Transcriptome-wide identification of transient RNA G-quadruplexes in human cells

Sunny Y. Yang, Pauline Lejault, Sandy Chevrier, Romain Boidot, A. Gordon Robertson, Judy M.Y. Wong & David Monchaud

Guanine-rich RNA sequences can fold into four-stranded structures, termed G-quadruplexes (G4-RNAs), whose biological roles are poorly understood, and in vivo existence is debated. To profile biologically relevant G4-RNA in the human transcriptome, we report here on G4RP-seq, which combines G4-RNA-specific precipitation (G4RP) with sequencing. This protocol comprises a chemical crosslinking step, followed by affinity capture with the G4-specific small-molecule ligand/probe BioTASQ, and target identification by sequencing, allowing for capturing global snapshots of transiently folded G4-RNAs. We detect widespread G4-RNA targets within the transcriptome, indicative of transient G4 formation in living human cells. Using G4RP-seq, we also demonstrate that G4-stabilizing ligands (BRACO-19 and RHPS4) can change the G4 transcriptomic landscape, most notably in long non-coding RNAs. G4RP-seq thus provides a method for studying the G4-RNA landscape, as well as ways of considering the mechanisms underlying G4-RNA formation, and the activity of G4-stabilizing ligands.
The biological functions and cellular regulations of RNA structures are dependent on their secondary and tertiary levels. RNA can adopt intricate bulged, stem–loop structures involving duplex-, triplex-, and quadruplex-RNA motifs. G-quadruplexes (G4s) are structures formed by Hoogsteen bonding of four guanines to form planar guanine quartet (G-quartet) units, which π-stack on each other, to assemble into columnar four-stranded structures with the central cavity stabilized by monovalent cations (i.e., $K^+$, $Na^+$). G4 folding is spontaneous in vitro and results in a highly stable structure. While both single-stranded DNA and RNA can fold into G4s, the latter is less studied, even though G4 formation in vitro can fold more readily due to their predominant single-stranded nature in vivo.

Formation of G4-RNA has recently been implicated in key RNA metabolism events, including the regulation of RNA processing and translation. To better understand the roles of G4-RNA in cell biology, there has been a strong interest in mapping the distribution of G4-RNAs within the human transcriptome. However, to date, in vitro and in vivo evidence appear to be contradictory, limiting interpretation of the relevance of G4-RNA. In vitro transcription experiments suggest that opportunities exist for G4-RNA formation during transcriptional folding of nascent RNA. Kwok et al. used reverse transcription (RT)-stall-coupled with next-generation sequencing to map thousands of G4-RNA sites in vitro, showing widespread potential G4-forming sites within the human transcriptome. In contrast, Guo and Bartel showed through in vivo DMS-mediated RNA modification with RT-stall-sequencing that G4s are nearly entirely in an unfolded state in vivo in mammalian cells. This observation is surprising since it has widely been assumed that G4-RNAs are formed in vivo, at least transiently. Conversely, evidence for in vivo G4 formation has been provided by cellular imaging studies using G4-specific antibodies and probes. To reconcile these results, we hypothesized that G4-RNAs must be able to form, at least transiently, in live human cells, and that the identity of these G4-RNAs may provide valuable insights into their regulatory mechanisms and functions. However, because none of the above methods are suitable for capturing transient G4s, an alternative approach was needed.

Here, we report on a small-molecule-based approach to assessing the existence and the identity of transient G4-RNAs in the human transcriptome. We designed a biotinylated version of a previously characterized G4-specific ligand/probe, template-assembled synthetic G-quartet (TASQ), which self-assembles into a synthetic G-quartet upon association with a G4 target through end-quartet stacking. We developed a protocol, G4R- seq (G4-RNA-specific precious and sequencing) using BioTASQ to capture G4-RNAs from human breast cancer cells in log-phase growth. Using this protocol to characterize in vivo transcriptomic landscapes, we showed that more G4s are present in gene transcripts that are GC-rich and have higher densities of predicted G4 motifs. We also evaluated G4 ligand-induced changes to the G4-RNA landscape following treatments with the G4 ligand, BRACO-19 or RHPS4, showing both similarities and differences in their respective induction profiles. Our data show that G4-RNAs can be ligand-induced in diverse RNA entities that include long non-coding RNAs; further, differential G4-RNAs induced by G4 ligands suggest that specific G4 structure–ligand interactions could be exploited.

**Results**

BioTASQ selectively captures G4 targets in vitro. To support affinity purification and identification of functional transcriptomic G4-RNA targets, we added a biotin tag to the biomimetic quadruplex ligands known as TASQ (Fig. 1a). Known for their high G4 selectivity and, for some of them, their ability to track G4-RNAs in live cells (N-TASQ), detailed synthesis and characterization of the biotinylated TASQ, or BioTASQ, can be found in Supplementary Figure 1 and Methods. We first evaluated the G4-interacting properties of BioTASQ via a fluorescence resonance energy transfer (FRET)-melting assay (Supplementary Figure 2). Detailed melting experiments were performed with BioTASQ against a panel of dual-labeled nucleic acid sequences that included: (a) three G4-DNAs (F-Myc-T and F-kit-T, found in the promoter regions of MYC and KIT gene, respectively, and F21T, the human telomeric sequence), (b) one G4-RNA (F-TERRA-T, the human telomeric transcript), and (c) one duplex-DNA as a control (F-DS-T) (Supplementary Figure 2). Introduction of a biotin tag negatively impacted G4-affinity of BioTASQ (further confirmed by ESI-MS measurements) compared to the non-biotinylated parent. However, it did not affect the ability of BioTASQ to discriminate G4s over duplexes.

Next, we assessed whether BioTASQ could efficiently capture G4s from solutions in vitro. We used fluorescein (F)-labeled nucleic acids since fluorescence signal measurements allow for convenient and sensitive detection of the ligand/G4 assemblies. We performed these experiments with: (a) three G4-DNAs (F-Myc and F-SRC, two sequences found in the promoter regions of MYC and SRC gene, respectively, (b) F-22AG, the human telomeric sequence); and (c) one G4-RNA as a control (F-Duplex). Labeled oligonucleotides (1 μM) were incubated with BioTASQ (20 μM) and streptavidin-coated magnetic beads (MagneSphere®, 25 μL). After overnight incubation at 25°C, streptavidin beads were precipitated, the supernatant removed, and the beads resuspended in denaturing buffer before thermal denaturation (8 min at 90°C). After separation from the beads (via centrifugation and magnet immobilization), fluorescent signals from the supernatant solutions were measured to quantify nucleic acid capture. Our results confirmed the efficiency of BioTASQ-mediated G4 capture (Fig. 1b) and showed that the level of G4s recovered was dependent on both the G4 nucleic acid type and topology. G4-DNA was enriched 4.1–20.7-fold, whereas G4-RNA was enriched between 10.9- and 23.8-fold when compared to controls, while duplex-DNA was not enriched. BioTASQ appeared to have stronger preferences for certain G4 topologies, as type I (or ‘parallel’) G4s displayed better enrichment than type II (or ‘mixed-hybrid’) G4s (with 16.6-, 20.7-, 20.6-, 10.9-, and 23.8-fold enhancement for F-Myc, F-SRC, F-TERRA, F-TRF2, and F-NRAS, respectively, versus 4.1-fold enhancement for F-22AG). The preference of BioTASQ for type I G4, which displays accessible external G-quartets, is expected, given that TASQs are sterically demanding ligands that require accessible, loopless G-quartets for binding G4 targets efficiently. This property represents a limitation to the use of BioTASQ for G4 detection, especially for G4-DNAs, which have higher conformational diversity than G4-RNAs. We have not yet tested the affinity of BioTASQ on the recently reported antiparallel G4-RNA. While the cellular prevalence of G4-RNA with antiparallel topology remains to be established, they may provide key insights into the topological preference of TASQ ligands. We confirmed that the streptavidin bead/BioTASQ system did not extract duplex-DNA (0.7-fold). We further confirmed G4 selectivity of BioTASQ via competitive pull-down experiments, which we performed with
F-SRC and F-NRAS (1 μM) in the presence of an excess of duplex-DNA (either ds17 or ds26, 20 μM) or DNA extracts (calf thymus DNA, ctDNA, 100 μM, expressed in base pairs). The capture efficiencies of the fluorescently labeled G4-RNA were not significantly affected by an excess of synthetic duplexes (90–112% with F-SRC, 74–82% with F-NRAS) or with DNA extracts (76% and 90% with F-SRC and F-NRAS, respectively). We also tested BioTASQ/streptavidin association with an excess of either biotin (80 μM, to compete with streptavidin interaction) or PNADOTASQ (10 μM, to compete with G4 interaction) (Fig. 1c). Together, these results show the strong ternary interaction between G4s, BioTASQ, and streptavidin (beads), which provided the basis for the development of our G4RP protocol (described below).

**G4RP isolates G4-RNA targets from human cell extracts.** After confirming that BioTASQ could interact with and capture G4s in vitro, we then assessed whether it could capture G4 targets from human cell extracts. For this, we developed the G4-RNA-specific precipitation (G4RP) protocol, a modified version of the commonly used RNA-immunoprecipitation (RIP) protocol\(^{30}\). MCF7 cells were first crosslinked with formaldehyde to halt biological processes and stabilize transient structural interactions. Harvested cells were then sonicated briefly to release cellular content. Cell lysates were incubated with a high concentration of BioTASQ (100 μM) overnight (Supplementary Figure 3C) before affinity purification with magnetic streptavidin beads.

We first used RT quantitative PCR (RT-qPCR) with genespecific primers to confirm the efficiency of the G4RP protocol\(^{31}\). G4RP-qPCR analysis of RNAs extracted from the BioTASQ-enriched fractions showed that non-specific binding was negligible (black bars) (Fig. 2a, b) while demonstrating the enrichment of two known G4-forming mRNAs, i.e., VEGFA and NRAS\(^{27}\) (gray bars) (Fig. 2a, b).

To confirm bona fide G4 formation in target mRNA sequences, we collected MCF7 cell extracts following treatments with two well-established G4-stabilizing ligands, BRACO-19
and RHPS4\textsuperscript{32,33}. We chose the treatment ligand doses to be between the respective IC\textsubscript{15} and IC\textsubscript{25} (doses that are growth inhibitory in 15–25% of the cells; BRACO-19: 5 µg/mL and RHPS4: 1.5 µM), as determined by growth kinetics profiling of MCF7 in the presence of G4 ligands, compared with vehicle controls (Supplementary Figure 4). Treatment with G4 ligands significantly increased G4RP-qPCR signals by 6.5-fold, 3.6-fold, and 4.4-fold in BRACO-19-treated cells and by 10.2-fold, 4.2-fold, and 2.5-fold in RHPS4-treated samples in three selected G4-rich regions, VEGFA, NRAS, and TERF2, respectively, compared to the untreated control (Fig. 2c; Supplementary Figure 5), demonstrating that BioTASQ can specifically enrich for G4-containing RNA sequences. HPRT1 was selected as an unstructured RNA control, as this housekeeping mRNA is expected to have low G4-forming potential\textsuperscript{34}. Neither BRACO-19 nor RHPS4 treatment induced a significant change in BioTASQ-captured HPRT1 signals, confirming that these G4 ligands were selective for G4-rich targets. Collectively, these results indicate that the G4RP protocol is suitable for the purification and identification of G4-containing RNAs from human cell extracts, as well as the quantification of the G4 ligand-induced changes.

We performed G4RP in samples that were crosslinked to preserve the transiently formed cellular G4s while preventing the induction of G4 formation in vitro, as crosslinked nucleic acids would be in an immobilized state. We used the reported formaldehyde concentration and crosslinking conditions where over 90% of nuclear DNA are immobilized\textsuperscript{35} and anticipated that the crosslink efficiency for cellular RNA to be similar. As controls and to illustrate the importance of this crosslinking step, G4RP was performed in non-crosslinked samples. We selected three targets from the top- and bottom-ranked transcripts obtained with our G4RP-seq results (see below). BioTASQ enrichment of these targets was quantified using RT-qPCR and compared between the non-crosslinked and crosslinked samples (Supplementary Figure 6). We observed a loss of difference in BioTASQ enrichment in the non-crosslinked samples, despite the overall higher signals. These increased signals in non-crosslinked samples were likely G4s that were formed in vitro, due to the high concentration of BioTASQ, arguing that the crosslinking step is necessary to immobilize cellular RNA structures and to minimize the effects of in vitro G4 formation and destabilization through the biochemical evaluation steps. As naked, non-crosslinked RNA targets are susceptible to the in vitro G4 stabilization effects of high concentration of BioTASQ during the incubation steps, the true differences between the in vivo transient levels of the top and bottom ranking targets would be masked.

G4RP-Seq identifies transcriptome-wide transient G4-RNAs.

To survey the baseline in vivo G4 transcriptomic landscape, we performed G4RP followed by sequencing (G4RP-seq) in human breast cancer cells harvested at log-phase growth. Due to the crosslinking step in the G4RP protocol, we expected to capture global levels of transient G4s in the transcriptome. Notably, as G4 ligand treatment resulted in gene expression changes in human cells\textsuperscript{36}, we needed to account for these input differences; therefore, an internal input control was included for each treatment condition. To enrich for diverse G4-forming RNAs, we elected to remove ribosomal RNA targets at the cDNA library preparation step, as rRNA constitute over 80% of total cellular RNAs. Ribodepletion instead of poly-A selection was used, as we anticipated that non-poly-A RNAs, including many non-coding RNAs (ncRNAs), not only harbor but contribute to a substantial proportion of cellular G4s\textsuperscript{37}. For the sequencing analysis, two comparisons were required: BioTASQ versus input (which gives normalized global levels of transient G4s for a specific transcript as Enrichment Score (ES)), and G4 ligand-treated versus untreated cells (given as Enrichment Score Change (ASE)) (Supplementary Figure 7). After mapping reads to the hg19 reference human genome assembly with HISAT2\textsuperscript{38}, normalization and differential gene expression analyses were performed with HTSeq/DESeq2 pipeline\textsuperscript{39,40}, by comparing the BioTASQ-enriched samples with the corresponding inputs\textsuperscript{41} ( Supplementary Data 1). Of note, the relative enrichment levels by BioTASQ (given as ES) were not direct quantitative readouts of G4 formation but instead indicate the relative propensity of a specific transcript to fold into G4. ES values allow ranking of transcripts by relative G4-folding status under specific experimental conditions.

We observed BioTASQ enrichment of many gene transcripts, suggesting the existence of widespread G4-RNAs, at least transiently, in live human cells. This observation was expected since we were capturing a snapshot of the G4-RNA landscape in which some G4-rich sequences were in folded states, while others were in unfolded states. While the sequencing depth was not high enough to detect subtle changes in individual G4-forming sequences, we were able to confidently determine gene-level changes by focusing on a subset of highly abundant transcripts, filtered by high normalized mean read counts (as an estimate of relative expression level). We compared ES for each condition to generate initial lists of gene transcripts filtered by a minimal abundance threshold (>50 normalized mean read counts) (Supplementary Data 1). To ensure high-confidence hits with substantive changes, we further filtered the list and included only those with high transcript read abundance (≥500 mean read counts; approximately the top 5% of the list), which we used for
downstream analyses. Gene transcripts in this filtered list were then ES-ranked for downstream analyses (Supplementary Data 1D).

ES, a gene-specific ratio of the BioTASQ signal normalized to the corresponding input signal, moderately positively correlated with their respective G/C content (Pearson correlation = 0.43, p < 0.0001) (Fig. 3b). ES was uncorrelated to gene length (Pearson correlation = 0.02, p = 0.89) (Fig. 3c). To evaluate whether the ES is related to the density of potential G4 sequences, Quadbase2 was then used to assess the number of predicted G4 (pG4) motifs

Fig. 3 Characterization of the baseline level of G4-RNA landscape using G4RP-seq. a Top 10 highly abundant transcripts (filtered by at least 500 base-read counts) with the lowest BioTASQ enrichment (blue bars, normalized to the input in the untreated sample) or with the highest BioTASQ enrichment (red bars) ranked by Enrichment Scores. b Regression plot of BioTASQ-Enrichment Score for each transcript versus its corresponding G/C content ($R^2 = 0.187$, p < 0.001, significant non-zero relationship). c Regression plot of BioTASQ-Enrichment Score for each transcript versus its corresponding gene length ($R^2 = 0.00005$, p = 0.89, non-significant relationship). d Number of pG4 motifs (calculated by Quadbase2 using mid stringency G3L1–7) to gene length ratio plotted against the subset of highly abundant transcripts ranked by their BioTASQ enrichment. The bar graph (inset) shows the average pG4 motif/gene length ratio between the top 100 ranked transcripts versus the bottom 100 ranked transcripts (p < 0.001, significant difference, Student’s t-test)
(i.e. sequences with the canonical G3L1–7 G4 sequence) in each gene transcript. The ratios of the number of pG4s to gene length were calculated and then plotted against the transcripts ranked by their ES (Fig. 3d). Comparison of the ratios between the top and bottom 100 ES-ranked transcripts showed significantly higher values (2.2-fold difference) for top-ranked transcripts. Results from gene ontology analysis of the top and bottom 100 ES-ranked transcripts are summarized in Supplementary Table 1. Together, our bioinformatics analyses confirmed that transcripts with higher ES tend to have higher G/C content and higher pG4 density.

We also found that highly expressed lncRNAs had some of the lowest relative G4 levels, as shown by their low ES (Fig. 3a). These bottom-ranked transcripts included well-known lncRNAs such as MALAT1, RPPH1, RMRP, and XIST. The presence of MALAT1 on this list was unexpected, as it has been previously reported as a G4-forming lncRNA in vitro.11,10 Interestingly, residual rRNAs (that escaped from the ribodepletion step during library preparation) were also among the most depleted within this list.

G4RP-seq identifies ligand-induced changes in G4-RNA profiles. We evaluated ligand-induced changes in the G4-RNA landscapes by applying G4RP-seq to samples treated with G4 ligands and calculating an Enrichment Score Change (ΔES) from the ratio of treated versus non-treated samples. By filtering the initial list ΔES > 1.75, we found BioTASQ enrichment in 251 and 463 gene transcripts to be highly induced by BRACO-19 and RHPS4, respectively (Supplementary Data 1B). Results from gene ontology analysis performed on the lists of top genes are summarized in Supplementary Table 2.

Among the list of gene transcripts with a high ligand-induced increase (ΔES > 1.75) and at least 500 mean read counts, the lncRNAs MALAT1, RPPH1, and XIST were highly ranked in both BRACO-19- and RHPS4-induced gene lists (Fig. 4a). Some RNAs previously reported to harbor G4s in vitro were also identified, including NEAT1 and NBR2.9,27 We found no correlation between the ligand-induced ΔES and the read counts of the transcripts (Pearson correlation = 0.02, p = 0.69) (Supplementary Figure 8). The ligand-induced ΔES of each transcript was compared to the corresponding G/C content of the transcripts, which interestingly showed a negative correlation (Fig. 4b). When the transcripts were ordered by ΔES, the pG4 density appeared to be distributed toward lower scores (1.7-fold and 2.7-fold difference between the average ratio of top and bottom 100 ranked transcripts for BRACO-19 and RHPS4, respectively) (bottom panel) (Fig. 4b). Overall, our observations suggest that transcripts with higher pG4 density were more likely to be captured in a folded state in the absence of ligands, resulting in ΔES being lower due to the higher baseline level of G4 formation. In contrast, transcripts with lower pG4 density were more likely to be unfolded in the absence of ligands and to have their G4 structures stabilized in the presence of ligands, leading to a higher ΔES.

When we compared the absolute number of pG4 motifs (i.e. without normalization to gene length) between the three treatment conditions ranked by their respective ES, we observed differential changes in pG4 profiles between the two G4 ligand treatments (Fig. 4c). pG4 scores generated using different stringency of searches (G2L1–10, G3L1–5, and G3L1–7) showed similar trends (Supplementary Figure 9). BioTASQ-captured targets generated from BRACO-19-treated samples exhibited higher levels of pG4-dependent enrichment regardless of the search stringency, conceivably due to the broader range of intramolecular G4s (longer loops, 2-quartet G4s, etc.) stabilized by this ligand. On the other hand, targets generated from RHPS4-treated samples showed lower levels of pG4-dependent enrichment and a pG4-dependency could only be observed when the plots were obtained using G4 motif searches with the lower stringency. We reasoned that RHPS4-binding preference could be selective towards sequences with lower numbers of G4 motifs or highly specific G4 sequences that are less prevalent within the transcriptome. However, we cannot rule out the possibility of intermolecular G4s, as computational algorithms are currently unable to predict these structures. Given this caveat, the lack of pG4-dependent enrichment in samples treated with RHPS4 could be partially explained by a preferential ligand-induced stabilization of intermolecular G4s. The overlap between the gene lists for the two G4 ligand treatments demonstrated that they have differential G4-induction profiles, in agreement with their differential in vitro G4-structure-specific binding profiles (Fig. 4d)44.

We further validated our findings from G4RP-seq by performing qPCR on separate G4RP samples obtained with biological repeat experiments, using primers specific for the top three common lncRNA targets: MALAT1, XIST, and RPPH1.47 Fold change differences between G4RP-qPCR and G4RP-seq were seen and expected due to the small qPCR region amplified. Nevertheless, the qPCR results were consistent with those obtained from sequencing, confirming that these lncRNAs were targets of G4 ligands (Fig. 4e; Supplementary Figure 10). Circular dichroism (CD) and thermal differential spectra (TDS) analyses of the three selected pG4 regions of MALAT1, XIST, and RPPH1 were consistent with the formation of parallel-type G4 structures in vitro in the selected pG4 motif sequences extracted from these genes (Supplementary Figure 11). In summary, we observed that the BRACO-19 and RHPS4 treatments in MCF7 cells similarly induced G4 stabilization in several highly expressed lncRNA targets, but the treatments also displayed distinct ligand specificity towards other RNA targets.

Discussion
It has been widely assumed that G4s must be formed in vivo, at least transiently, due to the high structural stability of G4 nucleic acids and the favorable intracellular potassium concentration. The extent to which G4-RNA formation occurs in vivo in human cells is debated, since it has been recently reported that G4s are globally unfolded in the mammalian transcriptome.10 To address this conundrum, we developed and reported here on G4RP-seq, a protocol that provides evidence supporting the existence of transient G4-RNAs in the in vivo human transcriptome. We also used G4RP-seq to generate readouts of the changes in transcriptome-wide G4-RNA landscape upon treatments with G4 ligands.

Our first key finding of global transient G4-RNA formation suggests an alternative perspective on Guo and Bartel’s10 lack of detected G4-RNAs by DMS/RT-profiling, an observation that can be interpreted in two ways: either G4s are seldomly formed in vivo, or G4s are formed but are unfolded quickly by destabilizing mechanisms such as through the actions of helicases. DMS/RT profiling may be reporting on an unfolded G4 landscape since it measures the biological endpoint in which all G4s are eventually unfolded. A recent study corroborated this by demonstrating that dynamic folding and unfolding of some G4-RNAs in live cells could occur within seconds.36 We postulated that even if the G4 equilibrium heavily favors an unfolded G4 state, global snapshots of transiently folded G4s could still be captured by a chemical crosslinking step. Our observation of widespread G4-RNAs using a snapshot approach, combined with Bartel’s observation of endpoint globally unfolded G4-RNAs, suggest that in vivo G4-RNAs can form continuously and are rapidly resolved...
to an unfolded state. Further, we anticipate that the equilibrium between an RNA’s transient folded and unfolded states can be influenced by competing factors such as the transcription rate of the RNA, the availability of helicases31, the chaperone activities of RNA-binding proteins52, and the structural stabilization by G4 ligands53 (Fig. 5). Importantly, the effectiveness of capturing these transient G4-RNAs within transcripts, as measured by BioTASQ enrichment, is correlated with their G/C content and pG4 density. This observation suggests that gene transcripts with a higher density of pG4 motifs would have a higher probability of having at least one of these motifs captured in a folded state with our crosslinking snapshot approach. It is currently undetermined whether such transient G4s have any biological roles or are simply stochastic passenger events that arise from the chemical
properties of guanines. Strikingly, IncRNAs appeared to avoid G4-formation and were found mainly in an unfolded state, in the absence of G4 ligands. The functions and interactions of IncRNAs with other molecules (DNA, RNA, and protein) are dependent on their folding into higher-order structures, and the formation of G4 may interfere with such folding. Consistent with this, a study on hnRNP F binding and G4 formation showed that hnRNP F ribonucleoprotein formation prevents G4 formation/reformation by sequestering the G-tracts in a single-stranded RNA state. As the rate of protein-RNA complex formation is faster than that for G4 formation, the shift of cellular G4-RNAs toward an unfolded state could be partially due to interactions with RNA-binding proteins.

The second key finding from our report is the characterization of G4 ligand treatment effects on the G4-RNA folding state within the in vivo human transcriptome. While previous studies have characterized G4 ligand-induced changes for human G4-formation and were found mainly in an unfolded state, in the absence of G4 ligands. The functions and interactions of IncRNAs with other molecules (DNA, RNA, and protein) are dependent on their folding into higher-order structures, and the formation of G4 may interfere with such folding. Consistent with this, a study on hnRNP F binding and G4 formation showed that hnRNP F ribonucleoprotein formation prevents G4 formation/reformation by sequestering the G-tracts in a single-stranded RNA state. As the rate of protein-RNA complex formation is faster than that for G4 formation, the shift of cellular G4-RNAs toward an unfolded state could be partially due to interactions with RNA-binding proteins.
changes, could be used as a quantitative method in the iterative chemical design of new G4 ligands.

**Methods**

**FRET-melting experiments.** The sequences of the oligonucleotides used herein are listed in Supplementary Table 3. The preparation of these sequences is described in the Supplementary Methods. Experiments were performed in a 96-well format using a Mx3005P qPFR machine (Agilent) equipped with FAM filters (λex = 492 nm; λem = 516 nm) in 100 µL (final volume) of 10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl (F21T, F-DS-T) or plus 1 mM KCl/99 mM LiCl (F-Myc-T, F-Kit-T, F-TERRA-T) with 0.2 µM of labeled oligonucleotide and 0, 1, 5, or 10 µM of BioTASQ. Competitive experiments were carried out with labeled oligonucleotide (0.2 µM), 10 µM BioTASQ, and increasing amounts (0, 15 and 50 equiv.) of the unlabeled competitor ds26. After the first equilibration step (25 °C, 30 s), a stepwise increase of 1 °C every 30 s for 65 cycles to reach 90 °C was performed, and measurements were made after each cycle. Final data were analyzed using Excel (Microsoft Corp.) and OriginPro®9.1 (OriginLab Corp.). The emission of FAM was normalized (0 to 1), and T1/2 was defined as the temperature for which the normalized emission is 0.5; ΔT1/2 values are means of triplicates.

**In vitro pull-down assay.** The sequences of the oligonucleotides used herein are listed in Supplementary Table 3. The preparation of these sequences is described in the Supplementary Methods. The in vitro quadruplex capture experiments were performed in 250 µL final volume as follows: first, the streptavidin MagneSphere® beads (Promega) were washed three times with 10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl (Caco.K) buffer. To this end, 200 µL of the commercial solution of beads were centrifuged, taken up in 200 µL of Caco.K, and this washing step was repeated three times. After the original solutions being reconstituted (as 250-µL mixtures in Caco.K), 25 µL of MagnetSphere® beads were added to solutions comprising BioTASQ (20 µM) and FAM-labeled oligonucleotides (1 µM). The mixtures were stirred overnight at 25 °C; then, the mixtures were added to solutions comprising BioTASQ (20 µM) and FAM-labeled oligonucleotides (1 µM), and MagnetSphere® beads (25 µL) along with the unlabeled competitor ds26. After a first equilibration step (25 °C, 30 s), a stepwise increase of 1 °C every 30 s for 65 cycles to reach 90 °C was performed, and measurements were made after each cycle. Final data were analyzed using Excel (Microsoft Corp.) and OriginPro®9.1 (OriginLab Corp.). The emission of FAM was normalized (0 to 1), and T1/2 was defined as the temperature for which the normalized emission is 0.5; ΔT1/2 values are means of triplicates.

**Cell line and culture.** Human breast cancer cells MCF7 was obtained from American Type Culture Collection (ATCC). The cells were cultured in Dulbecco’s modified Eagle’s medium (Gibco) supplemented with 5% synthetic fetal bovine serum (FetalClone III; GE Life Sciences) and 100 µg penicillin–streptomycin mixture (Gibco). Cells were incubated at 37 °C in a humidified, 5% CO2 atmosphere-controlled incubator (HERAcell). Standard cell-culturing procedures were employed including aspiration and washing with phosphate-buffered saline (PBS, Gibco). Cells were trypsinized using Trypsin-EDTA (0.25%) (Gibco). Cell counting was performed using a Coulter Counter (Beckman Coulter).

**Dose-response profiling of G4 ligands.** MCF7 cells were seeded at 3000/well in a 96-well flat bottom plate. The cells were treated with a series of BRACO-19 or RHPS4 concentrations, made from serial dilutions. The cells were then monitored in the Essen Bioscience IncuCyte ZOOM live-cell monitoring system®. Phase contrast was used to measure cellular proliferation under a range of doses. The maximum and minimum achievable confluency values under these conditions were used for normalization. Three biological replicates were used to produce the dose–response curve. Doses between LD15 to LD25 were calculated from the dose–response curve profiles (Supplementary Figure 4) and used for subsequent experiments.

**G4 RNA-specific precipitation (G4RP).** MCF7 cells were seeded at 3.5 × 10[5] cells per 10-cm dish before treatment with either vehicle (PBS), BRACO 5 µg/ml (LD50), or RHPS4 1.5 µM (LD50) for 72h. Cells were then crosslinked using 1% formaldehyde/PBS for 5 min at 25 °C and the crosslinking was then quenched with formaldehyde/PBS. The chromatin was then harvested using a Dounce homogenizer and subjected to sonication (5 cycles of 15 s each; 150 W) to sequentially decrease the size of the chromatin. The chromatin fractions were then analyzed using a Qubit fluorometer (Invitrogen) and a Nanodrop spectrophotometer (Thermo Scientific).
0.125 M glycine for 5 min. Cells were scraped and resuspended in G4RP buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, RNase inhibitor (Roche), hemoglobin protease inhibitor cocktail). Cells were then sonicated using a Covaris m220 Ultrasonicator using default settings at 10% duty for 2 min. The sonicated fractions were then incubated with 100 µM BioTASQ (or 100 µM biotin for negative controls) overnight at 4°C. Five percent of the sonicate was collected as input control. Ten micrograms of streptavidin-magnetic beads (Promega) was added and the extract was incubated for 2 h at 4°C. Magnetic beads were then washed four times in G4RP buffer for 5 min. The beads were then incubated at 70 °C for 1 h to reverse crosslink. TRIZOL was then used to extract the RNA from the beads using the manufacturer’s instructions.

RT-qPCR

The primer sets used for RT-qPCR are listed in Supplementary Table 4. Extracted RNA was reverse transcribed with Superscript III (Thermos) and random hexamer primers using the manufacturer’s standard protocol to generate cDNA. cDNAs were quantified using 2× SYBR green mix (Bimake) with three technical replicates. C(t) values of pull-down samples were normalized to the input control. Three biological replicates were used for all qPCR-based quantifications. Exon-spanning primers for quantifying mRNA levels were derived from Primerbank31.

G4RP sequencing (G4RP-seq) and analysis

g4RP samples were DNAase-treated, briefly thermally fragmented and ribo-depleted using the Illumina Ribo-Zero RNA removal kit. RNA library preparation was performed using the Illumina TruSeq RNA Library Prep kit by following the manufacturer’s instructions. Two replicates (non-treated, BRACO-19 treated and RHP54) along with an input for each condition were paired end sequenced at 2 × 75 bp using NextSeq 500. The sequenced reads were mapped to the human reference genome hg19 assembly using HISAT237. Exon feature count and annotation to genes were performed using 2× SYBR green mix (Bimake) with three technical replicates. C(t) values of pull-down samples were normalized to the input control. Three biological replicates were used for all qPCR-based quantifications. Exon-spanning primers for quantifying mRNA levels were derived from Primerbank31.

Statistics

Graphs were produced by Microsoft Excel or GraphPad Prism. Statistical testing for multiple groups of dataset was performed using one-way or two-way ANOVA, and multiple comparisons corrected by the Bonferroni’s method. Statistical comparisons of the average of two groups were performed using two-tailed Student’s t-test. Linear regression goodness of fit was determined from coefficient of correlation (R²) or Pearson correlation (r), and non-zero slope significance was given as p-value <0.05. All p-values <0.05 were considered significant unless specified otherwise.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All relevant data are available on request from the authors upon reasonable request. The sequencing data have been deposited into NCBI (Accession number 26404).

References

1. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 505, 696 (2014).
2. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. Nat. Rev. Genet. 15, 423–437 (2014).
3. Leppa, K., Das, R. & Barna, M. Functional 5′-UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat. Rev. Mol. Cell Biol. 19, 158–174 (2018).
4. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Understanding the transcriptome through RNA structure. Nat. Rev. Genet. 12, 641–655 (2011).
5. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. Nat. Rev. Genet. 13, 770–780 (2012).
6. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 43, 8627–8637 (2015).
7. Cammas, A. & Millevoi, S. RNA G-quadruplexes: emerging mechanisms in disease. Nucleic Acids Res. 45, 1584–1595 (2017).
8. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-quadruplexes in biology: principles and molecular mechanisms. J. Mol. Biol. 429, 2127–2147 (2017).
9. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. G4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat. Meth. 13, 841 (2016).
10. Gun, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science 353, aa5371 (2016).
11. Bif, G., Di Antonio, M., Tannahil, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat. Chem. 6, 75–80 (2014).
12. Laguerre, A. et al. Visualization of RNA-quadruplexes in live cells. J. Am. Chem. Soc. 137, 8521–8525 (2015).
13. Endoh, T. et al. Real-time monitoring of G-quadruplex formation during transcription. Anal. Chem. 88, 1984–1986 (2016).
14. Haudecoeur, R., Stefan, L., Denat, F. & Monchaud, D. A model of smart G-quadruplex ligands via an optimised FRET melting assay. Biochimie 115, 194–202 (2015).
15. Gabelica, V., Rosa, F. & De Pauw, E. A simple method to determine electrospray response factors of noncovalent complexes. Anal. Chem. 81, 6708–6715 (2009).
16. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nat. Rev. Mol. Cell Biol. 18, 279 (2017).
17. Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nat. Rev. Drug Discov. 10, 261–275 (2011).
18. Arzalin, C. M., Reichenbach, P., Khoriati, L., Giulotto, E. & Lengner, J. Telomeric repeat–containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318, 798–801 (2007).
19. Montero, J. J., De Silanes, I. L., Grafa, O. & Blasco, M. A. Telomeric RNAs are essential to maintain telomeres. Nat. Commun. 7, 12534 (2016).
20. Haudecoeur, R., Stefan, L. & Monchaud, D. Multitasking water-soluble synthetic G-Quartets: from preferential RNA-quadruplex interaction to biocatalytic activity. Chem. Eur. J. 19, 12739–12747 (2013).
21. Mueller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated G-quadruplex isolation from human cells. Nat. Chem. 2, 1095–1098 (2010).
22. Siddiqui-Jain, A., Grant, C. L., Beurs, D. J. & Hurley, L. H. Direct evidence for 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 (2002).
23. Rodríguez, R. et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. 8, 301–310 (2012).
24. Bugaut, A. & Balasubramanian, S. 5′-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. 40, 4727–4741 (2012).
25. Karasiosit, A. I. et al. Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. Angew. Chem. Int. Ed. 50, 10645–10648 (2011).
26. Xian, C., Ishizuka, T. & Xu, Y. Antiparallel RNA G-quadruplex formed by human telomere RNA containing bromoguanosine. Sci. Rep. 7, 6695 (2017).
27. Selth, L. A., Gilbert, C. & Svejstrup, J. Q. RNA immunoprecipitation to determine RNA-protein associations in vivo. Cold Spring Harb. Protoc. 2009, pdb5234 (2009).
28. Wang, X., Scandolario, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. Nucleic Acids Res. 40, D1144–D1149 (2011).
29. Moore, M. J. B. et al. Trisubstituted acridines as G-quadruplex telomere targeting agents. Effects of extensions of the 3.6- and 9-side chains on quadruplex binding, telomerase activity, and cell proliferation. J. Med. Chem. 59, 582–599 (2006).
30. Salvati, E. et al. Telomere damage induced by the G-quadruplex ligand RHP45 is an antitumor effect. J. Clin. Invest. 117, 3236–3247 (2007).
31. Miglietta, G. et al. RNA G-quadruplexes in Kirsten ras (KRAS) oncogene as targets for small molecules inhibiting translation. J. Med. Chem. 60, 9448–9461 (2017).
32. Hoffmann, A., Freyß, L., Smithl, M. & AubelD. T. Formaldehyde crosslinking: a tool for the study of chromatin complexes. J. Biol. Chem. 290, 26404–26411 (2015).
36. Verma, A., Yadav, V. K., Basundra, R., Kumar, A. & Chowdhury, S. Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. Nucleic Acids Res. 37, 4194–4204 (2009).

37. Zhao, W. et al. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics 15, 419 (2014).

38. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 13, 462–464 (2016).

39. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).

40. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).

41. Anders, S. et al. Count-based differential expression analysis of RNA-sequencing data using DESeq2. Genome Biol. 16, 131 (2015).

42. Dhapola, P. & Chowdhury, S. QuadBase2: web server for multiplexed guanine quadruplex mining and visualization. Nucleic Acid Res. 44, W277–W283 (2016).

43. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–W97 (2016).

44. Neidle, S. The structures of quadruplex nucleic acids and their drug complexes. Curr. Opin. Struct. Biol. 19, 239–250 (2009).

45. Ma, K. X. et al. Long noncoding RNA MALAT1 associates with the malignant status and poor prognosis in glioma. Tumor Biol. 36, 3355–3359 (2015).

46. Gilbert, S. L., Pehrsen, J. R. & Sharp, P. A. XIST RNA associates with specific regions of the inactive X chromatin. J. Biol. Chem. 275, 36491–36494 (2000).

47. Cai, Y. et al. Rpph1 upregulates CDC42 expression and promotes hippocampal neuron dendritic spine formation by competing with miR-330-3p. Front. Mol. Neurosci. 10, 27 (2017).

48. Mergny, J. L. et al. Thermal difference spectra: a specific signature for nucleic acid structures. Nucleic Acids Res. 33, e138–e138 (2005).

49. Kwok, C. K., Marsico, G. & Balasubramanian, S. Detecting RNA G-quadruplexes (rG4s) in the transcriptome. Cold Spring Harb. Perspect. Biol. 10, a032284 (2018).

50. Chen, X. C. et al. Tracking the dynamic folding and unfolding of RNA G-quadruplexes in live cells. Angew. Chem. Int. Ed. https://doi.org/10.1002/anie.201801999 (2018).

51. Mendoza, O., Bourdoncle, A., Boulé, J.-B., Brosh, R. M. & Mergny, J.-L. G-quadruplexes and helicases. Nucleic Acids Res. 44, 1989–2006 (2016).

52. Brázdíl, V., Háromíková, L., Liu, J. C. & Foja, M. DNA and RNA quadruplex-binding proteins. Int. J. Mol. Sci. 15, 17493–17517 (2014).

53. Neidle, S. Quadruplex nucleic acids as targets for anticancer therapeutics. Nat. Rev. Chem. 1, 0041 (2017).

54. Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. Nat. Struct. Mol. Biol. 20, 300 (2013).

55. Samatanga, B., Dominguez, C., Jeleasrov, I. & Allain, F. H. T. The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. Nucleic Acids Res. 41, 2505–2516 (2012).

56. Decorsière, A., Cayrel, A., Vagner, S. & Millevoi, S. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3’-end processing and function during DNA damage. Genes Dev. 25, 220–225 (2011).

57. Chambers, V. S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. 33, 877–881 (2015).

58. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. Nucleic Acids Res. 36, 5482–5515 (2008).

59. Chen, G. et al. LncRNA-Disease: a database for long non-coding RNA-associated diseases. Nucleic Acids Res. 41, D983–D986 (2012).

60. Engreitz, J. M. et al. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell 159, 188–199 (2014).

61. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc. 13, 551 (2018).

62. Sharma, E., Sterne-Weiler, T., O’Hanlon, D. & Blencowe, B. J. Global mapping of human RNA-RNA interactions. Mol. Cell 62, 618–626 (2016).

63. Yang, S. Y., Amor, S., Laguerre, A., Wong, J. M. & Monchaud, D. Real-time and quantitative fluorescent live-cell imaging with quadruplex-specific red-edge probe (G4-REP). Biochem. Biophys. Acta 1861, 1312–1320 (2017).