# Epigenomic Features of DNA G-Quadruplexes and Their Roles in Regulating Rice Gene Transcription

Yilong Feng, Shentong Tao, Pengyue Zhang, Francesco Rota Sperti, Guanqing Liu, Xuejiao Cheng, Tao Zhang, Hengxiu Yu, Xiu-E Wang, Caiyan Chen, et al.

# Epigenomic Features of DNA G-Quadruplexes and Their Roles in Regulating Rice Gene Transcription

Yilong Feng[1,#], Shentong Tao[1,#], Pengyue Zhang[1,#], Francesco Rota Sperti[2], Guanqing Liu[3,4], Xuejiao Cheng[1], Tao Zhang[3,4], Hengxiu, Yu[3], Xiu-e Wang[1], Caiyan Chen[5], David Monchaud[2], Wenli Zhang[1*,‡]

[1]State Key Laboratory for Crop Genetics and Germplasm Enhancement, JCIC-MCP, CIC-MCP, Nanjing Agricultural University, No.1 Weigang, Nanjing, Jiangsu 210095, P. R. China. [2]Institut de Chimie Moleculaire, ICMUB CNRS UMR 6302, UBFC Dijon, France. [3]Key Laboratory of Plant Functional Genomics of the Ministry of Education, College of Agriculture, Yangzhou University, Yangzhou 225009, China. [4]Jiangsu Key Laboratory of Crop Genetics and Physiology, Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, Agricultural College of Yangzhou University, Yangzhou, Jiangsu 225009, China. [5]Institute of Subtropical Agriculture, Chinese Academy of Sciences, Changsha, Hunan, 410125, P.R. China.

**Short Title: Functional Characterization of G4 in Rice**

[#] These authors contributed equally to this work

[*] Author for communication: wzhang25@njau.edu.cn

‡ Senior author.

**One-sentence Summary**

Rice DNA G-quadruplexes exhibit genic-dependent enrichment of epigenomic signatures and play important roles in modulating gene transcription.

**Author Contributions:**

W.L.Z. conceived and designed the study; Y.L.F. performed BG4-DNA-IP, PCR and dot blotting experiments. P.Y.Z. and S.T.T. analyzed the data; F.R.S. prepared the BioTASQ compound. X.J.C. helped with the genomic DNA preparation. X.E.W and C.Y.C. helped with generation of the Cas9 rice line. G.Q.L and T.Z. helped with construction of JBrowser. H.X.Y. helped with preparation of transgenic lines. W.L.Z., D.M., Y.L.F., P.Y.Z and S.T.T. interpreted the results. W.L.Z. and D.M. wrote the manuscript with contributions from all authors.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Author (https://academic.oup.com/plphys/pages/general-instructions) is: Wenli Zhang (wzhang25@njau.edu.cn)

**Abstract**

A DNA G-quadruplex (G4) is a non-canonical four-stranded nucleic acid structure involved in many biological processes in mammals. The current knowledge on plant DNA G4s, however, is limited; whether and how DNA G4s impact gene expression in plants is still largely unknown. Here, we applied a protocol referred to as BG4-DNA-IP-seq followed by a comprehensive characterization of DNA G4s in rice (*Oryza sativa* L.); we next integrated dG4s (experimentally detectable G4s) with existing omics data and found that dG4s exhibited differential DNA methylation between TE (transposable element) and non-TE (non-transposable element) genes. dG4 regions displayed genic-dependent enrichment of epigenomic signatures; finally, we showed that these sites displayed a positive association with expression of DNA G4-containing genes when located at promoters, and a negative association when located in the gene body, suggesting localization-dependent promotional/repressive roles of DNA G4s in regulating gene transcription. This study reveals interrelations between DNA G4s and epigenomic signatures, as well as implicates DNA G4s in modulating gene transcription in rice. Our study provides valuable resources for the functional characterization or bioengineering of some of key DNA G4s in rice.

**Key words:** G4, global mapping, regulation of gene transcription, epigenomic signatures, rice

## Introduction

A DNA G-quadruplex (G4) folds from guanine (G)-rich nucleic acid sequences harboring a specific G4 motif. It contains the self-assembly of four Gs forming co-planar G-quartets. It was first characterized in 1962 by X-ray studies of GMP gels (Gellert et al., 1962). DNA G4 was discovered using DNA oligonucleotides with sequences from immunoglobulin switching regions (Sen and Gilbert, 1988) and telomeres (Sundquist and Klug, 1989) *in vitro*. It has been reported that DNA G4s can be formed in different functional genomic contexts in mammalian and plant genomes (Hansel-Hertsch et al., 2016), including the functional heterochromatic regions, such as telomeres (Liu et al., 2016; Wu et al., 2020), G-rich repetitive sequences, like rDNA (Wallgren et al., 2016), genic regions (Hansel-Hertsch et al., 2016), and retrotransposons with long terminal repeats (LTR) in plants (Lexa et al., 2014).

DNA G4s have been found or expected to play important roles in various biological processes in eukaryotes (Yadav et al., 2017; Varshney et al., 2020), such as stress response (Garg et al., 2016; Fleming et al., 2019), DNA replication (Sparks et al., 2019), maintenance of telomeres (Bao et al., 2019) and regulation of gene transcription (Garg et al., 2016). Proteins for specific binding of DNA or RNA G4s

have been characterized in mammals (Mishra et al., 2016), bacteria (*Escherichia Coli; E. Coli*), yeast (*Saccharomyces cerevisiae; S. cerevisiae*) and plants (*Arabidopsis thaliana; A. thaliana*) (Kang and Henderson, 2002). For instance, *Nucleoside Diphosphate Kinase1* (*ZmNDPK1*) encodes a DNA G4-binding protein in maize (*Zea mays* L.) that exhibits a similar structure to the human homolog NM23-H2 for binding of G4 (Kopylov et al., 2015), indicating existence of possibly conserved regulatory mechanisms between mammals and plants. Several high-throughput methodologies have been developed for global mapping of DNA G4s in mammals, thus revolutionizing the advance of DNA G4 studies (Varshney et al., 2020): this includes chromatin immunoprecipitation coupled with sequencing (ChIP-seq) methods using either the G4-specific antibodies BG4 (Hansel-Hertsch et al., 2020) and D1 (Liu et al., 2016), or an artificial, truncated DHX36 protein termed G4 probe ChIP (G4P-ChIP) (Zheng et al., 2020). Molecular tools have also been implemented such as the template-assembled synthetic G-quartet (TASQ)-related G4-RNA-specific precipitation and sequencing (G4RP-seq) (Yang et al., 2018) and the pyridostatin (PDS)-mediated polymerase stop assays (G4-seq) (Marsico et al., 2019) for DNA G4s identification, and (rG4-seq) for identification of RNA G4s (Kwok et al., 2016).

BG4 is a single chain antibody from a phage-display library screen for G4-binding proteins (Biffi et al., 2013). It is now widely used to visualize DNA G4 structures located in human cells as well as in rice (*Oryza sativa* L.) nuclei (Zhang et al., 2018; Fang et al., 2019). BG4 does not exhibit DNA G4 conformation-related binding, it has

been applied to map DNA G4 structures in humans (Hansel-Hertsch et al., 2016; Hansel-Hertsch et al., 2020). Another antibody known as D1 can be alternatively used for performing ChIP-seq experiments (Liu et al., 2016). Collectively, the very important variation of the DNA G4 landscapes highlighted by different but complementary techniques (> 8,000 DNA G4 sites by D1-ChIP, > 10,000 by BG4-ChIP, and > 120,000 by G4P-ChIP) brightly illustrate the difficulties of identifying DNA G4s *in vivo* (Spiegel et al., 2020).

Putative G4-forming sequences (PQFSs) of DNA have been computationally identified in several plant species (Lexa et al., 2014; Garg et al., 2016; Ge et al., 2019) , including DNA PQFSs in *A. thaliana* (Takahashi et al., 2012), maize (Andorf et al., 2014), rice (Wang et al., 2015), wheat (*Triticum aestivum* L.) (Cagirici and Sen, 2020) and barley (*Hordeum vulgare* L.) (Cagirici et al., 2021). DNA G4s seem to be prevalent across the plant kingdom but variations in genomic distribution has been observed between monocots and *A. thaliana* (Cagirici and Sen, 2020). Plant DNA G4s have been proposed to be involved in various physiological processes such as regulation of gene expression and translation, plant growth and development and stress responses (Kwok et al., 2015; Garg et al., 2016; Yadav et al., 2017; Cho et al., 2018; Griffin and Bass, 2018; Cagirici and Sen, 2020) but the folding of these PQFSs *in vivo* still needs to be experimentally validated.

G4-seq was employed to identify DNA G4s in *A. thaliana* (Marsico et al., 2019), more to demonstrate the applicability of the method to various species than to provide

insights into plant DNA G4 biology. Similarly, rG4-seq was used to profile RNA G4s in *A. thaliana* (Yang et al., 2020), but their actual cellular roles were not fully addressed. In this study, we decided to perform BG4-based IP-seq investigations, combined with epigenetics analyses, in the aim of unravelling the functional relevance of DNA G4s in rice.

## Results

### Application of BG4-DNA-IP-seq for global mapping of G4s

The binding affinity of BG4 to plant DNA G4s was first confirmed in rice using dot blotting assays (**Supplemental Figure S1A**). A non-neglectable cross-reactivity of BG4 to i-motifs, another four-stranded DNA structure that folds from C-rich strands harbouring a specific motif capable of forming i-motif at acidic pH (pH=5.5) (Gehring et al., 1993), was reported (Zeraati et al., 2018). We also assessed whether BG4 interacts with i-motifs in G4-favorable conditions ($K^+$+ PEG, pH=7.5). After conducting a dot blotting assay using reported G4- and i-motif-forming sequences (Zeraati et al., 2018), we found that, in our conditions, BG4 was specific to G4s (**Supplemental Figure S1B**). We then applied BG4-mediated IP on rice seedlings by following the published procedures (Hansel-Hertsch et al., 2018), which includes fragmentation of genomic DNA, incubation with BG4 antibody and isolation of BG4-bound G4s (**Figure 1A**). Again, we observed that dot-signal intensity of genomic DNA obtained in $K^+$+PEG conditions was approximately 2.4- and 1.4-fold stronger

147 than those obtained in either $K^+$- or PEG-rich conditions, respectively (**Supplemental**
148 **Figure S1A**). As controls, we used a synthetic rice dG4 (experimentally detectable
149 G4s), for which the obtained signal was significantly positive, and both $H_2O$ and a
150 synthetic AT-rich oligonucleotide, which were both almost undetectable
151 (**Supplemental Figure S1A**). Moreover, the presence of G4s under the $K^+$+PEG
152 condition was also confirmed using a dot blotting assay with BioTASQ, a small
153 molecule that specifically binds with DNA/RNA G4s in humans (Renard et al., 2019)
154 (**Supplemental Figure S1C**). We then sequenced two replicated BG4-DNA-IP-seq
155 libraries under the $K^+$+PEG condition (**Supplemental Table S1**). Given a high
156 correlation between the two replicates (r = 0.89) (**Figure 1B**), we merged repI and repII
157 to re-identify total G4 peaks, and obtained 47,417, 29,430 and 27,042 dG4s relative to
158 input, IgG and anti-Flag DNA IP-seq as controls (**Figure 1C**), respectively, and 23,685
159 common dG4 peaks (**Figure 1D, E**). We thus decided to use these common dG4 peaks
160 for downstream assays.
161
162 **Validation of dG4s**
163 To verify the accuracy of the 23,685 common dG4s, we analyzed the sequence
164 features of dG4 peaks. We identified various PQFS patterns associated with dG4s
165 (**Figure 2A**), which were also detected from computationally predicted G4s (PQFS)
166 (Marsico et al., 2019) using fastaRegexFinder.py script (Fujimoto et al., 2020). We
167 observed a similar trend regarding the percentage of each pattern between dG4s and

predicted G4s: for instance, among the five patterns $G_{2+}L_{1-12}$, $G_{3+}L_{1-3}$, $G_{3+}L_{4-5}$, $G_{3+}L_{6-7}$ and $G_{3+}L_{8-12}$, the highest percentage was obtained for $G_{2+}L_{1-12}$ and the lowest for $G_{3+}L_{6-7}$. Next, after randomly shuffling 23,685 dG4s across the genome, we generated random regions with the same size as the common dG4 peaks and calculated the fold enrichment of PQFSs in dG4 peaks as compared to random regions. We found that all PQFSs were significantly enriched in the BG4 assay, especially for $G_3+L_{6-12}$ (**Figure 2B**), indicating that BG4 does bind to genomic G4-forming sequences.

We then randomly selected 17 G4 positive loci and 2 G4 negative loci for conducting BG4-DNA-IP-qPCR (**Supplemental Table S2.xlsx**), and found that all positive loci were enriched relative to the negative locus (**Figure 2C**). To further demonstrate the relevance of our approach, we reexamined 9 G4 positive loci and 1 G4 negative locus that were previously experimentally validated (Garg et al., 2016), and found that 8 of them (7 positive, 1 negative) matched our data (**Figure 2D**). Collectively, all above results confirm that the quality of our dG4 data can be justified for the downstream assay.

**Genomic distribution and sequence features of dG4s and udG4s**

Using a PQFS-based formula (see Materials and Methods), we identified *ca*. 1,800,000 total PQFSs, hereafter termed predicted G4s (**Supplemental Table S3.xlsx**) in the rice genome. A comparison with our dG4 results (above) indicates that the experimentally determined G4s cover ca. 5% of the predicted G4s (**Figure 3A**). These

results are reminiscent of the studies performed with the human genome, in which *ca*. 1% of all sequences detected by G4-seq (Chambers *et al.*, 2015) was detected by G4 ChIP-seq. (Hansel-Hertsch *et al.*, 2016). We decided to include over 95% of the sequences that are predicted but not experimentally detected in our subsequent analyses, referring to them as udG4s (experimentally undetectable G4s). In addition, 417 dG4s contained non-canonical PQFS (**Supplemental Table S4.xlsx**), such as the $(G3_{+}L1\text{-}12)3+G3$ pattern with at least one regular GGG run interrupted by a non-G bulge (e.g. GAG or GGAG representing one bulged stem of 2 or 3 stacked Gs).

We then compared genomic distributions and the PQFS size between dG4s and udG4s: dG4s-(PQFS) were more prevalent (1.5-fold) in promoters and in 5' UTRs (4.2-fold), but less present in exons (0.5-fold) compared to udG4s-(PQFS) (**Figure 3B**). On average, dG4s-(PQFS) were 51% longer than udG4s-(PQFS) (**Figure 3C**) and intervals between two neighboring PQFSs were 43% shorter in dG4s-(PQFS) compared to udG4s-(PQFS) (**Figure 3D**). These analyses suggest that a large size of PQFSs with a shorter distance between two neighboring PQFSs is likely prone to G4 formation or facilitates G4 stabilization or binding to BG4.

After calculating the GC content, GC and AT skews around ± 1 kb of the midpoint of d/udG4s regions, we also observed that dG4s had 12% higher GC content at ± 100 bp around the center of the PQFS than udG4s (**Figure 3E, top panel**), and they exhibited more conspicuous GC- and AT-skews than udG4s (**Figure 3E, middle and bottom panels**). We then conducted motif identification by extracting the DNA sequence

around the G4 peaks: we notably found repetitive GAG and GGCGG motifs in the putative basic pentacysteine (BPC5) and ethylene-responsive factor 105 (ERF105) transcription factor binding sites documented in *A. thaliana* (Simonini et al., 2012) (**Figure 3F**), suggesting that G4s may play regulatory roles in various plant processes.

Collectively, all above analyses indicate that dG4s have distinct genomic distribution and sequence features relative to udG4s.

**Interrelationship between G4s and DNA methylation**

It was previously reported that G4s can influence the epigenetic status of a genome, notably *via* the modulation of the DNA methylome (Mao et al., 2018). G4s can trigger hypomethylation through the sequestration of DNA methyltransferase. We thus investigated the G4 methylation levels by counting methylated cytosines at ± 1 kb around the center of dG4s ($n$ = 23,685) and udG4s ($n$ = 23,685), with similar C% using published bisulfite sequencing (BS-seq) data (SRP043447) (Hu et al., 2014) (**Supplemental Figure S2A**). We found that dG4s exhibited 58.6% less CG and 48.9% less CHG methylation levels compared to udG4s, but 17.5% higher CHH methylation levels at ± 100 bp around the center (**Figure 4A**). When dividing dG4s into high ($n$ = 5,000) and low ($n$ = 5,000) peak intensity with similar C% (**Supplemental Figure S2B**), the difference in methylation levels was subtle but significant: G4s with high peak intensity tended to have less CG and CHG (7.3 and 2.4%, respectively) and more

CHH methylation (12.4%) than G4s with low peak intensity (**Supplemental Figure S2C**), which is consistent with hypomethylated G4s in humans (Mao et al., 2018).

We next examined the relationship between G4s and DNA N6-methyladenine (D-6mA) sites, as adenine methylation is associated with gene expression, mostly *via* its ability to destabilize duplex stems. To this end, we integrated dG4 data with published DNA-6mA IP data (GSE103145) (Zhou et al., 2018) and detected 4,399 (18.6%) dG4s that colocalize with D-6mA sites. We observed that dG4s with D-6mA displayed 9.0% higher BG4-DNA-IP intensity (expressed in Reads Per Kilobase per Million mapped reads, RPKM) than those without D-6mA (**Figure 4B)**, and a similar tendency was observed for udG4s (**Figure 4B, right)**. We plotted RPKM of D-6mA around ± 1 kb of the center of both dG4s and udG4s peaks with similar A% (**Supplemental Figure S2D, E, F**). We found that dG4s had lower D-6mA levels around the center of the peak, compared to the flanking regions, but had an overall 27% higher D-6mA level than udG4s (**Figure 4C**). When comparing D-6mA levels between dG4s with high ($n$ = 5,000) and low ($n$ = 5,000) peak intensity with similar A% (**Supplemental Figure S2G**), we found that dG4s with high intensity had approximately 36.6% higher D-6mA levels (**Supplemental Figure S2H**). All above analyses demonstrate that dG4s have less CG and CHG methylation but they tend to have more CHH and D-6mA methylation.

To assess whether DNA methylation can affect G4 formation on a genome-wide scale, we next extracted genomic DNA from DMSO (Control, CK) and

zebularine-treated rice seedlings (a DNA demethylating agent), and from a mutant with CRISPR-Cas9-mediated knockout of the 6mA methyltransferase gene, *LOC_Os01g16180*. We then conducted BG4-, anti-5mC and anti-6mA-related DNA dot blotting assays (n = 2): we found a reduction of the global DNA methylation levels (*ca*. 40%) when rice seedling are treated with zebularine, combined with a global increase of G4 intensity (*ca*. 25%) (**Figure 4D, right**), thus confirming the negative association between DNA-5mC and G4 folding. Conversely, we found a global decrease of both DNA-6mA levels (*ca*. 24%) and G4 intensity (ca. 31%) in the *LOC_Os01g16180* mutant, (**Figure 4D, left**), indicating a positive association between DNA-6mA levels and G4 intensities. We further conducted BG4-DNA-IP-seq using zebularine treated DNA, referred to as Zebu hypoDNA. We detected more Zebu hypoDNA biased dG4s (n = 1,906) than CK biased dG4s (n = 6,16) (**Figure 4E, left**). In particular, we found that Zebu hypoDNA biased dG4s had a lower p-value relative to CK biased dG4s (**Figure 4E, right**), indicating that the peak density of Zebu hypoDNA biased dG4s changed more as compared to CK biased dG4s. Thus, the IP-seq result is overall consistent with the dot blotting result in Figure 4D. The detailed relationships between DNA methylation and G4 formation need to be further investigated. Hence, these results show distinct impacts of DNA-5mC and DNA-6mA on G4 formation *in vitro*.

**Differential methylation in transposable and non-transposable element genes**

To investigate whether G4s affect the methylation levels of G4-overlapping genes, which exhibit at least 1 bp overlapping G4 regions from 1 kb upstream of the transcription start sites (TSSs) to the entire gene bodies, we re-analyzed the published BS-seq data mentioned above, and counted DNA-5mC methylation levels across ± 1 kb of the TSSs and the transcription termination sites (TTSs) of both transposable element (TE) genes (TEGs, *n* = 1,996) and non-transposable element (non-TE) genes (non-TEGs, *n* = 13,352), with both dG4s or udG4s containing similar C% (**Supplemental Figure S3A**). A distinct profile of DNA methylation was observed between TEGs and non-TEGs (**Figure 5A**): we found that non-TEGs with dG4s exhibited 29.8% higher CG methylation in gene bodies, but lower downstream of the TTS (**Figure 5A, top left**), a subtle difference in CHG methylation (**Figure 5A, middle left**) and 28.2% higher CHH methylation upstream of the TSSs (**Figure 5A, bottom left**). In contrast, we found that TEGs with dG4s (*n* = 1,996) exhibited less CG and CHG methylation levels (16.1 and 21.4%, respectively) across the entire regions, but no difference in CHH methylation across the downstream of TSSs compared with TEGs with udG4s (*n* = 1,996) (**Figure 5A, right**).

We performed a similar analysis for D-6mA enrichment between TEGs and non-TEGs with both dG4s and udG4s containing similar A% (**Supplemental Figure S3B**). We found that non-TEGs or TEGs with dG4s had 14.6% and 83.1% higher enrichment of D-6mA, respectively, both upstream and downstream of the TSSs compared to the corresponding ones with udG4s (**Figure 5B**). We then wondered

whether the G4 formation is affected by DNA methylation. This was investigated by plotting normalized BG4-DNA-IP read density ± 1 kb around TEGs, ± 0.5 kb of the TTSs of non-TEGs, TEGs or non-TEGs with high/low DNA 5mC levels with similar CG% (**Supplemental Figure S4A, S4B**), and with/without D-6mA peaks containing similar AT% (**Supplemental Figure S4C, S4D**). We found that TEGs with low 5mC levels (**Figure 5C**) or with D-6mA peaks (**Figure 5D**) exhibited higher BG4-DNA-IP read density (140 and 118.5%, respectively) than the corresponding TEGs with high 5mC levels or without D-6mA peaks. Similarly, we found that non-TEGs with low 5mC levels (**Figure 5C**) or with D-6mA peaks (**Figure 5D**) exhibited higher BG4-DNA-IP read density (163.8 and 31.5%, respectively) than the corresponding non-TEGs with high 5mC levels or without D-6mA peaks. Thus, all above analyses indicate the intimate relationship between DNA methylation and G4 formation within TEGs/non-TEGs.

**Effects of G4s on expression levels of G4-overlapping genes**

We next decided to investigate the possible effects of G4s on the transcription of G4-containing genes. To this end, we compared the expression levels of TEGs and non-TEGs with both dG4s and udG4s using the published RNA-seq data (GSE33265) (Wu et al., 2011). We found that the mean expression levels of both TEGs and non-TEGs with dG4s were 3.6- and 5.5-fold higher than those with udG4s, respectively (**Supplemental Figure S5A**) and directly corresponded to dG4 intensity

(**Supplemental Figure S5B**). We divided TEGs and non-TEGs into 6 subgroups according to fragments per kilobase per million mapped fragments (FPKM) values, and found that the percentage of dG4-overlapping genes (TEGs and non-TEGs) was positively corresponded with the gene expression levels (**Supplemental Figure S5C**). As shown in Supplemental Figure S5D, genes with dG4s in the 5'UTR and downstream of the TTS exhibited the highest and lowest mean expression levels, respectively. Furthermore, we created a JBrowser (https://bioinfor.yzu.edu.cn/jbrowse2/bg4dnaip/) to visualize rice G4s for researchers who are interested in checking if there is a G4 in their target genes or validating the functions of promoter/gene body G4s in regulating expression of their target genes using CRISPR-mediated disruption of G4 structures. Therefore, these results suggest that dG4s have promotional roles in regulating the transcription of dG4-overlapping genes, and presumed contributions of dG4s in regulating gene transcription vary according to the location in or near different parts of gene structures.

Next, we further divided non-TEGs into 4 subgroups according to FPKM values and plotted their BG4-DNA-IP read density (RPKM) values, which represent G4 intensity around ± 1 kb of the TSSs (**Figure 6A**). Strikingly, the read intensity of BG4-DNA-IP at TSSs exhibited a positive association with the expression levels of dG4-containing genes, while that of BG4-DNA-IP in gene bodies was negatively corresponded with the expression levels of dG4-containing genes (**Figure 6A**). The heat map shown in Figure 6B further confirmed that BG4-DNA-IP reads were enriched

335 at the TSSs of highly expressed genes, while distributed over the whole gene length of
336 lowly or non-expressed genes. A similar trend was observed between the intensity of
337 udG4s in promoters/gene bodies and the expression levels of corresponding genes
338 (**Supplemental Figure S6A, S6B**).

339 To assess whether DNA sequences with higher G content have higher probability to
340 fold into G4s, we reassessed the relationship of BG4-DNA-IP read density between
341 dG4s and udG4s and the expression levels of the corresponding genes with similar G%.
342 As seen in Supplemental Figure S6C and S6D, this is indeed the case at promoters
343 (**Supplemental Figure S6C**), and no association was found among the different
344 subtypes of genes with udG4s (**Supplemental Figure S6D**). These results are in line
345 with what we observed for dG4s and udG4s at promoters and dG4s in gene bodies
346 (**Figure 6A; Supplemental Figure S6A**). Also, it has been reported that antisense
347 PQFSs hotspots are prevalent at the TSS, the 5' of intron 1 and around the mRNA AUG
348 in maize (Andorf et al., 2014). To assess positional distributions of antisense and sense
349 PQFSs around specific gene features, we plotted the total number of PQFSs that
350 overlap each base for the observed genes and simulated sequences across ± 1 kb of the
351 TSS, or ± 100 bp of the boundary of AUG or the 1$^{st}$ Exon-Intron. We found that
352 antisense PQFSs were highly enriched at TSSs (**Fig. 6C**), the boundary of the mRNA
353 AUG (**Fig. 6D**) and the 1$^{st}$ Exon-Intron (**Fig. 6E**) when compared with sense PQFSs
354 and simulated sequences. Thus, these analyses imply involvement of G4s in
355 transcription, splicing and translation in rice and maize.

Altogether, the above analyses indicate that dG4s have promotional and repressive roles in regulating expression of G4-overlapping genes; and each type of genic G4 may be involved in some distinct biological process.

**Potential interrelationship between epigenomic signatures and G4s**

It was then of interest to examine whether genomic regions at sequences detected by the BG4 assay are associated with any specific *in vivo* marks. To this end, we integrated dG4 data with existing omics data, including R-loop (DRIP-seq, GSE111944)(Fang et al., 2019), DHSs (DNase I hypersensitive sites, DNase-seq, GSE26734)(Zhang et al., 2012), ATAC-seq (GSE144564) (Liang et al., 2021) and 12 histone marks (**Supplemental Table S5**) (Zhang et al., 2012; Lu et al., 2015; Fang et al., 2016; Tan et al., 2016). We performed a fold enrichment assay by comparing each mark distributed in promoter, gene body, terminal and intergenic regions with dG4s relative to udG4s (**Figure 7A**). We observed distinct epigenomic mark enrichment among subgenomic dG4s: for example, R-loop and H3K4me3 were highly enriched in all subgenomic dG4s; dG4s at promoters were hypomethylated while those at terminal and intergenic regions had highly methylated CG and CHH (**Figure 7A**). These results indicate that dG4s corresponding to various subgenomic regions *in vivo* exhibit potential divergent enrichment of epigenomic marks.

To further confirm genomic regions capable of forming G4s detected by the BG4 assay associated with *in vivo* related marks, we conducted histone ChIP or DNase I

treatment followed by BG4-DNA-IP-qPCR assays (for G4s) and BG4-DNA-IP followed by S9.6-based DRIP-qPCR assay (for R-loops) (**Supplemental Figure S7)**. Positive loci with G4 overlapping the mark examined exhibited significant enrichment of marks examined, including H3K36me3 (**Figure 7B**), H3K4me3 (**Figure 7C**), H3K27me3 (**Figure 7D**), DH (**Figure 7E**) and R-loops (**Figure 7F**).

Collectively, these analyses imply that sequences capable of forming G4s may interact with distinct epigenomic signatures *in vivo*.

## Discussion

### BG4-DNA-IP-seq for *in vitro* G4 identification

Currently, high throughput methodologies for DNA G4 identification include polymerase stop assays (G4-seq) (Marsico et al., 2019), depending on stability and folding potentials of G4s; antibody or protein based ChIP-seq, D1-ChIP (Liu et al., 2016), BG4-ChIP (Hansel-Hertsch et al., 2020), and G4P-ChIP (Zheng et al., 2020). In addition, similar to G4RP-seq for RNA G4 identification (Yang et al., 2018), BioTASQ can be potentially used for identification of DNA G4s in the future. Those approaches depend on the specificity and binding efficiency of antibody or protein or the small molecule to G4s. Combinations of different approaches help unveil comprehensive landscapes of DNA G4s in the genome.

Our study showed that BG4 can be applied to identify rice G4s *in vitro*. All G4s detected *in vitro* (dG4s) are useful to provide references for *in vivo* investigations.

However, we were surprised that more than 95% of predicted G4s (PQFSs, also termed as udG4s) were undetectable by the BG4 assay(**Figure 3A**). The possible explanations for such a discrepancy could originate in the fact that dG4 overlapping PQFSs have shorter distance between two neighboring PQFSs within larger genomic regions as compared to dG4- overlapping PQFSs (**Figure 3C and 3D**); this suggests that K+PEG may not favour G4 formation in udG4. It has been reported that PQFSs with G3 L1-7 pattern form among the most stable G4 structures *in vitro* (Bugaut and Balasubramanian, 2008; Mullen et al., 2010); this popular motif has been widely used for *in silico* identification of potential canonical G4s across animal and plant genomes (Huppert and Balasubramanian, 2005; Andorf et al., 2014; Wang et al., 2015; Cagirici and Sen, 2020). We found that G3L1-12 subtypes were more enriched than G2L1-12 motifs (**Figure 2B**). Furthermore, those udG4s either only occur *in silico* or some of those PQFSs may preferentially form G4s in different cell types or tissues instead of seedlings as we examined. Tissue type G4 formation deserves to be further investigated.

Some of the PQFSs may form non-canonical G4s to which BG4 binds with a very low efficiency. Our study showed that 417 BG4-bound G4s had several types of non-canonical PQFS patterns (**Supplemental Table S4.xlsx**), such as the ($G3_{+}L1$-12)3+G3 pattern with at least one regular GGG run interrupted by a non-G bulge (106/417). A subset of non-canonical G4s have been detected *in vivo* or *in vitro* in both animals and plants, including bulged G4s (Mukundan and Phan, 2013; Chambers

et al., 2015; Kopylov et al., 2019), (3+1) hybrid G4s (Sengar et al., 2019), and G4s with a GGXO (G-G-8-oxoguanine-xanthine) tetrad (Cheong et al., 2015). Potential non-canonical G4 conformation thus needs to be further investigated, which will help to deepen the understanding of G4 biology.

It is also possible that some udG4s can form stable G4s but fail to be efficiently bound by the BG4 assay *in vitro*, which is the limitation of the use of BG4. *In vivo* study showed that approximately 15% of exon G4s can be detected by G4 antibody D1-related ChIP-seq, but undetected by BG4 ChIP-seq (Hansel-Hertsch et al., 2016; Liu et al., 2016), reflecting possible intrinsic variations in specificity of G4 antibodies for G4 detection. In addition, some PQFSs may preferentially form G4s in different conditions, such as PDS, $Na^+$ or other cations instead of $K^+$+PEG. $Na^+$ or $K^+$-specific G4 formation of synthesized oligoes has been detected in rice using circular dichroism (CD) spectroscopy (Garg et al., 2016).

**Regulatory roles of G4s in rice**

With this caveat in mind, we demonstrated that rice G4s do indeed play key regulatory roles in different biological processes (Yadav et al., 2017; Kim, 2019) such as transcription, splicing and translation in rice and maize (Andorf et al., 2014) (**Figure 6C to 6E**). Compared to *A. thaliana* (**Supplemental Figure S8A**) (Marsico et al., 2019), rice G4s display distinct genomic distributions, being more present at promoters (36.1 versus 21.7%) but less present in exons (16.5 versus 51.2%). These variations

indicate possible species-dependent G4 distributions and functions, which might be caused by differences in genome size or complexity.

**Dual roles of G4s in regulating gene transcription**

To go a step further, we showed that promoter G4s positively correspond with transcription levels of G4-containing genes in both rice and *A. thaliana* (**Supplemental Figure S8B**). This indicates a functional conservation of plant G4s at the promoter, which might be due to favorable spatial environments, which facilitate access or binding of trans-factors to promoters for transcriptional initiation, or to the fact that G4s can serve as docking sites for transcriptional factors (TFs) or G4-binding proteins, thereby indirectly recruiting trans-factors or transcriptional factors essential for gene transcription. Recent results lend credence to these hypotheses demonstrating that G4s can serve as common binding sites for binding of various TFs in human, especially for G4s found in the promoters of highly-expressed genes, in which TFs exhibited binding affinity to G4s comparable to typical double-stranded DNA *loci* (Spiegel et al., 2021).

Our study showed that the GGC-box motif within G4s can be recognized by ERF105 (Hao et al., 1998). ERF105 has been found to be involved in cold and high light response in *A. thaliana* (Vogel et al., 2014; Bolt et al., 2017) and resistance to *Exserohilum turcicum* in maize (Zang et al., 2020). BPC TFs have been reported to function in tissue-specific gene expression through binding promoters of target genes in *A. thaliana* (Simonini et al., 2012). Therefore, G4s may function in biotic and abiotic

responses or regulation of gene transcription through recruitment of some key TFs in plants. It has been reported that G4s can act as regulatory loci in multiple species (Du et al., 2009; Brooks and Hurley, 2010; Hansel-Hertsch et al., 2016), such as zinc-finger transcription factor with potential G4 binding in human (Kumar et al., 2011) and speculated to exist in maize (Andorf et al., 2014; Kopylov et al., 2015), rice and *A. thaliana* (Yadav et al., 2017; Griffin and Bass, 2018). G4s can thus function as cis elements to regulate transcription of genes that are involved in embryonic development (David et al., 2016).

In contrast, G4s in gene bodies negatively correspond with expression levels of G4-containing genes. Gene body G4s may create a steric hindrance that affects transcriptional elongation through blocking the movement of transcriptional machinery, or preventing access of elongation trans-factors. G4-related RNA polymerase II promoter-proximal transcriptional pausing has been reported in human genes (Eddy et al., 2011). Preferential enrichment of G4s in 1$^{st}$ exon-intron junctions (**Figure 6E**) may facilitate the occurrence of alternative splicing through blocking RNA polymerase II processivity. *In vitro* study showed that G4s formed in the non-transcribed DNA strand block the movement of mammalian RNA polymerase II (Sun and Hurley, 2010), suggesting that G4 structures act as physical obstacles to block the movement of polymerases during DNA transactions (Kim, 2019). G4s thus display apparently different roles, being alternatively activators (*e.g.*, recruiting TFs) or repressors (*e.g.*, blocking polymerase processivity). To date, the wealth of data

available on the prevalence of the G4 landscapes across multiple species provides precious insights into G4 biology, but does not allow for precisely understanding and/or predicting their cellular roles. The comparison between rice and *A. thaliana* described here is an example of such functional divergence: while it could be tempting to attribute this difference to a higher proportion of exonic G4s in *A. thaliana* (**Supplemental Figure S8A, B**, 51.2% versus 16.5% in rice), establishing such a functional link in a reliable manner will require further investigations. In addition, it is also possible that post-transcriptional mechanisms mediated by (pre)mRNA G4s could be involved in regulation of translation of related mRNAs. 5'UTR and CDS (coding sequence)-related G4s can function as a translational inhibitor of some mRNAs (Nie et al., 2015). RNA G4s generally suppress translation efficiency of mRNA from TEs (transposable elements) in *A. thaliana* (Yang et al., 2020).

**G4s in combination with epigenomic signatures may regulate expression of G4-overlapping genes**

We next focused on the relationship between G4s and epigenomic marks, with the hope of gaining precious insights into the roles that G4s may play in gene regulation *in vivo*. We found here that DNA-5mC methylation levels negatively correspond with G4 formation, while the DNA-6mA levels positively correspond with G4 formation in rice (**Figure 4**). These results are in line with the most recent epigenetic advances as G4s hamper proper C-methylation (by enzyme sequestration) (Mao et al., 2018) and are

favored at transcriptionally active sites (associated to A-methylation) (Hansel-Hertsch et al., 2016). Additionally, *C9ORF72* (C9) RNA G4 has been found to weakly bind Polycomb Repressive Complex 2 (PRC2), thereby modulating enrichment of H3K27me3 (Wang et al., 2019). We speculate that enrichment of H3K27me3 at or around dG4s at the gene body (**Figure 7A**) in our study is possibly mediated by recruitment of PRC2 through binding DNA G4. Moreover, G4s have been proposed to provide the docking site for recruitment of histone-modifiers (Reina and Cavalieri, 2020). Changes of H3K4me3 and H3K9/14ac at the promoter of B lymphocyte alloantigen (*BU-1*) have been reported to be associated with the presence of G4 in regulator of expression of virion protein 1(rev1)-deficient DT 40 cells (Schiavone et al., 2014). While we recognize that the detailed relationship between DNA methylation and G4s needs to be further studied in plants, we believe that our study contributes to decipher some of the G4 roles, notably through the demonstration that G4-forming sequences in rice interfere with specific histone marks *in vivo*.

## Materials and methods

### Plant materials

Seeds of rice (*Oryza. sativa* L.) cultivar Nipponbare (Japonica) were immerged in tap water at room temperature (RT) for three days. Uniformly germinated seeds were sowed into the nutrient soil and grown in a greenhouse with a controlled temperature of 28-30°C and a 14h/10h light-dark cycle. Two-week-old rice seedlings (the entire aerial

parts) were collected at 10:00 AM. The seedlings with 1-1.5 cm cut in length were cross-linked using 1%(v/v) formaldehyde in HEPES buffer pH=8.0 (20mM HEPES, 1mM EDTA, 100mM NaCl and 1mM PMSF) at 23-25°C for 10min under a vacuum condition. The excessive formaldehyde was quenched by adding 2M glycine to achieve a final concentration of 0.125 M under a vacuum condition for an additional 5min. The cross-linked seedlings were ground to a fine powder in liquid nitrogen. The ground powder can be immediately used for the preparation of genomic DNA, nuclei, and chromatin or stored at -80 °C for later use.

For zebularine treatment, a chemical for reduction of global DNA methylation through specifically inhibiting DNA methyltransferase in eukaryotic genomes (Zhou et al., 2002), 100 of 5-d-old rice seedlings were grown in 80 μM zebularine (Sigma-Aldrich Z4775) dissolved in DMSO or DMSO only (as control) for three days. The zebularine-treated or untreated seedlings were collected for extracting genomic DNA for dot blotting assays.

**Generation of the rice transgenic line**

The putative rice 6mA modifier gene, *LOC_Os01g16180* that encodes a homologous MTA-70 domain protein responsible for 6mA modification in *A. thaliana* (Luo et al., 2014), was selected for preparing a CRISPR-cas9 construct (**Supplemental Table S2**). The construct was introduced into calli induced from mature rice cultivar

544 “Nipponbare” seeds through AGL1 *Agrobacterium tumefaciens* infection.
545 Homozygous transgenic plants were grown in the field for collecting more seeds.
546
547 **Dot blotting assays**
548 The expression vector pSANG10-3F BG4 (plasmid no. 55756, Addgene) was used
549 for preparation of the recombinant BG4-FLAG-fused protein by CHEMGEN
550 BIOTECH (Shanghai). For the dot blotting assay, genomic DNA in G4-stabilizing
551 buffer (150mM KCl, 40% (w/v) PEG 200 and 10mM Tris-HCl, pH=7.5), including
552 DNA from WT or control, zebularine-treated seedlings and the 6mA-related transgenic
553 line, was denatured at 95°C for 5min. The Amersham Hybond-N+-nylon membrane
554 containing denatured DNA was pre-blocked in 5% (w/v) milk for 45min at RT, and
555 then incubated with the recombinant BG4 protein overnight at 4 °C followed by
556 incubation with anti-FLAG antibody for an additional 1.5h. The remaining steps were
557 the same as described before (Fang et al., 2019). Similar procedures were used for
558 detecting total levels of DNA-5mC and DNA-6mA by using anti-5mC (Aviva Systems
559 Biology, AMM99021) and anti-m6A (abcam, ab151230), respectively. Each blot was
560 repeated at least two times. We conducted a similar BG4 related dot blotting assay for
561 i-Motif using reported oligos (**Supplemental Table S2.xlsx**).
562
563 **Analysis of GC/AT content and GC/AT skew**

GC or AT content and GC/AT skew were calculated as follows: GC skew = (G - C)/(G + C), AT skew = (A - T)/(A + T); GC content = (C + G)/(A+T+C+G), AT content = (A + T)/(A+T+C+G), the ± 1 kb regions around dG4 peak centers were divided into 50 bp windows. Random regions were selected from the remaining genomic regions lacking G4s (command shuffleBed in bedtools, option -noOverlapping). Similar analyses were conducted for udG4s.

To acquire similar sequence content loci between two groups, a portion of random loci were first selected for one group, then loci with less than 5% sequence content variation were selected for the other group.

**Public data analyses**

mRNA-seq (mRNA sequencing, GSE33265), DNase-seq (DNase I sequencing, GSE26734), DRIP-seq (DNA-RNA immunoprecipitation followed by high-throughput DNA sequencing, GSE111944), BS-seq (bisulfite sequencing, SRP043447), histone ChIP-seq and DNA-6mA IP-seq (6mA immunoprecipitation sequencing, GSE103145) are published data (**Supplemental Table S5**). Unique reads with mapQ > 20 from each data were used for downstream assays.

**DNase-seq:** DNase-seq (Boyle et al., 2008) was used to identify DNase I hypersensitive sites (DHSs) with a 200 bp bandwidth for integrating with G4 data. The FDR was the ratio of DHSs identified from 10 random data sets divided by DHSs from DNase-seq. A cutoff in F-seq for controlling the FDR was <0.05.

**DRIP-seq:** DRIP-seq (DNA-RNA immunoprecipitation followed by high-throughput DNA sequencing) allows for high-resolution, genome-wide R-loops mapping (Sanz and Chedin, 2019) DRIP-seq data were here analyzed following our published pipelines (Fang et al., 2019). Briefly, W (Watson)- and C (Crick)-R-loop reads were sorted by flags of R2 reads equal to 163 and 147, respectively, for R-loop peak calling. MACS2 (Zhang et al., 2008) was used to call R-loop peaks. The command and parameters were as follows: macs2 callpeak --keep-dup 1 -g 3.8e+8 -B --extsize 200 -q 0.01 --nomodel --SPMR.

**BS-seq:** Bisulfite-sequencing (BS-seq) (Hu et al., 2014) was used here to further examine the relationship between DNA-5mC and G4s. To this end, bismark was used to map clean data to the rice reference genome (MSU7.0) (Krueger and Andrews, 2011). The bismark methylation_extractor program was used to calculate the methylated cytosines of total uniquely mappable reads. Each position with the total number of all (C+T) no less than 5 was included for estimating the DNA methylation levels. Each region was split into 10 bp windows to calculate the DNA methylation levels in different regions. The average DNA methylation ratios of all cytosine sites in each window were used to represent the DNA methylation levels.

**DNA-6mA IP-seq:** 6mA IP-seq data were analyzed following the published procedures (Zhou et al., 2018) to examine the interrelationship between DNA-6mA and G4s. MACS2 was used to call 6mA peaks by comparing the IP data with the input data.

**ChIP-seq and RNA-seq:** Both datasets were analyzed following the published procedures (Fang et al., 2019). MACS2 was used to call enriched peaks of each mark by comparing the IP data with the input data.

**BG4-DNA-IP-seq:** 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 (Burrows-Wheeler Aligner) (mem algorithm, version 0.7.17) with default parameters. Only reads with alignment length greater than 50 were used for G4 peak calling and further analyses. G4 Peak calling was performed using MACS (version 2.1.1) with the following command and parameters: macs2 callpeak -g 3.8e+8 -f BAM --extsize 200 -p 1e-5 –nomodel. The input, IgG and anti-FLAG IP-seq data were used as controls. G4 peaks shared among BG4-DNA-IP date relative to either of three controls were considered as G4 peaks with high confidence (command intersect of the bedtools package). The Spearman's rank correlation coefficient was calculated using plotCorrelation program of deepTools.

**BG4-DNA-IP-seq or -qPCR**

Preparation and purification of rice nuclei pellet were performed as described previously (Zhang et al., 2012). The purified nuclei were fragmented into sizes ranging from 100-500 bp in sonication buffer (50mM Tris-HCl, 10mM EDTA and 1% (w/v) SDS) using the water-based Biorupter (Diagnode) followed by genomic DNA

625 extraction and purification using phenol/chloroform extraction and cold ethanol
626 precipitation.

627 A total of 5μg fragmented genomic DNA was diluted in G4-stabilizing buffer
628 (150mM KCl, 40% (w/v) PEG 200 and 10mM Tris-HCl, pH=7.5), denatured at 95°C
629 for 5min, then reassociated by letting the temperature slowly drop down to RT. The
630 re-associated DNA was diluted with G4-IP incubation buffer (50mM HEPES, 150mM
631 KCl, 1mM $MgCl_2$, 130nM $CaCl_2$, 1% (w/v) BSA, 40% (w/v) PEG 200 and Complete
632 mini, pH=7.5). 3μg of BG4 protein was incubated for 4h, then 3μg of anti-FLAG
633 antibody (D110005, BBI) was added for an additional 4h at 4°C. 30 μl of washed
634 protein G Dynabeads (10004D, Invitrogen) was incubated for another 4h at 4°C. The
635 washed BG4-bound DNA was eluted with 200μl elution buffer (0.1M $NaHCO_3$ and 1%
636 (w/v) SDS) at 65°C for two times for 15min each. For the BG4-DNA-IP-seq assay,
637 two biologically replicated BG4-IPed DNA or one replicate of Input/IgG-/anti-FLAG
638 only-IPed control DNA was used for the library preparation for paired-end mode
639 sequencing on Illumina NovaSeq platform. All sequencing libraries were constructed
640 using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB, E7645S).

641 For the BG4-DNA-IP-qPCR, 1μl of input and IPed DNA (2ng/ μl each) was used as
642 DNA templates. The enrichment of IPed DNA was calculated using the $2^{(\Delta\Delta Ct)}$ method
643 and expressed as fold change over the corresponding input. Each primer set was
644 repeated three times in each qPCR. All primer sequences are listed in Supplemental
645 Table S2.xlsx.

**Histone ChIP or DNase I treatment followed by BG4-DNA-IP-qPCR assay**

H3K4me3 (07-473, Millipore), H3K27me3 (07-449, Millipore) and H3K36me3 (ab9050, abcam)-related ChIP were conducted as described previously (Zheng et al., 2019). Each ChIPed DNA was used for the BG4-DNA-IP-qPCR assay. The PCR primer sequences are listed in Supplemental Table S2.xlsx.

DNase I treatment was conducted as described previously (Zhang et al., 2012). DNase I-treated DNA fragments were used for the BG4-DNA-IP-qPCR assay. The PCR primer sequences are listed in Supplemental Table S2.xlsx.

**BG4-DNA-IP followed by DRIP-qPCR assay**

BG4-DNA-IP was conducted as described above. BG4-bound DNA was used for the S9.6-based DRIP-qPCR assay, which was conducted as described previously (Fang et al., 2019). The PCR primer sequences are listed in Supplemental Table S2.xlsx.

**PQFSs fold enrichment analyses**

The whole genome sequence was computationally scanned using fastaRegexFinder.py (https://github.com/dariober/bioinformatics-cafe/blob/master/fastaRegexFinder.py) (Fujimoto et al., 2020) for identifying putative G-quadruplex-forming sequences

(PQFSs). The subtypes of PQFSs were defined by loop length and G repeats following the published procedures (Marsico et al., 2019).

The fold enrichment of each subtype of PQFSs was calculated by comparing with random controls with the same size distribution as the G4 peaks across the genome (bedtools shuffle command, observed values divided by average of 1,000 randomization values).

**Identification of non-canonical PQFSs**

Non-canonical PQFSs within dG4 peaks were identified using fastaRegexFinder.py with a customized regular expression. For example, modified (G3+L1-12)3+G3 with 1-3 [gG]{3,} fragment changing to [gG]\w[gG], representing a non-G bulge in the middle of a G3 run, for identification of (G3+L1-12)3+G3-like pattern with a nucleotide variation in at least one regular GGG run. A total of 14 types of modified regular expressions were used for screening non-canonical PQFSs associated with dG4s (**Supplemental Table S4.xlsx**).

**Normalization of read counts**

The ± 1 kb upstream and downstream of G4 peaks was divided into 50 bp windows, and the G4 peaks were equally divided into 20 bins for normalization. The number of reads per sliding window was first divided by the window length and then by the number of all uniquely mappable reads within the genome (Mb). For all mapped

reads, their positions in the rice genome were used to determine the midpoint of the sequence fragment.

**Motif prediction**

DNA sequences from the center of a G4 peak spanning ± 100 bp were used for motif analyses. G4-related motifs were identified using MEME-ChIP (http://meme-suite.org/tools/meme-chip) (Machanick and Bailey, 2011) with parameters as options minimum width 5 and maximum width 20. The identified motifs were used to further screen *A. thaliana* databases to match putative TF-binding sites (Tomtom tool). Only the top two or three significantly enriched motifs with the highest *E*-values are listed in the text.

**Gene ontology (GO)**

GO term analysis of genic-related G4 overlapping genes was conducted using online tools in Agrigo v 2.0 (http://systemsbiology.cau.edu.cn/agriGOv2/ ) (Tian et al., 2017) with *Oryza sativa* annotation.

Statistical Analyses

Wilcoxon rank-sum test was performed in R environment using ggpubr package. Permutation test was performed using scipy package in python environment. One-way ANOVA analysis was conducted using GraphPad Prism 6. Variance analysis was

performed using GraphPad Prism6. One-way LSD test was performed in R environment using oneway.test function.

**Accession numbers**

The BG4-DNA-IP-seq data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE132775. Published data used in this study are listed in Supplemental Table S5.

**Supplemental data**

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Validation of BG4 antibody and comparison of G4s identified under different conditions.

**Supplemental Figure S2.** A and C content normalization and DNA methylation within normalized G4s.

**Supplemental Figure S3.** A or C normalization.

**Supplemental Figure S4.** Comparisons of GC and AT content with different methylation levels.

**Supplemental Figure S5.** Effects of d/udG4s on expression of G4-overlapping genes.

**Supplemental Figure S6.** Relationships between d/udG4s and transcription of d/udG4-overlapping genes.

**Supplemental Figure S7.** Schematic illustration of H3K4/27/36me3-related ChIP, DNase I treatment followed by BG4-DNA-IP-qPCR assay.

**Supplemental Figure S8.** Genomic distribution of OQs and relationship of OQs with expression of overlapping genes in *A. thaliana*.

**Supplemental Table S1.** Summary of sequencing data analyzed in this study.

**Supplemental Table S2.** Summary of primer information and PCR conditions for PCR or qPCR assays in this study. (xlsx)

**Supplemental Table S3.** Summary of G3 PQFSs with genomic coordinates for dG4 and udG4 in rice. (xlsx)

**Supplemental Table S4.** Summary of non-canonical PQFSs with genomic coordinates for dG4 in rice. (xlsx)

**Supplemental Table S5.** Summary of published data sets used in this study.

**Funding**

This research was supported by grants from the National Natural Science Foundation of China (32070561, U20A2030 and 31470443). The open funds of the Key Laboratory of Plant Functional Genomics of the Ministry of Education (ML202001). This work was also supported by the CNRS, the Agence Nationale de la Recherche (ANR-17-CE17-0010-01), and the European Union (PO FEDER-FSE Bourgogne 2014/2020 programs, FEDER n° BG0021532).

**Acknowledgements**

We thank the Bioinformatics Center, Nanjing Agricultural University for providing computing facilities for data processing and analyses.

## Figure Legends

**Figure 1 Global identification of G4s in rice seedlings using BG4-DNA-IP-seq. A**, Schematic representation of BG4-DNA-IP-seq workflow. **B**, Correlation analyses of two BG4-DNA-IP-seq reads in G4-favoring conditions (in the presence of $K^+$ and PEG). r stands for Spearman's rank correlation coefficient. **C**, Venn plots illustrating overlaps of BG4-DNA-IP-seq peaks relative to controls (input, IgG and anti-flag). **D**, Venn plots highlighting the 23,685 common BG4-DNA-IP-seq peaks. **E**, A representative Integrative Genomics Viewer (IGV) snapshot across a 113 kb window from rice chromosome 1 illustrating the reproducibility of BG4-DNA-IP-seq (in $K^+$+PEG conditions). Each G4 peak was marked with a solid rectangular blue box; green vertical bars indicate the presence of PQFSs.

**Figure 2 Validation of dG4s. A**, Comparison of dG4 overlapping PQFS with predicted G4s: *x*-axis represents PQFS patterns, for instance, $G_{2+}L_{1-12}$ representing PQFS containing two or more tetrads with loop length up to 12 nucleotides; $G_{3+}L_{1-7}$ representing PQFS containing at least three tetrads with loop length up to 7 nucleotides; $G_{3+}L_{8-12}$ representing extended PQFS containing at least three tetrads with at least one

loop with length from 8 to 12 nucleotides, and so on; *y*-axis represents the number of dG4 overlapping PQFSs (blue) *versus* predicted PQFSs (green). **B**, Fold enrichment of dG4 overlapping PQFS over random sequences: *x*-axis represents PQFS patterns; *y*-axis represents fold enrichment (observed values divided by average of 1,000 randomization values). Permutation test was conducted to determine the significance of the difference when the observed value was >90% or <90% of the permutation value. **C**, BG4-DNA-IP-qPCR assay for ten positive G4 *loci* (P1-P10) and one non-G4 control (N) (primers listed in Supplemental Table S2). Fold change indicated the enrichment level of each positive locus over the negative locus. Significance was determined by one-way ANOVA analysis, *** p-value < 0.01, ** p-value < 0.05. **D**, Reexamination of nine G4 *loci* (Os4-7, Os9-11, Os13 and Os18) and one non-G4 *locus* (Os3), previously validated by circular dichroism (CD) spectroscopy in the presence of $K^+$ in *Oryza sativa*.

**Figure 3 Characterization of dG4s and udG4s. A**, Venn diagram showing predicted PQFSs *versus* experimentally (dG4 peaks) and non-experimentally detected G4s (udG4s). **B**, Sub-genomic distributions of both dG4s-(PQFS) and udG4s-(PQFS). **C**, Mean length of both dG4s-(PQFS) and udG4s-(PQFS). **D**, Distance between two neighboring PQFSs for both dG4s-(PQFS) and udG4s-(PQFS). **E**, Strand-specific GC contents, GC and AT skews calculated around ± 1 kb from the center of dG4s (red), udG4s (black) and random peaks (grey). **F**, Motif discovery using MEME for dG4s;

794 only the top three significantly enriched G4 motifs are listed. Significance in C and D
795 was determined using Wilcoxon rank-sum test. ** p-value < 0.05.
796

797 **Figure 4 Interrelation between G4s and DNA methylation. A**, CG, CHG, and CHH
798 methylation levels of dG4s (*n* = 23,685) and udG4s (*n* = 23,685) with similar C content.
799 The heat map at the top indicates Wilcoxon rank-sum test and the color key represents
800 the p-values. **B**, Distribution of BG4-DNA-IP read density (Reads Per Kilobase per
801 Million mapped reads, RPKM) of dG4s and udG4 with 6mA (6mA+, *n* = 4,399, blue),
802 without 6mA (6mA-, *n* = 27,826, red) around ± 1 kb of G4s or PQFS regions. The heat
803 map at the top indicates Wilcoxon rank-sum test and the color key represents the
804 p-values **C**, Distribution of RPKM of 6mA across ± 1 kb of the center of dG4s and
805 udG4s with similar A content. The heat map at the top indicates Wilcoxon rank-sum
806 test and the color key represents the p-values. **D**, Dot blot assays for anti-5mC,
807 anti-6mA and BG4 antibodies, as indicated (n = 2); significance was determined using
808 variance analysis. *** p-value < 0.001, ** p-value < 0.01 and * p-value < 0.05. **E**, MA
809 plot on the left illustrating CK and Zebu hypoDNA biased dG4s identified from
810 BG4-DNA-IP-seq using DNA from CK and zebularine treated sample as the ones we
811 used in the dot blotting assay; box plot on the right showing the p-value of Zebu
812 hypoDNA biased dG4s is significantly lower than that of CK biased dG4s. Significance
813 was determined using Wilcoxon rank-sum test. ** p-value < 0.01.
814

**Figure 5 Relation between G4s and DNA methylation of dG4s in transposable element genes (TEGs) and non-TEGs. A**, CG, CHG, and CHH methylation levels in TEGs and non-TEGs associated with dG4s (red) and udG4s (blue) with similar C content. The heat map at the top indicates Wilcoxon rank-sum test and the color key represents the p-values. **B**, D-6mA methylation levels in non-TEGs (top panel) and TEGs (bottom panel) associated with dG4s (red) and udG4s (blue). Read density (RPKM) of DNA 6mA was plotted across 1 kb upstream of TSSs to 1 kb downstream of TTSs of genes with dG4s and udG4s. The heat map at the top indicates Wilcoxon rank-sum test and the color key represents the p-values. **C**, Plots showing the profile of BG4-DNA-IP reads ±1 kb around TEGs (left) and from 0.5 kb upstream of TSSs to 0.5 kb downstream of TSSs of non-TEGs (right) with high and low 5mC methylation levels (n = 500), as indicated. **D**, Plots showing the profile of BG4-DNA-IP reads ±1 kb around TEGs (left) and from 0.5 kb upstream of TSSs to 0.5 kb downstream of TSSs of non-TEGs (right) with or without D-6mA peaks (n = 500), as indicated. Significance in c and d was determined using Wilcoxon rank-sum test. ** p-value < 0.05.

**Figure 6 Relationships between dG4s and transcription of dG4-containing genes. A**, Plots showing the profile of read density (RPKM) of BG4-DNA-IP from 1 kb upstream of TSSs to 1 kb downstream of TTSs of dG4s in non-TEGs with different Fragments Per Kilobase per Million mapped fragments (FPKM) values (high, medium, low and non-expressed). **B**, The heat map showing distribution of RPKM of

BG4-DNA-IP from 1 kb upstream of TSSs to 1 kb downstream of TTSs of dG4-containing genes, ranked according to their FPKM. **(C, D and E)** Histograms showing the abundance of antisense (blue) and sense (red) PQFSs at specific positions plotted for the observed gene sets or randomized sequences (expected) for the TSS (**C**), the mRNA AUG (**D**) and the 1st Exon-Intron boundary (**E**).

**Figure 7 Epigenomic signatures of dG4-containing genes. A**, Heat map showing enrichment of each mark for dG4s relative to udG4s in each indicated subgenomic region, the color key represents the fold enrichment of dG4s relative to udG4s. Permutation test was conducted to determine the significance of the difference when the observed value was >90% or <90% of the permutation value. **(B, C and D)** H3K36me3-ChIP (**B**), H3K4me3-ChIP (**C**) and H3K27me3-ChIP (**D**) coupled with BG4-DNA-IP-qPCR for five positive loci (with mark and G4) and one control (N, with G4 but without mark). **E**, DNase I treatment coupled with the BG4-DNA-IP-qPCR assay for five positive loci (D1-D5, with DNase I hypersensitive site (DHS) and G4) and one control (N, with DHS but without G4). **F**, BG4-DNA-IP coupled with S9.6-based DNA-RNA immunoprecipitation (DRIP-qPCR) for six positive loci (RL1-RL6, with R-loop and G4) and one control (N, with R-loop but without G4). The fold change in each qPCR run (n = 2) is expressed relative to the control. Primer sequences for qPCR are listed in Supplemental Table S2. *** p-value < 0.001, ** p-value < 0.01 and * p-value < 0.05 determined by variance analysis.

858 ***Conflict of interest***. None declared

859

860 **References**


861 **Andorf CM, Kopylov M, Dobbs D, Koch KE, Stroupe ME, Lawrence CJ, Bass**
862 **HW** (2014) G-quadruplex (G4) motifs in the maize (*Zea mays* L.) genome are
863 enriched at specific locations in thousands of genes coupled to energy status,
864 hypoxia, low sugar, and nutrient deprivation. J Genet Genomics **41:** 627-647

865 **Bao HL, Liu HS, Xu Y** (2019) Hybrid-type and two-tetrad antiparallel telomere
866 DNA G-quadruplex structures in living human cells. Nucleic Acids Res **47:**
867 4940-4947

868 **Biffi G, Tannahill D, McCafferty J, Balasubramanian S** (2013) Quantitative
869 visualization of DNA G-quadruplex structures in human cells. Nat Chem **5:**
870 182-186

871 **Bolt S, Zuther E, Zintl S, Hincha DK, Schmulling T** (2017) ERF105 is a
872 transcription factor gene of *Arabidopsis thaliana* required for freezing tolerance
873 and cold acclimation. Plant Cell Environ **40:** 108-120

874 **Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS,**
875 **Crawford GE** (2008) High-resolution mapping and characterization of open
876 chromatin across the genome. Cell **132:** 311-322

877 **Brooks TA, Hurley LH** (2010) Targeting MYC expression through G-quadruplexes.
878 Genes Cancer **1:** 641-649

**Bugaut A, Balasubramanian S** (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. Biochemistry **47:** 689-697

**Cagirici HB, Budak H, Sen TZ** (2021) Genome-wide discovery of G-quadruplexes in barley. Sci Rep **11:** 7876

**Cagirici HB, Sen TZ** (2020) Genome-wide discovery of G-quadruplexes in wheat: distribution and putative functional roles. G3 **10:** 2021-2032

**Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S** (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat Biotechnol **33:** 877-881

**Cheong VV, Heddi B, Lech CJ, Phan AT** (2015) Xanthine and 8-oxoguanine in G-quadruplexes: formation of a G.G.X.O tetrad. Nucleic Acids Res **43:** 10506-10514

**Cho H, Cho HS, Nam H, Jo H, Yoon J, Park C, Dang TVT, Kim E, Jeong J, Park S, Wallner ES, Youn H, Park J, Jeon J, Ryu H, Greb T, Choi K, Lee Y, Jang SK, Ban C, Hwang I** (2018) Translational control of phloem development by RNA G-quadruplex-JULGI determines plant sink strength. Nat Plants **4:** 376-390

**David AP, Margarit E, Domizi P, Banchio C, Armas P, Calcaterra NB** (2016) G-quadruplexes as novel cis-elements controlling transcription during embryonic development. Nucleic Acids Res **44:** 4163-4173

**Du Z, Zhao Y, Li N** (2009) Genome-wide colonization of gene regulatory elements by G4 DNA motifs. Nucleic Acids Res **37:** 6784-6798

**Eddy J, Vallur AC, Varma S, Liu H, Reinhold WC, Pommier Y, Maizels N** (2011) G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. Nucleic Acids Res **39:** 4975-4983

**Fang Y, Chen L, Lin K, Feng Y, Zhang P, Pan X, Sanders J, Wu Y, Wang XE, Su Z, Chen C, Wei H, Zhang W** (2019) Characterization of functional relationships of R-loops with gene transcription and epigenetic modifications in rice. Genome Res **29:** 1287-1297

**Fang Y, Wang X, Wang L, Pan X, Xiao J, Wang XE, Wu Y, Zhang W** (2016) Functional characterization of open chromatin in bidirectional promoters of rice. Sci Rep **6:** 32088

**Fleming AM, Zhu J, Ding Y, Burrows CJ** (2019) Location dependence of the transcriptional response of a potential G-quadruplex in gene promoters under oxidative stress. Nucleic Acids Res **47:** 5049-5060

**Fujimoto A, Fujita M, Hasegawa T, Wong JH, Maejima K, Oku-Sasaki A, Nakano K, Shiraishi Y, Miyano S, Yamamoto G, Akagi K, Imoto S, Nakagawa H** (2020) Comprehensive analysis of indels in whole-genome microsatellite regions and microsatellite instability across 21 cancer types. Genome Res **30:** 334-346

**Garg R, Aggarwal J, Thakkar B** (2016) Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. Sci Rep **6:** 28211

**Ge F, Wang Y, Li H, Zhang R, Wang X, Li Q, Liang Z, Yang L** (2019) Plant-GQ: an integrative database of G-quadruplex in plant. J Comput Biol **26:** 1013-1019

**Gehring K, Leroy JL, Gueron M** (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. Nature **363:** 561-565

**Gellert M, Lipsett MN, Davies DR** (1962) Helix formation by guanylic acid. Proc Natl Acad Sci U S A **48:** 2013-2018

**Griffin BD, Bass HW** (2018) Review: Plant G-quadruplex (G4) motifs in DNA and RNA; abundant, intriguing sequences of unknown function. Plant Sci **269:** 143-147

**Hansel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M, Tannahill D, Balasubramanian S** (2016) G-quadruplex structures mark human regulatory chromatin. Nat Genet **48:** 1267-1272

**Hansel-Hertsch R, Simeone A, Shea A, Hui WWI, Zyner KG, Marsico G, Rueda OM, Bruna A, Martin A, Zhang X, Adhikari S, Tannahill D, Caldas C, Balasubramanian S** (2020) Landscape of G-quadruplex DNA structural regions in breast cancer. Nat Genet **52:** 878-883

**Hansel-Hertsch R, Spiegel J, Marsico G, Tannahill D, Balasubramanian S** (2018) Genome-wide mapping of endogenous G-quadruplex DNA structures by

chromatin immunoprecipitation and high-throughput sequencing. Nat Protoc **13:** 551-564

**Hao D, Ohme-Takagi M, Sarai A** (1998) Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J Biol Chem **273:** 26857-26861

**Hu L, Li N, Xu C, Zhong S, Lin X, Yang J, Zhou T, Yuliang A, Wu Y, Chen YR, Cao X, Zemach A, Rustgi S, von Wettstein D, Liu B** (2014) Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. Proc Natl Acad Sci U S A **111:** 10642-10647

**Huppert JL, Balasubramanian S** (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res **33:** 2908-2916

**Kang SG, Henderson E** (2002) Identification of non-telomeric G4-DNA binding proteins in human, *E. coli*, yeast, and *Arabidopsis*. Mol Cells **14:** 404-410

**Kim N** (2019) The interplay between G-quadruplex and transcription. Curr Med Chem **26:** 2898-2917

**Kopylov M, Bass HW, Stroupe ME** (2015) The maize (*Zea mays* L.) nucleoside diphosphate kinase1 (*ZmNDPK1*) gene encodes a human NM23-H2 homologue that binds and stabilizes G-quadruplex DNA. Biochemistry **54:** 1743-1757

**Kopylov M, Jackson TM, Stroupe ME** (2019) Bulged and canonical G-quadruplex conformations determine NDPK binding specificity. Molecules **24:** 1988

**Krueger F, Andrews SR** (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **27:** 1571-1572

**Kumar P, Yadav VK, Baral A, Kumar P, Saha D, Chowdhury S** (2011) Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. Nucleic Acids Res **39:** 8005-8016

**Kwok CK, Ding Y, Shahid S, Assmann SM, Bevilacqua PC** (2015) A stable RNA G-quadruplex within the 5'-UTR of *Arabidopsis thaliana* ATR mRNA inhibits translation. Biochem J **467:** 91-102

**Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S** (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods **13:** 841-844

**Lexa M, Kejnovsky E, Steflova P, Konvalinova H, Vorlickova M, Vyskot B** (2014) Quadruplex-forming sequences occupy discrete regions inside plant LTR retrotransposons. Nucleic Acids Res **42:** 968-978

**Liang Z, Zhang Q, Ji C, Hu G, Zhang P, Wang Y, Yang L, Gu X** (2021) Reorganization of the 3D chromatin architecture of rice genomes during heat stress. BMC Biol **19:** 53

**Liu HY, Zhao Q, Zhang TP, Wu Y, Xiong YX, Wang SK, Ge YL, He JH, Lv P, Ou TM, Tan JH, Li D, Gu LQ, Ren J, Zhao Y, Huang ZS** (2016) Conformation

selective antibody enables genome profiling and leads to discovery of parallel G-quadruplex in human telomeres. Cell Chem Biol **23:** 1261-1270

**Lu L, Chen X, Sanders D, Qian S, Zhong X** (2015) High-resolution mapping of H4K16 and H3K23 acetylation reveals conserved and unique distribution patterns in *Arabidopsis* and rice. Epigenetics **10:** 1044-1053

**Luo GZ, MacQueen A, Zheng G, Duan H, Dore LC, Lu Z, Liu J, Chen K, Jia G, Bergelson J, He C** (2014) Unique features of the m6A methylome in *Arabidopsis thaliana*. Nat Commun **5:** 5630

**Machanick P, Bailey TL** (2011) MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics **27:** 1696-1697

**Mao SQ, Ghanbarian AT, Spiegel J, Martinez Cuesta S, Beraldi D, Di Antonio M, Marsico G, Hansel-Hertsch R, Tannahill D, Balasubramanian S** (2018) DNA G-quadruplex structures mold the DNA methylome. Nat Struct Mol Biol **25:** 951-957

**Marsico G, Chambers VS, Sahakyan AB, McCauley P, Boutell JM, Di Antonio M, Balasubramanian S** (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. Nucleic Acids Res **47:** 3862-3874

**Mishra SK, Tawani A, Mishra A, Kumar A** (2016) G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. Sci Rep **6:** 38144

**Mukundan VT, Phan AT** (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. J Am Chem Soc **135:** 5017-5028

**Mullen MA, Olson KJ, Dallaire P, Major F, Assmann SM, Bevilacqua PC** (2010) RNA G-Quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. Nucleic Acids Res **38:** 8149-8163

**Nie J, Jiang M, Zhang X, Tang H, Jin H, Huang X, Yuan B, Zhang C, Lai JC, Nagamine Y, Pan D, Wang W, Yang Z** (2015) Post-transcriptional regulation of Nkx2-5 by RHAU in heart development. Cell Rep **13:** 723-732

**Reina C, Cavalieri V** (2020) Epigenetic modulation of chromatin states and gene expression by G-quadruplex structures. Int J Mol Sci **21:** 4172

**Renard I, Grandmougin M, Roux A, Yang SY, Lejault P, Pirrotta M, Wong JMY, 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

**Sanz LA, Chedin F** (2019) High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc **14:** 1734-1755

**Schiavone D, Guilbaud G, Murat P, Papadopoulou C, Sarkies P, Prioleau MN, Balasubramanian S, Sale JE** (2014) Determinants of G-quadruplex-induced epigenetic instability in REV1-deficient cells. EMBO J **33:** 2507-2520

**Sen D, Gilbert W** (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature **334:** 364-366

**Sengar A, Vandana JJ, Chambers VS, Di Antonio M, Winnerdy FR, Balasubramanian S, Phan AT** (2019) Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter. Nucleic Acids Res **47:** 1564-1572

**Simonini S, Roig-Villanova I, Gregis V, Colombo B, Colombo L, Kater MM** (2012) Basic pentacysteine proteins mediate MADS domain complex binding to the DNA for tissue-specific expression of target genes in *Arabidopsis*. Plant Cell **24:** 4163-4172

**Sparks MA, Singh SP, Burgers PM, Galletto R** (2019) Complementary roles of Pif1 helicase and single stranded DNA binding proteins in stimulating DNA replication through G-quadruplexes. Nucleic Acids Res **47:** 8595-8605

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

**Spiegel J, Cuesta SM, Adhikari S, Hansel-Hertsch R, Tannahill D, Balasubramanian S** (2021) G-quadruplexes are transcription factor binding hubs in human chromatin. Genome Biol **22:** 117

**Sun D, Hurley LH** (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. Methods Mol Biol **608:** 65-79

**Sundquist WI, Klug A** (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature **342:** 825-829

**Takahashi H, Nakagawa A, Kojima S, Takahashi A, Cha BY, Woo JT, Nagai K, Machida Y, Machida C** (2012) Discovery of novel rules for G-quadruplex-forming sequences in plants by using bioinformatics methods. J Biosci Bioeng **114:** 570-575

**Tan F, Zhou C, Zhou Q, Zhou S, Yang W, Zhao Y, Li G, Zhou DX** (2016) Analysis of chromatin regulators reveals specific features of rice DNA methylation pathways. Plant Physiol **171:** 2041-2054

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

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

**Vogel MO, Moore M, Konig K, Pecher P, Alsharafa K, Lee J, Dietz KJ** (2014) Fast retrograde signaling in response to high light involves metabolite export, MITOGEN-ACTIVATED PROTEIN KINASE6, and AP2/ERF transcription factors in *Arabidopsis*. Plant Cell **26:** 1151-1165

**Wallgren M, Mohammad JB, Yan KP, Pourbozorgi-Langroudi P, Ebrahimi M, Sabouri N** (2016) G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures in vitro and are unwound by the Pfh1 DNA helicase. Nucleic Acids Res **44:** 6213-6231

**Wang X, Goodrich KJ, Conlon EG, Gao J, Erbse AH, Manley JL, Cech TR** (2019) C9orf72 and triplet repeat disorder RNAs: G-quadruplex formation, binding to PRC2 and implications for disease mechanisms. RNA **25:** 935-947

**Wang Y, Zhao M, Zhang Q, Zhu GF, Li FF, Du LF** (2015) Genomic distribution and possible functional roles of putative G-quadruplex motifs in two subspecies of Oryza sativa. Comput Biol Chem **56:** 122-130

**Wu WQ, Zhang ML, Song CP** (2020) A comprehensive evaluation of a typical plant telomeric G-quadruplex (G4) DNA reveals the dynamics of G4 formation, rearrangement, and unfolding. J Biol Chem **295:** 5461-5469

**Wu Y, Kikuchi S, Yan H, Zhang W, Rosenbaum H, Iniguez AL, Jiang J** (2011) Euchromatic subdomains in rice centromeres are associated with genes and transcription. Plant Cell **23:** 4054-4064

**Yadav V, Hemansi, Kim N, Tuteja N, Yadav P** (2017) G-quadruplex in plants: A ubiquitous regulatory element and its biological relevance. Front Plant Sci **8:** 1163

**Yang SY, Lejault P, Chevrier S, Boidot R, Robertson AG, Wong JMY, Monchaud D** (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. Nat Commun **9:** 4730

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

**Zang Z, Lv Y, Liu S, Yang W, Ci J, Ren X, Wang Z, Wu H, Ma W, Jiang L, Yang W** (2020) A novel ERF Transcription Factor, *ZmERF105*, positively regulates maize resistance to exserohilum turcicum. Front Plant Sci **11:** 850

**Zeraati M, Langley DB, Schofield P, Moye AL, Rouet R, Hughes WE, Bryan TM, Dinger ME, Christ D** (2018) I-motif DNA structures are formed in the nuclei of human cells. Nat Chem **10:** 631-637

**Zhang S, Sun H, Wang L, Liu Y, Chen H, Li Q, Guan A, Liu M, Tang Y** (2018) Real-time monitoring of DNA G-quadruplexes in living cells with a small-molecule fluorescent probe. Nucleic Acids Res **46:** 7522-7532

**Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J** (2012) High-resolution mapping of open chromatin in the rice genome. Genome Res **22:** 151-162

**Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS** (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol **9:** R137

**Zheng D, Wang L, Chen L, Pan X, Lin K, Fang Y, Wang XE, Zhang W** (2019) Salt-responsive genes are differentially regulated at the chromatin levels between seedlings and roots in rice. Plant Cell Physiol **60:** 1790-1803

**Zheng KW, Zhang JY, He YD, Gong JY, Wen CJ, Chen JN, Hao YH, Zhao Y, Tan Z** (2020) Detection of genomic G-quadruplexes in living cells using a small artificial protein. Nucleic Acids Res **48:** 11706-11720

**Zhou C, Wang C, Liu H, Zhou Q, Liu Q, Guo Y, Peng T, Song J, Zhang J, Chen L, Zhao Y, Zeng Z, Zhou DX** (2018) Identification and analysis of adenine N(6)-methylation sites in the rice genome. Nat Plants **4:** 554-563

**Zhou L, Cheng X, Connolly BA, Dickman MJ, Hurd PJ, Hornby DP** (2002) Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. J Mol Biol **321:** 591-599

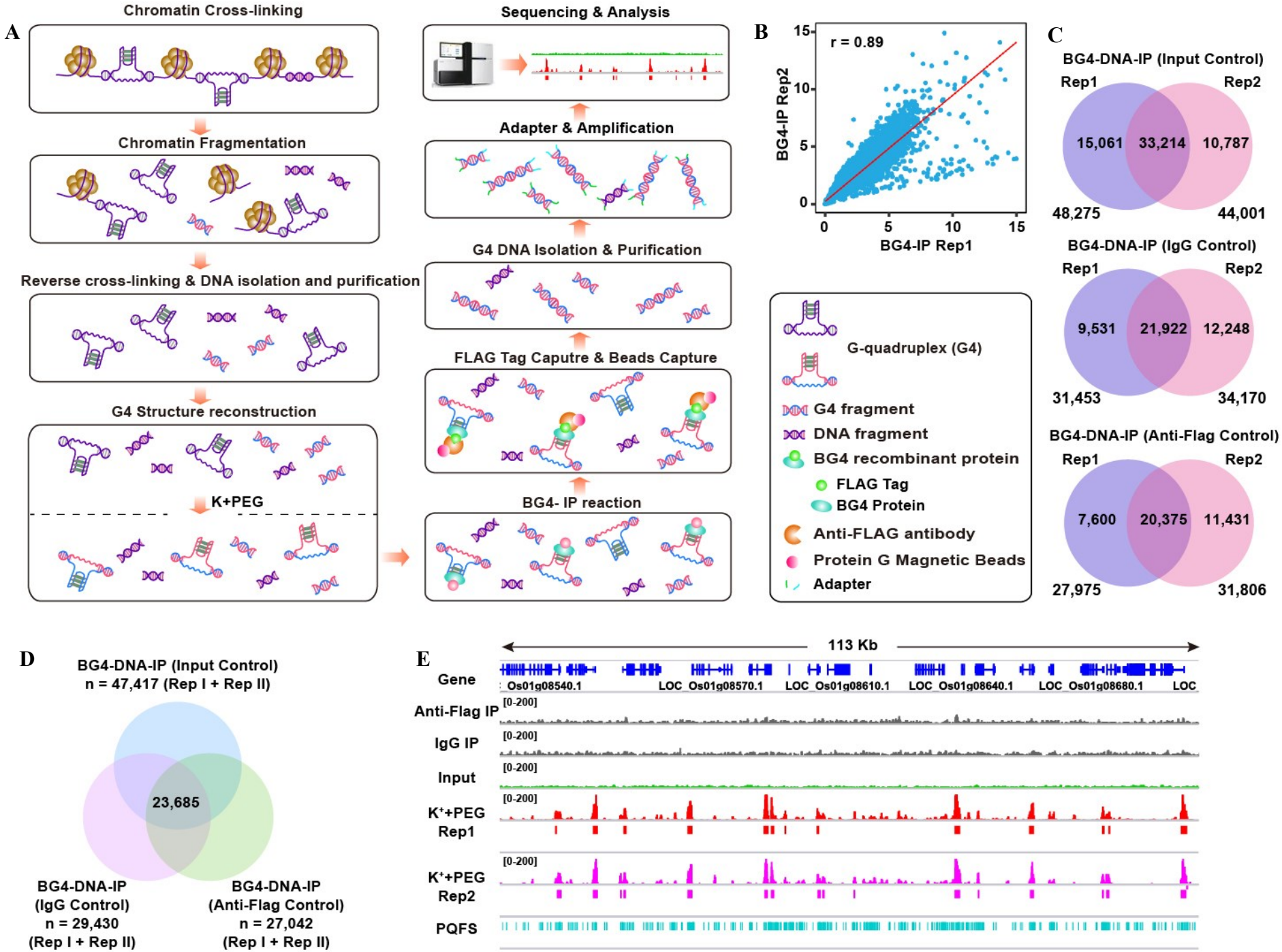


**Figure 1 Global identification of G4s in rice seedlings using BG4-DNA-IP-seq.** **A**, Schematic representation of BG4-DNA-IP-seq workflow. **B**, Correlation analyses of two BG4-DNA-IP-seq reads in G4-favoring conditions (in the presence of $K^+$ and PEG). r stands for Spearman's rank correlation coefficient. **C**, Venn plots illustrating overlaps of BG4-DNA-IP-seq peaks relative to controls (input, IgG and anti-flag). **D**, Venn plots highlighting the 23,685 common BG4-DNA-IP-seq peaks. **E**, A representative Integrative Genomics Viewer (IGV) snapshot across a 113 kb window from rice chromosome 1 illustrating the reproducibility of BG4-DNA-IP-seq (in $K^+$+PEG conditions). Each G4 peak was marked with a solid rectangular blue box; green vertical bars indicate the presence of PQFSs.

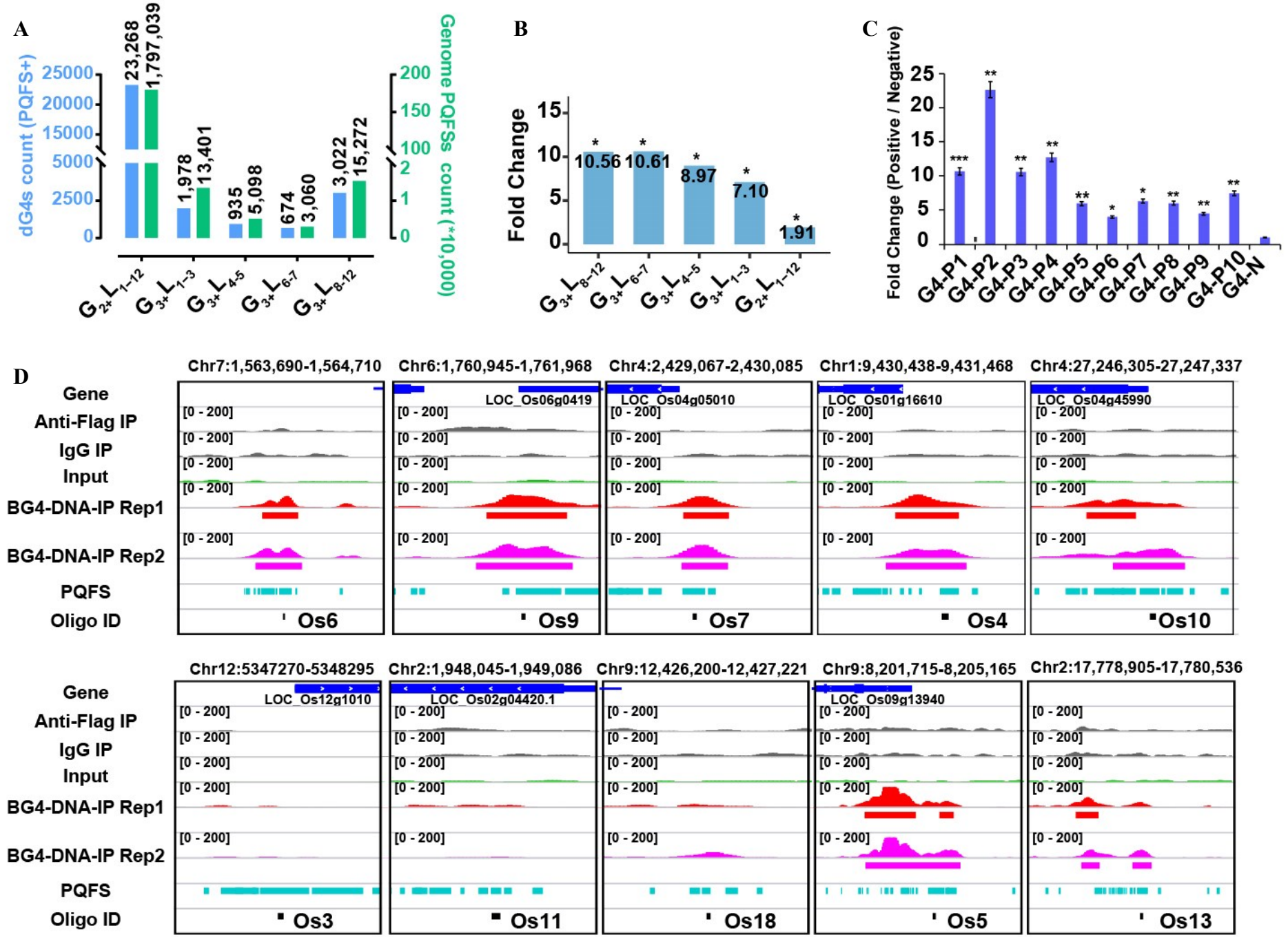


**Figure 2 Validation of dG4s. A**, Comparison of dG4 overlapping PQFS with predicted G4s: *x*-axis represents PQFS patterns, for instance, $G_{2+}L_{1-12}$ representing PQFS containing two or more tetrads with loop length up to 12 nucleotides; $G_{3+}L_{1-7}$ representing PQFS containing at least three tetrads with loop length up to 7 nucleotides; $G_{3+}L_{8-12}$ representing extended PQFS containing at least three tetrads with at least one loop with length from 8 to 12 nucleotides, and so on; *y*-axis represents the number of dG4 overlapping PQFSs (blue) *versus* predicted PQFSs (green). **B**, Fold enrichment of dG4 overlapping PQFS over random sequences: *x*-axis represents PQFS patterns; *y*-axis represents fold enrichment (observed values divided by average of 1,000 randomization values). Permutation test was conducted to determine the significance of the difference when the observed value was >90% or <90% of the permutation value. **C** , BG4-DNA-IP-qPCR assay for ten positive G4 *loci* (P1-P10) and one non-G4 control (N) (primers listed in Supplemental Table S2). Fold change with mean ± SD indicated the enrichment level of each positive locus over the negative locus. Significance was determined by one-way ANOVA analysis, *** p-value < 0.01, ** p-value < 0.05. **D**, Reexamination of nine G4 *loci* (Os4-7, Os9-11, Os13 and Os18) and one non-G4 *locus* (Os3), previously validated by circular dichroism (CD) spectroscopy in the presence of $K^+$ in *Oryza sativa*.

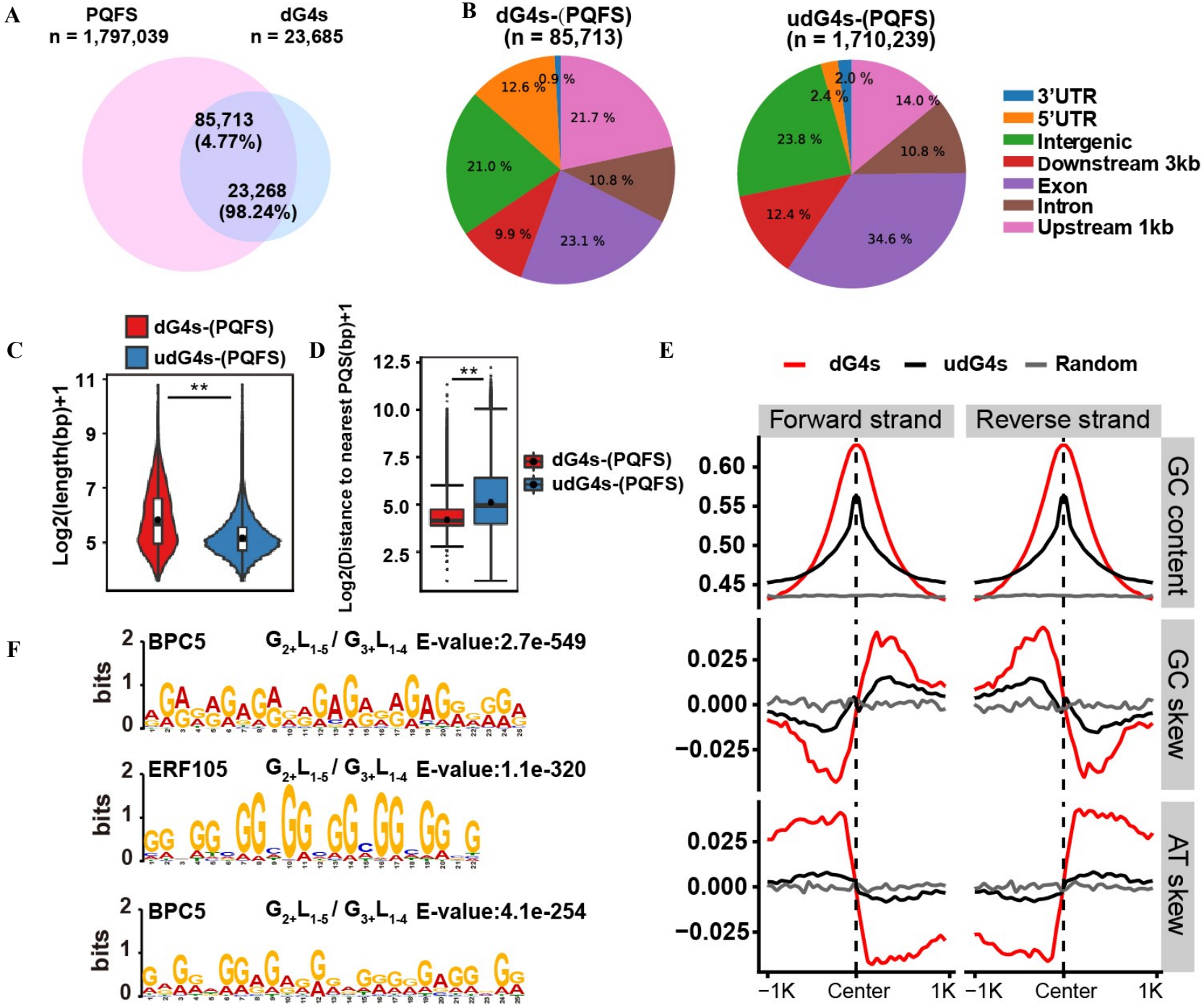


**Figure 3 Characterization of dG4s and udG4s. A**, Venn diagram showing predicted PQFSs *versus* experimentally (dG4 peaks) and non-experimentally detected G4s (udG4s). **B**, Sub-genomic distributions of both dG4s-(PQFS) and udG4s-(PQFS). **C**, Mean length of both dG4s-(PQFS) and udG4s-(PQFS). **D**, Distance between two neighboring PQFSs for both dG4s-(PQFS) and udG4s-(PQFS). Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. **E**, Strand-specific GC contents, GC and AT skews calculated around ± 1 kb from center of dG4s (red), udG4s (black) and random peaks (grey). **F**, Motif discovery using MEME for dG4s; only top three significantly enriched G4 motifs are listed. Significance in C and D was determined using Wilcoxon rank-sum test. ** p-value < 0.05.

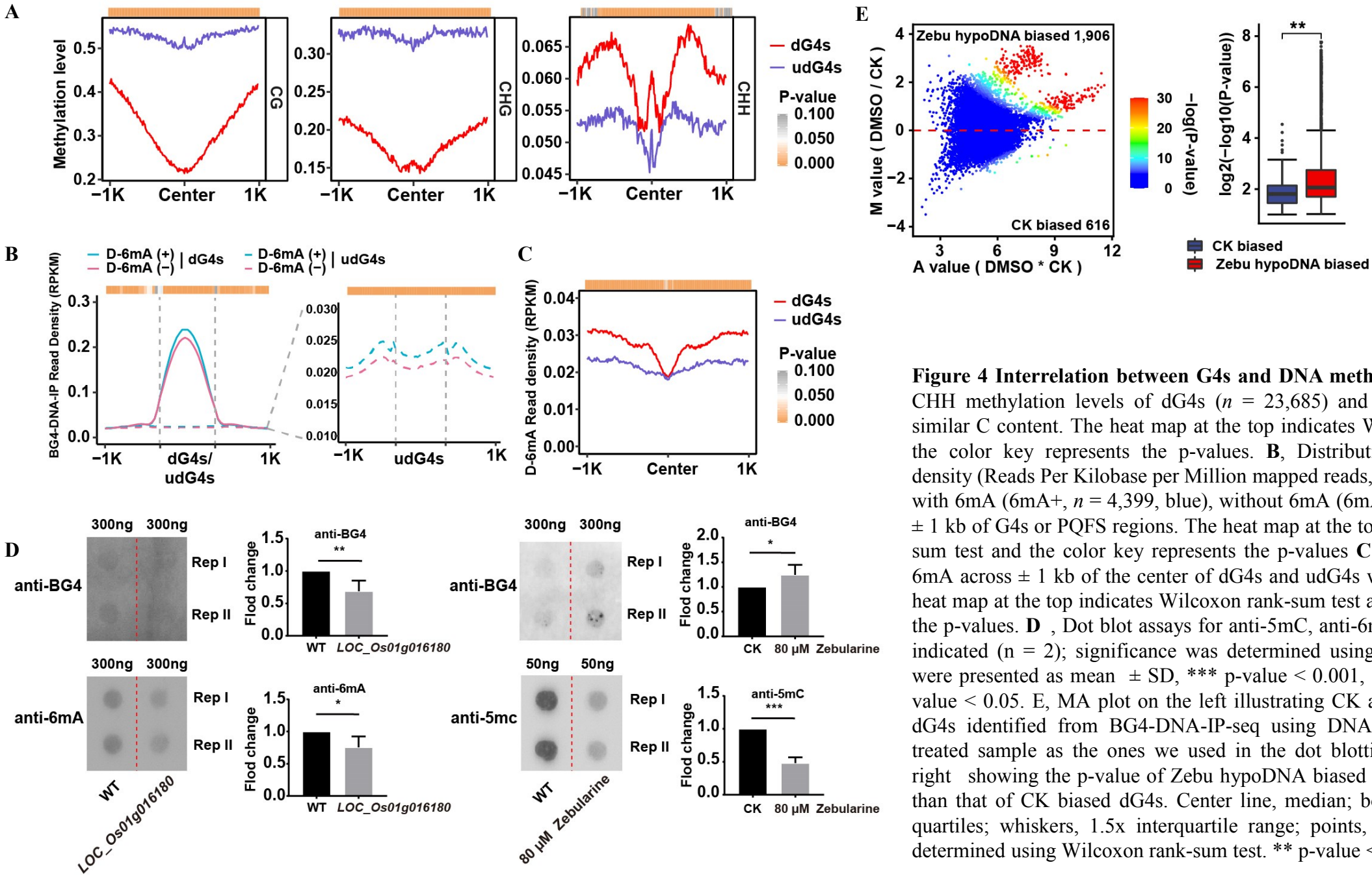


**Figure 4 Interrelation between G4s and DNA methylation. A**, CG, CHG, and CHH methylation levels of dG4s ($n$ = 23,685) and udG4s ($n$ = 23,685) with similar C content. The heat map at the top indicates Wilcoxon rank-sum test and the color key represents the p-values. **B**, Distribution of BG4-DNA-IP read density (Reads Per Kilobase per Million mapped reads, RPKM) of dG4s and udG4s with 6mA (6mA+, $n$ = 4,399, blue), without 6mA (6mA-, $n$ = 27,826, red) around ± 1 kb of G4s or PQFS regions. The heat map at the top indicates Wilcoxon rank-sum test and the color key represents the p-values **C**, Distribution of RPKM of 6mA across ± 1 kb of the center of dG4s and udG4s with similar A content. The heat map at the top indicates Wilcoxon rank-sum test and the color key represents the p-values. **D** , Dot blot assays for anti-5mC, anti-6mA and BG4 antibodies, as indicated (n = 2); significance was determined using variance analysis. Values were presented as mean ± SD, *** p-value < 0.001, ** p-value < 0.01 and * p-value < 0.05. E, MA plot on the left illustrating CK and Zebu hypoDNA biased dG4s identified from BG4-DNA-IP-seq using DNA from CK and zebularine treated sample as the ones we used in the dot blotting assay; box plot on the right showing the p-value of Zebu hypoDNA biased dG4s is significantly lower than that of CK biased dG4s. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. Significance was determined using Wilcoxon rank-sum test. ** p-value < 0.01.

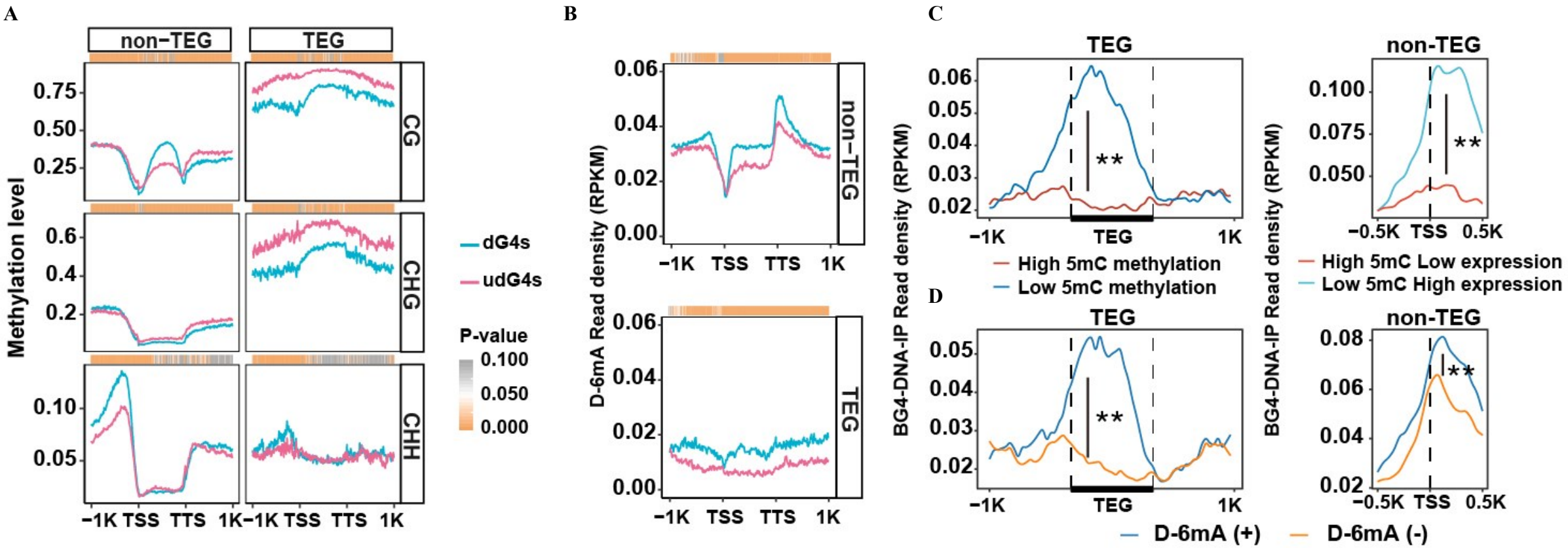


**Figure 5 Relation between G4s and DNA methylation of dG4s in transposable element genes (TEGs) and non-TEGs. A**, CG, CHG, and CHH methylation levels in TEGs and non-TEGs associated with dG4s (red) and udG4s (blue) with similar C content. The heat map at the top indicates Wilcoxon rank-sum test and the color key represents the p-values. **B**, D-6mA methylation levels in non-TEGs (top panel) and TEGs (bottom panel) associated with dG4s (red) and udG4s (blue). Read density (RPKM) of DNA 6mA was plotted across 1 kb upstream of TSSs to 1 kb downstream of TTSs of genes with dG4s and udG4s. The heat map at the top indicates Wilcoxon rank-sum test and the color key represents the p-values. **C**, Plots showing the profile of BG4-DNA-IP reads ±1 kb around TEGs (left) and from 0.5 kb upstream of TSSs to 0.5 kb downstream of TSSs of non-TEGs (right) with high and low 5mC methylation levels (n = 500), as indicated. **D**, Plots showing the profile of BG4-DNA-IP reads ±1 kb around TEGs (left) and from 0.5 kb upstream of TSSs to 0.5 kb downstream of TSSs of non-TEGs (right) with or without D-6mA peaks (n = 500), as indicated. Significance in c and d was determined using Wilcoxon rank-sum test. ** p-value < 0.05.

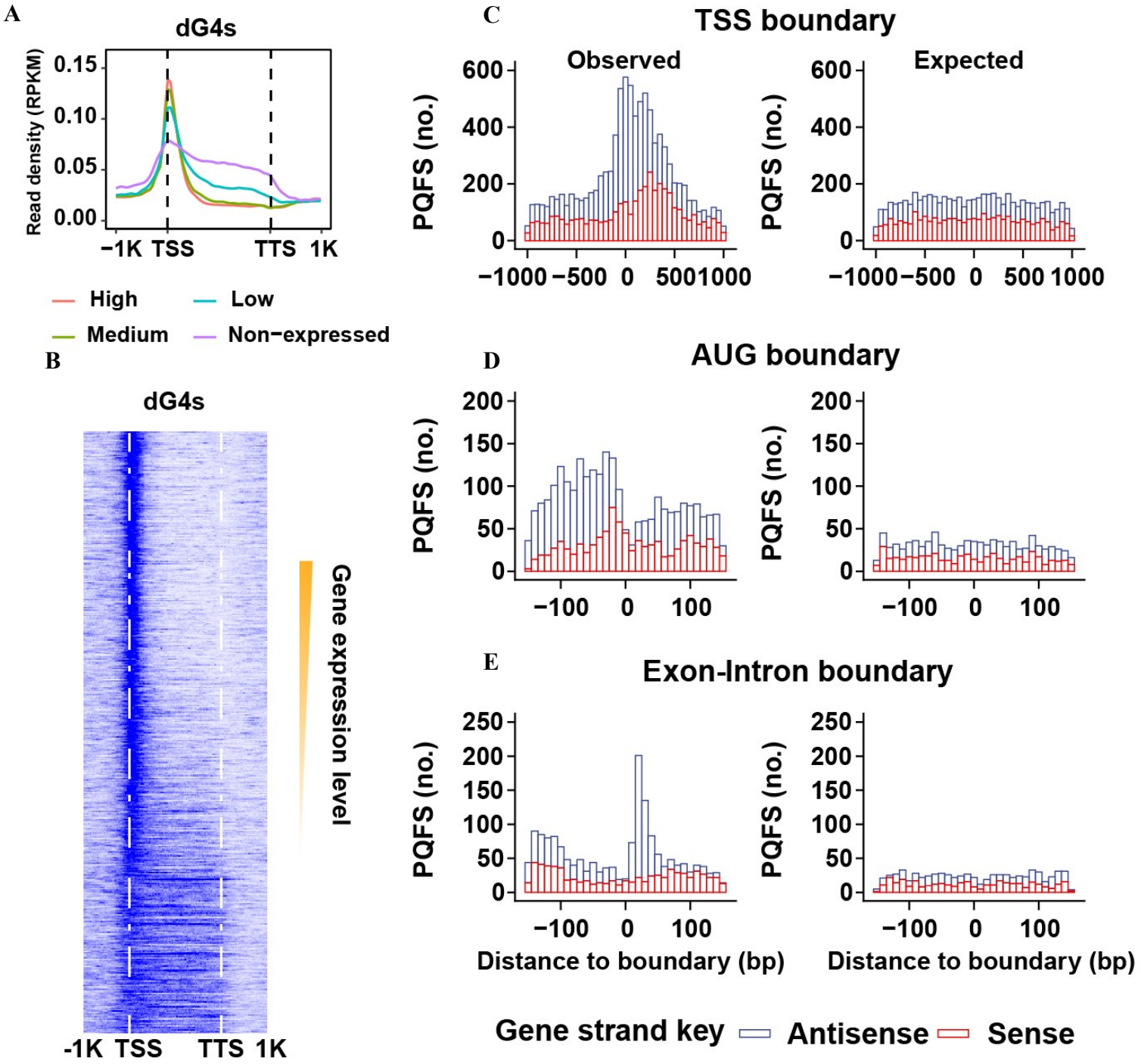


**Figure 6 Relationships between dG4s and transcription of dG4-containing genes. A**, Plots showing the profile of read density (RPKM) of BG4-DNA-IP from 1 kb upstream of TSSs to 1 kb downstream of TTSs of dG4s in non-TEGs with different Fragments Per Kilobase per Million mapped fragments (FPKM) values (high, medium, low and non-expressed). **B**, The heat map showing distribution of RPKM of BG4-DNA-IP from 1 kb upstream of TSSs to 1 kb downstream of TTSs of dG4-containing genes, ranked according to their FPKM. **(C, D and E)** Histograms showing the abundance of antisense (blue) and sense (red) PQFSs at specific positions plotted for the observed gene sets or randomized sequences (expected) for the TSS (**C**), the mRNA AUG (**D**) and the 1st Exon-Intron boundary (**E**).

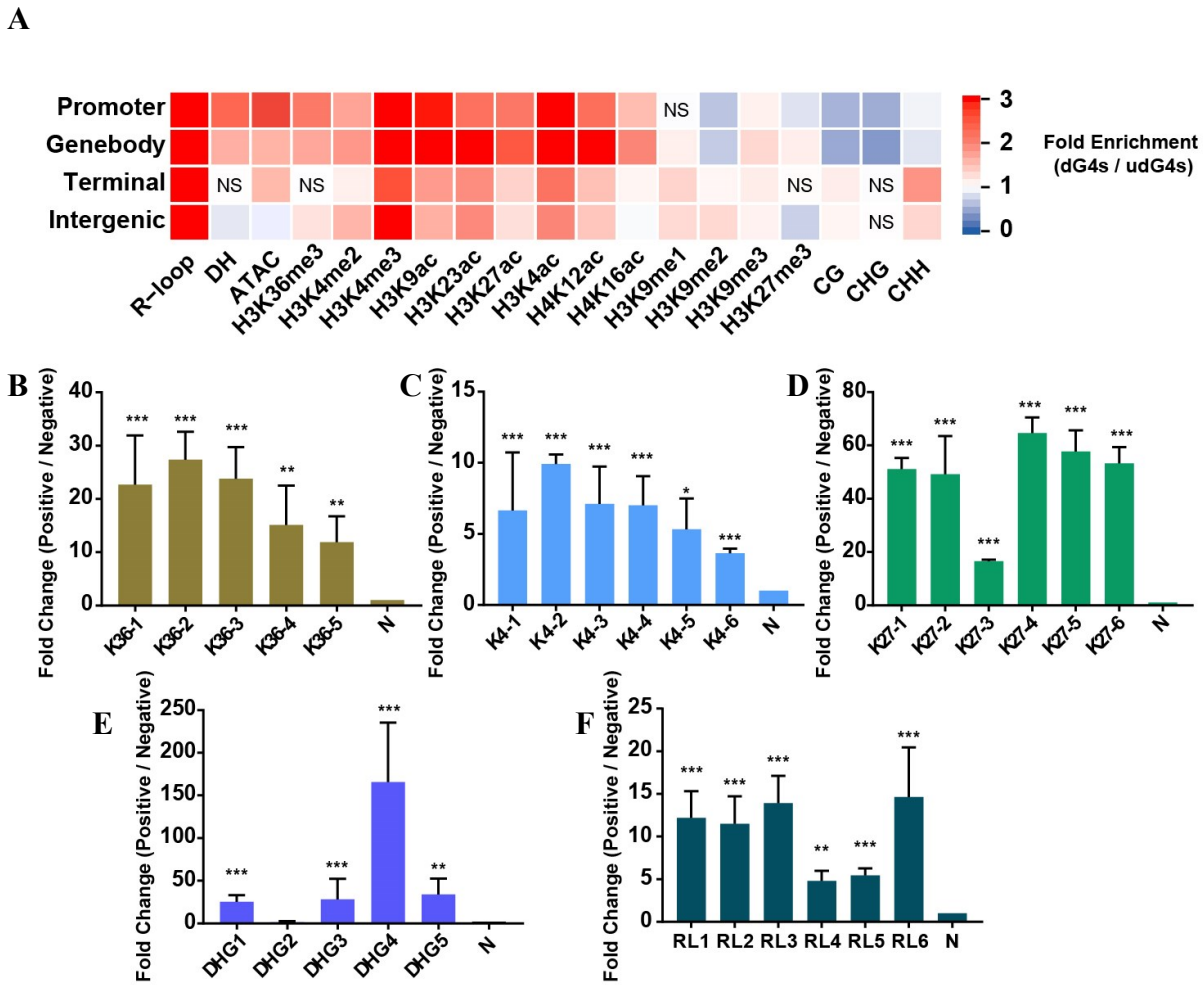


**Figure 7 Epigenomic signatures of dG4-containing genes. A**, Heat map showing enrichment of each mark for dG4s relative to udG4s in each indicated subgenomic region, the color key represents the fold enrichment of dG4s relative to udG4s. Permutation test was conducted to determine the significance of the difference when the observed value was >90% or <90% of the permutation value. **(B, C and D)** H3K36me3-ChIP (**B**), H3K4me3-ChIP (**C**) and H3K27me3-ChIP (**D**) coupled with BG4-DNA-IP-qPCR for five positive loci (with mark and G4) and one control (N, with G4 but without mark). **E**, DNase I treatment coupled with the BG4-DNA-IP-qPCR assay for five positive loci (D1-D5, with DNase I hypersensitive site (DHS) and G4) and one control (N, with DHS but without G4). **F**, BG4-DNA-IP coupled with S9.6-based DNA-RNA immunoprecipitation (DRIP-qPCR) for six positive loci (RL1-RL6, with R-loop and G4) and one control (N, with R-loop but without G4). The fold change in each qPCR run (n = 2) is expressed relative to the control. Primer sequences for qPCR are listed in Supplemental Table S2. B-F, values were presented as mean ± SD, *** p-value < 0.001, ** p-value < 0.01 and * p-value < 0.05 determined by variance analysis.

## Parsed Citations


**Andorf CM, Kopylov M, Dobbs D, Koch KE, Stroupe ME, Lawrence CJ, Bass HW (2014) G-quadruplex (G4) motifs in the maize (Zea mays L.) genome are enriched at specific locations in thousands of genes coupled to energy status, hypoxia, low sugar, and nutrient deprivation. J Genet Genomics 41: 627-647**
Google Scholar: Author Only Title Only Author and Title

**Bao HL, Liu HS, Xu Y (2019) Hybrid-type and two-tetrad antiparallel telomere DNA G-quadruplex structures in living human cells. Nucleic Acids Res 47: 4940-4947**
Google Scholar: Author Only Title Only Author and Title

**Biffi G, Tannahill D, McCafferty J, Balasubramanian S (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem 5: 182-186**
Google Scholar: Author Only Title Only Author and Title

**Bolt S, Zuther E, Zintl S, Hincha DK, Schmulling T (2017) ERF105 is a transcription factor gene of Arabidopsis thaliana required for freezing tolerance and cold acclimation. Plant Cell Environ 40: 108-120**
Google Scholar: Author Only Title Only Author and Title

**Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132: 311-322**
Google Scholar: Author Only Title Only Author and Title

**Brooks TA, Hurley LH (2010) Targeting MYC expression through G-quadruplexes. Genes Cancer 1: 641-649**
Google Scholar: Author Only Title Only Author and Title

**Bugaut A, Balasubramanian S (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. Biochemistry 47: 689-697**
Google Scholar: Author Only Title Only Author and Title

**Cagirici HB, Budak H, Sen TZ (2021) Genome-wide discovery of G-quadruplexes in barley. Sci Rep 11: 7876**
Google Scholar: Author Only Title Only Author and Title

**Cagirici HB, Sen TZ (2020) Genome-wide discovery of G-quadruplexes in wheat: distribution and putative functional roles. G3 10: 2021-2032**
Google Scholar: Author Only Title Only Author and Title

**Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat Biotechnol 33: 877-881**
Google Scholar: Author Only Title Only Author and Title

**Cheong VV, Heddi B, Lech CJ, Phan AT (2015) Xanthine and 8-oxoguanine in G-quadruplexes: formation of a G.G.X.O tetrad. Nucleic Acids Res 43: 10506-10514**
Google Scholar: Author Only Title Only Author and Title

**Cho H, Cho HS, Nam H, Jo H, Yoon J, Park C, Dang TVT, Kim E, Jeong J, Park S, Wallner ES, Youn H, Park J, Jeon J, Ryu H, Greb T, Choi K, Lee Y, Jang SK, Ban C, Hwang I (2018) Translational control of phloem development by RNA G-quadruplex-JULGI determines plant sink strength. Nat Plants 4: 376-390**
Google Scholar: Author Only Title Only Author and Title

**David AP, Margarit E, Domizi P, Banchio C, Armas P, Calcaterra NB (2016) G-quadruplexes as novel cis-elements controlling transcription during embryonic development. Nucleic Acids Res 44: 4163-4173**
Google Scholar: Author Only Title Only Author and Title

**Du Z, Zhao Y, Li N (2009) Genome-wide colonization of gene regulatory elements by G4 DNA motifs. Nucleic Acids Res 37: 6784-6798**
Google Scholar: Author Only Title Only Author and Title

**Eddy J, Vallur AC, Varma S, Liu H, Reinhold WC, Pommier Y, Maizels N (2011) G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. Nucleic Acids Res 39: 4975-4983**
Google Scholar: Author Only Title Only Author and Title

**Fang Y, Chen L, Lin K, Feng Y, Zhang P, Pan X, Sanders J, Wu Y, Wang XE, Su Z, Chen C, Wei H, Zhang W (2019) Characterization of functional relationships of R-loops with gene transcription and epigenetic modifications in rice. Genome Res 29: 1287-1297**
Google Scholar: Author Only Title Only Author and Title

**Fang Y, Wang X, Wang L, Pan X, Xiao J, Wang XE, Wu Y, Zhang W (2016) Functional characterization of open chromatin in bidirectional promoters of rice. Sci Rep 6: 32088**
Google Scholar: Author Only Title Only Author and Title

**Fleming AM, Zhu J, Ding Y, Burrows CJ (2019) Location dependence of the transcriptional response of a potential G-quadruplex in gene promoters under oxidative stress. Nucleic Acids Res 47: 5049-5060**
Google Scholar: Author Only Title Only Author and Title

**Fujimoto A, Fujita M, Hasegawa T, Wong JH, Maejima K, Oku-Sasaki A, Nakano K, Shiraishi Y, Miyano S, Yamamoto G, Akagi K, Imoto S, Nakagawa H (2020) Comprehensive analysis of indels in whole-genome microsatellite regions and microsatellite instability across 21 cancer types. Genome Res 30: 334-346**
Google Scholar: Author Only Title Only Author and Title

**Garg R, Aggarwal J, Thakkar B (2016) Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. Sci Rep 6: 28211**
Google Scholar: Author Only Title Only Author and Title

**Ge F, Wang Y, Li H, Zhang R, Wang X, Li Q, Liang Z, Yang L (2019) Plant-GQ: an integrative database of G-quadruplex in plant. J Comput Biol 26: 1013-1019**
Google Scholar: Author Only Title Only Author and Title

**Gehring K, Leroy JL, Gueron M (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. Nature 363: 561-565**
Google Scholar: Author Only Title Only Author and Title

**Gellert M, Lipsett MN, Davies DR (1962) Helix formation by guanylic acid. Proc Natl Acad Sci U S A 48: 2013-2018**
Google Scholar: Author Only Title Only Author and Title

**Griffin BD, Bass HW (2018) Review: Plant G-quadruplex (G4) motifs in DNA and RNA; abundant, intriguing sequences of unknown function. Plant Sci 269: 143-147**
Google Scholar: Author Only Title Only Author and Title

**Hansel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M, Tannahill D, Balasubramanian S (2016) G-quadruplex structures mark human regulatory chromatin. Nat Genet 48: 1267-1272**
Google Scholar: Author Only Title Only Author and Title

**Hansel-Hertsch R, Simeone A, Shea A, Hui WWI, Zyner KG, Marsico G, Rueda OM, Bruna A, Martin A, Zhang X, Adhikari S, Tannahill D, Caldas C, Balasubramanian S (2020) Landscape of G-quadruplex DNA structural regions in breast cancer. Nat Genet 52: 878-883**
Google Scholar: Author Only Title Only Author and Title

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

**Hao D, Ohme-Takagi M, Sarai A (1998) Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J Biol Chem 273: 26857-26861**
Google Scholar: Author Only Title Only Author and Title

**Hu L, Li N, Xu C, Zhong S, Lin X, Yang J, Zhou T, Yuliang A, Wu Y, Chen YR, Cao X, Zemach A, Rustgi S, von Wettstein D, Liu B (2014) Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. Proc Natl Acad Sci U S A 111: 10642-10647**
Google Scholar: Author Only Title Only Author and Title

**Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res 33: 2908-2916**
Google Scholar: Author Only Title Only Author and Title

**Kang SG, Henderson E (2002) Identification of non-telomeric G4-DNA binding proteins in human, E. coli, yeast, and Arabidopsis. Mol Cells 14: 404-410**
Google Scholar: Author Only Title Only Author and Title

**Kim N (2019) The interplay between G-quadruplex and transcription. Curr Med Chem 26: 2898-2917**
Google Scholar: Author Only Title Only Author and Title

**Kopylov M, Bass HW, Stroupe ME (2015) The maize (Zea mays L.) nucleoside diphosphate kinase1 (ZmNDPK1) gene encodes a human NM23-H2 homologue that binds and stabilizes G-quadruplex DNA. Biochemistry 54: 1743-1757**
Google Scholar: Author Only Title Only Author and Title

**Kopylov M, Jackson TM, Stroupe ME (2019) Bulged and canonical G-quadruplex conformations determine NDPK binding specificity. Molecules 24: 1988**
Google Scholar: Author Only Title Only Author and Title

**Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27: 1571-1572**
Google Scholar: Author Only Title Only Author and Title

**Kumar P, Yadav VK, Baral A, Kumar P, Saha D, Chowdhury S (2011) Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. Nucleic Acids Res 39: 8005-8016**
Google Scholar: Author Only Title Only Author and Title

**Kwok CK, Ding Y, Shahid S, Assmann SM, Bevilacqua PC (2015) A stable RNA G-quadruplex within the 5'-UTR of Arabidopsis thaliana ATR mRNA inhibits translation. Biochem J 467: 91-102**
Google Scholar: Author Only Title Only Author and Title

**Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S (2016) rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods 13: 841-844**
Google Scholar: Author Only Title Only Author and Title

**Lexa M, Kejnovsky E, Steflova P, Konvalinova H, Vorlickova M, Vyskot B (2014) Quadruplex-forming sequences occupy discrete regions inside plant LTR retrotransposons. Nucleic Acids Res 42: 968-978**
Google Scholar: Author Only Title Only Author and Title

**Liang Z, Zhang Q, Ji C, Hu G, Zhang P, Wang Y, Yang L, Gu X (2021) Reorganization of the 3D chromatin architecture of rice genomes during heat stress. BMC Biol 19: 53**
Google Scholar: Author Only Title Only Author and Title

**Liu HY, Zhao Q, Zhang TP, Wu Y, Xiong YX, Wang SK, Ge YL, He JH, Lv P, Ou TM, Tan JH, Li D, Gu LQ, Ren J, Zhao Y, Huang ZS (2016) Conformation selective antibody enables genome profiling and leads to discovery of parallel G-quadruplex in human telomeres. Cell Chem Biol 23: 1261-1270**
Google Scholar: Author Only Title Only Author and Title

**Lu L, Chen X, Sanders D, Qian S, Zhong X (2015) High-resolution mapping of H4K16 and H3K23 acetylation reveals conserved and unique distribution patterns in Arabidopsis and rice. Epigenetics 10: 1044-1053**
Google Scholar: Author Only Title Only Author and Title

**Luo GZ, MacQueen A, Zheng G, Duan H, Dore LC, Lu Z, Liu J, Chen K, Jia G, Bergelson J, He C (2014) Unique features of the m6A methylome in Arabidopsis thaliana. Nat Commun 5: 5630**
Google Scholar: Author Only Title Only Author and Title

**Machanick P, Bailey TL (2011) MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics 27: 1696-1697**
Google Scholar: Author Only Title Only Author and Title

**Mao SQ, Ghanbarian AT, Spiegel J, Martinez Cuesta S, Beraldi D, Di Antonio M, Marsico G, Hansel-Hertsch R, Tannahill D, Balasubramanian S (2018) DNA G-quadruplex structures mold the DNA methylome. Nat Struct Mol Biol 25: 951-957**
Google Scholar: Author Only Title Only Author and Title

**Marsico G, Chambers VS, Sahakyan AB, McCauley P, Boutell JM, Di Antonio M, Balasubramanian S (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. Nucleic Acids Res 47: 3862-3874**
Google Scholar: Author Only Title Only Author and Title

**Mishra SK, Tawani A, Mishra A, Kumar A (2016) G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. Sci Rep 6: 38144**
Google Scholar: Author Only Title Only Author and Title

**Mukundan VT, Phan AT (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. J Am Chem Soc 135: 5017-5028**
Google Scholar: Author Only Title Only Author and Title

**Mullen MA, Olson KJ, Dallaire P, Major F, Assmann SM, Bevilacqua PC (2010) RNA G-Quadruplexes in the model plant species Arabidopsis thaliana: prevalence and possible functional roles. Nucleic Acids Res 38: 8149-8163**
Google Scholar: Author Only Title Only Author and Title

**Nie J, Jiang M, Zhang X, Tang H, Jin H, Huang X, Yuan B, Zhang C, Lai JC, Nagamine Y, Pan D, Wang W, Yang Z (2015) Post-transcriptional regulation of Nkx2-5 by RHAU in heart development. Cell Rep 13: 723-732**
Google Scholar: Author Only Title Only Author and Title

**Reina C, Cavalieri V (2020) Epigenetic modulation of chromatin states and gene expression by G-quadruplex structures. Int J Mol Sci 21: 4172**
Google Scholar: Author Only Title Only Author and Title

**Renard I, Grandmougin M, Roux A, Yang SY, Lejault P, Pirrotta M, Wong JMY, 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**
Google Scholar: Author Only Title Only Author and Title

**Sanz LA, Chedin F (2019) High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc 14: 1734-1755**
Google Scholar: Author Only Title Only Author and Title

**Schiavone D, Guilbaud G, Murat P, Papadopoulou C, Sarkies P, Prioleau MN, Balasubramanian S, Sale JE (2014) Determinants of G-quadruplex-induced epigenetic instability in REV1-deficient cells. EMBO J 33: 2507-2520**
Google Scholar: Author Only Title Only Author and Title

**Sen D, Gilbert W (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature 334: 364-366**
Google Scholar: Author Only Title Only Author and Title

**Sengar A, Vandana JJ, Chambers VS, Di Antonio M, Winnerdy FR, Balasubramanian S, Phan AT (2019) Structure of a (3+1) hybrid G-quadruplex in the PARP1 promoter. Nucleic Acids Res 47: 1564-1572**

Google Scholar: Author Only Title Only Author and Title

**Simonini S, Roig-Villanova I, Gregis V, Colombo B, Colombo L, Kater MM (2012) Basic pentacysteine proteins mediate MADS domain complex binding to the DNA for tissue-specific expression of target genes in Arabidopsis. Plant Cell 24: 4163-4172**
Google Scholar: Author Only Title Only Author and Title

**Sparks MA, Singh SP, Burgers PM, Galletto R (2019) Complementary roles of Pif1 helicase and single stranded DNA binding proteins in stimulating DNA replication through G-quadruplexes. Nucleic Acids Res 47: 8595-8605**
Google Scholar: Author Only Title Only Author and Title

**Spiegel J, Adhikari S, Balasubramanian S (2020) The structure and function of DNA G-quadruplexes. Trends Chem 2: 123-136**
Google Scholar: Author Only Title Only Author and Title

**Spiegel J, Cuesta SM, Adhikari S, Hansel-Hertsch R, Tannahill D, Balasubramanian S (2021) G-quadruplexes are transcription factor binding hubs in human chromatin. Genome Biol 22: 117**
Google Scholar: Author Only Title Only Author and Title

**Sun D, Hurley LH (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. Methods Mol Biol 608: 65-79**
Google Scholar: Author Only Title Only Author and Title

**Sundquist WI, Klug A (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature 342: 825-829**
Google Scholar: Author Only Title Only Author and Title

**Takahashi H, Nakagawa A, Kojima S, Takahashi A, Cha BY, Woo JT, Nagai K, Machida Y, Machida C (2012) Discovery of novel rules for G-quadruplex-forming sequences in plants by using bioinformatics methods. J Biosci Bioeng 114: 570-575**
Google Scholar: Author Only Title Only Author and Title

**Tan F, Zhou C, Zhou Q, Zhou S, Yang W, Zhao Y, Li G, Zhou DX (2016) Analysis of chromatin regulators reveals specific features of rice DNA methylation pathways. Plant Physiol 171: 2041-2054**
Google Scholar: Author Only Title Only Author and Title

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

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

**Vogel MO, Moore M, Konig K, Pecher P, Alsharafa K, Lee J, Dietz KJ (2014) Fast retrograde signaling in response to high light involves metabolite export, MITOGEN-ACTIVATED PROTEIN KINASE6, and AP2/ERF transcription factors in Arabidopsis. Plant Cell 26: 1151-1165**
Google Scholar: Author Only Title Only Author and Title

**Wallgren M, Mohammad JB, Yan KP, Pourbozorgi-Langroudi P, Ebrahimi M, Sabouri N (2016) G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures in vitro and are unwound by the Pfh1 DNA helicase. Nucleic Acids Res 44: 6213-6231**
Google Scholar: Author Only Title Only Author and Title

**Wang X, Goodrich KJ, Conlon EG, Gao J, Erbse AH, Manley JL, Cech TR (2019) C9orf72 and triplet repeat disorder RNAs: G-quadruplex formation, binding to PRC2 and implications for disease mechanisms. RNA 25: 935-947**
Google Scholar: Author Only Title Only Author and Title

**Wang Y, Zhao M, Zhang Q, Zhu GF, Li FF, Du LF (2015) Genomic distribution and possible functional roles of putative G-quadruplex motifs in two subspecies of Oryza sativa. Comput Biol Chem 56: 122-130**
Google Scholar: Author Only Title Only Author and Title

**Wu WQ, Zhang ML, Song CP (2020) A comprehensive evaluation of a typical plant telomeric G-quadruplex (G4) DNA reveals the dynamics of G4 formation, rearrangement, and unfolding. J Biol Chem 295: 5461-5469**
Google Scholar: Author Only Title Only Author and Title

**Wu Y, Kikuchi S, Yan H, Zhang W, Rosenbaum H, Iniguez AL, Jiang J (2011) Euchromatic subdomains in rice centromeres are associated with genes and transcription. Plant Cell 23: 4054-4064**
Google Scholar: Author Only Title Only Author and Title

**Yadav V, Hemansi, Kim N, Tuteja N, Yadav P (2017) G-quadruplex in plants: A ubiquitous regulatory element and its biological relevance. Front Plant Sci 8: 1163**
Google Scholar: Author Only Title Only Author and Title

**Yang SY, Lejault P, Chevrier S, Boidot R, Robertson AG, Wong JMY, Monchaud D (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. Nat Commun 9: 4730**
Google Scholar: Author Only Title Only Author and Title

**Yang X, Cheema J, Zhang Y, Deng H, Duncan S, Umar MI, Zhao J, Liu Q, Cao X, Kwok CK, Ding Y (2020) RNA G-quadruplex structures exist and function in vivo in plants. Genome Biol 21: 226**
Google Scholar: Author Only Title Only Author and Title

**Zang Z, Lv Y, Liu S, Yang W, Ci J, Ren X, Wang Z, Wu H, Ma W, Jiang L, Yang W (2020) A novel ERF Transcription Factor, ZmERF105, positively regulates maize resistance to exserohilum turcicum. Front Plant Sci 11: 850**
Google Scholar: Author Only Title Only Author and Title

**Zeraati M, Langley DB, Schofield P, Moye AL, Rouet R, Hughes WE, Bryan TM, Dinger ME, Christ D (2018) I-motif DNA structures are formed in the nuclei of human cells. Nat Chem 10: 631-637**
Google Scholar: Author Only Title Only Author and Title

**Zhang S, Sun H, Wang L, Liu Y, Chen H, Li Q, Guan A, Liu M, Tang Y (2018) Real-time monitoring of DNA G-quadruplexes in living cells with a small-molecule fluorescent probe. Nucleic Acids Res 46: 7522-7532**
Google Scholar: Author Only Title Only Author and Title

**Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J (2012) High-resolution mapping of open chromatin in the rice genome. Genome Res 22: 151-162**
Google Scholar: Author Only Title Only Author and Title

**Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9: R137**
Google Scholar: Author Only Title Only Author and Title

**Zheng D, Wang L, Chen L, Pan X, Lin K, Fang Y, Wang XE, Zhang W (2019) Salt-responsive genes are differentially regulated at the chromatin levels between seedlings and roots in rice. Plant Cell Physiol 60: 1790-1803**
Google Scholar: Author Only Title Only Author and Title

**Zheng KW, Zhang JY, He YD, Gong JY, Wen CJ, Chen JN, Hao YH, Zhao Y, Tan Z (2020) Detection of genomic G-quadruplexes in living cells using a small artificial protein. Nucleic Acids Res 48: 11706-11720**
Google Scholar: Author Only Title Only Author and Title

**Zhou C, Wang C, Liu H, Zhou Q, Liu Q, Guo Y, Peng T, Song J, Zhang J, Chen L, Zhao Y, Zeng Z, Zhou DX (2018) Identification and analysis of adenine N(6)-methylation sites in the rice genome. Nat Plants 4: 554-563**
Google Scholar: Author Only Title Only Author and Title

**Zhou L, Cheng X, Connolly BA, Dickman MJ, Hurd PJ, Hornby DP (2002) Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. J Mol Biol 321: 591-599**
Google Scholar: Author Only Title Only Author and Title