regions of highest breakage at the CFSs FRA3B, FRA16D and FRA7K (Becker et al., 2002; Helmrich et al., 2007; Krummel et al., 2000; Wang et al., 1997).

formation, R-loop structures play physiological roles in immunoglobulin class switch recombination in human B cells (Yu et al., 2003) and in the promotion of transcription termination on the human β -actin gene (Skourti-Stathaki et al., 2011).

A second ribonuclease H, RNase H2, is able to hydrolyze RNA:DNA hybrids. RNase H2 removes the RNA primers during lagging-strand DNA synthesis. It also functions in the removal of single ribonucleoside monophosphates (rNMPs), which were misincorporated by DNA polymerases into duplex DNA (Rydberg and Game, 2002). Deficiency in RNase H2 and consequent retention of rNMPs results in replicative stress and genome instability (Nick McElhinny et al., 2010). Besides RNase H2, Top1 activity was shown to be involved in rNMPs removal from genomic DNA (Kim et al., 2011).

In human, genomic regions with increased replication fork stalling and DNA break formation are known as Common Fragile Sites (CFSs) (Durkin and Glover, 2007). CFSs are preferred regions of homologous recombination in normal cells (Hirsch, 1991) and they are hot spots for chromosomal rearrangements in the initial phase of oncogenic transformation (Gorgoulis et al., 2005). Deletion boundaries in many types of cancer map to CFS regions (Bignell et al., 2010; McAvoy et al., 2007), and CFS instability is involved in the amplification of oncogenes (Ciullo et al., 2002; Myllykangas et al., 2006). In normal cells, breakage at CFSs increases upon exposure to mild replication stress. Nevertheless, the molecular mechanism that links this replication stress induced fragility to specific genomic regions remains largely unclear. A growing number of CFSs have been mapped by hybridization experiments to the coding regions of long genes (Helmrich et al., 2006; Smith et al., 2007). Here we show that transcription of very long genes takes more than one cell cycle to complete, and that CFS instability is based on a spatial and temporal overlap of transcription and replication within these genes.

RESULTS

Long Genes Are Transcribed during More Than One Cell Cycle

To analyze the transcription timing of very long genes and to understand whether their transcription would interfere with replication and genome stability, we first identified 92 human genes, which are more than 800 kb in length (called long genes hereafter) (see Table S1). Next we analyzed the transcription timing of such long genes during the different phases of the cell cycle. To this end, we FACS separated nonsynchronized B-lymphoblasts into four different cell cycle fractions and quantified their pre-mRNA levels by RT-qPCR using oligonucleotides that hybridize to the nascent transcripts at every 195 kb on average. In this way we analyzed the specific pre-mRNA levels throughout the FHIT (1.5 Mb), WWOX (1.1 Mb) and IMMP2L (0.9 Mb) genes (Figure 1A). Primers hybridizing to the 5' ends of the three open reading frames (ORFs) identified transcripts mainly in the G2/M-phase, suggesting that all tested transcripts were mostly initiated in the G2/M-phase of the cell cycle. Using primer sets designed to more downstream regions we found that these genes continue to be transcribed throughout one complete round of the cell cycle and complete transcription in the next G1 or early S phase. We then tested the transcription timing of smaller genes. GAPDH (4 kb), CDKN1A/p21 (11 kb), C3orf67 (308 kb) and DOCK4 (480 kb) were all transcribed mainly within the G1-phase (Figure 1B). Cyclin B1 pre-mRNA was detected

Table 1. Long Genes' Transcription Rates				
Genes	Primer Positions	Primer Spacing (kb)	Region Transcribed during Cell Cycle Phases (Estimated Time)	Transcr Speed (nt/s)
FHIT	intron 1-intron 2	207	G2/M - G1;early S (2 hr)	28
	intron 2-intron 3	220	G1;early S - mid/late S (2 hr)	30
	intron 3-intron 4	200	mid/late S - late S; G2/M (2 hr)	28
	intron 4-intron 5	247	late S;G2/M - G2/M (2 hr)	34
	intron 5-intron 6	452	G2/M - G1 (3.5 hr)	36
	intron 6-intron 7	167	G1 - early S (1.5 hr)	31
	whole FHIT gene	1502	13 hr	32
wwox	intron 4-intron 5	178	G2/M - G1 (2 hr)	25
	intron 5-intron 6	113	G1 - G1;early S (1 hr)	31
	intron 6-intron 8a	245	G1;early S - late S (3.5 hr)	19
	intron 8a-intron 8b	542	late S - G1 (4.5 hr)	33
	whole WWOX gene	1113	11h	28
IMMP2L	intron 1-intron 2	73	G2/M - G2/M;G1 (1 hr)	20
	intron 2-intron 3a	298	G1 - mid S (3 hr)	28
	intron 3a-intron 3b	127	mid S - late S (1.5 hr)	24
	intron 3b-intron 4	175	late S - G2/M (2 hr)	24
	intron 4-intron 5	222	G2/M - G1;early S (3.5 hr)	18
	whole IMMP2L gene	899	11 hr	23

Positions of the primers and ratios of native transcripts within the cell cycle fractions are shown in Figure 1A. Cell cycle kinetics were analyzed by FACS (data not shown). Intragenic transcription speed was calculated by relating the primer distances to the duration of phase progression.

in the late S and G2/M phase fractions, reaching a maximum level in G2/M. Our observations agree with the previously described Cyclin B1 expression pattern (Kakino et al., 1996) and validate the methodology of our transcription timing analysis. The fact that *C3orf67*, neighboring the *FHIT* gene, and *DOCK4*, neighboring the *IMMP2L* gene, were found to be transcribed within the G1-phase (Figure 1B) indicates that the results described for *FHIT* and *IMMP2L* are indeed specific for these ORFs. Thus, our transcription timing analysis suggest that Pol II needs a very long time to complete the synthesis of *FHIT*, *WWOX* and *IMMP2L* full-length transcripts and that their transcription extends into the S phase of a subsequent cell cycle.

To estimate the transcription speed along these long genes, we measured the cell cycle kinetics of B-lymphoblasts by FACS analysis at different time points after release from an aphidicolin block (data not shown). Knowing that these cells need ~10 hr to complete one cell cycle, we estimated *IMMP2L* and *WWOX* genes to be transcribed within 11 hr, and *FHIT* within 13 hr. Based on these observations, the average transcription speed throughout the *IMMP2L*, *WWOX* and *FHIT* coding regions were calculated to be 23, 28 and 32 nt/s, respectively (Table 1). Intragenic transcription speeds between primer pairs were estimated to vary from 18 to 36 nt/s. These data show that long genes are transcribed with kinetics similar to smaller genes (Pérez-Ortín et al., 2007; Singh and Padgett, 2009).

Long Genes Exhibit CFS Instability Only When They Are Transcribed

As long genes are transcribed over more than one complete cell cycle, they are inevitably transcribed during S phase. Interestingly, human *FHIT*, *WWOX* and *IMMP2L* gene regions are known to harbor CFSs, called FRA3B, FRA16D and FRA7K, respectively. Thus, it is tempting to speculate that CFS formation may be provoked by collisions of transcribing RNA Pol IIs with replication forks in human cells. Importantly, the centers of breakage in *FHIT*, *WWOX* and *IMMP2L* have been narrowed down to regions including S phase transcribed areas (Helmrich et al., 2007; Krummel et al., 2000; Wang et al., 1997) (Figure 1A).

To test whether transcription is involved in CFS instability, we compared CFS break formation and expression levels of the underlying long genes. We choose B-lymphoblasts and myoblasts since the two cell types differ in their expression profiles of several long genes. *FHIT* and *WWOX* were expressed in human B-lymphoblasts, but not in myoblasts (Figure 2A), whereas *IMMP2L* mRNA was detected in both cell types, but to different levels. We also analyzed two of the longest human genes *CNTNAP2* (2.3 Mb) and *DMD* (2.2 Mb) (Table S1) of which the mRNAs were only detected in myoblasts (Figure 2A).

We next compared the CFS breakage frequencies of the five genes between the two cell types upon exposure to mild replication stress by aphidicolin. Metaphase spreads were hybridized with BAC probes mapping to the ORFs of the studied genes (Figure 2B). Each of them formed CFS breaks only in the cell type where the given gene was transcribed, but not in the cells in which it was silent. At the chromosomal locations of expressed smaller genes we did not detect any CFS breaks (Figure S1A and data not shown). Moreover, the *FHIT* and the *WWOX* genes are not transcribed in fibroblasts, where they were shown not to form CFSs (Figure S1B; Letessier et al., 2011). Even though this data shows that fragility is related to long genes, we did not find the level of instability to be simply proportional to the size of a gene, as we did not detect increasing lesion frequencies with



Figure 2. CFS Breakage Depends on the Active Transcription of Long Genes

(A) Upper panel: RT-qPCR quantification of expression levels of five long genes (as indicated) in B-lymphoblasts and myoblasts. mRNA levels relative to Cyclophylin B mRNA are represented (±standard deviations). Lower panel: FISH mapping of CFS break levels on the long genes (see panel B). The break frequency for each gene is given in % of total lesion numbers (2–4 independent experiments are represented with n = 30 cells). For smaller control genes, see Figure S1.

(B) Examples of CFS break mapping on the long genes by two color FISH. Two different BAC probes labeled with red and green fluorescent dies and hybridizing to 5' and 3' gene boundaries have been used. BACs were hybridized to metaphase chromosomes of aphidicolin treated B-lymphoblasts and myoblasts. DNA breaks map to the respective gene when a chromatin gap is bordered by red and green BAC hybridization at either side.

increasing gene size from 0.9 Mb (*IMMP2L*) to 2.3 Mb (*CNTNAP2*). Rather we find a good correlation between the five gene's expression levels and their respective CFS break frequencies (Pearson correlation coefficient r = 0.70 (p < 0.005), see also Table S2). Thus, we demonstrate the existence of a direct link between CFS formation and transcription.

Long Genes Replicate Late Independently of Their Expression Status

As break hot spots in the *FHIT*, *WWOX* and *IMMP2L* genes map to late S phase transcribed areas (Figure 1A), we wanted to test whether replication occurs simultaneously with transcription in these regions and whether the replication timing would change upon active transcription. To this end we compared the replication timings of the five long genes between B-lymphoblasts and myoblasts, where either the genes are transcribed or not (see above). We used BrdU-pulse labeling and FACS-sorting of the cells into G1, early S, mid early S, mid late S, late S, and G2/M fractions, followed by immunoprecipitation of BrdU-labeled DNA and real-time q-PCR to quantify the BrdU-containing freshly replicated DNA. As a control, we tested the replication timing of the highly expressed small nonfragile genes *Cyclin B1, GAPDH* and *p21*, and found them to replicate in early S phase (Figure 3A). These data agree with the observation that regions of actively transcribed genes tend to be correlated with early replication (White et al., 2004; Woodfine et al., 2004). Two further control genes HPRT and FMR1, replicated as previously reported in mid-early (Wang et al., 1999) and in late S phase (Palakodeti et al., 2004), respectively. Then, we analyzed the replication timing along the long CFS-associated genes and found that all five ORFs are either entirely late S phase replicating, or contain late S phase replicating areas close to the hotspots of breakage (Figure 3B; see Figures 1A and S2). Note that CFS break sites were found to be late replicating in both cell types, independently of transcription, and fragility. In line with our results, FHIT was recently described to be a late replicating gene in cells with various FRA3B fragility levels (Letessier et al., 2011). Our results show that late replication is a common feature of long genes and is independent of transcription. Importantly, FHIT, WWOX and IMMP2L CFSs map to regions where transcription and replication act simultaneously in late S phase (compare Figures 1A and 3B). In contrast, the early S phase replication of the Cyclin B1 gene does not overlap with its late S phase transcription (Figures 1B and 3A).

By showing that CFS breakage occurs only when transcription and replication take place at the same time on the same template, our results also suggest that CFS instability might be caused by a clash of replication and transcription machineries.

Myoblast **B-lymphoblast** 120 120 100 100 of BrdU positives of BrdU positives G2/M 80 80 Iate S I mid late S 60 60 mid early S early S 40 40 % % **G**1 20 20 0 GAPOH GAPOH HPRI Cyclin B1 0 Cyclin B1 FMRI 222 A HPRIFNRI 2º B CFS Genes **B-lymphoblast** 120 120 120 120 100 100 100 100 80 80 80 80 % of BrdU positives 60 60 60 60 40 40 40 40 I IT 20 20 20 20 0 n Myoblast 120 120 120 100 100 100 100 80 80 80 80 % of BrdU positives 60 60 60 60 40 40 40 40 20 20 20 20 0

A Control Genes





Figure 3. Replication Timing throughout Five CFS Gene Regions Is Late in S Phase

(A and B) Cells were separated in six cell cycle fractions (including four S phase fractions) and freshly replicated BrdU-containing DNA was isolated. Colored bars (as indicated) show DNA quantification in the given cell cycle fractions as % of total BrdU-containing DNA obtained in two independent experiments. Replication timing of the short control genes Cyclin B1, GAPDH, p21, HPRT, and FMR1 (A). Replication timing was monitored by using different primer pairs across the long CFS-associated genes (B). Transcribed (and break forming) long CFS-associated genes are shown on a gray, and nontranscribed genes on a white background. See also Figure S2. In panel (A) and (B), error bars represent ± standard deviations.

-11 Mb

DMD

₩₩₩

-

RNA:DNA Hybrids Form at Sites of Cotemporal Transcription and Replication

The presence of an RNA transcript stably bound to DNA in proximity to a replication fork may form an obstacle for the fork, and thus induce genomic instability.

To investigate the existence of RNA:DNA hybrids at CFSs, we adapted a method originally developed in yeast (Huertas and Aguilera, 2003) to mammalian cells (Experimental Procedures). To isolate RNA:DNA hybrids, nucleic acids from either nontreated or aphidicolin treated cells were incubated with RNase

1 Mb



Figure 4. In Vivo Detection of RNA:DNA Hybrids

Total nucleic acids were isolated from aphidicolin treated (+ aphidicolin) or untreated (- aphidicolin) lymphoblasts and digested with RNase A (Hybrids) or both RNases A and H to eliminate DNA/RNA hybrids (background) prior to DNase I treatment. In (A), the reverse transcriptase coupled qPCR quantification of RNA forming hybrids on 18S region is shown in respect to the total RNA. Error bars represent standard deviations (±) calculated on several replicates of PCR reactions done on two independent experiments. Asterisk (*) indicates significant difference (p = 0.055), defined by unpaired t test. In (B and C), nucleic acid samples were treated with different nucleases as indicated above the panels (the expected nucleic acids are depicted in the bottom panels), immobilized on nylon membranes and analyzed by slot blot hybridization using labeled DNA probes from either the *FHIT*-FRA3B (B) or the A in order to degrade free RNAs, prior to DNase I digestion. As negative control a sample was also treated with RNase H to degrade hybrids. As the formation of RNA:DNA hybrids was described in the region of the yeast 18S gene (El Hage et al., 2010), we tested the presence of such hybrids on the human ribosomal 18S RNA gene region. To this end, the different nuclease treated samples were used for RT-qPCR amplification with the corresponding primers. Our results show the formation of RNA:DNA hybrids at the human 18S RNA gene locus and their induction under replication stress (Figure 4A), indicating also that the modified method works in mammalian cells. We then guantified the RNA:DNA hybrids at FRA3B in FHIT intron 4. Slot-blot hybridization with α^{32} P-labeled PCR probes from the FRA3B region shows the presence of RNA:DNA hybrids at the CFS in nonstressed cells (Figure 4B, -aphidicolin). Upon aphidicolin treatment RNA:DNA hybrids increase about five times (Figures 4B, 4D, and S3, +aphidicolin). In contrast, at the transcribed nonfragile Cyclin B1 gene we did not detect RNA:DNA hybrids (Figure 4C). These data indicate that R-loops form specifically at CFSs.

As RNA:DNA hybrids are hydrolyzed by the endogenous RNase H1 enzyme we wanted to test its effect on the formation of CFSs. We either overexpressed or knocked down RNase H1 and counted the DNA break frequencies on mitotic chromosomes, originating mostly from transfected cells (Figures S4A and S4D). Spontaneous break numbers decreased in cells, which were transfected with an RNase H1 expression vector and conversely, breaks increased significantly in cells in which RNase H1 was knocked down by siRNA treatment (Figure 5A, left panel). We exposed the transfected cells to aphidicolin and observed also a positive effect of RNase H1 in preventing genome-wide DNA breakage caused by replication stress (Figure 5A, right panel). Importantly, when tested by two color FISH and cytogenetic lesion mapping, RNase H1 overexpression induced a decrease of chromosome breaks at the long CFS-associated genes FHIT, WWOX and IMMP2L. In contrast. RNase H1 knockdown led to an increase of breakage at these CFS regions (Figure 5B). Under the same conditions no chromosome breaks were detected at the nonexpressed long genes, CNTNAP2 and DMD, or at shorter nonfragile genes, p21 and Cyclin B1. These data show unambiguously that RNA:DNA duplex formation at the CFS-associated long genes participates in the high DNA break rate at CFSs.

RNase H2, Topoisomerase I, and Fast Pol II Elongation Are Not Implicated in CFS Stabilization

To test whether misincorporated ribonucleotides are implicated in CFS formation (see Introduction), we depleted either RNase H2 or TOP1 in B-lymphoblasts (Figures S4B and S4C) and quantified the resulting chromosome break levels. Genome-wide break levels were not affected by loss of RNase H2, but

nonfragile *Cyclin B1* (C) loci. See also Figure S3. In (D), histograms show the quantification of total FHIT pre-mRNA signal divided by the total DNA signal (left panel) and RNA:DNA hybrids forming in the *FHIT*-FRA3B region in the presence of aphidicolin compared to the nontreated sample (right panel). The mean values (\pm) obtained from two independent biological replicates (see also Figure S3) as data ranges are indicated.



Figure 5. RNase H1 Functions to Stabilize CFSs

(A) B-lymphoblasts were untransfected, mock transfected, transfected with an RNase H1 expression vector or with a mixture of four RNase H1 siRNAs (see Figure S4). Aphidicolin was added to half of the cells during 15 hr before harvest. Genome-wide DNA break frequencies were analyzed on metaphase spreads. Asterisk (*) indicates significant differences (p < 0.001), defined by unpaired t test. (B) Break frequencies at the long *FHIT*, *WWOX*, and *IMMP2L* genes were defined by a combination of two color FISH (see Figure 2B) and cytogenetic lesion mapping on aphidicolin treated cells. Asterisk (*) indicates significant decrease (p < 0.05), defined by unpaired t test.

(C and D) Genome-wide (C) and FHIT, WWOX or IMMP2L gene specific (D) break frequencies after siRNA knockdown (see Figure S4) of RNase H2 or TOP1. The experiment was carried out as described in (A). Asterisk (*) indicates significant decrease (p < 0.01), defined by unpaired t test.

(E) B-lymphoblasts were transfected with expression vectors coding for human α -amanitin sensitive (WTs), resistant (WTr), or α -amanitin resistant slow elongating mutant (C4r) forms of Pol II large subunits. Cells were treated or not with α -amanitin and aphidicolin (as indicated). DNA break frequencies (A–E) were determined on metaphase chromosomes of three experiments (with n = 30 cells each) and are shown as mean values. In panels (A–E) error bars represent ± standard deviations.

increased upon TOP1 knock-down by about 2.6 fold in untreated cells and about 1.7 fold in cells under replication stress (Figure 5C). Nevertheless, neither RNase H2 nor TOP1 knock-down affected the instability at long CFS-associated genes in aphidicolin treated cells (Figure 5D). These data prove that ribo-

nucleotides do not preferentially incorporate into CFS regions, suggesting that the mechanism leading to CFS instability under replication stress is independent from rNMP removal.

To test whether a slowed Pol II complex would increase R-loop-mediated breakage, we transfected B-lymphocytes with an expression vector encoding a α -amanitin resistant slow elongating variant of the largest subunit of Pol II (Rpb1), called C4r (de la Mata et al., 2003) (Figure 5E). As control, cells were transfected with expression vectors that encode wild-type Rpb1, either as a α -amanitin resistant (WTr) or as a sensitive form (WTs). 52 hr following transfection cells, which survive only in the presence of the slow Pol II (C4r), did not show significantly increased chromosome break frequencies (Figure 5E). Therefore, a slow Pol II elongation rate is not linked with RNA:DNA hybrid stabilization and CFS formation.

DISCUSSION

Collisions of transcription complexes with replication forks cause genetic instability in bacteria and yeast (Prado and Aguilera, 2005; Torres et al., 2004). In order to avoid such collisions, replication and transcription are spatially and temporally coordinated in eukaryotic cells. In this study we show that long genes are prone to DNA breakage in human cells due to the impossibility of separating transcription and replication at large transcription units. Our data indicates that transcript formation of genes, which span more than 800 kb extends into a second cell cycle and also into S phase. Unlike shorter genes, which are mainly transcribed in G1, the long genes initiate transcription in G2 or mitosis phase (Figure 1). Transcript formation continues while the cells pass through one cell cycle and Pol II reaches the termination site of these genes in G1 or early S phase of a following cell cycle. The attachment of forming pre-mRNAs to mitotic chromosomes may interfere with chromosome condensation, resulting in decondensed chromatin at CFSs. The fact that the 5' ends of the studied long genes are transcribed in G2/M phase (Figure 1) suggests that a new round of transcription will not start until replication is finished. We argue that ongoing replication acts as a negative regulator of transcription initiation on the long genes. However, RNA polymerases that are already productively elongating on these long genes at the time of replication initiation will proceed with elongation and thus, have a high risk to collide with the replication machinery.

The timing of replication is one important manner of coordinating transcription and replication. For instance, actively transcribed regions are typically early replicating (White et al., 2004; Woodfine et al., 2004). Our study shows that long genes form rather an exception to this rule, as we found all five actively transcribed long genes to replicate late. It was shown recently, that fragile regions within the long FHIT and WWOX genes are devoid of replication initiation events in B-lymphoblasts, whereas low levels of FRA3B and FRA16D breakage in fibroblasts are linked with origin activation throughout the whole gene region (Letessier et al., 2011). When we compared FHIT and WWOX expression levels between B-lymphoblasts and fibroblasts, we found transcription to be linked with fragility and with a lack of replication initiation (Figure S1B). A similar finding has been reported in yeast, where the ARS605 replication origin, being located within the open reading frame of MSH4 gene, is switched off by the activation of MSH4 transcription (Mori and Shirahige, 2007). Thus, it seems that ongoing transcription acts as a suppressor to replication initiation. This conclusion is further supported by the fact that CFSs are unable to activate replication origins for the rescue of perturbed forks caused by mild replication stress (Ozeri-Galai et al., 2011).

DNA polymerase impediment at highly transcribed genes was described to depend on an opposite orientation (French, 1992; Deshpande and Newlon, 1996), or to be independent of the direction of transcription and replication (Azvolinsky et al., 2009). In order to determine the directionality of collisions at human CFSs, we searched for obvious replication fork movements through the long genes by using public data on genomewide replication timing profiles (http://replicationdomain.org/; Weddington et al., 2008). All the analyzed CFSs map into large regions of consistent late replication, with no apparent replication initiation sites nearby (Figure S2 and Letessier et al., 2011). Therefore, it is impossible to determine the directionality of replication with respect to transcription on the studied long genes within a given cell.

When pre-mRNA processing is perturbed, R-loops are formed and replication forks are stalled thereby causing DNA doublestranded breaks (Gómez-González et al., 2009; Tuduri et al., 2009). Our experiments indicate that without interfering in RNA processing, R-loop dependent DNA breaks accumulate in long genes, especially when replication is mildly inhibited. In contrast to the stressed replication machinery, slowing down of Pol II elongation does not increase R-loop mediated breakage (Figure 5E). Therefore, RNA:DNA hybrids can form, and/or be stabilized, when the transcriptional elongation complexes are blocked, e.g., by collisions with the slowed replication machinery (Figures 5A, 5C, and 5E) or at natural Pol II pause sites at promoter proximal and 5' termination regions of the human β -actin gene (Skourti-Stathaki et al., 2011).

Even though rNMPs within genomic DNA have been associated with genomic instability in yeast, the depletion of neither RNase H2, the ribonuclease which removes rNMPs from duplex DNA, nor TOP1, the topoisomerase which nicks the DNA to allow rNMPs removal (Kim et al., 2011), did enhance instability at long CFS-associated genes in aphidicolin treated cells (Figure 5D). Therefore, rNMP-associated deletions occur independently of replication/transcription collisions. Nevertheless, loss of TOP1 caused a general increase in chromosomal lesions, and was previously shown to enhance the break rate of CFSs in replication nonstressed cells (Tuduri et al., 2009). We conclude that CFS instability is favored by hindered replication fork progression, which can arise either from a slowed DNA polymerase or from loss of TOP1 activity, separately. In agreement, another topoisomerase, TOP2, is implicated in maintaining genome integrity at regions of S phase transcription (Bermejo et al., 2009).

It has been shown in vitro that the *E. coli* replisome can progress through collision sites by displacing the RNA polymerase from the DNA and using the transcript as a replication primer (Pomerantz and O'Donnell, 2008). In bacteria in vivo, the replication machinery resolves potentially deleterious encounters at head-on and co-directional collisions between replication and transcription machineries (Merrikh et al., 2011). In human cells, it seems that even a mild inhibition of replication, like the one used in this study, can inhibit such a collision rescue mechanism. As a result, Pol II transcription elongation machinery stays



Figure 6. Under Replication Stress RNA Pol II and DNA Pol Collisions in Long Genes Result in R-Loop Formation and CFS Instability (A and B) The 1.5 Mb *FHIT* gene is shown as an example. Red arrows in the cell cycle schemes (on the left) represent the cell cycle stages where RNA Pol II transcription proceeds in the long gene. (A) Silent genes do not entail obstacles to the replication machinery and thus do not exhibit DNA instability. (B) In its active form, the long gene initiates transcription in the G2 or mitosis phase. RNA Pol II elongation continues throughout G1. When Pol II transcription progresses through S phase, it encounters a replication fork. The CFS break rate is low under nonstressed conditions, probably because the replication machinery is involved in resolving such collisions. But when an elongating Pol II collides with a perturbed replication fork including slowed DNA polymerases, stable R-loops form at sites of blocked Pol II, thereby facilitating CFS-related instability.

blocked, stable RNA:DNA hybrids form, CFSs appear driving the cell into genomic instability and tumor formation. Similar results have been proposed in bacteria (Dutta et al., 2011). While several studies indicate that transcriptional stalling provokes replication fork arrest (reviewed in Pomerantz and O'Donnell, 2010), our data shows that replication fork progression is necessary for preventing transcriptional pausing and maintaining genomic stability.

We show that collisions between transcription and replication, RNA:DNA hybrid formation and consequent DNA breakage are surprisingly frequent at long genes. Frequent transcriptional pausing and R-loop formation has also been defined at active rRNA genes (El Hage et al., 2010; and this study). To protect against genomic instability cells have evolved several mechanisms. By minimizing the R-loop formation on collision sites during S phase, RNase H1 reduces DNA fragmentation. Moreover, stalled replication forks are efficiently targeted by the checkpoint kinase ATR (ataxia-telangiectasia and Rad3-related). In agreement, loss of ATR or ATR downstream targets is linked with replication fork collapse and CFS break formation (Casper et al., 2002; Cha and Kleckner, 2002; Durkin et al., 2006; Houlard et al., 2011). Thus, the known function of ATR to restart replication forks may be involved in preventing R-loop stabilization at collision sites. Unrepaired CFS lesions in G1phase are targeted by 53BP1-OPT domains (Harrigan et al., 2011). All these studies prove the complexity of combined repair mechanisms in order to protect cells from CFS derived genomic instability.

Based on our data, we propose a model for human CFS formation (Figure 6). Genes of 800 kb or more in size produce their transcripts over more than one cell cycle, consequently extending transcription into the next S phase. At these sites of concurrent replication and transcription, a fully functional replication fork would help to resolve collisions between transcription and replication complexes. However, an impaired DNA polymerase (in the presence of replication stress) will block the elongating RNA polymerase and thus, lead to increased R-loop formation at Pol II pause sites. Such collisions result in CFS breakage and consequent genomic instability. Thus, we have identified RNA:DNA hybrids as possible therapeutic targets to reduce CFS-dependent genomic instability in oncogenic transformation.

EXPERIMENTAL PROCEDURES

Identification of Large Genes

Genome wide RefSeq genes were extracted from the UCSC Genome Browser (Assembly GRCh37/hg19) and sorted by gene length. Each gene was counted only once, with the size of its longest known transcript.

RNA Quantification

GM14907 and MHN cells were collected during their exponential growth phase. Total RNA was extracted with Tri Reagent (Molecular Research). 1.5 μ g of total RNA was reverse-transcribed with SuperScript II Reverse Transcriptase (Invitrogen). Real-time quantitative PCR reactions were set up using the QuantiTect SYBR Green Master Mix (Quiagen) and run on a LightCycler 480 (Roche). DNA contamination was quantified in reverse transcriptase free reactions and was subtracted from the RNA values. Primer sequences are listed in Table S3. RNA levels were calculated relative to Cyclophylin B mRNA quantities.

Transcription Timing Analysis

Living GM14907 cells were incubated with the DNA stain Hoechst 33342 (Sigma) and cells were flow sorted into G1, early S, late S2 and G2/M phase fractions (1×10^6 cells each). Total RNA was isolated, reverse-transcribed and quantified as described above. Primary transcripts were quantified by real time PCR by using intron specific primers (see Table S4).

Transcription Speed Calculation

The cell cycle kinetics of GM14907 cells was analyzed by a 20h 4µg/ml aphidicolin block followed by propidium iodide staining and cell cycle analysis using a FACS Calibur Analyzer at various times postrelease. Primer positions were mapped according to the UCSC Genome Browser (GRCh37/hg19 assembly) and inter-primer distances were calculated. Intragenic transcription speed values were calculated by dividing the primer distances by the duration of passing from one cell cycle phase to another.

In Vivo Detection of RNA:DNA Hybrids

DNA and RNA were isolated using a modification of the method used in yeast (Huertas and Aguilera, 2003). 1x10⁹ GM14907 cells were grown in the presence or absence of 0.4 μM aphidicolin for 15 hr, washed twice with PBS and resuspended in 50 mM Tris-HCL (pH7.9), 1 mM EDTA, 1 mM DTT, 0.5 mM PMSF prior to douncing with a loose-fitting Dounce pestle. After centrifugation, pellets were resuspended in 1% CTAB, 700 mM NaCl, 50 mM Tris-HCL (pH 7.6), 12.5 mM EDTA dounced and treated with 1 mg/ml of proteinase K. All buffers were prepared with DEPC water. After 2 hr at 37°C, nucleic acids were isolated and treated with 5 units RNase H and/or 15 μg RNase A. One percent of the sample treated with both the RNases was stored as DNA sample before DNase I was used to remove genomic DNA contaminations (Huertas and Aguilera, 2003). Real-time qPCR on the ribosomal 18S RNA gene region (primer sequences were 5'-aaacggctaccacatccaag-3' and 5'-cctccaatggatcctcgtta-3') was used as a positive control. All samples were immobilized on Hybond-N membranes and hybridized at 42°C with a α^{32} P-labeled PCR probe from the *FHIT* intronic FRA3B (primer sequences were 5'-gtaagaagtgcctttcacctc-3' and 5'-cactgggactgtctgctttcca-3') and the Cyclin B1 (primer sequences were 5'-ccctccagaaattggtgact-3' and 5'-ccccctaatgtacccatacg-3') regions. Densitometric analysis was performed using Typhoon Imager (GE Healthcare) and Optiquant software. Background was substracted from R-loop levels and the values were related to total RNA quantities.

Plasmids and Transfection

GM14907 cells were electroporated with 960 μ F, 320 V. The expression vectors for human α -amanitin resistant Pol II large subunit (WTr), slow mutant (C4r) and α -amanitin sensitive Pol II large subunit Rpb1 (WTs) have been previously described (de la Mata et al., 2003). Cells were also electroporated with the pCMV6-XL5-RNaseH1 vector (Origene), the pHygEGFP reporter vector (Clontech), or different siRNAs pools (containing four different siRNAs each) from the Dharmacon SMARTpool: hRNaseH1 siRNA (ref. number L-012595-01), hRNaseH2 siRNA (ref. number L-003535-01), hTOP1 siRNA (ref. number L-005278-00), or Scramble siRNA (ref. number D-001810-10) as control. Cells were analyzed 50 hr after transfection. The wild-type Pol II in the Rpb1 transfected cells was inhibited by exposure to α -amanitin (5 μ g/ml) for 15 hr and DNA polymerase was hindered by addition of aphidicolin (0.4 μ M) for 15 hr

Western Blot Analysis

Whole-cell protein extracts were prepared from GM14907 cells lysed in 400mM KCI, 20mM Tris-HCI (pH 7.5), 2mM DTT, 20% glycerol and protease inhibitor C-Complete (Roche). Extracts were separated by electrophoresis on 10% poly-acrylamide gel (Invitrogen) and electroblotted onto nitrocellulose membrane (Protran, Norway); immunoblots were incubated overnight with anti-TOP1 (Abcam, ab3825) polyclonal antibody and anti-RNase H1 (Abcam, ab56560), anti-RNase H2 (Abcam, ab92876), and GAPDH (Santa Cruz Biotechnology, sc-32233) monoclonal antibodies. Immunoreactivity was determined by using the ECL method (GE Healthcare) according to the manufacturer's instructions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, four tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.molcel.2011.10.013.

ACKNOWLEDGMENTS

We are grateful to A. Coquelle and P. Pasero for discussions and for the RNase H1 expression vector, to T. Decker for advice on B-lymphoblast transfection, to D.M. Gilbert for discussion, and to A.R. Kornblihtt for the Pol II large subunit expression vectors. We thank A.R. Krebs, M.E. Torres-Padilla, E. Soutoglou, and M. Rebagliati for critical reading of the manuscript. A.H. was supported by postdoctoral fellowships from the Deutsche Forschungsgemeinschaft, and AH and MB by La Fondation pour la Recherche Médicale. This work was funded by grants from ANR (GenomATAC; ANR-09-BLAN-0266), AICR (09-0258), EU (EPIDIACAN) and CNRS (LEA-SkinChroma) grants.

Received: March 15, 2011 Revised: July 27, 2011 Accepted: October 7, 2011 Published: December 22, 2011

REFERENCES

Aguilera, A., and Gómez-González, B. (2008). Genome instability: a mechanistic view of its causes and consequences. Nat. Rev. Genet. 9, 204–217.

Azvolinsky, A., Giresi, P.G., Lieb, J.D., and Zakian, V.A. (2009). Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. Mol. Cell *34*, 722–734.

Becker, N.A., Thorland, E.C., Denison, S.R., Phillips, L.A., and Smith, D.I. (2002). Evidence that instability within the FRA3B region extends four megabases. Oncogene *21*, 8713–8722.

Bermejo, R., Capra, T., Gonzalez-Huici, V., Fachinetti, D., Cocito, A., Natoli, G., Katou, Y., Mori, H., Kurokawa, K., Shirahige, K., and Foiani, M. (2009). Genome-organizing factors Top2 and Hmo1 prevent chromosome fragility at sites of S phase transcription. Cell *138*, 870–884.

Bignell, G.R., Greenman, C.D., Davies, H., Butler, A.P., Edkins, S., Andrews, J.M., Buck, G., Chen, L., Beare, D., Latimer, C., et al. (2010). Signatures of mutation and selection in the cancer genome. Nature *463*, 893–898.

Brewer, B.J. (1988). When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. Cell 53, 679–686.

Casper, A.M., Nghiem, P., Arlt, M.F., and Glover, T.W. (2002). ATR regulates fragile site stability. Cell *111*, 779–789.

Cha, R.S., and Kleckner, N. (2002). ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. Science 297, 602–606.

Ciullo, M., Debily, M.A., Rozier, L., Autiero, M., Billault, A., Mayau, V., El Marhomy, S., Guardiola, J., Bernheim, A., Coullin, P., et al. (2002). Initiation of the breakage-fusion-bridge mechanism through common fragile site activation in human breast cancer cells: the model of PIP gene duplication from a break at FRA7I. Hum. Mol. Genet. *11*, 2887–2894.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. Nat. Struct. Mol. Biol. *14*, 796–806.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol. Cell *12*, 525–532.

Deshpande, A.M., and Newlon, C.S. (1996). DNA replication fork pause sites dependent on transcription. Science *272*, 1030–1033.

Durkin, S.G., and Glover, T.W. (2007). Chromosome fragile sites. Annu. Rev. Genet. *41*, 169–192.

Durkin, S.G., Arlt, M.F., Howlett, N.G., and Glover, T.W. (2006). Depletion of CHK1, but not CHK2, induces chromosomal instability and breaks at common fragile sites. Oncogene *25*, 4381–4388.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in E. coli. Cell *146*, 533–543.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev. *24*, 1546–1558.

French, S. (1992). Consequences of replication fork movement through transcription units in vivo. Science *258*, 1362–1365.

Gómez-González, B., Felipe-Abrio, I., and Aguilera, A. (2009). The S-phase checkpoint is required to respond to R-loops accumulated in THO mutants. Mol. Cell. Biol. *29*, 5203–5213.

Gorgoulis, V.G., Vassiliou, L.V., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., Ditullio, R.A., Jr., Kastrinakis, N.G., Levy, B., et al. (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature *434*, 907–913.

Gottipati, P., Cassel, T.N., Savolainen, L., and Helleday, T. (2008). Transcription-associated recombination is dependent on replication in Mammalian cells. Mol. Cell. Biol. 28, 154–164.

Harrigan, J.A., Belotserkovskaya, R., Coates, J., Dimitrova, D.S., Polo, S.E., Bradshaw, C.R., Fraser, P., and Jackson, S.P. (2011). Replication stress induces 53BP1-containing OPT domains in G1 cells. J. Cell Biol. 193, 97–108.

Helmrich, A., Stout-Weider, K., Hermann, K., Schrock, E., and Heiden, T. (2006). Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. Genome Res. *16*, 1222–1230.

Helmrich, A., Stout-Weider, K., Matthaei, A., Hermann, K., Heiden, T., and Schrock, E. (2007). Identification of the human/mouse syntenic common fragile site FRA7K/Fra12C1—relation of FRA7K and other human common fragile sites on chromosome 7 to evolutionary breakpoints. Int. J. Cancer *120*, 48–54.

Hirsch, B. (1991). Sister chromatid exchanges are preferentially induced at expressed and nonexpressed common fragile sites. Hum. Genet. 87, 302–306.

Houlard, M., Artus, J., Léguillier, T., Vandormael-Pournin, S., and Cohen-Tannoudji, M. (2011). DNA-RNA hybrids contribute to the replication dependent genomic instability induced by Omcg1 deficiency. Cell Cycle 10, 108-117.

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

Kakino, S., Sasaki, K., Kurose, A., and Ito, H. (1996). Intracellular localization of cyclin B1 during the cell cycle in glioma cells. Cytometry 24, 49–54.

Kim, N., Huang, S.N., Williams, J.S., Li, Y.C., Clark, A.B., Cho, J.E., Kunkel, T.A., Pommier, Y., and Jinks-Robertson, S. (2011). Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. Science *332*, 1561–1564.

Krummel, K.A., Roberts, L.R., Kawakami, M., Glover, T.W., and Smith, D.I. (2000). The characterization of the common fragile site FRA16D and its involvement in multiple myeloma translocations. Genomics *69*, 37–46.

Letessier, A., Millot, G.A., Koundrioukoff, S., Lachagès, A.M., Vogt, N., Hansen, R.S., Malfoy, B., Brison, O., and Debatisse, M. (2011). Cell-typespecific replication initiation programs set fragility of the FRA3B fragile site. Nature 470, 120–123.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365–378.

McAvoy, S., Ganapathiraju, S.C., Ducharme-Smith, A.L., Pritchett, J.R., Kosari, F., Perez, D.S., Zhu, Y., James, C.D., and Smith, D.I. (2007). Non-random inactivation of large common fragile site genes in different cancers. Cytogenet. Genome Res. *118*, 260–269.

Merrikh, H., Machón, C., Grainger, W.H., Grossman, A.D., and Soultanas, P. (2011). Co-directional replication-transcription conflicts lead to replication restart. Nature 470, 554–557.

Mirkin, E.V., and Mirkin, S.M. (2005). Mechanisms of transcription-replication collisions in bacteria. Mol. Cell. Biol. *25*, 888–895.

Mirkin, E.V., and Mirkin, S.M. (2007). Replication fork stalling at natural impediments. Microbiol. Mol. Biol. Rev. 71, 13–35.

Mori, S., and Shirahige, K. (2007). Perturbation of the activity of replication origin by meiosis-specific transcription. J. Biol. Chem. 282, 4447–4452.

Myllykangas, S., Himberg, J., Böhling, T., Nagy, B., Hollmén, J., and Knuutila, S. (2006). DNA copy number amplification profiling of human neoplasms. Oncogene *25*, 7324–7332.

Nick McElhinny, S.A., Kumar, D., Clark, A.B., Watt, D.L., Watts, B.E., Lundström, E.B., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Genome instability due to ribonucleotide incorporation into DNA. Nat. Chem. Biol. *6*, 774–781.

Ozeri-Galai, E., Lebofsky, R., Rahat, A., Bester, A.C., Bensimon, A., and Kerem, B. (2011). Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. Mol. Cell *43*, 122–131.

Palakodeti, A., Han, Y., Jiang, Y., and Le Beau, M.M. (2004). The role of late/ slow replication of the FRA16D in common fragile site induction. Genes Chromosomes Cancer *39*, 71–76.

Pérez-Ortín, J.E., Alepuz, P.M., and Moreno, J. (2007). Genomics and gene transcription kinetics in yeast. Trends Genet. 23, 250–257.

Pomerantz, R.T., and O'Donnell, M. (2008). The replisome uses mRNA as a primer after colliding with RNA polymerase. Nature *456*, 762–766.

Pomerantz, R.T., and O'Donnell, M. (2010). Polymerase trafficking: A role for transcription factors in preventing replication fork arrest. Transcr 1, 136–139.

Prado, F., and Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. EMBO J. *24*, 1267–1276.

Reaban, M.E., Lebowitz, J., and Griffin, J.A. (1994). Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. J. Biol. Chem. *269*, 21850–21857.

Rydberg, B., and Game, J. (2002). Excision of misincorporated ribonucleotides in DNA by RNase H (type 2) and FEN-1 in cell-free extracts. Proc. Natl. Acad. Sci. USA *99*, 16654–16659.

Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nat. Struct. Mol. Biol. *16*, 1128–1133.

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.

Smith, D.I., McAvoy, S., Zhu, Y., and Perez, D.S. (2007). Large common fragile site genes and cancer. Semin. Cancer Biol. *17*, 31–41.

Tennyson, C.N., Klamut, H.J., and Worton, R.G. (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat. Genet. *9*, 184–190.

Torres, J.Z., Schnakenberg, S.L., and Zakian, V.A. (2004). Saccharomyces cerevisiae Rrm3p DNA helicase promotes genome integrity by preventing replication fork stalling: viability of rrm3 cells requires the intra-S-phase check-point and fork restart activities. Mol. Cell. Biol. *24*, 3198–3212.

Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., et al. (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat. Cell Biol. *11*, 1315–1324.

Vieira, K.F., Levings, P.P., Hill, M.A., Crusselle, V.J., Kang, S.H., Engel, J.D., and Bungert, J. (2004). Recruitment of transcription complexes to the betaglobin gene locus in vivo and in vitro. J. Biol. Chem. *279*, 50350–50357. Wang, L., Paradee, W., Mullins, C., Shridhar, R., Rosati, R., Wilke, C.M., Glover, T.W., and Smith, D.I. (1997). Aphidicolin-induced FRA3B breakpoints cluster in two distinct regions. Genomics *41*, 485–488.

Wang, L., Darling, J., Zhang, J.S., Huang, H., Liu, W., and Smith, D.I. (1999). Allele-specific late replication and fragility of the most active common fragile site, FRA3B. Hum. Mol. Genet. *8*, 431–437.

Weddington, N., Stuy, A., Hiratani, I., Ryba, T., Yokochi, T., and Gilbert, D.M. (2008). ReplicationDomain: a visualization tool and comparative database for genome-wide replication timing data. BMC Bioinformatics *9*, 530.

Wei, X., Samarabandu, J., Devdhar, R.S., Siegel, A.J., Acharya, R., and Berezney, R. (1998). Segregation of transcription and replication sites into higher order domains. Science *281*, 1502–1506.

White, E.J., Emanuelsson, O., Scalzo, D., Royce, T., Kosak, S., Oakeley, E.J., Weissman, S., Gerstein, M., Groudine, M., Snyder, M., and Schübeler, D. (2004). DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. Proc. Natl. Acad. Sci. USA *101*, 17771–17776.

Woodfine, K., Fiegler, H., Beare, D.M., Collins, J.E., McCann, O.T., Young, B.D., Debernardi, S., Mott, R., Dunham, I., and Carter, N.P. (2004). Replication timing of the human genome. Hum. Mol. Genet. *13*, 191–202.

Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat. Immunol. *4*, 442–451.