

# Duplex DNA from Sites of Helicase-Polymerase Uncoupling Links Non-B DNA Structure Formation to Replicative Stress

CAMILLE AMPARO<sup>1,2</sup>, JARROD CLARK<sup>1,2</sup>, VICTORIA BEDELL<sup>3</sup>,  
JOYCE L. MURATA-COLLINS<sup>3</sup>, MARIANNA MARTELLA<sup>4,5</sup>, FLAVIA PICHIORRI<sup>4,5</sup>, EMILY F. WARNER<sup>6</sup>,  
MAHMOUD ABDELHAMID<sup>6</sup>, ZOË A. E. WALLER<sup>6</sup> and STEVEN S. SMITH<sup>2,5</sup>

<sup>1</sup>*Division of Urology, City of Hope National Medical Center, Duarte, CA, U.S.A.;*

<sup>2</sup>*Beckman Research Institute, City of Hope, Duarte, CA, U.S.A.;*

<sup>3</sup>*Division of Cytogenetics, City of Hope National Medical Center, Duarte, CA, U.S.A.;*

<sup>4</sup>*Judy and Bernard Briskin Center for Multiple Myeloma Research,  
City of Hope National Medical Center, Duarte, CA, U.S.A.;*

<sup>5</sup>*Hematological Malignancies and Translational Science,  
City of Hope National Medical Center, Duarte, CA, U.S.A.;*

<sup>6</sup>*School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich, U.K.*

**Abstract.** *Background: Replication impediments can produce helicase-polymerase uncoupling allowing lagging strand synthesis to continue for as much as 6 kb from the site of the impediment. Materials and Methods: We developed a cloning procedure designed to recover fragments from lagging strand near the helicase halt site. Results: A total of 62% of clones from a p53-deficient tumor cell line (PC3) and 33% of the clones from a primary cell line (HPS-191) were within 5 kb of a G-quadruplex forming sequence. Analyses of a RACK7 gene sequence, that was cloned multiple times from the PC3 line, revealed multiple deletions in region about 1 kb from the cloned region that was present in a non-B conformation. Sequences from the region formed G-quadruplex and i-motif structures under physiological conditions. Conclusion: Defects in components of non-B structure suppression systems (e.g. p53 helicase targeting) promote replication-linked damage selectively targeted to sequences prone to G-quadruplex and i-motif formation.*

Non-B DNA structure formation must be suppressed during replication in order for replication to proceed properly. Many such structures occur at DNA sequences that exhibit GC-

skew where cytidine residues are largely confined to one strand and guanidine residues are largely confined to the other. In addition to R-Loops (1), these sequences are capable of forming G-quadruplexes (2, 3) triple helices (4) and i-motifs (5). At physiological pH, certain sequences are capable of forming several different structures (6, 7), however, the most commonly seen structures in these regions are the R-Loop, G-quadruplex and, less frequently, the triple helix (7) and the i-motif (8). DNA damage associated with R-loop formation is generally associated with transcription blocks (9, 10), but can be associated with replication. There is also considerable evidence that G-quadruplex formation can cause both epigenetic (11, 12) and genetic damage associated with DNA replication when G-quadruplex suppression systems are compromised (13-15). Here it is important to point out that G4-seq (60) results confirm the earlier report (16) that the frequency of G-quadruplex forming sequences increases with biological complexity suggesting that these sequences have been under positive selective pressure during evolution of higher organisms. To facilitate this increased frequency, the systems that promote replication through G-quadruplex structures or unwind these structures during replication and transcription of DNA appear to have evolved increased sophistication and redundancy in higher organisms. For example the Y polymerases REV1, Pol  $\kappa$  and Pol  $\eta$  are known to process G-quadruplex motifs when replication is slowed by them (17). Further, several helicases including Pif1 (15), and the superfamily 2 helicases WRN (18), BLM (19) and FANCI (20) have all been shown to unwind G-quadruplex DNA structures *in vitro*. Even so, WRN (21), BLM (22), Pif1 (23), and FANCI (24) are also capable of unwinding RNA:DNA hybrids *in vitro*.

This article is freely accessible online.

*Correspondence to:* Steven S. Smith, BRC, Room 2018, City of Hope National Medical Center, 1500 E. Duarte Rd., Duarte, CA 91010, U.S.A. E-mail: ssmith@coh.org

*Key Words:* CMG helicase uncoupling, replication stress, G-quadruplex, i-motif.

In one of the most carefully studied instances, loss of function in the FANCD1 homolog *dog-1* in *C. elegans*, produces selective deletion mutagenesis at sites where persistent G-quadruplex and i-motif structures can occur (14). In this case, deletion was apparently produced *via* a polymerase theta end-joining mechanism (25, 26). These data and those of Paeschke *et al.* (15), suggest that when non-B suppression is compromised the resulting structures can produce strong impediments to leading strand progress during replication. Leading strand impediments are known to cause Cdc45-Mcm2-7-GINs (CMG) helicase complex and lagging strand synthesis to become uncoupled from leading strand synthesis (27), permitting lagging strand synthesis to continue for an extended distance beyond a stalled replication fork (28). We hypothesized that the structure formed due to non-B structure induced uncoupling could be exploited to identify sites of recurring non-B structure. As can be seen in Figure 1, short double-stranded DNA fragments originating from the lagging strand at or near the CMG halt site should be produced after shear fragmentation or mild chemical fragmentation of genomic DNA.

In this report, we tested this hypothesis using ligation-mediated PCR to clone spontaneously- and chemically-induced DNA fragments from isolated genomic DNA and then mapped the isolated sequences to the human genome. A comparison of cloned sequences from a primary prostate stromal cell line (HPS-19I) with a normal karyotype with those of a prostate tumor cell line with a heavily rearranged karyotype (PC3) showed that the tumor cell line yielded about twice as many clones adjacent to a sequence with G-quadruplex forming potential as the normal cell line. Further, overlapping clones originating from the same genomic sequences were obtained only in the tumor cell line. One sequence, which was obtained multiple times in the screen of the tumor cell line was examined in detail. A region upstream from the cloned sequences, was shown to be accessible to bisulfite modification under non-denaturing conditions indicating that it was in a non-B conformation in genomic DNA (29). *In vitro* analysis of oligodeoxynucleotides corresponding to the site of chemical attack showed that a stable G-quadruplex structure can form on the G-rich strand and that an i-motif structure can form on the C-rich strand under physiological conditions. Sequencing analyses of the region showed that the sequence was subject to both deletion as seen in *C. elegans* (14), as well as apparent slippage and deamination mutagenesis downstream from the site of quadruplex formation. Models of processes that can generate these effects are discussed.

## Materials and Methods

**Cell culture and DNA isolation.** Human prostate cancer cells (PC3) were cultured as previously described (30). Human prostate stromal cells (HPS-19I) were a gift of Prof. David R. Rowley Baylor

University College of Medicine, Houston, TX, USA. They were cultured as described (31). DNA was isolated from each cell line using the QIAamp® Blood Mini kit (Qiagen Sciences, Frederick, MD, USA) using the instructions supplied by the manufacturer.

**Karyotyping.** Cells were harvested, fixed with Carnoy fixative, dropped onto glass slides and aged overnight at 60°C. Following standard GTG banding, 50 metaphases were captured with GenASIS Bandview software (Applied Spectral Imaging, Carlsbad, CA, USA). A 24-color SKY spectral karyotypic was performed, using a standard protocol. Briefly, after RNase and Pepsin treatment, DNA in metaphases was denatured at 70°C for 2 min followed by three ice-cold ethanol washes with 70%, 80%, and 90% ethanol. Denatured SKY™ 24-color probes were hybridized to the slide preparation for 48 h at 37°C. After a wash with formamide containing SSC and incubations with 1xSSC and 4xSSC, images were captured under Vectashield DAPI using the HiSKY system (Applied Spectral Imaging, Carlsbad, CA).

**Preparation of random libraries with clones from uncoupled replication forks.** Ligation-mediated PCR was used to amplify DNA fragments spontaneously generated during DNA isolation, or after bisulfite treatment of genomic DNA at 37°C (29) to promote chemical breakage of the DNA (32) at single stranded regions. Under these conditions, bisulfite cannot attack duplex DNA so the final treatment with sodium hydroxide was omitted to preserve duplex structure.

Two DNA preparations from each cell line were treated with CIP (Calf Intestinal Phosphatase, New England Biolabs, USA) in a 30 µl reaction containing 10 U of CIP, in 50 mM Potassium Acetate, 20 mM Tris-acetate 10 mM Magnesium Acetate 100 µg/ml BSA pH 7.9 for 1 h at 37°C to prevent ligation concatamers. However, concatamers were not observed when CIP treatment was omitted, indicating that CIP treatment was unnecessary. Two linkers were employed to form the duplex linker: Linker\_1 (5'-AGAAGCTTGAATTCGAGCAGTCAG-3') was annealed to linker\_2 (5'-CTGCTCGAATTCAAGCTTCT-3'). A total of 2.5 µl of 100 µM of each oligodeoxynucleotide was brought to a final volume of 45 µl and incubated for 2 min at 94°C, 5 min at 70°C, and for 5 min at 50°C. The duplex thus formed was allowed to cool to room temperature. The duplex was then diluted to a final volume of 250 µl for a 1 µM duplex linker and ligated to 15 µl isolated DNA that had been treated with of T4 polymerase to repair ends for blunt end ligation to the linkers by using 5 units of T4 ligase (New England Biolabs) and 1.5 µl 10x ligase buffer (New England Biolabs), 1 µl 1 µM Linker, to a final volume of 15 µl. The reaction was then incubated overnight at 4°C. The ligated DNA was then treated with the Qiagen Reaction cleanup kit, and the DNA was eluted in 15 µl of pure water. The eluted DNA was then used for PCR, with 0.25 U of Hotstar Taq (Qiagen), 10 µl of 10x Taq Buffer (Qiagen), 8 µl 25 mM MgCl<sub>2</sub>, 1.6 µl of 10 mM dNTPs (Roche), 1 µl 100 µM linker 2 in a final volume of 100 µl. The PCR conditions were 55°C 2 min, 72°C 5 min, 94°C 10 min, 24 cycles of 94°C 1 min, 55°C 1 min, 72°C 1 min, and then 72°C for 5 min, with a 4°C hold. After amplification, 2 µl of the PCR reaction was cloned using the TOPO® TA cloning® Kit (Invitrogen, Waltham, MA, USA) following the manufacturer's instructions. The resulting colonies were grown in liquid culture, and plasmids were isolated and sequenced as previously described (11).

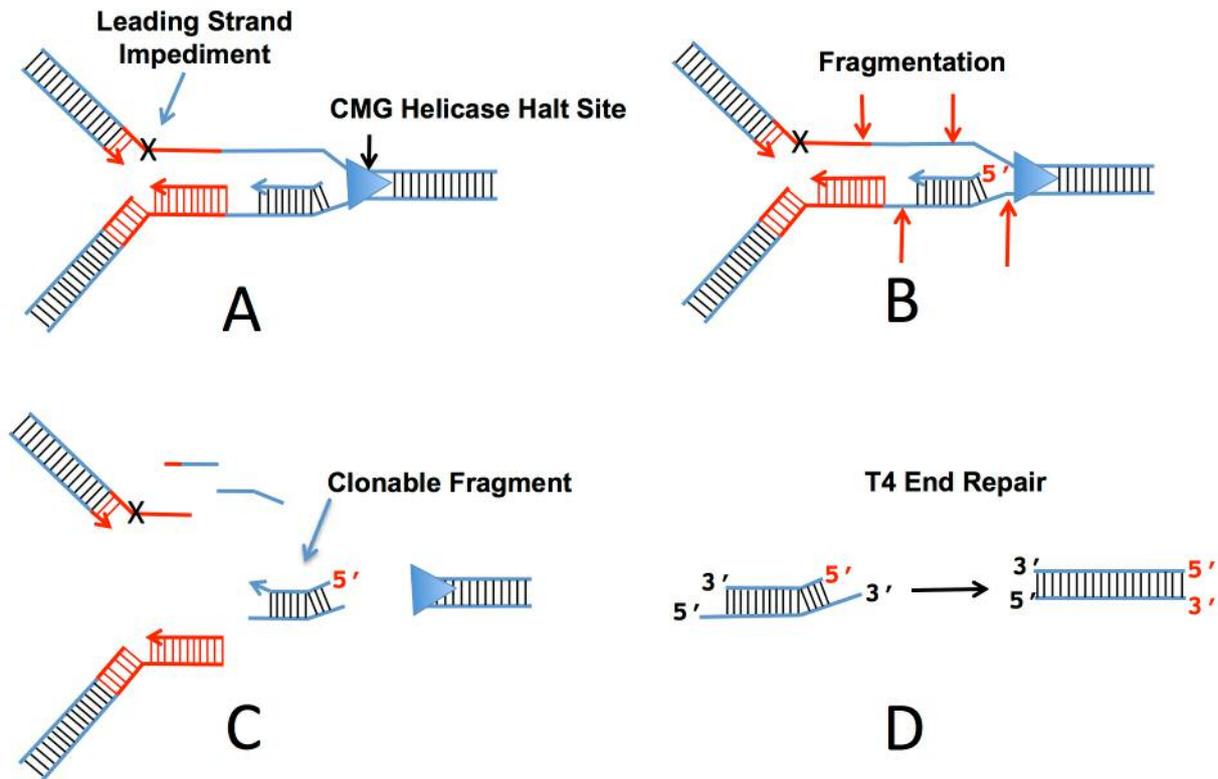


Figure 1. Fork stalling at a replication impediment permits cloning of adjacent DNA. (A) DNA damage, a non-B DNA structure or an R-Loop can halt the leading strand. However, the CMG helicase complex and lagging strand synthesis can continue for a considerable distance (up to 6 kb) downstream from site of the halted leading strand (27, 28). (B) If the unrepaired lesion persists for a time, shear and/or chemically induced cleavage (red arrows) within the lesion can lead to a variety of clonable fragments from the intermediate (C). T4 end repair (D) permits LMPCR amplification and cloning of duplex fragments that identify the region downstream the leading strand halt site. Region with non-B structure potential is shown in red.

**Cleavage frequency.** A scanned profile of the molecular weight distribution of purified PC3 DNA separated by capillary electrophoresis was obtained by using a DNA 7500 microfluidic chip (Agilent Technologies). The number average molecular weight of the DNA preparation was determined from the scan as previously described (30).

**Direct PCR from the RACK7 region (bisulfite-treated PC3 DNA).** Each of the colored regions in Figure 2 were amplified from PC3 cell line DNA after non-denaturing bisulfite treatment (29) to detect possible open DNA structures. In each case, 1  $\mu$ g of native DNA from each cell line was treated with bisulfite using the EZ DNA Methylation Kit (Zymo Research), also as previously described (32), except that the initial sodium hydroxide denaturation step was omitted, and the temperature of the 16 h incubation period was reduced from 55°C to 37°C so as to preserve the native secondary structure of the isolated DNA. Initial exposure of the DNA was to the bisulfite reagent adjusted to pH 5.3, as described in (11). The reaction was cleaned up by use of the Zymo spin columns. The manufacturer's instructions were followed exactly in these experiments. The bisulfite-treated DNA was then eluted in 15  $\mu$ l of pure water, and all of it used for PCR amplification. Pretreatment of DNA from RNase H was carried out in some cases as described in (33).

PCR Primers for the region 3' to the cloned region (Bright Green in Figure 2) were as follows: Forward: 5'-CGTTCCATTGGAACAGCACCTGACTATG-3'; Reverse: 5'-ATTTATGGTTTGGGAGTTATTGGGAAGAC-3'. PCR Primers for the Cloned Region (Yellow in Figure 2) were: Forward 5'-ACTCCTGAGAGTACTGTACAG-3'; and Reverse: 5'-GTCTACCTTCTAATTCTGCCT-3'. PCR primers for the region 5' to the cloned region (Turquoise in Figure 2) were: Forward 5'-GGTTGGCCAGTTGTGTACCTCACAGTGG-3'; and Reverse 5'-GCTCCAGCCTGGGCAACAAAGCAACAGT-3'. Forward and reverse primers were used to form a solution containing 9  $\mu$ M of each primer. PCR amplification master mix contained with 0.25 U of Hotstar Taq (Qiagen), 2.5  $\mu$ l of 10x Taq Buffer (Qiagen), 0.95  $\mu$ l 5X Q solution (Qiagen) 2  $\mu$ l 25 mM MgCl<sub>2</sub>, 0.8  $\mu$ l of 10 mM dNTPs (Roche) and 2.5  $\mu$ l 9  $\mu$ M primers to obtain a final volume of 9  $\mu$ l. Each reaction contained 9  $\mu$ l master mix and 200 ng of target DNA brought to a final volume of 25  $\mu$ l with molecular Biology Water (Sigma). The PCR conditions were 95°C 10 min, 35 cycles of 95°C 1 min, 64.5°C 1 min, 72°C 1 min, for 35 cycles and then 72°C for 5 min, with a 4°C hold.

**Direct sequencing of the RACK7 region (native DNA).** DNA from the 5'-region (Turquoise in Figure 2) of RACK7 was amplified without bisulfite treatment from PC3 DNA specimens as follows: PCR primers for this region were: Forward: 5'-GGTTGGCCAGTTG



Figure 2. RACK7 Sequence in the vicinity of the multiply cloned region. The sequence for RACK7 from Homo sapiens chromosome 20 depicted here is that given in the human genome sequence (GRCh38.p12 Primary Assembly). Native DNA regions examined for cytosine conversion by bisulfite under non-denaturing conditions are highlighted in color. Turquoise: 5' region tested for bisulfite conversion under non-denaturing conditions. Red: Region of intense bisulfite conversion. Yellow: Central region tested for bisulfite conversion under non-denaturing conditions. This region also contains the seven overlapping independently cloned sequences cloned in the screen. The arrows highlighted in gray indicate the common 3' breakpoint present in six of the seven clones. Breakpoints for the seventh clone are highlighted in red. The remaining six arrows indicate 5' breakpoints for the clones with a common 3' end. Bright Green: 3' region tested for bisulfite conversion.

TGTACTTCACAGTGG-3'; and Reverse 5'-GCTCCAGCCTGG GCAACAAAGCAACAGT-3'. PCR reactions contained 12.5 µl 2X KAPA2G Fast Multiplex mix (Kapabiosystems, Wilmington, MA, USA), 0.55 µl 9µM of each primer, 200ng target DNA, made up to a final volume of 25 µl with Molecular Biology Water (Sigma). PCR cycling conditions were 95°C for 3 min, followed by 35 cycles of 95°C for 1 min, 60°C for 30 sec, 72°C for 38 sec followed by a final extension at 72°C for 26 sec.

**Cloning and sequencing of RACK7 region PCR products.** DNA amplicons from 20 µl of each RACK7 region PCR reaction were separated on 2% agarose gels (40 mM Tris-Acetate, 10 mM EDTA pH 7.8) and stained with 4 µg/ml Ethidium Bromide and destained in water. Excised fragments of the expected length (710 bp for the 3' turquoise region in Figure 2, 640 bp for the yellow cloned region in Figure 2, and 430 bp for the bright green 5'-region in Figure 2) were purified using the Qiagen Gel Extraction Kit as described by the manufacturer. Lastly, 2 µl of the purified amplicon was cloned using the TOPO® TA cloning® Kit (Invitrogen, Waltham MA, USA) following the manufacturer's instructions. The resulting colonies were grown in liquid culture, and plasmids were isolated and sequenced as previously described (11).

**DNA sequence alignment software.** Cloned DNA sequences in FASTA format we aligned against the reference sequence for RACK7 from Homo sapiens chromosome 20, GRCh38.p12 Primary Assembly, using two software packages: MAFFT (34) and T-Coffee (35) with similar results showing variable length deletions and TCC→TTC OR TAC mutations. Sequence data sets depicted in Figure 3 were constructed from MView (36) renderings of the sequence alignments obtained with MAFFT.

Circular dichroism and UV melting/annealing analysis of model oligodeoxynucleotides. CD spectra and melting and annealing data were collected as previously described (37). Briefly,

oligodeoxynucleotide sequences (10 µM for CD, 2.5 µM for melting/annealing) were diluted in buffer containing 10 mM sodium cacodylate and either 100 mM or 25 mM potassium chloride at the specified pH, thermally annealed by heating in a heat block at 95°C for 5 min, and slowly cooled to room temperature overnight. CD data were recorded using a Jasco J-810 spectropolarimeter and values were corrected for the molar ellipticity reading at 320 nm. The transitional pH (pHT) of the i-motif was calculated from the inflection point of the fitted ellipticity at 288 nm. All UV melting/annealing experiments were carried out at pH 6.5 (i-motif) or 7.0 (G-quadruplex) and were recorded using a Jasco V-730 spectrophotometer. Melting ( $T_m$ ) and annealing ( $T_a$ ) temperatures were determined using the first-derivative method (37).

**Results**

A total of 96 clones were sequenced and mapped to chromosomal locations in the human genome: 43 from IPS19I and 53 from PC3 (Tables I and II). Motifs with non-B structure potential were observed within 5 kb of the cloned sequence in 60% of the clones from IPS-19I and 74% of the clones from PC3. However, clones within 5 kb of a sequence conforming to the G-quadruplex folding rule (2) were twice as frequent in PC3 compared to IPS19I (Table III). This was consistent with the karyotypes of the two cell lines showing that IPS-19I possessed a normal karyotype (Figure 4) suggesting a normal phenotype, while the hyperdiploid and heavily rearranged karyotype of the p53 deficient PC3 cell line (Figures 5 and 6) indicated that ongoing chromosomal damage might be occurring in this tumor cell line. Moreover, clones containing a chromosomal rearrangement (a fusion between chromosome 20 and Chromosome 7), and clones

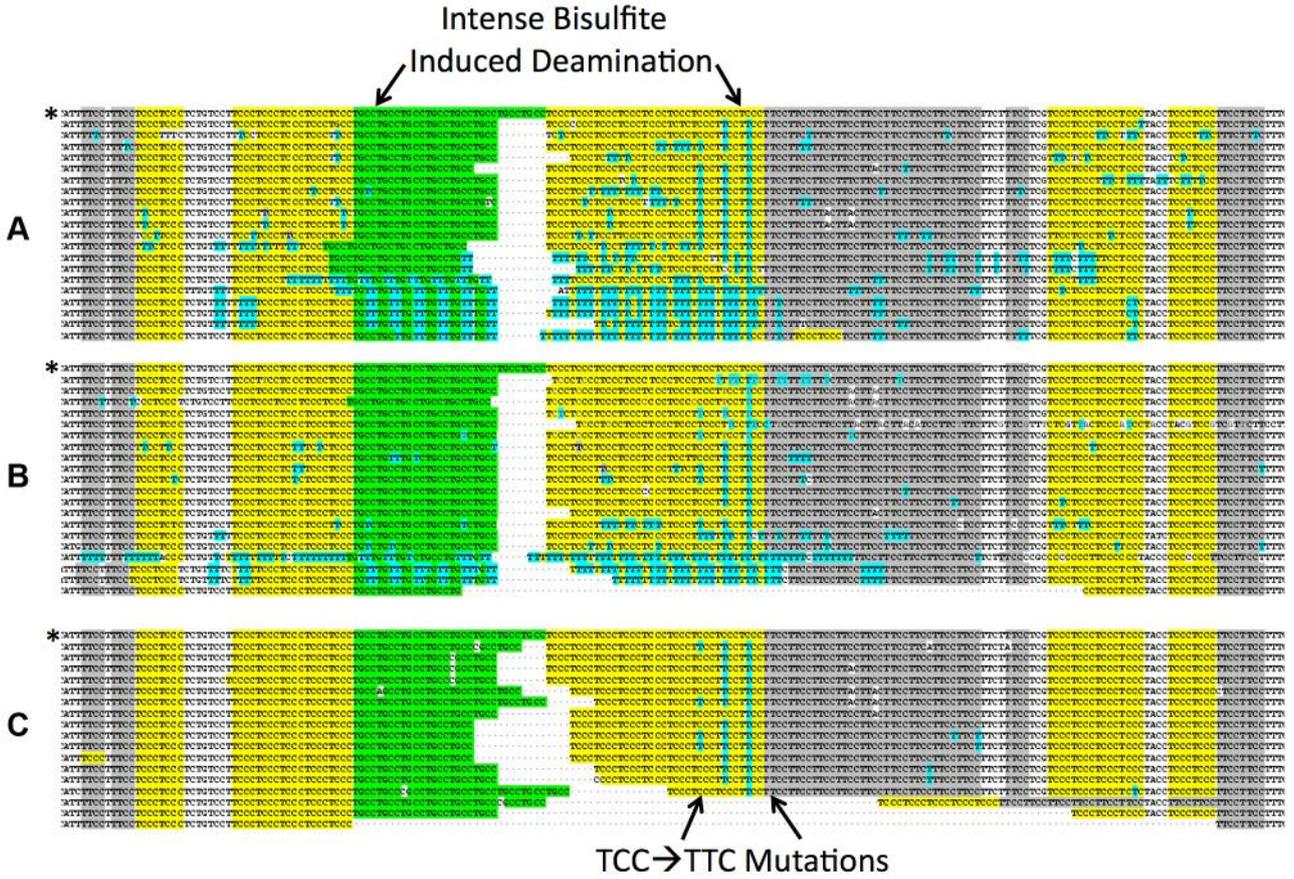


Figure 3. RACK7 sequences obtained from PC3 DNA in the region 5' to the independently cloned sequences. Each panel is colored to highlight tetrameric repeat elements in the concatenated repeat defining the region (Gray: TTCC; Bright Green: CCTG; Yellow: TCCC). Panel A: Native sequences obtained after bisulfite treatment under non-denaturing conditions colored to show bisulfite converted C→T sites in Turquoise. Note that the region of intense bisulfite conversion in this panel lies within a region of GC-skew defined by the concatemer in the genomic sequence (\*) with the general sequence (TCCC)<sub>5</sub>-(CCTG)<sub>8</sub>-CC-(TCCC)<sub>9</sub>-(TTCC)<sub>9</sub>. The highest degree of C→T conversion occurs in the (CCTG)<sub>8</sub>-CC-(TCCC)<sub>9</sub> region. Panel B: Native sequences obtained after RNase H treatment followed by bisulfite treatment under non-denaturing conditions colored to show bisulfite converted C→T sites in Turquoise. Panel C: Native sequences obtained without RNase H or bisulfite treatment. Focused TCC→TTC mutations in the native sequences are colored in Turquoise. (\*) indicates the genomic sequence at top of each panel. Each of the isolated sequences contained a deletion (shown as a gap in the panels) relative to the genomic sequence.

originating from the same chromosomal location were observed only in the PC3 cell line (Table II).

In the PC3 cell line, two clones mapped to the same chromosomal location on chromosome 7, two clones mapped to the same location on chromosome 11 and seven clones mapped to the same chromosomal location in an intron of the RACK7 gene on chromosome 20 (Figure 2). Four of the seven clones were obtained from untreated DNA. Three of the seven clones were obtained from DNA that had been treated with sodium bisulfite to facilitate breakage. We estimated the probability that clones could originate from the same location by random shear as follows. Using the method also described in (30) we determined that spontaneous cleavage frequency for the PC3 DNA preparation to be

1/8800bp from the scanned molecular weight profile for purified PC3 DNA used in the preparation of the PC3 DNA library (not shown).

For a genome of length  $\Lambda$ , it has been shown (38, 39) that the weight fraction ( $F_w$ ) between DNA fragment lengths  $L_1$  and  $L_2$  that is produced by random breakage at a frequency  $f$  is given by:

$$F_w(L) = \frac{\int_0^{L_1} Lf^2 e^{-fL} dL}{\int_0^{\Lambda} Lf^2 e^{-fL} dL}$$

Table I. *HPS-191 cell line.*

Chromosome	Cytpogenetic location	Treatment*	Adjacent motif @ distance**
1	1p31.1	B-C-	No motif
1	1q31.1	B+C+	No motif
1	1q23.3	B+C+	GT <sub>14</sub> @ 3087
2	2p25.3	B+C-	Multi GQ @ (3826,3937,3610)
2	2p25.2	B+C+	GQ @ 1766
2	2q33.1	B+C+	No motif
3	3p26.1	B+C-	Internal GQ & GQ @ 940
3	3p22	B+C+	Long TA <sub>n</sub> @ 714
4	4q12	B+C-	No motif
4	4q28.1	B+C-	GAA cluster @ 2419
4	4q34	B+C-	CA <sub>18</sub> @ 1950
4	4q32.1	B+C+	GQ @ 981
5	5q31.1	B+C-	GQ @ 4819
6	6q21	B-C-	No motif
6	6p21.1	B+C-	No motif
6	6p25.1	B+C+	Two internal GQs & flanking GQs @ (1648,897)
7	7q21.11	B+C-	TG cluster @ 2374
7	7q31.32	B+C-	No motif
8	8q11.22	B-C-	GQ @ 548 (TAA) <sub>14</sub> @ 3446
8	8p21.1	B-C-	GQ @ 3654
8	8q23.1	B-C-	No motif
9	9q33	B+C+	Pyr <sub>29</sub> @ 909 and Pyr <sub>49</sub> @ 1217
10	10q26.3	B+C+	Multi GQ @ (188, 1089,1219)
10	10q22.3	B+C+	No motif
11	11q14.1	B+C-	TG cluster @ 856
11	11q21	B+C+	GQ @ 319
13	13q31.1	B+C+	No motif
14	14q21.1	B-C-	No motif
14	14q31.1	B+C-	TG cluster @ 1500
14	14q22.3	B+C+	GA cluster @ 973
14	14q23.3	B+C+	No motif
15	15q21	B+C+	No motif
16	16q11.2	B-C-	Multi GQ @ (3449, 3759)
16	16p13.2	B+C+	TG cluster @ 2412
17	17q22	B+C-	No motif
18	18p11.31	B+C-	T <sub>28</sub> @ 2508
18	18q11.1	B+C-	TG cluster @ 2194
20	20q11.21	B+C-	Internal GQ
20	20p12.3	B+C+	GQ @ 2791
21	21p12	B-C-	No motif
22	22q13.31	B+C-	Multi GQ @ (2200, 2583)
X	Xp11.2	B-C-	No motif
X	Xq26.3	B+C-	No motif

\*B+: DNA treated with bisulfite under native conditions prior to isolation. B-: DNA not treated with bisulfite under native conditions prior to isolation. C+: DNA treated with CIP under native conditions prior to isolation. C-: DNA not treated with CIP under native conditions prior to isolation. \*\*GQ: DNA sequence conforming to the G-Quadruplex folding rule (2).

Table II. *PC3 cell line.*

Chromosome	Cytpogenetic location	Treatment*	Adjacent motif @ distance**
1	1p22.2	B-C-	GQ @ 1645bp
1	1p31.1	B-C-	GQ @ 4209
1	1p22.1	B+C-	Internal GQ
1	1q32.3	B+C-	GQ @ 3923
2	2q33.3	B-C-	GQ @ 54
2	2q22.2	B-C-	T <sub>25</sub> @ 1716, GT <sub>11</sub> @ 2530
2	2q12.3	B-C+	Multiple GA triplets @ 970
2	2p16.3	B+C-	No motifs
3	3p12.3	B-C-	GQ @ 3744
3	3q27.3	B+C+	Multi GQ @ 2952
3	3p21.31	B+C+	GQ @ 3448
3	3q27.1	B+C+	No motifs
5	5q21.3	B-C+	T <sub>21</sub> @ 2528, CA <sub>19</sub> @ 232, (GGAA) <sub>15</sub> @ 2596
6	6p22.2	B-C-	TG <sub>n</sub> cluster @ 3010
7	7q34	B-C-	No motifs
7	7p22-p11.2	B+C-	No motifs
7	7q22.1	B+C+	TG <sub>n</sub> , TA <sub>n</sub> , GQ @ 349
7	7q22.1	B+C+	TG <sub>n</sub> , TA <sub>n</sub> , GQ @ 350
7	7p22	B+C+	GQ @ 3248
8	8q21.11	B-C-	No motifs
8	8q24.1	B+C-	GQ @ 2985
8	8q22.2	B+C+	Multi GQ @ 2217
10	10q22.2	B-C-	No motifs
10	10q22.1	B+C-	GQ @ 5
10	10p11.23	B+C-	GQ @ 192
11	11q21	B-C-	GQ @ 3373
11	11p15.4	B-C+	Multi GQ @ 3742
11	11p15.4	B+C+	Multi GQ @ 3742
11	11p13	B-C+	No motifs
11	11p14.1	B+C-	GQ @ 1775
11	11q14.1	B+C+	GQ @ 1822
11	11p15.4	B+C+	Multi GQ @ 3742
12	12q12	B-C+	No motifs
12	12q13.13	B+C-	GQ @ 492
12	12q12	B+C-	No motifs
13	13q21.2	B-C-	GA <sub>3</sub> cluster @ 1146
13	13q12.11	B-C+	No motifs
13	13q22	B+C-	GQ @ 3923
17	17q25.3	B-C-	GQ @ 603, GQ @ 1582 Flank
17	17q21.3	B-C-	GQ @ 1181, GQ @ 666 Flank
18	18q12.2	B+C-	GQ @ 4217
20	20p12.1	B-C-	No motifs
20	20q11.23	B-C-	Multi GQ @ 360
20	20q11.23	B-C-	Multi GQ @ 360
20	20q11.23	B-C-	Multi GQ @ 360
20	20p11.23	B-C-	Multi GQ @ 360
20	20q11.23	B-C+	Multi GQ @ 360
20	20q11.23	B+C-	Multi GQ @ 360
20	20q11.23	B+C+	Multi GQ @ 360
20	20q11.23	B+C+	Multi GQ @ 360
X	Xp22.33	B-C-	Multiple Long T <sub>n</sub> runs
X	Xq26.3	B+C-	GQ @ (2767, 4206)
X	Xp11.23	B+C-	TG <sub>n</sub> cluster @ 2767
Y	Yp11.2	B-C-	No motifs

\*B+: DNA treated with bisulfite under native conditions prior to isolation. B-: DNA not treated with bisulfite under native conditions prior to isolation. C+: DNA treated with CIP under native conditions prior to isolation. C-: DNA not treated with CIP under native conditions prior to isolation. \*\*GQ: DNA sequence conforming to the G-Quadruplex folding rule (2).

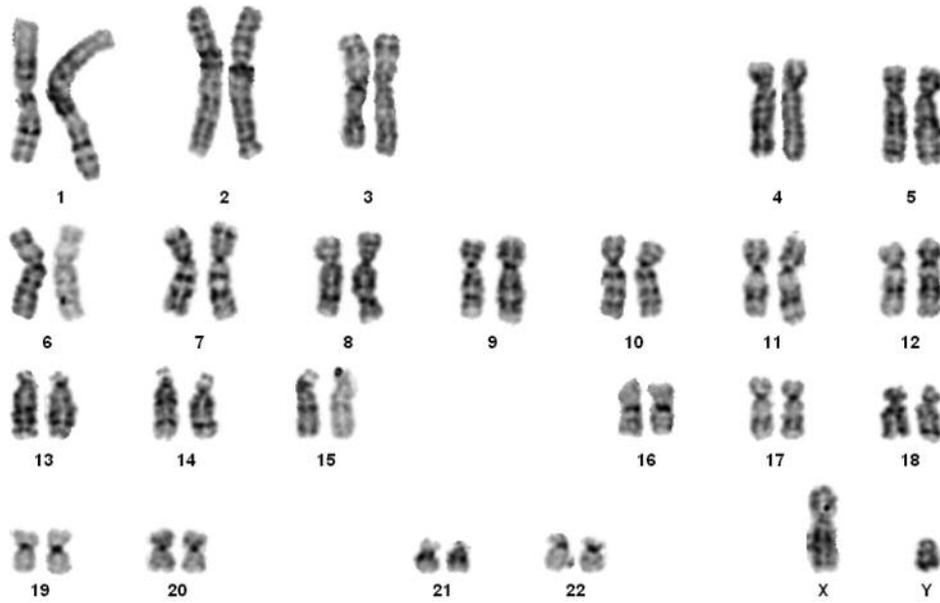


Figure 4. G-Banding karyotype of the HPS-191 cell line. The karyotype of the low passage human prostate stromal cell line, HPS-191, was normal, with a mean chromosome number of 46 [2n]. In eighteen mitotic cells analyzed, including one tetraploid cell, there was no significant numerical chromosomal abnormality and no structural aberration of any chromosome detectable within the limits of cytogenetic resolution.

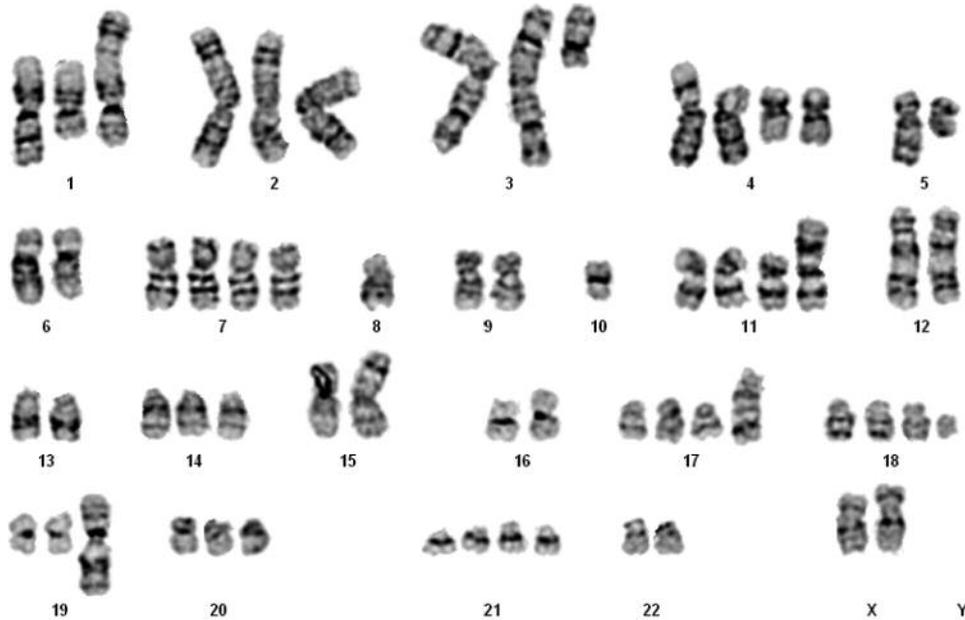


Figure 5. G-Banding karyotype of the PC3 Cell line. The G-banding karyotype of the high passage established prostate cancer cell line, PC3, was abnormal. All mitotic cells analyzed by conventional G-banding were near triploid and clonally abnormal.

Since the procedure used for pCR™ 2.1-TOPO® cloning is designed to clone fragments less than 1,000 bp, the inserts we observed were all in this range. The weight fraction predicted by random breakage in this range for a cleavage

frequency  $f=1/8800$  is about 0.006 and could represent almost 1% of the genome. Thus, the twenty independent isolates obtained without bisulfite or CIP treatment could be produced by random breakage. At the measured cleavage

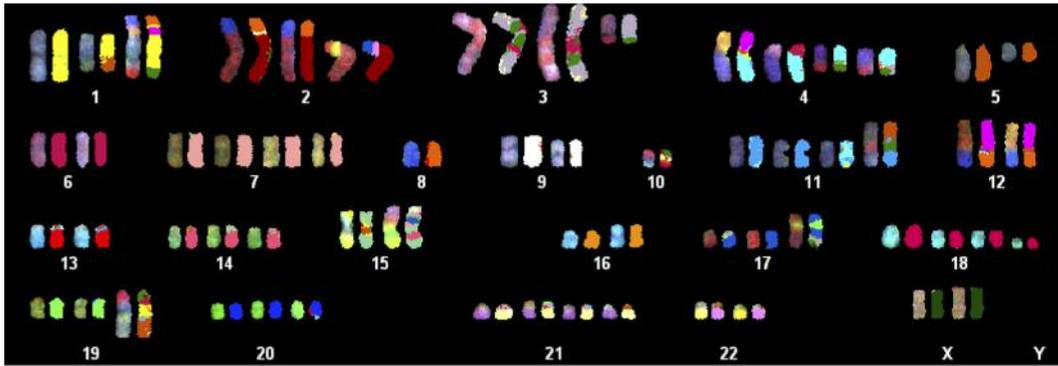


Figure 6. P53 SKY spectral karyotype. The analysis of the complex karyotype of the PC3 cell line was corroborated with SKY spectral karyotyping. The complete analysis of the hyperdiploid stemline showed that it contained recurrent structural rearrangements involving chromosomes 1, 2, 3, 4, 5, 6, 8, 10, 11, 12, 15, 16, 17, 18, and 19. Numerical gains of chromosomes X, 1, 2, 3, 4, 7, 11, 14, 16, 17, 18, 19, 20, and 21 were present. Of note, the only chromosome homologues that did not have structural rearrangements were chromosomes 7, 9, 13, 14, 20, 21, 22, and X. The Y chromosome was apparently lost in this cell line. The mean chromosome number was: 62, with the following ISCN (International System for Human Cytogenetic Nomenclature) description, based on both conventional cytogenetic analysis and SKY spectral karyotyping :62,<2n>, X,+X,-Y,+der(1)t(8;12;1;10), der(2)t(8;2)x2, +der(2)t(17;22;2), der(3)t(3;6;3;6;3)x2, +der(3;10)(p10;p10)x2, der(4;12)(q10;q10), der(4;6)(q10;p10), +der(4)t(4;10)(p10;q10)x2, +i(5)(p10),+7, +7, -8,+11,+der(11)t(5;10;11), der(12)t(12;8)(q24;q13)x2, +14,der(15)t(8;15)(q22;p11.2), der(15)t(5,15)(q13;p11.2), +16,der(16)t(10;1;16), +der(17)t(3;17;15;17),+18,+i(18)(p10), +der(19)t(5;19;5),+20,+20,+21,+21[5].

Table III. Summary of cloning results.

Cell Line*	Number of clones obtained	Fraction conforming to the G-Quadruplex folding rule within 5 KB <sup>§</sup>	Fraction with non-B motif within 5KB	Average distance to any motif	Regions cloned more than once
HPS191 (primary)	43	14/43=33%	26/43=60%	1778bp	0
PC3 (tumor)	53	33/53=62%	39/53=74%	1777bp	3

\*See Tables I and II for clone details. <sup>§</sup>The difference between the values for the two cell lines is statistically significant in a two-sided *t*-test at the *p*<0.01 level.

frequency, a hyper diploid PC3 genome (Figure 5) comprising about  $7.5 \times 10^9$  bp of DNA would have at least  $2 \times 10^{11}$  fragments/ $\mu$ g DNA. Of these, about  $1.7 \times 10^9$  bp representing a random sampling of the genome would be in the size range clonable by pCR™ 2.1-TOPO®. Consequently, each of the single isolates could be the result of random cleavage unless biologically frangible sites associated with replication stress (e.g. non-B sequence potential) are more frequent.

On the other hand, sequences obtained more than once cannot be the result of random cleavage. This is most clearly seen for the case of the sequences four sequences obtained from untreated DNA that appear to have been produced by random shear from RACK7 gene. In this case four independently isolated overlapping fragments originated from the same 656 bp region of RACK7. The weight fraction of the PC3 genome that falls between 0 bp and 656 bp for

randomly cleaved DNA is expected to be 0.0026. Since the cloned sequences all fall within a 656 bp region of RACK7, if we assume that there are only 6 copies of the region present in the PC3 hyper-diploid genome based on Giemsa (Figure 5) and SKY Spectral analysis (Figure 6), then the random probability of cloning fragments ranging in size from 0 bp to 656 bp from that 656 bp region of the RACK7 gene sequence with 100% cloning efficiency is given by:

$$p = 6 \times 656 \text{ bp} / 7.504 \times 10^9 \text{ bp} \times 0.0026 / 0.006 = 2.2 \times 10^{-7}$$

The probability of cloning it *m* times in *n* trials is given by:

$$P = \binom{n}{m} p^m (1-p)^{n-m}$$

For 4 clones in 20 trials  $p = 1.2 \times 10^{-23}$

The observed cloning frequency of  $4/20 = 0.25$  clearly rules this out.

Given that the probability that random breakage during DNA isolation would produce multiple clones from the same site is vanishingly small (see Materials and Methods) it is likely that this region represents a site of recurring DNA damage that renders it particularly frangible, and thus subject to repeated isolation by the cloning scheme used here. Further, six of the seven clones had a common breakpoint its 3' end and a randomly placed break point at its 5' end (Figure 2) suggesting that a recurring biologically induced gap occurs at the 3' site and that the clonable duplex fragments were produced by hydrodynamic shear during DNA isolation or by bisulfite induced breakage at the gap in native DNA prior DNA isolation.

Regions that persist as non-B structures in native DNA have been shown to be accessible to chemical modification (6, 40, 41). When we used published methodology (29) for the chemical deamination of cytosine residues by sodium bisulfite under non-denaturing conditions, we found that deamination only occurred in the region upstream of the cloned region (Turquoise in Figure 2) and not in the cloned region itself (Yellow in Figure 2) or the region downstream (Green in Figure 2) of the cloned region (data not shown). The sequence in this intronic region of RACK7 was not rearranged in the prostate cancer cell line, however the bisulfite accessible region contained deletions of different lengths (Figure 3). Variation among individual isolates of the region suggests that the deletion process may be ongoing and cumulative. Further deletions did not block bisulfite access since deletions were observed even in sequences containing heavily deaminated regions after treatment with bisulfite under non-denaturing conditions (Figure 3A). We also noted TCC→TTC or TAC mutations characteristic of the AID and APOBEC family of single-strand specific DNA cytosine deaminases downstream from the deletions in untreated DNA (Figure 3C).

Studies of *dog-1* induced deletions in *C. elegans* (42) found that deletions are generally sharply confined to the 3' end of the G-rich strand conforming to the general G-quadruplex folding rule (2) ( $G_{3+N_{1-7}}G_{3+N_{1-7}}G_{3+N_{1-7}}G_{3+N_{1-7}}$ ). The data in Figure 3 display the C-rich strand from this region of the human RACK7 gene, since the extreme GC-skew in the region precludes extensive bisulfite mediated C→T deamination on the G-rich strand. Deletions tended to be more commonly confined to the 5' end of the (TCCC)<sub>9</sub> repeat of the C-rich strand (Figure 3). On the complementary G-rich strand the deletions are confined to what would be the 3' end of the complementary (GGGA)<sub>9</sub> repeat that conforms to the general G-quadruplex folding rule (2). Thus, the positioning of the deletions is similar to those observed in *dog-1* deficient *C. elegans*.

Our cloning strategy was based the possibility that a variety of non-B structures could uncouple lagging strand replication at a leading strand impediment (Figure 1). Although non-B DNA structures can produce this effect

hybrid structures like the R-Loop are also candidates. DRIP methods report a weak R-loop signal in this region in epithelial cells, but not in fibroblast or leukemia cell lines (43). RDIP-Seq reports weak R-loop signal in both epithelial and fibroblast cell lines (44), and surprisingly DRIPc reports weak R-Loop signal only on the minus strand in a pluripotent cell line (45). Even so, an R-Loop in RACK7 would have to violate the general rule that the nascent RNA strand is the G-rich strand in the region of GC-Skew (10, 46) because transcription of RACK7 produces a nascent C-rich strand RNA. Nevertheless, we tested the region for bisulfite sensitivity after pretreating the DNA with RNase H. The intense bisulfite induced deamination seen on the C-rich strand under non-denaturing (Figure 3A) is only moderately reduced by pretreatment of the DNA with RNase H (Figure 3B), suggesting that the open structure in this region is not completely abolished by treatment with RNase H as is generally seen with R-Loops. The most intense deamination occurs within the (CCTG)<sub>8</sub>-CC-(TCCC)<sub>9</sub> region of the repetitive element (Figure 3A and 3B) with or without RNase H pretreatment. Assuming that each sequence is an independent isolate of a representative pool, analysis of the sequence sets produced by bisulfite mediated deamination with or without RNase H pretreatment shows that the difference between the two sets is not statistically significant at the  $p < 0.05$  level in either one sided or two-sided t-tests.

In order to study the nature of the structures that can form in the region of bisulfite accessibility, we performed biophysical analysis on model oligodeoxynucleotides corresponding each strand from this region using circular dichroism (CD) and UV thermal melting/annealing analysis (Figure 7). The CD spectrum of the model oligodeoxynucleotide from the G-rich strand shows a strong positive signal at 263 nm, indicative of a parallel G-quadruplex structure (Figure 7B). Consistent with this observation, the UV melting/annealing experiments (Figure 7A) revealed a highly stable structure in physiological-like conditions, such that it would not melt under the conditions of the experiment. Consequently, we had to reduce the potassium cation concentration from 100 mM, similar to physiological conditions, to 25 mM. Only then were we able to observe a melting transition ( $T_m = 81^\circ\text{C}$ ). This suggests the G-quadruplex is highly stable in conditions that mimic physiological pH and cation concentration. This data also clearly support the formation of a G-quadruplex associated with deletions at its 3' end under physiological conditions. The stability of the structure indicated by its high  $T_m$  is consistent with its formation even within sequences that have undergone short deletion near its 3' end, as evidenced by the bisulfite accessibility of the region in those the sequences carrying deletions depicted in Figure 3.

We also studied a model oligodeoxynucleotide from the C-rich strand for i-motif structures most commonly detected in C-rich strand sequences. Such structures involve C:C<sup>+</sup> base pairs that are frequently most stable at acidic pHs.

Consequently, we expected it to carry a lower potential for non-B structure formation under physiological conditions than is expected for the complementary G-rich strand. Analysis of a 37-mer from this region indicated this could form into an i-motif structure, indicated by a strong positive signal at 288 nm. Plotting the ellipticity of this peak at 288 nm against pH provided a transitional pH of 6.9, indicating that this sequence forms i-motif at neutral pH (Figure 7D). This structure also appears to be stable under near-physiological conditions (Figure 7C) with melting transitions at 37.6 and 41.6°C. Clearly both G-quadruplex and i-motif structures can potentially form in the same region, although previous biophysical experiments indicate that the formation of each structure is mutually exclusive (47) in a constrained duplex, suggesting that it may be that only one of the two structures would be present at this specific site at a given time.

## Discussion

The model used in the cloning design (Figure 1) is based on evidence demonstrating that CMG uncoupling permits lagging strand synthesis to continue past a replication impediment for as much as 6 kb beyond the leading strand halt site (27, 28). The model is supported by the cloning results reported herein. Those results show that a sequence conforming to the G-quadruplex folding rule (2) within a region of GC-skew was present within 5 kb of the cloned sequence in 33% of the sequences cloned from the normal cell line and 62% of the sequences from the tumor cell line. Several studies (13-15) suggest that DNA Quadruplex formation can provide an impediment to DNA replication that can lead to DNA damage in cells compromised in one or more of the numerous systems that aid replication and repair at these structure prone sequences. The two-fold increase in G-quadruplex linked clones in the tumor cell line compared to the normal cell line could be associated with the absence of functional p53 in the tumor cell line (48). Functional p53 is important in binding at least one quadruplex resolving helicase (BLM) (19) and perhaps another (WRN) (18) at sites of DNA damage and repair (49, 50), where they appear to cooperate with FANCD1 to resolve G-quadruplex replication impediments (51). However, both BLM and WRN have also been shown to resolve RNA:DNA hybrids *in vitro* (21, 22) so the p53 lesion in the PC3 cell line does not exclude the possibility that a fraction of the sites we have cloned reside near replication impediments caused by R-Loops.

In addition to Quadruplex forming structures, other non-B structures that are not characterized by GC-skew are capable of causing DNA damage and influencing repair (52). Consistent with this possibility, our data showed that 60% of clones were within 5 kb of a sequence capable of non-B

structure formation in the normal cell line while 74% of the clones from the tumor cell line were within 5 kb of a non-B structure forming sequence. Of particular interest are the representatives of the multiply cloned sequences from the tumor cell line. Only the repair compromised tumor cell line yielded multiple clones from the same region, suggesting that these sites represent recurring replication impediments in the tumor cell line. Consistent with this possibility seven overlapping clones from the same region of RACK7 were obtained from the tumor cell line. Furthermore, six of those seven clones carried an identical breakpoint at one end. Since this breakpoint (gray arrows in Figure 2) is about 1 kb 3' to the site of the open structures detected with bisulfite treatment of native DNA (red region in Figure 2) it is also consistent with our cloning model where the common breakpoint would represent a site at or very near the uncoupled CMG halt site with the leading strand replication impediment at the site of the non-B structure.

The involvement of quadruplex structures at the putative replication impediment in RACK7 is suggested by the data in Figures 2 and 3, and the biophysical studies on the representative oligodeoxynucleotides from the region (Figure 7). First, nearly all sequences recovered from this region contain a deletion in the region of bisulfite accessibility in native DNA, and neither the deletions nor the RNase H treatment fully block the formation of the open structure detected by bisulfite treatment of native DNA. Those sequences recovered from bisulfite treated DNA (with or without RNase H pretreatment) for which bisulfite mediated deamination was detected retain a nearly full copy of the (GGGA)<sub>9</sub>/(TCCC)<sub>9</sub> duplex capable of G-quadruplex and i-motif quadruplex formation.

In many cases of R-Loop formation the shortening of the duplex region caused by the RNA:DNA hybrid's A-Form structure is thought to bring regions of the looped out G-rich strand into juxtaposition where G-quadruplex formation can occur (53). Given the stability of the G-quadruplex it may persist after an R-Loop or transcription process has been resolved. In the case of the structure at RACK7 under study here, if the RNase H results are interpreted to mean that non-canonical R-Loop forms with between the nascent C-Rich strand RNA transcript and the G-rich strand DNA generating a C-rich loop, then a quadruplex in the form the i-motif structure that is stable at neutral pH and physiological temperature at the (TCCC)<sub>9</sub> sequence in the loop would again be involved. RNase H treatment could remove the hybridized RNA while permitting the i-motif to persist in sequences with deletions that do not remove the (TCCC)<sub>9</sub> region. Alternatively the i-motif structure in the loop might slow RNase H action. Here again the data are consistent with our cloning model where the putative i-motif and an R-Loop provide the generic replication impediment depicted Figure 1.

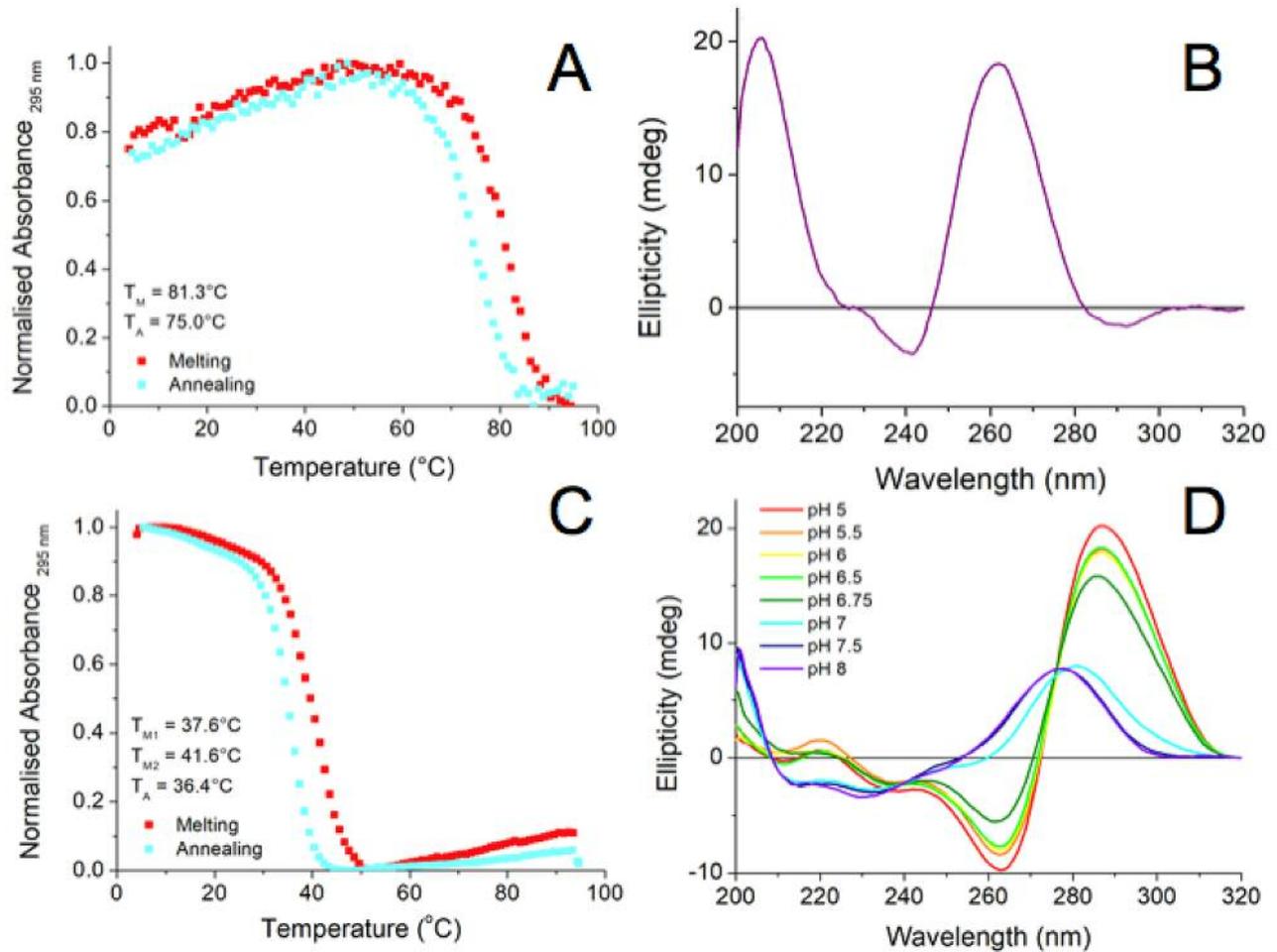


Figure 7. UV and circular dichroism analysis of model oligodeoxynucleotides. 36-mer  $(GGGA)_9$  corresponding to the region with G-quadruplex forming potential in the G-rich strand of RACK7 (top panels) and model 37-mer  $(TCCC)_9T$  corresponding to the i-motif region in the complementary C-rich strand (bottom panels). (A) UV melting/annealing experiments with  $2.5 \mu M (GGGA)_9$  in 10 mM sodium cacodylate with 25 mM KCl at pH 7.0. (B) The CD spectrum of  $10 \mu M (GGGA)_9$  in 10 mM sodium cacodylate with 100 mM KCl at pH 7.0. (C). UV melting/annealing experiments with  $2.5 \mu M (TCCC)_9T$  in 10 mM sodium cacodylate with 100 mM KCl at pH 6.5. (D) CD spectra of  $10 \mu M (TCCC)_9T$  in 10 mM sodium cacodylate with 100 mM KCl at pH 5.0, 5.5, 6.0, 6.5, 6.75, 7.0, 7.5 and 8.0. Inset is a plot of ellipticity at 288 nm versus pH.

Additional support for the single stranded nature of the C-rich strand comes from the detection of TCC $\rightarrow$ TTC or TAC transitions downstream from the deletions. Transition mutations of this type are characteristic of the APOBEC family (54), and it is reasonable to suspect that AID, and APOBEC family of deaminases may be attacking single-stranded DNA in the cells that contain these extensively remodeled sequences. Moreover, these enzymes are known to prefer low pH (55) and single-stranded DNA. The observation is also consistent with the preference of APOBEC for single-stranded DNA at replication forks and DSBs (56), and the preference of the AID DNA deaminase for single-stranded DNA formed near G-quadruplex model oligodeoxynucleotides (57). Clearly, deletion of the quadruplex forming sequences

would preclude their formation, but it is also important to note that dC $\rightarrow$ dU mutagenesis would have a similar effect. For example, the i-motif is destabilized by the introduction of a damaged base (37).

Finally, in nearly all aspects, the data from the human RACK 7 region mirrors the data obtained from *dog-1* deficient *C. elegans* (42, 58). Their data rules out XPF-1 mediated excision repair or MUS-81 linked homologous recombination repair in favor of a bypass mechanism (28, 59) or a replication fork merging mechanism that would generate a persistent gap on the leading strand opposite putative G-quadruplex (42, 58). Subsequent replication of the region would generate double strand breaks that they showed required DNA polymerase theta (25) end joining repair to

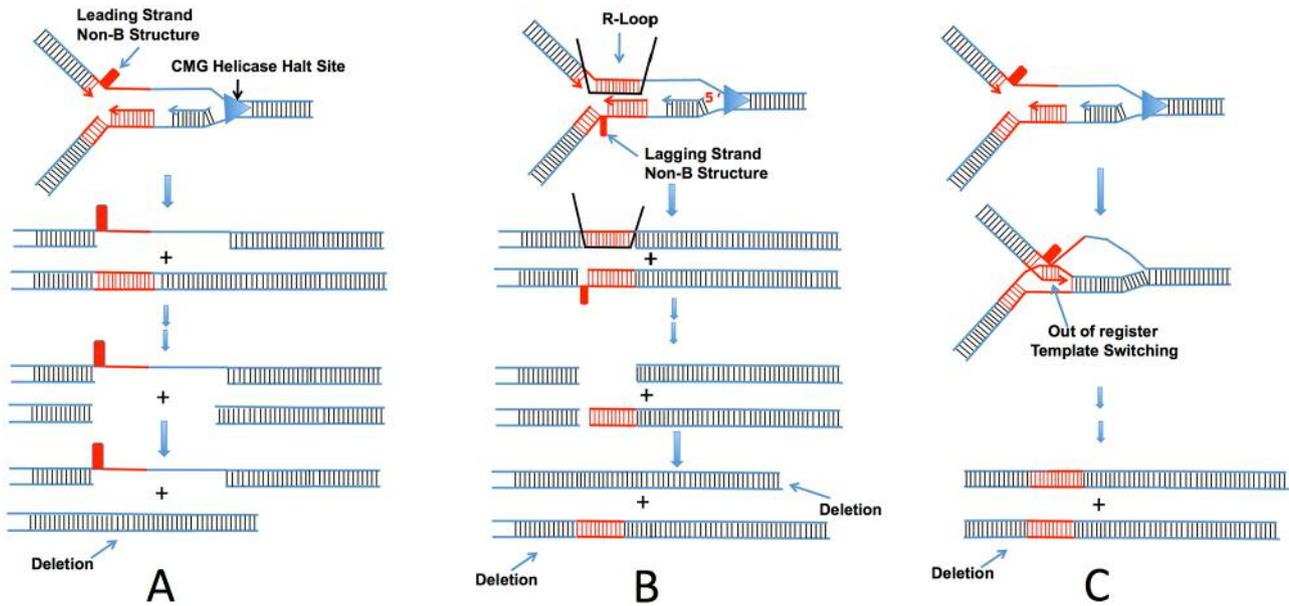


Figure 8. DNA synthesis at non-B DNA replication impediments. (A) Bypass synthesis or merging with a distal replication fork at a persistent G-quadruplex structure generates a persistent gapped structure that on further replication generates double strand breaks repaired by end-joining that result in deletions (26). (B) Bypass at a persistent R-Loop involving an i-motif on the single strand produces two types of persistent gaps if G-Quadruplex or i-motif structures form on the looped out strand. Further replication generates two types of deletion. (C) Out of register template switching restarts the replication fork and also generates a deletion after further replication.

create the observed deletions (26). Although the enzymatic properties of DOG-1 do not appear to have been studied in detail, human FNACJ is known to unwind RNA:DNA hybrids (24). Moreover, R-Loops have been observed to impair replication in *C. elegans* (60). Hence a very similar mechanism in which *dog-1* mutants fail to resolve R-Loop replication impediments is not ruled out by the data on deletions in regions conforming to the G-quadruplex folding rule in *C. elegans*. Models for the genesis of deletions at RACK7 include bypass as well as template switching mechanisms that can generate the deletions associated with replication impediments at sites of non-B structure formation (Figure 8).

**Comparison with G4-seq.** G4-seq (61, 62), provides an effective way of globally mapping G-quadruplex sequence potential, by using high-throughput next generation sequencing to detect sites that can adopt G-quadruplex structures when single-stranded DNA is exposed to K<sup>+</sup> and G-quadruplex stabilizing ligands during sequencing *in vitro*. On the other hand, G4-seq does not offer a means for determining which of the sequences identified *in vitro* actually play a role *in vivo*. In this regard it is important to compare the number of sequences identified by G4-seq with the immunohistochemistry results demonstrating the presence of

G-quadruplex in the nucleus of mammalian cells (63). In those experiments the highest number of fluorescent foci in the nucleus was seen during S-phase, consistent with replication dependent formation of G-quadruplex (63). Even so, only about 35 foci/nucleus were detected at any one time in S-phase. G4-seq returns more than 716,000 distinct genomic sites capable of G-quadruplex formation (61). Consequently, even if the fluorescent foci represent replication factories (64) containing several hundred growing replication forks, less than 2% of the sequences identified with G4-seq would be present at any one time during replication and most of these would be resolved properly by G-quadruplex suppression systems.

The approach we describe herein offers a much-needed method for supplementing G4-seq data by directly detecting those quadruplex forming sequences that have blocked DNA polymerase progression *in vivo* in a given cell type. Moreover, while G4-seq is focused on G-quadruplex, and our results support the idea that G-quadruplex formation is a major contributor to replication stress, our method also has the potential to identify other non-B structure forming sequences that induce replication stress linked helicase uncoupling.

Finally, G4-seq requires access to next-generation sequencing, which may not be available to many labs at reasonable cost, while the method we describe can be performed with simple low-cost cloning and sequencing tools.

## Conclusion

All of the data presented above are consistent with the cloning model depicted in Figure 1. Moreover, elevated frequency of cloned sequences adjacent to sites of G-quadruplex forming potential in the repair compromised tumor cell line supports the idea that endogenous G-quadruplex and i-motif formation is an important source of genetic instability in this cell line. Further, the detailed analysis of the site within RACK7 from the tumor cell line strongly suggests that deletions and mutations are linked to recurrent quadruplex DNA formation at that site. This suggests that the cloning procedure we describe permits the *in vivo* identification of sequences that often produce DNA polymerase-helicase uncoupling during replication where the intercession of damage prone repair processes is required. As such, the method provides a valuable approach to identifying those sequences with G-quadruplex potential that actually produce replication impediments *in vivo*.

## Conflicts of Interest

None to be declared.

## Authors' Contributions

C. A., J. C., V. B., J. M., M. M., E. W. and M. A. conducted experiments, F.P. designed experiments, Z.A. E. W. designed and conducted experiments, S. S. S. designed and conducted experiments and wrote the paper.

## Acknowledgements

This work was supported by: The Biotechnology and Biological Sciences Research Council (BB/L02229X/1) to Z.A.E.W, by a grant 5R01-CA102521 to S.S.S. from the U.S. National Cancer Institute of the National Institutes of Health, and by the Ensign Foundation. E.F.W was supported by a Wellcome Trust grant (204515/Z/16/Z). Research reported in this publication also included work performed in the Integrative Genomics and Bioinformatics Core and the Cytogenetics Core supported by the National Cancer Institute of the National Institutes of Health under award number P30CA033572. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

- Ginno PA, Lott PL, Christensen HC, Korf I and Chedin F: R-loop formation is a distinctive characteristic of unmethylated human cpg island promoters. *Mol Cell* 45(6): 814-825, 2012. PMID: 22387027. DOI: 10.1016/j.molcel.2012.01.017
- Huppert JL and Balasubramanian S: Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* 33(9): 2908-2916, 2005. PMID: 15914667. DOI: 10.1093/nar/gki609
- Huppert JL and Balasubramanian S: G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* 35(2): 406-413, 2007. PMID: 17169996. DOI: 10.1093/nar/gkl1057
- Wang G and Vasquez KM: Effects of replication and transcription on DNA structure-related genetic instability. *Genes (Basel)* 8(1), 2017. PMID: 28067787. DOI: 10.3390/genes8010017
- Day HA, Pavlou P and Waller ZA: I-motif DNA: Structure, stability and targeting with ligands. *Bioorg Med Chem* 22(16): 4407-4418, 2014. PMID: 24957878. DOI: 10.1016/j.bmc.2014.05.047
- Siddiqui-Jain A, Grand CL, Bearss DJ and Hurley LH: Direct evidence for a g-quadruplex in a promoter region and its targeting with a small molecule to repress c-myc transcription. *Proc Natl Acad Sci USA* 99(18): 11593-11598, 2002. PMID: 12195017. DOI: 0.1073/pnas.182256799
- Del Mundo IMA, Zewail-Foote M, Kerwin SM and Vasquez KM: Alternative DNA structure formation in the mutagenic human c-myc promoter. *Nucleic Acids Res* 45(8): 4929-4943, 2017. PMID: 28334873. DOI: 10.1093/nar/gkx100
- Wright EP, Huppert JL and Waller ZAE: Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res* 45(6): 2951-2959, 2017. PMID: 28180276. DOI: 10.1093/nar/gkx090
- Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A and Cimprich KA: Transcription-coupled nucleotide excision repair factors promote r-loop-induced genome instability. *Mol Cell* 56(6): 777-785, 2014. PMID: 25435140. DOI: 10.1016/j.molcel.2014.10.020
- Sollier J and Cimprich KA: Breaking bad: R-loops and genome integrity. *Trends Cell Biol* 25(9): 514-522, 2015. PMID: 26045257. DOI: 10.1016/j.tcb.2015.05.003
- Clark J and Smith SS: Secondary structure at a hot spot for DNA methylation in DNA from human breast cancers. *Cancer Genomics Proteomics* 5(5): 241-251, 2008. PMID: 19129555.
- Guilbaud G, Murat P, Recolin B, Campbell BC, Maiter A, Sale JE and Balasubramanian S: Local epigenetic reprogramming induced by g-quadruplex ligands. *Nat Chem* 9(11): 1110-1117, 2017. PMID: 29064488. DOI: 10.1038/nchem.2828
- Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, Teulade-Fichou MP, Foiani M and Nicolas A: G-quadruplex-induced instability during leading-strand replication. *EMBO J* 30(19): 4033-4046, 2011. PMID: 21873979. DOI: 10.1038/emboj.2011.316
- Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E and Tijsterman M: Mutagenic capacity of endogenous g4 DNA underlies genome instability in fancj-defective c. Elegans. *Curr Biol* 18(12): 900-905, 2008. PMID: 18538569. DOI: 10.1016/j.cub.2008.05.013
- Paeschke K, Capra JA and Zakian VA: DNA replication through g-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* pif1 DNA helicase. *Cell* 145(5): 678-691, 2011. PMID: 21620135. DOI: 10.1016/j.cell.2011.04.015
- Smith SS: Evolutionary expansion of structurally complex DNA sequences. *Cancer Genomics Proteomics* 7(4): 207-215, 2010. PMID: 20656986.
- Boyer AS, Grgurevic S, Cazaux C and Hoffmann JS: The human specialized DNA polymerases and non-b DNA: Vital relationships to preserve genome integrity. *J Mol Biol* 425(23): 4767-4781, 2013. PMID: 24095858. DOI: 10.1016/j.jmb.2013.09.022
- Fry M and Loeb LA: Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile x syndrome repeat sequence d(cgg)n. *J Biol Chem* 274(18): 12797-12802, 1999. PMID: 10212265. DOI: 10.1074/jbc.274.18.12797

- 19 Huber MD, Lee DC and Maizels N: G4 DNA unwinding by blm and sgs1p: Substrate specificity and substrate-specific inhibition. *Nucleic Acids Res* 30(18): 3954-3961, 2002. PMID: 12235379. DOI: 10.1093/nar/gkf530
- 20 Wu Y, Shin-ya K and Brosh RM, Jr.: Fancj helicase defective in fanconia anemia and breast cancer unwinds g-quadruplex DNA to defend genomic stability. *Mol Cell Biol* 28(12): 4116-4128, 2008. PMID: 218426915. DOI: 10.1128/MCB.02210-07
- 21 Chakraborty P and Grosse F: Wrn helicase unwinds okazaki fragment-like hybrids in a reaction stimulated by the human dhx9 helicase. *Nucleic Acids Res* 38(14): 4722-4730, 2010. PMID: 20385589. DOI: 10.1093/nar/gkq240
- 22 Grierson PM, Lillard K, Behbehani GK, Combs KA, Bhattacharyya S, Acharya S and Groden J: Blm helicase facilitates rna polymerase i-mediated ribosomal rna transcription. *Hum Mol Genet* 21(5): 1172-1183, 2012. PMID: 22106380. DOI: 10.1093/hmg/ddr545
- 23 Chib S, Byrd AK and Raney KD: Yeast helicase pif1 unwinds rna:DNA hybrids with higher processivity than DNA:DNA duplexes. *J Biol Chem* 291(11): 5889-5901, 2016. PMID: 26733194. DOI: 10.1074/jbc.M115.688648
- 24 Cantor S, Drapkin R, Zhang F, Lin Y, Han J, Pamidi S and Livingston DM: The brca1-associated protein bach1 is a DNA helicase targeted by clinically relevant inactivating mutations. *Proc Natl Acad Sci USA* 101(8): 2357-2362, 2004. PMID: 14983014. DOI: 10.1073/pnas.0308717101
- 25 He P and Yang W: Template and primer requirements for DNA pol theta-mediated end joining. *Proc Natl Acad Sci USA* 115(30): 7747-7752, 2018. PMID: 29987024. DOI: 10.1073/pnas.1807329115
- 26 Koole W, van Schendel R, Karambelas AE, van Heteren JT, Okihara KL and Tijsterman M: A polymerase theta-dependent repair pathway suppresses extensive genomic instability at endogenous g4 DNA sites. *Nat Commun* 5: 3216, 2014. PMID: 24496117. DOI: 10.1038/ncomms4216
- 27 Taylor MRG and Yeeles JTP: Dynamics of replication fork progression following helicase-polymerase uncoupling in eukaryotes. *J Mol Biol* 431: 2040-2049, 2019. PMID: 30894292. DOI: 10.1016/j.jmb.2019.03.011
- 28 Yeeles JTP, Poli J, Marians KJ and Pasero P: Rescuing stalled or damaged replication forks. *Cold Spring Harb Perspect Biol* 5: a012815, 2013. PMID: 23637285. DOI: 10.1101/cshperspect.a012815
- 29 Raghavan SC, Tsai A, Hsieh CL and Lieber MR: Analysis of non-b DNA structure at chromosomal sites in the mammalian genome. *Methods Enzymol* 409: 301-316, 2006. PMID: 16793408. DOI: 10.1016/S0076-6879(05)09017-8
- 30 Shevchuk T, Kretzner L, Munson K, Axume J, Clark J, Dyachenko OV, Caudill M, Buryanov Y and Smith SS: Transgene-induced ccwgg methylation does not alter cg methylation patterning in human kidney cells. *Nucleic Acids Res* 33(19): 6124-6136, 2005. PMID: 16246913. DOI: 10.1093/nar/gki920
- 31 Tuxhorn JA, Ayala GE, Smith MJ, Smith VC, Dang TD and Rowley DR: Reactive stroma in human prostate cancer: Induction of myofibroblast phenotype and extracellular matrix remodeling. *Clin Cancer Res* 8(9): 2912-2923, 2002. PMID: 1231536.
- 32 Munson K, Clark J, Lamparska-Kupsik K and Smith SS: Recovery of bisulfite-converted genomic sequences in the methylation-sensitive qpcr. *Nucleic Acids Res* 35(9): 2893-2903, 2007. PMID: 17439964. DOI: 10.1093/nar/gkm055
- 33 Yu K, Roy D, Huang FT and Lieber MR: Detection and structural analysis of r-loops. *Methods Enzymol* 409: 316-329, 2006. PMID: 16793409. DOI: 10.1016/S0076-6879(05)09018-X
- 34 Katoh K, Misawa K, Kuma K and Miyata T: Mafft: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res* 30(14): 3059-3066, 2002. PMID: 12136088. DOI: 10.1093/nar/gkf436
- 35 Notredame C, Higgins DG and Heringa J: T-coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302(1): 205-217, 2000. PMID: 10964570. DOI: 10.1006/jmbi.2000.4042
- 36 Brown NP, Leroy C and Sander C: Mview: A web-compatible database search or multiple alignment viewer. *Bioinformatics* 14(4): 380-381, 1998. PMID: 9632837. DOI: 10.1093/bioinformatics/14.4.380
- 37 Wright EP, Lamparska K, Smith SS and Waller ZAE: Substitution of cytosine with guanylyurea decreases the stability of i-motif DNA. *Biochemistry* 56(36): 4879-4883, 2017. PMID: 28853563. DOI: 10.1021/acs.biochem.7b00628
- 38 Botchan M, McKenna G and Sharp PA: Cleavage of mouse DNA by a restriction enzyme as a clue to the arrangement of genes. *Cold Spring Harb Symp Quant Biol* 38: 383-395, 1974. PMID: 4364785. DOI: 10.1101/sqb.1974.038.01.041
- 39 Hamer DH and Thomas CA, Jr.: The cleavage of *Drosophila melanogaster* DNA by restriction endonucleases. *Chromosoma* 49: 243-267, 1975. DOI: 10.1007/BF00361069
- 40 Hansel-Hertsch R, Spiegel J, Marsico G, Tannahill D and Balasubramanian S: Genome-wide mapping of endogenous g-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat Protoc* 13(3): 551-564, 2018. PMID: 29470465. DOI: 10.1038/nprot.2017.150
- 41 Raghavan SC, Swanson PC, Wu X, Hsieh CL and Lieber MR: A non-b-DNA structure at the bcl-2 major breakpoint region is cleaved by the rag complex. *Nature* 428(6978): 88-93, 2004. PMID: 14999286. DOI: 10.1038/nature02355
- 42 Lemmens B, van Schendel R and Tijsterman M: Mutagenic consequences of a single g-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat Commun* 6: 8909, 2015. PMID: 26563448. DOI: 10.1038/ncomms9909
- 43 Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X and Chedin F: Prevalent, dynamic, and conserved r-loop structures associate with specific epigenomic signatures in mammals. *Mol Cell* 63(1): 167-178, 2016. PMID: 27373332. DOI: 10.1016/j.molcel.2016.05.032
- 44 Nadel J, Athanasiadou R, Lemetre C, Wijetunga NA, P OB, Sato H, Zhang Z, Jeddelloh J, Montagna C, Golden A, Seoighe C and Grealley JM: RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin* 8: 46, 2015. PMID: 26579211. DOI: 10.1186/s13072-015-0040-640
- 45 Jenjaroenpun P, Wongsurawat T, Sutheworapong S and Kuznetsov VA: R-loopdb: A database for r-loop forming sequences (rlfs) and r-loops. *Nucleic Acids Res* 45(D1): D119-D127, 2017. PMID: 27899586. DOI: 10.1093/nar/gkw1054
- 46 Chedin F: Nascent connections: R-loops and chromatin patterning. *Trends Genet* 32(12): 828-838, 2016. PMID: 27793359. DOI: 10.1016/j.tig.2016.10.002
- 47 Cui Y, Kong D, Ghimire C, Xu C and Mao H: Mutually exclusive formation of g-quadruplex and i-motif is a general

- phenomenon governed by steric hindrance in duplex DNA. *Biochemistry* 55(15): 2291-2299, 2016. PMID: 27027664. DOI: 10.1021/acs.biochem.6b00016
- 48 Carroll AG, Voeller HJ, Sugars L and Gelmann EP: P53 oncogene mutations in three human prostate cancer cell lines. *Prostate* 23(2): 123-134, 1993. PMID: 8104329. DOI: 10.1002/pros.2990230206
- 49 Wang XW, Tseng A, Ellis NA, Spillare EA, Linke SP, Robles AI, Seker H, Yang Q, Hu P, Beresten S, Bemmels NA, Garfield S and Harris CC: Functional interaction of p53 and blm DNA helicase in apoptosis. *J Biol Chem* 276(35): 32948-32955, 2001. PMID: 11399766. DOI: 10.1074/jbc.M103298200
- 50 Yang Q, Zhang R, Wang XW, Spillare EA, Linke SP, Subramanian D, Griffith JD, Li JL, Hickson ID, Shen JC, Loeb LA, Mazur SJ, Appella E, Brosh RM, Jr., Karmakar P, Bohr VA and Harris CC: The processing of holliday junctions by blm and wrn helicases is regulated by p53. *J Biol Chem* 277(35): 31980-31987, 2002. PMID: 12080066. DOI: 10.1074/jbc.M204111200
- 51 Sarkies P, Murat P, Phillips LG, Patel KJ, Balasubramanian S and Sale JE: Fancj coordinates two pathways that maintain epigenetic stability at g-quadruplex DNA. *Nucleic Acids Res* 40(4): 1485-1498, 2012. PMID: 22021381. DOI: 10.1093/nar/gkr868
- 52 Wang G and Vasquez KM: Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst)* 19: 143-151, 2014. PMID: 24767258. DOI: 10.1016/j.dnarep.2014.03.017
- 53 Duquette ML, Handa P, Vincent JA, Taylor AF and Maizels N: Intracellular transcription of g-rich dnas induces formation of g-loops, novel structures containing g4 DNA. *Genes Dev* 18(13): 1618-1629, 2004. PMID: 15231739. DOI: 10.1016/j.dnarep.2014.03.017
- 54 Hoopes JI, Cortez LM, Mertz TM, Malc EP, Mieczkowski PA and Roberts SA: Apobec3a and apobec3b preferentially deaminate the lagging strand template during DNA replication. *Cell Rep* 14(6): 1273-1282, 2016. PMID: 26832400. DOI: 10.1016/j.celrep.2016.01.021
- 55 Ito F, Fu Y, Kao SA, Yang H and Chen XS: Family-wide comparative analysis of cytidine and methylcytidine deamination by eleven human apobec proteins. *J Mol Biol* 429(12): 1787-1799, 2017. PMID: 28479091. DOI: 10.1016/j.jmb.2017.04.021
- 56 Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, Klimczak LJ, Kryukov GV, Malc E, Mieczkowski PA, Resnick MA and Gordenin DA: Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol Cell* 46(4): 424-435, 2012. PMID: 22607975. DOI: 10.1016/j.molcel.2012.03.030
- 57 Qiao Q, Wang L, Meng FL, Hwang JK, Alt FW and Wu H: Aid recognizes structured DNA for class switch recombination. *Mol Cell* 67(3): 361-373 e364, 2017. PMID: 28757211. DOI: 10.1016/j.molcel.2017.06.034
- 58 van Kregten M and Tijsterman M: The repair of g-quadruplex-induced DNA damage. *Exp Cell Res* 329(1): 178-183, 2014. PMID: 25193076. DOI: 10.1016/j.yexcr.2014.08.038
- 59 Langston LD and O'Donnell M: DNA replication: Keep moving and don't mind the gap. *Mol Cell* 23(2): 155-160, 2006. PMID: 16857582. DOI: 10.1016/j.molcel.2006.05.034
- 60 Castellano-Pozo M, Garcia-Muse T and Aguilera A: R-loops cause replication impairment and genome instability during meiosis. *EMBO Rep* 13(10): 923-929, 2012. PMID: 22878416. DOI: 10.1038/embor.2012.119
- 61 Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP and Balasubramanian S: High-throughput sequencing of DNA g-quadruplex structures in the human genome. *Nat Biotechnol* 33(8): 877-881, 2015. PMID: 26192317. DOI: 10.1038/nbt.3295
- 62 Marsico G, Chambers VS, Sahakyan AB, McCauley P, Boutell JM, Antonio MD and Balasubramanian S: Whole genome experimental maps of DNA g-quadruplexes in multiple species. *Nucleic Acids Res* 47(8): 3862-3874, 2019. PMID: 308926626. DOI: 10.1093/nar/gkz179
- 63 Biffi GT, Tannahill D, McCafferty J and Balasubramanian S: Quantitative visualization of DNA g-quadruplex structures in human cells. *Nat Chem* 5: 182-186, 2013. PMID: 23422559. DOI: 10.1038/nchem.1548
- 64 Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N and Oda M: Eukaryotic chromosome DNA replication: Where, when, and how? *Annu Rev Biochem* 79: 89-130, 2010. PMID: 20373915. DOI: 10.1146/annurev.biochem.052308.103205

*Received January 7, 2020*  
*Revised January 25, 2020*  
*Accepted January 27, 2020*