iScience

CellPress
OPEN ACCESS

Article

Side-by-side comparison of G-quadruplex (G4) capture efficiency of the antibody BG4 versus the small-molecule ligands TASQs

Yilong Feng, Zexue He, Zhenyu Luo, Francesco Rota Sperti, Ibai E. Valverde, Wenli Zhang, David Monchaud

wzhang25@njau.edu.cn (W.Z.)
david.monchaud@cnrs.fr (D.M.)

Highlights

Sequencing-based techniques are pivotal to understand G-quadruplex-DNA biology

Prior to sequencing, G4s must be isolated by affinity capture *in vitro* or *in vivo*

We compare here G4 chemoprecipitation by TASQs and immunoprecipitation by BG4

G4DP-seq (TASQ) and BG4-DNA-IP-seq (BG4) offer complementary views on G4 landscapes

Feng et al., iScience *26*, 106846
June 16, 2023
https://doi.org/10.1016/j.isci.2023.106846

iScience

CellPress
OPEN ACCESS

Article

Side-by-side comparison of G-quadruplex (G4) capture efficiency of the antibody BG4 versus the small-molecule ligands TASQs

Yilong Feng,[1,3] Zexue He,[1,3] Zhenyu Luo,[1,3] Francesco Rota Sperti,[2] Ibai E. Valverde,[2] Wenli Zhang,[1,*] and David Monchaud[2,4,*]

SUMMARY

The search for G-quadruplex (G4)-forming sequences across the genome is motivated by their involvement in key cellular processes and their putative roles in dysregulations underlying human genetic diseases. Sequencing-based methods have been developed to assess the prevalence of DNA G4s genome wide, including G4-seq to detect G4s in purified DNA (*in vitro*) using the G4 stabilizer PDS, and G4 chromatin immunoprecipitation sequencing (G4 ChIP-seq) to detect G4s in *in situ* fixed chromatin (*in vivo*) using the G4-specific antibody BG4. We recently reported on G4-RNA precipitation and sequencing (G4RP-seq) to assess the *in vivo* prevalence of RNA G4 landscapes transcriptome wide using the small molecule BioTASQ. Here, we apply this technique for mapping DNA G4s in plants (rice) and compare the efficiency of this new technique, G4-DNA precipitation and sequencing, G4DP-seq, to that of BG4-DNA-IP-seq that we developed for mapping of DNA G4s in rice using BG4. By doing so, we compare the G4 capture ability of small-sized ligands (BioTASQ and BioCyTASQ) versus the antibody BG4.

INTRODUCTION

G-quadruplex-DNA (G4-DNA or G4) are DNA higher-order structures that originate in the folding of guanine (G)-rich DNA sequences, when freed from the duplex constraint, into a four-stranded structure.[1,2] The stability of these structures is provided by both the self-assembly of Gs to form G-quartets and the self-stacking of several contiguous G-quartets (Figure 1A).[3] Gs are thus not randomly distributed along the G4-forming sequences (G4FSs) but gathered into G-runs usually comprising 2 to 4 Gs. The repetitive nature of these sequences makes them readily detectable genome wide: the bioinformatics processing of the human genome in the search for the general sequence $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$ (where N is any intervening nucleobase, x ranging from 1 to 7) led to the identification of >300,000 putative G4FSs.[4,5] The length of the connecting loops (the x value) was then extended ≤ 15, ≤ 21 and tehn ≤ 25, which mechanically increased the number of G4FSs to >1,000,000.[6–9]

This high G4 density prompted researchers to demonstrate their existence *in vitro*. To this end, the G4 sequencing (G4-seq) technique[6] was developed and applied to purified single-stranded DNA, using G4-promoting conditions (*i.e.*, in a buffer with a high K^+ content or in the presence of the G4-stabilizing small-molecule pyridostatin,[10] PDS). The folded and stable G4s were then detected through a polymerase stop assay, in which the stabilized G4s pause the polymerase processivity, which creates an erroneous incorporation of nucleotides in the downstream sequence. The resulting, highly mismatched sequences are then easily visualizable through a Phred quality score analysis, which led to the detection of the polymerase stop sites, that is, the G4 sites. G4-seq identified >500,000 G4FSs in K^+-rich buffers and >700,000 G4FSs in presence of PDS within the human genome and was subsequently applied to 12 different species (including human, mouse, bacteria, and *Arabidopsis thaliana*, a model plant species)[11] to track and highlight the diversity of abundance and location of G4FS (being particularly prevalent in mammals).

The existence of G4s *in cella* was then scrutinized: to this end, an antibody (Ab)-based G4-immunoprecipitation protocol was developed, the G4 chromatin immunoprecipitation sequencing (G4 ChIP-seq),[12,13] using fragmented chromatin from fixed human keratinocyte cells from which folded G4s were precipitated using BG4,[14] a G4-specific single-chain Ab (Figure 1B). G4 ChIP-seq identified *ca.* 10,000 G4s (*e.g.*, 9,000,

[1]State Key Laboratory for Crop Genetics and Germplasm Enhancement and Utilization, CIC-MCP, Nanjing Agricultural University, Nanjing, P.R. China

[2]Institut de Chimie Moléculaire, ICMUB CNRS UMR 6302, Université de Bourgogne, Dijon, France

[3]These authors contributed equally

[4]Lead contact

*Correspondence: wzhang25@njau.edu.cn (W.Z.), david.monchaud@cnrs.fr (D.M.)

https://doi.org/10.1016/j.isci.2023.106846

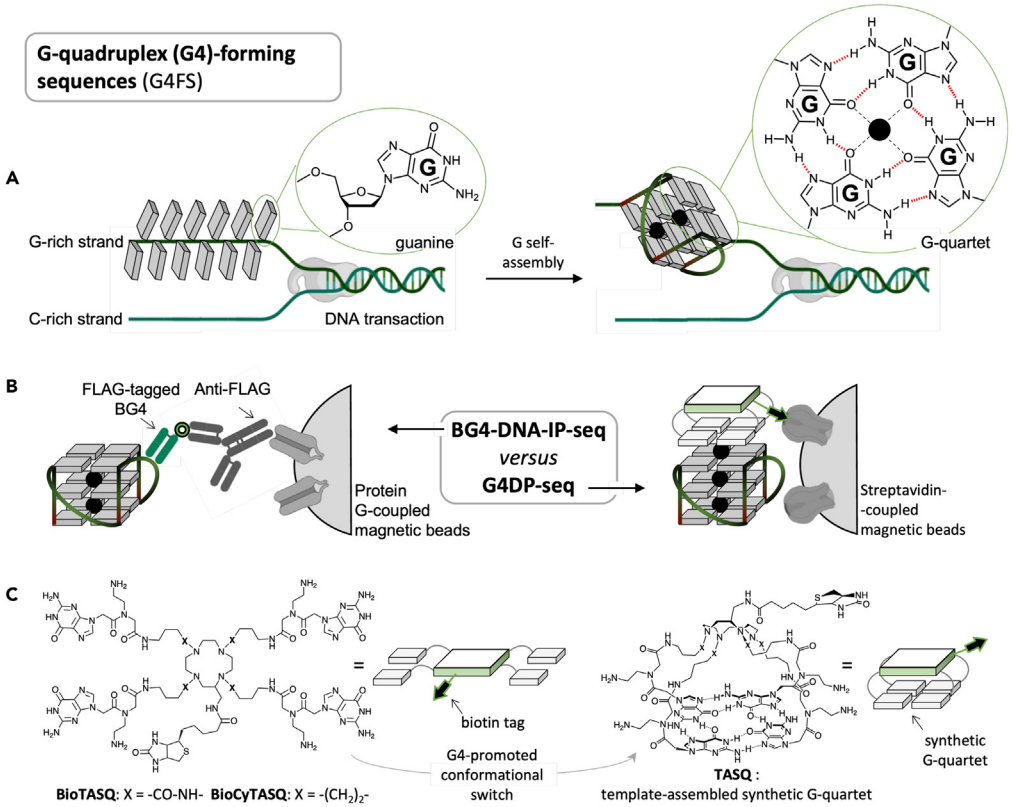

Figure 1. Guanine (G)-rich sequences fold into G-quadruplex (G4)-DNA

Schematic representations of a G-quadruplex (G4)-folding from a guanine (G)-rich DNA sequence (A) and of the two affinity capture techniques studied here (B), the BG4-DNA-IP-seq performed with the antibody BG4 and the G4DP-seq performed with two biotinylated small molecules, BioTASQ and BioCyTASQ (C). The black balls represent monovalent cations.

13,000, and 15,000 G4 ChIP-seq peaks in leukemia K562, breast cancer MCF7, and osteosarcoma U2OS cells, respectively).[12,15] This represents a minor fraction (2%) of G4FSs detected by G4-seq, likely due to the repressive role of the chromatin packing. This technique was then extended to various cancer cells and tissues, including K562 chronic myelogenous leukemia cells[16,17] and breast cancer patient-derived tumor xenograft (PDTX) models.[18] The prevalence of G4s *in vivo* was then confirmed by a series of G4 CUT&Tag (cleavage under targets and tagmentation) experiments performed in both human and mouse cells (*ca.* 9,000 G4s in both MCF7 and embryonic stem cells, ESCs) with BG4[15,19] and more recently with the endogenously expressed, G4-specific nanobody SG4 (ca. 7,000 G4s in HEK293T).[20]

We also used BG4 to assess the prevalence of G4s in the genome of plants (rice). We developed a technique termed BG4-DNA-IP-seq,[21] which was implemented with purified and fragmented rice genomic DNA in G4-promoting conditions (*i.e.*, in a buffer with a high K^+ content) from which folded G4s were precipitated using BG4. We identified *ca.* 20,000 G4FSs, which represent a minor fraction (5%) of *in silico* predicted G4FSs but far higher than the number of G4FSs detected by G4-seq in another model plant species, the *A. thaliana* (*ca.* 2,000 G4FSs), as a result of the difference in the implemented technique.

The reliability of the BG4-DNA-IP-seq protocol, along with the straightforward access to rice genomic DNA, makes this system ideal for comparing the pull-down efficiency of the Ab BG4 versus the small-molecular baits BioTASQ[22,23] and its new derivative BioCyTASQ[24] (Figure 1C). These two TASQs (template-assembled synthetic G-quartets)[25,26] are biotinylated biomimetic ligands that were successfully employed for isolating G4-RNAs from solution (pull-down protocol). BioTASQ was applied to fish G4-RNA out from human cell lysates in a protocol referred to as G4RP-seq (G4-RNA precipitation and sequencing),[22,27] which belongs to the toolbox of sequencing-based techniques used to interrogate G4-RNA prevalence,[28]

comprising rG4-seq,[29] dimethyl sulfate (DMS)-seq,[30] Keth-seq,[31] and SHALiPE-seq.[32] BioCyTASQ, which is now commercially available, is structurally simpler (and thus synthetically more accessible) than BioTASQ and was used only in G4RP-RT-qPCR experiments.[33] We reasoned that the ability of both TASQs to isolate G4-DNA deserves not only to be investigated in sequencing-based experiments but also to be compared to that of BG4 when used in a ChIP-seq-like protocol. We thus report here on these investigations, which provide the very first comparison of the efficiency of the immuno- versus chemoprecipitation of G4s, under strictly identical experimental conditions.

RESULTS

BG4-IP-seq and G4DP-seq protocols

The BG4-DNA-IP-seq protocol[21] was implemented with purified rice genomic DNA, after fragmentation (sonication) into 100- to 500-bp sequences in size. These DNA fragments were then thermally treated (10 min at 90°C, followed by a slow return to 25°C in G4-stabilizing conditions, 150 mM K^+) in order to maximize G4 folding. BG4-DNA-IP-seq relies on the use of an FLAG-tagged BG4 (expressed from the pSANG10-3F-BG4 plasmid, 3 μg for 5 μg of DNA), followed by the addition of an anti-BG4 Ab (anti-FLAG, 3 μg) and then protein G-coupled magnetic beads for pulling down the G4/BG4/anti-FLAG/beads complexes. This protocol was then adapted to the TASQ molecular tools: in reference to their use as baits for G4-RNA in G4RP-seq,[22,27] we refer to this protocol as G4DP-seq (G4-DNA precipitation and sequencing). G4DP-seq thus relies on the treatment of the prepared DNA fragments (G4-optimized) with TASQ (either BioTASQ or BioCyTASQ, 100 μM), followed by the addition of streptavidin-coated magnetic beads for pulling the G4/TASQ/bead complexes down. In both instances (BG4-DNA-IP-seq and G4DP-seq), thermal elution steps (65°C twice, 15 min each) followed by DNA purification (PhOH/$CHCl_3$ extraction and then EtOH precipitation) provide samples ready for library preparation. The binding specificity of BioTASQ or BioCyTASQ to G4s was validated by dot blotting experiments (Figure S1).

BG4-DNA-IP-seq and G4DP-seq results

The reliability of both BG4-DNA-IP-seq and G4DP-seq protocols was assessed by sequencing two independent biological libraries for each condition. Raw FASTQ reads were aligned to the MSU v7.0 rice genome reference, and only reads with a high mapping quality (mapq score >10) were kept for further analyses (Tables 1 and S1). Only unique reads were directly used for G4 peak calling using MACS2[34] without being relative to any control. The use of BG4 led to a high and consistent number of G4 peaks (77,986 for Rep. 1 and 73,476 for Rep. 2), with 52,088 common peaks (Figure 2A, **left**; Table 1); BioTASQ to a variable number of G4 peaks (63,202 for Rep. 1 and 101,765 for Rep. 2), with 59,271 common peaks; and BioCyTASQ to a low number of G4 peaks (47,551 for Rep. 1 and 23,943 for Rep. 2), with 17,747 common peaks. Of note, some of these G4s were validated through qPCR analyses (Figure S2, Tables S2 and S8). The ratio of common peaks (*i.e.*, 71, 94, and 74%) testifies to the good reproducibility of the BG4-DNA-IP-seq and G4DP-seq protocols, which was confirmed by a heatmap of Spearman correlation values of reads between biological replicates (Figure S3). To further demonstrate the reproducibility of G4DP-seq, we randomly selected the same number of reads (*ca.* 24 million reads for BioTASQ and 12 million reads for BioCyTASQ, Table S3) for G4 peak calling: the peaks found for each replicate were then compared, and we found a far better correlation between replicates (96 and 86% for BioTASQ and BioCyTASQ, respectively, Figure S4), which further supports G4DP-seq reproducibility.

Next, all unique reads from biological replicates were merged for G4 peak calling using MACS2, which was performed against three controls for BG4 (input, immunoglobulin G [IgG], and anti-FLAG alone) and two controls for TASQ (input and biotin) (Tables 1 and S4). As seen in Figure 2B, 29,469, 77,096, and 49,898 high-confidence G4 peaks (*i.e.*, compared to all controls) were identified for BG4, BioTASQ, and BioCyTASQ, respectively. This indicates that both Ab and ligands capture G4s efficiently. The genomic distributions of these peaks are globally comparable (Figure 2C), the most abundantly detected G4s belonging to promoters, distal intergenic regions, and exons, although with some variations in the distribution (28, 28, and 16% for BG4; 22, 21, and 34% for BioTASQ; and 26, 22, and 29% for BioCyTASQ). To better highlight the differences between them, a pairwise comparison (Figure 2D) shows that the 9,917 common G4s (detected by both Ab and TASQs, *vide infra*) are primarily located in promoters (33%) and distal intergenic regions (32%), while those enriched specifically by each bait are exonic G4s with TASQs (41 and 42% for BioTASQ and BioCyTASQ, respectively) and promoter G4s with BG4 (32%). In brief, TASQ-specific G4s were distributed more in exons but less in promoters and introns as compared to BG4-specific ones (Figure 2D). To discard any bias regarding a possible influence of the sequencing depth on genomic

Table 1. Reads, peaks, and G4-enrichment for both *in vitro* and *in vivo* sequence-based experiments

Molecular tools	Repli-cate	Clean reads	Unique reads	Peaks (raw)	Common peaks	Merged peaks vs. input	Merged peaks vs. IgG/biotin	Merged peaks vs. anti-Flag	Common merged peaks
in vitro BG4-DNA-IP-seq and G4DP-seq experiments									
BG4	#1	55.5M	21.1M	78.0k	52,088	79,501	72,267	29,634	29,469
	#2	80.7M	37.1M	73.5k					
BioTASQ	#1	42.3M	23.0M	63.2k	59,271	95,581	80,389	–	77,096
	#2	80.8M	42.6M	101.8k					
BioCyTASQ	#1	43.4M	25.0M	47.6k	17,747	54,415	51,199	–	49,898
	#2	23.2M	11.9M	23.9k					
Molecular tools	Repli-cate	Clean reads	Unique reads	Peaks (raw)	Common peaks	Merged peaks vs. input	Merged peaks vs. biotin	Merged peaks vs. anti-Flag	Common merged peaks
in vivo G4DP-seq experiments									
BioTASQ	#1	51.4M	36.2M	94.0k	77,502	58,037	81,699	–	57,500
	#2	54.3M	39.5M	85.7k					

G4-enrichment					Specific G4 peaks		
Molecular tools	Common merged peaks	Specific peaks	PG4FS	G4 density		BioTASQ vs. BioCyTASQ	BioTASQ *in vitro* vs. *in vivo*
BG4	29,469	10,268	25,317	1,79	Common	73,283 (85%)	57,606 (74%)
BioTASQ	77,096	30,578	62,236	2,63	Former	11,458 (13%)	7,447 (9%)
BioCyTASQ	49,898	12,833	83,228	2,68	Latter	1,797 (2%)	12,487 (16%)
common	–	9,917	14,388	2,12			

distribution, we conducted G4 peak calling using the same sequencing depth: as seen in Figure S5, the genomic distributions obtained were similar, indicating that the variation in promoter/exon distributions between TASQs and BG4 likely originates in a difference of sequence (that is, of G4 structure), which could drive epitope recognition by BG4 and accessibility to the external G-quartet (nature, length, and distribution of loops) for TASQs.

This difference in G4 recognition was confirmed by a closer examination of exonic G4s (Figure S6): we found 25,886, 14,625, and 4,629 G4 peaks identified by BioTASQ, BioCyTASQ, and BG4, respectively. We then analyzed the putative G4FS (PG4FS) content of these peaks: we bioinformatically searched for PG4FS of general $G_{\geq 2}N_xG_{\geq 2}N_xG_{\geq 2}N_xG_{\geq 2}$ sequence (x between 1 and 12, or $(G_{2+}N_{1\text{-}12})$) using QuadParser[5] with the published regular expression syntax '([gG]{2,}\w{1,12}){3,}[gG]{2,}'. We found that >97% of them correspond to 2-quartet (and more) PG4FSs (whole-genome number: 1,797,039), without notable differences between TASQs and BG4. We then lengthened the G tracts from 2 to 3 (general sequence: $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$ (or $(G_{3+}N_{1\text{-}12})$), using the syntax '([gG]{3,}\w{1,12}){3,}[gG]{3,}') and found that 17, 25, and 47% correspond to 3-quartet (and more) PG4FSs (whole-genome number: 85,067) for BioTASQ, BioCyTASQ, and BG4, respectively. These results highlight the better efficiency of BG4 to isolate 3-quartet G4s as compared to TASQs.

To further investigate whether the two TASQs recognize different G4s, we plotted normalized read counts obtained by G4DP-seq performed with either BioTASQ or BioCyTASQ across ±2.0 kb from transcription start sites (TSSs) to transcription terminate sites (TTSs) of genes (Figure S7): the read intensity was higher with BioTASQ directly downstream of the TSSs as compared to BioCyTASQ, indicating that both TASQs do not recognize exactly the same G4s, despite a high structural similarity. We also assessed the accuracy of G4 peaks identified with G4DP-seq performed with either BioTASQ or BioCyTASQ focusing on 10 previously validated G4 *loci* (Figure S8):[21] we found that 6 and 7 of them were detected by G4DP-seq performed with BioTASQ and BioCyTASQ, respectively, which lends credence to the reliability of the G4DP protocol. Finally, to go a step further, we bioinformatically searched for PG4FS of $(G_{2+}N_{1\text{-}12})$ sequence in the G4 peaks enriched with either BioTASQ or BioCyTASQ (Figure S9): we found that G4 peaks identified with BioCyTASQ were more enriched in PG4FS motifs than BioTASQ, with 325,374 versus 277,889 PG4FS for

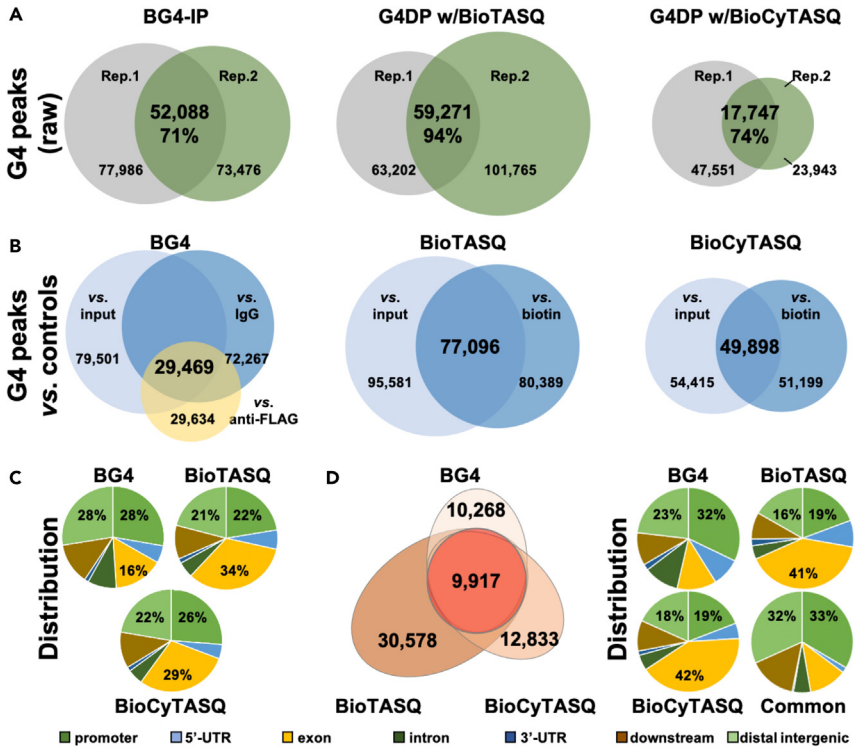

Figure 2. Comparison of the detection of G4-forming sequences by BG4 and TASQs *in vitro*

(A) Venn plots illustrating the reproducibility of G4 peaks identified using BG4-DNA-IP-seq (performed with BG4) and G4DP-seq (performed with either BioTASQ or BioCyTASQ).

(B) G4 peak calling performed against controls (input, IgG and anti-FLAG for BG4-DNA-IP-seq; input and biotin for G4DP-seq) that leads to the identification of high-confidence G4 peaks.

(C) Genomic distributions of the high-confidence G4 peaks.

(D) Pairwise comparisons (left) of G4 peaks and genomic distributions of the 10,268 BG4-specific peaks, 30,578 BioTASQ-specific peaks, 12,833 BioCyTASQ-specific peaks, and 9,917 common peaks, along with the genomic distributions (right) of the BG4-, BioTASQ-, and BioCyTASQ-specific peaks and of the common peaks.

BioCyTASQ and BioTASQ, respectively, which represent 99% of BioCyTASQ G4 peaks having at least one PG4FS versus 98% of BioTASQ G4 peaks. Lengthening the G tracts from 2 to 3 ($G_{3+}N_{1-12}$) decreased these proportions to 53% and 31% for BioCyTASQ and BioTASQ, respectively.

Comparisons of BG4-DNA-IP-seq and G4DP-seq G4 peaks

We then focused on the common and specific G4 peaks identified by BG4-DNA-IP-seq and G4DP-seq seen in Figure 2D. We calculated the GC content and found that TASQ G4 peaks displayed a higher GC content than BG4 G4 peaks (Figure S10A). We then conducted *de novo* motif identification using multiple expectation-maximizations for motif elicitation (MEME) analyses[35] and detected motifs that correspond mostly to two-quartet G4s (Figure 3A) for the 9,917 common peaks, 30,578 BioTASQ-specific peaks, 12,833 BioCyTASQ-specific peaks, and 10,268 BG4-specific peaks, without any bias for particular G4 structural types (Figure S10B).

This observation lends credence to the interchangeable use of either capture tools. However, some discrepancies could be noted: for instance, the CGGCGG motifs, known to be bound by the AP2/EREBP transcription factor (TF),[36] were more enriched in TASQ peaks than in BG4 peaks, indicating some subtle difference between molecular tools used to capture G4s; also, the peaks uniquely enriched by either of the TASQs exhibited notable differences, being enriched in atypical motifs (*e.g.*, GGTGAA and CGTGGC) with BioCyTASQ and in more classical motifs (*e.g.*, GGAGGA or GGCGGC) with BioTASQ (Figure S11). A QuadParser analysis ($G_{2+}N_{1-12}$) of these common/specific peaks revealed a better efficiency of both TASQs (with 62,236 and 83,228 PG4FS identified by BioTASQ and BioCyTASQ, respectively) as compared

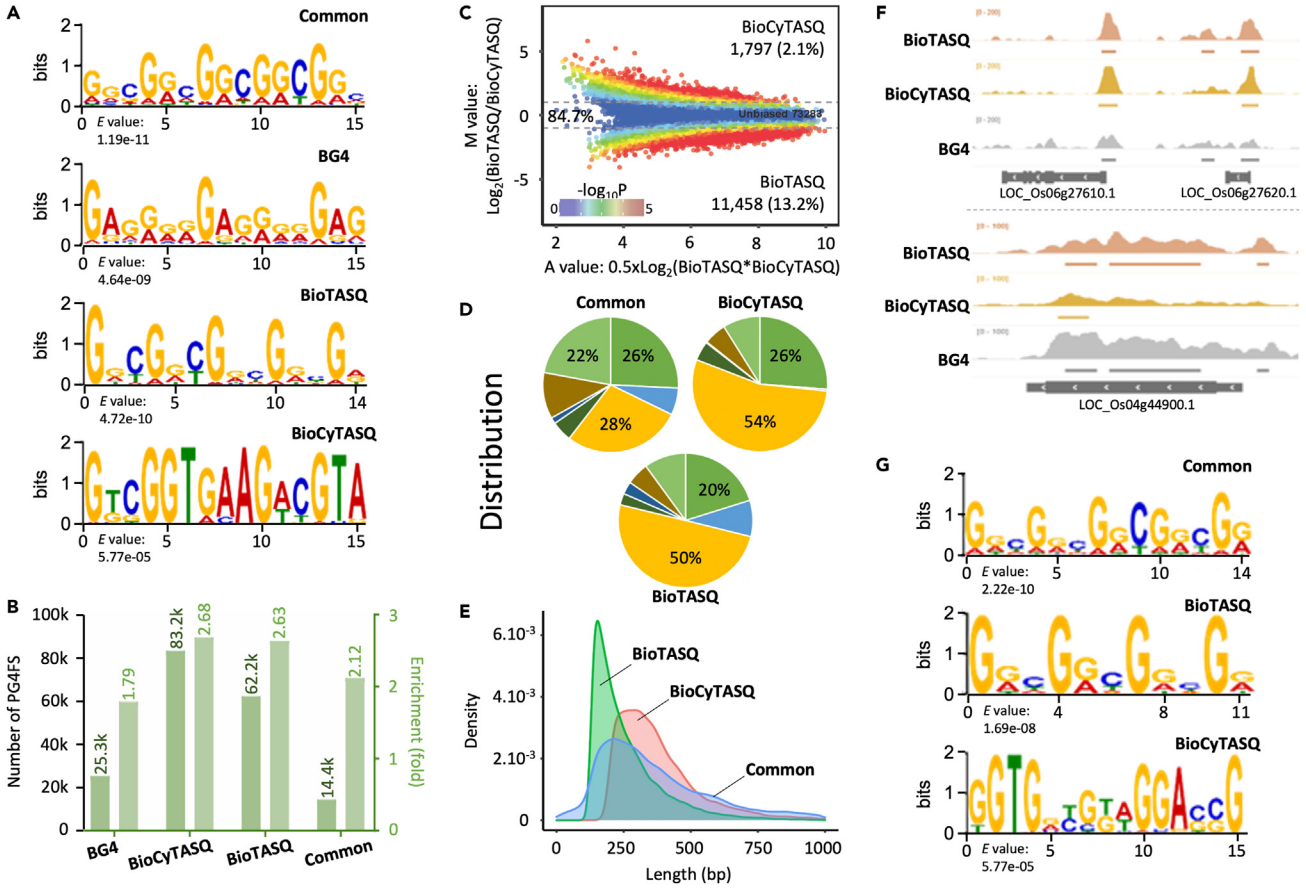

Figure 3. G4DP-seq using BioTASQ or BioCyTASQ *in vitro*

(A) Motif discovery using MEME for BG4-DNA-IP-seq and G4DP-seq peak datasets, performed with the 10,268 BG4-specific peaks, 30,578 BioTASQ-specific peaks, 12,833 BioCyTASQ-specific peaks, and the 9,917 common peaks.

(B) Number and fold enrichment of putative G4-forming sequences (PG4FSs) identified by QuadParser for both BG4-DNA-IP-seq and G4DP-seq G4 peaks.

(C) MA plot illustrating the significant overlap (>80%) between G4 peaks detected using G4DP-seq with both BioTASQ and BioCyTASQ.

(D) Genomic distribution of the specific/common G4 peaks from panel C.

(E) Length distribution of G4 peaks detected using G4DP-seq with either BioTASQ (green) or BioCyTASQ (pink) and of common peaks (blue).

(F) Example of integrative genomics viewer (IGV) screenshot of the rice (*Oryza sativa*) chromosomes 4 and 6 for BG4-DNA-IP-seq (gray) and G4DP-seq performed with BioTASQ (pink) and BioCyTASQ (yellow); G4 peaks are indicated with colored rectangles.

(G) MEME analyses for G4DP-seq performed with the 11,458 BioTASQ-specific peaks, 1,797 BioCyTASQ-specific peaks, and the 77,283 common peaks from panel C.

to BG4 (25,317 PG4FS) (Figure 3B, Table 1), which might originate in the G4-stabilizing properties of TASQs (they do not, however, promote G4 folding,[24] as further discussed below). These elevated numbers originate in both the wide range search ($G_{2+}N_{1-12}$) and the average length of the G4 peaks (0–1,000 bp-long, *vide infra*); several PG4FSs could thus be found on the same G4 peak. To further investigate this, the G4 density in each G4 peak was calculated relative to *ad hoc* controls (random sequences of similar size, n = 100) and confirmed the better performances of TASQs (enrichment score [ES] = 2.63 and 2.68 for BioTASQ and BioCyTASQ, respectively) as compared to BG4 (ES = 1.79) (Figure 3B, Table 1). Of note, a similar analysis performed with the 9,917 common G4 peaks confirms their high G4 content (14,388 PG4FS, ES = 2.12).

In order to validate that these sequences do actually fold into G4 structures, we performed both dot blotting and circular dichroim (CD) experiments. These investigations were performed with 16 sequences (Table S5) identified in both *in vivo* and *in vitro* conditions (further discussed below) by both TASQs and BG4, using the G4P protein, known to bind to G4 structures with a high selectivity.[37] We used the artificial G4P protein as an orthogonal technique as these G4s were isolated by both TASQs and BG4. Results seen

in Figure S12A indicated that 14 out of 16 sequences positively responded to blotting. We then recorded the CD spectrum of 1 invalidated sequences along with 4 validated ones: all of them displayed typical G4 CD spectra, with minima at *ca.* 240 nm and maxima at *ca.* 260 nm, which correspond to parallel G4 signatures. This series of results, which highlight that the CD technique is more sensitive than the dot blot assay, confirm that the G4 sites detected using TASQs and BG4, in both *in vivo* and *in vitro* conditions, do actually fold into stable G4 structures.

BioTASQ- versus BioCyTASQ-based G4DP-seq

It was thus of interest to further compare the properties of the two TASQs. As indicated above, the G4 peaks enriched by both TASQs contain a high number of PG4FS, with a very low rate of false positives (<2%, which may contain irregular PG4FSs). The overlap between detected G4 peaks is illustrated by a Bland-Altman plot[38] (or MA plot: M refers to *minus* and corresponds to the log2 ratios on the y axis; A refers to *average* and corresponds to the mean values on the x axis). This visualization (Figure 3C, Table 1) compares the agreement between the two G4DP-seq datasets (BioTASQ versus BioCyTASQ) and reveals a quite homogeneous distribution of the peaks with 85% (n = 73,283) common peaks ($-1 < M < 1$; $-\log_{10}(P) < 2$), along with some BioTASQ-specific peaks (n = 11,458, 13%) and BioCyTASQ-specific peaks (n = 1,797, 2%).

A closer look at the genomic distributions (Figure 3D) revealed some differences as they showed that TASQ-specific peaks are majorly found in exons (54 and 50% for BioCyTASQ and BioTASQ, respectively) and promoters (26 and 20%), while common peaks are equally distributed in promoters, exons, and distal intergenic regions (26, 28, and 22%, respectively). Also, the distribution of length of the G4 peaks is wider for BioCyTASQ (average length: 367 bp) than for BioTASQ (average length: 236 bp), which underlines a significant difference in capture efficiency (Figure 3E). We thus performed a visual inspection of the integrative genomics viewer (IGV) screenshot for two regions (Chr. 4 and 6) of the *Oryza sativa* genome (Figure 3F): this inspection revealed a similar pattern of G4 peaks captured by TASQs and BG4, with a slightly better signal-to-noise ratio for BioCyTASQ as compared to BioTASQ. These differences are interesting as they highlight that subtle structural differences (here, only the nature of the guanine arms differs: X = -CONH- versus $-(CH_2)_2-$ for BioTASQ and BioCyTASQ, respectively, Figure 1) lead to different G4 recognition. This was already exemplified when comparing BioCyTASQ and its clicked analogue BioTriazoTASQ, the former interacting more readily with DNA and the latter with RNA *in vitro*.[39] These effects are not easily rationalized but might originate in a combination of stability and accessibility of the intramolecularly folded synthetic G-quartet, indirectly governed by the flexibility of the guanine arms.

Altogether, these results highlight that BioCyTASQ should be preferred for the capture of DNA G4s *in vitro*. To confirm this, we used plotEnrichment (deepTools) to quantify the percentage of reads in the conserved G4FS (common to all methods) and BG4-/TASQ-biased G4FSs: the results seen in Figure S13 confirmed that BioCyTASQ captures G4s more efficiently. Finally, to further compare the properties of the two TASQs, we conducted an MEME analysis with the 77,283 common peaks, 1,797 BioCyTASQ-enriched peaks, and 11,458 BioTASQ-enriched peaks (Figure 3G): we observed classical G4 motifs along with motifs known to bind to TFs, including AP2/EREBP (common and BioTASQ peaks) and HMG (BioCyTASQ peaks).[36]

In vitro versus *in vivo* G4DP-seq

These results led us to investigate whether TASQs could be suited for the capture of G4s *in vivo*. The G4DP-seq protocol was thus adapted using BioTASQ as bait since it was used for the G4RP-seq, developed to trap G4-RNA in *in vivo*-like conditions. Rice seedlings were cross-linked using formaldehyde (1%) before being ground in liquid nitrogen to disrupt cell walls and isolate nuclei (several washing steps), which were subsequently lysed for isolating genomic DNA. Cross-linked chromatin was thus fragmented (100- to 500-bp), which was followed by BioTASQ incubation (4°C, overnight) and addition of streptavidin-coated beads for pulling down the cross-linked G4/TASQ/bead complexes. After thermal elution and cross-link reversal steps, DNA was precipitated for preparation of the sequencing libraries as described above.

Two biological replicates were sequenced in each conditions (*in vitro* versus *in vivo*, Tables 1, S6, and S7), and a heatmap of Spearman correlation values confirms the high correlation between them (Figure S14). Similar to what was done with BioTASQ in *in vitro* conditions (n = 77,096, Figure 2B), clean reads were

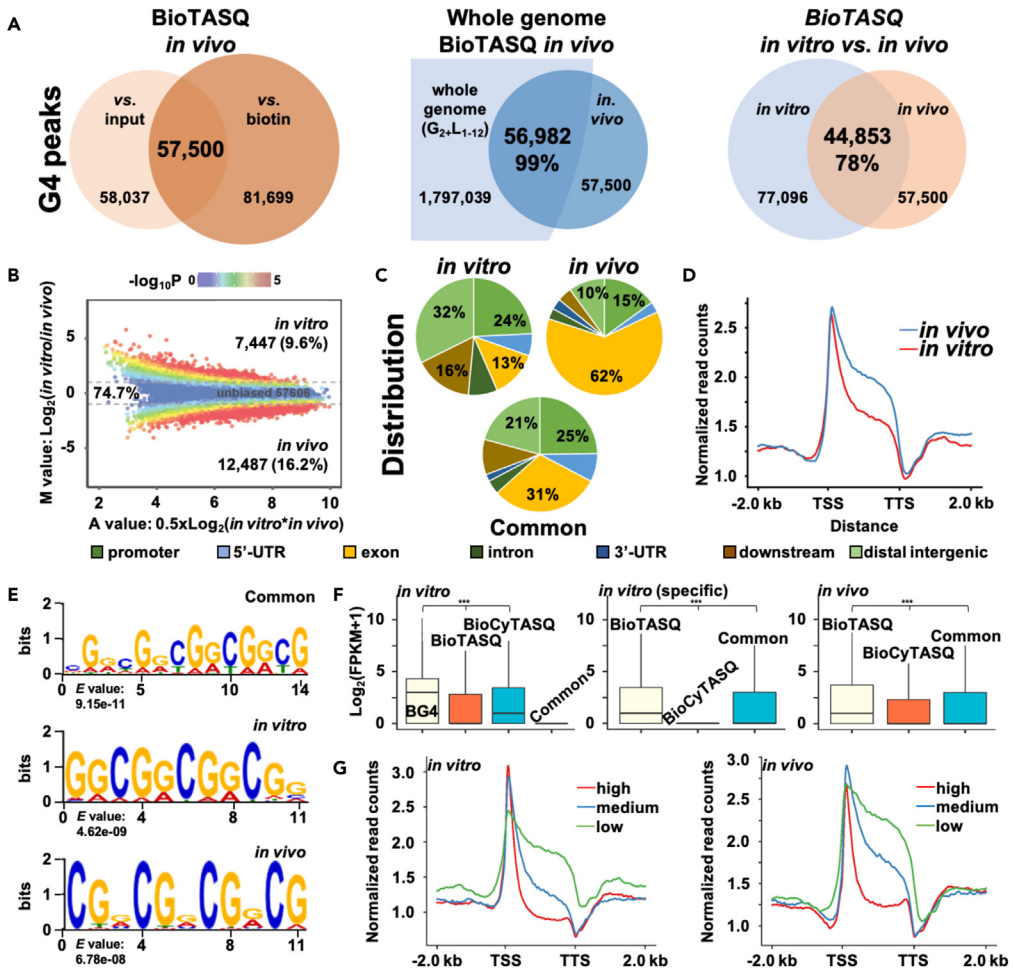

Figure 4. G4DP-seq using BioTASQ *in vitro* or *in vivo*

(A) Venn plots illustrating the G4 peak calling against two controls (input and biotin) for G4DP-seq performed *in vivo* (left panel), the number of PG4FSs identified by QuadParser for G4DP-seq performed *in vivo* compared to genome-wide PG4FSs (center), and the comparison between *in vitro* and *in vivo* G4DP-seq datasets (right panel).

(B) MA plot illustrating the significant overlap (>70%) between G4 peak datasets of G4DP-seq performed *in vitro* or *in vivo*.

(C) Genomic distribution of the specific/common G4 peaks from panel B.

(D) Profiling of *in vivo* and *in vitro* G4DP read counts across ±2 kb from the transcription start sites (TSSs) to the transcription terminate sites (TTSs) of genes.

(E) MEME analysis for G4DP-seq performed *in vitro* (32,243 peaks) and *in vivo* (12,647 peaks), along with the 44,853 common peaks from panel A.

(F) Expression levels of genes associated with common and specific G4 peaks seen in Figure 2D (left panel), Figure 3C (center), and Figure 4B (right panel); *** $p < 0.001$, ** $p < 0.01$ and * $p < 0.05$ (Wilcoxon test).

(G) Profiling of *in vitro* and *in vivo* G4DP-seq read counts across ±2 kb from the TSSs to the TTSs of genes with different expression levels (high, middle, and low, FPKM values).

used for *in vivo* G4 peak calling using MACS2 against two controls (input and biotin). This led to 57,500 high-confidence G4 peaks (Figure 4A, Table 1), with 99% (n = 56,982) with at least one ($G_{2+}N_{1-12}$) PG4FS and 35% (n = 20,077) with at least one ($G_{3+}N_{1-12}$) PG4FS. This high content represents only a minor fraction (*ca.* 3%) of the ($G_{2+}N_{1-12}$) PG4FSs detected genome wide (*ca.* 1,800,000 putative G4 motifs). The difference (*ca.* 25%) between the numbers of G4 peaks identified with BioTASQ *in vivo* and *in vitro* (57,500 and 77,096, respectively) indicated that the *in vitro* conditions are more permissive as to G4 folding than *in vivo* conditions (Figure 4B), which can be related to the ability of TASQ to promote G4-folding, while it has not been detected experimentally. We found that 78% of G4 peaks (n = 44,853) were common between *in vitro* and *in vivo* conditions (Figure 4A), indicative of a strong consistency for G4 identification in both conditions. We then conducted an MA analysis and identified 9% (n = 7,447) of *in vitro*-specific peaks, 16% (n = 12,487) of

in vivo-specific peaks, and 74% (n = 57,606) of common G4s (Figure 4B, Table 1), further suggesting the agreement between the two G4DP-seq datasets. Of note, some of these G4s were validated through qPCR analyses (Figure S15 and Table S8). Quite satisfyingly, all G4s validated *in vitro* (Figure S2) and *in vivo* (Figure S15) were more enriched relative to their controls. However, we noticed that some of the *in vivo*-detected G4s were more enriched than the *in vitro*-detected ones, which could be ascribed to either a higher G4 folding potential or more stable G4 folds *in vivo*. These hypotheses, even if appealing, remain speculative without further G4 genomic information. The integrative genomics viewer (IGV) illustrates representative genomic loci in Chr.1 for *in vitro*/*in vivo* and common G4s (Figure S16).

A significant difference was found when analyzing the genomic distributions of these peaks: while the distribution for *in vitro* G4 peaks was in line with those previously obtained (with notable enrichments in promoters and distal intergenic regions, 24 and 32%, respectively, Figure 4C), G4s identified by *in vivo* G4DP-seq were particularly enriched in exons (62%). This was further demonstrated by profiling the normalized G4 read counts across genomic regions spanning from 2 kb before the TSS to 2 kb after the TTS of the 55,801 Nipponbare genes, thus encompassing the 7 genomic regions of interest used above (promoters, 5′- and 3′-UTRs, exons and introns, downstream, and distal intergenic regions). The strong enrichment of read counts between TSS and TTS for *in vivo* conditions (Figure 4D) is likely due to an over-abundance of exonic G4s. This finding, which strongly suggests key roles of G4s in exons in a functional cellular context, is in line with their recent involvement in alternative splicing events at the transcripts level.[40,41]

Finally, the G4 motif occurrence within the 32,243 peaks detected *in vitro*, 12,647 peaks detected *in vivo*, and the 44,853 common peaks (from Figure 4A) was investigated via MEME analyses: as seen in Figure 4E, the most enriched motifs were again found to correspond to two-quartet G4s; the particular G-richness observed for the common peaks is interesting as these G4s were trapped and enriched by two orthogonal techniques (*in vitro* versus *in vivo* G4DP-seq), which is, again, quite unique and leaves no doubt about their G4 nature. Again, the motifs known to be bound by TFs binding were detected (*e.g.*, GGCGGC for AP2/EREBP). Also, we calculated the GC content, GC skew, and AT skew of these peaks (Figure S17) and found that *in vitro*-biased G4s exhibited a slightly more conspicuous GC skews at the center of the peaks as compared to *in vivo*-biased G4s.

Functional relevance of G4s detected *in vitro* and *in vivo*

It was finally of interest to correlate identified *in vitro* and *in vivo* G4s with the expression of related genes. We started with the comparison of the expression level of genes associated with *in vitro* G4s detected using BG4 (n = 10,268), BioTASQ (n = 30,578), and BioCyTASQ (n = 12,833) and the common peaks (n = 9,917) displayed in Figure 2D. We found that the mean expression level of genes related to G4s detected by BG4 was higher than those detected by BioTASQ and BioCyTASQ (with BioTASQ > BioCyTASQ), while that related to common G4s was not expressed (Figure 4F, left panel). Next, we compared the expression levels of genes associated with *in vitro* G4s detected by both TASQs (n = 73,283), solely by BioTASQ (n = 11,458), or BioCyTASQ (n = 1,797) displayed in Figure 3C. Again, the expression levels of genes related to BioTASQ G4s were higher than those of BioCyTASQ G4s, found to be lower than those of common G4s (Figure 4F, center). Finally, we compared the expression levels of genes associated with G4s detected by BioTASQ, either solely *in vitro* (n = 7,447) or solely *in vivo* (n = 12,487) or in both conditions (n = 57,606) displayed in Figure 4B. The expression levels of genes related to *in vitro*-detected G4s are higher than those of *in vivo*-detected G4s, found to be lower than those of common G4s (Figure 4F, right panel). These results thus highlighted that both molecules have distinct binding capacity for G4 *loci in vitro* or *in vivo*; the choice of molecular tools used for fishing genomic G4s might thus have a substantial consequence in terms of G4 cellular functions, which might be taken into account when interpreting the data of BG4-DNA-IP-seq and/or G4DP-seq data. This was confirmed by a gene ontology (GO) term enrichment analysis (Figures S18–S20), which indicated that the different molecular tools provided distinct GO term functions, with a higher impact of BioCyTASQ, found to enrich G4s belonging to genes associated DNA and RNA binding along with a series of metabolic and cellular processes (in contrast to BioTASQ that leads to modest effects only). To go a step further, we plotted the normalized read counts of G4DP-seq using BG4 or BioTASQ *in vitro* and *in vivo* across $\pm$ 2 kb from the TSS to the TTS of genes with different expression levels (high, middle, and low, FPKM values) (Figures 4G and S21): for *in vitro*-detected G4s, those at the TSSs exhibited a positive correlation with expression while those within gene body and at the TTS negatively impacted expression levels. This brightly highlights the dual roles of G4s in the regulation of gene expression, which is strongly dependent on their location. In contrast, for *in vivo*-detected G4s, those at

the TSS did not exhibit a direct association with gene expression levels, while those in gene body exhibited a negative correlation with expression levels. This observation was surprising as it is now firmly established that G4s are positively correlated with gene transcription.[17] This indicates that the use of TASQs for detecting G4s *in vivo* must be subjected to caution as they are certainly not responsive enough in these conditions. To further investigated this, we examined the association between G4s detected *in vivo* by BioTASQ and histone marks (Figures S22): the BioTASQ-enrichment score was found to be high with highly expressed genes and active marks (H3K4me3) and low with lowly expressed genes and repressive marks (H3K27me3). These results highlighted the need to exercise caution when using *in vivo* data to draw reliable mechanistic conclusions. Of note, we compared the profiles obtained with both TASQ and BG4 *in vitro* and found similar patterns (Figure S21). This result indicates again that the roles that detected G4s may play on gene expression are more dependent on the technique used to detect them (*in vitro* versus *in vivo*) than the molecular tools used to detect them (TASQs versus BG4).

DISCUSSION

Chemical biology and chemical genetics rely on the use of molecular tools that uniquely allow for interrogating biological systems at the cellular level.[42,43] Scientists have access to a large and diverse portfolio of tools, enabling them to select the most suited molecular devices for their intended applications. Quite often, Abs rank high in this selection, owing to their established high affinity and selectivity for their targets. This explains why the Ab BG4 is increasingly used to interrogate G4 biology at the cell/tissue level.[12–14,21,44–53] Despite the fact that, by definition, small molecules benefit from more straightforward chemical access, purification, and characterization, they still suffer from a globally accepted lower target affinity and specificity when compared to Abs. However, with a few notable exceptions, including our recent investigations in which we compared chemo-detection (N-TASQ) versus immuno-detection (BG4) of G4s in ALT^+ versus $TERT^+$ cancer cells,[50] or in *in vitro* single-molecule mechanical unfolding experiments,[54] no direct comparison of the performances of BG4 versus multivalent G4 ligands has been attempted in more complicated experimental setups, particularly the chemo- versus immuno-precipitation of G4s prior to their identification by sequencing. Our study aims at providing such a comparison.

To this end, the performances of BG4 were compared to that of two small molecules, BioTASQ and BioCyTASQ, under similar experimental conditions. These performances were compared in terms of mapping quality (clean reads) and G4 peaks calling along with both the genomic distribution and the G-rich nature (G4 motifs) of the identified G4s. Three types of comparison were undertaken:

1) The Ab and TASQs were used to fish G4s out from cell lysates in *in vitro* conditions, where the G4FSs present in the fragmented DNA are properly folded prior to being precipitated by affinity capture. This comparison did not aim at demonstrating the superiority of one tool over the other but at providing a fair comparison of their properties at a genome-wide scale. This side-by-side mapping highlights that chemicals and biologicals cannot be used interchangeably as their G4 recognition differs by the nature of the isolated G4s (in terms of both genomic location and G4-forming sequences). However, both tools can be used interchangeably if the intended investigations aim at comparing G4 landscapes between species or upon treatments, as both BG4-DNA-IP-seq and G4DP-seq offer an accurate and global overview of G4 signatures.
2) We then compared the performances of the two TASQs, BioTASQ and BioCyTASQ, under identical experimental conditions: the majority (85%) of G4s were commonly identified by both TASQs, and their genomic distribution was coherent (mostly exonic G4s); the only noticeable difference was a better signal-to-noise ratio obtained with BioCyTASQ, implying that this TASQ should be privileged for every new investigations.
3) The G4s fished out from cell lysates using BioTASQ in both *in vitro* and *in vivo*-like conditions, where naturally folded G4s are fixed (cross-linked) in live-cell conditions prior to cell lysis, chromatin fragmentation, and affinity capture steps, were compared: these investigations confirmed the suppressive role of heterochromatin for G4 formation in *in vivo* conditions as a 25% difference in the number of G4 peaks was found. However, a majority (74%) of G4s were commonly identified in both conditions, indicating the suitability of the G4DP-seq method to both approaches, and the bias toward a strong G4 enrichment in exons was confirmed to be prevalent *in vivo*.

Globally speaking, both BG4-DNA-IP-seq and G4DP-seq provide an accurate portrayal of G4 landscape and biology. The relevance of the use of the two small molecules implemented here, BioTASQ and BioCyTASQ, as reliable surrogates for BG4 is supported by side-by-side comparisons, in different experimental conditions. In light of the recent advances in the field of G4 genetics,[15,19,20] the strategic interest of G4 mapping *in vitro* could be questionable; we demonstrated here the scope of application of TASQ-based approaches as these probes can be implemented in both *in vitro* and *in vivo* conditions. These results open also new perspectives for the use of these biotinylated TASQs in Chem-map protocols,[55] which could lead to alternative information about G4 biology, on the basis of live-cell incubation with TASQs. These investigations, which are beyond the scope of the present study, could nevertheless be considered now that the performances and versatility of TASQs have been demonstrated under different setups. We thus believe that biotinylated TASQs are high-value additions in the arsenal of tools for deciphering G4 biology from different angles and in complementary ways.

Limitations of the study

The G4DP-seq has certain limitations, notably related to the TASQs themselves, as they may act as ligands and could promote G4 formation (at least to some extent), thus favoring certain G4s beyond their true natural abundance. When we compared *in vitro* and *in vivo* conditions, we indeed observed a higher number of G4s *in vitro*, while the number of common G4s is high (75%). Another possible concern is a G4 structure selection of TASQs: they interact more strongly with parallel G4s in *in vitro* experiments, and we cannot rule out that such a bias does not occur in the G4DP conditions. Beyond TASQ, a limitation of G4DP is its reliance on a sonication step, with the classical issue of under- versus over-sonication (the former could possibly decrease the G4 capture efficiency while the latter could possibly disrupt G4 structure), which both limit the mapping resolution. Also, the sequencing depth used herein was certainly insufficient to gain reliable insights into low-abundance PG4FS. Finally, G4DP-seq has now to be performed in mammalian cells to make a straightforward comparison with BG4-based G4 ChIP-seq/CUT&Tag possible.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Genetic material
  - BioTASQ/BioCyTASQ/G4P based DNA dot blotting assays
  - G4DP-seq protocol *in vitro*
  - G4DP-seq protocol *in vivo*
  - qPCR validation
  - Analysis of sequencing data
  - Motif prediction
  - PFQSs identification and fold-enrichment analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106846.

ACKNOWLEDGMENTS

We thank the Bioinformatic Center in Nanjing Agricultural University for providing facilities to assist sequencing data analysis. This research was supported by grants from the National Natural Science Foundation of China (32070561, 32201782, and U20A2030) and the Jiangsu Funding Program for Excellent Postdoctoral Talent (2022ZB346). This work was also supported by the CNRS, the Agence Nationale de la Recherche (ANR-22-CE44-0039-01), and the European Union (PO FEDER-FSE Bourgogne 2014/2020 programs, FEDER no. BG0021532). We thank Judy M. Y. Wong (UBC Vancouver, CA) for critical reading of the manuscript.

AUTHOR CONTRIBUTIONS

Y.F. developed G4DP-seq and performed the experiments; Z.H. and Z.L. designed and implemented the bioinformatic analyses; F.R.S. synthesized both BioTASQ and BioCyTASQ; I.E.V. designed and optimized TASQs' design and synthesis; W.Z. and D.M. designed the study, interpreted the data, and wrote the manuscript.

DECLARATION OF INTERESTS

The CNRS (F.R.S., I.E.V., and D.M.) has licensed BioCyTASQ to Merck KGaA for commercialization.

Received: November 28, 2022
Revised: March 20, 2023
Accepted: May 4, 2023
Published: May 6, 2023

REFERENCES

1. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., and Balasubramanian, S. (2020). The regulation and functions of DNA and RNA G-quadruplexes. Nat. Rev. Mol. Cell Biol. *21*, 459–474.

2. Spiegel, J., Adhikari, S., and Balasubramanian, S. (2020). The structure and function of DNA G-quadruplexes. Trends Chem. *2*, 123–136.

3. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. *34*, 5402–5415.

4. Todd, A.K., Johnston, M., and Neidle, S. (2005). Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. *33*, 2901–2907.

5. Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res. *33*, 2908–2916.

6. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P., and Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. *33*, 877–881.

7. Guédin, A., Gros, J., Alberti, P., and Mergny, J.-L. (2010). How long is too long? Effects of loop size on G-quadruplex stability. Nucleic Acids Res. *38*, 7858–7868.

8. Bedrat, A., Lacroix, L., and Mergny, J.-L. (2016). Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. *44*, 1746–1759.

9. Lombardi, E.P., and Londoño-Vallejo, A. (2020). A guide to computational methods for G-quadruplex prediction. Nucleic Acids Res. *48*, 1603.

10. Rodriguez, R., Müller, S., Yeoman, J.A., Trentesaux, C., Riou, J.-F., and Balasubramanian, S. (2008). A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. J. Am. Chem. Soc. *130*, 15758–15759.

11. Marsico, G., Chambers, V.S., Sahakyan, A.B., McCauley, P., Boutell, J.M., Antonio, M.D., and Balasubramanian, S. (2019). Whole genome experimental maps of DNA G-quadruplexes in multiple species. Nucleic Acids Res. *47*, 3862–3874.

12. Hänsel-Hertsch, R., Beraldi, D., Lensing, S.V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., et al. (2016). G-quadruplex structures mark human regulatory chromatin. Nat. Genet. *48*, 1267–1272.

13. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D., and Balasubramanian, S. (2018). Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc. *13*, 551–564.

14. Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. *5*, 182–186.

15. Hui, W.W.I., Simeone, A., Zyner, K.G., Tannahill, D., and Balasubramanian, S. (2021). Single-cell mapping of DNA G-quadruplex structures in human cancer cells. Sci. Rep. *11*, 23641.

16. Mao, S.-Q., Ghanbarian, A.T., Spiegel, J., Martínez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D., and Balasubramanian, S. (2018). DNA G-quadruplex structures mold the DNA methylome. Nat. Struct. Mol. Biol. *25*, 951–957.

17. Shen, J., Varshney, D., Simeone, A., Zhang, X., Adhikari, S., Tannahill, D., and Balasubramanian, S. (2021). Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. Genome Biol. *22*, 143.

18. Hänsel-Hertsch, R., Simeone, A., Shea, A., Hui, W.W.I., Zyner, K.G., Marsico, G., Rueda, O.M., Bruna, A., Martin, A., Zhang, X., et al. (2020). Landscape of G-quadruplex DNA structural regions in breast cancer. Nat. Genet. *52*, 878–883.

19. Lyu, J., Shao, R., Kwong Yung, P.Y., and Elsässer, S.J. (2021). Genome-wide mapping of G-quadruplex structures with CUT& Tag. Nucleic Acids Res. *50*, e13.

20. Galli, S., Melidis, L., Flynn, S.M., Varshney, D., Simeone, A., Spiegel, J., Madden, S.K., Tannahill, D., and Balasubramanian, S. (2022). DNA G-quadruplex recognition in vitro and in live cells by a structure-specific nanobody. J. Am. Chem. Soc. *144*, 23096–23103.

21. Feng, Y., Tao, S., Zhang, P., Rota Sperti, F., Liu, G., Cheng, X., Zhang, T., Yu, H., Wang, X.-e., Chen, C., et al. (2022). Epigenomic features of DNA G-quadruplexes and their roles in regulating rice gene transcription. Plant Physiol. *188*, 1632–1648.

22. Yang, S.Y., Lejault, P., Chevrier, S., Boidot, R., Robertson, A.G., Wong, J.M.Y., and Monchaud, D. (2018). Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. Nat. Commun. *9*, 4730.

23. Renard, I., Grandmougin, M., Roux, A., Yang, S.Y., Lejault, P., Pirrotta, M., Wong, J.M.Y., and Monchaud, D. (2019). Small-molecule affinity capture of DNA/RNA quadruplexes and their identification in vitro and in vivo through the G4RP protocol. Nucleic Acids Res. *47*, 5502–5510.

24. Rota Sperti, F., Charbonnier, T., Lejault, P., Zell, J., Bernhard, C., Valverde, I.E., and Monchaud, D. (2021). Biomimetic, smart, and multivalent ligands for G-quadruplex isolation and bioorthogonal imaging. ACS Chem. Biol. *16*, 905–914.

25. Stefan, L., and Monchaud, D. (2019). Applications of guanine quartets in nanotechnology and chemical biology. Nat. Rev. Chem *3*, 650–668.

26. Monchaud, D. (2023). Template-assembled synthetic G-quartets (TASQs): multiTASQing molecular tools for investigating DNA and RNA G-quadruplex biology. Acc. Chem. Res. *56*, 350–362.

27. Yang, S.Y., Monchaud, D., and Wong, J.M.Y. (2022). Global mapping of RNA G-quadruplexes (G4-RNAs) using G4RP-seq. Nat. Protoc. *17*, 870–889.

28. Lyu, K., Chow, E.Y.-C., Mou, X., Chan, T.-F., and Kwok, C.K. (2021). RNA G-quadruplexes (rG4s): genomics and biological functions. Nucleic Acids Res. *49*, 5426–5450.

29. Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., and Balasubramanian, S. (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat. Meth. *13*, 841–844.

30. Guo, J.U., and Bartel, D.P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science *353*, aaf5371.

31. Weng, X., Gong, J., Chen, Y., Wu, T., Wang, F., Yang, S., Yuan, Y., Luo, G., Chen, K., Hu, L., et al. (2020). Keth-seq for transcriptome-wide RNA structure mapping. Nat. Chem. Biol. *16*, 489–492.

32. Yang, X., Cheema, J., Zhang, Y., Deng, H., Duncan, S., Umar, M.I., Zhao, J., Liu, Q., Cao, X., Kwok, C.K., and Ding, Y. (2020). RNA G-quadruplex structures exist and function in vivo in plants. Genome Biol. *21*, 226.

33. Rota Sperti, F., Mitteaux, J., Zell, J., Pipier, A., Valverde, I.E., and Monchaud, D. (2023). The multivalent G-quadruplex (G4)-ligands MultiTASQs allow for versatile click chemistry-based investigations. RSC Chem. Biol. https://doi.org/10.1039/D3CB00009E

34. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.

35. Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

36. O'Malley, R.C., Huang, S.S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. Cell *165*, 1280–1292.

37. Zheng, K.-w., Zhang, J.-y., He, Y.-d., Gong, J.-y., Wen, C.-j., Chen, J.-n., Hao, Y.-h., Zhao, Y., and Tan, Z. (2020). Detection of genomic G-quadruplexes in living cells using a small artificial protein. Nucleic Acids Res. *48*, 11706–11720.

38. Altman, D.G., and Bland, J.M. (1983). Measurement in medicine: the analysis of method comparison studies. J. R. Stat. Soc., Ser. D Stat. *32*, 307–317.

39. Rota Sperti, F., Dupouy, B., Mitteaux, J., Pipier, A., Pirrotta, M., Chéron, N., Valverde, I.E., and Monchaud, D. (2022). Click-Chemistry-based biomimetic ligands efficiently capture G-quadruplexes in vitro and help localize them at DNA damage sites in human cells. JACS Au *2*, 1588–1595.

40. Georgakopoulos-Soares, I., Parada, G.E., Wong, H.Y., Medhi, R., Furlan, G., Munita, R., Miska, E.A., Kwok, C.K., and Hemberg, M. (2022). Alternative splicing modulation by G-quadruplexes. Nat. Commun. *13*, 2404.

41. Zhang, J., Harvey, S.E., and Cheng, C. (2019). A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes. Nucleic Acids Res. *47*, 3667–3679.

42. Altmann, K.-H., Buchner, J., Kessler, H., Diederich, F., Kräutler, B., Lippard, S., Liskamp, R., Müller, K., Nolan, E.M., Samorì, B., et al. (2009). The state of the art of chemical biology. Chembiochem *10*, 16–29.

43. Schreiber, S.L. (2005). Small molecules: the missing link in the central dogma. Nat. Chem. Biol. *1*, 64–66.

44. Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat. Chem. *6*, 75–80.

45. Biffi, G., Tannahill, D., Miller, J., Howat, W.J., and Balasubramanian, S. (2014). Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. PLoS One *9*, e102711.

46. Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N.J., Santos, N.D., Silvester, J., Wei, V., Garcia, J., et al. (2017). CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. Nat. Commun. *8*, 14432.

47. Wang, Y., Yang, J., Wild, A.T., Wu, W.H., Shah, R., Danussi, C., Riggins, G.J., Kannan, K., Sulman, E.P., Chan, T.A., and Huse, J.T. (2019). G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. Nat. Commun. *10*, 943.

48. Zhang, M., Wang, B., Li, T., Liu, R., Xiao, Y., Geng, X., Li, G., Liu, Q., Price, C.M., Liu, Y., and Wang, F. (2019). Mammalian CST averts replication failure by preventing G-quadruplex accumulation. Nucleic Acids Res. *47*, 5243–5259.

49. Xu, Y.-Z., Jenjaroenpun, P., Wongsurawat, T., Byrum, S.D., Shponka, V., Tannahill, D., Chavez, E.A., Hung, S.S., Steidl, C., Balasubramanian, S., et al. (2020). Activation-induced cytidine deaminase localizes to G-quadruplex motifs at mutation hotspots in lymphoma. NAR Cancer *2*, zcaa029.

50. Yang, S.Y., Chang, E.Y.C., Lim, J., Kwan, H.H., Monchaud, D., Yip, S., Stirling, P.C., and Wong, J.M.Y. (2021). G-quadruplexes mark alternative lengthening of telomeres. NAR Cancer *3*, zcab031.

51. De Magis, A., Kastl, M., Brossart, P., Heine, A., and Paeschke, K. (2021). BG-flow, a new flow cytometry tool for G-quadruplex quantification in fixed cells. BMC Biol. *19*, 45.

52. Masson, T., Landras Guetta, C., Laigre, E., Cucchiarini, A., Duchambon, P., Teulade-Fichou, M.-P., and Verga, D. (2021). BrdU immuno-tagged G-quadruplex ligands: a new ligand-guided immunofluorescence approach for tracking G-quadruplexes in cells. Nucleic Acids Res. *49*, 12644–12660.

53. Lago, S., Nadai, M., Cernilogar, F.M., Kazerani, M., Domíniguez Moreno, H., Schotta, G., and Richter, S.N. (2021). Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. Nat. Commun. *12*, 3885.

54. Yangyuoru, P.M., Di Antonio, M., Ghimire, C., Biffi, G., Balasubramanian, S., and Mao, H. (2015). Dual binding of an antibody and a small molecule increases the stability of TERRA G-quadruplex. Angew. Chem. Int. Ed. *54*, 910–913.

55. Yu, Z., Spiegel, J., Melidis, L., Hui, W.W.I., Zhang, X., Radzevičius, A., and Balasubramanian, S. (2023). Chem-map profiles drug binding to chromatin in cells. Nat. Biotechnol.

56. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.

57. Jung, Y., and Han, D. (2022). BWA-MEME: BWA-MEM emulated with a machine learning approach. Bioinformatics *38*, 2404–2413.

58. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

59. Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. *45*, W122–W129.

60. Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J. (2012). High-resolution mapping of open chromatin in the rice genome. Genome Res. *22*, 151–162.

61. Zhang, W., and Jiang, J. (2015). Genome-wide mapping of DNase I hypersensitive sites in plants. In Plant Functional Genomics (Springer), pp. 71–89.

62. Fang, Y., Chen, L., Lin, K., Feng, Y., Zhang, P., Pan, X., Sanders, J., Wu, Y., Wang, X.-e., Su, Z., et al. (2019). Characterization of functional relationships of R-loops with gene transcription and epigenetic modifications in rice. Genome Res. *29*, 1287–1297.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
BioTASQ	This paper	N/A
BioCyTASQ	Sigma-Aldrich	SCT246
HRP conjugated Rabbit anti-Biotin	BBI	D111135-0100
anti-FLAG antibody	BBI	D110005
G4P Recombination Protein	This paper	N/A
Dynabeads™ M-280 streptavidin	Invitrogen™	11206D
Biological samples		
Nipponbare (Japonica)	This papper	N/A
Chemicals, peptides, and recombinant proteins		
HEPES	Sigma-Aldrich	H7006-100G; CAS No.:75277-39-3
0.5M EDTA	Ambion™	AM9261;
5M NaCl	Ambion™	AM9760G
PMSF	Sigma-Aldrich	P7626 CAS No.: 329-98-6
Glycine	Sigma-Aldrich	G7126-500G CAS No.: 56-40-6
Spermine	Sigma-Aldrich	85590-5G CAS No.: 71-44-3
Sperminine	Sigma-Aldrich	85578-5G CAS No.: 334-50-9
Triton™ X-100	Sigma-Aldrich	T8787-100ML CAS No.: 9036-19-5
UltraPure™ 1 M Tris-HCl Buffer, pH 7.5	Invitrogen™	15567027
KCl	Sigma-Aldrich	P9333-500G CAS No.: 7447-40-7
NON-Fat Powdered Milk	BBI	A600669-0250
20% SDS	Invitrogen™	AM9820
DTT	BBI	A620058 CAS No.:3483-12-3
TERGITOL™ (Type NP-40)	Sigma-Aldrich	NP40S-100ML
Biotin	Sigma-Aldrich	B4639 CAS No.: 58-85-5
$NaHCO_3$	Sigma-Aldrich	S6297-1KG CAS No.: 144-55-8
Phenol	Sigma-Aldrich	P4557-400ML CAS No.: 108-95-2
Chloroform	Sigma-Aldrich	C2432-500ML CAS No.: 67-66-3
Critical commercial assays		
NEBNext® Ultra™ II DNA Library Prep Kit for Illumina	NEB	E7645S
Deposited data		
Raw and preprocessed data	This paper	GSE132775
Rice reference genome	MSU v7.0	http://rice.plantbiology.msu.edu/
Oligonucleotides		
Primers for qPCR assay, see Table S5	This paper	N/A
Oligos for CD and dot blot assay, see Table S6	This paper	N/A
Software and algorithms		
fastp version 0.21.0	Chen et al.[56]	https://github.com/OpenGene/fastp
BWA (mem algorithm, version 0.7.17)	Jung et al.[57]	https://github.com/lh3/bwa

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SAMtools(version 1.5, option -markdup)	Li et al.[58]	https://www.htslib.org/
MACS2 (version 2.1.1)	Zhang et al.[34]	https://pypi.org/project/MACS2/2.1.4/
MEME-ChIP	P. Machanick et al.[35]	http://meme-suite.org/tools/meme-chip
QuadParser	python	https://github.com/dariober/bioinformaticscafe/blob/master/fastaRegexFinder.py
AgriGO	Tian et al.[59]	http://systemsbiology.cau.edu.cn/agriGOv2/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. David Monchaud, david.monchaud@cnrs.fr.

Materials availability

This study did not generate new unique reagents.

Data and code availability

The BG4-DNA-IP-seq data set is available at GSE132775; the G4DP-seq data sets generated in this study are available at the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE200171 (secure token: mxwdquwcljuhvaj).

METHOD DETAILS

Genetic material

Seeds of rice (*Oryza sativa*) cultivar Nipponbare (Japonica) were pregerminated at 25°C for 3 d under dark condition. Uniformly germinated seeds were transferred to pots containing the nutrient soil and grown in a greenhouse with an automatically controlled condition: 28 to 30°C and a 14 h/10 h light/dark cycle. For G4DP-seq *in vivo*: two-week-old rice seedlings were cut into 1-1.5 cm in length and merged into 1% of formaldehyde (v/v) in HEPES buffer pH = 8.0 (20 mM HEPES, 1 mM EDTA, 100 mM NaCl and 1 mM PMSF) for cross-linking at 25°C for 10 min *in vacuo*. After quenching the excess of formaldehyde by adding 0.125 M final concentration of glycine followed by vacuuming for additional 5 min, the cross-linked leaves were rinsed with autoclaved ddH_2O and air-dried. For G4DP-seq *in vitro* and *in vivo*: cross-linked or native leaves were ground into fine powder using liquid nitrogen; after several washing steps using NIB (nuclear isolation buffer, containing spermine, spermidine and mercaptoethanol), NWB (nuclear washing buffer, containing triton X-100) and NDB (nuclear digestion buffer),[60,61] chromatin (for cross-linked leaves) and genomic DNA (for native leaves) were isolated and used directly for downstream experiments (or stored at −80°C for later use).

BioTASQ/BioCyTASQ/G4P based DNA dot blotting assays

Two genomic DNA replicates were isolated from non-crosslinked rice materials. Biological replicates along with oligonucleotides forming iM ($^{5'}C_4AC_2T_2C_4AC_3TC_4AC_3TC_4^{3'}$) and G4 ($^{5'}TGAG_3TG_3TAG_3TG_3TA_2^{3'}$), and AT-rich oligonucleotide ($^{5'}ATATA_3T2ACACA_2TGAGTA_2TACA_2GTACAT_2AT_2^{3'}$) were denatured and re-associated in G4 reconstruction buffer (10 mM Tris-HCl, pH = 7.5, 150 mM KCl) at 95°C for 10 min before being slowly cooled down to RT. All the re-associated DNA samples were then loaded on Amersham Hybond-N^+-nylon membrane followed by pre-blocking in 5% milk for 30 min at RT, three technological repeats for each sample. The pre-blocked membrane was incubated with the BioTASQ/BioCyTASQ ligand (3 μM, *ca.* 1:300 dilution) overnight at 4°C upon gentle steering. After washing with 1XPBS (three times), the membrane was further incubated with anti-Biotin (HRP) antibody for an additional 1.5 h at 4°C upon gentle steering. For G4P protein, 2 μg anti-flag antibody was added for another 1.5 h, after washing by 1XPBS (three times), 2 μg anti-rabbit HRP antibody was added and the incubation was continued for 1 h before checking the immunosignal. The remaining procedures for immune-signal development were conducted according to standard procedures.[62]

G4DP-seq protocol *in vitro*

The purified nuclei, which were prepared using non cross-linked rice leaf tissue following the published procedures,[60] were used for isolation of high-quality genomic DNA. The purified genomic DNA was diluted with 1X sonication buffer (50 mM Tris-HCl, 10 mM EDTA and 1% SDS w/v, pH = 8.0), then fragmented into 100 - 500 bp in size using the water-based Biorupter (Diagnode). A total of 5 μg purified fragmented genomic DNA was diluted in a G4-stabilizing buffer (150 mM KCl and 10 mM Tris-HCl, pH = 7.5), denatured at 95°C for 10 min, then slowly cooled down to RT. BG4-DNA-IP-seq was performed as previously described;[18] regarding G4DP-seq: the prefolded DNA was diluted with G4DP buffer (10 mM Tris-HCl, 5 mM EDTA, 150 mM KCl, 0.5 mM DTT, 0.5% NP-40, pH = 7.4), then incubated with 100 μM of BioTASQ/BiocyTASQ (water solution) under gentle steering at 4°C overnight. 30 μl of streptavidin-coupled Dynabeads were prewashed 1 min using G4DP buffer (three times) and added to the reaction mixture and steered for 4 h at 4°C. After washings (three times) with G4DP buffer, the washed beads with BiocyTASQ/BioTASQ bound G4 DNA were eluted twice with 200 μL elution buffer (0.1 M $NaHCO_3$ and 1% SDS w/v) at 65°C for 15 min each. BioCyTASQ/BioTASQ-bound DNA and input DNA was purified using phenol/chloroform extraction followed by cold ethanol precipitation. Of note, to better reflect the nonspecific background of these methods, we did not use 50% (common conditions) but 100% of the DNA precipitated using IgG (BG4-DNA-IP-seq) or biotin controls (G4DP-seq) for library preparation and sequencing, with 1 or 2 additional PCR amplification cycles. All libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S) for paired-end mode sequencing on Illumina NovaSeq platform.

G4DP-seq protocol *in vivo*

Nuclei were prepared and purified from cross-linked rice leaf tissue following published procedures.[60] The purified nuclei were then fragmented into sizes ranging from 100 to 500 bp in sonication buffer (10 mM Tris-HCl, 5 mM EDTA, 150 mM KCl, 0.5 mM DTT, 0.5% NP-40 and 0.5% SDS, pH = 7.4) using the water-based Biorupter. After being centrifuged at 12,000 rpm at 4°C for 10 min, the supernatant containing fragmented chromatin was carefully transferred to a new 1.5 mL tube and kept on ice. The fragmented chromatin was diluted with incubation buffer (10 mM Tris-HCl, 5 mM EDTA, 150 mM KCl, 0.5 mM DTT and 0.5% NP-40, pH = 7.4) until the final concentration of SDS was below 0.1%. After keeping 1/10 volume of diluted supernatant as input, the remaining volume was incubated with 30 μL of BioTASQ/biotin (control) at 4°C overnight, then incubated with 30 μL of Dynabeads, which were prewashed using incubation buffer, for another 4 h at 4°C. The washed beads with BioTASQ-bound DNA were eluted with 200 μL elution buffer twice at 65°C for 15 min each, then were reverse cross-linked at 65°C overnight after the addition of NaCl (final concentration: 0.2 M). The de-crosslinked DNA was purified by incubation with RNase A at 37°C for 45 min, then phenol/chloroform extraction and cold ethanol precipitation. Precipitated DNA and input DNA was used for library preparation as described above.

qPCR validation

For both the G4DP experiments performed *in vitro* and *in vitro*, 2 μL of input and precipitated DNA (3 ng/μL each) was used as template. The enrichment of precipitated DNA was calculated by using the $2^{(\Delta\Delta Ct)}$ method and expressed as fold-change over the input. Each experiment was repeated three times; all primer sequences are listed in Tables S2 and S8. Significance test was performed by using One-way Anova analysis in SPSS Statistics 13 program, which p-value < 0.05 means significant ($p < 0.001$ ***, $P < 0.01$ ** and $p < 0.05$ *).

Analysis of sequencing data

Raw sequencing data were quality-checked and cleaned using fastp[56] (version 0.21.0). All cleaned reads were aligned to the MSU v7.0 reference genome http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/) using BWA[57] (mem algorithm, version 0.7.17) with default parameters. SAMtools[58] (version 1.5, option -markdup) was used for complete removal of any PCR duplicates. MACS2[34] (version 2.1.1) was used for G4 peak calling using reads with alignment length greater than 50. The command and parameters for G4 peak calling were: macs2 callpeak -g 3.8e+8 -f BAM –nomodel -q 0.01. MACS2 software was used for peak calling with the -c parameter relative to the input, IgG and anti-FLAG as input files for peak calling, respectively. Biologically replicated G4 peaks were considered as G4 peaks with high confidence (command intersect of the bedtools package). The plot Correlation program of deepTools was used for calculating the Spearman's rank correlation coefficients between biological replicates. The GC- and AT-skews were calculated using the following formula: for the GC-skew: [(G-C)/(G+C)]*100%; for the AT-skew:

[(A-T)/(A+T)]*100%. Gene Ontology (GO) analyses of genes were analyzed using AgriGO (http://systemsbiology.cau.edu.cn/agriGOv2/).

Motif prediction

G4 motifs within G4 peaks were predicted using MEME-ChIP (http://meme-suite.org/tools/meme-chip)[35] with the following parameters: minimum width 5 bp, maximum width 25 bp. Only the top significantly enriched motifs (*i.e.*, with the highest *E*-values) are shown in Figures 3 and 4.

PFQSs identification and fold-enrichment analyses

Putative G4-forming sequences (PG4FSs) were identified by screening the whole genome sequences using QuadParser (https://github.com/dariober/bioinformaticscafe/blob/master/fastaRegexFinder.py). The fold-enrichment of PG4FSs ($G_{2+}N_{1-12}$) was calculated relative to random controls across the genome (bedtools shuffle command, observed values divided by average of 100 randomizations values).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis is based on free software and R, which are described in the STAR Methods and key resources table.