



HAL
open science

Omics studies of DNA G-/C-quadruplexes in plants

Mengqi Li, Ruiping Tian, David Monchaud, Wenli Zhang

► **To cite this version:**

Mengqi Li, Ruiping Tian, David Monchaud, Wenli Zhang. Omics studies of DNA G-/C-quadruplexes in plants. Trends in Genetics, 2022, 38 (10), pp.999-1002. 10.1016/j.tig.2022.06.005 . hal-03836052

HAL Id: hal-03836052

<https://hal.science/hal-03836052>

Submitted on 17 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Omics studies of DNA G-/C-quadruplexes in plants

Mengqi Li,^{1,3} Ruiping Tian,^{1,3} David Monchaud,² and Wenli Zhang^{1,*}

Abstract. Genome-wide studies of DNA G- and C-quadruplexes (G4s and i-motifs, respectively) can boost the pace of progress towards a comprehensive understanding of their biological implications and practical applications in plants. We summarize the current state of knowledge about omics studies in order to highlight the current challenges and propose future directions to take studies of plant quadruplexes to the next step.

Introduction

DNA G-quadruplexes (G4s) (see Glossary) and i-motifs (iMs) are two main types of quadruplexes (i.e., four-stranded DNA structures) located within G- or C-rich genomic DNA regions, respectively. The distribution of Gs/Cs in these regions is not random – they must be gathered in runs (comprising usually two to four units) and several runs must be consecutive (with four runs or more), which makes these sequences known as putative G4-forming sequences (PQFSs) or putative iM-forming sequences (PiMFSs), respectively, readily identifiable. Both quadruplexes are physically prevalent in gene promoters and functional chromatin regions, such as telomeres and centromeres. They are functionally implicated in critical cellular processes in humans, including maintenance of genome stability, regulation of gene expression, and human diseases [1–4]. Their formation and biological relevance *in vivo* are being increasingly accepted for G4s and emerging for iMs in the human genome, but remain largely unknown in the plant genome. Further and massive effort must thus be invested to decipher the exact roles of quadruplexes, particularly iMs, in the regulation of gene transcription on a genome-wide scale in both humans and plants. The advent of high-

throughput sequencing experimental methodologies has provided a key turning point to improve our understanding of their genomic/epigenomic features and subsequent biological implications in plants. Here, we summarize the omics studies developed for both DNA quadruplexes, discuss the challenges already met and those that remain to be addressed for in vivo applications, and propose future directions to advance their studies and applications in plants.

Existing methodologies for the identification of DNA G4s or iMs in plants. The experimental methodologies for the identification of both quadruplexes can be generally classified into low- and high-throughput methodologies (Figure 1). The former comprise structural (NMR, CD, X-ray crystallography, etc.), biochemical (PAGE, equilibrium dialysis, PCR-based assays, etc.), and biophysical investigations (SPR, ESI-MS, UV- and fluorescence-based assays, etc.). While being time- and labor-consuming, these techniques are still largely used for studying individual genomic 'loci' and useful for both validating high-throughput methodologies and studying quadruplex–molecule interactions. The latter can be classified into theoretical investigations (computer-aided prediction) and experimental identification (sequencing-based methods). Several customized algorithms or scripts have been applied for genome-wide prediction of quadruplexes in both mammals and plants [5,6] (Figure 1). It is worth mentioning that even if in silico prediction is an efficient way to identify PQFSs and PiMFSs, it eventually requires an experimental validation.

Sequencing-based approaches have been applied for the identification of DNA G4s both in vitro (i.e., in purified DNA) and in vivo (i.e., in a chromatin context) in humans: G4-seq allows for the detection of G4s in vitro, relying on the use of a small-molecule ligand (pyridostatin) coupled with G4-mediated polymerase stop assays [7,8]. G4 ChIP-seq allows for the detection of G4s in vivo thanks to the immunoprecipitation of G4s by the antibody BG4 [9], while G4P-ChIP-seq relies on the use of an artificially truncated G4 helicase DHX36 to precipitate G4s [10], and G4DP-seq relies on the use of a small molecular ligand BioTASQ for G4 affinity capture [11] (Figure 1). However, these approaches suffer from a limitation related to the different G4 binding capacity between antibodies (e.g., BG4) and proteins (e.g., helicase). Only BG4-DNA-IP-seq [12] and G4-seq [7] have been adapted for the identification of DNA G4s in vitro in rice and Arabidopsis, which paves the way for a global characterization of plant DNA G4s

(Figure 1). More sequencing and chemical biology methodologies are still needed for in vitro/in vivo G4 studies that can deepen our understanding of their biological relevance in plants.

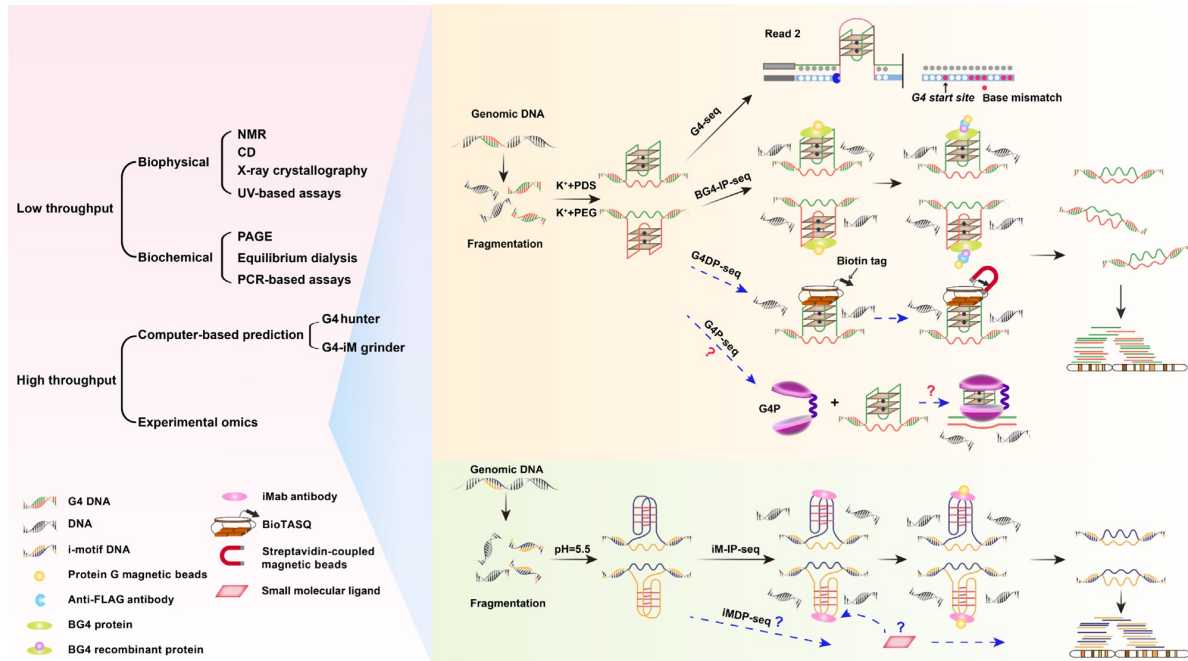


Figure 1. Summary of methodologies for study of DNA G4s or iMs. They include low- and high-throughput methodologies. Low-throughput ones can be divided into biophysical and biochemical methods. High-throughput ones can be classified into computer-based prediction [5,6] and experimental omics studies. Experimental omics studies include small-molecule ligand coupled with DNA polymerases top assay (G4-seq) [7,8], antibody [BG4-(Ch)IP-seq, iM-IP-seq] [9,12,14,15], truncated native protein (G4P-seq) [10], and small-molecule ligand affinity capture (G4DP-seq) [11] sequencing. Black arrows indicate methods used in plants, red question mark indicates methods used in humans but not in plants, and the blue question mark indicates the potential method not applied in both humans and plants. Abbreviations: DNA G4s, DNA G-quadruplexes; iMs, i-motifs.

The global identification of DNA iMs lags far behind that of G4s in both humans and plants. The advent of the iMab antibody, specific for binding to human iMs [13], achieves a big step towards genome-wide studies of iMs in both humans and plants [14,15]. The iMab antibody-based iM-IP-seq methodology was recently developed for the identification of in vitro DNA iMs in the rice genome [14] (Figure 1). This can be considered as a breakthrough and opens the way for genome-wide studies of DNA iMs in both mammals and plants, but in vivo studies are still missing. Further massive effort must thus be invested to push the boundaries of such omics studies, which aim at advancing our understanding of their functional relevance in plants.

Potential biological implications of DNA G4s or iMs in plants. Recently obtained results help better understand the G4 functions in the regulation of gene transcription, DNA damage and repair, chromatin organization, interaction with epigenetic marks, diseases, and pharmacogenomics in humans [1]. However, global DNA G4 studies in plants were only reported in Arabidopsis [7] and rice [12]. BG4-DNA-IP-seq uniquely uncovered genomic-dependent functions of DNA G4s in the regulation of gene transcription and potential interrelations between G4s and epigenomic signatures in rice [12] (Figure 2). It also raised questions about mechanistic studies in plants, such as those about the interrelationships between DNA or histone modifications and G4 formation. Studies of G4s in related mutants will help address those questions.

It was originally assumed that iMs could only form under acidic conditions, due to the requirement of the hemi-protonation of Cs that is required for the formation of C-C⁺ base pairs. This might explain why the studies of iMs under physiological cellular functions have been long in coming. The recent characterization of the iM-specific iMab has invigorated iM studies as this antibody provided the first strong evidence of iM formation under physiological conditions in human cells [13]. This undoubtedly generates a renewed interest in reconsidering iM functions in human cells, in addition to their involvement in the maintenance of genome stability [2] and the regulation of gene transcription (e.g., HIV-1 and BmPOUM2) [3,4], which needs to be further verified at a genome-wide level. Similar to BG4-DNA-IP-seq described above, the recently developed iM-IP-seq provided the first global mapping of iMs in rice [14], revealing its potential functions in the regulation of gene transcription and genome stability through regulation of TE (transposable element) activity. This very first genome-wide characterization of iM landscapes in plants (Figure 2) pioneered iM genomic studies but also raised several questions that still await further exploration, including those related to the interrelationships between DNA methylation or histone modifications and iM formation, or how iMs act as cis-regulatory elements (CREs) or act as platforms for recruitment of trans-acting factors in the regulation of gene transcription or TE activities in plants.

In vivo omics studies are urgently needed in plants. The omics studies in plants discussed above primarily reflect folding potentials of PQFSs or PiMFSs in vitro, which cannot truly pinpoint quadruplex formation within the chromatin context (i.e., in vivo).

An illustrative example is the comparison of G4-seq and G4 ChIP-seq datasets that brightly demonstrates the suppressive role of chromatin for G4 formation in cells (the number of G4s detected in vivo representing ~1% only of the G4s detected in vitro). Therefore, more efforts must be invested in profiling the G4/iM landscapes in vivo in plants. There are several possible ways that deserve to be assayed, which involve: (i) the discovery of more accurate antibodies against G4/iM along with the adaptation and optimization of the related ChIP-seq methods developed in human cells and implemented for the identification of quadruplexes in plants; (ii) the development of multivalent small molecules (ligands) implementable as molecular baits for chemoprecipitation of G4/iM followed by sequencing, or Chem-seq (as performed with BioTASQ for plant G4s). Of note, while dozens of G4 ligands have been designed and used, including ad hoc molecular G4 baits, reliable iM ligands are yet to be found, which somehow curtails the development of this field; and (iii) the production of artificially truncated plant-specific G4/iM-binding proteins for ChIP-seq-like investigations (as performed with the truncated DHX36 for human G4s). Comprehensive comparisons of G4/iM landscapes using plant cells that transiently or stably express tagged G4/iM-binding proteins could help decipher the dynamics of G4/iM folding/ unfolding related to the expression of G4/ iM-binding proteins across the genome.

Any advance that will enable us to portray in vivo quadruplex landscapes more accurately will greatly improve our understanding of the biological functions of G4/iM in developmental stages, phenotypic changes, genome stability, chromatin structure and stress responses (as a function of environmental stimuli including salt, heat and acidic conditions).

Concluding remarks

While still in their early stages as compared to those performed in human cells, the studies of G4s and iMs in plants are nevertheless fascinating and exciting. The handful of studies reported to date have provided invaluable insights into the functional relevance of G4s/iMs in plant cells but also raised a plethora of key questions that impatiently await further investigations. For example, how the dynamics of these quadruplexes are linked to development changes or to stress responses? Do they act as autonomous genetic regulators or cooperatively with chromatin marks to regulate

gene expression during normal growth, different development stages and stress responses? To unravel these conundrums, *in vivo* omics studies are urgently needed as they will boost functional studies, help design potential sensors usable in climate change adaptation and provide key 'loci' for genomic editing towards desirable agronomic traits formation. Please see the online supplemental information for an additional reading list.

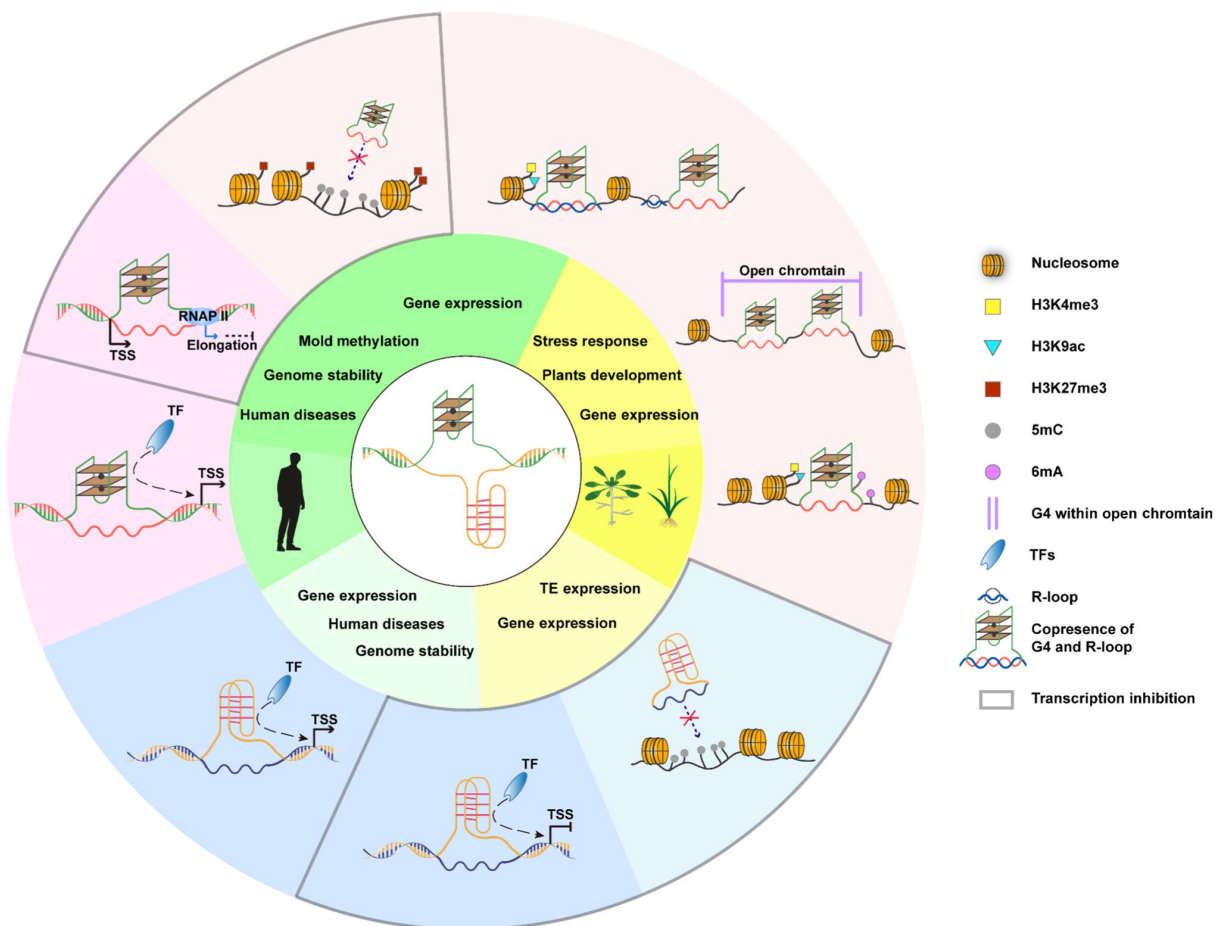


Figure 2. Biological relevance of DNA G4s and iMs in humans and plants. They include diseases, genome stability, intersect with epigenetic marks and regulation of gene expression in human (green) [1–4,7,15], and various stress responses, plant development and regulation of gene expression, TE activities in plants (yellow) [7,12,14]. Abbreviations: DNA G4s, DNA G-quadruplexes; iMs, i-motifs; TE, transposable element; TF, transcription factor; TSS, transcription start site.

Acknowledgments

We apologize to researchers whose work could not be cited due to space restrictions. This work is supported by the National Natural Science Foundation of China (32070561 and U20A2030), the National Key Research and Development Program of China (2020YFE0202900), the Fundamental Research Funds for the Central

Universities (KYZZ2022003), and the Graduate Research and Practice Innovation Program of Jiangsu Province (KYCX20_0579). We thank the Bioinformatics Center in Nanjing Agricultural University for providing facilities to assist with sequencing data analyses.

Author contributions

W.Z. conceptualized the idea and provided financial support. M.L. and R.T. drafted the manuscript. M.L. prepared the figures. D.M. edited the manuscript. W.Z. rewrote and finalized the manuscript. All authors proofread the final manuscript.

Declaration of interests

The authors declare that they have no competing interests.

Glossary

Cis-regulatory elements (CREs): a type of DNA (typically, noncoding DNA) sequences contains intrinsic binding sites for TFs (transcription factors) and/or other functional proteins that are essential for the regulation of gene transcription.

DNA G-quadruplexes (G4s): noncanonical DNA quadruple helix structure with guanine-rich nucleic acid sequences formed through Hoogsteen hydrogen bonding.

DNA i-motifs (iMs): noncanonical DNA quadruplexes with cytosine-rich nucleic acid sequences held by hemiprotonated cytosine–cytosine (C) base pairs.

DNA methylation: heritable epigenetic mark, which includes 5-methylcytosine (5mC), 6-methyladenine (6mA), 4-methylcytosine (4mC), etc. High-throughput sequencing: also referred to as next-generation sequencing, high-output sequencing methods that generate genome-scale data. Histone modification: covalent post-translational modification to the N-tails of histones, which includes methylation, phosphorylation, acetylation, ubiquitylation, sumoylation, etc.

Putative G4-forming sequences (PQFSs): guanine-rich nucleotide sequences can be computationally predicted to form a G4 structure. Putative i-motif-forming sequences (PiMFSs): cytosine-rich nucleotide sequences can be computationally predicted to form an i-motif structure. Transcription start sites (TSSs): genomic location at the 5' end of a gene where transcription starts.

References

1. Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* 43, 8627–8637
2. Bacolla, A. and Wells, R.D. (2004) Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* 279, 47411–47414
3. Ruggiero, E. et al. (2019) A dynamic i-motif with a duplex stem-loop in the long terminal repeat promoter of the HIV-1 proviral genome modulates viral transcription. *Nucleic Acids Res.* 47, 11057–11068
4. Niu, K. et al. (2018) BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in *Bombyx mori*. *Nucleic Acids Res.* 46, 1710–1723
5. Bedrat, A. et al. (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746–1759
6. Belmonte-Reche, E. and Morales, J.C. (2020) G4-iM Grinder: when size and frequency matter. G-Quadruplex, i-Motif and higher order structure search and analysis tool. *NAR Genom. Bioinform.* 2, lqz005
7. Marsico, G. et al. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* 47, 3862–3874
8. Chambers, V.S. et al. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33, 877–881
9. Hänsel-Hertsch, R. et al. (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* 48, 1267–1272
10. Zheng, K.W. et al. (2020) Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res.* 48, 11706–11720
11. Feng, Y. et al. (2022) Chemo- versus immuno-precipitation of G-quadruplex-DNA (G4DNA): a direct comparison of the efficiency of the antibody BG4 versus the small-molecule ligands TASQs for G4 affinity capture. *Res. Square* Published online June 15, 2022. <https://doi.org/10.21203/rs.3.rs-1717207/v1>
12. Feng, Y. et al. (2022) Epigenomic features of DNA G-quadruplexes and their roles in regulating rice gene transcription. *Plant Physiol.* 188, 1632–1648
13. Zeraati, M. et al. (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* 10, 631–637
14. Ma, X. et al. (2022) Genome-wide characterization of i-motifs and their potential roles in the stability and evolution of transposable elements in rice. *Nucleic Acids Res.* 50, 3226–3238
15. Peña Martinez, C.D. et al. (2022) Human genomic DNA is widely interspersed with i-motif structures. *bioRxiv* Published online April 14, 2022. <https://doi.org/10.1101/2022.04.14.488274>