

Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression

Himesh Fernando¹, Sven Sewitz¹, Jeremy Darot², Simon Tavaré², Julian Leon Huppert³ and Shankar Balasubramanian^{1,4,*}

¹Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, ²CRUK Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, ³Cavendish Laboratory, Cambridge CB3 0HE and ⁴School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK

Received July 17, 2009; Revised August 20, 2009; Accepted August 22, 2009

ABSTRACT

G-quadruplex nucleic acids have been proposed to play a role in a number of fundamental biological processes that include transcription and translation. We have developed a single-chain antibody that is selective for G-quadruplex DNA over double-stranded DNA, and here show that when it is expressed in human cells, it significantly affects the expression of a wide variety of genes, in a manner that correlates with the presence of predicted G-quadruplexes. We observe cases where gene expression is increased or decreased, and that there are apparent interactions with G-quadruplex motifs at the beginning and end of the genes, and on either strand. The outcomes of this genome-wide study demonstrate that G-quadruplex recognition by the antibody has physiological consequences, and provides insights into some of the complexity associated with G-quadruplex-based regulation.

INTRODUCTION

G-quadruplexes are highly stable alternative DNA structures formed by tetrads of guanines that interact *via* Hoogsteen hydrogen bonds and are stabilized by monovalent cations (1–4). A number of conformationally diverse G-quadruplex structures have been determined by NMR or X-ray crystallography (5–8). G-quadruplexes have been shown to occur at telomeres (9,10), and have potential as a target class for therapeutics (11–13). Algorithms have been developed to predict which genomic sequences will form G-quadruplexes (14), and many have been predicted to form in gene

promoters (15), a number of which have been the subject of detailed investigation (16). These include oncogenes such as *c-myc* (17), *c-kit* (18,19) and *k-ras* (20), where G-quadruplex formation has been shown to control gene transcription levels. Further support for the G-quadruplex-promoter hypothesis has been provided by genomic studies which have shown a link between the presence of a G-quadruplex motif and expression levels of the associated genes (21). The detailed molecular mechanism(s) by which DNA G-quadruplexes could influence gene expression is the subject of investigation by a number of research groups. One model is that G-quadruplex formation interferes with protein-DNA interactions (22). However, it is also conceivable that more than one G-quadruplex-related mechanism may operate, depending on the specific gene in question. There is evidence that gene expression can be perturbed by small molecules (13,23) or proteins (24,25) that target G-quadruplex nucleic acids, and indeed there are many natural proteins that interact specifically with G-quadruplexes (22), suggestive of natural G-quadruplex-related mechanisms.

G-quadruplex DNA structures have been found computationally in many positions in the genome considered to be of importance for gene regulation. For promoter quadruplexes (i.e. near to the TSS), there are numerous experimental studies on individual cases to support their structure and formation *in vitro* (3,8,26–28) and that their formation may influence gene regulation in cells (17,20,29).

Recent attempts to elucidate the biological function of DNA G-quadruplexes in promoter regions have been focused on the use of small molecule ligands, selected for their ability to bind and stabilize or disrupt these structures (20,29,30). There have been experiments described in the literature that demonstrated either

*To whom correspondence should be addressed. Tel: +44 1223 336347; Fax: +44 1223 336913; Email: sb10031@cam.ac.uk
Correspondence may also be addressed to Julian Leon Huppert. Tel: +44 1223 337256; Fax: +44 1223 337000; E-mail: jlh29@cam.ac.uk.

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Table 1. PQS densities, and proportions of genes with a PQS, in four genomic regions: 200 bp before the TSS ('promoter region'), 200 bp after the TSS ('Beginning of the gene'), 100 bp before the TES ('End of the gene') and 100 bp after the end of the TES ('After the gene')

Gene type	Strand	All genes			Differentially expressed genes			Upregulated genes			Downregulated genes		
		Both	Coding	Template	Both	Coding	Template	Both	Coding	Template	Both	Coding	Template
PQS density (# per kb)	Promoter region	1.48	0.74	0.74	1.81	0.90	0.91	1.98	1.03	0.95	1.61	0.75	0.86
	Beginning of the gene	0.99	0.57	0.42	1.32	0.72	0.60	1.40	0.76	0.64	1.23	0.68	0.55
	End of the gene	0.18	0.08	0.10	0.28	0.10	0.19	0.42	0.14	0.28	0.11	0.04	0.08
	After the gene	0.30	0.21	0.09	0.50	0.39	0.11	0.50	0.37	0.13	0.49	0.42	0.08
Proportion of genes with a PQS (%)	Promoter region	24.2	12.9	12.9	29.2	15.5	15.5	31.5	17.0	16.8	26.5	13.7	14.0
	Beginning of the gene	17.7	10.6	7.8	23.0	13.1	11.1	24.3	13.6	11.8	21.5	12.5	10.2
	End of the gene	1.8	0.8	1.0	2.8	1.0	1.9	4.1	1.4	2.8	1.1	0.4	0.8
	After the gene	2.9	2.1	0.9	4.9	3.9	1.1	5.0	3.7	1.3	4.8	4.0	0.8

specific cases (29–31) or genome wide studies of changes in gene expression by a G-quadruplex binding small molecule (32,33). We have previously shown that artificial proteins that recognize G-quadruplex DNA with nanomolar affinity can be selected using phage display of libraries of zinc fingers (24,34,35) and more recently using single-chain antibody libraries (25). In the current work, we have employed a single-chain G-quadruplex binding antibody (HF1) that was selected from a phage library by methods analogous to those used previously (25), but with negative selection against only duplex DNA. Antibody HF1 shows broad affinity to a range of DNA G-quadruplexes that we had selected from the promoters of various genes: *ckit2*, $K_d = 54$ nM; *GRINA*, $K_d = 4.5$ μ M; *AGRN*, $K_d = 300$ nM; *SPIN1*, $K_d = 16$ μ M; *ASPSCR1*, $K_d = 23$ μ M (see 'Materials and Methods' section and Table 1 of Supplementary Data for details). In contrast, HF1 shows no detectable binding to double stranded DNA (dsDNA, Supplementary Table S1). As such, it represents a highly structure-specific tool, which can be used to probe the presence and function of G-quadruplexes within the genome. Binding events of the antibody to G-quadruplexes located in promoters might be expected to interfere with the regulatory mechanism and result in changes of the levels of gene expression of those genes containing a quadruplex in their promoter. We therefore conducted a genome-wide expression study to test the promoter hypothesis in human cells in which we have exogenously expressed antibody HF1.

In addition to promoter-specific effects, we also sought to investigate other effects that this antibody may have. In particular, we observed effects either side of the site of transcriptional termination. These experimental data are in line with predictions made in our recent computational study, which suggested the possible involvement of G-quadruplexes in transcription termination (36).

MATERIALS AND METHODS

Selection and cloning of the scFv antibody

The single-chain antibody HF1 was selected from the Tomlinson J phage library by screening against the biotinylated c-Kit 2 G-quadruplex sequence immobilized

onto streptavidin-coated immunotubes as described in ref. (25) with the alteration that three rounds of panning were performed and duplex DNA was used as the competitor. The DNA sequence was then PCR amplified from the parental pIT2 vector using the primers 'pIT2 fwd PCR':

5'-CATGTATACCTGCGCGCGGAGGTGCAGCTGTTG-3'

and 'pIT2 rev PCR':

5'-GCAGAATTCTCAATTCAGATCCTCTTCTGAGAT-3'.

The scFv coding DNA was inserted into the mammalian expression vector construct pIRES2-EGFP (Clontech) by using the restriction sites *Bss*HII and *Eco*RI and adding a nuclear localization sequence (NLS), which were appended to the DNA encoding the scFv using PCR amplification from the phagemid. Sequencing results confirmed integrity of the resulting construct (Supplementary Figure S2), and fluorescence microscopy of a NLS-HF1-EGFP fusion construct confirmed the nuclear localization of the scFv (Supplementary Figure S3).

Expression of scFv and ELISA against G-quadruplexes

Expression of the soluble scFv was performed as described in ref. (25). We carried out ELISA measurements, as described in ref. (25), to determine the K_d of HF1 against a range of promoter derived G-quadruplexes: *ckit2* (5' GAC CCG GGC GGG CGC GAG GGA GGG GAG G 3'); *GRINA* (5' CGA GGG GTG GGG TCG GGC GGT GGG CGC 3'); *AGRN* (5' CTC CGG GCG GGC CGG GGG AGG GAG GAG GGA GGG GGC GGG AGG GGG AAG 3'); *SPIN1* (5' CAG TGG GGC GGG GCG GGG CCG GCG AGG GCG GGG CCT 3'); *ASPSCR1* (5' CCA AGG GCG GGG CCG GGG CGC CGT GGG TGG GGA AAG GGC GGG GCC T 3') (Supplementary Table S1).

Transfection of mammalian cells

We employed human gastric carcinoma cells HGC-27, for their good general handling qualities and ease of transfection. Transient transfections of HGC-27 were done in T-75 flasks using Lipofectamine2000 transfection reagent (Invitrogen), using 12 ml of cells at a density of 1×10^5 cells/ml grown overnight. Thirty micrograms of

plasmid DNA (HF1 expression plasmid or empty vector pIRES-EGFP) and 75 μ l of transfection reagent were used per T-75 flask and transfected according to the manufacturer's protocol in serum-free medium. Cells were harvested 15 h after transfection. The medium was removed and the flask was washed with 10 ml 1 \times PBS three times. Five ml Trypsin-EDTA solution was added to the flask and incubated at 37°C for 2–3 min. When the cells had dislodged from the flask bottom, 20 ml of EMEM medium + 10% FCS was added into the flask to stop trypsin digestion. The contents were gently mixed and transferred into a 50 ml Falcon tube. The tube was spun at 300 g for 5 min and resuspended in 10 ml PBS. The cells were spun again and resuspended in PBS containing 0.1% FCS. The cell suspension was analysed FACS after the addition of propidium iodide (to stain the non-viable cells), and sorted to select for transfected cells.

RNA extraction

Flow sorted cells harboring the transfected plasmid were used for total RNA extraction. Cells were disrupted with QIAshredders and total RNA was isolated using the RNeasy kit from Qiagen according to the manufacturers instructions. Purified RNA was analysed at the Centre for Microarray Resources at the Department of Pathology, University of Cambridge.

Microarray data processing

Gene expression was profiled in triplicate (three treatments, three controls) on an Illumina Human 6 version 2 BeadArray (37). The bead-level data were analysed using the R *beadarray* package (38), including image sharpening, local background subtraction, bead summarization and quantile normalization between arrays. Replicate 2J (HF1 treatment, array E) was found to be an outlier and discarded from the analysis prior to normalization. Differential expression was assessed using the R *limma* package (39), taking into account bead variances as weights, and using a corrected *P*-value threshold of 0.001. We only considered probes that mapped to genes that have an Ensembl gene id, 'known' and 'protein coding' status, using the R *biomaRt* package (40) and the Illumina probe annotation provided in Ensembl. When multiple probes mapped to a single gene, only the probe with the lowest corrected *P*-value was kept. The Ensembl annotation was used to define the transcription start site (TSS) and end site of each gene.

Identification of PQS

Putative quadruplex sequences (PQS) were identified in the human genome (NCBI 36) using the quadparser algorithm (14), which is available online at <http://www.quadruplex.org/?view=quadparser>. Briefly, we used the default parameters for this program, which searches for sequences of the form $G_3+N_{1-7}G_3+N_{1-7}G_3+N_{1-7}G_3+$ on either strand of the sequence given. The PQS density distributions between different groups of genes were compared in different regions, and the significance of those differences was assessed using a

two-sample *t*-test, allowing unequal variances. Control sequences, such as those with only three GGG-repeats, or that were A or T rich, did not produce equivalent results (data not shown).

RESULTS AND DISCUSSION

The single-chain antibody HF1 was selected from the Tomlinson J phage library as described in 'Materials and Methods' section. The human cell line HGC-27 was transfected with the plasmid encoding the gene for the HF1 antibody under the control of a CMV promoter. The parental vector pIRES2-EGFP, which lacked the HF1 gene, was employed in a control transfection. The expression levels were assessed by fluorescence microscopy and FACS analysis of the co-expressed EGFP. The levels were in line with control data from a CMV promoter, indicative of strong expression. For experiments using both HF1 and the control, total RNA was extracted 15 h after transfection ('Materials and Methods' section). All transfections were done in triplicate and flow sorted to collect only the cells harboring the expression plasmid, which also expressed EGFP. A sample of cDNA from each triplicate was hybridized to a separate Illumina Human 6 version 2 BeadArray. The microarray data were analysed using the R packages *beadarray* and *limma* ('Materials and Methods' section).

The analysis showed that of 18 105 known protein-coding genes, 1767 were significantly differentially expressed (DE) in the HF1 samples versus the control samples, using a corrected *P*-value of 10^{-3} . Of these DE genes, 973 (55%) were up-regulated compared with the control, and the remaining 794 (45%) were down-regulated.

The subject of our analysis was to investigate whether the presence of putative G-quadruplex sequences (PQS) in regulatory regions was related to the differential expression of genes. The principal region of study was a window 200 bp either side of the TSS, which is involved in transcription initiation. We also considered the 100 bp region either side of the transcription end site (TES), to explore regulatory effects that may be associated with transcription termination (36). In our analysis, we elected to limit the ranges for each of these regions (Figures 1 and 2), to capture most of the deviations of interest, without including too many distal sequences that would add to the noise in the data (for an analysis of genomic regions not considered here Supplementary Figure S1). Previous results have shown that these regions contain peaks in incidental G-quadruplex density (15,36). A single base-pair resolution analysis of the PQS densities did not yield additional insights compared with the region-level analysis. Of note for comparative purposes, the previously calculated density of PQS in the whole genome is 0.115 and 0.153 PQS/kbase when considering only the fraction of DNA that is ever transcribed as pre-mRNA.

Promoter region—200 bp upstream of the TSS

The promoter region of genes is known to have a high PQS density compared with other regions of the human

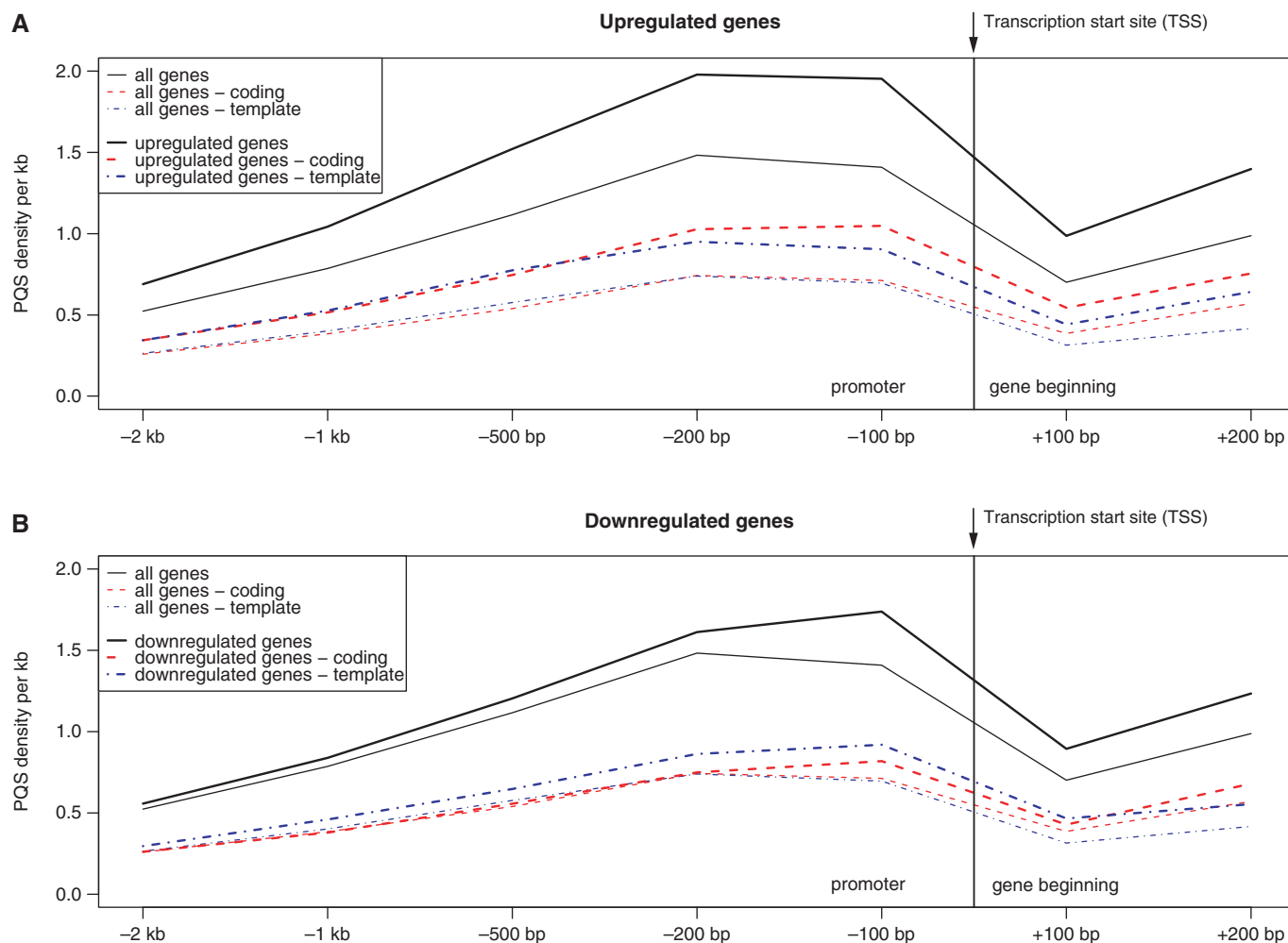


Figure 1. PQS density at the promoter region. (A) Up-regulated genes (bold) compared with all genes. (B) Down-regulated genes (bold) compared with all genes. Red lines indicate data for the coding strand, blue lines for the template strand, black for the sum. The numbers are relative to the TSS.

genome (15). In our experiment, we found that the promoters of differentially expressed (DE) genes were significantly enriched in PQS compared with all genes (1.81 PQS/kb compared with 1.48 PQS/kb, a 22% enrichment, $P = 3 \times 10^{-6}$ using a two-sample *t*-test) (full details in Table 1; graphical representations in Figure 1). This enrichment was higher for up-regulated genes (1.98 PQS/kb) compared with down-regulated genes (1.61 PQS/kb) (significant difference, $P = 0.01$). There was no significant strand asymmetry; G-quadruplexes located on the coding and template strands were equivalent (at significance level 0.05).

Beginning of the gene—200 bp downstream of the TSS

We then considered the first 200 bp of each gene transcript, and found broadly similar results to those obtained for the 200 bp upstream of the TSS, but at generally lower PQS densities. DE gene were enriched in PQS in this region (1.32 PQS/kb for DE genes, compared with 0.99 PQS/kb for all genes, a 33% enrichment, $P = 6 \times 10^{-9}$). There was a slight excess of enrichment

for up-regulated genes (1.40 PQS/kb) as compared with down-regulated genes (1.23 PQS/kb), although this was not statistically significant at the 0.05 level. No significant strand asymmetry was observed.

These results show clearly that both upstream and downstream of the TSS, differentially expressed genes are significantly enriched for PQS, even when taking into consideration the general enrichment of PQS proximal to the TSS for genes. This is consistent with previous studies of these regions that showed they are important for regulation of gene expression (15,17–21,41). Whilst we cannot rule out the existence of cases where a G-quadruplex on the template strand causes an effect via an RNA quadruplex, the strand symmetry suggests the major effect is at the DNA level. These results are consistent with a direct interaction between HF1 with DNA G-quadruplexes that changes the level of expression of the adjacent gene. The data indicate that the antibody acts preferentially on genes that contain a PQS in the promoter sequence. This finding is in line with the hypothesis that DNA promoter

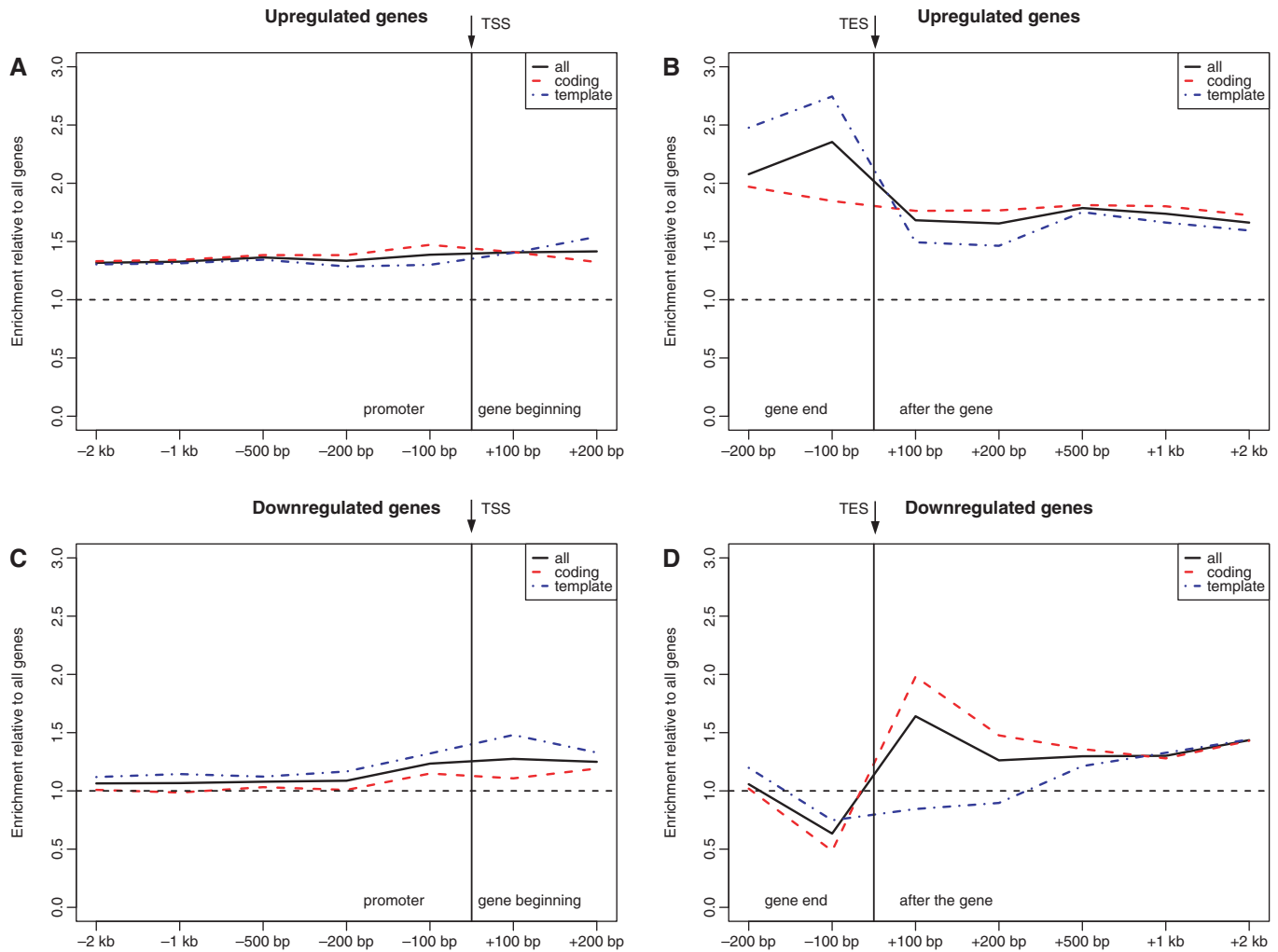


Figure 2. Relative enrichment of PQS in the regions studied. (A and B) Up-regulated genes. Enrichment of PQS in the coding strand (red), the template strand (blue) and both (black) of up-regulated genes, compared with all known genes. (C and D) Down-regulated genes, colours as in (A). All enrichments are compared with PQS densities in specified regions and strands for all genes. Numbers are relative to the TSS or TES.

G-quadruplexes are involved with transcription. Collectively, these data suggest that G-quadruplexes on either strand, upstream or downstream of the TSS can have a regulatory effect which can lead to up- or down-regulation.

End of the gene—100 bp upstream of the TES

The last 100 bp of each gene has a fairly low density of PQS, but DE genes showed significant PQS enrichment patterns there. Only 1.8% of all genes had a PQS in this region (i.e. a density of 0.18 PQS/kb), whereas 2.8% of DE genes (a density of 0.28 PQS/kb) had a PQS in this region, giving an enrichment of 56% ($P = 6 \times 10^{-3}$). In relative terms, this enrichment was much larger than in the promoter region. Intriguingly, this enrichment appeared to be confined to up-regulated genes, with four times as many genes with a PQS in this region being up-regulated (40 genes) as compared with down-regulated (nine genes). The PQS density in this region for up-regulated genes was 0.42 PQS/kb, compared with 0.11 PQS/kb for

down-regulated genes, and 0.18 PQS/kb for all genes ($P = 5 \times 10^{-5}$ comparing up-regulated genes to down-regulated genes). PQS were more likely to be found on the template strand in this region than on the coding strand, when analysing all genes, (30% excess, $P = 0.02$), but among differentially expressed genes this became a 90% excess (1% of coding, versus 1.9% of template strands contain a PQS).

After the gene—100 bp downstream of the TES

The first 100 bp after the end of the gene is another region with relatively low PQS density but striking enrichment patterns. However, 5.0% of DE genes, as compared with only 3.0% of all genes, had a PQS in this region, an enrichment of 68%. The PQS density in this region for DE genes was higher at 0.50 PQS/kbase, as compared with 0.30 PQS/kb for all genes ($P = 5 \times 10^{-5}$). In contrast to the last 100 bp of the gene, PQS were much more common here on the coding strand than the template strand in all genes, with densities of 0.21 and

0.09 PQS/kb, respectively ($P = 2 \times 10^{-20}$). DE genes showed an even greater strand bias, with densities of 0.39 and 0.11 PQS/kb, respectively ($P = 3 \times 10^{-5}$ for the coding strand). For down-regulated genes, the PQS density on the coding strand was 0.42 PQS/kb, whereas the template strand had a density of only 0.08 PQS/kb, which was even lower than that for all genes.

We considered PQS at the 3'-end of the gene because of the proposed involvement of this region in transcription termination (36). PQS are relatively rare in this region, however, the significant enrichment of PQS apparent in differentially expressed genes in both of these regions is clear and suggestive of an interaction of HF1 with quadruplexes. A previous study on quadruplexes in UTRs had also shown a marked increase in overall PQS frequencies immediately after the end of genes (36). This was determined computationally, based purely on sequence identity. Our genome-wide data support a role of G-quadruplexes in molecular mechanisms occurring at the end of genes. As termination of transcription is a finely regulated process and closely linked to initiation of transcription (42), interfering at this stage of gene expression could give rise to altered mRNA levels, as is manifest here.

Strand bias of PQS density around the TES

An interesting observation is the marked asymmetry in PQS frequencies between both DNA strands at very localized regions at the end of genes. In up-regulated genes, our results show particularly high levels of PQS enrichment in the template strand just before the end of the gene. This PQS density drops greatly in sequences just 3'-end of the gene (Figure 2B).

In down-regulated genes, the distribution of PQS shows an inverse trend as compared with up-regulated genes. Within the last 100 bp of the UTR, the density of PQS is less than the average across all genes. This is reversed immediately after the end of the gene, where the PQS density rises above average in the coding strand (Figure 2D). This marked asymmetry of PQS density with regards to both the strand as well as the position relative to the end of the gene is striking and provides evidence that the effects of quadruplexes are highly dependent on the precise location within the genome. The fact that both up- and down-regulation can be observed indicates that G-quadruplexes are part of a *bona fide* regulatory system.

In our analysis, we have not factored in that some differentially expressed genes might affect the expression of others, for example by acting as transcription factors to genes genetically downstream. Such downstream genes would, in principle, be affected independently of the presence of PQS in their vicinity, causing an apparent dilution of any quadruplex-specific trends. Indeed, it is known that transcription factors are more likely to have PQS in their promoters than other gene classes (15). Thus, the 1767 differentially expressed genes include genes whose expression levels are directly or indirectly affected by HF1. Therefore the levels of enrichment we report are likely to be an underestimate. While we have described the perturbation of gene expression by the quadruplex binding

protein HF1, we cannot on the basis of these data alone unequivocally define the explicit role of the natural quadruplex forming elements in the absence of HF1.

CONCLUSION

We have shown that an engineered G-quadruplex binding antibody can have an effect on gene expression in human cells. The current genome-wide study reveals a correlation between altered gene expression and the incidence of PQS at either the promoter or the terminus of genes. Overall, this genome-wide study supports the involvement of PQS in gene regulation, whilst suggesting that the sense in which G-quadruplex elements affects gene expression may well depend on the context of each quadruplex.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Cambridge Genomic Services (CGS), who performed the microarray experiments, and Julien Bauer and Emily Clemente from CGS for valuable help. J.L.H. is an RCUK Academic Fellow and J.D. was supported by an EMBL studentship.

FUNDING

Cancer Research UK (grant number C9681/A5709), project funding from the BBSRC (grant number BB/E012752/1) and a Cambridge Commonwealth Trust studentship to H.F. Funding for open access charge: Cancer Research UK (grant no. C9681/A5709). S.T. and J.D. acknowledge the support of the University of Cambridge, Cancer Research UK, and Hutchison Whampoa.

Conflict of interest statement. None declared.

REFERENCES

1. Neidle, S. and Balasubramanian, S. (2006) *Quadruplex Nucleic Acids*. RSC Publishing, Cambridge, UK.
2. Huppert, J.L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.*, **37**, 1375–1384.
3. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
4. Patel, D.J., Phan, A.T. and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
5. Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
6. Haider, S., Parkinson, G.N. and Neidle, S. (2002) Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex. *J. Mol. Biol.*, **320**, 189–200.
7. Haider, S.M., Parkinson, G.N. and Neidle, S. (2003) Structure of a G-quadruplex-ligand complex. *J. Mol. Biol.*, **326**, 117–125.
8. Phan, A.T., Kuryavyi, V., Burge, S., Neidle, S. and Patel, D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.

9. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. and Plückthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylochyia lemnae* macronuclei. *Proc. Natl Acad. Sci. USA*, **98**, 8572–8577.
10. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. and Lipps, H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.*, **12**, 847–854.
11. Oganessian, L. and Bryan, T.M. (2007) Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *Bioessays*, **29**, 155–165.
12. De Cian, A., Grellier, P., Mouray, E., Depoix, D., Bertrand, H., Monchaud, D., Teulade-Fichou, M.P., Mergny, J.L. and Alberti, P. (2008) Plasmodium telomeric sequences: structure, stability and quadruplex targeting by small compounds. *ChemBiochem*, **9**, 2730–2739.
13. Balasubramanian, S. and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.*, **13**, 345–353.
14. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
15. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
16. Qin, Y. and Hurley, L.H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie*, **90**, 1149–1171.
17. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
18. Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene. *J. Am. Chem. Soc.*, **127**, 10584–10589.
19. Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S. and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human *c-kit* oncogene. *Biochemistry*, **45**, 7854–7860.
20. Cogoi, S. and Xodo, L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.
21. Du, Z., Zhao, Y. and Li, N. (2008) Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.*, **18**, 233–241.
22. Fry, M. (2007) Tetraplex DNA and its interacting proteins. *Front. Biosci.*, **12**, 4336–4351.
23. Ou, T.M., Lu, Y.J., Tan, J.H., Huang, Z.S., Wong, K.Y. and Gu, L.Q. (2008) G-quadruplexes: targets in anticancer drug design. *Chem. Med. Chem.*, **3**, 690–713.
24. Isalan, M., Patel, S.D., Balasubramanian, S. and Choo, Y. (2001) Selection of zinc fingers that bind single-stranded telomeric DNA in the G-quadruplex conformation. *Biochemistry*, **40**, 830–836.
25. Fernando, H., Rodriguez, R. and Balasubramanian, S. (2008) Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry*, **47**, 9365–9371.
26. Phan, A.T., Modi, Y.S. and Patel, D.J. (2004) Propellor-type parallel-stranded G-quadruplexes in the human *c-myc* promoter. *J. Am. Chem. Soc.*, **126**, 8710–8716.
27. Kuryavyi, V., Majumdar, A., Shallop, A., Chernichenko, N., Skripkin, E., Jones, R. and Patel, D.J. (2001) A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G(syn)-G(syn)-G(anti)-G(anti) tetrads flanked by a G-(T-T) Triad and a T-T-T triple. *J. Mol. Biol.*, **310**, 181–194.
28. Dexheimer, T.S., Sun, D. and Hurley, L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.*, **128**, 5404–5415.
29. Bejugam, M., Sewitz, S., Shirude, P.S., Rodriguez, R., Shahid, R. and Balasubramanian, S. (2007) Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of *c-kit* oncogene expression. *J. Am. Chem. Soc.*, **129**, 12926–12927.
30. Grand, C.L., Han, H., Munoz, R.M., Weitman, S., Von Hoff, D.D., Hurley, L.H. and Bearss, D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.*, **1**, 565–573.
31. Ou, T.M., Lu, Y.J., Zhang, C., Huang, Z.S., Wang, X.D., Tan, J.H., Chen, Y., Ma, D.L., Wong, K.Y., Tang, J.C. et al. (2007) Stabilization of G-quadruplex DNA and down-regulation of oncogene *c-myc* by quindoline derivatives. *J. Med. Chem.*, **50**, 1465–1474.
32. Verma, A., Halder, K., Halder, R., Yadav, V.K., Rawal, P., Thakur, R.K., Mohd, F., Sharma, A. and Chowdhury, S. (2008) 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.
33. Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.S. and Johnson, F.B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **36**, 144–156.
34. Ladame, S., Schouten, J.A., Roldan, J., Redman, J.E., Neidle, S. and Balasubramanian, S. (2006) Exploring the recognition of quadruplex DNA by an engineered Cys2-His2 zinc finger protein. *Biochemistry*, **45**, 1393–1399.
35. Patel, S.D., Isalan, M., Gavory, G., Ladame, S., Choo, Y. and Balasubramanian, S. (2004) Inhibition of human telomerase activity by an engineered zinc finger protein that binds G-quadruplexes. *Biochemistry*, **43**, 13452–13458.
36. Huppert, J.L., Bugaut, A., Kumari, S. and Balasubramanian, S. (2008) G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.*, **36**, 6260–6268.
37. Kuhn, K., Baker, S.C., Chudin, E., Lieu, M.H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T.K. and Chee, M.S. (2004) A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res.*, **14**, 2347–2356.
38. Dunning, M.J., Smith, M.L., Ritchie, M.E. and Tavaré, S. (2007) Beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics*, **23**, 2183–2184.
39. Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article 3.
40. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A. and Huber, W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.
41. Simonsson, T., Pecinka, P. and Kubista, M. (1998) DNA tetraplex formation in the control region of *c-myc*. *Nucleic Acids Res.*, **26**, 1167–1172.
42. Proudfoot, N. (2004) New perspectives on connecting messenger RNA 3'-end formation to transcription. *Curr. Opin. Cell Biol.*, **16**, 272–278.