Genome-Wide Formation of DNA:RNA Hybrid G-Quadruplexes in Live Yeast Cells

Chen-xia Ren[a,c], Rui-fang Duan[a,c], Yu-hua Hao[b], Zheng Tan[a,b,1]

[a]Center for Healthy Aging, Central Laboratory, Changzhi Medical College, Changzhi 046000, Shanxi, P. R. China

[b]State Key Laboratory of Membrane Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, P. R. China

[c]Contributed equally.

[1]Address correspondence to Zheng Tan, E-mail: z.tan@ioz.ac.cn

Abstract

Four-stranded G-quadruplex (G4) structures formed by guanine-rich nucleic acids play a role in many essential physiological processes. To date, studies have focused almost exclusively on monomeric intramolecular G4s with three or more G-tetrads. Based on the analysis of Okazaki fragment syntheses and G4 probing, we report here a massive genome-wide formation of a hybrid type of G4s (hG4s), composed of G-tracts from both DNA and RNA, in the genome of live yeast cells. We found that hG4s can form with as few as one G-tract in DNA, increasing the total number of G4 sites from 38 to 127841. Furthermore, we also effectively detected hG4s of two G-tetrads, bringing the total number of G4 sites to a final of 587694, an increase of >15,000-fold. We even found that hG4s dominate in DNAs capable of forming the canonical intramolecular DNA G4s by themselves. Taken together, our results reveal a previously unrecognized rule governing G4 formation in eukaryotic genomes that makes hG4s the overwhelmingly dominant G4 species. They are present in all genes, with greater structural diversity and a wider dynamic range. More importantly, the involvement of RNA gives hG4s the ability to function in a transcription-dependent manner. These new findings thus emphasize a complete renewal and expansion of our current perspective on the existence and functional role of G4s in eukaryotic cells.

Keywords: DNA:RNA hybrid G-quadruplex, Okazaki fragment, genome, eukaryotic cell, transcription, replication

Significance

Our work reveals a massive and genome-wide formation of a virtually unnoticed, but in reality overwhelmingly dominant type of G-quadruplexes (G4s), i.e. DNA:RNA hybrid G-quadruplexes (hG4s), in living yeast cells. The involvement of RNA confers several unique properties to hG4s: 1) it allows genomic DNA to form G4s with as few as one G-tract, resulting in complete hG4 dominance over canonical monomeric DNA G4s (dG4s) and their ubiquitous presence in genes; 2) it increases the structural diversity of G4s by offering different combinations of DNA and RNA G-tracts; and 3) it couples G4 formation to transcription, such that hG4s function in a transcription-dependent manner. The discovery of hG4s reveals the physiological reality of the prevalence and functionality of G4s in the context of transcription.

Introduction

Guanine-rich nucleic acids can form four-stranded G-quadruplex (G4) structures by combining four guanine-tracts (G-tracts) into a stack of guanine-tetrads (G-tetrads) (Figure 1A). Because of their topological diversity, physical stability, and unique genomic location, G4s are involved in a variety of cellular activities, ranging from replication, transcription, and translation to genome instability (1,2). Putative G-quadruplex sequences (PQSs) are widely distributed in the genomes of both prokaryotic and eukaryotic organisms. They are not randomly distributed but are highly enriched in regulatory regions, especially in promoters of higher species, suggesting their role in the regulation of gene expression (3,4).

In principle, the four G-tracts of a G4 can be supplied by a single or multiple nucleic acid strands, forming either an intramolecular or an intermolecular architecture. To date, the canonical intramolecular DNA or RNA G4s formed by four G-tracts from a single strand (Figure 1A) have been the exclusive focus of G4 biology (see a recent review (4)). Recently, G4s have been detected in the genomes of living animal cells (5). Ten years ago, we found in our *in vitro* studies that a multimeric DNA:RNA hybrid type of G4s (hG4s) (Figure 1, B-C) could form in transcribed DNA by simultaneous recruitment of G-tracts from a non-template DNA strand and RNA transcripts, respectively (6,7). In these and our subsequent studies (6-15), we characterized hG4s in terms of their formation mechanism, kinetics, stability, structural diversity, competition with DNA G4s (dG4s), and other features. We found that even though a transcribed DNA can form a dG4 by itself, hG4s can still dominate (11), which is explained by *in vitro* studies from other groups and ours showing that hG4s are more stable and fold faster than the corresponding dG4s (15-19). However, we had only proven their formation in living bacterial cells (11); it is still not known whether such hG4s can be formed in eukaryotic cells.

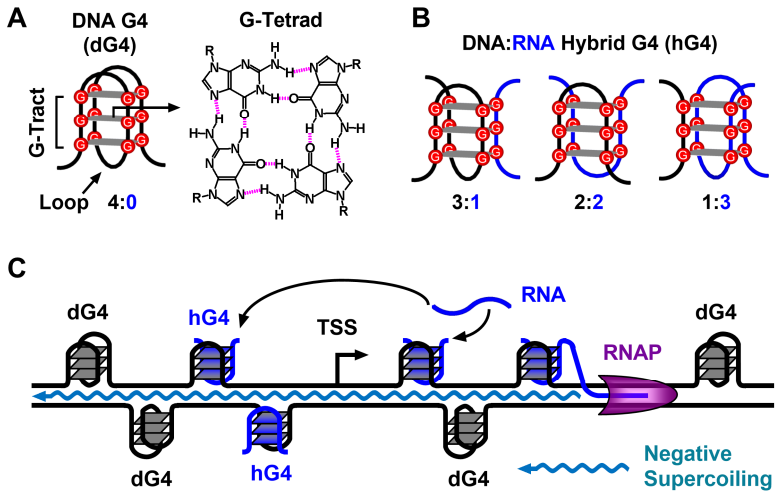

Figure 1. Intramolecular and intermolecular G4 structures and their formation in a DNA duplex transcribed *in vitro*. (A) An intramolecular DNA G4 composed of four G-tracts by stacking of G-tetrads. (B) DNA:RNA hybrid G4s that can form in transcribed DNA with the joint participation of G-tracts from DNA and RNA. (C) G4 formation can be triggered by a transcribing RNA polymerase (RNAP) on its upstream side by negative supercoiling and on its downstream side by a transcriptional bubble. See the text for more information.

Because of the involvement of G-tracts from RNA, DNA can theoretically form an hG4 with as few as one G-tract (Figure 1B). Thus, the hG4 PQS is far more abundant than the dG4 PQS. For example, the yeast genome has >3,300 times more hG4 PQS than dG4 PQS, if we only count those that can form a G4 with three or more G-tetrads. More importantly, hG4 formation is proportional to transcriptional activity according to our *in vitro* studies (12,20). Thus, hG4-mediated regulation is coupled to transcription. Collectively, the hG4s should represent a distinct genotype that is drastically more abundant and capable of playing a role in close association with transcription. However, the uncertainty about their existence has prevented researchers from exploring and understanding this unique genotype in the eukaryotic kingdoms. Therefore, an answer to the question of whether hG4s can form in eukaryotic cells would reveal the *in vivo* reality of a far more prevalent genotype as well as a universal regulatory mechanism that has largely gone unnoticed. Here, we present evidence for the genome-wide formation of hG4s in yeast cells and discuss their unique physiological implications.

Materials and Methods

Polymerase and exonuclease stop assay

The two assays were performed with the DNAs listed in Table S1 as previously described using Bsu DNA Polymerase, Large Fragment (NEB, USA) (21) and Exonuclease Exo I (NEB, USA) (22).

Data from public repositories

Yeast chromosome sequences (sacCer3) and bed files of yeast SGD genes were downloaded from the UCSC Genome Browser (http://genome-asia.ucsc.edu). Five paired-end sequencing datasets of Okazaki fragments (OKFs) in bed or bedpe format (23-27) were downloaded from the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/) at the National Center for Biotechnology Information (NCBI) website. Data in bedpe format were converted to bed format using a custom Perl script adapted from bedpe2bed.pl, available at https://github.com/srinivasramachandran/sam2bed/blob/main/bedpe2bed.pl.

Identification of PQS and control motifs in genomes

Chromosome sequences were searched for PQS with one or more G-tracts using the regular expression G{n,}(.{1,7}?G{n,}){0,}, where n was 3 or 2, as described (8). The resulting PQSs were grouped into multiple bed files according to the number of G-tracts, and each bed file was further split into two bed files containing the PQSs on the forward and reverse strands, respectively. PQSs able to for G4s of only two G-tetrad layers were first identified using the regular expression G{2,}(.{1,7}?G{2,}){0,}, and those that matched G{3,} were discarded. Motifs containing one or more $A_{\geq 3}$ tracts were identified using the regular expression A{3,}(.{1,7}?A{3,}){0,}. Motifs containing one or more GNG tracts, where N denotes any non-G nucleotide, were identified using the regular expression [^G]{7}G[^G]G(.{1,7}? G[^G]G){0,} [^G]{7}.

Identification of Orphan PQS

PQS Bed files were processed by the Bedtools merge software (28) with the following arguments: -d n -c 4,5,6 -o count,collapse,distinct, where “-d n” indicated the maximum distance n between

PQSs to be merged. Unmerged PQSs flagged with a count of one were collected as orphaned PQSs.

Survey of PQS load in genes

Bed files of genes were expanded by 120 bp on the upstream side to cover the promoters (29,30) using the Bedtools slop software (28). PQSs overlapping with the expanded gene regions by at least one bp were counted using Bedtools map software (28).

Distribution of Okazaki fragments (OKFs) across PQSs

Bed files of OKFs were converted to bedgraph format using the Bedtools genomeCoverageBed software (28) to represent the full-length (-bg -strand +/-), the 3' (-bg -3 -strand +/-), or 5' (-bg -5 -strand +/-) ends of the OKFs, respectively. The resulting bedgraph files were then converted to bigwig format using the Bedtools bedGraphToBigWig software (28) with a bin size of 10 nucleotides (nts) unless otherwise indicated. The distribution of OKFs and their 3' or 5' ends across the 3' end of PQSs was profiled using the computeMatrix and plotProfile tools of the Deeptools software (31) using the bigwig files and the relevant PQS bed files. The final profiling result of each dataset represented the average of all samples in each study.

To profile the 5' ends of those OKFs whose 3' ends fell within ±20 nts of the 3' ends of PQSs, a new bed file was built from the original with the start and end coordinates of the PQSs appended as two additional columns to the new bed file. The new bed file was filtered using the Bedtools intersect software (28) to pool those 3' ends within ±20 nts of the 3' ends of the PQSs. The two columns representing the start and end of the PQSs were then restored to columns 2 and 3, respectively, and further processed as described in the previous section.

Plasmid construction for G4P-ChIP

The coding sequence of the G4 probe (G4P) protein was amplified from the pNLS-G4P-IRES2-EGFP (5) using a PCR primer pair of 5'-ATTAAGCTTATGCCCAAGAAGAAGCGGAAG-3'and 5'-GTGGATCCTTACTTGTCATCGTCATCCTT-3'and inserted into the pYES2 plasmid between the Hind III and BamH I sites.

G4P-Seq

The recombinant plasmid pYES2-G4P-3xFLAG was transformed into *S. cerevisiae* strain BY4741 using the lithium acetate/PEG transformation method (32). Recombinant colonies were selected after 72 hrs of incubation at 30 °C on SC-Ura (synthetic drop-out agar medium without uraci). Strains expressing G4P were cultured in SD-Ura (synthetic complete medium without uracil supplemented with 2% (w/v) glucose) liquid medium at 30 °C until the OD_{600} reached 3 and harvested by centrifugation followed by three washes in SG-Ura (synthetic complete medium without uracil supplemented with 2% (w/v) galactose) liquid medium. Cell pellets were suspended in liquid SG-Ura to an OD_{600} of 0.6 and incubated at 30 °C for GAL1 promoter induction.

After induction for 6 hrs, cells were processed for ChIP-seq by adapting the protocol as previously described (33), except that the lysis was performed using a Tissuelyser-48L for 9 cycles of 1 min on and 3 min off (samples were kept on ice during the off cycle). The lysate was

pelleted as described, resuspended in 600 μl of lysis buffer, and sonicated in a Covaris M220 ultrasonicator for 20 min (10% duty, 75 W power, 200 cycles/burst, 7 °C).

For sequencing, 1% of the lysate was saved as input, and the remainder was incubated with 50 μl of anti-FLAG M2 magnetic beads (Sigma-Aldrich) at 4 °C overnight and further processed as described. 2×150 bp paired-end reads were generated by sequencing the input DNA and DNA recognized by G4P on an Illumina HiSeq (Genewiz, Suzhou, China).

Distribution of G4P across PQS

Profiling of the distribution of G4P binding across regions of interest was carried out as describe (5).

Brief background of G4 formation *in vitro*

Our group has so far performed most of the studies (6-15) on hG4s (6-19), including their discovery, identification, mechanism of formation, effect on transcription, etc. A co-transcriptional formation of G4s is dynamic and can be driven by different mechanisms. On the downstream side of a moving RNA polymerase (RNAP), it is triggered in the non-template strand when an approaching transcription bubble is 7 nucleotides (nts) away from a PQS (34) (Figure 1C). A G4 can also form behind a moving RNAP, driven by a negative supercoiling wave (20,35), which can propagate upward and induce G4 formation over a range of up to several thousand nts (35) (Figure 1C). For a hG4, it can form during the second round of transcription, when the RNA on the template strand produced in the previous round of transcription is displaced to supply the necessary G-tracts (9). An hG4 can also form between G-tracts in DNA on the upstream side of transcription and G-tracts in RNA produced by transcription on the downstream side (13). Compared to a corresponding dG4, the formation of an hG4 was favored by a faster folding rate and greater mechanical stability (15), such that hG4s can still dominate, even when a DNA itself is capable of forming a dG4 (11). Taken together, these findings suggest that hG4s can form regardless of where G-tracts are located in a genome, as long as RNAs with sufficient G-tracts are provided from the bulk RNA pool in a cell.

Results

G4 detection based on stop of Okazaki fragment (OKF) synthesis

Protein translocation along a DNA strand is physically hindered by a G4 (36). For this reason, a polymerase stop assay (21) has been developed and widely used for the *in vitro* detection of G4s in both DNA (21,37,38) and RNA (39,40). In this method, a short DNA primer is annealed to a target strand, followed by an extension with DNA polymerase. The extension stops when it encounters a G4 to signal the presence of the G4. As shown in Figure 2 (A and B), a yeast or human DNA G4 could halt the progress of both DNA polymerase and exonuclease. In contrast, both reactions reached their full length when the two dG4 PQSs were mutated to prevent G4 formation. In Figure 2C and 2D, the DNA template and substrate both contained seven G-tracts, allowing only one dG4 to form, but at four alternative positions. In this case, four DNA bands representing one dG4 at four different positions were detected in both assays. This feature was used to determine the G-tracts involved in G4 formation (11,22).

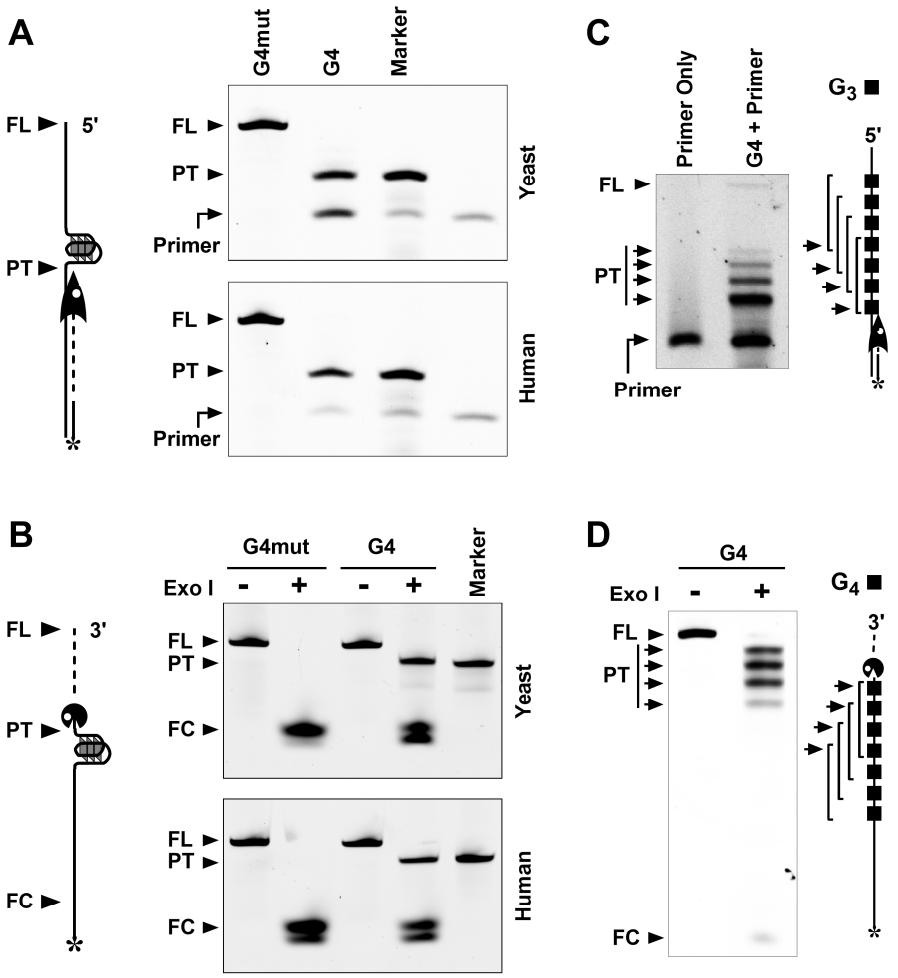

Figure 2. Stop of DNA polymerase and exonuclease by an intramolecular dG4. (A) DNA polymerase stop assay showing premature termination (PT) by a G4 and full-length (FL) extension with a G4 mutant. (B) DNA exonuclease stop assay showing premature termination (PT) of DNA cleavage by a G4 and full length cleavage (FC) with a G4 mutant. The G4 PQS was from the promoter of the human c-KIT gene and from between the yeast ATO3 and EFT2 genes, respectively. (C) DNA polymerase stop assay showing premature termination (PT) by a G4 that could form at multiple positions (brackets). (D) DNA exonuclease stop assay showing premature termination (PT) of DNA cleavage by a G4 that could form at multiple positions (brackets).

To detect G4 formation in live yeast cells, we adopted the *in vitro* polymerase stop principle by monitoring the progression of Okizaki fragment (OKF) syntheses that are primed by RNA rather than DNA. OKFs are short DNA fragments generated in a replication fork on the lagging strand during DNA replication (41). Over the past few years, OKF syntheses in yeast have been investigated for other topics using the OK-seq technique in several independent studies, which provided the OKF data for this work.

Retardation of OKF syntheses by PQSs

We first identified the PQSs in the yeast genome and classified them into four groups, 4G3+, 3G3+, 2G3+, and 1G3+, based on the number of $G_{\geq 3}$ tracts. While the 4G3+ PQSs are capable of forming dG4s, the 3G3+, 2G3+, and 1G3+ PQSs can form hG4s by recruiting additional G-tracts from RNA transcripts, as shown in Figure 1B. We then profiled the distribution of OKFs relative to the 3' ends of each of the four groups of PQSs using the OK-seq data obtained from the GEO database (Figure 3A). Regardless of the type of G4 the PQSs could form, all the profiles showed a positive peak downstream of the PQS on the PQS-bearing strand (Figure 3, B-E, red lines). The positive peaks all showed an abrupt drop at the 3' end of the PQSs, suggesting that dG4s may have formed in the 4G3+ (Figure 3B) and hG4s in the 3G3+, 2G3+, and 1G3+ PQSs (Figure 3, C-E), respectively, halting the progression of the OKF synthesis according to the principle of the polymerase stop assay.

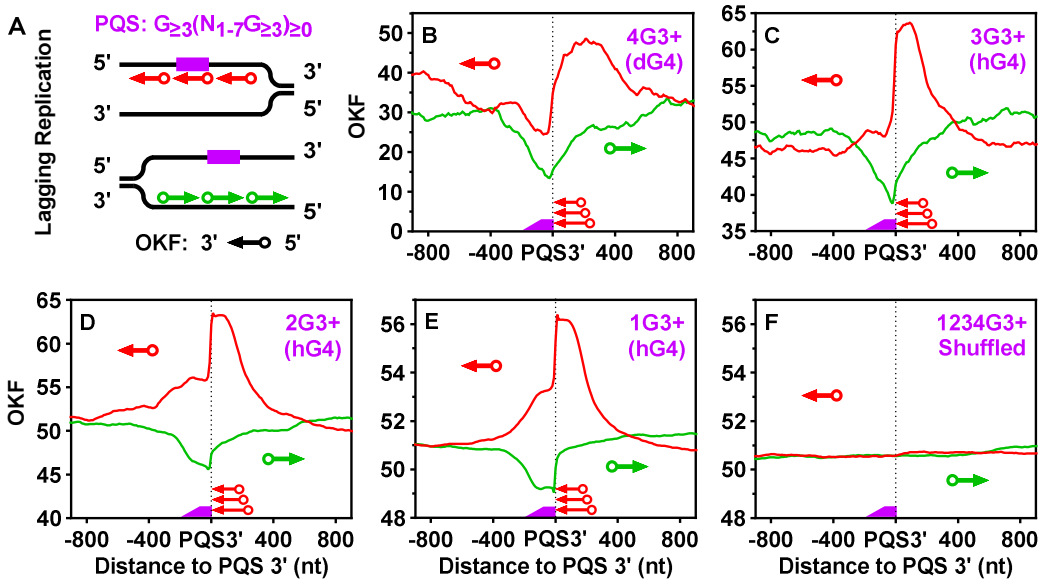

Figure 3. Distribution of Okazaki fragments (OKFs) across the 3' end of PQSs with 4 to 1 $G_{\geq 3}$ tracts that could form either dG4s or hG4s of three or more G-tetrads. (A) Scheme of OKF profiling on PQS-bearing strands and strands opposite to the PQS. The two color codes are used in this and the remaining figures to represent the corresponding OKFs. (B) OKFs across the 3' end of 4G3+ PQSs capable of forming intramolecular dG4s. (C-E) OKFs across the 3' end of 3G3+, 2G3+, and 1G3+ PQSs capable of forming only hG4s. (F) OKFs across the 3' end of randomly shuffled total PQSs. Data for this and the remaining figures in the main text are from GSE115897 (23) unless otherwise indicated. The "3+" in panels B-F indicates that the G-tracts in the PQS had three or more consecutive guanines; the number(s) before the "G" indicate the number of G-tracts in the PQS; text in parentheses indicates the type of G4 the PQS can form. Bin size: 10 nts.

In contrast to the PQSs, the complementary cytosine-rich (C-rich) motifs were all associated with a negative OKF peak (Figure 3, B-E, green lines), since they did not form G4s, while their

flanking regions had scattered background G4 formation. The two oppositely polarized OKF peaks were specifically associated with PQSs because they both disappeared when the coordinates of the PQSs were randomly shuffled across the genome and the OKF signal was reprofiled over the resulting fake PQSs (Figure 3F). Similar profiles were also observed in four other independent datasets from the GEO database (Figure S1-S4), indicating that the accumulation of OKF synthesis at the 3' end of PQSs is a universal phenomenon.

To further evaluate the G-tract-dependence, we performed the same analysis with motifs in which the guanines were replaced by adenines. In this case, a negative peak was observed on both strands for the OKF (Figure S5), as expected, because these A/T-rich motifs were both unable to form a G4, while their flanks could. A mutation in the middle of a G-tract is an effective way to disrupt G4 formation. So we also tested motifs containing one or more GNG tracts, where N denotes any nucleotide other than G. These motifs are also unable to form a G4 either and therefore also showed a negative OKF peak on both DNA strands (Figure S6). Shuffling these two types of motifs on the genome both resulted in the disappearance of the negative peak, as expected (Figure S5F and Figure S6F). Collectively, the results supported the formation of hG4s in the G-tract-containing motifs (Figure 3, Figure S1-S4).

Verification of G4 formation with a G4 probe (G4P) protein

The G-tract-dependent arrest of OKF synthesis at the PQSs suggested the formation of G4s in the PQSs. To verify this, we next expressed a small G4 probe (G4P) protein in yeast and performed G4P-seq analysis (Figure 4), as we have done for human and other animal cells (5). Carrying two G4-binding domains and with a small molecular weight of only 6.7 kDa (Figure 4A), the G4P is highly specific for G4 recognition. Consistent with the OKF signal (Figure 3), the G4P showed enrichment at the PQSs, which could form either dG4s (Figure 4B) or hG4s (Figure 4, C-E). Similarly, the enrichment disappeared when the coordinates of the PQSs were randomly shuffled over the genome and the G4P signal was reprofiled (Figure 4F). We also analyzed the G4P signal across the A/T-rich and GNG motifs, respectively. In accordance with the OKF signals around these motifs (Figure S5 and Figure S6), the distribution of G4P binding all featured a negative peak (Figure S7 and Figure S8), further confirming the formation of hG4s and dG4s in the PQSs.

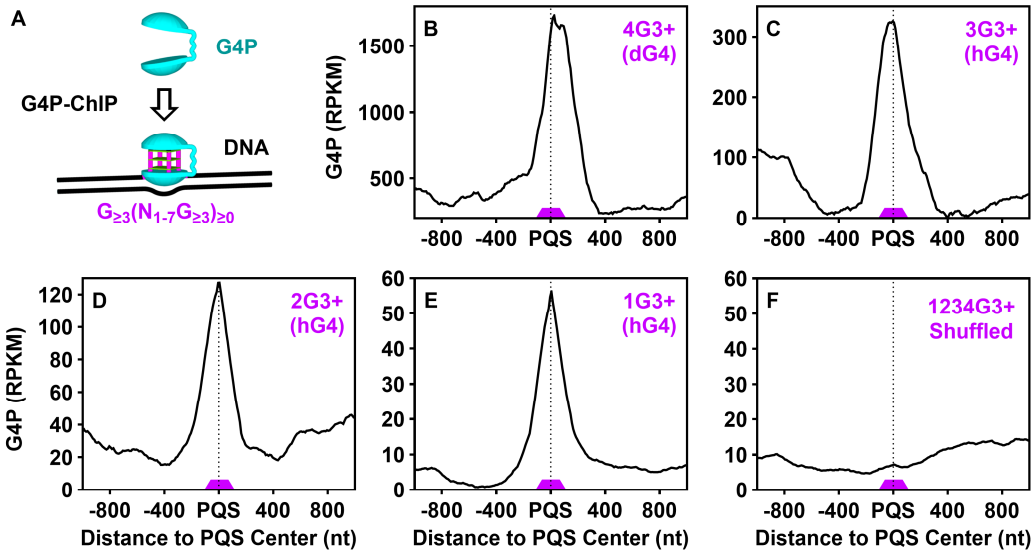

Figure 4. Detection of G4 by G4P-ChIP in PQSs with 4 to 1 $G_{\geq3}$ tracts that could form either dG4s or hG4s of three or more G-tetrads. PQSs are the same as in Figure 3.

Formation of hG4s and dG4s of two G-tetrad layers

While stable G4s of three or more G-tetrads have been the focus of attention, we have previously observed the formation of hG4s of only two G-tetrads in DNA duplexes transcribed *in vitro* (6). Such hG4s are much less stable than those with three or more G-tetrads (6,12). We were curious to see if such hG4s would also form in living cells. To this end, we performed analyses of the OKF and G4P signals around such PQSs with only GG tracts. The results also displayed a piling up of OKF (Figure 5) and an enrichment of G4P (Figure 6), proving the formation of such hG4s and dG4s in the genome. Based on this finding, our subsequent analysis of G4s would include those with two G-tetrads.

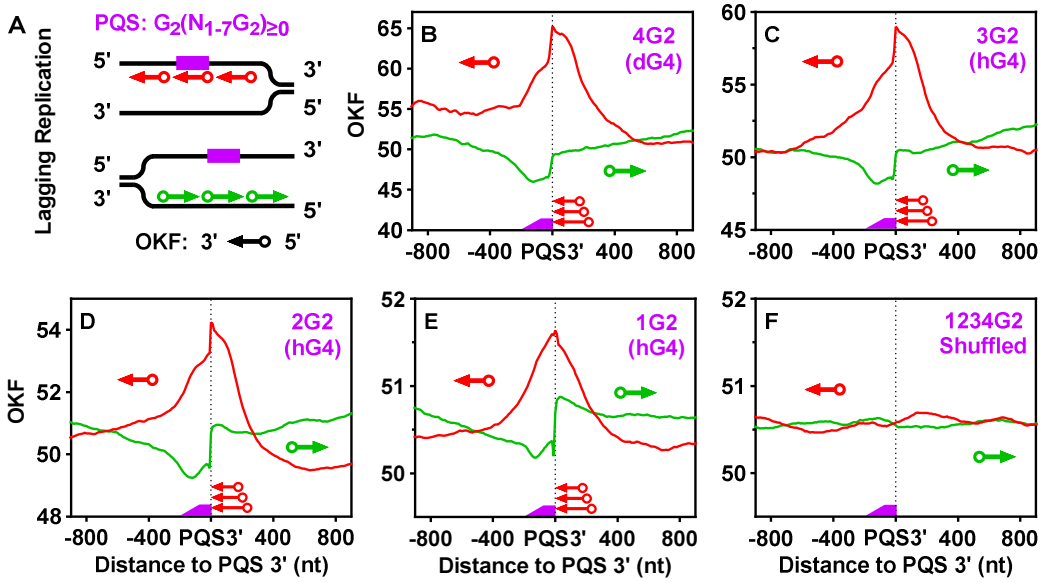

Figure 5. Distribution of OKFs across the 3' end of PQS with 4 to 1 G_2 tracts that could form either dG4s or hG4s of only two G-tetrads. (A) Scheme of OKF profiling on PQS-bearing strands and strands opposite to the PQS. (B) OKFs across the 3' end of 4G2 PQSs capable of forming intramolecular dG4s. (C-E) OKFs across the 3' end of 3G2, 2G2, and 1G2 PQSs capable of forming hG4s only. (F) OKFs across the 3' end of randomly shuffled 1234G2 PQSs.

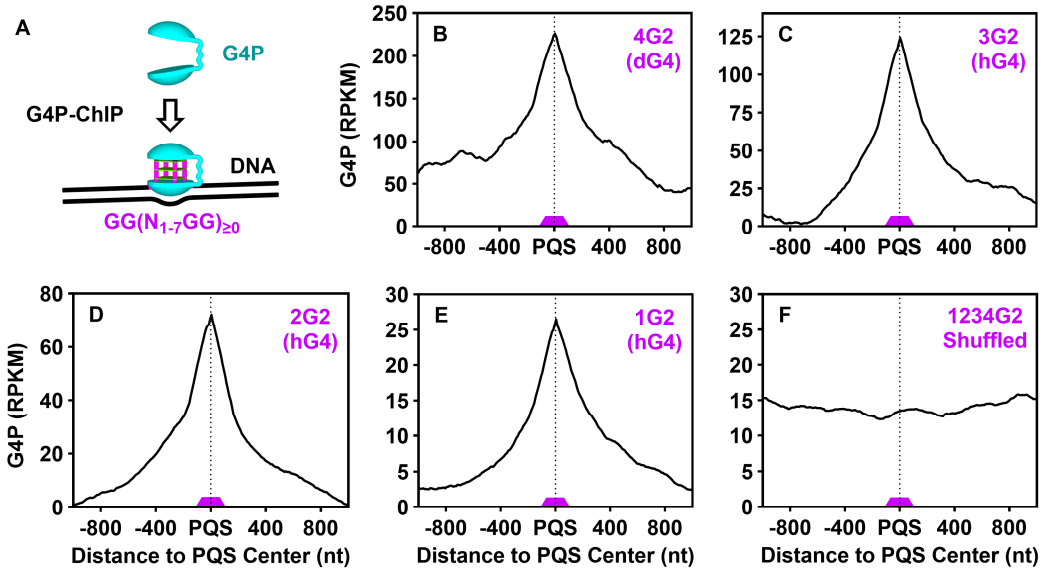

Figure 6. Detection of G4 by G4P-ChIP in PQSs with 4 to 1 G_2 tracts that could form either dG4s or hG4s of only two G-tetrads. PQSs are the same as in Figure 5.

Alignment of the 3' end of OKFs at the G4s

The OKFs are of different lengths, which is also a source of variation that contributes to the OKF profiles. To determine the exact stop position of the OKFs, we extracted the coordinates of the 3' ends of the OKFs and plotted their distribution across the PQSs (Figure 7A, Figure 8A, arrowhead). A sharp peak was seen at the 3' end of the PQSs, regardless of whether the PQSs were able to form dG4 (Figure 7B and Figure 8B) or hG4 of only two (Figure 8, C-E) or more (Figure 7, C-E) G-tetrads. Here, the sharp peak should represent an interception of OKF synthesis at the G4s. To prove this, we collected those OKFs whose 3' ends were within ±20 nts of the 3' end of the PQSs and profiled them together with their 5' ends. The results (Figure 9) showed that the 5' ends of these OKFs were distributed over a wide range of ~500 nts downstream of the PQSs. This fact indicated that the syntheses of these OKFs were initiated over a broad range and then intercepted by the G4s in the PQSs. On the other hand, the sharp peak of the OKF 3'-end with an almost flat background makes it a much clearer indicator of G4 formation.

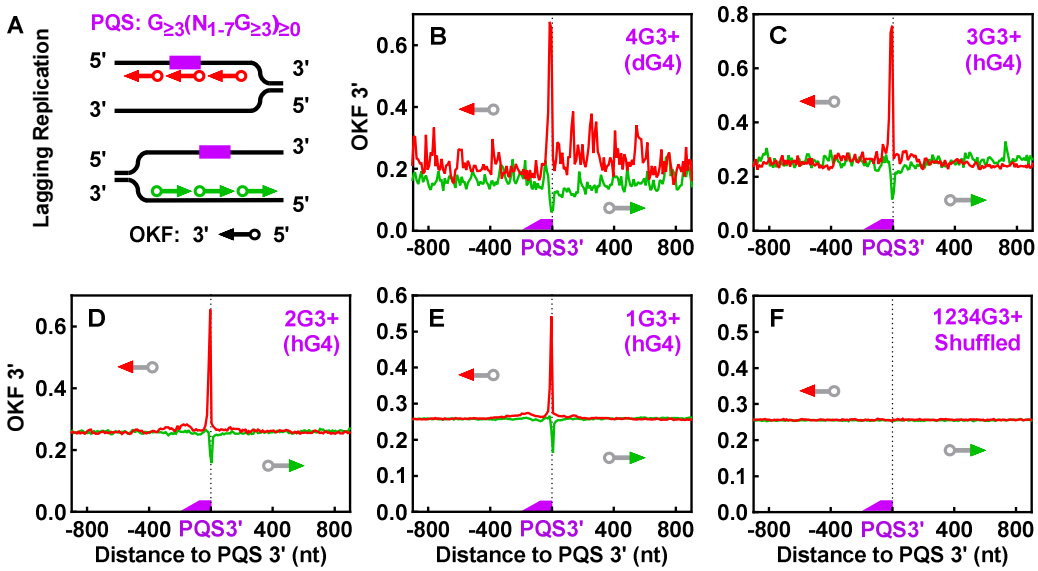

Figure 7. Distribution of the 3' end (A, arrowhead) of OKFs across the 3' end of PQSs that could form either dG4s (B) or hG4s (C-E) of three or more G-tetrads. (F) Distribution across the 3' end of randomly shuffled 1234G3+ PQSs. PQSs are the same as in Figure 3.

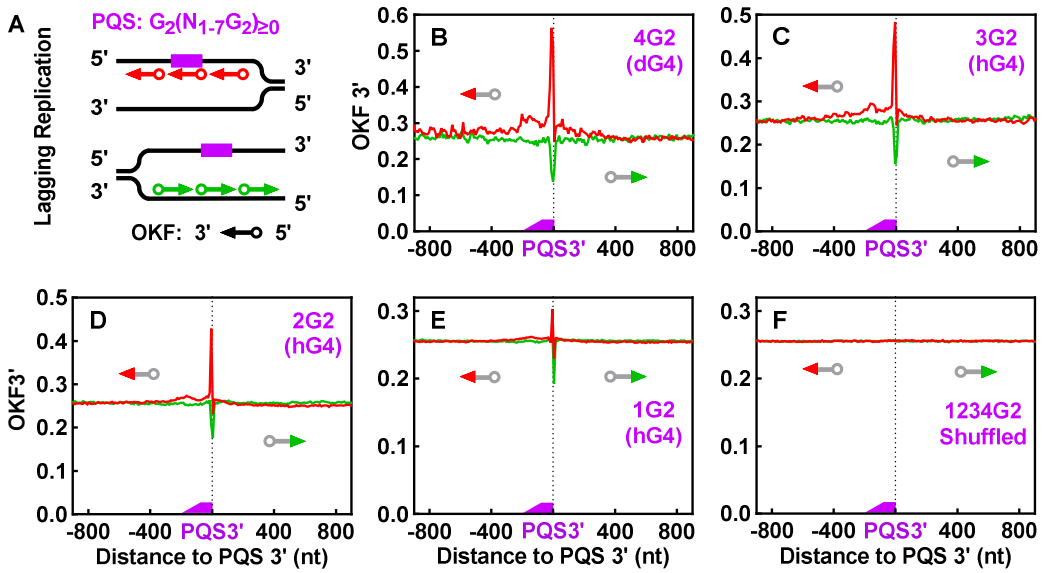

Figure 8. Distribution of the 3' end (A, arrowhead) of OKFs across the 3' end of PQSs that could form either dG4s (B) or hG4s (C-E) of only two G-tetrads. (F) Distribution across the 3' end of randomly shuffled 1234G2 PQSs. PQSs are the same as in Figure 5.

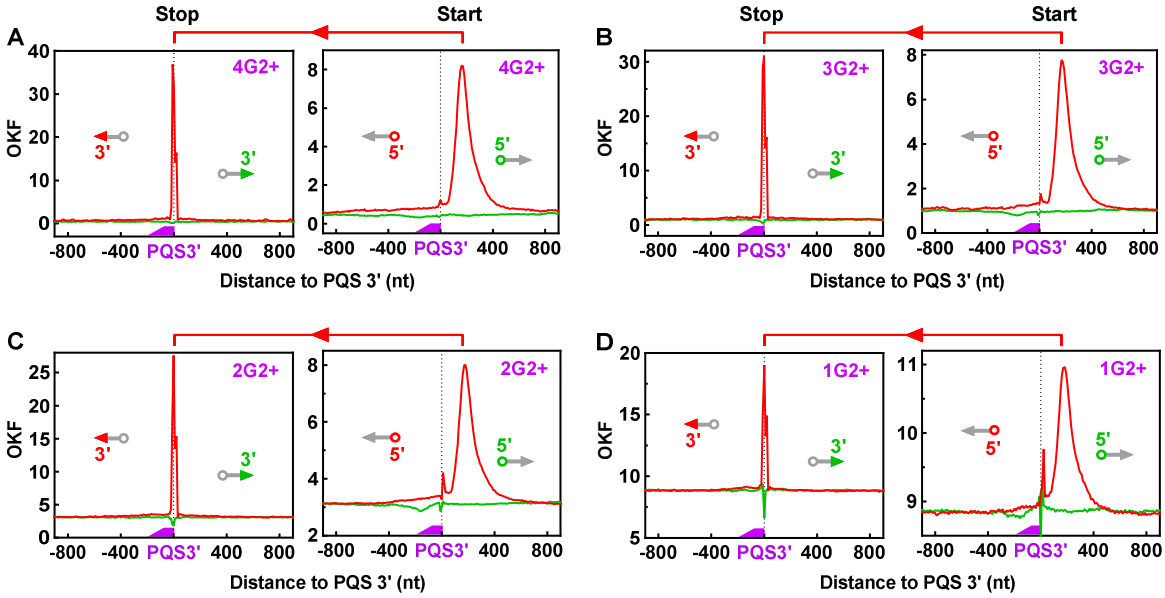

Figure 9. Interception of OKF synthesis at the 3' end of PQSs that could form either (A) dG4s or (B-D) hG4s of two or more G-tetrads. OKFs whose 3' ends were within the 3' end ± 20 nts of the PQSs were collected, and the distribution of their 3' and 5' ends was plotted across the 3' end of the PQS, respectively. Bin size: 5 nts.

Distinguishing hG4 from dG4

In the polymerase stop assay, the number of stop sites is determined by the number of G4s formed with the available G-tracts (11) (Figure 2C). For a single DNA G-tract, only one hG4 can be formed with three RNA G-tracts, resulting in one stop for polymerase extension. For a DNA PQS with four G-tracts, only a single stop is induced if only one dG4 forms because it consumes all four G-tracts (Figure 2A). However, when hG4s form, a maximum of four stops can be generated, with each hG4 involving one G-tract from DNA and three from RNA. This principle has allowed us to distinguish hG4 from dG4 in transcribed DNA *in vitro* and in bacteria (11). Here, we used the same principle to further identify the G4s based on the number of stop signals in OKF synthesis.

In Figure 10, the extension of the OKFs was traced at 1-nt resolution, spanning the 80-nt range, centered on the PQSs carrying 1-4 GG tracts, respectively, each with loops of identical size. For the PQSs with a single GG tract, a single major OKF peak was detected, consistent with the formation of a single hG4 composed of one GG tract from DNA and three from RNA (Figure 10A). In Figure 10B, we listed the representative combinations of G-tracts from DNA and RNA in the formation of different forms of G4s and the resulting stops of OKF synthesis. In Figure 10C, OKF syntheses across the PQSs with two GG tracts resulted in two major peaks that became clearer when the loop was increased from 1 to 7 nts for better resolution, which was consistent with the formation of two hG4s composed of one GG tract from DNA and three from RNA. The same rule extended to the PQSs with three GG tracts (Figure 10D), for which three OKF peaks were detected accordingly.

PQSs with four GG tracts were expected to result in one OKF peak if all the four GG tracts were used to form a dG4. Interestingly, however, these PQSs also showed the number of OKF peaks equal to the number of GG tracts (Figure 10E). This behavior was confirmed in the other four datasets from the GEO database (Figure 11), indicating that a DNR:RNA G-tract combination of 4:12 was adopted to form four hG4s (Figure 10B, bottom scheme), although we could not exclude other combinations shown in Figure 10B (bottom schemes). Consistent results were also obtained with the GGG (Figure S9) and GGGG (Figure S10) tracts, respectively.

Taken together, our analyses in this section indicate, first, that even a single GG tract in the yeast genome can form an hG4 and, second, that a PQS with four G tracts can still form hG4s, even though it can form an intramolecular dG4 merely by itself. Judging from the well-resolved four peaks, the hG4s in Figure 10E and Figure 11 could be the major form of G4. If not, a major OKF peak would be expected at the first G-tract from the downstream side of the PQSs. A PQS in a chromosome is constrained in a DNA duplex, where the folding of G-tracts should be more difficult than those in a more flexible RNA strand. The dominance of hG4 in such a case was observed in our previous study using linear duplex and plasmid DNA transcribed *in vitro* with T7 RNA polymerase (11,15), which was explained by the observation that hG4 formation is favored by faster kinetics and greater stability (15).

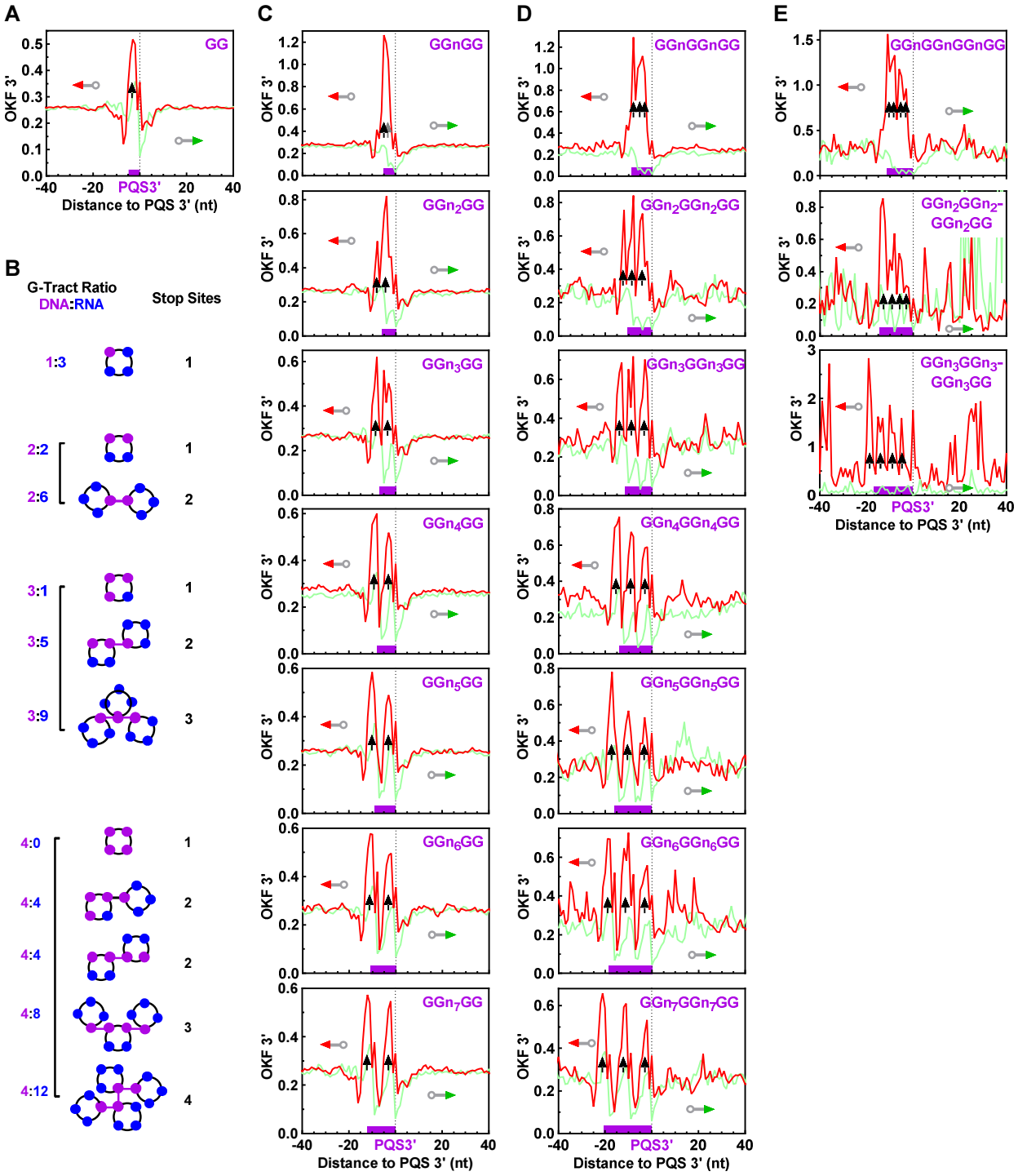

Figure 10. OKF-3' profiling showing the involvement of DNA G-tracts in hG4 formation in PQSs with (A) one, (C) two, (D) three, or (E) four GG tracts. (B) Representative combinations of DNA and RNA G-tracts in G4 formation. "n" denotes any nucleotide other than guanine. Bin size: 1 nt.

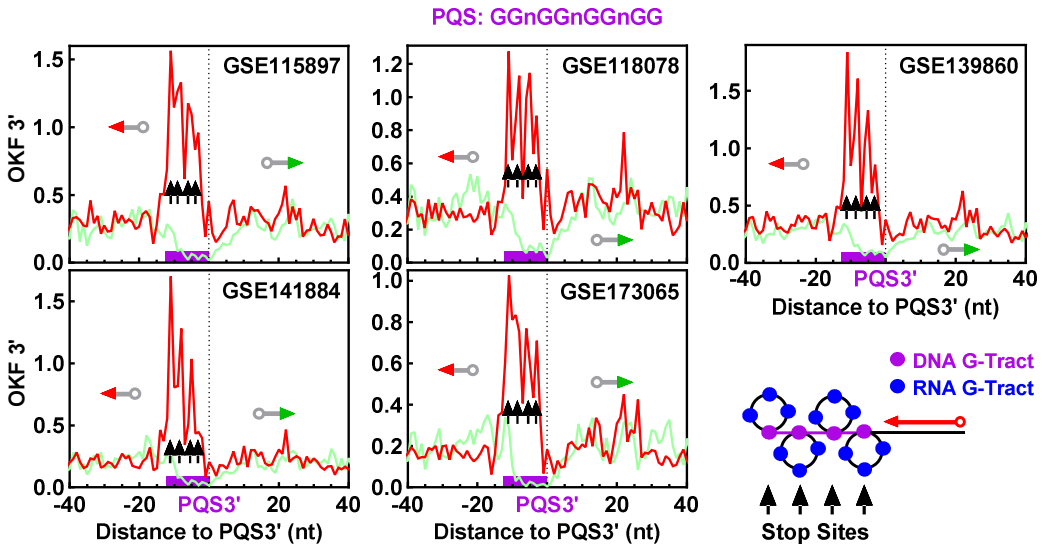

Figure 11. OKF-3' profiling showing the involvement of DNA G-tracts in the formation of hG4s in the PQSs containing four GG tracts and 1-nt loops. The source of the OKF data is indicated in the panels. Bin size: 1 nt.

Formation of hG4 in orphan PQSs

We still had one concern at this point, since the PQS was defined by a consensus that limits the loop between G-tracts to a maximum of seven nts, a general rule adopted by the G4 community. When considering a PQS with fewer than four G-tracts, it is possible, although very unlikely, that neighboring G-tracts beyond the loop limit could be involved to form a dG4 rather than a hG4. With this concern, we identified three groups of PQS carrying a single run of GG, GGG, and GGGG, respectively, separated from all other G-tracts of two or more guanines by a defined number of GG-free flanks ranging from 20 to 100 nts.

These G-tracts, termed "orphan G-tracts", were analyzed for the OKF stop signal and G4P binding, respectively (Figure 12). For all these G-tracts, a single major OKF-3' peak was detected as an indication of G4 formation. In particular, the peak magnitude did not change as the isolation flanks increased from 20 to 100 nts (Figure 12A, panels from left to right), ruling out the possibility of dG4 formation. In contrast, the peak magnitude increased as the G-tract size increased from 2 to 4 nts (Figure 12A, panels from top to bottom), consistent with the well-documented fact that G4s with longer G-tracts are more stable (42). The formation of hG4s was further confirmed by the enrichment of G4P at these orphan G-tracts (Figure 12B). Like other ChIP-seq techniques, G4P-seq has a resolution of a few hundred nucleotides. Even at this level, a single G4P peak was still visible for most of the orphan G-tracts. These results from the orphan G-tracts also confirmed the ability of a single G-tract to form an hG4, not to mention those with two or three G-tracts.

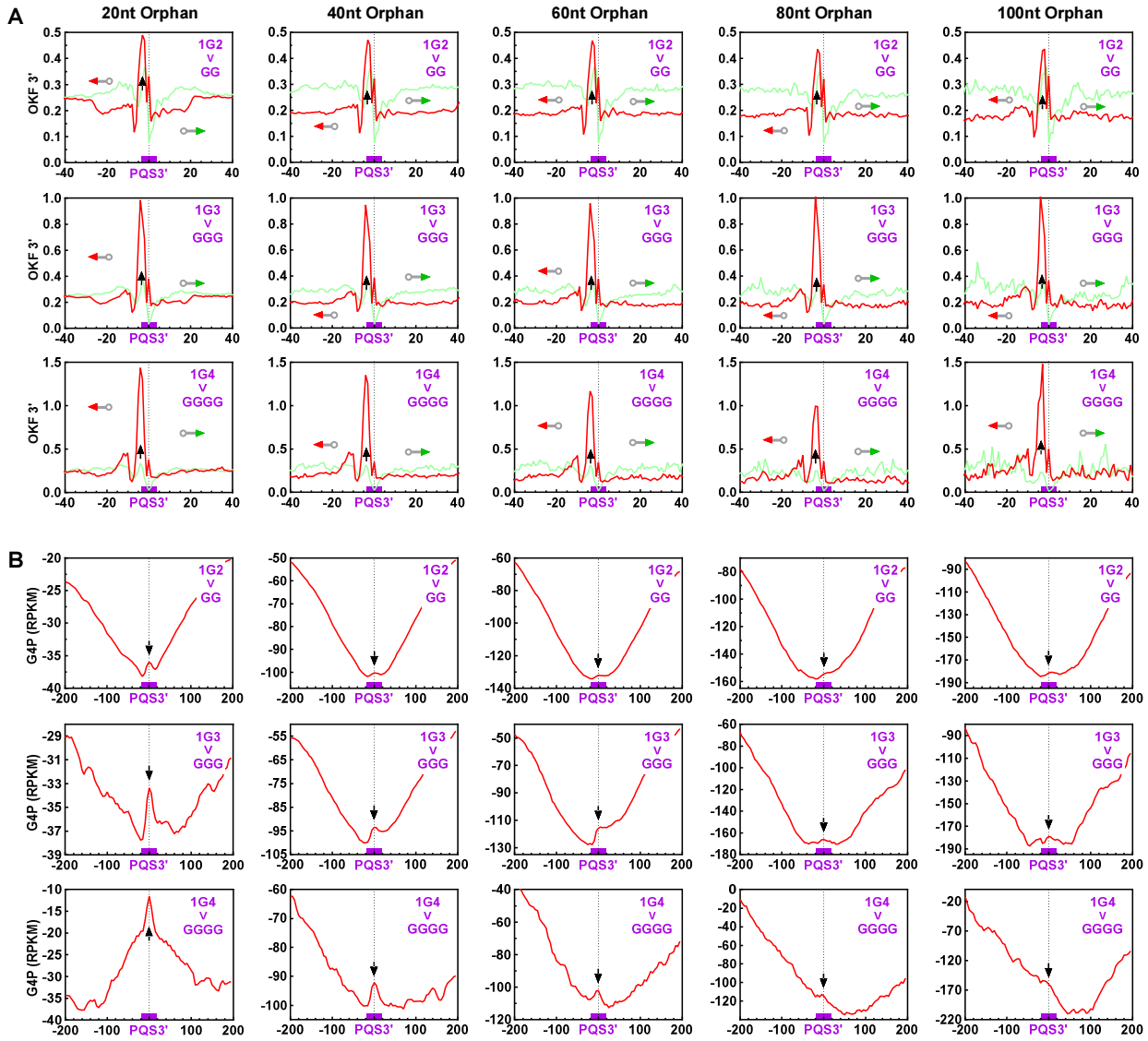

Figure 12. hG4 formation in orphan G-tracts isolated from neighboring G-tracts by at least the indicated number of nucleotides. (A) Detection by an OKF pause peak. (B) Detection by G4P.

Survey of G4 sites in the yeast genome

So far, our analyses have revealed a distinctive picture that goes far beyond our previous perception of the nature and extent of G4 formation in the yeast genome. To get an overview, we surveyed the abundance of G4 sites in the genome and genes. Since promoter activities are mostly concentrated in the 120 nts region upstream of the TSSs in yeast (29,30), we merged this region with the gene bodies. Based on the canonical consensus, the yeast genome can only form up to 38 dG4s. The discovery of hG4 formation increased the total number of G4 sites to 587694, resulting in a >15,000-fold increase, of which the canonical dG4s account for less than 2%

(Figure 13A). Meanwhile, the prevalence of hG4s in genes is >50-fold higher than that of dG4s (Figure 13, B versus C), suggesting a more dominant role of hG4s in transcriptional regulation.

We also profiled the distribution of PQS, OKF, and G4P signals across genes. Figure 13 (D-E) shows that G4 sites had a similar density across the whole genome with a significant drop near the transcription start sites (TSSs) and transcription end sites (TESs). In contrast, the OKF-3' signal peaked at the TSSs (Figure 13F), probably due to a locally greater production of RNA transcripts. However, the G4P binding instead showed a gradual decrease towards the TSSs (Figure 13G), which could be attributed to a greater activity of protein interaction and translocation involved in transcription initiation.

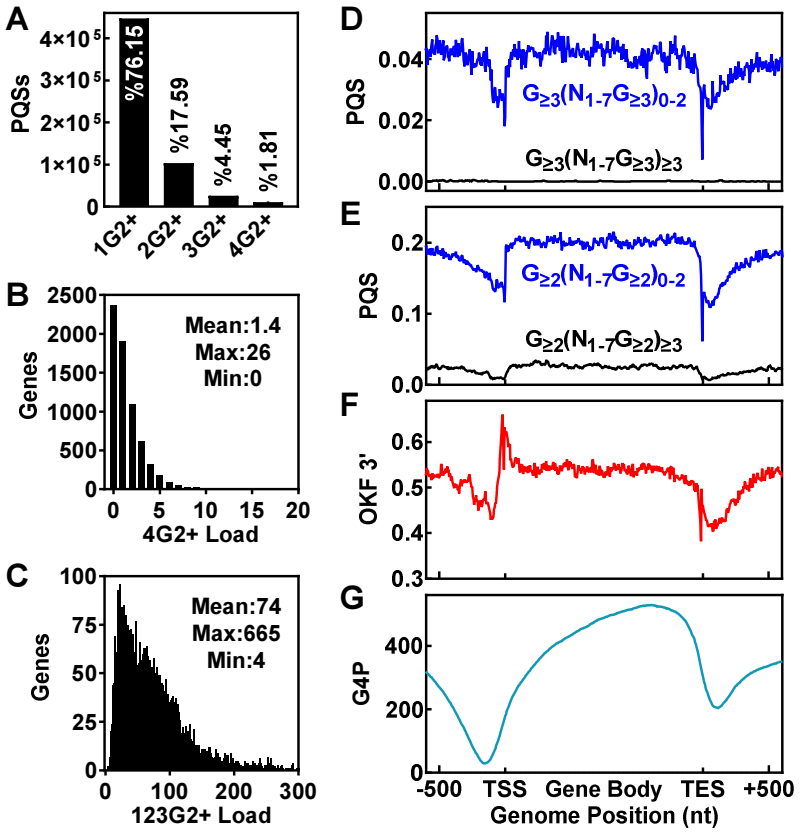

Figure 13. Overview of hG4 sites in the yeast genome and genes. (A) Count of total PQS with different numbers of G-tracts. (B) PQSs capable of forming dG4s in genes. (C) PQS capable of forming only hG4s in genes. (D) Distribution of PQS capable of forming only hG4s (blue line) and dG4s (black line) of three or more G-tetrads across genes. (E) Distribution of PQS capable of forming only hG4s (blue line) and dG4s (black line) of only two G-tetrads across genes. (F) Distribution of 3' ends of OKFs across genes. (G) Binding of G4P across genes.

Discussion

We have successfully applied the principle of the polymerase stop assay and the G4P to detect the formation of hG4s in the yeast genome, providing the first evidence for the existence of hG4s in living eukaryotic cells. Due to the relaxation of the G-tracts requirement, hG4s are found to be the overwhelmingly dominant G4 species. The biological roles of G4s are quite diverse (22,43). Compared to the canonical dG4s, the hG4s have several distinctive features. First, their formation sites are orders of magnitude more abundant, as an hG4 can form with as few as one DNA G-tract of two guanines. Second, the different combinations of G-tracts from DNA and RNA (Figure 10B) (11) give them greater structural variability. Most importantly, hG4 formation is coupled to transcription (12,20); thus, hG4-mediated events will be transcription-dependent.

G4s have been implicated in transcriptional regulation (44). It has been postulated that hG4s serve as constitutional cis-regulatory elements in transcription (7). The transcriptional-dependence of hG4 formation may give it a unique role in creating a feedback loop for transcriptional regulation. G4s have a general role of acting as "roadblocks" to impede protein translocation along a DNA strand (36,45). This results in transcriptional repression under both *in vitro* and *in vivo* conditions when a G4 is present within a gene body (7,11,46). Therefore, greater transcription will result in greater RNA level and consequently greater hG4 formation, which in turn will result in greater transcriptional repression. Due to the ubiquitous presence of hG4 formation in genes, this negative feedback may serve as a common mechanism for maintaining transcriptional homeostasis and stability.

Our study also illustrates how hG4s directly affect replication. The impediment to the progression of DNA polymerase demonstrates how hG4s, as a type of "roadblock" on such a large scale, can affect protein translocation along a DNA molecule (e.g., Figure 2), a crucial event in genomic metabolism. Such effects may contribute to genome instability and explain why G4-stabilizing ligands cause DNA damage in a replication-dependent manner (47). In reality, an hG4 can affect any protein-DNA recognition or interaction, such as binding and diffusion.

The formation of hG4s of only two G-tetrads greatly diversifies the stability of G4 structures. In our previous studies, hG4s with two G-tetrads showed a lifetime of minutes (6), while those with three G-tetrads could last for hours (12) if left alone. Thus, these two sets of structures may function in a more responsive and persistent manner, respectively. Since the involvement of RNA G-tracts promotes G4 formation in both kinetics and stability (15), the different ratio of DNA:RNA G-tracts in an hG4 should add another source of diversity.

The formation of hG4s may also have played a role in evolution. PQSs capable of forming a dG4 are progressively selected in evolution, leading to a higher frequency in higher species (8). From this perspective, the formation of hG4s in PQSs with 1-3 G-tracts with as few as two guanines could provide the necessary and smooth transitions to facilitate the adaptation and selection of the intramolecular dG4s. Otherwise, the abrupt transition from nothing to the formation of dG4s of four G-tracts would be difficult to justify rationally.

From a technical point of view, the analysis of the OKF synthesis should provide a very useful tool for the detection of G4 formation in the genomes of living cells. As shown in this work, the analysis of OKF can be performed at 1-nt resolution. The opening of a DNA duplex is a prerequisite for G4 formation. Our previous work has shown that G4 formation can be triggered by an approaching transcription bubble (34). A similar mechanism may also operate in a moving

replication fork, such that OKF synthesis and G4 formation occur concurrently in close proximity, resulting in greater sensitivity. However, this mechanism may preferentially detect G4s in the replicating DNA. Therefore, the actual hG4 formation in cells may be more robust than that represented by OKF synthesis.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant # 21977094 and 22037004) and the Shanxi "1331 Project".

References

1. Kwok, C.K. and Merrick, C.J. (2017) G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends Biotechnol.*, **35**, 997-1013.
2. Masai, H. and Tanaka, T. (2020) G-quadruplex DNA and RNA: Their roles in regulation of DNA replication and other biological functions. *Biochem. Biophys. Res. Commun.*, **531**, 25-38.
3. Tian, T., Chen, Y.-Q., Wang, S.-R. and Zhou, X. (2018) G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem*, **4**, 1314-1344.
4. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. and Balasubramanian, S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, **21**, 459-474.
5. Zheng, K.W., Zhang, J.Y., He, Y.D., Gong, J.Y., Wen, C.J., Chen, J.N., Hao, Y.H., Zhao, Y. and Tan, Z. (2020) Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res.*, **48**, 11706-11720.
6. Xiao, S., Zhang, J.Y., Wu, J., Wu, R.Y., Xia, Y., Zheng, K.W., Hao, Y.H., Zhou, X. and Tan, Z. (2014) Formation of DNA:RNA hybrid G-quadruplexes of two G-quartet layers in transcription: expansion of the prevalence and diversity of G-quadruplexes in genomes. *Angew. Chem. Int. Ed. Engl.*, **53**, 13110-13114.
7. Zheng, K.W., Xiao, S., Liu, J.Q., Zhang, J.Y., Hao, Y.H. and Tan, Z. (2013) Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.*, **41**, 5533-5541.
8. Xiao, S., Zhang, J.Y., Zheng, K.W., Hao, Y.H. and Tan, Z. (2013) Bioinformatic analysis reveals an evolutional selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals. *Nucleic Acids Res.*, **41**, 10379-10390.
9. Zhang, J.Y., Zheng, K.W., Xiao, S., Hao, Y.H. and Tan, Z. (2014) Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA. *J. Am. Chem. Soc.*, **136**, 1381-1390.
10. Zheng, K.W., Wu, R.Y., He, Y.D., Xiao, S., Zhang, J.Y., Liu, J.Q., Hao, Y.H. and Tan, Z. (2014) A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the

mitochondrial transcription termination at the DNA replication priming site. *Nucleic Acids Res.*, **42**, 10832-10844.

11. Wu, R.Y., Zheng, K.W., Zhang, J.Y., Hao, Y.H. and Tan, Z. (2015) Formation of DNA:RNA hybrid G-quadruplex in bacterial cells and its dominance over the intramolecular DNA G-quadruplex in mediating transcription termination. *Angew. Chem. Int. Ed. Engl.*, **54**, 2447-2451.
12. Zhao, Y., Zhang, J.Y., Zhang, Z.Y., Tong, T.J., Hao, Y.H. and Tan, Z. (2017) Real-Time Detection Reveals Responsive Cotranscriptional Formation of Persistent Intramolecular DNA and Intermolecular DNA:RNA Hybrid G-Quadruplexes Stabilized by R-Loop. *Anal. Chem.*, **89**, 6036-6042.
13. Zhang, J.Y., Xia, Y., Hao, Y.H. and Tan, Z. (2020) DNA:RNA hybrid G-quadruplex formation upstream of transcription start site. *Sci Rep*, **10**, 7429.
14. Zheng, K.W., Zhang, J.Y. and Tan, Z. (2019) Characterization of Co-transcriptional Formation of G-Quadruplexes in Double-Stranded DNA. *Methods Mol. Biol.*, **2035**, 243-255.
15. Shrestha, P., Xiao, S., Dhakal, S., Tan, Z. and Mao, H. (2014) Nascent RNA transcripts facilitate the formation of G-quadruplexes. *Nucleic Acids Res.*, **42**, 7236-7246.
16. Xu, Y., Ishizuka, T., Yang, J., Ito, K., Katada, H., Komiyama, M. and Hayashi, T. (2012) Oligonucleotide models of telomeric DNA and RNA form a Hybrid G-quadruplex structure as a potential component of telomeres. *J. Biol. Chem.*, **287**, 41787-41796.
17. Xu, Y., Suzuki, Y., Ishizuka, T., Xiao, C.D., Liu, X., Hayashi, T. and Komiyama, M. (2014) Finding a human telomere DNA-RNA hybrid G-quadruplex formed by human telomeric 6-mer RNA and 16-mer DNA using click chemistry: a protective structure for telomere end. *Bioorg. Med. Chem.*, **22**, 4419-4421.
18. Bao, H.L. and Xu, Y. (2020) Telomeric DNA-RNA-hybrid G-quadruplex exists in environmental conditions of HeLa cells. *Chem. Commun. (Camb.)*, **56**, 6547-6550.
19. Choi, B.E. and Lee, H.T. (2022) DNA-RNA hybrid G-quadruplex tends to form near the 3' end of telomere overhang. *Biophys. J.*, **121**, 2962-2980.
20. Xia, Y., Zheng, K.W., He, Y.D., Liu, H.H., Wen, C.J., Hao, Y.H. and Tan, Z. (2018) Transmission of dynamic supercoiling in linear and multi-way branched DNAs and its regulation revealed by a fluorescent G-quadruplex torsion sensor. *Nucleic Acids Res.*, **46**, 7418-7424.
21. Han, H., Hurley, L.H. and Salazar, M. (1999) A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.*, **27**, 537-542.
22. Gong, J.Y., Wen, C.J., Tang, M.L., Duan, R.F., Chen, J.N., Zhang, J.Y., Zheng, K.W., He, Y.D., Hao, Y.H., Yu, Q. *et al.* (2021) G-quadruplex structural variations in human genome associated with single-nucleotide variations and their impact on gene activity. *Proc. Natl. Acad. Sci. USA*, **118**, e2013230118.
23. Porcella, S.Y., Koussa, N.C., Tang, C.P., Kramer, D.N., Srivastava, P. and Smith, D.J. (2020) Separable, Ctf4-mediated recruitment of DNA Polymerase alpha for initiation of DNA synthesis at replication origins and lagging-strand priming during replication elongation. *PLoS Genet.*, **16**, e1008755.

24. Koussa, N.C. and Smith, D.J. (2021) Post-replicative nick translation occurs on the lagging strand during prolonged depletion of DNA ligase I in Saccharomyces cerevisiae. *G3 (Bethesda)*, **11**.

25. Koussa, N.C. and Smith, D.J. (2021) Limiting DNA polymerase delta alters replication dynamics and leads to a dependence on checkpoint activation and recombination-mediated DNA repair. *PLoS Genet.*, **17**, e1009322.

26. Yeung, R. and Smith, D.J. (2020) Determinants of Replication-Fork Pausing at tRNA Genes in Saccharomyces cerevisiae. *Genetics*, **214**, 825-838.

27. Kahli, M., Osmundson, J.S., Yeung, R. and Smith, D.J. (2019) Processing of eukaryotic Okazaki fragments by redundant nucleases can be uncoupled from ongoing DNA replication in vivo. *Nucleic Acids Res.*, **47**, 1814-1822.

28. Quinlan, A.R. (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*, **47**, 11 12 11-34.

29. Lubliner, S., Keren, L. and Segal, E. (2013) Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res.*, **41**, 5569-5581.

30. Lin, Z., Wu, W.S., Liang, H., Woo, Y. and Li, W.H. (2010) The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. *BMC Genomics*, **11**, 581.

31. Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187-191.

32. Gietz, R.D., Schiestl, R.H., Willems, A.R. and Woods, R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, **11**, 355-360.

33. Lohse, M.B., Kongsomboonvech, P., Madrigal, M., Hernday, A.D. and Nobile, C.J. (2016) Genome-Wide Chromatin Immunoprecipitation in Candida albicans and Other Yeasts. *Methods Mol. Biol.*, **1361**, 161-184.

34. Liu, J.Q., Xiao, S., Hao, Y.H. and Tan, Z. (2015) Strand-Biased Formation of G-Quadruplexes in DNA Duplexes Transcribed with T7 RNA Polymerase. *Angew. Chem. Int. Ed. Engl.*, **54**, 8992-8996.

35. Zhang, C., Liu, H.H., Zheng, K.W., Hao, Y.H. and Tan, Z. (2013) DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix. *Nucleic Acids Res.*, **41**, 7144-7152.

36. Liu, J.Q., Chen, C.Y., Xue, Y., Hao, Y.H. and Tan, Z. (2010) G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates. *J. Am. Chem. Soc.*, **132**, 10521-10527.

37. Kumari, R., Nambiar, M., Shanbagh, S. and Raghavan, S.C. (2015) Detection of G-quadruplex DNA using primer extension as a tool. *PLoS ONE*, **10**, e0119722.

38. Sun, D. and Hurley, L.H. (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol. Biol.*, **608**, 65-79.

39. Kwok, C.K. and Balasubramanian, S. (2015) Targeted Detection of G-Quadruplexes in Cellular RNAs. *Angew. Chem. Int. Ed. Engl.*, **54**, 6751-6754.

40. Hagihara, M., Yoneda, K., Yabuuchi, H., Okuno, Y. and Nakatani, K. (2010) A reverse transcriptase stop assay revealed diverse quadruplex formations in UTRs in mRNA. *Bioorg. Med. Chem. Lett.*, **20**, 2350-2353.

41. Balakrishnan, L. and Bambara, R.A. (2013) Okazaki fragment metabolism. *Cold Spring Harb. Perspect. Biol.*, **5**.

42. Rachwal, P.A., Brown, T. and Fox, K.R. (2007) Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry*, **46**, 3036-3044.

43. Robinson, J., Raguseo, F., Nuccio, S.P., Liano, D. and Di Antonio, M. (2021) DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.*, **49**, 8419-8431.

44. Kim, N. (2019) The Interplay between G-quadruplex and Transcription. *Curr. Med. Chem.*, **26**, 2898-2917.

45. Belotserkovskii, B.P., Liu, R., Tornaletti, S., Krasilnikova, M.M., Mirkin, S.M. and Hanawalt, P.C. (2010) Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl. Acad. Sci. USA*, **107**, 12816-12821.

46. Endoh, T., Kawasaki, Y. and Sugimoto, N. (2013) Suppression of gene expression by G-quadruplexes in open reading frames depends on G-quadruplex stability. *Angew. Chem. Int. Ed. Engl.*, **52**, 5522-5526.

47. Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N.J., Santos, N.D., Silvester, J., Wei, V., Garcia, J. *et al.* (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.*, **8**, 14432.

Supporting Information for

Genome-Wide Formation of DNA:RNA Hybrid G-Quadruplexes in Live Yeast Cells

Chen-xia Ren[a,c], Rui-fang Duan[a,c], Yu-hua Hao[b], Zheng Tan[a,b,1]
aCenter for Healthy Aging, Central Laboratory, Changzhi Medical College, Changzhi 046000, Shanxi, P. R. China

[b]State Key Laboratory of Membrane Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, P. R. China

[c]Contributed equally.

[1]Addressed correspondence to: Zheng Tan, E-mail: z.tan@ioz.ac.cn

Table S1. DNAs used in the DNA polymerase and exonuclease stop assay.

DNA polymerase stop assay	
DNA	Sequence (5'-3')
G4 (Yeast)	ctctcttaac**GGGG**a**GGGG**aa**GGGG**a**GGGG**tgtatacactgaagcgatcgtgatac
G4mut (Yeast)	ctctcttaac**G**cc**G**a**GGGG**aa**GGGG**a**GGGG**tgtatacactgaagcgatcgtgatac
G4 (Human)	ctctcttaac**GGG**c**GGG**cgcga**GGG**a**GGG**tgtatacactgaagcgatcgtgatac
G4mut (Human)	ctctcttaac**GGG**c**G**a**G**cgcga**GGG**a**GGG**tgtatacactgaagcgatcgtgatac
Primer	FAM-gtatcacgatcgctt
Marker	tgtatacactgaagcgatcgtgatac
G4	ctctcttaac**GGG**tt**GGG**tt**GGG**tt**GGG**tt**GGG**tt**GGG**tt**GGG**tgtatacactgtttatcgaacggtg
G4mut	ctctcttaac**G**t**G**tt**G**t**G**tt**G**t**G**tt**G**t**G**tt**G**t**G**tt**G**t**G**tgtatacactgtttatcgaacggtg
Primer	FAM-caccgttcgataaac
DNA exonuclease stop assay	
DNA	Sequence (5'-3')
G4 (Yeast)	FAM-tttttttttttt**GGGG**a**GGGG**aa**GGGG**a**GGGG**tttttttttt
G4mut (Yeast)	FAM-tttttttttttt**G**cc**G**a**GGGG**aa**GGGG**a**GGGG**tttttttttt
Marker (Yeast)	FAM-tttttttttttt**GGGG**a**GGGG**aa**GGGG**a**GGGG**
G4 (Human)	FAM-tttttttttttt**GGG**a**GGG**cgcga**GGG**a**GGG**tttttttttt
G4mut (Human)	FAM-tttttttttttt**GGG**c**G**a**G**cgcga**GGG**a**GGG**tttttttttt
Marker (Human)	FAM-tttttttttttt**GGG**c**GGG**cgcga**GGG**a**GGG**
G4	FAM-tttttttttttttt**GGGG**tt**GGGG**tt**GGGG**tt**GGGG**tt**GGGG**tt**GGGG**tt**GGGG**tttttttttt

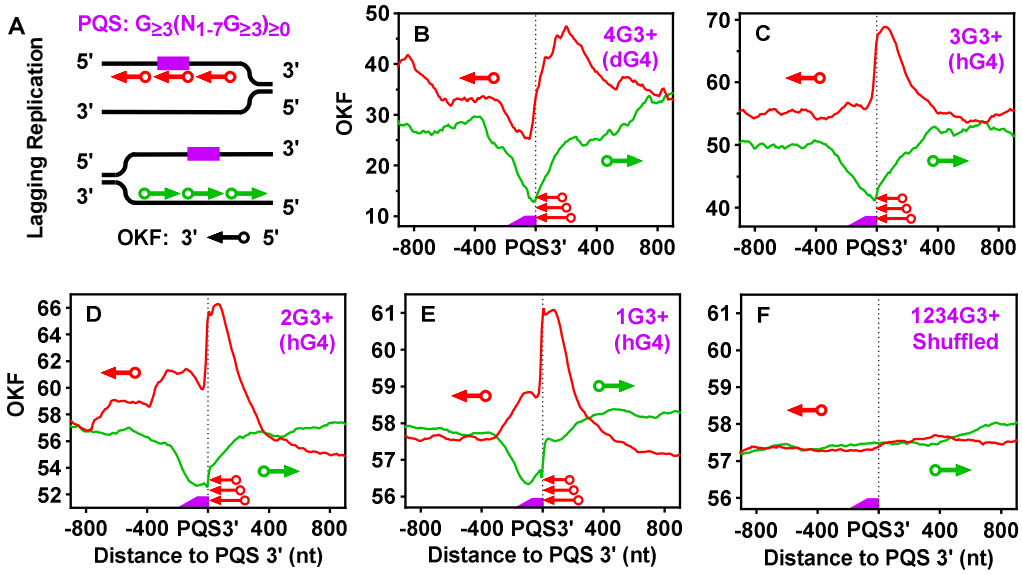

Figure S1. Distribution of OKFs across the 3' end of various PQSs with 4 to 1 $G_{\geq 3}$ tracts that could form either dG4s or hG4s of three or more G-tetrads. OKF-seq data from GSE118078. Other information is the same as in Figure 3.

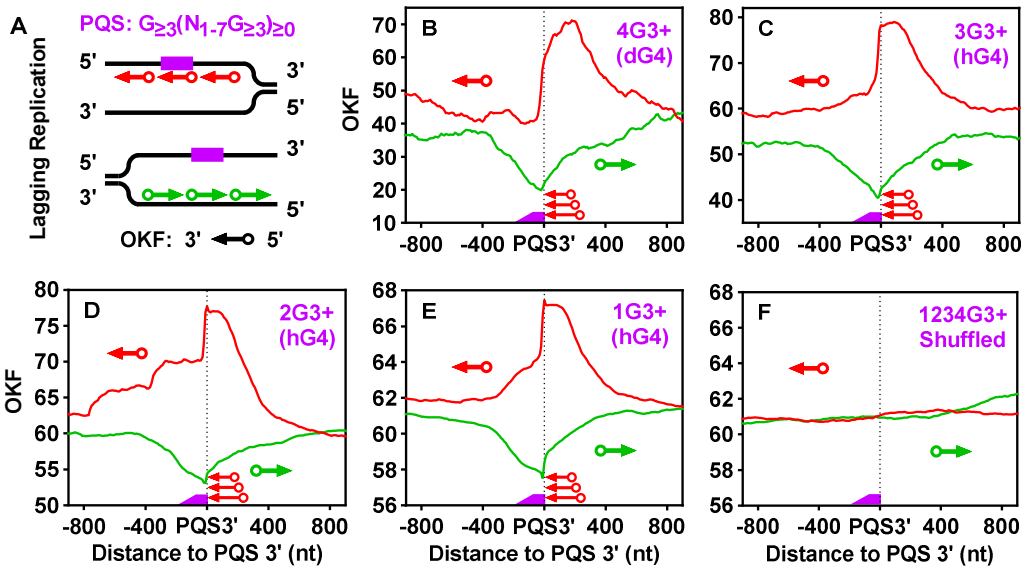

Figure S2. Distribution of OKFs across the 3' end of various PQSs with 4 to 1 $G_{\geq 3}$ tracts that could form either dG4s or hG4s of three or more G-tetrads. OKF-seq data from GSE139860. Other information is the same as in Figure 3.

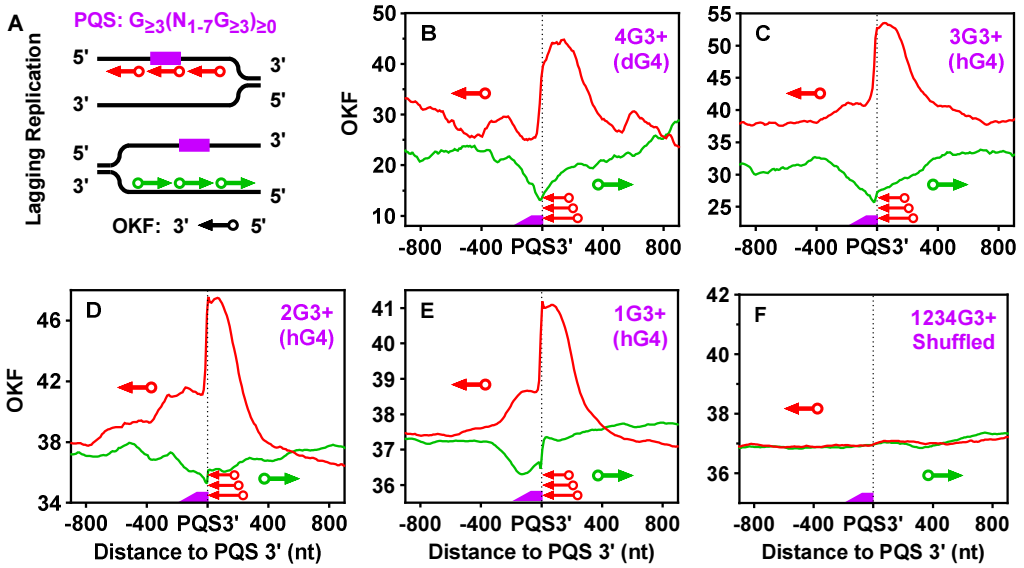

Figure S3. Distribution of OKFs across the 3' end of various PQSs with 4 to 1 $G_{\geq 3}$ tracts that could form either dG4s or hG4s of three or more G-tetrads. OKF-seq data from GSE141884. Other information is the same as in Figure 3.

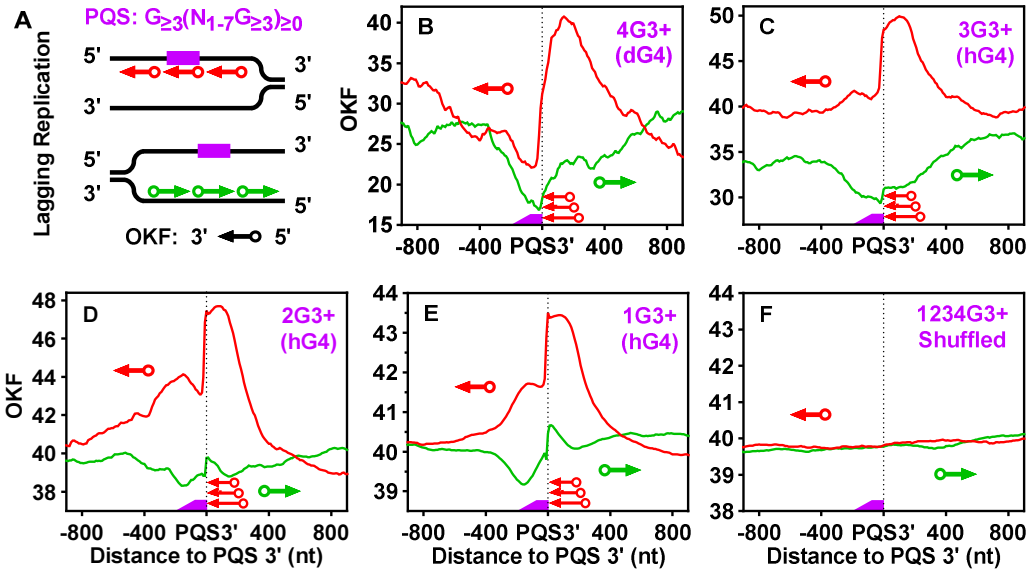

Figure S4. Distribution of OKFs across the 3' end of various PQSs with 4 to 1 $G_{\geq 3}$ tracts that could form either dG4s or hG4s of three or more G-tetrads. OKF-seq data from GSE173065. Other information is the same as in Figure 3.

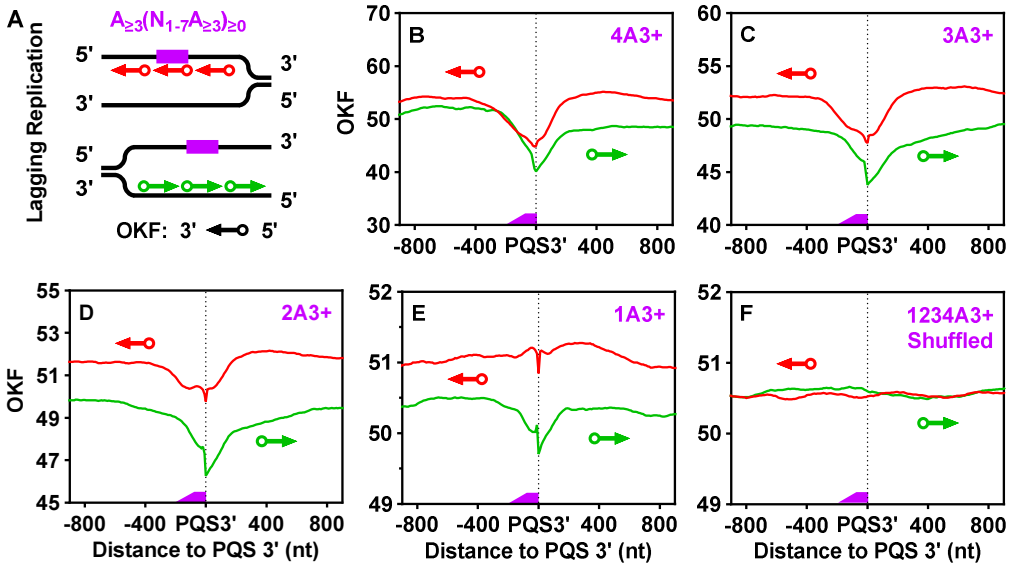

Figure S5. Distribution of OKFs across the 3' ends of motifs containing 1 to 4 or more $A_{\geq3}$ tracts that were unable to form a G4. Other information is the same as in Figure 3.

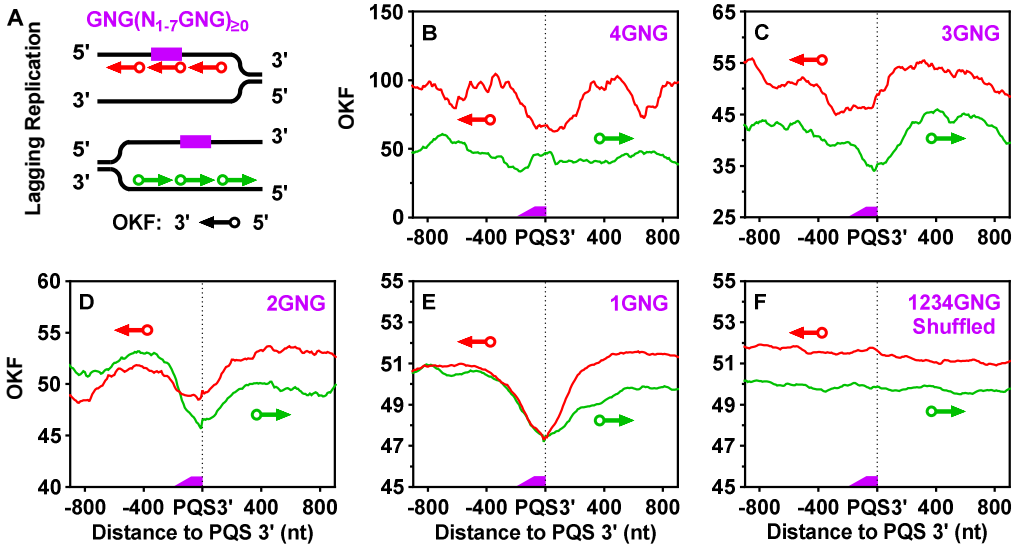

Figure S6. Distribution of OKFs across the 3' ends of motifs containing 1 to 4 or more GNG tracts that are unable to form a G4. Other information is the same as in Figure 3.

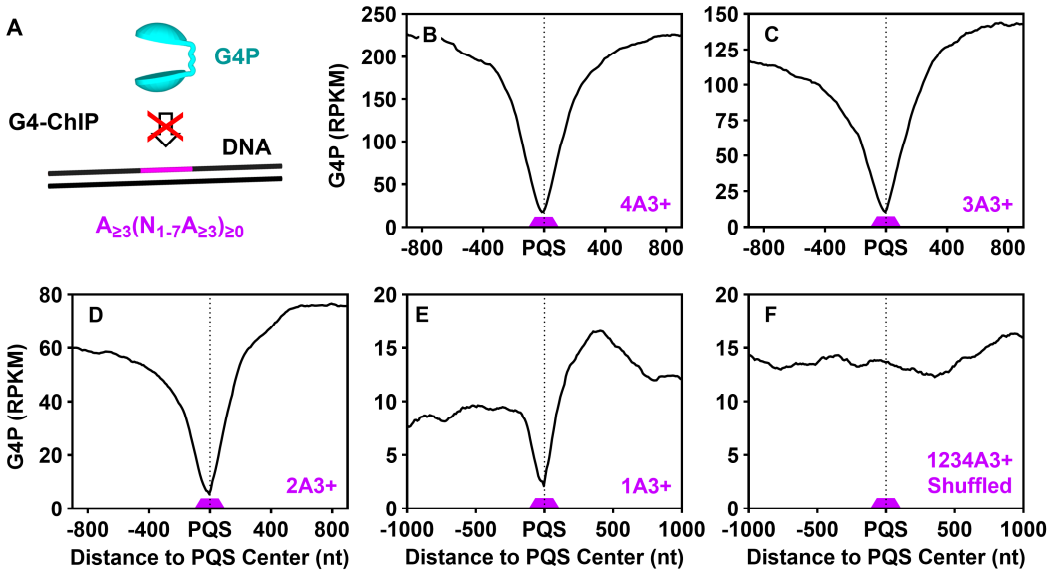

Figure S7. Binding profiles of G4P across PQSs of one or more $A_{\geq3}$ tracts that are unable to form a G4.

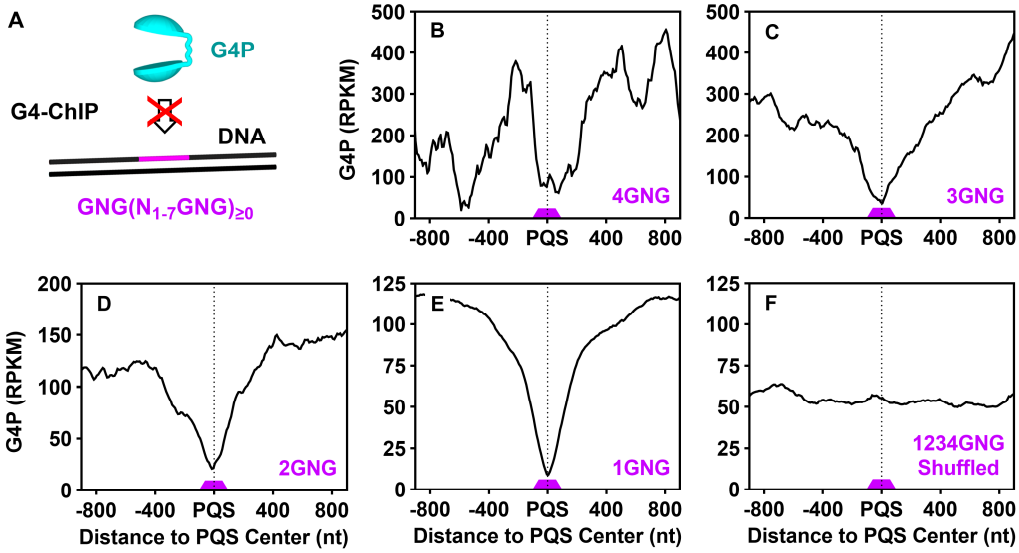

Figure S8. Binding profiles of G4P across PQSs of one or more GNG tracts that are unable to form a G4.

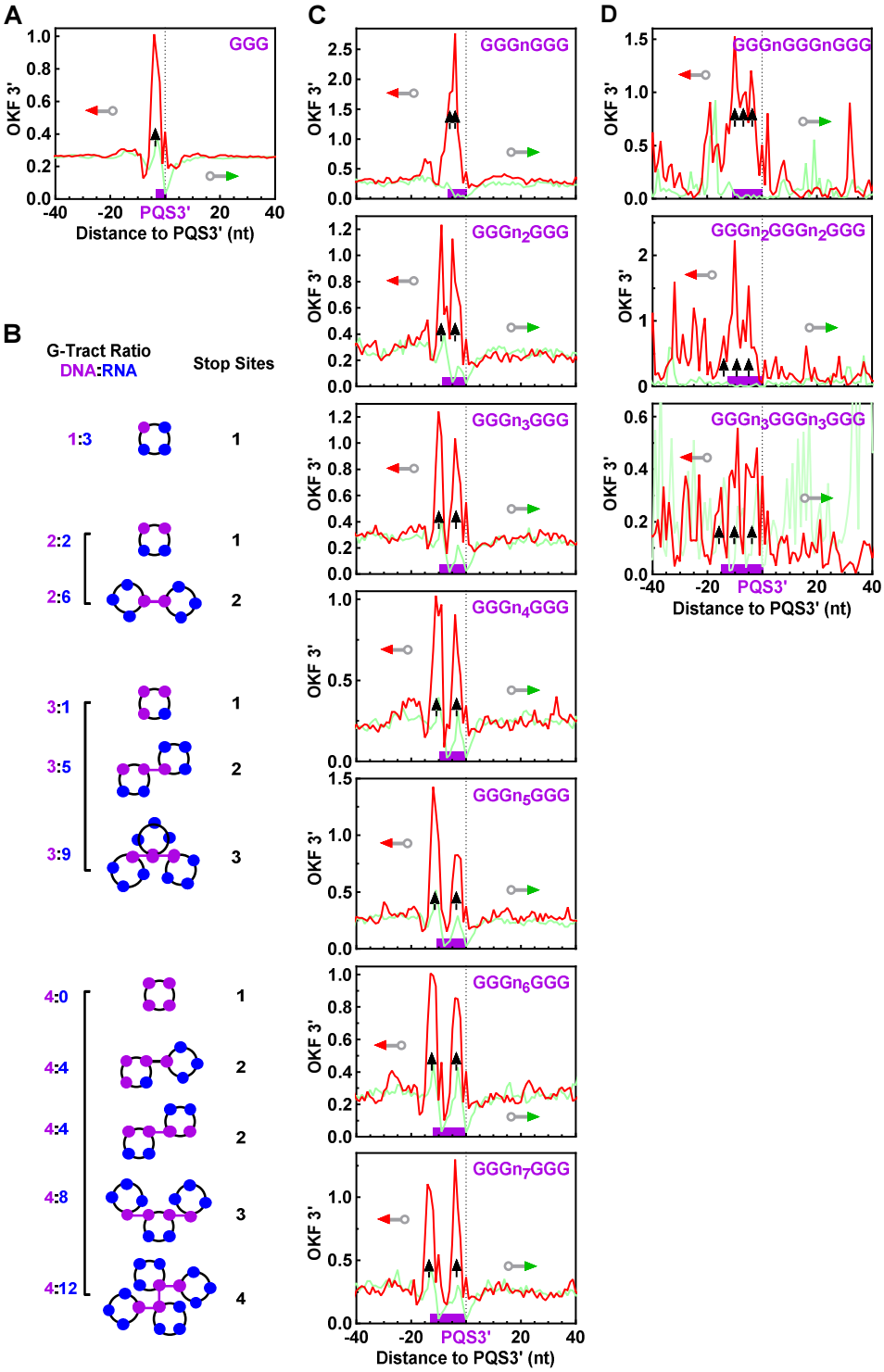

Figure S9. OKF-3' profiling showing the involvement of G-tracts from DNA and RNA in the joining formation of hG4s in PQSs with (A) one, (C) two, (D) three, or (E) four GGG tracts. (B) Representative combinations of DNA and RNA G-tracts in G4 formation. Bin size: 1 nt.

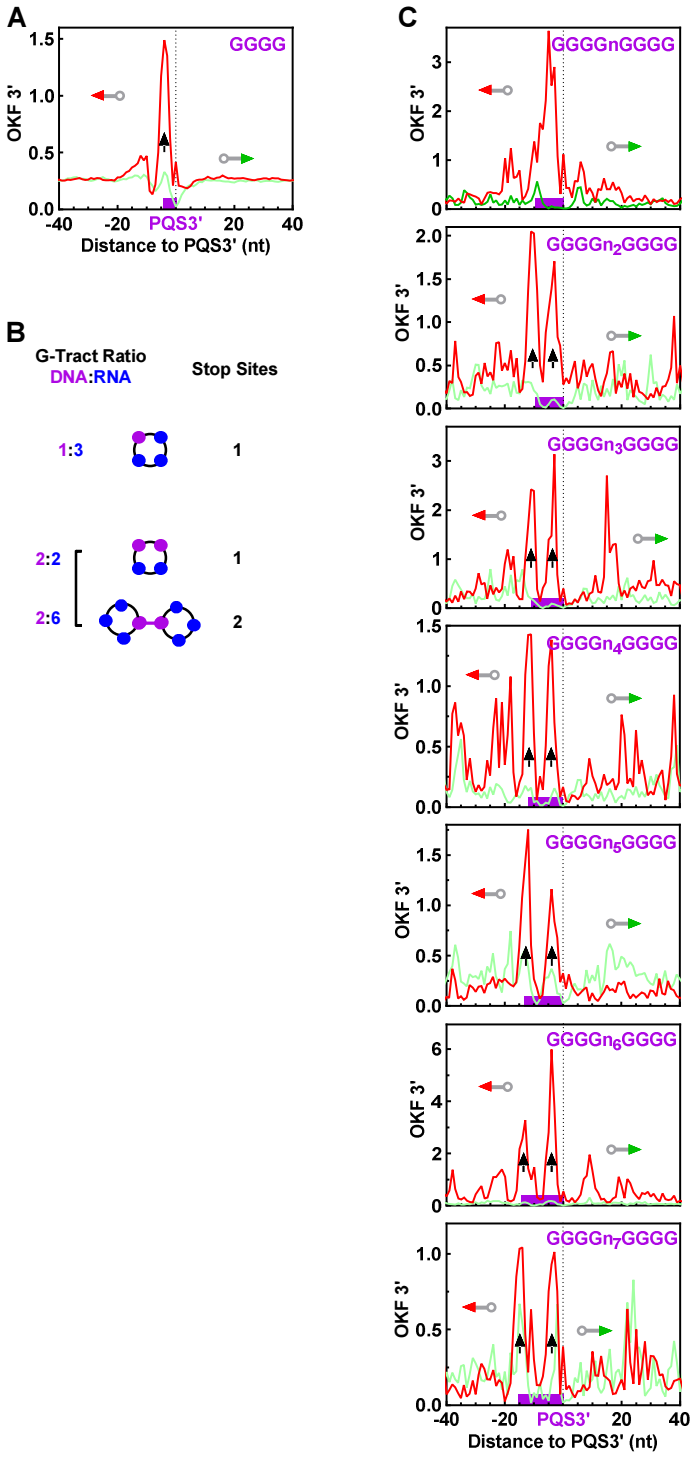

Figure S10. OKF-3' profiling showing the involvement of G-tracts from DNA and RNA in the joining formation of hG4s in PQSs with (A) one or (C) two GGGG tracts. (B) Representative combinations of DNA and RNA G-tracts in G4 formation. Bin size: 1 nt.