frontiers in Genetics

ORIGINAL RESEARCH
published: 11 May 2015
doi: 10.3389/fgene.2015.00177

Sites of instability in the human *TCF3* (*E2A*) gene adopt G-quadruplex DNA structures *in vitro*

Jonathan D. Williams[1], Sara Fleetwood[1], Alexandra Berroyer[1], Nayun Kim[2] and Erik D. Larson[1]*

[1] School of Biological Sciences, Illinois State University, Normal, IL, USA, [2] Department of Microbiology and Molecular Genetics, University of Texas Health Science Center at Houston, Houston, TX, USA

The formation of highly stable four-stranded DNA, called G-quadruplex (G4), promotes site-specific genome instability. G4 DNA structures fold from repetitive guanine sequences, and increasing experimental evidence connects G4 sequence motifs with specific gene rearrangements. The human transcription factor 3 (*TCF3*) gene (also termed *E2A*) is subject to genetic instability associated with severe disease, most notably a common translocation event t(1;19) associated with acute lymphoblastic leukemia. The sites of instability in *TCF3* are not randomly distributed, but focused to certain sequences. We asked if G4 DNA formation could explain why *TCF3* is prone to recombination and mutagenesis. Here we demonstrate that sequences surrounding the major t(1;19) break site and a region associated with copy number variations both contain G4 sequence motifs. The motifs identified readily adopt G4 DNA structures that are stable enough to interfere with DNA synthesis in physiological salt conditions *in vitro*. When introduced into the yeast genome, *TCF3* G4 motifs promoted gross chromosomal rearrangements in a transcription-dependent manner. Our results provide a molecular rationale for the site-specific instability of human *TCF3,* suggesting that G4 DNA structures contribute to oncogenic DNA breaks and recombination.

Keywords: G-quadruplex, G4, t(1;19), *TCF3*, *PBX1*

OPEN ACCESS

Edited by:
David Azorsa,
Ronald A. Matricaria Institute of Molecular Medicine at Phoenix Children's Hospital, USA

Reviewed by:
Karen Usdin,
National Institutes of Health, USA
Eiman Aleem,
Phoenix Children's Hospital and University of Arizona College of Medicine-Phoenix, Germany

****Correspondence:***
Erik D. Larson,
School of Biological Sciences, Illinois State University, Campus Box 4120, Normal, IL 61790-4120, USA
elarson@ilstu.edu

Specialty section:
This article was submitted to Cancer Genetics, a section of the journal Frontiers in Genetics

Received: *18 February 2015*
Accepted: *25 April 2015*
Published: *11 May 2015*

Citation:
Williams JD, Fleetwood S, Berroyer A, Kim N and Larson ED (2015) Sites of instability in the human TCF3 (E2A) gene adopt G-quadruplex DNA structures in vitro. Front. Genet. 6:177. doi: 10.3389/fgene.2015.00177

Introduction

Transcription factor 3, also called E2A, is a key regulatory protein that institutes transcriptional programs for proper B and T cell differentiation (Kee et al., 2000; Miyazaki et al., 2014). Given well-established regulatory roles in transcription activation, it is not surprising that disruption of *TCF3* or its gene product is associated with malignant transformation. For instance, a translocation event t(1;19) between the *TCF3* and *PBX1* genes results in the expression of a TCF3–PBX1 chimera, which is commonly found in ALL (Hunger, 1996; Aspland et al., 2001; Pui et al., 2004). Genomic studies of Burkitt's lymphomas found *TCF3* to be among the most mutated genes (Schmitz et al., 2012) and 70% of patient samples were identified with heterozygous *TCF3* deletions in Sezary syndrome cells, an aggressive T cell lymphoma (Steininger et al., 2011). In addition to pre-B cell cancers, TCF3–PBX1 fusion transcripts have been identified in non-small cell lung cancer (Mo et al., 2013), indicating that the impact of this rearrangement likely extends beyond the

Abbreviations: ALL, acute lymphoblastic leukemia; CD, circular dichroism; CNV, copy number variant; G4, guanine quadruplex; *PBX1*, Pre-B cell leukemia homeoboX 1 gene; *TCF3*, transcription factor 3 (*E2A*).

immune system. While *TCF3* appears to be a hot spot for DNA breaks, and broadly associated with oncogenesis, the mechanisms responsible for driving this genetic instability are undefined.

The distribution of chromosomal breakpoints involved in the t(1;19) translocation suggest that certain sequences are involved in promoting instability. Previous characterization of t(1:19) breakpoints revealed that most of the recombination junctions (16 of 24) cluster in a 5 bp sequence window in *TCF3* (Wiemels et al., 2002). Those breaks were associated with CpG sequences in *TCF3*, but not *PBX1* (Tsai et al., 2008), and these same recombination sites are surrounded by transposable element repeats, specifically a MER20 transposon (Rodić et al., 2013). The non-random distribution of break sites and proximity to DNA repeat motifs suggest that the recombination events involving *TCF3* are influenced at either the DNA sequence or structural level.

DNA repeats can promote instability because of their ability to adopt structural conformations that interfere with normal DNA transactions (Lobachev et al., 2007; Zhao et al., 2010; van Kregten and Tijsterman, 2014). In particular, guanine repeats readily fold into highly stable four-stranded structures called G4 DNA, which promote mutagenesis and recombination (Brooks et al., 2010; Wu and Brosh, 2010; Bochman et al., 2012; Tarsounas and Tijsterman, 2013; van Kregten and Tijsterman, 2014). G4 DNA structures are stabilized by hydrogen bonding between four guanine bases to create a single "tetrad" of guanines. When tandem guanine repeats are present, stacks of tetrads can form within or among DNA strands to build the four-stranded, or quadruplex, structure. The size, stability, and specific type of G4 structure depends upon the characteristics of the repeat sequence and aqueous conditions (Burge et al., 2006).

G-quadruplex DNA structures have been identified in loci involved in both induced and spontaneous genome instability (Maizels and Gray, 2013). At the guanine-rich immunoglobulin switch regions, programmed recombination requires transcriptional activation, and transcribed switch regions can form loops that are stabilized by RNA/DNA hybrids on one strand and G4 DNA on the other (Duquette et al., 2004). Factors involved in this recombination pathway specifically recognize G4 DNA (Larson et al., 2005). In addition to induced recombination events, G4 DNA structures may lead to instability at other guanine-rich genomic loci. In a few examples, G4 DNA structures fold from sequences flanking the translocation break sites in the *HOX11* gene (Nambiar et al., 2013). Similarly, sequences near the break sites in *BCL2* that lead to t(14;18) translocations form G4 structures *in vitro* (Nambiar et al., 2011). This is probably not a feature of just a few unstable genes because in lymphoid cancers G4 sequence motifs were found at multiple rearrangement sites, including the *TCF3* gene (Katapadi et al., 2012). Experimental systems have also directly connected instability of guanine-rich DNA with the formation of G4 structures. For instance, chromosomal rearrangements at guanine-rich human minisatellite sequences were dependent upon G4 formation (Lopes et al., 2011; Piazza et al., 2012), and addition of G4-stabilizing ligands increased the instability at the G4 repeats, but not for other types of sequence repeats (Piazza et al., 2010). Chromosomal rearrangement assays have also been developed to characterize specific loci and factors involved in G4-mediated genome instability (Piazza et al., 2012; Paeschke et al., 2013; Yadav et al., 2014).

Considering the emerging evidence connecting guanine-rich sequences and G4 DNA structures with genome instability, we asked if the known rearrangements of the *TCF3* gene correlate with G4 structure formation. Here, we applied a comprehensive computational analysis of *TCF3* and the translocation partner *PBX1* and found that G4 sequence motifs reside near the major t(1;19) breakpoints. Using multiple methods, we demonstrate here that those sequences fold into highly stable G4 structures *in vitro* and induced site-specific instability in yeast. We also characterize a second site in *TCF3* for G4 formation that is not involved with the t(1;19) translocation, but is instead associated with CNVs. Our results indicate that the site-specific instability of human *TCF3* is governed at least in part by a capacity to form structures at those sites.

Results and Discussion

G4 Sequence Motifs Surround Regions of Instability in *TCF3*

Based on a growing body of evidence implicating G4 DNA structures with site-specific rearrangements, we asked if formation of G4 DNA structures could explain the apparent genetic instability of the *TCF3* gene. The positions of 30 different breakpoints in *TCF3* spanning 4 kb, and 16 breakpoints in *PBX1* spanning 12 kb are documented in the Translocations In Cancer database (TICdb; Novo et al., 2007) and the Leiber database (Tsai et al., 2008). We used these sites to analyze and map break positions in relation to G4 sequence motifs. This was accomplished by applying a web-based server program for G4 structure prediction to overlay G4 sequences with known break site positions.

Quadruplex forming G-rich sequences (QGRS mapper) is a web-based server program developed to score nucleic acid sequences for G4 forming potential (Kikin et al., 2006). QGRS mapper converts input sequences into scores representing the likelihood for G4 formation, with a maximum score value of 180 given for a run of guanines that is 27 nt long (Kikin et al., 2006). The location of G4 motifs within the queried sequence is an output of the program (Kikin et al., 2006). **Figures 1A,B** diagrams the output for QGRS scoring of *TCF3* and *PBX1* genes. Both strands were analyzed for non-overlapping motifs. Two regions of *TCF3* are associated with genomic instability; the t(1;19) translocation site (which forms the TCF3–PBX1 chimera), and a region rich in breaks connected to CNVs. Both of these regions are intronic. QGRS mapper identified 25 different G4-capable sequences at the CNV site, and 15 G4 sequences surrounding the major t(1;19) break site (**Figure 1A**). An additional region of high guanine repeat density can be found in the 5′ end of the gene, but this site did not correlate with any known breakpoints, suggesting the genetic positioning of G4 sequences influences structure formation, or that instability at that site is not connected with diseases cataloged in existing databases. *PBX1* also contains G4 motifs near the t(1;19) translocation site, but G4 motifs occur at much lower density compared to *TCF3*, with 2 found for *PBX1* (**Figure 1B**) compared to 15 for *TCF3* (**Figure 1A**).

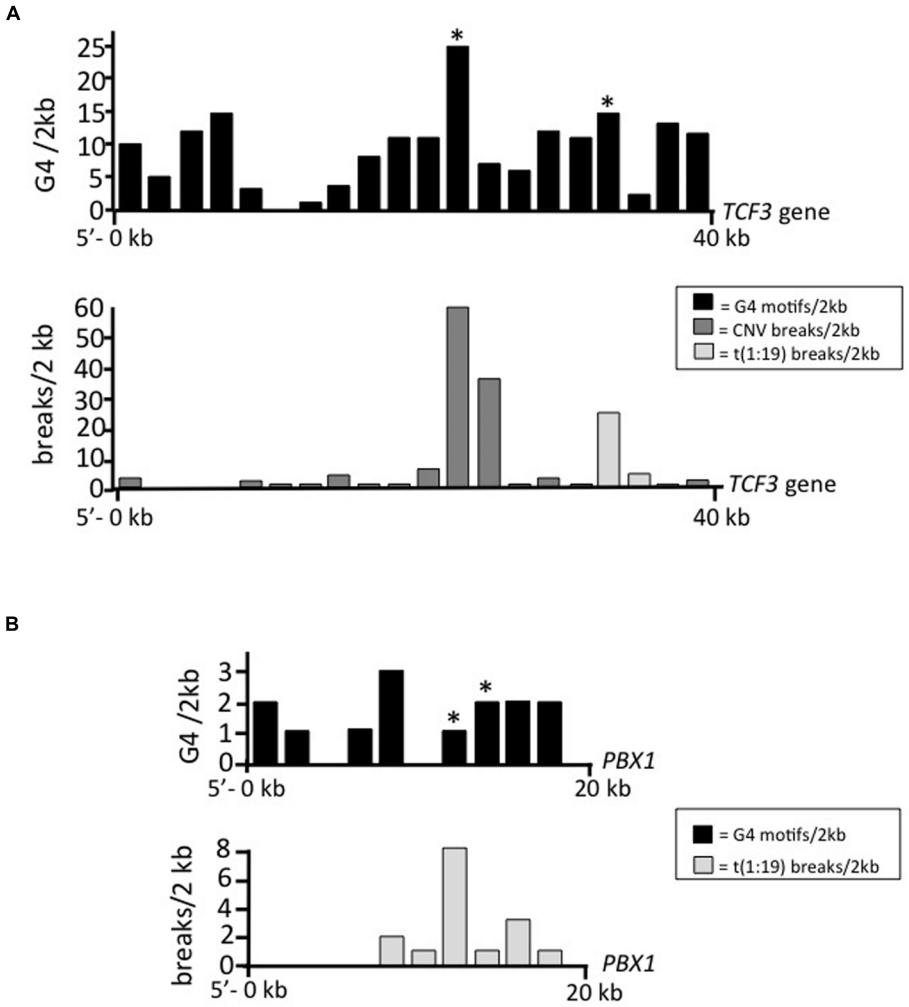

FIGURE 1 | Genome instability coincides with G4 motifs in *TCF3* and *PBX1*. (A,B) The number of individual non-overlapping G4 sequence motifs (black) identified in 2 kb windows (Y-axis) graphed according to genetic position (X-axis) within the **(A)** *TCF3* gene or **(B)** 20 kb region of *PBX1* intron associated with the t(1;19) translocation, displayed in the 5′–3′ direction. Regions corresponding to motifs tested for G4 formation are indicated (*). The bar graphs below the *TCF3* and *PBX1* gene diagrams show the relative locations of break sites (light gray) and CNVs (dark gray) identified using 2 kb sequence windows.

TCF3 and *PBX1* G4 Motifs Support G4 Structure Formation *In Vitro*

Considering the breadth of experimental evidence linking site-specific instability with G4 DNA structures (Brooks et al., 2010; Wu and Brosh, 2010; Bochman et al., 2012; Tarsounas and Tijsterman, 2013; van Kregten and Tijsterman, 2014), we asked if the guanine-rich motifs we identified using QGRS mapper also form stable G4 structures in physiological salt and pH conditions. We selected multiple G4 sequence motifs for each break site locus. This includes one repeating G4 motif from a large guanine-rich intron connected with CNVs in *TCF3* (T-lg), three motifs flanking the t(1;19) breakpoint [T-5′, T-3′, and T-3′(2)], and two motifs next to the *PBX1* t(1;19) breakpoints (P-1 and P-2). Sequence and QGRS scores for each motif we tested are listed in **Table 1**. Importantly, the sequence "T-5′ " is 20 bases 5′ and "T-3′" is just 2 bases 3′ of a t(1;19) breakpoint cluster. T-3′(2) is a second G4 motif located ∼1200 bp from the major break site cluster. P-1 and P-2 are 70 and 750 bp from a t(1;19) break site cluster, respectively. However, unlike the breakpoints in the *TCF3* locus, the breakpoints in intron 1 of *PBX1,* involved with TCF3–PBX1 fusions, are more broadly distributed (Wiemels et al., 2002), suggesting that the G4 sequence motifs residing within the *PBX1* locus may not instigate translocations with *TCF3* to same degree as the *TCF3* G4 sequences.

Each single-stranded guanine-rich motif (termed "G4") was synthesized along with a companion control in which tandem guanine repeats of three or more were disrupted by substituting thymine, thereby greatly reducing the potential for G4 formation (termed "GT"; Supplementary Table S1). GT control and G4 motif oligonucleotides co-migrated on denaturing PAGE (**Figure 2A**, top), as expected. G4 DNA is stabilized by K^+ or Na^+ ions (Williamson et al., 1989; Sen and Gilbert, 1990). In

the presence of 100 mM KCl, all of the *TCF3* and *PBX1* G4 oligonucleotides migrated as larger and smaller species compared to the GT interrupted control, which migrated as a single product in native PAGE (**Figure 2A**, bottom). The slow migrating species are consistent with multi-molecular structures, and species migrating faster than the GT control are consistent with mono-molecular, or self-pairing, conformations. GT control oligonucleotides retained identical mobility patterns in Native PAGE experiments independent of the presence of KCl (not shown), as expected. T-5′ retains some self-complementarity even when the guanine repeats were disrupted by thymine (GT control), likely explaining the faster mobility pattern observed in Native PAGE. Consistent with a mono-molecular structural conformation, scrambling the T-5′ sequence (T-5′-S) reduced its mobility compared to the companion GT and G4 samples (**Figure 2A**, bottom left). Together, we conclude that the guanine-rich sequences derived from *TCF3* and *PBX1* break site regions adopt alternative DNA conformations in the presence of K^+. This is consistent with G4 DNA.

We further tested the ability of each *TCF3* and *PBX1* sequence motif to adopt G4 DNA structures in solution using circular dichroism (CD), comparing spectra of those oligonucleotides with controls that cannot adopt stable G4 conformations. CD measures the differential absorption of circularly polarized light by chiral molecules (called ellipticity). It is best suited for identifying the presence of structures, like G4, but not sensitive enough for atomic-level resolution. G4 DNA structures produce characteristic CD spectra, with positive peaks at either 264 or 295 nm and negative dips at 240 or 265 nm, respectively, depending on the type of quadruplex (Balagurumoorthy et al., 1992; Đapić et al., 2003; Kypr et al., 2009; Vorlíčková et al., 2012). The CD spectra for all *TCF3* motifs tested showed molar ellipticities that peak at ∼264 nm and a dip at ∼240 nm (**Figure 2B**), consistent with G4 DNA. Interestingly, T-5′ shows a shallow and broadened peak that extends beyond 280 nm (**Figure 2B**, top left), probably reflecting the presence of non-G4, or B-form variant, structures (Kypr et al., 2009), which is consistent with Native PAGE analysis for this oligonucleotide (**Figure 2A**). CD analysis of T-5′ at a twofold higher concentration resulted in larger 260 nm maximum and 240 nm minimum peaks (Supplementary Figure S1), suggesting that at the lower concentration B-form conformations (i.e., self pairing) either reduces the potential for or competes with G4 DNA structure formation. Either way, the T-5′ sequence appears to adopt multiple DNA conformations that deviate from standard duplex. *PBX1* sequences P-1 and P-2 both showed CD spectra comparable to *TCF3* G4 DNA (**Figure 2B**, top right). As expected, interruption of the tandem guanine repeats by thymine substitution (Supplementary Table S1) eliminated the characteristic CD spectra for G4 (**Figure 2B**, bottom). We conclude that the sequences surrounding major break site regions in *TCF3* and *PBX1* adopt G4 DNA conformations in solution.

TCF3 and *PBX1* G4 Structures Block DNA Synthesis *In Vitro*

The precise mechanisms by which formation of G4 DNA structures induce genome instability are not defined. However, regions of repetitive DNA that become transiently denatured have an opportunity to interact, or self pair, to form structures that interfere with DNA metabolism (Zhao et al., 2010; Lopes et al., 2011; van Kregten and Tijsterman, 2014). Presumably, highly stable non-duplex DNA structures are more likely to interfere with DNA transactions compared with those that are less stable. To test that model for *TCF3* G4 sequence motifs we next employed primer extension assays using genomic sequences surrounding the major break sites. This assay has been well described for characterizing G4 formation, showing that K^+ ions support guanine-dependent G4 formation and polymerase pausing, while Li^+ and NH_4^+ ions do not (Weitzmann et al., 1996; Sun and Hurley, 2010; Ehrat et al., 2012). Using that system, we expected to find K^+-dependent polymerase pausing on templates containing *TCF3* and *PBX1* break site sequences. Templates for this assay contained the G4 motifs shown in **Table 1**, but also some additional genomic sequence (Supplementary Figure S2). We cloned each sequence into the plasmid pCR2.1 (Invitrogen, Carlsbad, CA, USA) in both orientations with respect to an F1 origin and the primer-binding site. Single-stranded phagemids were then isolated for each orientation (cytosine-rich or guanine-rich sequences) and used as single-stranded templates for primer extension reactions catalyzed by Klenow polymerase. Independent of specific template, polymerase extension reactions showed full-length product when the cytosine-rich strands (complements to the G4 motifs) were assayed, and synthesis was independent of the salt used, as expected (**Figure 3A**, left and Supplementary Figure S3). In contrast, extension reactions prematurely paused on the guanine-rich templates when K^+, but not Li^+, was present (**Figure 3A**, right). Although fully consistent with G4 DNA, we further tested the dependence of stalling on guanine by using thymine substitution mutagenesis to disrupt

TABLE 1 | *TCF3* and *PBX1* G4 sequences.

Name	Sequence	QGRS score
T-5′-G4	CCA**GGGG**ACACT**GGG**TGATGTCT**GGGG**ACATCTACAGTTGTCA**GGG**CTGA**GGGG**AGC	62
T-3′-G4	AGA**GGG**AGAGA**GGG**AA**GGGGGG**A**GGG**C**GGGG**CA**GGG**CAG	72
T-3′(2)-G4	A**GGG**AGT**GGGG**ACGTGAAT**GGGG**TGCGA**GGGG**C**GGGG**TG	101
T-Ig-G4	A**GGGGG**TGAGGC**GGG**AA**GGGG**ACAGCAGAACTCAC**GGGG**T	61
P-1-G4	GGT**GGGGG**CAGGTT**GGG**A**GGGG**AGGA**GGG**CAGATCTACA**GGG**A**GGG**TGG	70
P-2-G4	GCT**GGGG**T**GGGG**A**GGG**AAGAGATGA**GGGGG**A**GGG**AGA	66

*Names given for each sequence are in the far left column. Guanines that are predicted to be involved in G4 formation are bolded. QGRS scores are listed in the right column. The T-Ig G4 sequence is not associated with t(1;19) break sites, it is a repeat motif located at the CNV site diagramed in **Figure 1**.*

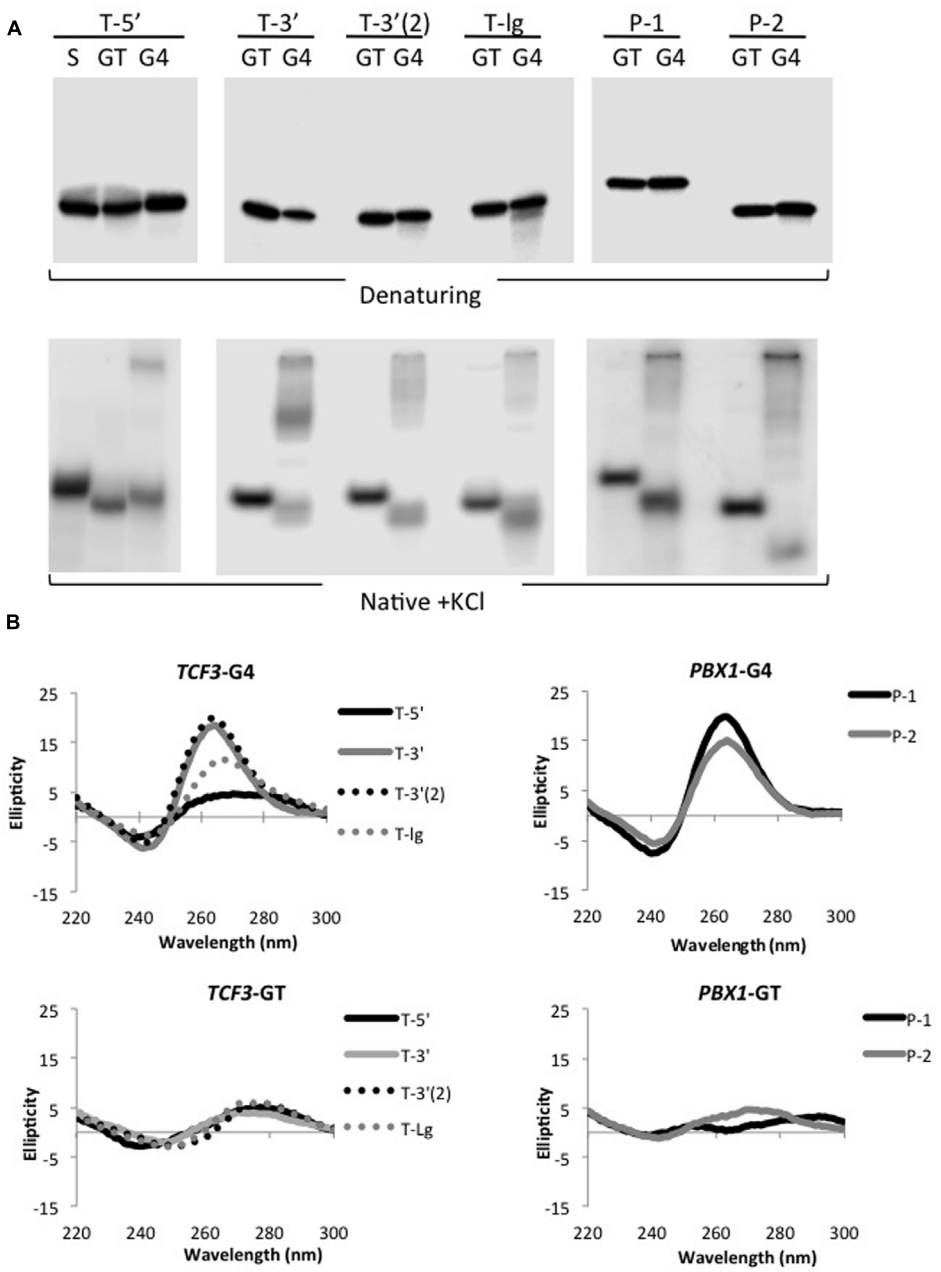

FIGURE 2 | Sequences from *TCF3* and *PBX1* adopt G4 conformations in solution. (A) Phosphorimages showing 5′ end labeled guanine-rich oligonucleotides (G4) and corresponding thymine substituted controls (GT) resolved by PAGE. T-5′ also includes an additional scrambled control (S), where nucleotides were reordered to remove the potential for stable hairpin folding. Migration of each radiolabeled oligonucleotide upon denaturing PAGE (top) and native PAGE (bottom) is shown. **(B)** CD spectra for *TCF3* (top left) and *PBX1* (top right) G4 motifs. CD spectra are shown for thymine-substituted (GT) controls for *TCF3* (bottom left) and *PBX1* motifs (bottom right).

the repeats in the T-5′, P-1, and P-2 templates (Supplementary Figure S2). Full extension was restored when the G4 motifs were disrupted (Supplementary Figure S4), essentially matching results for the C-rich complements (**Figure 3A**), and demonstrating that the guanine repeats are needed for the polymerase pausing. Therefore, this stalling of synthesis argues that G4 structures can form from the *TCF3* and *PBX1* break site sequences.

We next asked if the G4 structures formed within *TCF3* and *PBX1* sequences are thermally stable, and thereby capable of adopting difficult to resolve structural conformations

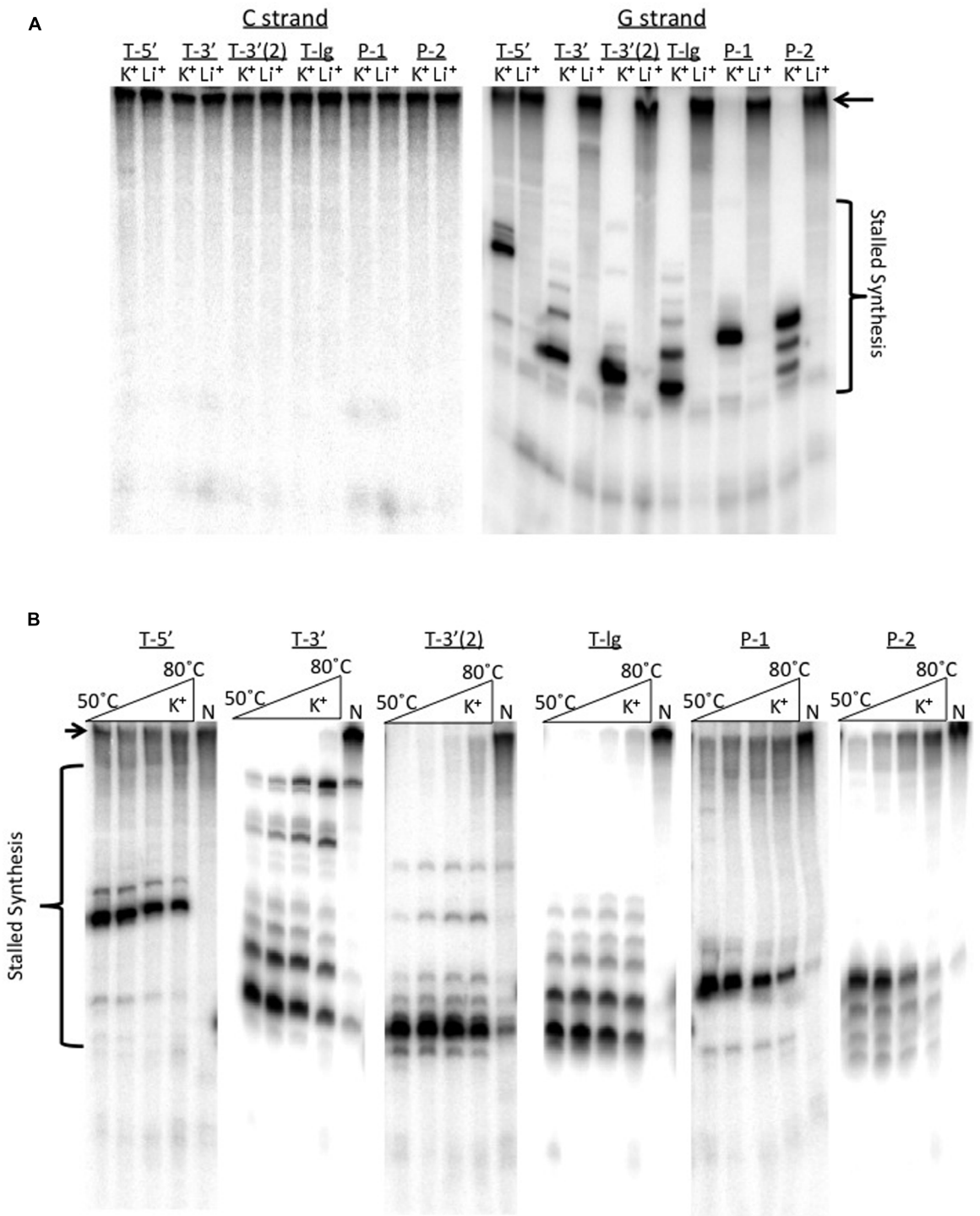

FIGURE 3 | Guanine-rich templates from *TCF3* and *PBX1* block DNA synthesis *in vitro*. (A) Klenow polymerase extension assays using templates from the cytosine-rich strand (C-strand, left) or guanine-rich strand (G-strand, right) for each G4 sequence motif, resolved by denaturing PAGE. Reactions were performed in either KCl (K^+) or LiCl (Li^+). Shown are bands for stalled DNA synthesis (bracket) and full-length extension product (black arrow). G4 sequences begin at the bottom of the brackets. T-3′ differs from the motif listed in **Table 1**, it includes 472 bp of surrounding genomic sequence (Supplementary Figure S2). **(B)** Primer extension reactions used Taq polymerase across a temperature range (50–80°C) in either KCl (K^+) or $(NH_4)_2SO_4$ (N) salt conditions on guanine-rich templates from *TCF3* and *PBX1*. Bands corresponding to stalled DNA synthesis (bracket) or full-length extension (black arrow) are shown.

in the cell. Taq polymerase, used in standard PCR, has optimal activity around 75°C (Lawyer et al., 1993). One would predict that at elevated temperatures, beyond 37°C, any polymerase pausing would reflect the presence of highly stable template blockades. Based on that logic, we replicated the primer extension experiments described above with Taq polymerase at ranging temperatures in reactions containing either K^+ or NH_4^+ salt. Resolution of the Taq extension products by denaturing PAGE revealed polymerase-stalling patterns similar to that of Klenow (**Figure 3B**). Importantly, bands corresponding to stalled synthesis on the guanine-rich templates were only marginally altered when reaction temperatures reached 80°C, suggesting that the replication blockades formed within the *TCF3* and *PBX1* templates are thermally stable (**Figure 3B**). Full extension was observed when NH_4^+ was substituted for K^+ (**Figure 3B**) or when the C-rich strand served as the template (Supplementary Figure S3), as expected. We conclude that the G-rich sequence motifs located proximal to the *TCF3* and *PBX1* t(1;19) translocation sites (**Figure 1**) and to the *TCF3* CNV break sites (**Figure 1A**) all form G4 structures capable of interfering with DNA synthesis *in vitro*.

TCF3 Break Site G4 Motifs and Instability

Although our *in vitro* results suggests that sequences proximal to *TCF3* breaks adopt G4 DNA, it does not necessarily connect structure formation with site-specific instability in the cell. Therefore, we next asked if the G4 motifs surrounding the *TCF3* t(1;19) break site promote instability in yeast using a previously described genetic assay. In these systems, a model G4-forming sequence from the murine Sμ Ig switch region was shown to enhance ectopic recombination (Kim and Jinks-Robertson, 2011) and gross chromosome rearrangements (GCRs; Yadav et al., 2014). The genome instability occurring at the Sμ Ig switch region was influenced by conditions that impact G4 structure formation, such as high transcription rate, orientation of the sequence with respect to the promoter, and disruption of structure metabolizing enzymes like Sgs1, a G4 specific helicase (Huber et al., 2002), or Top1 topoisomerase (Yadav et al., 2014). In order to test whether G4 motifs identified at the *TCF3* translocation breakpoints can also induce genome instability, we modified the GCR assay, which selects for the simultaneous loss of the *CAN1* and *URA3* genes located telomeric to a reporter cassette that is integrated on Chromosome V. This cassette consists of *LYS2* gene transcribed from the tetracycline-repressible promoter *pTET*. A 472 bp sequence (T-3′-G4 Supplementary Figure S2) surrounding the *TCF3* t(1;19) G4 motif was introduced into this cassette so that the G-rich strand is positioned on either the non-transcribed (T-G-Top) or transcribed (T-G-Bottom) strand with respect to the *pTET* promoter. In wild type backgrounds under high transcription conditions (WT), the rates of GCR for T-G-Top and T-G-Bottom did not differ significantly (**Figure 4A**). GCR rates increased significantly for both T-G-Top and T-G-Bottom by 69- and 36-fold, respectively, in Top1-deficient yeast strains. This is similar to the G4-associated genetic instability observed for the model G4 sequence, Ig Sμ (Yadav et al., 2014). When the transcription from *pTET* was repressed by addition of tetracycline analog doxycycline (+DX), the rates of GCR for T-G-Top and T-G-Btm were both reduced by ∼4-fold compared to the high transcription conditions (**Figure 4A**). The difference in instability between T-G-Top and T-G-Bottom in both transcription conditions ranged from ∼2.7 to 3-fold. Therefore, in this assay T-G-Top, which is in a G4 favorable orientation, promoted the formation of DNA breaks leading to chromosomal rearrangements. However, transcriptional orientation had less of an impact on instability compared to the model Sμ G4 sequence, with about a 28-fold difference between the two Sμ transcriptional orientations (Yadav et al., 2014) compared to threefold difference for *TCF3* under high transcription conditions (**Figure 4A**). This probably reflects the difference in the density of G-runs and the relative sizes of the *TCF3* and Sμ sequences used in the assay. Regardless, the t(1;19) G4 break site sequence from *TCF3* displayed co-transcriptional genetic instability, fully consistent with the model that G4 DNA formation promotes *TCF3* instability.

The second site of instability in *TCF3* (T-lg) is associated with CNVs (**Figure 1**), and those sequences readily adopted G4 structures *in vitro* (**Figures 2** and **3**). We next asked if this site in *TCF3* contains signatures for G4-mediated instability by using the existing genome variation databases. T-lg is extensively repetitive (with 106 GGG repeats), so it is reasonable to predict that the CNVs mapping to this region are associated with DNA breaks related to the G4 motifs and G4 structure formation. If so, we would expect to find DNA break signatures proximal to the tandem guanine repeats. Therefore, we examined the entire *TCF3* gene using the human database for short genetic variation (dbSNP; Sherry et al., 2001) available on Ensembl77 (Kersey et al., 2014), looking specifically to map the positions of short (<100 bp) insertions and deletions. At the CNV site (T-lg), there are approximately 8-fold more insertions and deletions (indels) compared to the rest of *TCF3* (**Figure 4B**). The locations of the sequence variations are not random, with 93% of indels directly next to or inside of a tandem guanine repeat (Supplementary Figure S5). Fine mapping of deletion positions shows that all directly flank or reside within the same guanine repeat sequence, which is part of a sequence motif repeated 32 times in the T-lg sequence (Supplementary Figure S6). Insertion mutations were also distributed within two nucleotides of the tandem guanine repeats, but there are fewer insertions overall compared to deletions (Supplementary Figure S6). Although it is not immediately clear what the pattern or type of sequence variation indicates with regard to the mechanism(s) of instability, the loss, and gain of sequence coincides with repetitive guanines.

We favor a model whereby guanine-rich DNA in *TCF3* adopts G4 DNA structures, and failure to resolve those structures promotes DNA breaks. *TCF3* instability due to G4 formation is manifested in the available genome databases as copy number variations and translocation events. This is significant because it is currently unclear why the *TCF3* gene is unstable. An alternate possibility to consider is that the guanine repeats in *TCF3* alone are unstable, or that instability is otherwise independent of G4 structure formation. When long guanine-rich sequences are positioned on the non-transcribed strand the cytosine rich complement can participate in RNA–DNA

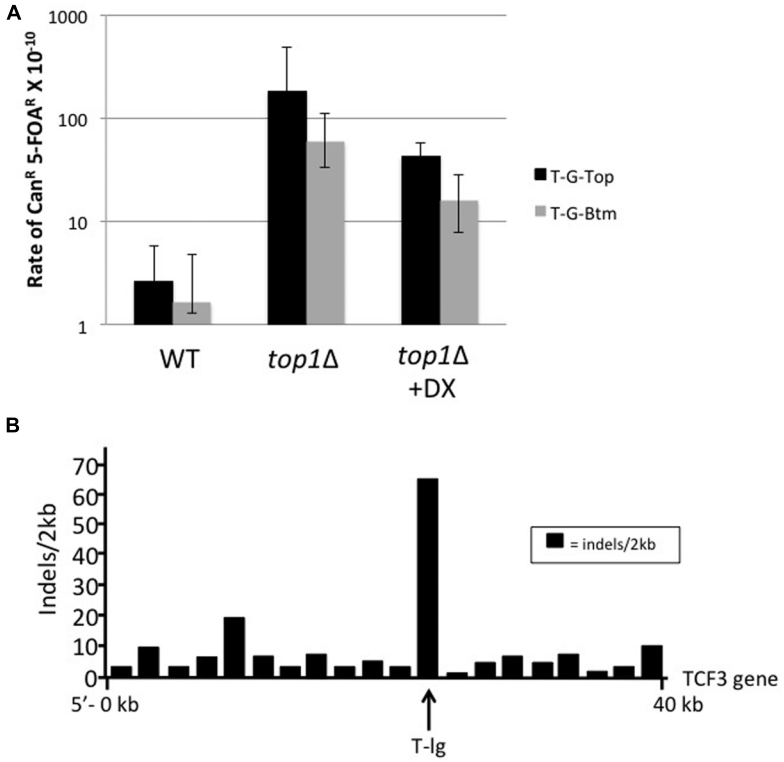

FIGURE 4 | *TCF3* G4 motifs promote genetic instability in yeast. **(A)** Rate of gross chromosomal rearrangements for the t(1;19) major break site sequence (T-3′-G4 motif) in the *pTET-lys2 T-G-Top* (solid black bar) or *T-G-Btm* (gray bar) for WT or *top1* deletion yeast strains. Where indicated (+DX), transcription was repressed by addition of doxycycline to 2 mg/l. Error bars indicate 95% confidence intervals. **(B)** *TCF3* intron with high CNV (T-Ig) also shows high levels of small insertions and deletions (<100 bp). Insertion and deletions (indels) identified in *TCF3* using 2 kb sequence windows are mapped (indels/2 kb) according to genetic position.

hybridizations during transcription to form structures called R-loops, which promote instability (Hamperl and Cimprich, 2014). Indeed, recent results suggest that R-loops are cleaved by factors required for transcription-coupled nucleotide excision repair, providing a potential mechanism for instability at guanine-rich DNA (Sollier et al., 2014). G4 can form on the unpaired strand, and those structures were observed in electron microscopy studies of transcribed switch regions (Duquette et al., 2004), although there is also evidence that R-loop stability may not depend upon G4 formation (Roy et al., 2008). We did not directly test the capacity of various *TCF3* sequences to form R-loop structures in this work. However, we did observe that instability was influenced by transcription and the transcriptional orientation of T-3′ had a modest affect (∼3-fold) on gross chromosomal loss in yeast (**Figure 4A**). Therefore, we cannot exclude the possibility that R-loop formation itself contributes to *TCF3* instability at some level.

Our findings add to a growing list of oncogenic rearrangements that have been connected with G4 DNA structures. G4-forming sequences are found in the *BCL-2* and *HOX11* genes near oncogenic translocation sites (Nambiar et al., 2011, 2013). And a region of the unstable *c-MYC* oncogene also adopts G4 *in vitro* (Simonsson et al., 1998). G4 formation may be common among oncogenes participating in rearrangements because previous analyses of sequences at translocation breakpoint regions revealed G4 sequences in multiple genes associated with lymphoid cancers, including *TCF3* (Katapadi et al., 2012). Results presented here demonstrate that guanine repeat sequences at two different sites of instability in the *TCF3* gene are able to fold into stable G4 structures *in vitro*, suggesting a role for this structure in promoting the genetic instability of *TCF3*. Our results provide one molecular rationale for the apparent instability of *TCF3*. Mutagenesis and recombination at *TCF3* may not be due to the sequences found at the break sites *per se*, but rather the capacity of these sequences to transform into G4 DNA conformations.

Materials and Methods

Sequence Analysis

The *TCF3*/*PBX1* t(1:19) breakpoint sequences from the Lieber database (Wiemels et al., 2002; Tsai et al., 2008) and the TICdb (Novo et al., 2007) were mapped onto *TCF3* and *PBX1* genome sequences. CNVs for *TCF3* were downloaded from the database of genomic structural variation (dbVAR) on NCBI.org (Lappalainen et al., 2013). All CNVs' breakpoints (>99 bp) were mapped to *TCF3*'s genomic location and confirmed using

Ensembl release 77 (Kersey et al., 2014). For the identification of G4 sequence motifs, *TCF3* and *PBX1* sequences were analyzed using QGRS mapper (Kikin et al., 2006) with the following filters: a maximum loop length of 45 nucleotides, minimum G group of 3, and a loop size 0–36 nucleotides. The output of that analysis was mapped to *TCF3* and *PBX1* genes. G4 sequence motifs with a QGRS score of at least 42 were presented as the number of independent G4 sequences identified in 2 kb non-overlapping windows. Guanine repeat motifs used in structure analysis were selected based on their proximity to the t(1;19) breakpoint clusters in *TCF3* and *PBX1* (T-5′, T-3′, T-3′(2), P-1, and P-2), and to breaks associated with copy number variations identified for *TCF3* (T-lg). The location of all dbSNP database insertions and deletions (Sherry et al., 2001) were mapped to *TCF3* using Enseml release 77 (Kersey et al., 2014) and density was calculated by number of insertion or deletion events per 2 kb.

G4 Folding and PAGE Analysis

G4 oligonucleotides for native PAGE and CD analysis were synthesized and PAGE purified by Operon (Huntsville, AL, USA). Sequences are shown in **Table 1** and Supplementary Table S1. For Native PAGE, oligonucleotides were 5′ end labeled using T4 PNK (New England Biolabs, Ipswich, MA, USA) and [γ-P^{32}]ATP (MP Biomedicals, Solon, OH, USA) at 37°C for 30 min. Unincorporated label was removed by Illustra Microspin G-25 spin chromatography (GE healthcare, Pittsburgh, PA, USA). G4 structures were formed in reactions containing 100 mM KCl in Tris-EDTA buffer by incubating in a small >90°C water bath that was allowed to cool slowly to room temperature. Samples were then incubated an additional hour at 37°C. Native PAGE experiments used 16% polyacrylamide (37:1) containing 0.5 X TBE with 100 mM KCl in the gel and run buffer. Oligonucleotides were resolved by electrophoresis at 100 V at room temperature for 6 h. Denaturing PAGE of radiolabeled oligonucleotides used 16% polyacrylamide (19:1) gels made with 7 M urea and 0.5X TBE. Prior to loading, samples were denatured in 90% formamide and heated to 90°C for 20 min. DNA was resolved by electrophoresis at 400 V for 1.5 h. Images were captured by phosphorimaging using a Molecular Dynamics Storm 840 phosphorimager (Amersham/GE). Each PAGE analysis was repeated at least three times and representative images are shown.

Circular Dichroism

Circular dichroism analysis was performed using an Aviv model 215 CD spectrometer at 37°C. Spectra were taken in 1 cm path quartz cells containing 12 μM G4 or GT oligonucleotide in 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, and 100 mM KCl. The molar ellipticity was measured from 220 to 300 nm and recorded for three scans in 1 nm increments at a 1 s averaging time.

Primer Extension Assays

Phagemids for extension assays were obtained by reconstituting the genomic sequence via overlapping PCR using semi-complimentary primers in a standard PCR reaction. PCR products were gel purified and TOPO cloned (Invitrogen) into pCR2.1 in both orientations and verified by sequencing. Templates for extension assays included the sequences shown in **Table 1** and were the following sizes; T-5′ (161 bp), T-3′ (472 bp), T-3′(2) (168 bp), T-lg (124 bp), P-1 (95 bp), P-2 (92 bp; Supplementary Figure S2). Closed circular single stranded DNA was obtained using M13K07 helper phage (NEB) according to the manufacturer's instructions.

Polymerase extension assays were performed essentially as described (Ehrat et al., 2012) and based on previously described assays (Weitzmann et al., 1996; Sun and Hurley, 2010). Single-stranded phagemid templates were primed with a ^{32}P 5′ end labeled M13 forward primer, which was extended with Klenow or Taq polymerase (NEB). In Klenow reactions, KCl or LiCl was added to a final concentration of 25 mM. Klenow extension reactions took place at 37°C for 8 min on single stranded template primed with 5′ end-labeled M13 forward (−20) primer. Taq reactions used identical conditions, except that temperatures ranged from 50 to 80°C for 9 min in buffers containing either $(NH_4)_2SO_4$ or KCl salt. Extension reactions were halted by the addition of an equal volume of 90% formamide and 1 mM EDTA followed by heating to 90°C for 20 min. Products of polymerase extension were resolved by 8% denaturing PAGE (19:1) with 7 M urea and 0.5X TBE, at 700 V at room temperature. Gels were then dried and images were captured by phosphorimaging with a Molecular Dynamics Storm 840 phosphorimager (Amersham/GE). Primer extension reactions were repeated three times and representative images are shown.

Gross Chromosomal Rearrangements Assay

The plasmids containing *lys2 T-G-Top* and *lys2 T-G-Btm* cassettes were constructed by inserting the 472 bp T-3′ sequence (Human genome 1:1618084-1618556) into the BglII site located +390 nt position within the *S. cerevisiae LYS2* gene in two different orientations relative to transcription start site. Using the standard two-step allele replacement protocol, these constructs were used to replace the wild type *LYS2* gene located proximal to *CAN1* on the left arm of the chromosome V. The rates of GCR occurring at the chromosome V were determined according to the previously described procedure (Yadav et al., 2014). Briefly, individual colonies were inoculated into the rich media (1% yeast extract, 2% peptone, and 2% glucose – YEPD) and cultured to saturation at 30°C. Appropriate dilution of the cultures was plated either on the YEPD media for determination of total cell numbers or on the selective media (synthetic complete media supplemented with canavanine (60 mg/l) and 5-Fluoroorotic acid (1 g/l) for determination of the cells that lost *CAN1* and *URA3* genes. For each strain, 24–32 cultures were used to calculate rate and 95% confidence levels either by Lea-Coulson method of median (*top1*Δ, high transcription; Spell and Jinks-Robertson, 2004) or by P_0 method (*wt*, high transcription, *top1*Δ low transcription; Foster, 2006).

Author Contributions

JW, experimental design, experimentation, and manuscript preparation. AB, experimental design and experimentation. SF,

experimental design, experimentation, and manuscript preparation. EL, research design and manuscript preparation. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to recognize Dr. Brad Dyniewski, Dr. Brad Johnson, and Dr. Glen Borchert for their help in identifying and cloning the *TCF3*-lg sequence. The authors recognize support from the School of Biological Sciences at Illinois State University and the National Institutes of Health, R15CA182978 to EL.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00177/abstract

References

Aspland, S. E., Bendall, H. H., and Murre, C. (2001). The role of E2A-PBX1 in leukemogenesis. *Oncogene* 20, 5708–5717. doi: 10.1038/sj.onc.1204592

Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., and Sasisekharan, V. (1992). Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* 20, 4061–4067. doi: 10.1093/nar/20.15.4061

Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Gen.* 13, 770–780. doi: 10.1038/nrg3296

Brooks, T. A., Kendrick, S., and Hurley, L. (2010). Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* 277, 3459–3469. doi: 10.1111/j.1742-4658.2010.07759.x

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* 34, 5402–5415. doi: 10.1093/nar/gkl655

Đapić, V., Abdomerović, V., Marrington, R., Peberdy, J., Rodger, A., Trent, J. O., et al. (2003). Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res.* 31, 2097–2107. doi: 10.1093/nar/gkg316

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* 18, 1618–1629. doi: 10.1101/gad.1200804

Ehrat, E. A., Johnson, B. R., Williams, J. D., Borchert, G. M., and Larson, E. D. (2012). G-quadruplex recognition activities of *E. Coli* MutS. *BMC Mol. Biol.* 13:23. doi: 10.1186/1471-2199-13-23

Foster, P. L. (2006). Methods for determinine spontaneous mutation rates. *Methods Enzym.* 409, 195–213. doi: 10.1016/s0076-6879(05)09012-9

Hamperl, S., and Cimprich, K. A. (2014). The contribution of co-transcriptional RNA: DNA hybrid structures to DNA damage and genome instability. *DNA Repair* 19, 84–94. doi: 10.1016/j.dnarep.2014.03.023

Huber, M. D., Lee, D. C., and Maizels, N. (2002). G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.* 30, 3954–3961. doi: 10.1093/nar/gkf530

Hunger, S. P. (1996). Chromosomal translocations involving the E2A gene in acute lymphoblastic leukemia: clinical features and molecular pathogenesis. *Blood* 87, 1211–1224. doi: 10.1093/nar/gkf530

Katapadi, V. K., Nambiar, M., and Raghavan, S. C. (2012). Potential G-quadruplex formation at breakpoint regions of chromosomal translocations in cancer may explain their fragility. *Genomics* 100, 72–80. doi: 10.1016/j.ygeno.2012.05.008

Kee, B. L., Quong, M. W., and Murre, C. (2000). E2A proteins: essential regulators at multiple stages of B-cell development. *Immunol. Rev.* 175, 138–149. doi: 10.1111/j.1600-065x.2000.imr017514.x

Kersey, P. J., Allen, J. E., Christensen, M., Davis, P., Falin, L. J., Grabmueller, C., et al. (2014). Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Res.* 42, D546–D552. doi: 10.1093/nar/gkt979

Kikin, O., D'Antonio, L., and Bagga, P. S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 34, W676–W682. doi: 10.1093/nar/gkl253

Kim, N., and Jinks-Robertson, S. (2011). Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair* 10, 953–960. doi: 10.1016/j.dnarep.2011.07.002

Kypr, J., Kejnovská, I., Renčiuk, D., and Vorlíčková, M. (2009). Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* 37, 1713–1725. doi: 10.1093/nar/gkp026

Lappalainen, I., Lopez, J., Skipper, L., Hefferon, T., Spalding, J. D., Garner, J., et al. (2013). DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res.* 41, D936–D941. doi: 10.1093/nar/gks1213

Larson, E. D., Duquette, M. L., Cummings, W. J., Streiff, R. J., and Maizels, N. (2005). MutSα binds to and promotes synapsis of transcriptionally activated immunoglobulin switch regions. *Curr. Biol.* 15, 470–474. doi: 10.1016/j.cub.2004.12.077

Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D., et al. (1993). High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5′ to 3′ exonuclease activity. *Genome Res.* 2, 275–287. doi: 10.1016/j.cub.2004.12.077

Lobachev, K. S., Rattray, A., and Narayanan, V. (2007). Hairpin-and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. *Front. Biosci.* 12:4208. doi: 10.2741/2381

Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M. P., et al. (2011). G-quadruplex-induced instability during leading-strand replication. *EMBO J.* 30, 4033–4046. doi: 10.1038/emboj.2011.316

Maizels, N., and Gray, L. T. (2013). The G4 genome. *PLoS Gen.* 9:e1003468. doi: 10.1371/journal.pgen.1003468

Miyazaki, K., Miyazaki, M., and Murre, C. (2014). The establishment of B versus T cell identity. *Trends Immunol.* 35, 205–210. doi: 10.1016/j.it.2014.02.009

Mo, M. L., Chen, Z., Zhou, H. M., Li, H., Hirata, T., Jablons, D. M., et al. (2013). Detection of E2A-PBX1 fusion transcripts in human non-small-cell lung cancer. *J. Exp. Clin. Can. Res.* 32, 29. doi: 10.1186/1756-9966-32-29

Nambiar, M., Goldsmith, G., Moorthy, B. T., Lieber, M. R., Joshi, M. V., Choudhary, B., et al. (2011). Formation of a G-quadruplex at the BCL2 major breakpoint region of the t (14; 18) translocation in follicular lymphoma. *Nucleic Acids Res.* 39, 936–948. doi: 10.1093/nar/gkq824

Nambiar, M., Srivastava, M., Gopalakrishnan, V., Sankaran, S. K., and Raghavan, S. C. (2013). G-quadruplex structures formed at the HOX11 breakpoint region contribute to its fragility during t (10; 14) translocation in T-cell leukemia. *Mol. Cell. Biol.* 33, 4266–4281. doi: 10.1128/mcb.00540-13

Novo, F. J., de Mendíbil, I. O., and Vizmanos, J. L. (2007). TICdb: a collection of gene-mapped translocation breakpoints in cancer. *BMC Genomics* 8:33. doi: 10.1186/1471-2164-8-33

Paeschke, K., Bochman, M. L., Garcia, P. D., Cejka, P., Friedman, K. L., Kowalczykowski, S. C., et al. (2013). Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* 497, 458–462. doi: 10.1038/nature12149

Piazza, A., Boulé, J. B., Lopes, J., Mingo, K., Largy, E., Teulade-Fichou, M. P., et al. (2010). Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 38, 4337–4348. doi: 10.1093/nar/gkq136

Piazza, A., Serero, A., Boule, J. B., Legoix-Ne, P., Lopes, J., and Nicolas, A. (2012). Stimulation of gross chromosomal rearrangements by the human

CEB1 and CEB25 minisatellites in *Saccharomyces cerevisiae* depends on G-quadruplexes or Cdc13. *PLoS Gen.* 8:e1003033. doi: 10.1371/journal.pgen.1003033

Pui, C. H., Relling, M. V., and Downing, J. R. (2004). Acute lymphoblastic leukemia. *New Eng. J. Med.* 350, 1535–1548. doi: 10.1056/nejmra023001

Rodić, N., Zampella, J. G., Cornish, T. C., Wheelan, S. J., and Burns, K. H. (2013). Translocation junctions in TCF3-PBX1 acute lymphoblastic leukemia/lymphoma cluster near transposable elements. *Mobile DNA* 4:22. doi: 10.1186/1759-8753-4-22

Roy, D., Yu, K., and Lieber, M. R. (2008). Mechanism of R-loop formation at immunoglobulin class switch regions. *Mol. Cell. Biol.* 28, 50–60. doi: 10.1128/mcb.01251-07

Schmitz, R., Young, R. M., Ceribelli, M., Jhavar, S., Xiao, W., Zhang, M., et al. (2012). Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* 490, 116–120. doi: 10.1038/nature11378

Sen, D., and Gilbert, W. (1990). A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* 344, 410–414. doi: 10.1038/344410a0

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311. doi: 10.1093/nar/29.1.308

Simonsson, T., Pecinka, P., and Kubista, M. (1998). DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* 26, 1167–1172. doi: 10.1093/nar/26.5.1167

Sollier, J., Stork, C., Garcia-Rubio, M. L., Paulsen, R. D., Aguilera, A., and Cimprich, K. A. (2014).Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* 56, 777–785. doi: 10.1016/j.molcel.2014.10.020

Spell, R. M., and Jinks-Robertson, S. (2004). "Determination of mitotic recombination rates by fluctuation analysis in *Saccharomyces cerevisiae*," in *Genetic Recombination*, ed. A. S. Waldman (Totowa, NJ: Humana Press), 3–12. doi: 10.1385/1-59259-761-0:003

Steininger, A., Möbs, M., Ullmann, R., Köchert, K., Kreher, S., Lamprecht, B., et al. (2011). Genomic loss of the putative tumor suppressor gene E2A in human lymphoma. *J. Exp. Med.* 208, 1585–1593. doi: 10.1084/jem.20101785

Sun, D., and Hurley, L. H. (2010). "Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay," in *G-Quadruplex DNA*, ed. P. Baumann (Totowa, NJ: Humana Press), 65–79. doi: 10.1007/978-1-59745-363-9-5

Tarsounas, M., and Tijsterman, M. (2013). Genomes and G-quadruplexes: for better or for worse. *J. Mol. Biol.* 425, 4782–4789. doi: 10.1016/j.jmb.2013.09.026

Tsai, A. G., Lu, H., Raghavan, S. C., Muschen, M., Hsieh, C. L., and Lieber, M. R. (2008). Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* 135, 1130–1142. doi: 10.1016/j.cell.2008.10.035

van Kregten, M., and Tijsterman, M. (2014). The repair of G-quadruplex-induced DNA damage. *Exp. Cell Res.* 329, 178–183. doi: 10.1016/j.yexcr.2014.08.038

Vorlíčková, M., Kejnovská, I., Sagi, J., Renčiuk, D., Bednářová, K., Motlová, J., et al. (2012). Circular dichroism and guanine quadruplexes. *Methods* 57, 64–75. doi: 10.1016/j.ymeth.2012.03.011

Weitzmann, M. N., Woodford, K. J., and Usdin, K. (1996). The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. *J. Biol. Chem.* 271, 20958–20964. doi: 10.1074/jbc.271.34.20958

Wiemels, J. L., Leonard, B. C., Wang, Y., Segal, M. R., Hunger, S. P., Smith, M. T., et al. (2002). Site-specific translocation and evidence of postnatal origin of the t(1; 19) E2A-PBX1 fusion in childhood acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15101–15106. doi: 10.1073/pnas.222481199

Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* 59, 871–880. doi: 10.1016/0092-8674(89)90610-7

Wu, Y., and Brosh, R. M. (2010). G-quadruplex nucleic acids and human disease. *FEBS J.* 277, 3470–3488. doi: 10.1111/j.1742-4658.2010.07760.x

Yadav, P., Harcy, V., Argueso, J. L., Dominska, M., Jinks-Robertson, S., and Kim, N. (2014). Topoisomerase I plays a critical role in suppressing genome instability at a highly transcribed G-quadruplex-forming sequence. *PLoS Gen.* 10:e1004839. doi: 10.1371/journal.pgen.1004839

Zhao, J., Bacolla, A., Wang, G., and Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* 67, 43–62. doi: 10.1007/s00018-009-0131-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Williams, Fleetwood, Berroyer, Kim and Larson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.