

RESEARCH ARTICLE

# G-Quadruplexes Involving Both Strands of Genomic DNA Are Highly Abundant and Colocalize with Functional Sites in the Human Genome

Andrzej S Kudlicki<sup>1,2,3\*</sup>

**1** Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas, United States of America, **2** Institute for Translational Sciences, University of Texas Medical Branch, Galveston, Texas, United States of America, **3** Sealy Center for Molecular Medicine, University of Texas Medical Branch, Galveston, Texas, United States of America

\* [askudlic@utmb.edu](mailto:askudlic@utmb.edu)



**OPEN ACCESS**

**Citation:** Kudlicki AS (2016) G-Quadruplexes Involving Both Strands of Genomic DNA Are Highly Abundant and Colocalize with Functional Sites in the Human Genome. PLoS ONE 11(1): e0146174. doi:10.1371/journal.pone.0146174

**Editor:** Arthur J. Lustig, Tulane University Health Sciences Center, UNITED STATES

**Received:** July 9, 2015

**Accepted:** December 13, 2015

**Published:** January 4, 2016

**Copyright:** © 2016 Andrzej S Kudlicki. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This study was conducted with the support of the Institute for Translational Sciences at the University of Texas Medical Branch, supported in part by a Clinical and Translational Science Award (UL1TR000071 and UL1TR001439) from the National Center for Advancing Translational Sciences. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The G-quadruplex is a non-canonical DNA structure biologically significant in DNA replication, transcription and telomere stability. To date, only G4s with all guanines originating from the same strand of DNA have been considered in the context of the human nuclear genome. Here, I discuss interstrand topological configurations of G-quadruplex DNA, consisting of guanines from both strands of genomic DNA; an algorithm is presented for predicting such structures. I have identified over 550,000 non-overlapping interstrand G-quadruplex forming sequences in the human genome—significantly more than intrastrand configurations. Functional analysis of interstrand G-quadruplex sites shows strong association with transcription initiation, the results are consistent with the XPB and XPD transcriptional helicases binding only to G-quadruplex DNA with interstrand topology. Interstrand quadruplexes are also enriched in origin of replication sites. Several topology classes of interstrand quadruplex-forming sequences are possible, and different topologies are enriched in different types of structural elements. The list of interstrand quadruplex forming sequences, and the computer program used for their prediction are available at the web address <http://moment.utmb.edu/allquads>.

## Introduction

The G-quadruplex (G4) is a non-canonical DNA structure consisting of four strands stabilized by Hoogsteen bonds that has received significant attention in the recent years. G4s have been implicated in numerous cellular contexts and functions [1,2], including telomeres [3], *cis*-acting regulatory elements [4], transcription [5], and replication [6,7,8]. Four runs of guanine must be present in the DNA sequence from which a G4 is created [9,10]. While bimolecular or tetramolecular G-quadruplexes have been discussed in the context of short oligomers, or of



- $m/(G\{3,\},\{0,7\}(C\{3,\},\{1,7\})^2.\{1,7\}^2[(G\{3,\},\{1,7\})^3.\{1,7\}^3.\{0,7\}(C\{3,\},g\ m/(C\{3,\},\{0,7\}(G\{3,\},\{1,7\})^2.\{1,7\}^2(C\{3,\},\{1,7\})^3.\{1,7\}^3.\{0,7\}(G\{3,\},g\ m/(G\{3,\},\{0,7\}(C\{3,\},\{0,7\})^1.\{1,7\}^1(C\{3,\},\{1,7\})^3.\{0,7\}(G\{3,\},\{0,7\})^3/g\ m/(C\{3,\},\{0,7\}(G\{3,\},\{0,7\})^1.\{1,7\}^1|(G\{3,\},\{1,7\})^3.\{0,7\}(C\{3,\},\{0,7\})^3/g$

for the different topology classes of cross-strand G-quadruplexes. The first regular expression produces results nearly identical to the *Quadparser* software [12], with minor differences due to different implicit heuristics applied in situations where alternative or overlapping PQS sequences exist. Similarly, my approach to finding interstrand quadruplex-forming sequences differs from the approach of Cao et al. [13] in that here a separate one-step search is performed for each topology class, while Cao et al. first identify DNA intervals with quadruplex-forming potential and then characterize the topology of the possible quadruplex. As a result, in certain cases of partially overlapping quadruplex-forming sequences some of the alternative topology types may be missed by the two-step approach, although my one-step method requires additional processing of the results if only non-overlapping sequences are desired. A complete Perl program and the results for the hg19 human genome assembly are available as supplementary data, and from the supporting website <http://moment.utmb.edu/allquads> (the website also provides the results for the hg18 and hg38 assemblies). The program reads a fasta file and outputs PQS's in text format, one per line, including sequence id, topology class, position and the PQS sequence. The post-processing required to identify overlapping quadruplex-forming sequences is a straightforward task. It can be implemented as an algorithm with  $O(N \log N)$  computational complexity in the number of quadruplexes when the PQS and DS-PQS sequences are first sorted according to chromosomal coordinate, this function is implemented among others by the *intersectBed* command of the *bedtools* package [32].

## Analysis of functional associations

Following Gray et al. [5], I also predicted the potential PQS and DS-PQS sequences allowing for loops of up to 12 nt between the guanine tracts (replacing 7 with 12 in the regular expressions above), and used them in the functional analysis of PQS and DS-PQS sites compared to sequencing data on transcriptional initiation and origins of replication. The analysis used the hg19 build of the genome, with the exception of the *hf2* antibody pull-down results [19] that have been mapped to hg18. To analyze the prevalence of PQS and DS-PQS sequences in promoter regions, I applied the *allquads.pl* algorithm directly to the *upstream1000.fa* fasta file obtained from the UCSC genome database on February 17<sup>th</sup>, 2015. To infer functions enriched in the DS-PQS loci, I analyzed their overlaps with experimentally identified sites of transcription initiation and origins of replication. The binomial tests for enrichment were performed as in [5], using the *gsl\_cdf\_binomial\_Q* function included in the *Math::GSL::CDF* CPAN library [33].

The ChIP-seq, pull-down, G4-seq and nascent DNA sequencing peaks were obtained from GEO accession numbers GSE44849 (GSM1092544, GSM1092545), GSE28911 (GSM716435, GSM716437), GSE63874 and GSE45241 (GSM1099724, GSM1099725, GSM1099726, GSM1099727). Overlaps between peaks and G4's were defined if there was at least one base pair common to both features.

## Results

### Prevalence of interstrand quadruplex forming sequences

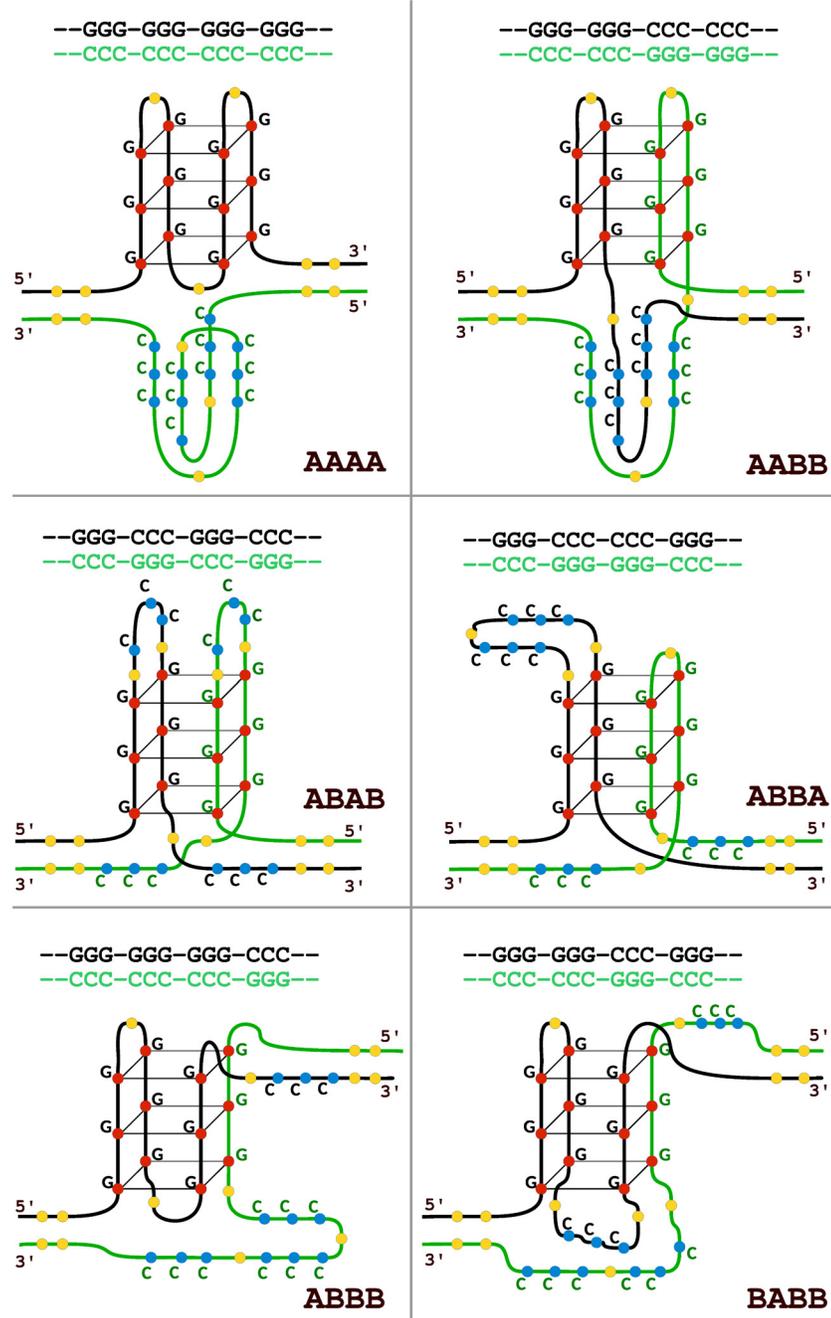
Nine classes of G4 topologies involving both DNA strands are possible in addition to the previously described case of all guanines located on one strand of dsDNA [13], or five if one ignores

the difference between quadruplex starting from the positive and one starting from the negative strand; example topological configurations are shown in [Fig 1](#). Depending on the order of guanine and cytosine tracts on either strand within the sequence, I will denote them as AABB (4), ABAA (8), ABAB (6), ABBA (7), ABBB (2), BAAA (1), BABA (5), BABB (9) and BBAA (3), numbers in parentheses correspond to pattern classes as defined by Cao et al. [13]. AAAAA stands for the widely discussed single-strand configuration; generally “A” represents a guanine tract and “B” a cytosine tract, counting from the 5' end of either strand, reverse complements are not distinguished (e.g. AABA and BABB are the same). Note that each of the ten classes allows several conformations (differing by polarity and arrangement of loops), however they cannot be distinguished based on sequence alone. It is likely that certain topologies will allow formation of a hybrid *i-motif* [4] in addition to the G4, depending on the lengths of the loops connecting the runs of guanine and cytosine. The *i-motif* requires a specific range of pH [34,35] and its significance in-vivo is thus limited, therefore although *i-motifs* in physiological conditions have been reported in certain cases [36,37,38], in this paper I will focus on the G-quadruplex only.

To identify putative double-strand-derived quadruplex sequences (DS-PQS) in the human genome, I implemented an algorithm representing the sequence search in terms of Perl-compatible regular expressions. The source code of the program *allquads.pl* is available as a supplementary material ([S1 File](#)) as well as from the supporting website <http://moment.utmb.edu/allquads>. A genome-wide search with loop lengths between 1 and 7 nucleotides (0 to 7 for loops between guanine runs on opposite strands) has revealed 897,935 DS-PQS sequences, of which 196,953 have an overlap of at least one nucleotide with one of the 374,834 single-strand (AAAA) PQS's. 150,294 DS-PQS are of the BAAA topology class (97,565 without overlap with a single-strand quadruplex), 152,329 (99,299 not overlapping with a single-strand PQS)–ABBB, 69,198 (56,445)–AABB, 142,890 (115,766)–ABAA, 55,735 (50,866)–ABAB, 96,163 (87,940)–ABBA, 49,558 (44,976)–BABA, 128,404 (101,561)–BABB and 53,364 (46,024)–BBAA. Notably, a significant asymmetry exists in the prevalence of some of the ‘mirror image’ configurations: AABB is more abundant than BBAA, ABAB is more abundant than BABA, and ABAA than BABB. Similar differences are present in the yeast genome, see fig. 4a of Cao et al. [13].

Taking into account the overlaps between DS-PQS's of different topologies, 550,977 independent DS-PQS sites are present in the human genome. Similarly, 832,540 independent groups of overlapping quadruplex-forming sequences of any type (interstrand or intrastrand) are found. The complete list of human DS-PQS sites is available in the online supplement ([S2 File](#)) and from the supporting website. The numbers of identified interstrand and intrastrand PQS sites per chromosome are shown in [Table 1](#), and detailed breakdown by topology class is listed in [S1 Table](#). While the average PQS density (per megabase) varies significantly from chromosome to chromosome, the ratios of numbers of intrastrand to interstrand PQS sequences for every chromosome are close to the genome average of 0.68, with the exception of the X and Y chromosomes that are depleted in DS-PQS sites and have intrastrand to interstrand ratios of 0.81 and 0.82 respectively. This statistically significant difference ( $20.7\sigma$  and  $9.3\sigma$  respectively, Poisson model) may reflect different functions of genes, different regulation, or different chromatin organization in the sex chromosomes compared to autosomes.

The presented prevalence of potentially quadruplex-forming sequences have been computed for the standard human genome. While the genomes of human cell lines used in research (such as HeLa or HEK293) differ from the standard genome assembly, most of the differences are translocations or copy number variations that do not have significant impact on the presence or absence of sequences potentially forming G-quadruplexes. Local polymorphisms specific to the cell lines are limited to a relatively small number of sequences, and are not expected



**Fig 1. Examples of topology classes of G-quadruplex structures within genomic DNA.** Examples of topological configurations of intrastrand and interstrand quadruplexes are shown schematically. For pairs of topology types that differ only by strand from which the sequence is derived (e.g. ABAB and BABA) only one is shown. Black, green—the Watson and Crick strands. Red: guanines, blue: cytosines, yellow: loops of up to seven nucleotides of any type. AAAA—the intrastrand topology, AABB, ABAB, ABBA, ABBB, BABB: interstrand configurations.

doi:10.1371/journal.pone.0146174.g001

**Table 1. Intrastrand and interstrand G-quadruplex sequences by human chromosome.**

Chromosome	Unique interstrand		Intrastrand		Intra/inter ratio
	count	per MB	count	per MB	
<b>Genome wide</b>	550,977	177.98	374,834	121.08	0.68
<b>chr1</b>	48,860	196.03	32,597	130.78	0.67
<b>chr2</b>	39,032	160.49	26,826	110.30	0.69
<b>chr3</b>	26,696	134.81	18,839	95.14	0.71
<b>chr4</b>	21,369	111.79	15,358	80.34	0.72
<b>chr5</b>	24,701	136.53	17,587	97.21	0.71
<b>chr6</b>	22,788	133.17	16,675	97.45	0.73
<b>chr7</b>	28,578	179.58	19,372	121.73	0.68
<b>chr8</b>	22,872	156.27	15,750	107.61	0.69
<b>chr9</b>	26,855	190.17	17,522	124.08	0.65
<b>chr10</b>	26,533	195.77	17,420	128.53	0.66
<b>chr11</b>	30,587	226.56	19,984	148.02	0.65
<b>chr12</b>	23,132	172.82	16,330	122.00	0.71
<b>chr13</b>	11,244	97.63	7,802	67.74	0.69
<b>chr14</b>	17,177	160.01	11,311	105.37	0.66
<b>chr15</b>	17,585	171.51	11,733	114.43	0.67
<b>chr16</b>	25,607	283.41	16,495	182.56	0.64
<b>chr17</b>	29,171	359.27	18,897	232.74	0.65
<b>chr18</b>	10,458	133.94	7,474	95.73	0.71
<b>chr19</b>	30,206	510.85	20,794	351.67	0.69
<b>chr20</b>	17,655	280.12	11,496	182.40	0.65
<b>chr21</b>	7,857	163.25	4,848	100.73	0.62
<b>chr22</b>	18,280	356.30	10,420	203.10	0.57
<b>chrX</b>	20,295	130.71	16,468	106.06	<b>0.81</b>
<b>chrY</b>	3,436	57.87	2,832	47.70	<b>0.82</b>

For each chromosome in the human nuclear genome (v. hg19), the numbers of intrastrand and interstrand potentially quadruplex-forming sequences are listed. Sequences with loops up to 7nt long and at least three guanines per tract are considered. In case of overlapping interstrand quadruplexes, only one is taken into account (unique DS-PQS).

doi:10.1371/journal.pone.0146174.t001

to affect the global statistics of PQS sites. For example, the polymorphisms specific to the HEK293T line, as identified by [39], do not overlap with any of the quadruplex-forming sites, either interstrand or intrastrand. Therefore, the results can be readily applied to the genomes of research cell lines.

### Functions of sites with potential to form interstrand G4s

The high abundance of DS-PQS sites opens the possibility that interstrand quadruplexes may play a role in a major cellular process. Indirect evidence in favor of such a role may be derived from association between DS-PQS sites and genomic loci with functional properties known to involve G4 structures. Intrastrand G-quadruplexes and G-quadruplex forming sequences have been reported to coincide with promoter regions and play a role in transcriptional regulation [4,28,40–46]. Indeed, at least one single-strand (AAAA) PQS is present within 45.0% regions 1-kb upstream of a transcription start site. Searching for DS-PQS sites reveals potential

interstrand quadruplexes in 52.5% of these sequences ( $p < 10^{-308}$ ; binomial); a total of 63.1% of human transcripts have at least one putative G-quadruplex of any type in their 1-kb upstream region.

Additional evidence for the role of G4s in transcription initiation has been provided by a recent ChIP-seq study mapping the binding sites of transcriptional helicases XPB and XPD [5]: approximately 20% of XPB and XPD ChIP-seq peaks overlap with a single-strand PQS (approximately 40% when the PQS definition is relaxed to include loops of up to 12nt connecting the guanine runs). I have analyzed the data in the context of quadruplex-forming sites of all types. The overlaps of XPB and XPD binding sites with all detected G4 sequences—including interstrand—are significantly higher: 45% and 48% respectively for standard loop length of up to 7nt; or 70% and 73% respectively for XPB and XPD when allowing for loops up to 12nt long (see details in Table 2 and S2 Table). This result, along with the enrichment in promoter regions described above, demonstrates a significant association of DS-PQS sites with transcriptional initiation ( $p < 10^{-308}$ ; binomial test for enrichment of DS-PQS both in XPB and in XPD peaks). Notably, the enrichment of DS-PQS's in transcriptional helicase binding sites is higher than for interstrand PQS's, and there is no enrichment at all of peaks containing an intrastrand PQS but no DS-PQS; this observation is consistent with XPB and XPD binding only at interstrand G4's; the enrichment of intrastrand PQS reported by [5] may be explained by intrastrand PQS overlapping with DS-PQS or present in some of the loci containing a DS-PQS.

G4 structures have also been associated with origins of replication. To investigate whether this also applies to cross-strand topologies, I considered the overlap between DS-PQS's and origins of replication that have been mapped by sequencing short nascent DNA in human MCF7 and K562 cells [47]. Again, while single-strand PQS's are significantly enriched in the replication origins (present respectively in 20% and 26% of the peaks in K562 and MCF7 libraries), DS-PQS's are even more prevalent (25% and 35%,  $p < 10^{-308}$  in both cases), resulting in 34% and 44% origins respectively overlapping with at least one G4 of any type, see Table 1.

In a recent study, a *hf2* antibody that binds to G4 but not to dsDNA was used in a genome-wide pull-down experiment to characterize stable G4 structures in the human genome [19], revealing that 12% of the *hf2* binding sites overlap with an intrastrand PQS. In *hf2* peaks common to two or more replicate experiments, the ratio is 23%. When both intrastrand and

**Table 2. Intrastrand and interstrand G-quadruplex forming sequences associated with functional sites in the human genome.**

Site	# in genome	Intrastrand PQS			DS-PQS				Any PQS			DS/SS ratio
		with G4	Enr.	%sites	with G4	Enr.	%sites	significance	with G4	Enr.	%sites	
<b>XPB peaks</b>	14,570	3,288	1.99	22.6%	6,052	2.49	41.5%	< 1e-308	6,995	1.91	48.0%	1.84
<b>XPB peaks</b>	21,555	4,466	2.05	20.7%	8,363	2.61	38.8%	< 1e-308	9,769	2.02	45.3%	1.87
<b>ORI MCF7</b>	94,195	24,117	5.05	25.6%	34,425	4.90	36.6%	< 1e-308	42,108	3.97	44.7%	1.43
<b>ORI K562</b>	62,971	12,583	5.08	20.0%	16,724	4.59	26.6%	< 1e-308	22,331	4.05	35.5%	1.33
<b>1kb upstream</b>	39,927	17,975	3.96	45.0%	20,976	3.15	52.5%	< 1e-308	25,199	2.50	63.1%	1.17
<b>Hf2 bind (1+)</b>	9,149	1,095	1.34	12.0%	754	0.63	8.2%	not enriched	1,582	0.87	17.3%	0.69
<b>Hf2 bind (2+)</b>	771	177	1.43	23.0%	128	0.70	16.6%	not enriched	256	0.93	33.2%	0.72
<b>G4seq Pds&amp;K<sup>+</sup></b>	490,269	177,119	39.8	36.2%	108,229	16.6	22.1%	< 1e-308	226,496	22.4	46.2%	0.61
<b>Unique G4s</b>		374,834			550,977				832,540			1.47

For each type of functional site (transcriptional helicase binding, origin of replication, promoter, *hf2* antibody binding), the number and percentage of sites with a G4 motif, and enrichment ratio of G4's within the sites are listed. Data are provided for intrastrand PQS's, inter-strand PQS's, and for PQS of any type, with loops up to 7nt long. For enriched DS-PQS's, a limit on binomial significance is listed. Rightmost column: ratio of number of sites with DS-PQS to number of sites with intrastrand PQS; for XPD and XPB binding sites it exceeds the whole-genome ratio (bottom row).

doi:10.1371/journal.pone.0146174.t002

interstrand PQS's are taken into account, up to 17% of *hf2* peaks (33% for two or more libraries) are associated with a potential quadruplex sequence. In this dataset, the ratio of interstrand to intrastrand G4's is lower than in the functional studies above, and there is no significant enrichment of DS-PQS's in the sequencing peaks. This result does not however contradict previous findings because the *hf2* antibody was designed and tested for specificity only to intramolecular G-quadruplexes, derived from a single strand of DNA.

The resulting low ratio of interstrand to intrastrand quadruplexes is similar in the more sensitive *G4-seq* study of Chambers et al. [48], who detect quadruplexes by analysing sequencing mismatches between conditions promoting and disfavoring G4 formation. The *G4-seq* approach to quadruplex detection involves separate analysis for each strand of DNA and thus appears to favour intrastrand quadruplexes. Nonetheless 108,229 quadruplex sites overlap with DS-PQS loci (16-fold enriched in DS-PQS sequences), including 49,377 observed interstrand quadruplexes not overlapping with an intrastrand PQS, corresponding to a very significant 9.1-fold enrichment (this calculation is based on quadruplexes observed by [48] simultaneously in the  $K^+$  and the PDS experiments). The enrichment provides evidence that the *G4-seq* method does detect DS-PQS sites that do not coincide with an intrastrand PQS.

### Enrichment of topology classes among DS-PQS associated with different functional loci

While the structures of interstrand G4s with different topologies are yet to be determined crystallographically, structural differences between them may be significant for the specific functions of the quadruplex structures. Specifically, if interstrand quadruplexes are functional in transcription initiation or in replication origin, different ratios of numbers of PQS's with particular topology classes may be expected in the quadruplex-forming sequences associated with such functional elements. The numbers of potentially quadruplex-forming sequences in each topological category, associated with each type of functional element are listed in [S3 Table](#), along with the fractions of all PQS's and all DS-PQS's that they constitute, and a comparison with the ratios computed genome-wide, irrespective of functional site. The enrichment calculation uses all predicted quadruplex-forming structures with a 7nt limit on loop length, including overlapping PQS's with different topologies, as any of the overlapping PQS's can be potentially functional. Generally, among the DS-PQS's coinciding with origin of replication sites, the BABB, ABBB and BAAA topologies are significantly enriched (between  $5\sigma$  and  $24\sigma$ ; asymptotic estimation for Poisson distribution), compared to the genome-wide prevalence. In the transcriptional helicase binding sites, the ABBA, BABB and BBAA topologies are enriched compared to their genome-wide abundances, while the intrastrand AAAA is consistently very strongly depleted ( $>18\sigma$ ). Interestingly, while some of the "mirrored" topologies have different abundances genome-wide (e.g. AABB vs. BBAA, or ABAA vs. BABB), their abundances in many functional sites are nearly equal, suggesting different mechanism of selection of quadruplex topologies in functional and non-functional loci. Generally, these results constitute evidence of functional preference of quadruplex-forming sequences of different topology classes, and suggest that the function depends on the topology and structure of the interstrand G-quadruplex formed within the genomic DNA.

### Discussion

By integrated sequence-based prediction with results of functional studies, I have shown that sequences potentially forming interstrand G-Quadruplexes, a nucleic acid structure previously not considered in higher eukaryotic nuclear DNA, are highly prevalent in the human genome and colocalize with functionally significant loci. Enrichments of interstrand and intrastrand

PQS's in the functional studies suggest that in DNA replication interstrand G4 conformations may have serve a function similar to intrastrand quadruplexes. In transcription initiation, the role of DS-PQS is, in the light of this analysis, even more prominent than that of intrastrand quadruplexes; possibly only interstrand G-quadruplexes are involved in recruitment of transcriptional helicases. Both single-strand and double-strand PQS's should be considered in future studies of these and other functions of G4s in the nuclear DNA.

## Supporting Information

**S1 File. Source code of the AllQuads program for predicting interstrand G4-forming sequences.**

(TAR)

**S2 File. The complete list of interstrand and intrastrand quadruplex sites in the human genome (hg19), with at least three guanines per tract and loops not longer than 7nt—(tar archive of compressed text files, one per chromosome).**

(TAR)

**S1 Table. Abundance of intrastrand and interstrand G-quadruplex sequences of different topology classes in human chromosomes (separate pdf file).**

(PDF)

**S2 Table. Detailed functional analysis of intrastrand and interstrand G-quadruplex sequences in human genome (separate pdf file).**

(PDF)

**S3 Table. Relative abundances of PQS's of different topology classes in the whole genome, and in the functional regions, calculated for all PQS's and for interstrand PQS only. XPD, XPB—transcriptional helicase binding sites; ORI—origins of replication; hf2 –antibody to intrastrand PQS, upstream1000 –promoter regions. Enrichments of ratios in functional sites compared to genome-wide proportions of PQS's with different topology classes suggest that different PQS topologies may be responsible for different functions. Bottom panels—z-transformed enrichments, compared to genome-wide ratios for all PQS's and for interstrand PQS only; negative numbers denote depletion. (separate pdf file).**

(PDF)

## Acknowledgments

This study was conducted with the support of the Institute for Translational Sciences at the University of Texas Medical Branch, supported in part by a Clinical and Translational Science Award (UL1TR000071 and UL1TR001439) from the National Center for Advancing Translational Sciences.

## Author Contributions

Conceived and designed the experiments: AK. Performed the experiments: AK. Analyzed the data: AK. Contributed reagents/materials/analysis tools: AK. Wrote the paper: AK.

## References

1. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet.* 13(11):770–80. Epub 2012/10/04. doi: [10.1038/nrg3296](https://doi.org/10.1038/nrg3296) PMID: [23032257](https://pubmed.ncbi.nlm.nih.gov/23032257/); PubMed Central PMCID: [PMC3725559](https://pubmed.ncbi.nlm.nih.gov/PMC3725559/).

2. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* 43 (18):8627–37. Epub 2015/09/10. doi: gkv862 [pii] doi: [10.1093/nar/gkv862](https://doi.org/10.1093/nar/gkv862) PMID: [26350216](https://pubmed.ncbi.nlm.nih.gov/26350216/); PubMed Central PMCID: PMC4605312.
3. Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol.* 2005; 12(10):847–54. Epub 2005/09/06. doi: nsmb982 [pii] doi: [10.1038/nsmb982](https://doi.org/10.1038/nsmb982) PMID: [16142245](https://pubmed.ncbi.nlm.nih.gov/16142245/).
4. Kendrick S, Hurley LH. The role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements. *Pure Appl Chem.* 82(8):1609–21. Epub 2010/01/01. doi: [10.1351/PAC-CON-09-09-29](https://doi.org/10.1351/PAC-CON-09-09-29) PMID: [21796223](https://pubmed.ncbi.nlm.nih.gov/21796223/); PubMed Central PMCID: PMC3142959.
5. Gray LT, Vallur AC, Eddy J, Maizels N. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat Chem Biol.* 10(4):313–8. Epub 2014/03/13. doi: nchembio.1475 [pii] doi: [10.1038/nchembio.1475](https://doi.org/10.1038/nchembio.1475) PMID: [24609361](https://pubmed.ncbi.nlm.nih.gov/24609361/); PubMed Central PMCID: PMC4006364.
6. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol.* 19(8):837–44. Epub 2012/07/04. doi: nsmb.2339 [pii] doi: [10.1038/nsmb.2339](https://doi.org/10.1038/nsmb.2339) PMID: [22751019](https://pubmed.ncbi.nlm.nih.gov/22751019/).
7. Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintome C, et al. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* 33(7):732–46. Epub 2014/02/14. doi: embj.201387506 [pii] doi: [10.1002/embj.201387506](https://doi.org/10.1002/embj.201387506) PMID: [24521668](https://pubmed.ncbi.nlm.nih.gov/24521668/); PubMed Central PMCID: PMC4000090.
8. Comoglio F, Schlumpf T, Schmid V, Rohs R, Beisel C, Paro R. High-Resolution Profiling of Drosophila Replication Start Sites Reveals a DNA Shape and Chromatin Signature of Metazoan Origins. *Cell Reports.* 2015; 11(5):821–34. ISI:000353902900015. doi: [10.1016/j.celrep.2015.03.070](https://doi.org/10.1016/j.celrep.2015.03.070) PMID: [25921534](https://pubmed.ncbi.nlm.nih.gov/25921534/)
9. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature.* 1988; 334(6180):364–6. Epub 1988/07/28. doi: [10.1038/334364a0](https://doi.org/10.1038/334364a0) PMID: [33932228](https://pubmed.ncbi.nlm.nih.gov/33932228/).
10. Maizels N, Gray LT. The G4 genome. *PLoS Genet.* 9(4):e1003468. Epub 2013/05/03. doi: [10.1371/journal.pgen.1003468](https://doi.org/10.1371/journal.pgen.1003468) PGENETICS-D-13-00197 [pii]. PMID: [23637633](https://pubmed.ncbi.nlm.nih.gov/23637633/); PubMed Central PMCID: PMC3630100.
11. Tarsounas M, Tijsterman M. Genomes and G-Quadruplexes: For Better or for Worse. *J Mol Biol.* 2013; 425(23):4782–9. ISI:000328522600013. doi: [10.1016/j.jmb.2013.09.026](https://doi.org/10.1016/j.jmb.2013.09.026) PMID: [24076189](https://pubmed.ncbi.nlm.nih.gov/24076189/)
12. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Research.* 2005; 33(9):2908–16. Epub 2005/05/26. doi: 33/9/2908 [pii] doi: [10.1093/nar/gki609](https://doi.org/10.1093/nar/gki609) PMID: [15914667](https://pubmed.ncbi.nlm.nih.gov/15914667/); PubMed Central PMCID: PMC1140081.
13. Cao K, Ryykin P, Johnson FB. Computational detection and analysis of sequences with duplex-derived interstrand G-quadruplex forming potential. *Methods.* 57(1):3–10. Epub 2012/06/02. doi: S1046-2023 (12)00117-X [pii] doi: [10.1016/j.ymeth.2012.05.002](https://doi.org/10.1016/j.ymeth.2012.05.002) PMID: [22652626](https://pubmed.ncbi.nlm.nih.gov/22652626/); PubMed Central PMCID: PMC3701776.
14. Todd AK, Johnston M, Neidle S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Research.* 2005; 33(9):2901–7. Epub 2005/05/26. doi: 33/9/2901 [pii] doi: [10.1093/nar/gki553](https://doi.org/10.1093/nar/gki553) PMID: [15914666](https://pubmed.ncbi.nlm.nih.gov/15914666/); PubMed Central PMCID: PMC1140077.
15. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Research.* 2006; 34(14):3887–96. doi: [10.1093/Nar/Gkl529](https://doi.org/10.1093/Nar/Gkl529) ISI:000240583800010. PMID: [16914419](https://pubmed.ncbi.nlm.nih.gov/16914419/)
16. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research.* 2006; 34(Web Server issue):W676–82. Epub 2006/07/18. doi: 34/suppl\_2/W676 [pii] doi: [10.1093/nar/gkl253](https://doi.org/10.1093/nar/gkl253) PMID: [16845096](https://pubmed.ncbi.nlm.nih.gov/16845096/); PubMed Central PMCID: PMC1538864.
17. Wong HM, Stegle O, Rodgers S, Huppert JL. A toolbox for predicting g-quadruplex formation and stability. *J Nucleic Acids.* 2010. Epub 2010/08/21. doi: [10.4061/2010/564946](https://doi.org/10.4061/2010/564946) PMID: [20725630](https://pubmed.ncbi.nlm.nih.gov/20725630/); PubMed Central PMCID: PMC2915886.
18. Zhang R, Lin Y, Zhang CT. Greglist: a database listing potential G-quadruplex regulated genes. *Nucleic Acids Research.* 2008; 36:D372–D6. doi: [10.1093/Nar/Gkm787](https://doi.org/10.1093/Nar/Gkm787) ISI:000252545400067. PMID: [17916572](https://pubmed.ncbi.nlm.nih.gov/17916572/)
19. Lam EY, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat Commun.* 4:1796. Epub 2013/05/09. doi: ncomms2792 [pii] doi: [10.1038/ncomms2792](https://doi.org/10.1038/ncomms2792) PMID: [23653208](https://pubmed.ncbi.nlm.nih.gov/23653208/); PubMed Central PMCID: PMC3736099.
20. Beaume N, Pathak R, Yadav VK, Kota S, Misra HS, Gautam HK, et al. Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: radioresistance of D. radiodurans

- involves G4 DNA-mediated regulation. *Nucleic Acids Research*. 41(1):76–89. Epub 2012/11/20. doi: gks1071 [pii] doi: [10.1093/nar/gks1071](https://doi.org/10.1093/nar/gks1071) PMID: [23161683](https://pubmed.ncbi.nlm.nih.gov/23161683/); PubMed Central PMCID: PMC3592403.
21. Rawal P, Kummarasetti VB, Ravindran J, Kumar N, Halder K, Sharma R, et al. Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res*. 2006; 16(5):644–55. Epub 2006/05/03. doi: 16/5/644 [pii] doi: [10.1101/gr.4508806](https://doi.org/10.1101/gr.4508806) PMID: [16651665](https://pubmed.ncbi.nlm.nih.gov/16651665/); PubMed Central PMCID: PMC1457047.
  22. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research*. 2006; 34(19):5402–15. Epub 2006/10/03. doi: gkl655 [pii] doi: [10.1093/nar/gkl655](https://doi.org/10.1093/nar/gkl655) PMID: [17012276](https://pubmed.ncbi.nlm.nih.gov/17012276/); PubMed Central PMCID: PMC1636468.
  23. Andorf CM, Kopylov M, Dobbs D, Koch KE, Stroupe ME, Lawrence CJ, et al. G-quadruplex (G4) motifs in the maize (*Zea mays* L.) genome are enriched at specific locations in thousands of genes coupled to energy status, hypoxia, low sugar, and nutrient deprivation. *J Genet Genomics*. 41(12):627–47. Epub 2014/12/21. doi: S1673-8527(14)00186-6 [pii] doi: [10.1016/j.jgg.2014.10.004](https://doi.org/10.1016/j.jgg.2014.10.004) PMID: [25527104](https://pubmed.ncbi.nlm.nih.gov/25527104/).
  24. Du XJ, Gertz EM, Wojtowicz D, Zhabinskaya D, Levens D, Benham CJ, et al. Potential non-B DNA regions in the human genome are associated with higher rates of nucleotide mutation and expression variation. *Nucleic Acids Research*. 2014; 42(20):12367–79. ISI:000347693200010. doi: [10.1093/nar/gku921](https://doi.org/10.1093/nar/gku921) PMID: [25336616](https://pubmed.ncbi.nlm.nih.gov/25336616/)
  25. Nguyen GH, Tang WL, Robles AI, Beyer RP, Gray LT, Welsh JA, et al. Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111(27):9905–10. ISI:000338514800050. doi: [10.1073/pnas.1404807111](https://doi.org/10.1073/pnas.1404807111) PMID: [24958861](https://pubmed.ncbi.nlm.nih.gov/24958861/)
  26. Lexa M, Kejnovsky E, Steflava P, Konvalinova H, Vorlickova M, Vyskot B. Quadruplex-forming sequences occupy discrete regions inside plant LTR retrotransposons. *Nucleic Acids Research*. 2014; 42(2):968–78. ISI:000331138100030. doi: [10.1093/nar/gkt893](https://doi.org/10.1093/nar/gkt893) PMID: [24106085](https://pubmed.ncbi.nlm.nih.gov/24106085/)
  27. Nakken S, Rognes T, Hovig E. The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic Acids Research*. 2009; 37(17):5749–56. ISI:000271569100015. doi: [10.1093/nar/gkp590](https://doi.org/10.1093/nar/gkp590) PMID: [19617376](https://pubmed.ncbi.nlm.nih.gov/19617376/)
  28. Verma A, Halder K, Halder R, Yadav VK, Rawal P, Thakur RK, et al. Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J Med Chem*. 2008; 51(18):5641–9. ISI:000259342700018. doi: [10.1021/jm800448a](https://doi.org/10.1021/jm800448a) PMID: [18767830](https://pubmed.ncbi.nlm.nih.gov/18767830/)
  29. Qin MY, Chen ZX, Luo QC, Wen Y, Zhang NX, Jiang HL, et al. Two-Quartet G-Quadruplexes Formed by DNA Sequences Containing Four Contiguous GG Runs. *J Phys Chem B*. 2015; 119(9):3706–13. ISI:000350840600010. doi: [10.1021/jp512914t](https://doi.org/10.1021/jp512914t) PMID: [25689673](https://pubmed.ncbi.nlm.nih.gov/25689673/)
  30. Dong DW, Pereira F, Barrett SP, Kolesar JE, Cao K, Damas J, et al. Association of G-quadruplex forming sequences with human mtDNA deletion breakpoints. *BMC Genomics*. 15:677. Epub 2014/08/16. doi: 1471-2164-15-677 [pii] doi: [10.1186/1471-2164-15-677](https://doi.org/10.1186/1471-2164-15-677) PMID: [25124333](https://pubmed.ncbi.nlm.nih.gov/25124333/); PubMed Central PMCID: PMC4153896.
  31. Friedl JEF. *Mastering regular expressions: powerful techniques for Perl and other tools*. 1st ed. Cambridge; Sebastopol: O'Reilly; 1997. xxiv, 342 p. p.
  32. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26(6):841–2. doi: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) ISI:000275243500019. PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/)
  33. Comprehensive Perl Archive Network website. Available: <http://www.cpan.org>.
  34. Phan AT, Mergny JL. Human telomeric DNA: G-quadruplex, i-motif and watson-crick double helix. *Nucleic Acids Research*. 2002; 30(21):4618–25. doi: [10.1093/Nar/Gkf597](https://doi.org/10.1093/Nar/Gkf597) ISI:000179038100005. PMID: [12409451](https://pubmed.ncbi.nlm.nih.gov/12409451/)
  35. Singh RP, Blossey R, Cleri F. Structure and Mechanical Characterization of DNA i-Motif Nanowires by Molecular Dynamics Simulation. *Biophysical Journal*. 2013; 105(12):2820–31. doi: [10.1016/j.bpj.2013.10.021](https://doi.org/10.1016/j.bpj.2013.10.021) ISI:000328597400027. PMID: [24359754](https://pubmed.ncbi.nlm.nih.gov/24359754/)
  36. Xu Y, Sugiyama H. Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Research*. 2006; 34(3):949–54. doi: [10.1093/nar/gkj485](https://doi.org/10.1093/nar/gkj485) ISI:000235606200017. PMID: [16464825](https://pubmed.ncbi.nlm.nih.gov/16464825/)
  37. Day HA, Pavlou P, Waller ZAE. i-Motif DNA: Structure, stability and targeting with ligands. *Bioorganic & Medicinal Chemistry*. 2014; 22(16):4407–18. doi: [10.1016/j.bmc.2014.05.047](https://doi.org/10.1016/j.bmc.2014.05.047) ISI:000340703500009.
  38. Phan AT, Leroy JL. Intramolecular i-motif structures of telomeric DNA. *Journal of Biomolecular Structure & Dynamics*. 2000:245–51. ISI:000165410200010.
  39. Lin YC, Boone M, Meuris L, Lemmens I, Van Roy N, Soete A, et al. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications*. 2014; 5 Artn 4767. doi: [10.1038/Ncomms5767](https://doi.org/10.1038/Ncomms5767) ISI:000342927700001.

40. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99(18):11593–8. ISI:000177843100014. PMID: [12195017](#)
41. Sun DY, Pourpak A, Beetz K, Hurley LH. Direct evidence for the formation of G-quadruplex in the proximal promoter region of the RET protooncogene and its targeting with a small molecule to repress RET protooncogene transcription. *Clin Cancer Res*. 2003; 9(16):6122s–3s. ISI:000187467300218.
42. Simonsson T, Pecinka P, Kubista M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Research*. 1998; 26(5):1167–72. ISI:000072363300005. PMID: [9469822](#)
43. Zhang C, Liu HH, Zheng KW, Hao YH, Tan Z. DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix. *Nucleic Acids Research*. 2013; 41(14):7144–52. ISI:000323050700038. doi: [10.1093/nar/gkt443](#) PMID: [23716646](#)
44. Thakur RK, Kumar P, Halder K, Verma A, Kar A, Parent JL, et al. Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Research*. 2009; 37(1):172–83. ISI:000262335700015. doi: [10.1093/nar/gkn919](#) PMID: [19033359](#)
45. Sun D, Hurley LH. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J Med Chem*. 2009; 52(9):2863–74. ISI:000265911800025. doi: [10.1021/jm900055s](#) PMID: [19385599](#)
46. Raiber EA, Kranaster R, Lam E, Nikan M, Balasubramanian S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Research*. 2012; 40(4):1499–508. ISI:000301069400016. doi: [10.1093/nar/gkr882](#) PMID: [22021377](#)
47. Martin MM, Ryan M, Kim R, Zakas AL, Fu H, Lin CM, et al. Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome Res*. 21(11):1822–32. Epub 2011/08/05. doi: [gr.124644.111](#) [pii] doi: [10.1101/gr.124644.111](#) PMID: [21813623](#); PubMed Central PMCID: PMC3205567.
48. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature Biotechnology*. 2015; 33(8):877–+. doi: [10.1038/nbt.3295](#) ISI:000359274900028. PMID: [26192317](#)