4194–4204 *Nucleic Acids Research, 2009, Vol. 37, No. 13* *Published online 11 February 2009*
doi:10.1093/nar/gkn1076

Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells

Anjali Verma[1], Vinod Kumar Yadav[2], Richa Basundra[1], Akinchan Kumar[1] and Shantanu Chowdhury[1,2,*]

[1]Proteomics and Structural Biology Unit and [2]G. N. Ramachandran Knowledge Centre for Genome Informatics, Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110 007, India

Received November 23, 2008; Revised and Accepted December 22, 2008

ABSTRACT

Guanine-rich DNA of a particular sequence adopts four-stranded structural forms known as G-quadruplex or G4 DNA. Though *in vitro* formation of G4 DNA is known for several years, *in vivo* presence of G4 DNA was only recently noted in eukaryote telomeres. Recent bioinformatics analyses showing prevalence of G4 DNA within promoters of human and related species seems to implicate G4 DNA in a genome-wide *cis*-regulatory role. Herein we demonstrate that G4 DNA may present regulatory sites on a genome-wide scale by showing widespread effect on gene expression in response to the established intracellular G4 DNA-binding ligands. This is particularly relevant to genes that harbor conserved potential G4 DNA (PG4 DNA) forming sequence across human, mouse and rat promoters of orthologous genes. Genes with conserved PG4 DNA in promoters show co-regulated expression in 79 human and 61 mouse normal tissues (z-score > 3.5; $P < 0.0001$). Conservation of G4 DNA across related species also emphasizes the biological importance of G4 DNA and its role in transcriptional regulation of genes; shedding light on a relatively novel mechanism of regulation of gene expression in eukaryotes.

INTRODUCTION

Since the discovery of DNA as a double-helical B-DNA molecule, many other non B- forms of DNA have been observed (1). Among these the four-stranded guanine-rich motif, G-quadruplex or G4 DNA adopted by particular G-rich sequences have gained substantial attention in recent times as potential targets for cancer therapy (1–4). The four-stranded structure of G4 DNA is stabilized by hydrogen-bonding between guanines present in a co-planar arrangement called tetrads (5). Though G4 DNA forms were first noted almost two decades back (6–8), evidence of *in vivo* formation of G4 DNA was observed only recently in telomeric ends in a cell-cycle-dependent fashion (9) along with other evidence of intracellular presence (10,11). Direct evidence of *in vivo* existence of G4 DNA at the telomeric regions in humans was shown by Chang *et al.* (10) by probing G4 DNA by a fluorescent compound BMVC (3,6-bis(1-methyl-4-vinylpyridinium)carbazole diiodide. BMVC fluorescence increases as it binds to DNA; additionally it gives distinct fluorescence emissions for binding with G4 DNA and duplex DNA (around 575 and 545 nm, respectively) thus enabling detection of G4 DNA-specific BMVC binding. This study also indicated other possible G4 DNA formation sites in the genome, besides the telomeric ends. In another independent study, G4 DNA existence inside cells was demonstrated using an antibody specific for telomeric G4 DNA in *Stylonychia lemnae* nuclei (11).

Telomeric G4 DNA as a potential therapeutic target in cancer has been reported by several groups (12). Telomeres are composed of tandem repeats of d(TTAGGG) which may extend up to 25 kb in length. These regions have a characteristic single-stranded overhang of about 100–150 bases with a propensity to form G4 DNA structures (12). Ligand-induced stabilization of G4 DNA at the telomeric ends inhibits telomerase (a ribonucleoprotein that is crucial for telomere maintenance) causing cell-death (13). In this context, molecules like PIPER (N,N′-bis[2-(1-piperidino)-ethyl]-3,4,9,10-perylenetetracarboxylic diimide) (14), the cationic porphyrin TMPyP4 (15) and anthraquinone analogs (16) which target G4 DNA and stabilize them are interesting potential anticancer drugs (13). TMPyP4 as an anticancer molecule is extensively studied (15,17,18), which is primarily due to its binding with telomeric G4 DNA resulting in inhibition of telomerase activity.

Role of G4 DNA in regulation of several genes has been either demonstrated or predicted. G-quartet formation in the insulin-linked polymorphic region upstream of the

*To whom correspondence should be addressed. Tel: +91 11 2766 6157; Fax: +91 11 2766 7471; Email: shantanuc@igib.res.in

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

insulin gene was observed to enhance transcription, where single/double mutations that disrupted the quartet structure gave reduced promoter activity (19). On the other hand, G4 DNA could also be a 'silencer element' as demonstrated by Siddiqui-Jain *et al*. (20) in case of regulation of the proto-oncogene *c-MYC*. Specific substitutions that disrupted the structure of G4 DNA within the *c-MYC* promoter gave increased promoter activity; while stabilization of the G4 DNA by TMPyP4 decreased transcription by several folds (20). Recently, it was also demonstrated that TMPyP4-mediated stabilization of G4 DNA found within the NHE of *PDGF-A* can silence *PDGF-A* expression (21). In a similar way, G4 DNA formation by d(CGG) repeats within the 5′-untranslated region of the first exon of *FMR1* gene could contribute to silencing of the gene resulting in the neurological disorder fragile X syndrome (22). However, in contrast to above cases, here TMPyP4 was observed to destabilize the G4 DNA motif (23). Therefore, effect of TMPyP4 on G4 DNA demonstrates the regulatory role of TMPyP4-mediated stabilization or destabilization of G4 DNA. Several other oncogenes, including c-*KIT* (24), *KRAS* (25), *VEGF* (26), *BCL-2* (27) and retinoblastoma susceptibility gene (*Rb*) (28) have also been shown to possess G4 DNA as a potential transcription regulatory element. Interestingly, recent studies show the role of G4 RNA in regulation of *NRAS* translation (29) and expression of other genes (30).

In addition to its role in transcription regulation, G4 DNA has been observed to be associated with several of the genetic diseases like Werner syndrome, Bloom syndrome and Rothmund–Thomson syndrome (31), which are characterized by chromosomal instability caused by unusual changes in RecQ helicase proteins like WRN, BLM and RECQ4. These and several other proteins like thrombin (32), MyoD (33,34), DNA topoisomerase 1 (35) and G Quartet Nuclease1 (GQN1) (36) that interact with G4 DNA in humans further indicate the physiological importance of G4 DNA.

It is interesting to note that recent whole-genome *in silico* analysis revealed high occurrence of G4 DNA in both prokaryotes (37) and eukaryotes (38–41). In prokaryotes, the study indicated that G4 DNA was conserved across different genomes, and its occurrence was enriched in regions near the transcription start sites (TSS) (37). In humans also *in silico* analysis showed that the occurrence of G4 DNA was higher in regulatory regions (promoters) as compared to others regions of the genome (39,40,42). These studies implicate G4 DNA in genome-wide regulation as a *cis*-regulatory element. In view of the above evidence, we researched the role of G4 DNA in genome-wide gene regulation in the present study by using whole-genome expression microarrays in two different cell lines when treated with the G4 DNA ligand TMPyP4.

MATERIAL AND METHODS

cDNA microarray and hybridization

cDNA Microarrays representing 19 200 single spotted clones of human genes that represent about 7029 unique genes from University Health Network Microarray Centre (UHN, Canada) were used in the present study. RNA was isolated from treated (100 μM TMPyP4; 48 hr) and untreated HeLa S3 cells and A549 cells using TRIZOL reagent (Sigma) as per manufacturer's protocol. Ten micrograms of RNA were converted to cDNA using Microarray cDNA Synthesis Kit (ROCHE), as per manufacturer's protocol. Purified cDNA was labeled with either Cy5 or Cy3 dyes (Amersham Biosciences) with the help of RNA target synthesis kit T7 (Roche) as per manufacturer's protocol. The untreated control samples were labeled with Cy3 and the treated samples were labeled with Cy5. Labeling was reversed in case of dye-swap reactions. Labeled product was purified with Microarray Target Purification Kit (Roche) as per manufacturer's protocol. For each cell line four replicate experiments were done, consisting of two standard replicate arrays and two dye-swap arrays. The labeled cRNA, both Cy3 and Cy5 labeled, were pooled together and precipitated with ammonium acetate and the dried pellet dissolved in 18 MΩ RNAse free water (Sigma). The labeled product was loaded onto the slides and the hybridization was done in presence of Dig Easy hybridization buffer (Roche), 10 mg/ml salmon testis DNA (Sigma) and 10 mg/ml yeast tRNA (Sigma), at 37°C for 16 h. Slides were washed three times (15 min each) with 1× SSC and 0.1% SDS at 50°C with occasional swirling, followed by three washes with 1× SSC at room temperature (15 min each). Final two washes were done with 0.1× SSC for 15 min each. The slides were dried and scanned at 10-μm resolution with GenePix 4000A Microarray Scanner (Molecular Devices), using lasers for both Cy3 and Cy5 dyes.

Microarray data filtering and analysis

Array images were scanned using GenePix 4000A Microarray Scanner (Molecular Devices), using both green and red lasers. The 16-bit TIFF images were preprocessed and quantified using Gene Pix Pro 6.0 (Molecular Devices). Data normalization was performed using Acuity 4.0 (Molecular Devices). Data were normalized so that the mean of the ratio of medians (635/532) of all features is equal to 1. Data were expressed as the log2 ratio of the samples and the reference for each spot on the array. Data were filtered to include only those elements which contained only a small percentage (<3) of saturated pixels, were not flagged bad or found absent (flags ≥ 0), had relatively uniform intensity and uniform background [Rgn R2 (635/532) ≥ 0.6] and were detectable above background (SNR ≥ 3). Analyzable spots in at least six of eight expression arrays performed (four each in HeLa S3 and A549) were retrieved for downstream analysis using Significance Analysis of Microarrays [SAM 2.21, Excel Add-In, Stanford (43)] under the conditions of one class response and 100 permutations, with data input parameters kept at the default values. The results from SAM analysis include a SAM score and false discovery rate (*q*-value). Differentially expressed genes were retrieved at a cutoff of $\leq$20% False Discovery Rate (FDR).

Quantitative reverse transcription polymerase chain reaction (RT-PCR)

Two-step real-time RT-PCR was performed to validate the microarray results as determined by SAM analysis. Cells were treated with the small molecule ligands TMPyP4, TpPy, BMVC or TMPyP2 and RNA was isolated using TRIZOL reagent (Sigma) and first-strand cDNA synthesis was done using 2.0 µg of total RNA with cDNA Reverse Transcription Kit (Applied Biosystems) at 37°C for 2 h. Relative quantitative real-time RT-PCR was subsequently performed with an Applied Biosystems 7900 system and SYBR Green Master Mix (Applied Biosystems), as per manufacturer's protocol. Reverse and forward primer sequences for the genes of interest were designed using Primer Express software. For each gene, real-time PCR was performed with samples from at least two different preparations using cDNA generated from the reverse transcription of 2 ug of total RNA, in a 10-µl reaction. Negative controls for each gene included no template (water) and no reverse transcriptase. The PCR amplification consisted of a 10-min denaturation at 95°C, followed by 40 cycles of amplification (15 s at 95°C, 60 s at 60°C). A standard curve was derived from serial dilutions with an external cDNA obtained from HeLa S3 or A549 cells for each gene. Relative concentrations were expressed in arbitrary units. After amplification, amplicons were melted and the resulting dissociation curve assessed to ensure a single product. Relative quantification of each of the genes of interest was performed against a housekeeping gene, β-2-microglobulin (*B2M*), according to Relative Standard Curve Method as described previously (44). Several commonly employed housekeeping genes were assessed; *B2M* was found to remain stable in expression in both untreated and TMPyP4, TpPy or BMVC treated cells. For all PCRs average of measurements done in triplicate is reported. Averaged cell viability was around 88% for 100 µM TMPyP4 or TMPyP2; 83% with 50 µM BMVC; 82% with 10 µM TpPy after 48 h (MTT assay, average of three independent observations).

PG4 DNA motif searching, genomic mapping and analysis

PG4 DNA sequences were searched and mapped to the promoters (±1 kb flanking TSS) of different genes as described earlier (37) and mapped to the human genome (UCSC hg16). Briefly, PG4 DNA motifs were searched with a program written using Perl. A general pattern: G_3-N_L-G_3-N_L-G_3-N_L-G_3, where G is guanine and N is any nucleotide including G. The loop length (L) was varied from one to seven. The same program was rerun with cytosine instead of guanine to identify PG4 DNA motifs on the minus strand and appropriately corrected for orientation before mapping their position in the context of genes.

Tissue-specific enrichment of gene set with PG4 DNA in promoters

Tissue specificity of genes harboring PG4 DNA within putative promoters was checked in 79 human and 61 mouse tissues (45). Analysis was largely based on a previously described method (46). The expression data of each gene across all tissues were first normalized to be mean 0 and variance 1 before ranking them as per their normalized expression level in each tissue, hence generating tissue-specific ranked gene lists. Enrichment of expression (either up- or downregulated) of microarray gene set in a particular tissue and its significance was analyzed from the ranked list of genes specific for that tissue after evaluating the nonrandomness of ranks of microarray gene list (sample set) within tissue-specific gene list (reference set), using the Mann–Whitney rank sum statistic (47). After summing the ranks of sample set in reference set, we tested the significance of this rank sum against the rank sum of control set (all genes in reference set, excluding sample set genes). If µ and α^2 are the mean and variance of the control set than enrichment (z-score) of our sample set is given by $(\mu - S)/\alpha$, which measures enrichment in terms of number of standard deviations away from the mean of the control set. A z-score of ≥ 3.5 is considered to be significant in the present study.

Pathway analysis

For GO analysis, UniGene symbols pertaining to differentially expressed genes were converted to UniProt IDs, using Gene name converter (http://genemerge.bioteam.net/convertgenenames.html). To examine enrichment or depletion of genes in the differentially expressed sets, the UniProt IDs were fed to the GO ToolBox [http://www.geneontology.com; (48)] to obtain the output using the options—ontology: biological process and molecular function; mode: all terms; reference: genome; evidence: all–all evidence; species: *Homo sapiens*; GO-stats; statistic test: hypergeometric; correction for multiple testing: Bonferroni. Gene names of 850 genes out of 1161 differentially expressed genes (both up- and downregulated) mapped to respective UniProt ID's, using Gene name converter, as explained above.

Search for orthologous genes with conserved PG4 DNA

Orthologous gene data were downloaded from (NCBI HomoloGene http://www.ncbi.nlm.nih.gov/). For every PG4 DNA found within ±1 kb upstream or downstream of TSS in a human promoter the respective orthologous mouse and rat promoter was searched for a conserved occurrence in a sequence window of 400 bases centered at the human PG4 DNA. Conserved occurrence was searched based on the PG4 DNA sequence pattern described above whereby conservation is at the level of the PG4 DNA structure and not necessarily primary sequence. A compendium of the orthologous genes and respective PG4 positions are given in the database—EuQuad (http://quadbase.igib.res.in) (40).

RESULTS AND DISCUSSION

Evidence of widespread gene regulation by PG4 DNA

To understand the physiological relevance of G4 DNA in gene regulation we probed the whole genome for

expression changes using TMPyP4. Several previous studies have shown that TMPyP4 is a potent G4 DNA-binding molecule within cells (15,18,20). In the most direct demonstration of the role of G4 DNA in gene regulation, it was shown that *c-MYC* (20) and *PDGF-A* (21) promoter activity was repressed on using TMPyP4 to stabilize G4 DNA found within promoters, while destabilization of G4 DNA in the *FMR1* gene could lead to increased transcription (49). Furthermore, the repressive effect of TMPyP4 on *c-MYC* promoter was lost on disrupting the G4 DNA (20). Recent studies by us and others show prevalence of G4 DNA within promoters of chicken (50), human and related species (39,40,42,51) and several bacteria (37) implicating widespread role of G4 DNA in regulation.

We performed whole-genome expression studies after TMPyP4 treatment for 48 h in two different cell lines—human cervix cancer cells (HeLa S3) and human lung cancer cells (A549). For every TMPyP4-treated experiment we first independently confirmed downregulation of *c-MYC* by real-time quantitative PCR before hybridization to microarray slides to ascertain that the previously known effect of TMPyP4 [i.e. suppression of *c-MYC* (52)] was evident. Four replicates in each cell line were performed and analyzed simultaneously, taking results from both cell lines together, using Significance Analysis of Microarray [SAM (43), see 'Material and Methods' section for details of analysis] to first establish the global effect of G4 DNA-regulated genes. One thousand one hundred and sixty-one genes were significantly up- or downregulated at $\leq$20% FDR as per the statistical significance test performed using SAM (Figure 1A, expression array results shown in Supplementary Figure 1 and http://quadbase.igib.res.in/Geneexpression.html). Out of the 1161 genes, 863 genes were upregulated while 298 genes downregulated (Figure 1B). It is expected that observed genome-wide expression changes will primarily be a result of both direct and indirect effects of TMPyP4-induced stabilization/destabilization of G4 DNA. In other words, some genes will be directly affected by binding of TMPyP4 to G4 DNA within promoters and these genes in turn may regulate several downstream genes. In order to find genes that were most likely to be regulated by G4 DNA, we examined the sequence around transcription start sites (TSS) using a previously published computational method that detects DNA sequence patterns liable to adopt G4 DNA forms (details of the algorithm used are provided in the methods) (37). We found a total of 38 757 PG4s within $\pm$1 kb of TSS in 25 706 annotated human genes where 1550 or (3.9%) of PG4s are within exons, 4363 (11.2%) are in introns and 8490 (21.9%) within the 5′UTRs. We noted 711 out of 1161 differentially expressed genes (526/863 upregulated and 185/298 downregulated) showed presence of at least one PG4 DNA forming sequence within $\pm$1 kb of TSS (Figure 1B). However, we noted that the PG4 density (no. of PG4 DNA motifs per gene) within 1 kb of TSS of genes that responded to TMPyP4 treatment was 1.05 fold higher than expected by chance (though statistically significant, see Supplementary Information) suggesting that the correlation between presence of PG4 DNA and the influence of TMPyP4 is complex. This is not surprising given that expression assays do not distinguish between direct versus indirect effects of treatment. Furthermore, the putative nature of G4 DNA limits the likelihood of all of them forming and/or responding to a binding molecule at the same time. On the other hand, it may also be possible that intracellular effect of TMPyP4 is limited. Keeping these confounding factors in mind we addressed this issue more directly using an experimental plan that involves validation of the genome-wide expression data by expression analysis of selected genes using RT-PCR using other known G4 DNA binding molecules.

In order to validate the microarray data, we used quantitative RT-PCR. 12 genes showing significant response in microarray were picked randomly and their expression measured by quantitative RT-PCR. In most cases (10 genes) results were consistent in terms of the up- or down-regulation observed in microarray (Figure 1C).

In order to ascertain that the effect of TMPyP4 was largely due to G4 DNA binding, we also used two additional molecules whose binding with G4 DNA have been reported. We selected the carbazole derivative BMVC (Figure 2A) which was shown recently to bind G4 DNA motifs *in vivo* within human telomeres (10,52). The second molecule, 5,10,15,20-tetrakis (R-pyridinio-*p*-methylphenyl)-21,23*H*-porphyrin tetrachloride (TpPy) is an analog of TMPyP4 with extended aromatic side chains (Figure 2A). In previous reports, TpPy was demonstrated to have specific binding affinity for the G-quartet due to the favorable $\pi-\pi$ stacking rendered by the aromatic side chain in TpPy. Quantitative RT-PCR was used to check the effect of BMVC and TpPy on gene expression. Twelve genes were used for this purpose including *c-MYC* (11 genes were from the ones randomly selected for qRT-PCR with TMPyP4; *DHCR24*, which gave no transcripts in A549 cells was excluded). We noted with interest that for BMVC (50 μM) and TpPy (10 μM) treatment, all the 12 genes showed repressed levels, which were similar to the repression noted for TMPyP4 treatment in 10 genes (Figure 2B). Treatments were done using 50 and 10 μM of BMVC and TpPy, respectively, as higher concentrations compromised cell viability. For two genes, *C5ORF6* and *MTCP1*, effect of TMPyP4 was opposite to that of TpPy and BMVC.

For further validation of these observations, we also used 5,10,15,20-tetra-(*N*-methyl-2-pyridyl)porphyrin (TMPyP2). TMPyP2 is positional isomer of TMPyP4 where the position of the cationic substitution on the periphery of the porphyrin is altered (Figure 2A). This results in significantly decreased affinity for G4 DNA and telomerase inhibition whereas its binding affinity for double-strand DNA is similar to that of TMPyP4 (54,55) Keeping this in mind, we hypothesized that TMPyP2 would be a suitable negative control for G4 DNA binding within cells *vis-à-vis* TMPyP4. This was tested using qRT-PCR of 12 genes (as selected above) after treatment with TMPyP2. We first noted that *c-MYC* expression did not change significantly on TMPyP2 treatment and was therefore considered as the endogenous control for relative quantitation of gene expression. Further, expression of eight other genes was very different from that observed

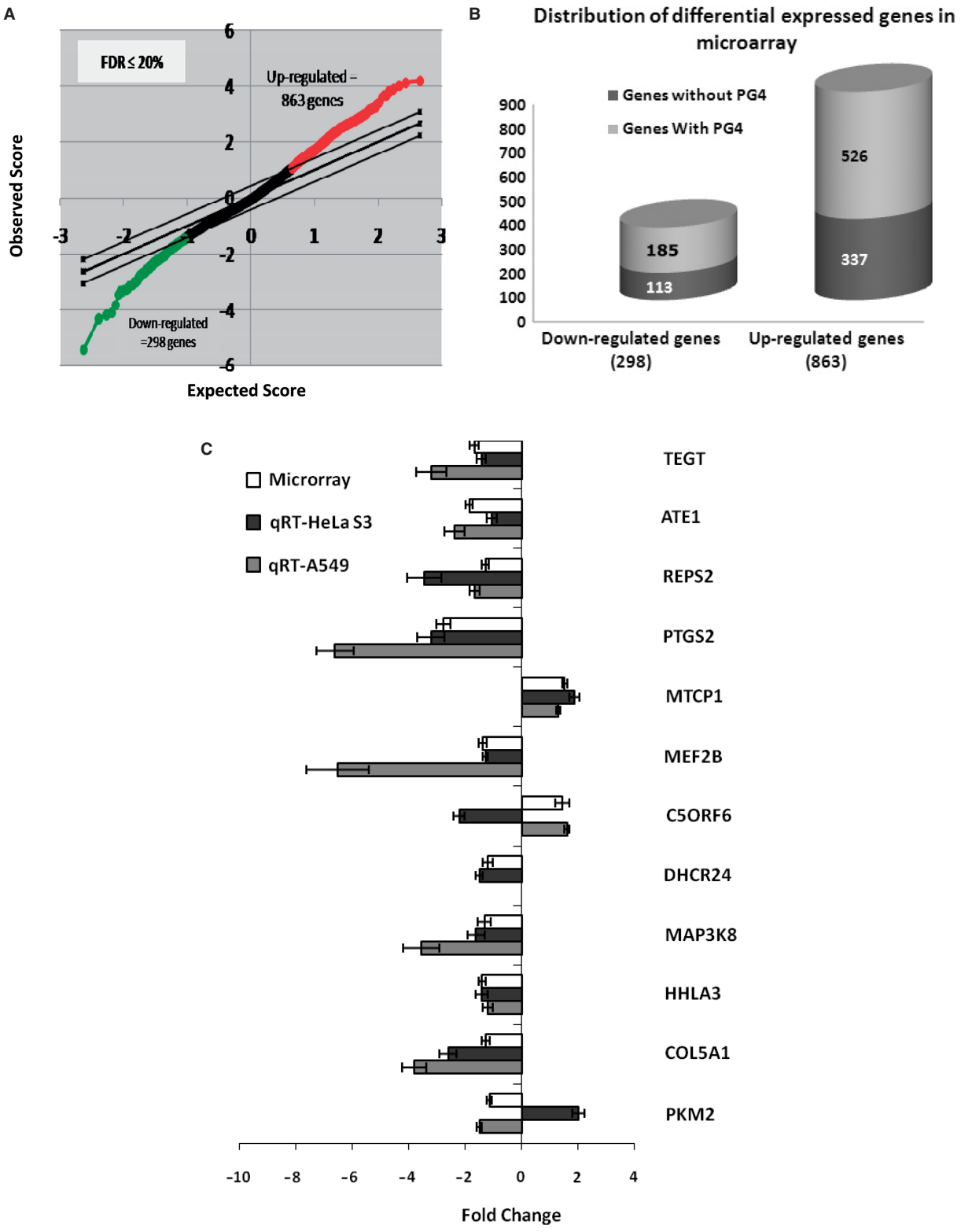

Figure 1. Genome-wide expression analysis shows wide-spread PG4 DNA-mediated gene regulation. (**A**) SAM plot showing overexpressed or repressed genes at ≤20% FDR; (**B**) distribution of significantly differentially expressed genes (≤20% FDR) in microarray based on presence/absence of PG4 within ±1 kb from transcription start sites. (**C**) Validation of microarray results by quantitative RT-PCR. Ten out of twelve reactions tested gave consistent change in expression on TMPyP4 treatment, in both cell lines. All selected genes (12) were at ≤20% FDR. Transcript level could not be detected for *DHCR24* gene in A549 cells.

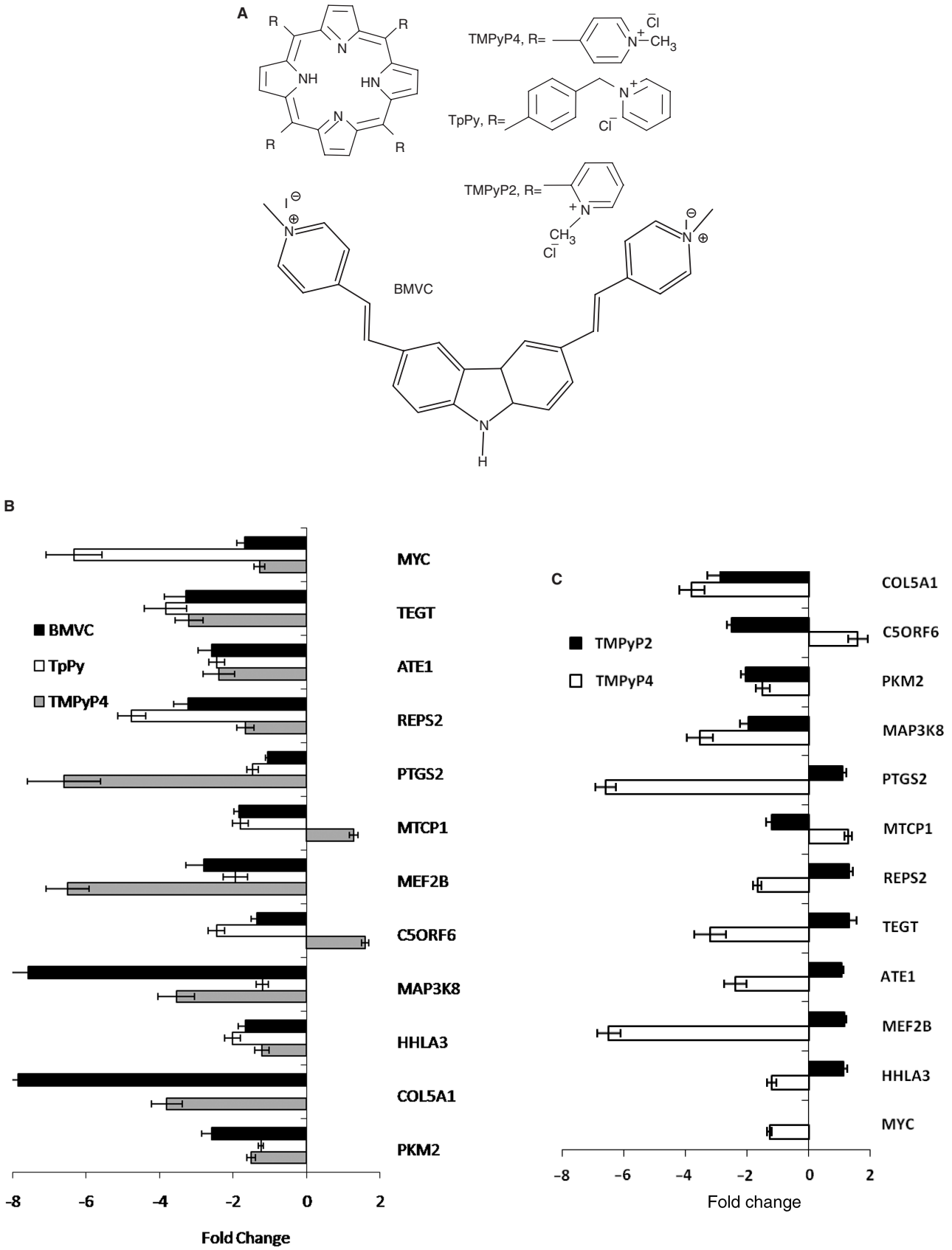

Figure 2. Gene expression change in presence of G4 DNA-binding ligands was consistent with TMPyP4 but not TMPyP2, which does not bind to G4 DNA. (**A**) Scheme showing the structure of ligands used. Quantitative RT-PCR results 48 h after treatment of A549 cells with BMVC/TpPy/TMPyP4 (**B**) or TMPyP4/TMPyP2 (**C**). Average of at least three measurements is given for each PCR.

in TMPyP4 treatment (Figure 2C). In contrast to TMPyP4 treatment, five of these genes were expressed at 1.2-fold or lower and two genes were expressed at ~1.3-fold on TMPyP2 treatment. However, we did note that three of the 12 genes showed expression that was similar to that of TMPyP4.

To understand the overall relevance in terms of biological function, we asked whether differentially expressed genes were significantly over- or underrepresented in any functional categories. The Gene Ontology (GO) database (http://www.geneontology.org/) was used for this analysis. Out of 1161 differentially expressed genes, 850 genes (for which uniprot IDs were available) were used for GO analysis. Out of 850 genes, 631 genes were upregulated and 219 downregulated in the microarray experiment. In this set of 850 genes, many were found to be significantly ($P < 0.05$ after Bonferroni correction for multiple hypothesis testing) overrepresented in functions related to intracellular transport, cellular localization and protein binding. In support of this, we noted a recent bioinformatics study which found overrepresentation of G4 DNA motifs in promoters of genes related to 'cation transport' (39), which is a subclassification of 'intracellular transport' observed within the differentially expressed genes analyzed by us. Another important functional class observed in the previous report was 'transcriptional activator activity'—we noted this at a slightly lower significance level ($P = 0.058$). On the other hand, functions related to receptor activity (transmembrane, rhodopsin-like and G-protein-coupled receptors) were significantly depleted in the GO result obtained from analysis of the 850 genes. Depletion of PG4 DNA harboring genes in functions related to G-protein-coupled- receptor protein signaling pathway ($P = 2.46E-06$) and receptor activity ($P = 3.071E-04$) was also observed before (39). All significant classes along with *P*-values are given in Supplementary Table 1. Enriched functional classes observed on using the genes that changed expression significantly (this study) were in many cases different from those observed in the earlier bioinformatics study based on presence of putative G4 DNA within 1 kb of TSS (56). This is not surprising given that the presence of putative G4 DNA may not always result in formation of the structural motif(s). Alternatively, TMPyP4 interaction with G4 DNA may not be a necessary condition for regulatory changes. In a recent bioinformatics study Du *et al.* (56) showed correlation of potential quadruplex sequence (PQS) present in downstream regions (+500 bases with respect to TSS) with increased gene expression, which was not noticed on considering upstream PQS within –500 bp of TSS. Based on this it was predicted that downstream G4 DNA may be more relevant in the context of gene expression. Our findings indicating experimentally determined gene expression that may be related to G4 DNA presents an interesting way to test the above prediction. We used all the genes that have shown significant expression change in our data and harbor PG4 DNA and asked whether differential presence of PG4 DNA could be seen 500 bases upstream or downstream of TSS. Any bias toward presence of downstream PG4 *vis-à-vis* upstream PG4 would be in line with the prediction made by Du *et al*. We noted 56.8% of all PG4s in the differentially expressed genes occurred within 500 bases downstream of TSS in contrast to 43.2% of PG4s, which were present upstream. However, this does not appear to be significant when compared to all human genes with PG4 DNA within 500 bases of TSS where about 54% of the PG4s are found downstream of TSS.

PG4 DNA is conserved within promoters of differentially expressed genes

In addition to testing the role of PG4 DNA in genome-wide expression we explored whether PG4 DNA within promoters of genes were conserved in other related species since biologically meaningful G4 DNA would most likely be conserved. In our expression experiments we noted that 711 genes were differentially expressed (≤20% FDR) and also harbored PG4 DNA within ±1 kb centered at TSS. Expression of these genes was more likely to be influenced by the presence of PG4 DNA. Therefore, we checked for conservation of PG4 DNA found within these 711 promoters. In keeping with our interest on the structural PG4 DNA, we considered conservation at the level of structure and not sequence; in other words, conserved elements would not have identical sequence but are most likely to result in the G4 motif structure nevertheless. Conservation of PG4 DNA that was found in human promoters (within ±1 kb flanking TSS) was checked within promoters of orthologous mouse and rat genes. We defined conservation as presence of PG4 DNA within ±200 bases flanking the PG4 DNA on human in corresponding gene pairs. The 200 base window considered on either side of the PG4 DNA occurrence was based on a previous study which showed that most *cis*-regulatory sites are conserved within ~400 bases in whole-genome comparison of human, mouse, rat and dog genomes (46). Orthologs in both mouse and rat was found for only 479 out of the 711 genes. Out of the 479 'orthologous' promoters thus identified, 71 promoters had at least one conserved PG4 DNA in all the three species, human, mouse and rat (57 expected by chance, $P = 0.6$; χ test). List of these genes and their significance level in the microarray analysis (FDR) (showing the reliability with which the changed level of expression was determined on treatment with TMPyP4) are given in Supplementary Table 2. In order to find out whether the 71 genes belonged to any particular functional class we performed GO analysis. Significant overrepresentation was observed in three biological processes, protein binding, nucleotide metabolism and biosynthesis ($P < 0.05$; after Bonferroni correction). Protein binding was also overrepresented in the GO analysis done with all 1161 genes differentially expressed on TMPyP4 treatment.

Intramolecular G4 DNA structures are single-stranded fold back forms and thus their formation would require strand separation within a duplex genome. In this context dynamics of G4 DNA folding has been addressed by various groups, under *in vitro* conditions in the presence of the second strand (57,58). Therefore, mere occurrence of sequence that forms G4 DNA may be of limited functional significance. In this context, conserved PG4 DNA

was most likely to be functionally relevant as *cis*-regulatory motifs. Based on these, we reasoned that presence of a unique functional regulatory site was expected to show largely coordinated expression of a group of genes across specific tissues (46). With this in mind, we checked tissue-specific expression of the 71 genes having conserved PG4 DNA within promoters in the expression profile of 79 human tissues reported earlier (45). For this analysis, expression of the group of 71 genes was compared against randomly picked genes to check whether the expression of the 71 genes when taken together changed in a statistically significant manner relative to random sets in a given tissue (see 'Material and Methods' section). Interestingly, 72 out of 79 tissues exhibited significantly changed expression (either up- or downregulation) of the 71 genes that harbor conserved PG4 DNA (z-score > 3.5; $P < 0.0001$ as shown in Figure 3). Genes (200) without a PG4 DNA (within ±5 kb of TSS) or no PG4 DNA within ±1 kb of TSS showed no significant enrichment (Figure 3, human tissue expression profile). Therefore, it is likely that PG4 DNA presence near TSS is affecting tissue-specific expression while presence of PG4 DNA in regions further upstream or downstream may not be directly involved. However, more work is required to rule out the effect of other cellular modulators [several G4 DNA-binding proteins are known (59) that may be involved in a tissue-specific fashion]. As the PG4 DNA were conserved within 71 'orthologous' mouse promoters, we further validated our observations by checking tissue-specific expression of the 71 mouse orthologous genes in mouse tissues. For this analysis we took gene expression microarray data available for 61 normal mouse tissues (45) and performed a similar analysis as discussed above. Here also we observed that the 71 genes with conserved PG4 DNA within promoters showed significantly changed expression in 56 out of the 61 tissues analyzed (z-score > 3.5, $P < 0.0001$; Figure 3) in line with our previous observation.

In a previous genome-wide comparative bioinformatics study involving *Escherichia coli* and 17 other bacterial genomes we observed striking enrichment of G4 DNA within promoters (37). *In silico* analysis also indicated that G4 DNA may regulate more than 1000 genes in the growth phase of *E. coli* when supercoiling results in formation of single strands (37). In support of this, formation of non-B DNA motifs have been demonstrated to regulate several operons in *E. coli* when the duplex character of the genome is destabilized giving single-stranded forms (60). Emerging evidence shows G4 DNA enrichment is wide-ranging in bacterial promoters [we studied 146 different microbes (40)]. Recent evidence also supports enrichment of G4 DNA within promoters of human and other related genomes (39,40,51). These findings along with demonstrated experimental evidence of the role of G4 DNA in regulation of the *c-MYC* (20) and *PDGF-A* (21) genes strongly implicate G4 DNA as regulatory elements. In a recently conducted genome-wide comparative bioinformatics study involving human, chimpanzee, mouse and rat, we asked whether G4 DNA motifs preserve characteristics of global *cis*-regulatory elements. We noted with interest that PG4 DNA found within human promoters were conserved with more than 700 mouse and rat promoters of orthologous genes; we also performed genome-wide expression experiments using TMPyP4 in cervical epithelial carcinoma (HeLa S3) cells where about 69 genes responded significantly at 48 h (42).

In the current study we first extend our previous findings by validating TMPyP4-mediated genome-wide expression in a second cell line (A549) wherein a combined analysis of the microarray results (considering both the cell lines together) demonstrated an extensive repertoire of genes that could be under G4 DNA regulation supporting previous bioinformatics predictions (39,40,51). Many genes, which changed expression also harbored PG4 DNA forming sequence (identified using bioinformatics approaches) in their promoters supporting the role of G4 DNA as a *cis*-regulatory element. To further support this, we examined and found that 71 genes that responded significantly in microarrays also conserved PG4 DNA within promoters of orthologous genes in related species (mouse and rat). In this study we noted both upregulation as well as downregulation of gene expression as a result of TMPyP4 treatment. It is interesting to consider the possibilities as both stabilization and destabilization of the G4 DNA motif on interaction with TMPyP4 have been observed. In cases where G4 DNA positively influences gene expression [e.g. insulin-linked polymorphin region upstream of insulin gene (19)], stabilization may lead to enhanced expression and conversely, disruption of the G4 DNA would result in down regulation. On the other hand, in cases like *c-MYC* (20) and *PDGF-A* (21) where the G4 DNA motif may be a negative regulator, TMPyP4-mediated stabilization would result in decreased promoter activity. Destabilization in this case would again result in upregulated expression of the gene. In all the above mechanistic models, influence of the G4 DNA motif as a *cis*-regulatory unit is central. Moreover, considering that other intracellular factors may associate directly or indirectly with G4 DNA many other mechanistic models can be envisaged.

Based on the findings presented here, it is tempting to speculate widespread influence of G4 DNA on gene regulation as a result of perturbation in chromatin, thereby suggesting a mechanistic link between local and global chromosomal structure to specific gene regulation events. This view is supported by observations indicating that non-B DNA structures, resulting from supercoiling induced during transcription, recruit structure-specific regulatory factors (61). Our study suggests such a possibility at a genome-wide level and warrants further experiments to specifically probe role of G4 DNA at a molecular level, particularly in connection with factors that may be involved directly with G4 DNA during transcription. In this context it is noteworthy that the nonmetastatic protein NM23-H2, which is a transcriptional activator of *c-MYC* (62), binds a G4 DNA structure present in the promoter of *c-MYC* (63). When considered with recent reports that add to the repertoire of experimentally characterized G4 DNA found within human gene promoters (20,21,25,64), an interesting possibility of another dimension to gene regulation is revealed.

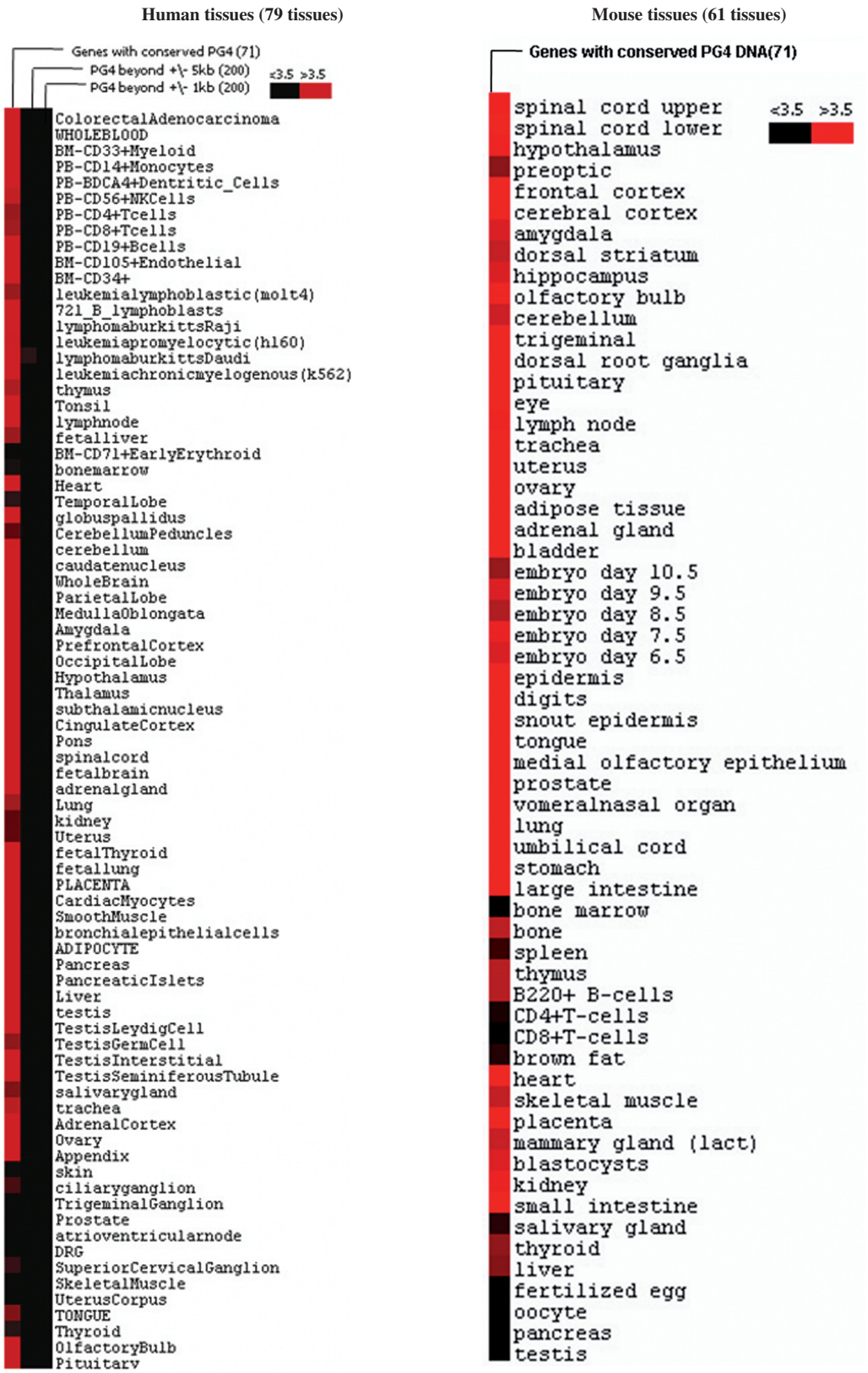

Figure 3. Tissue-specific expression of genes with conserved PG4 DNA in promoters. 71 genes show enriched expression (either up- or down-regulated) in 72/79 human and 56/61 mouse (normal) tissues. Statistical tissue-specific expression enrichment (z-score > 3.5, $P < 0.0001$) is represented in pseudo-color for both human and mouse tissues. Genes (200) without PG4 DNA in ±1 kb or ±5 kb (centered at TSS) do not show enrichment of expression in 79 human tissues (Figure 1).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The microarray data sets are available from GEO (Gene Expression Omnibus) under accession number GSE9710. Authors thank Prof. Ta-Chau Chang for generous gift of the ligand BMVC and anonymous referees for helpful comments.

FUNDING

Research fellowships by CSIR (to A.V., V.Y. and A.K.); Task Force Project CMM 0017 (to R.B.); research grant from Department of Biotechnology (PR6752) and Council of Scientific and Industrial Research (CMM0017), Government of India to SC.

Conflict of interest statement. None declared.

REFERENCES

1. Bates,P., Mergny,J.L. and Yang,D. (2007) Quartets in G-major. The First International Meeting on Quadruplex DNA. *EMBO Rep.*, **8**, 1003–1010.
2. 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.
3. Wilson,W.D. and Sugiyama,H. (2007) First International Meeting on Quadruplex DNA. *ACS Chem. Biol.*, **2**, 589–594.
4. Stuart,A.B. (2007) Ascent Of quadruplexes-nucleic acid structures become promising drug targets. *Chem. Eng. News*, **85**.
5. 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.
6. Balagurumoorthy,P., Brahmachari,S.K., Mohanty,D., Bansal,M. and Sasisekharan,V. (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.*, **20**, 4061–4067.
7. Balagurumoorthy,P. and Brahmachari,S.K. (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.*, **269**, 21858–21869.
8. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
9. 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.
10. Chang,C.C., Kuo,I.C., Ling,I.F., Chen,C.T., Chen,H.C., Lou,P.J., Lin,J.J. and Chang,T.C. (2004) Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative. *Anal. Chem.*, **76**, 4490–4494.
11. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Pluckthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl Acad. Sci. USA*, **98**, 8572–8577.
12. Bearss,D.J., Hurley,L.H. and Von Hoff,D.D. (2000) Telomere maintenance mechanisms as a target for drug development. *Oncogene*, **19**, 6632–6641.
13. Kerwin,S.M. (2000) G-Quadruplex DNA as a target for drug design. *Curr. Pharm. Des.*, **6**, 441–478.
14. Fedoroff,O.Y., Salazar,M., Han,H., Chemeris,V.V., Kerwin,S.M. and Hurley,L.H. (1998) NMR-Based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry*, **37**, 12367–12374.
15. Izbicka,E., Wheelhouse,R.T., Raymond,E., Davidson,K.K., Lawrence,R.A., Sun,D., Windle,B.E., Hurley,L.H. and Von Hoff,D.D. (1999) Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res.*, **59**, 639–644.
16. Sun,D., Thompson,B., Cathers,B.E., Salazar,M., Kerwin,S.M., Trent,J.O., Jenkins,T.C., Neidle,S. and Hurley,L.H. (1997) Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *J. Med. Chem.*, **40**, 2113–2116.
17. Dixon,I.M., Lopez,F., Esteve,J.P., Tejera,A.M., Blasco,M.A., Pratviel,G. and Meunier,B. (2005) Porphyrin derivatives for telomere binding and telomerase inhibition. *Chembiochem*, **6**, 123–132.
18. Seenisamy,J., Rezler,E.M., Powell,T.J., Tye,D., Gokhale,V., Joshi,C.S., Siddiqui-Jain,A. and Hurley,L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.*, **126**, 8702–8709.
19. Catasti,P., Chen,X., Moyzis,R.K., Bradbury,E.M. and Gupta,G. (1996) Structure–function correlations of the insulin-linked polymorphic region. *J. Mol. Biol.*, **264**, 534–545.
20. 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.
21. Qin,Y., Rezler,E.M., Gokhale,V., Sun,D. and Hurley,L.H. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.*, **35**, 7698–7713.
22. Fry,M. and Loeb,L.A. (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl Acad. Sci. USA*, **91**, 4950–4954.
23. Weisman-Shomer,P., Cohen,E., Hershco,I., Khateb,S., Wolfovitz-Barchad,O., Hurley,L.H. and Fry,M. (2003) The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n. *Nucleic Acids Res.*, **31**, 3963–3970.
24. 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.
25. 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.
26. Sun,D., Guo,K., Rusche,J.J. and Hurley,L.H. (2005) 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.*, **20**, 6070–6080.
27. Dai,J., Dexheimer,T.S., Chen,D., Carver,M., Ambrus,A., Jones,R.A. and Yang,D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.*, **128**, 1096–1098.
28. Xu,Y. and Sugiyama,H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.*, **34**, 949–954.
29. Kumari,S., Bugaut,A., Huppert,J.L. and Balasubramanian,S. (2007) An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.*, **3**, 218–221.
30. Wieland,M. and Hartig,J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, **14**, 757–763.
31. Li,J.L., Harrison,R.J., Reszka,A.P., Robert,M., Bohr,V.A., Neidle,S. and Hickson,I.D. (2001) Inhibition of the Bloom's and Werner's Syndrome Helicases by G-Quadruplex Interacting Ligands. *Biochemistry*, **40**, 15194–15202.
32. Bock,L.C., Griffin,L.C., Latham,J.A., Vermaas,E.H. and Toole,J.J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*, **355**, 564–566.
33. Yafe,A., Shklover,J., Weisman-Shomer,P., Bengal,E. and Fry,M. (2008) Differential binding of quadruplex structures of muscle-specific genes regulatory sequences by MyoD, MRF4 and myogenin. *Nucleic Acids Res.*, **36**, 3916–3925.
34. Walsh,K. and Gualberto,A. (1992) MyoD binds to the guanine tetrad nucleic acid structure. *J. Biol. Chem.*, **267**, 13714–13718.
35. Arimondo,P.B., Riou,J.F., Mergny,J.L., Tazi,J., Sun,J.S., Garestier,T. and Helene,C. (2000) Interaction of human DNA

topoisomerase I with G-quartet structures. *Nucleic Acids Res*., **28**, 4832–4838.
36. Sun,H., Yabuki,A. and Maizels,N. (2001) A human nuclease specific for G4 DNA. *Proc. Natl Acad. Sci. USA*, **98**, 12444–12449.
37. Rawal,P., Kummarasetti,V.B., Ravindran,J., Kumar,N., Halder,K., Sharma,R., Mukerji,M., Das,S.K. and Chowdhury,S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. *Genome Res*., **16**, 644–655.
38. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*., **33**, 2908–2916.
39. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*., **35**, 406–413.
40. Yadav,V.K., Abraham,J.K., Mani,P., Kulshrestha,R. and Chowdhury,S. (2007) QuadBase: genome-wide database of G4 DNA occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res*., **36**, 381–385.
41. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res*., **33**, 2901–2907.
42. 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.
43. Tusher,V.G., Tibshirani,R. and Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
44. Larionov,A., Krause,A. and Miller,W. (2005) A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics*, **6**, 62.
45. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al*. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci*., **101**, 6062–6067.
46. Xie,X., Lu,J., Kulbokas,E.J., Golub,T.R., Mootha,V., Lindblad-Toh,K., Lander,E.S. and Kellis,M. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
47. Gibbon,J.D. and Chakraborty,S. (2007) *Nonparametric Statistical Interference*. Marcel Dekker, New York.
48. Martin,D., Brun,C., Remy,E., Mouren,P., Thieffry,D. and Jacq,B. (2004) GOToolBox: functional analysis of gene datasets based on Gene Ontology. *Genome Biol*., **5**, R101.
49. Weisman-Shomer,P., Cohen,E. and Fry,M. (2002) Distinct domains in the CArG-box binding factor A destabilize tetraplex forms of the fragile X expanded sequence d(CGG)n. *Nucleic Acids Res*., **30**, 3672–3681.
50. Du,Z., Kong,P., Gao,Y. and Li,N. (2007) Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochem. Biophys. Res. Commun*., **354**, 1067–1070.
51. Zhao,Y., Du,Z. and Li,N. (2007) Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals. *FEBS Lett*., **581**, 1951–1956.
52. 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.
53. Yang,D.Y., Chang,T.C. and Sheu,S.Y. (2008) Interaction between human telomere and a carbazole derivative: a molecular dynamics simulation of a quadruplex stabilizer and telomerase inhibitor. *J. Phys. Chem. A*, **111**, 9224–9232.
54. Han,F.X., Wheelhouse,R.T. and Hurley,L.H. (1999) Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. *J. Am. Chem. Soc*., **121**, 3561–3570.
55. Han,H., Langley,D.R., Rangan,A. and Hurley,L.H. (2001) Selective Interactions of Cationic Porphyrins with G-Quadruplex Structures. *J. Am. Chem. Soc*., **123**, 8902–8913.
56. 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.
57. Halder,K. and Chowdhury,S. (2007) Quadruplex-coupled kinetics distinguishes ligand binding between G4 DNA motifs. *Biochemistry*, **46**, 14762–14770.
58. Halder,K. and Chowdhury,S. (2005) Kinetic resolution of bimolecular hybridization versus intramolecular folding in nucleic acids by surface plasmon resonance: application to G-quadruplex/duplex competition in human *c-myc* promoter. *Nucleic Acids Res*., **33**, 4466–4474.
59. Fry,M. (2007) Tetraplex DNA and its interacting proteins. *Front Biosci*., **12**, 4336–4351.
60. Hatfield,G.W. and Benham,C.J. (2002) DNA topology-mediated control of global gene expression in Escherichia coli. *Annu. Rev. Genet*., **36**, 175–203.
61. Kouzine,F., Sanford,S., Elisha-Feil,Z. and Levens,D. (2008) The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol*., **15**, 146–154.
62. Postel,E.H. (1999) Cleavage of DNA by human NM23-H2/nucleoside diphosphate kinase involves formation of a covalent protein-DNA complex. *J. Biol. Chem*., **274**, 22821–22829.
63. Thakur,R.K., Kumar,P., Halder,K., Verma,A., Kar,A., Parent,J.L., Basundra,R., Kumar,A. and Chowdhury,S. (2009) Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res*., **37**, 172–183
64. Sun,D., Liu,W.J., Guo,K., Rusche,J.J., Ebbinghaus,S., Gokhale,V. and Hurley,L.H. (2008) The proximal promoter region of the human vascular endothelial growth factor gene has a G-quadruplex structure that can be targeted by G-quadruplex-interactive agents. *Mol. Cancer Ther*., **7**, 880–889.