

ARTICLE

Received 17 Sep 2012 | Accepted 22 Mar 2013 | Published 30 Apr 2013

DOI: 10.1038/ncomms2792

G-quadruplex structures are stable and detectable in human genomic DNA

Enid Yi Ni Lam¹, Dario Beraldi¹, David Tannahill¹ & Shankar Balasubramanian^{1,2,3}

The G-quadruplex is an alternative DNA structural motif that is considered to be functionally important in the mammalian genome for transcriptional regulation, DNA replication and genome stability, but the nature and distribution of G-quadruplexes across the genome remains elusive. Here, we address the hypothesis that G-quadruplex structures exist within double-stranded genomic DNA and can be explicitly identified using a G-quadruplex-specific probe. An engineered antibody is employed to enrich for DNA containing G-quadruplex structures, followed by deep sequencing to detect and map G-quadruplexes at high resolution in genomic DNA from human breast adenocarcinoma cells. Our high sensitivity structure-based pull-down strategy enables the isolation of genomic DNA fragments bearing single, as well as multiple G-quadruplex structures. Stable G-quadruplex structures are found in sub-telomeres, gene bodies and gene regulatory regions. For a sample of identified target genes, we show that G-quadruplex-stabilizing ligands can modulate transcription. These results confirm the existence of G-quadruplex structures and their persistence in human genomic DNA.

¹Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ²The University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. ³School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK. Correspondence and requests for materials should be addressed to S.B. (email: sb10031@cam.ac.uk).

Guanine-rich DNA can form four-stranded DNA structures called G-quadruplexes¹. Genome-wide bioinformatics analyses show that G-quadruplex sequence motifs are prevalent in the human genome and are enriched in gene regulatory regions and gene bodies, and in repetitive sequences, such as telomeres^{2–5}. Such studies have sparked the need to map folded G-quadruplex structures carried by the genome in an explicit way. A number of studies have linked G-quadruplexes to key biological processes such as transcriptional regulation, DNA replication and genome stability, leading to their exploration as therapeutic targets^{6,7}. G-quadruplexes are stable under near-physiological conditions *in vitro*, and a central challenge in the field has been to establish whether such structures can form in genomic DNA. Support for G-quadruplex formation has emerged from the use of small molecule G-quadruplex-targeting ligands to perturb cellular function⁸. Such small molecules have been shown to localize at telomeres⁹ and have been used to enrich for human telomeric DNA¹⁰. Engineered proteins, including the hf2 single-chain antibody¹¹ and the Gq1 zinc finger protein¹² have been generated as molecular probes that are exquisitely G-quadruplex-structure specific. Indeed, a G-quadruplex antibody was previously employed to visualize G-quadruplex formation at the millions of telomeres present in the macronuclei of ciliates¹³ and in human cells¹⁴. Here, we sought to explicitly demonstrate that G-quadruplex structures exist within double-stranded genomic DNA using a G-quadruplex-specific probe. We report the presence and localization of stable G-quadruplex structures in genomic DNA isolated from human cells.

Results

The hf2 antibody pulls down DNA G-quadruplex structures.

We employed a G-quadruplex-specific antibody to enrich for genomic DNA fragments containing folded G-quadruplex structures, followed by deep sequencing of the isolated DNA to identify the regions comprising G-quadruplex structures. We used the single-chain antibody, hf2, which we previously showed binds to a range of folded DNA G-quadruplex structures formed by synthetic oligonucleotides, but with negligible binding to single- and double-stranded DNA¹¹. After re-affirming the binding specificity of hf2 by ELISA (Supplementary Fig. S1), we confirmed that in solution hf2 could pull-down DNA G-quadruplex structures formed by synthetic DNA oligonucleotides (Supplementary Figs S2 and S3). The specificity of hf2 to enrich G-quadruplex structures was then confirmed using mixtures of G-quadruplex and non-G-quadruplex single-stranded or double-stranded DNA oligonucleotides. The gel shown in Fig. 1a (Supplementary Fig. S3) confirms the specific capture of G-quadruplex-structured DNA (KIT-2) from such mixtures by hf2 without significant contaminating non-G-quadruplex DNA. We also showed that hf2 could specifically capture the telomeric G-quadruplex oligonucleotide, Htelo, in the presence of a 100-fold excess of double-stranded salmon sperm DNA (Fig. 1a), confirming the probe specificity and robust conditions for seeking such non-canonical structures from DNA of genomic complexity.

Genome-wide pull-down of G-quadruplexes. We then undertook to detect G-quadruplex structures in double-stranded genomic DNA isolated from human breast adenocarcinoma (MCF7) cells using the hf2 antibody as outlined schematically in Fig. 1b. If stable G-quadruplex structures exist within double-stranded genomic DNA, we reasoned that hf2 would selectively enrich such regions, leading to peaks in DNA sequencing that correspond to G-quadruplex motifs. The hf2 antibody probe was first immobilized onto protein-A Dynabeads then incubated with

sonicated human genomic DNA, followed by elution of the captured DNA fragments and subsequent deep sequencing. Pull-down conditions were optimized by assessing telomere enrichment, as there is already substantial evidence for G-quadruplex structure formation at telomeres^{13,15}. Selective qRT-PCR analysis of the repetitive telomeric sequences¹⁶ compared with a region of the oestrogen receptor enhancer lacking predicted G-quadruplexes, showed that telomeric sequences were enriched by >20-fold in the pull-down DNA relative to control, providing proof-of-principle of the overall approach. Further controls confirmed that the buffer conditions used during genomic DNA isolation and sonication also do not induce or destroy G-quadruplex structures (Supplementary Figs S4 and S5). The hf2-enriched genomic DNA library was then sequenced at high depth to identify DNA fragments containing a G-quadruplex structure. Sequencing reads from four independent libraries were aligned to the human genome and peaks called using the Model-based Analysis of ChIP-Seq (MACS) algorithm¹⁷. To identify enriched regions with high confidence and assess the reproducibility of the enrichment, we computed the quality metrics recently developed to assess ENCODE data¹⁸ (Supplementary Table S1, Supplementary Fig. S6). The quality metrics of the G-quadruplex pull-down libraries were generally within the ENCODE acceptable thresholds, and are comparable to transcription factor ChIP-Seq data.

Confirmation of G-quadruplex structures in identified peaks.

The sequence of the peaks across the genome was then examined for predicted G-quadruplex sequences by the G4-calculator algorithm¹⁹. The G4-calculator analyzes a fixed width sequence window for G-quadruplex-forming potential, defined as at least four runs of three or more guanines within 100 bases. The proportion of windows with G-quadruplex-forming potential is then computed for the entire length of the peak. The number of peaks with G-quadruplex-forming potential was significantly higher than expected by chance (Supplementary Table S2, Supplementary Figs S7 and S8). To identify consistently observed peaks, we considered only those in common between at least two of the four libraries. Of the enriched regions, the majority (568/768, 74.0%) had G-quadruplex-forming potential, and this compares favourably with the proportion of motif-containing peaks typically seen in ChIP-Seq for transcription factors, such as NRSF²⁰. Although G4-calculator indicates which peaks have G-quadruplex-forming potential, it does not specify their precise position within the peak. To accurately map the genomic location of predicted G-quadruplexes within the peaks, we therefore used an alternative G-quadruplex prediction algorithm, quadparser². Quadparser uses a more stringent consensus ($G_{3+} N_{1-7} G_{3+} N_{1-7} G_{3+} N_{1-7} G_{3+}$) by constraining loop lengths of the G-quadruplex to a maximum of seven bases. The number of peaks having a predicted G-quadruplex computed by quadparser was also found to be statistically significant (Supplementary Table S3, Supplementary Figs S9 and S10), giving 175 predicted G-quadruplex-containing peaks (Fig. 2, Supplementary Fig. S11, Supplementary Table S4).

As an independent evaluation of the binding specificity of hf2, we analysed the combined sequence reads from all libraries using the motif-finding algorithm, MEME²¹. This approach makes no *a priori* assumptions of the sequence types expected. Analysis of the top 200 peaks by enrichment over input showed that the most frequent MEME sequence motif calculated matches the G-quadruplex consensus (Fig. 3a, Supplementary Fig. S12), and is thus consistent with the enrichment of potential G-quadruplex structures by our pull-down strategy. When MEME was used on the 200 most enriched peaks called in the input library or 200

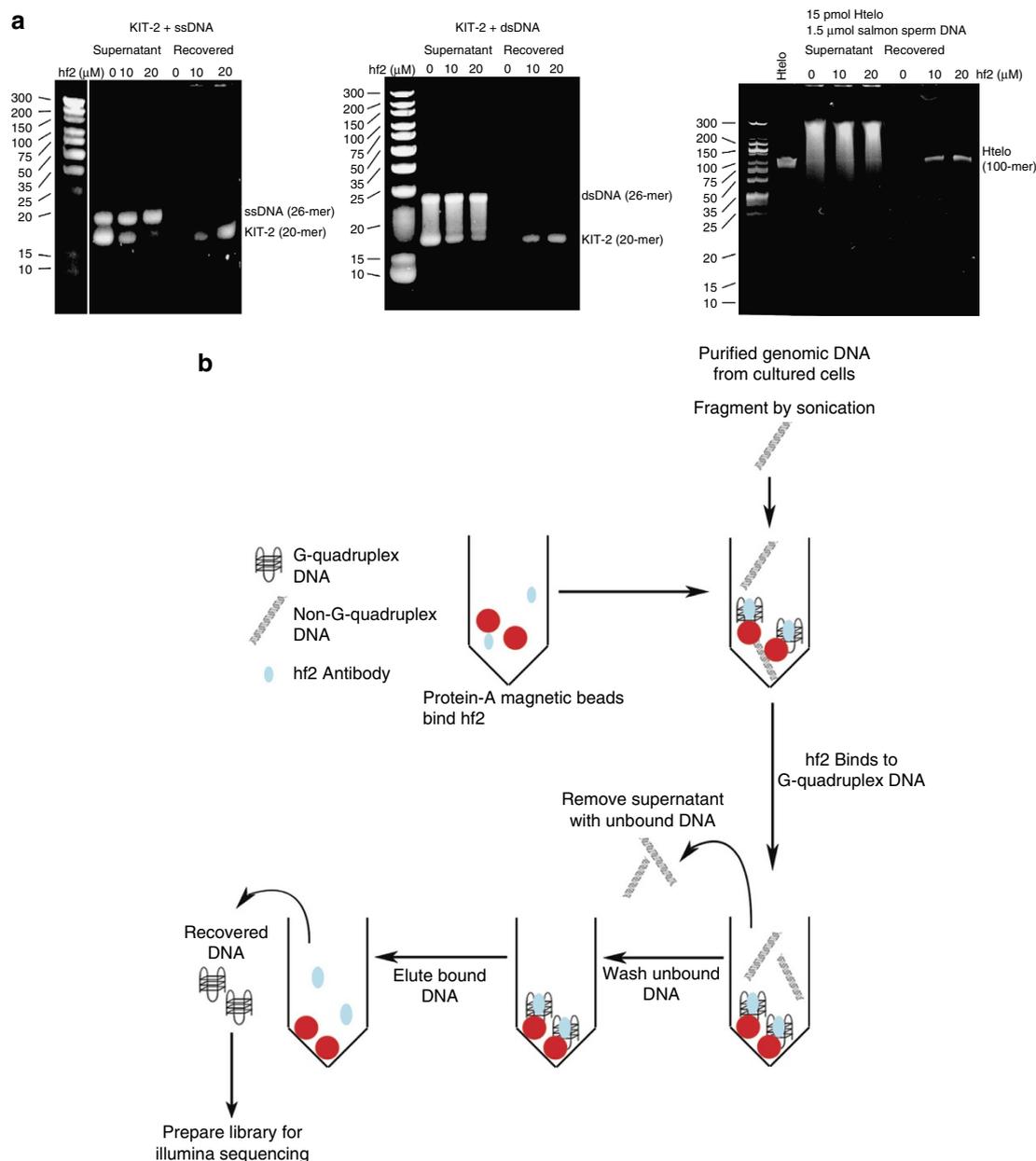


Figure 1 | The hf2 single-chain antibody specifically pulls down G-quadruplex oligonucleotides. (a) Pull-down of G-quadruplex oligonucleotides by hf2 analysed on TBE-urea gels. Left, KIT-2 G-quadruplex oligonucleotides but not single-stranded DNA are captured by hf2. Lane 1 shows the GeneRuler Ultra Low Range DNA Ladder (Thermo Scientific), lanes 2–4 show concentration-dependent depletion of KIT-2 quadruplex, but not single-stranded DNA (control) from the supernatant by hf2, lanes 5–7 show the specific recovery of KIT-2 quadruplex, but not single-stranded DNA with increasing hf2 concentration. Middle, KIT-2 G-quadruplex oligonucleotides but not double-stranded DNA are captured by hf2. Lane 1 shows the GeneRuler Ultra Low Range DNA Ladder, lanes 2–4 show concentration-dependent depletion of KIT-2 quadruplex, but not double-stranded DNA (control) from the supernatant by hf2, lanes 5–7 show the specific recovery of KIT-2 quadruplex, but not double-stranded DNA with increasing hf2 concentration. Right, Htelo G-quadruplex oligonucleotides are captured by hf2 in the presence of excess sonicated salmon sperm DNA. Lane 1 shows the Htelo oligonucleotide alone, lanes 2–4 show the unbound supernatant with different hf2 concentrations, lanes 5–7 show the specific recovery of KIT-2 quadruplex, but not double-stranded salmon sperm DNA. (b) Pull-down protocol used to isolate G-quadruplex DNA from genomic DNA with the hf2 antibody.

random sequences from the genome, similar motifs were not observed (Supplementary Fig. S12). As G-quadruplexes display characteristic circular dichroism (CD) spectroscopic signatures indicative of their structure²², we determined the structural characteristics of oligonucleotides with G-quadruplex-forming potential covering a set of pull-down peaks. Parallel G-quadruplexes display positive and negative peaks at 260 and 240 nm, respectively, while anti-parallel G-quadruplexes exhibit positive and negative peaks at 295 and 260 nm²³. We analysed

the CD spectra of a series of 44 non-overlapping oligonucleotides spanning all of the G-repeats, from two sub-telomeric peaks and two peaks elsewhere in the genome (Supplementary Table S5). Forty-two showed CD spectra with a peak at 295 nm. These spectra are consistent with the majority of the sequences folding into either a hybrid-type G-quadruplex structure with mixed parallel/anti-parallel strands or a mixture of parallel and anti-parallel G-quadruplexes (Fig. 3b, Supplementary Fig. S13).

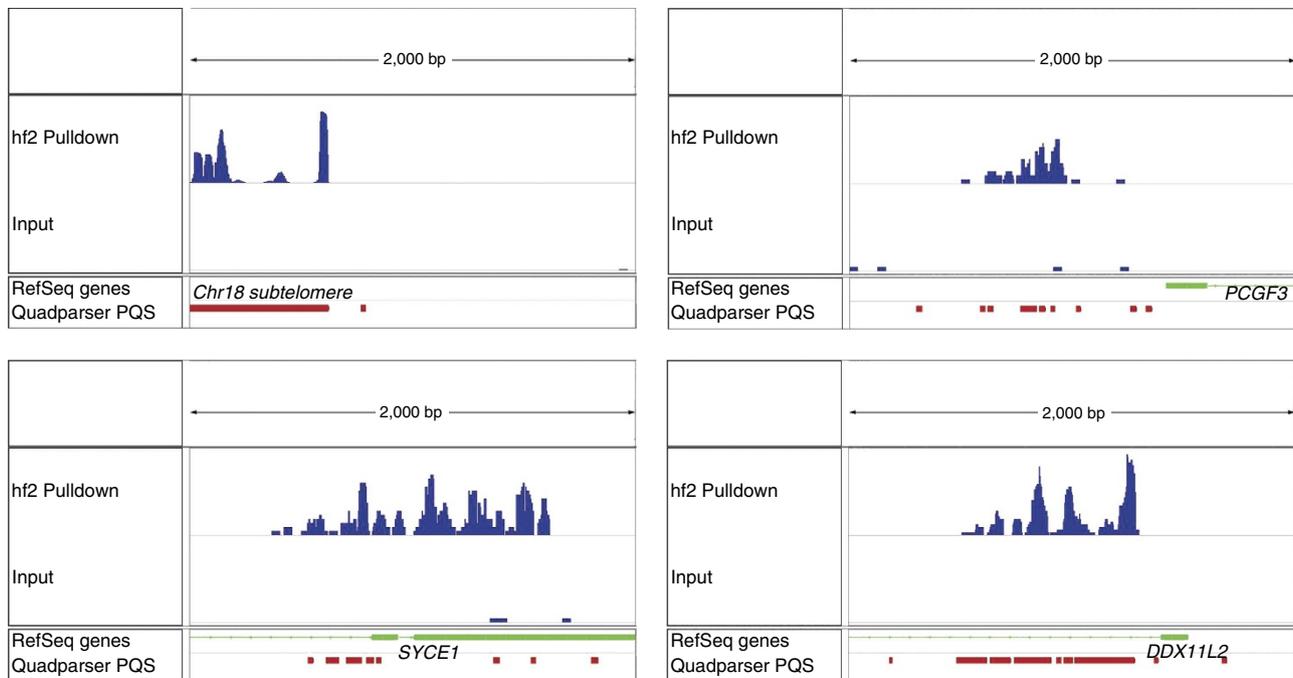


Figure 2 | Peaks identified by deep sequencing after pull-down with the anti-G-quadruplex hf2 antibody. Genome browser view of four peaks (blue) present compared with input and the overlap with G-quadruplex sequences predicted by quadparser (red). RefSeq gene is shown in green. The peaks map to different chromosomal locations including the sub-telomere (top left), gene promoter (top right), exon (bottom left) and intron (bottom right).

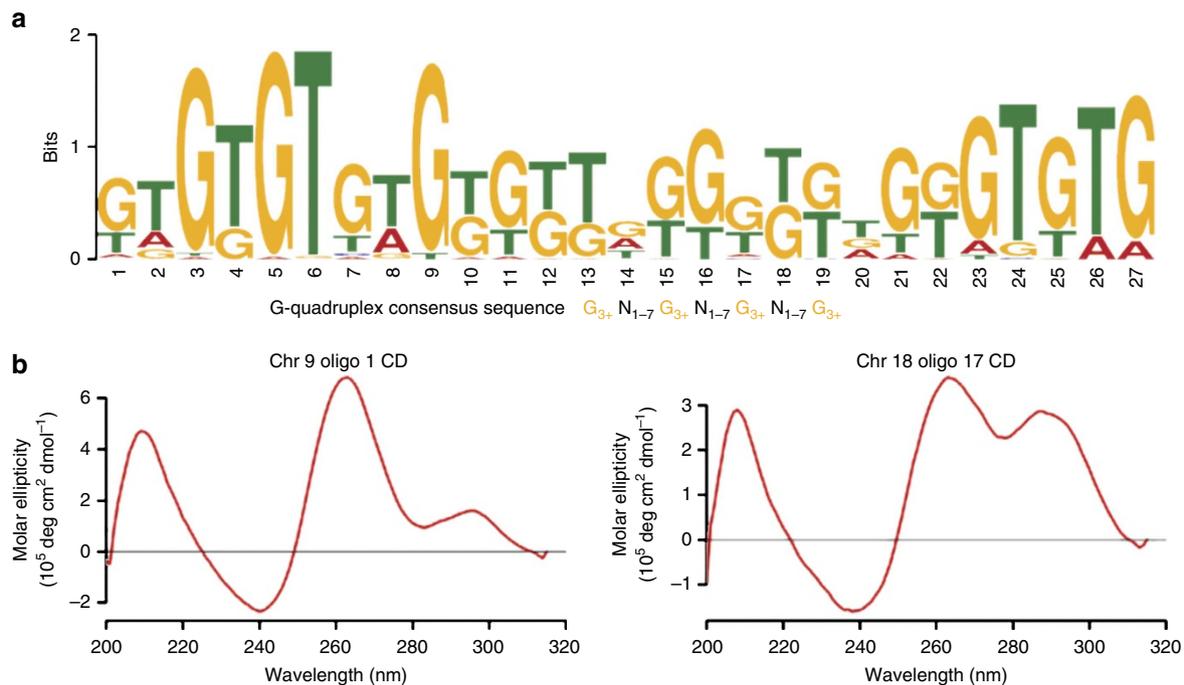


Figure 3 | Motif and CD analyses substantiate G-quadruplex identification. (a) Sequence logo of the most enriched motif as determined by MEME²¹ (e -value = $1.9e^{-190}$, expected value from MEME expectation maximization algorithm) found in the top 200 peaks ranked by enrichment (false discovery rate < 0.05). The G-quadruplex consensus sequence is shown here for comparison. (b) Examples of CD spectra for two oligonucleotides from the identified peaks. Parallel G-quadruplexes display a characteristic peak at 263 nm and a trough at 240 nm, whereas anti-parallel G-quadruplexes show a peak at 295 nm. The CD spectra show characteristics of both parallel and anti-parallel G-quadruplexes indicative of hybrid-type G-quadruplexes or a mixture of parallel and anti-parallel G-quadruplexes.

Regulation of identified genes by a G-quadruplex ligand. Having proven the presence of G-quadruplex structures in defined locations within genomic DNA, we next investigated the

consequence of G-quadruplex formation at some of these explicit sites in cells. We selected an example set of eight genes (*ABCG1*, *ACTN1*, *DYSF*, *ELL*, *LRP1*, *PVT1*, *STARD8* and *TOM1*), each

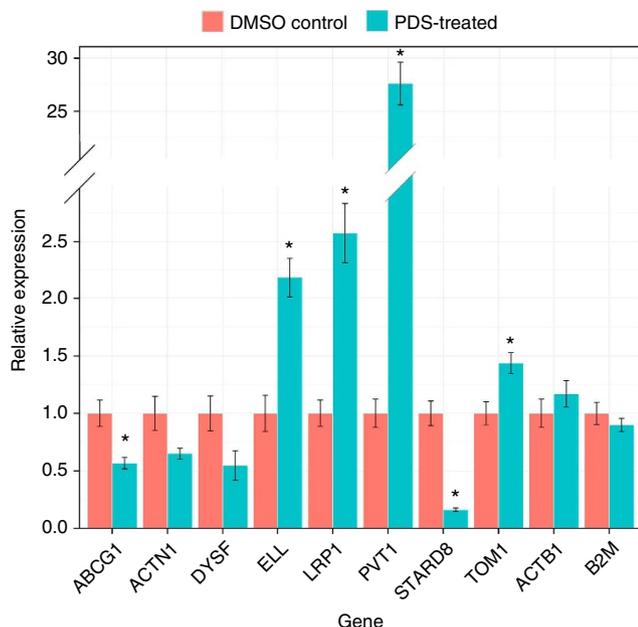


Figure 4 | Stabilizing small molecules modulate the expression of identified G-quadruplex-containing genes. qRT-PCR was used to examine the expression of selected genes, identified by hf2 pull-down to contain G-quadruplexes, in MCF7 cells that were treated in triplicate with the G-quadruplex-specific small molecule PDS or DMSO control. The mean and s.e. (error bars) of the relative expression levels of genes in PDS-treated and DMSO control are plotted on two different scales to show the different magnitudes of changes. Two genes in particular, *PVT1* and *STARD8*, show large changes in gene expression in PDS-treated cells compared with controls. The Student's *t*-test was used to calculate statistical significance between PDS-treated and control cells. Asterisks indicate statistically significant changes in gene expression with $P < 0.05$. The *P*-values for *ABCG1*, *ACTN1*, *DYSF*, *ELL*, *LRP1*, *PVT1*, *STARD8*, *TOM1*, *ACTB1* and *B2M* are 0.0251, 0.0868, 0.0840, 0.0069, 0.0051, 0.0002, 0.0015, 0.0323, 0.3720 and 0.4189.

with G-quadruplex structures, identified by hf2 enrichment, within 1 kb of the transcribed regions of the gene, in at least three libraries. G-quadruplex formation has been linked to the regulation of transcription and thus G-quadruplex stabilization by a ligand would be predicted to modulate transcription as has been exemplified in other studies^{24,25}. We treated MCF7 cells with the highly specific G-quadruplex-stabilizing ligand, pyridostatin (PDS)²⁶, and assessed changes in expression of the selected genes by qRT-PCR. Gene expression was normalized to the housekeeping gene *RPLP0*, which contains no predicted G-quadruplexes and does not show a peak with hf2 pull-down. Nested analysis of variance showed PDS treatment caused a significant change in the expression of the selected genes ($P = 4.9 \times 10^{-10}$). Six of the G-quadruplex-containing genes analysed had statistically significant changes in gene expression ($P < 0.05$, Student's *t*-test) caused by ligand treatment, the remaining two genes also showed downregulation (P -value < 0.09 , Student's *t*-test) (Fig. 4), while two control genes, *ACTB1* and *B2M*, that were not enriched by hf2, showed no changes in gene expression.

Discussion

The work presented here investigates the hypothesis that G-quadruplex structures are present and stable in human DNA. Using a structure-specific antibody, we have proven that

G-quadruplex structures indeed persist in genomic DNA isolated from human cancer cells. Furthermore, we have mapped their locations in the genome at high resolution using deep sequencing.

ChIP-Seq is a widely used method to map transcription factors and chromatin-associated protein binding sites in the genome, and metrics to assess data quality and analysis are becoming accepted¹⁸. Our G-quadruplex genomic DNA mapping approach has significant differences from the standard ChIP-Seq method for point-source peaks (for example, transcription factors). Therefore, we propose a workflow for G-quadruplex DNA sequencing experiment. First, before sequencing we recommend quantifying the telomere G-quadruplex enrichment by qRT-PCR, which should be preferably greater than 20-fold. Second, we suggest assessing the quality of G-quadruplex enriched libraries according to the ENCODE guidelines¹⁸ as exemplified by Supplementary Table S1. Third, a statistically significant proportion of the identified peaks should contain putative G-quadruplex sequences (as in Supplementary Figs S6 and S8). In the G-quadruplex mapping method, most peaks are only present in one library. This is due to the small number of *bona fide* peaks compared with the very large number of peaks that peak-calling algorithms generally produce, thus replicates are essential to identify consistently called peaks. We therefore recommend sequencing at least three independent biological replicates, and performing peak-calling with an established ChIP-Seq peak-caller, such as MACS. We suggest that only peaks below a *P*-value cutoff of 10^{-5} (from MACS) and present in at least two libraries be used for further analyses.

Computational methods have previously identified $> 370,000$ individual sequence motifs in the human genome with the potential to form a G-quadruplex². Such studies have raised the need to elucidate which of these potential structures actually form in the genome. Our work has definitively identified several genomic DNA regions that stably retain a folded G-quadruplex structure even after DNA isolation procedures. This particular study does not define an exhaustive list of all G-quadruplex structures that will form in the genome for a number of reasons. First, the hf2 antibody used in our experiments was selected against a particular G-quadruplex¹¹, and although it does bind to several other G-quadruplex sequences, it is unlikely to recognize all G-quadruplex folds and sequences with equal affinity. Second, we hypothesize that the formation of some G-quadruplexes will be temporally coupled to functional processes, such as transcription and replication, thus structure formation may only be transient¹⁴. For example, helicases present in cells, such as PIF1, BLM, WRN and FANCD1, are known to resolve G-quadruplex structures at replication forks^{27–30} and may be involved in controlling the lifetimes of such structure. Furthermore, the changing torsional stress in regions of the genome is also likely to have influence on the formation of such alternative structures during transcriptional events^{31,32}.

In our study, we have found examples of stable G-quadruplexes within 2 kb upstream of the transcriptional start site of several genes. This finding is in keeping with the predictions from bioinformatics studies that many of protein-coding genes may harbour a G-quadruplex in their upstream promoter regions^{19,33}. Extensive biophysical and cellular experiments have previously lent support for G-quadruplex formation in the promoters of specific genes including *KIT*³⁴, *BCL2*³⁵, *VEGF*³⁶, *MYC*³⁷ and *KRAS*³⁸. As we have provided evidence for the existence of stable G-quadruplexes in the promoter of a set of previously uninvestigated genes, these may prove to be valuable new targets for the exploration of G-quadruplex-mediated transcriptional regulation.

Informatics analyses have also highlighted an enrichment of G-quadruplex-forming potential in 5'-UTRs, first introns and

Pulldown of genomic DNA. Fifty microliter of Protein A Dynabeads were washed three times with PBS then incubated with 2.5 µg of hf2 in 0.5% BSA and 1 mg ml⁻¹ yeast tRNA in PBS overnight rotating at 4 °C. Beads were washed three times with 0.5% BSA then incubated with 400 µl of 300 ng µl⁻¹ sonicated genomic DNA. Following overnight incubation rotating at 4 °C, beads were washed six times with 10 mM Tris pH 7.4, 100 mM KCl, 0.1% tween then once with 10 mM Tris pH 7.4, 100 mM KCl. Bound DNA was eluted in 50 µl of 1% SDS, 0.1 M NaHCO₃ at 30 °C for 1 h then purified with Roche PCR purification columns. Recovered DNA was assessed for enrichment of telomeric sequences by qPCR with primers for telomeric repeats and a control genomic region in a regulatory region of the oestrogen receptor gene without predicted G-quadruplex sequences. Recovered DNA was used to prepare libraries for Illumina sequencing using the TruSeq DNA library kit. Library quality was assessed using an Agilent Bioanalyzer 2100 before single end 36 base sequencing on the Illumina MiSeq, GAIIX or HiSeq 2000.

Sequences of oligonucleotides for PCR. tel 1 GGTTTTGGAGGGTGAGGGTGAGGGTGAGGGTGAGGGT; tel 2 TCCGACTATCCCTATCCCTATCCCTATCCCTATCCCTA; ESR Fw GAAACAGCCCCAAATCTCAA; ESR Rv TTGTA GCCAGCAAGCAAATG.

Analysis of Illumina sequencing data. Four independent pull-down and one input control libraries were analysed. Fastq files generated by the Illumina pipeline (CASAVA 1.7 and OLB 1.9.4) were aligned against the Human Reference Genome (assembly hg18, NCBI build 36.1, March 2008) using bwa⁵⁰ with default parameters. Reads not assigned to a single genomic position (that is, with MAPQ < 15) were discarded. Reads overlapping regions with unusually large numbers of reads independent of the antibody used were also discarded (<http://hgdownload-test.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeMapability/wgEncodeDukeRegionsExcluded.bed6.gz>). Quality metrics for the four libraries sequence are summarized in Supplementary Table S1. To minimize biases in the detection of enriched regions caused by unequal library sizes⁵¹, libraries were down-sampled to eight million reads using Picard/DownsampleSam (<http://picard.sourceforge.net/command-line-overview.shtml#DownsampleSam>). Peak-calling was performed using MACS 1.4¹⁷ with default settings and the input library as control. The motif discovery tool MEME⁵² was used to identify common sequence motifs in the peaks.

Circular dichroism. Oligonucleotides (10 µM) for sequences predicted to form G-quadruplexes present in the peaks (Supplementary Table S4) were annealed in 10 mM Lithium cacodylate pH 7.2, 1 mM EDTA and 100 mM KCl by heating to 95 °C for 10 min followed by slow cooling to room temperature overnight. The CD spectra were measured on a Chirascan spectropolarimeter in a quartz cuvette with a 1-mm optical path length. Three scans were obtained from 200 to 315 nm at 25 °C for each sample with a step size of 1 nm, a time per point of 1 s and a bandwidth of 0.5 nm. The scans were averaged and the spectrum obtained with a buffer only sample was subtracted with zero-correction at 315 nm.

Pyridostatin treatment and qRT-PCR. MCF7 cells (2 × 10⁵ per well) were plated in six-well plates and cultured with 10 µM PDS or 0.1% DMSO for 24 h. Total RNA was isolated using the RNeasy mini kit (Qiagen, Crawley, UK) and 2 µg RNA used for cDNA synthesis using the Maxima reverse transcriptase (Fermentas) with random hexamers following the manufacturers' instructions. Quantitative real-time PCR (qRT-PCR) was performed using Fast SYBR PCR mix (Applied Biosystems, UK), with a BioRad CFX96 quantitative PCR machine. Cycling conditions were 95 °C for 20 s followed by 40 cycles of 3 s at 95 °C and 30 s at 60 °C. Pyridostatin-treated and control samples were analysed in triplicate and the results analysed with the BioRad CFX software.

Sequences of oligonucleotides used for qRT-PCR. ABCG1 Forward TCAGGG ACCTTTCCTATTCC; ABCG1 Reverse TTCCTTTCAGGAGGGTCTTGT; ACTB1 Forward Qiagen primer set QT00193473; ACTB1 Reverse Qiagen primer set QT00193473; ACTN1 Forward GGGTTATGATATTGGCAACGA; ACTN1 Reverse TTGGGGTCCACAATGCTC; B2M Forward Qiagen primer set QT00088935; B2M Reverse Qiagen primer set QT00088935; DYF Forward TTCAAAGCCTCAGACTTGG; DYF Reverse GGGACTGCCATAGAGG TTGA; ELL Forward CCGAAGTGCCATTGTGCATC; ELL Reverse CCGAACTG AACCTTCTTGG; LRP1 Forward CGCTGCATCAACACTCATGG; LRP1 Reverse AACGGTTCCTCGTCCAGTCAC; PVT1 Forward AGAATCCGTGTCTGGG AGAA; PVT1 Reverse TCCCCTTAATAGTTGGCTTCC; RPLP0 Forward CCT CGTGGAAGTGACACTGT; RPLP0 Reverse CTGTCTCCCTGGGCATCAC; STARD8 Forward GCCTCTTTTACGCTCCTCC; STARD8 Reverse TGGGAAG CACTTCACCTTCC; TOM1 Forward TGATGCTGGCTCTCACAGTC; TOM1 Reverse GGTCTCACACACTCT.

References

- Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA* **48**, 2013–2018 (1962).

- Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
- Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901–2907 (2005).
- Verma, A. *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.* **51**, 5641–5649 (2008).
- Eddy, J. & Maizels, N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* **36**, 1321–1333 (2008).
- Tauchi, T. *et al.* Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia. *Oncogene* **25**, 5719–5725 (2006).
- Drygin, D. *et al.* Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* **69**, 7653–7661 (2009).
- Balasubramanian, S. & Neidle, S. G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **13**, 345–353 (2009).
- Granotier, C. *et al.* Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.* **33**, 4182–4190 (2005).
- Müller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated G-quadruplex isolation from human cells. *Nat. Chem.* **2**, 1095–1098 (2010).
- Fernando, H., Rodriguez, R. & Balasubramanian, S. Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry* **47**, 9365–9371 (2008).
- Isalan, M., Patel, S. D., Balasubramanian, S. & Choo, Y. Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. *Biochemistry* **40**, 830–836 (2001).
- Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylochyia lemnae* macronuclei. *Proc. Natl Acad. Sci. USA* **98**, 8572–8577 (2001).
- Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative Visualization of DNA G-quadruplex Structures in Human Cells. *Nat. Chem.* **5**, 182–186 (2013).
- Tang, J. *et al.* G-quadruplex preferentially forms at the very 3' end of vertebrate telomeric DNA. *Nucleic Acids Res.* **36**, 1200–1208 (2008).
- Cawthon, R. M. Telomere measurement by quantitative PCR. *Nucleic Acids Res.* **30**, e47 (2002).
- Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
- Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
- Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887–3896 (2006).
- Wilbanks, E. G. & Facciotti, M. T. Evaluation of Algorithm Performance in ChIP-Seq Peak Detection. *PLoS ONE* **5**, e11471 (2010).
- Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
- Karsisiotis, A. I. *et al.* Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew Chem. Int. Ed. Engl.* **50**, 10645–10648 (2011).
- Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M. & Sasisekharan, V. Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* **20**, 4061–4067 (1992).
- Broxson, C., Beckett, J. & Tornaletti, S. Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myc gene. *Biochemistry* **50**, 4162–4172 (2011).
- Tornaletti, S., Park-Snyder, S. & Hanawalt, P. C. G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J. Biol. Chem.* **283**, 12756–12762 (2008).
- Rodriguez, R. *et al.* A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **130**, 15758–15759 (2008).
- Sanders, C. M. Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem. J.* **430**, 119–128 (2010).
- Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.* **273**, 27587–27592 (1998).
- Fry, M. & Loeb, L. A. Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.* **274**, 12797–12802 (1999).
- London, T. B. *et al.* FANCI is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.* **283**, 36132–36139 (2008).
- Kouzine, F., Liu, J., Sanford, S., Chung, H. J. & Levens, D. The dynamic response of upstream DNA to transcription-generated torsional stress. *Nat. Struct. Mol. Biol.* **11**, 1092–1100 (2004).

32. Sun, D. & Hurley, L. H. 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.* **52**, 2863–2874 (2009).
33. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406–413 (2007).
34. Rankin, S. *et al.* Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **127**, 10584–10589 (2005).
35. Dai, J., Chen, D., Jones, R. A., Hurley, L. H. & Yang, D. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **34**, 5133–5144 (2006).
36. Sun, D., Guo, K., Rusche, J. J. & Hurley, L. H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **33**, 6070–6080 (2005).
37. Simonsson, T., Pecinka, P. & Kubista, M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* **26**, 1167–1172 (1998).
38. Cogoi, S. & Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **34**, 2536–2549 (2006).
39. Huppert, J. L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* **36**, 6260–6268 (2008).
40. Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).
41. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
42. Rogakou, E. P., Boon, C., Redon, C. & Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J. Cell Biol.* **146**, 905–916 (1999).
43. Smith, J. S. *et al.* Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat. Struct. Mol. Biol.* **18**, 478–485 (2011).
44. Zaug, A. J., Podell, E. R. & Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl Acad. Sci. USA* **102**, 10864–10869 (2005).
45. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**, 718–720 (1991).
46. Law, M. J. *et al.* ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143**, 367–378 (2010).
47. Shirude, P. S., Okumus, B., Ying, L., Ha, T. & Balasubramanian, S. Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-kit. *J. Am. Chem. Soc.* **129**, 7484–7485 (2007).
48. Deng, H. & Braunlin, W. H. Duplex to quadruplex equilibrium of the self-complementary oligonucleotide d(GGGGCCCC). *Biopolymers* **35**, 677–681 (1995).
49. Kumar, N., Sahoo, B., Varun, K. A., Maiti, S. & Maiti, S. Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic Acids Res.* **36**, 4433–4442 (2008).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
51. Chen, Y. *et al.* Systematic evaluation of factors influencing ChIP-seq fidelity. *Nat. Meth.* **9**, 609–614 (2012).
52. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).

Acknowledgements

We thank the CRI Genomics core for their support with Illumina sequencing; Rory Stark and Ros Russell from the CRI Bioinformatics core for their assistance with bioinformatic analysis; Pierre Murat for his assistance with the circular dichroism analysis; Raphaël Rodriguez for his helpful comments on the manuscript. The Balasubramanian group is supported by core funding from Cancer Research UK.

Author contributions

E.L., D.T. and S.B. designed the study; E.L. performed the experiments; D.B. and E.L. performed the bioinformatic analyzes; E.L., D.B., D.T. and S.B. wrote the manuscript.

Additional information

Accession codes: Sequencing and processed data have been submitted to the NCBI Gene Expression Omnibus under accession code GSE45241.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Lam, E. Y. N. *et al.* G-Quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* **4**:1796 doi: 10.1038/ncomms2792 (2013).