DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential

Robert Hänsel-Hertsch[1], Marco Di Antonio[1,2] Shankar Balasubramanian[1,2,3,*]

1. Cancer Research UK, Cambridge Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK.
2. Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.
3. School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK.

***correspondence should be addressed to S.B. (sb10031@cam.ac.uk).**

Abstract

Single-stranded guanine rich DNA sequences can fold into four-stranded DNA structures called G-quadruplexes (G4s) that arise from self-stacking of two or more guanine-quartets. There has been considerable recent progress on the detection and mapping of G4 structures in the human genome and in biologically relevant contexts. These advancements have provided important new insights into their functions, for example in regulating transcription and genome stability, and their potential for therapeutic applications, much of which is aligned to predictions previously made in computational studies.

Introduction

Single-stranded guanine rich DNA sequences can fold into stable intramolecular and intermolecular four-stranded non-B DNA structures called G-quadruplexes (G4s, Figure 1)[1]. G4s arise from Hoogsteen hydrogen bonding of four guanines arranged within a planar quartet (G-quartet)[1]. Self-stacking of two or more G-quartets generates a G4 structure that is further stabilised by monovalent cations in the order (K^+>Na^+>NH_4^+>Li^+) (Figure 1)[1]. G4 formation has been observed in oligonucleotide sequences derived from the human genome, particularly gene promoters and telomeres. Such studies, along with chemical biological approaches using G4 targeted small molecules or antibodies, and also computational predictions, have suggested that G4s may be important in gene regulation and telomere biology.

This article will focus on recent, significant progress in detecting and mapping G4s in the human genome and the new insights into their functions and their potential for therapeutic applications. More comprehensive reviews on functional roles of DNA G4 can be found elsewhere[2,3].

Imaging and Mapping G4s

Computational G4 predictions using simple algorithms have suggested that over 300,000 sequence motifs (of the type $G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$) in the human genome have the potential to form a G4 structure[4,5]. A more recent algorithm has predicted the number of potential G4 sequences to be substantially higher[6]. These computational studies showed that G4 motifs are enriched in telomeres, promoters and the first intron of genes, but have highlighted the need to generate explicit experimental data about the existence and function(s) of G4s in biologically relevant contexts. G4-selective probes have been developed and employed to capture G4s in cells by fluorescence microscopy and also by DNA chromatin immuno-precipitation followed by sequencing.

***Cellular visualisation*.** One approach to visualise particular DNA structures in cells is to employ structure-selective molecular probes. Antibody proteins can have exquisite specificity in their recognition of molecular structures and are widely used to bind to and visualise proteins within cells or map their binding sites in DNA or RNA. Antibodies can be generated by immunisation or *via in vitro* affinity selection to recognise a particular DNA structure or chemical feature. The first reported visualisation of G4 formation in a biologically relevant context used a G4-selective single-chain antibody (scFV) probe (Sty3)[7] to show G4 formation at telomeres in the macronuclei of the ciliate *Stylonychia lemnae*[7]. The same antibody was used to elucidate the dynamic, formation and loss of telomeric G4 under the cooperative control of telomere-end-binding proteins, and a cell-cycle dependent phosphorylation of one of them[8]. Recently G4s have been visualised in human cells by immunofluorescence microscopy using two G4-specific antibodies BG4[9] and 1H6[10], each generated from separate labs, with the use of secondary and tertiary antibodies for signal sensitivity. These independent and complementary studies were each performed on *in situ* fixed nuclei and showed punctate staining of G4 in genomic DNA in the nuclei of a range of human cell lines. Cell synchronisation experiments

revealed cell cycle dependent G4 dynamics with the quantity of G4s reaching a maximum during the S-phase[9]. Immunofluorescence staining of metaphase chromosome spreads revealed G4s at telomeres with the majority of foci occurring away from the telomeres[9]. The number of detected G4-antibody *foci* increased after exposure of live human cells to G4 ligands, that include PDS, PhenDC3 and TMPyP4, demonstrating that such ligands do indeed trap out G4 structures once they form in cells (Figure 2b)[9,10]. The number of G4 foci in the presence of G4-stabilising ligand telomestatin in DT40 chicken cells was higher when FANCJ, a G4-specific helicase, was deficient, consistent with FANCJ controlling the susceptibility of G4s as molecular targets for ligands (Figure 2f)[10]. BG4 Immunofluorescence has shown colocalisation of human telomerase with a subset of endogenous telomeric G4 structures in cells, suggesting telomerase might be recruited to G4 to extend telomeres during meiosis[11] and presenting a alternative perspective to the earlier views that telomeric G4 structure may preclude telomerase recognition and action[12]. Synthetic small molecules that recognise G4s have also been employed to detect these DNA structures. A derivative of the G4-ligand Pyridostatin (PDS) called PDS-α enabled nuclear detection of G4s by bio-orthogonal ligation of a fluorophore to the ligand after cellular incubation and formaldehyde fixation[13]. PDS-α staining was significantly colocalised with the G4-helicase Pif1 in osteosarcoma (U2OS) cells, by super high-resolution spectroscopy, consistent with Pif1 processing of G4 structures in human cells[13]. The intrinsically fluorescent G4 ligands BMVC and DAOTA-M2, have also been used to visualise G4s suggesting higher G4 prevalence in some cancer cell lines compared to normal cells[14,15]. These studies have complemented earlier work that visualised accumulation of a radiolabeled small molecule G4 ligand at telomeres in human cells[16]. Be they antibodies or small molecules, it is a fundamental consequence that probes that bind to particular DNA structures can alter the intrinsic stability of those structures by the very act of binding. Thus, molecular probes can alter the apparent lifetimes of these dynamic structures, from their natural states. Probe-based observations of natural biological dynamics (e.g. during the cell cycle)[8,9] or perturbation experiments (by ligands or manipulation of key enzymes)[9,10,17], are helpful to visualise changes that are unlikely to be attributed to the binding effect of the probes.

***Genome mapping*.** A method for combining G4-dependent DNA polymerase stalling and next-generation sequencing, called G4-seq, has been developed to map G4 structures in purified, single-stranded DNA on a human genome scale[18]. Typically, genomic DNA isolated from cells, is sequenced first under conditions that do not favour G4 structure formation, then the same DNA fragments are re-sequenced under conditions that stabilize G4 structure formation, either by addition of K^+ or the G4-ligand PDS. G4-specific polymerase stalling is detected at specific sites during the second sequencing run by a precipitous loss of sequencing data quality, as compared to the first sequencing run. G4-seq identified over 700,000 G4s in the human genome, the majority (70%) of which comprised extra-long loops and/or bulges in their G-tracts, which precluded their prediction by earlier algorithms, e.g. ($G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$)[4]. Together with other studies[6], this suggests the breadth and number of potential G4s is greater than originally envisaged. The G4s were enriched in regulatory regions that included promoter, 5'UTR, splicing sites and were also overrepresented in cancer-related genes and in somatic copy number alterations (SCNAs) amplified in cancer genomes[18]. There are now a number of G4 predictor algorithms available that vary considerably in the details and the types of G4s that are captured. While, computational predictors and G4-seq provide a framework for understanding the potential for G4 structures to form in genomes, it is important to experimentally consider the profile and genome dynamics of G4 DNA in a biological context. A step towards this is to employ probes that bind to and enrich G4s from chromatin, followed by sequencing (e.g. Chromatin Immuno Precipitation Sequencing; ChIP-seq). An early attempt was to map the sites of DNA double-strand breaks (DSBs) induced by the G4-targeting ligand PDS in human immortalized fibroblast (MRC5-SV40) cells by ChIP-seq using an antibody for the DSB marker γH2AX in fixed chromatin[13]. A significant enrichment of DSBs was observed in particular genomic regions rich in computationally predicted G4 motifs consistent with the G4-ligand binding to G4 target structures and causing DSBs at those sites[13]. The binding sites of endogenous cellular proteins that can bind or resolve G4s *in vitro*, such as human ATRX[19] and XPB/ XPD[20] and yeast PIF-1 (an inactive mutant form)[21] and RIF-1[22], have been mapped by ChIP-seq to regions that comprise predicted G4 motifs that occur for example at telomeres and gene promoters (Figure 2g). Whilst the proteins that feature in such studies may also be capable of binding to other sequences

or structures, the data are consistent with hypotheses linking their biological functions to G4 structures or genomic regions that are enriched in G4 motifs. The mapping of G4 structures in chromatin was recently achieved using the G4 antibody BG4 as a G4 structure-specific ChIP-seq probe (G4 ChIP-seq) to map endogenous G4 structures in fixed chromatin in normal (NHEK) and spontaneously immortalized pre-cancerous (HaCaT) human epidermal keratinocyte cells (Figure 2e)[17]. In this study, about 10,000 and 1,000 G4s were detected in HaCaT and NHEKs, respectively, which is only ~1% of those identified by G4-seq, and by G4 predictors. This suggests G4 structure formation is largely suppressed in the context of chromatin, possibly due to chromatin-associated proteins and other proteins that control the duplex vs. non-duplex folded states of DNA. Most G4s observed were in regulatory, nucleosome-depleted chromatin regions that that were on average highly transcribed[17] and also significantly overlapped with G4 predicted sequences enriched in earlier ChIP-Seq mapping of the transcriptional helicases XPB/XPD[20]. Furthermore, endogenous G4s are enriched in promoter and 5'UTR regions of cancer-related genes and genes strongly associated with somatic copy number aberrations in cancer, such as *MYC*[23]. A perturbation experiment using the histone deacetylase inhibitor Entinostat, caused dynamic reprograming of the G4 landscape by G4 ChIP-seq with the loss and emergence of G4s showing, on average, a coupling to transcriptionally active chromatin[17]. The observation of G4 dynamics goes some way towards addressing the possibility that the G4s may be an artifact of antibody stabilization, which was discussed earlier in relation to antibody imaging of G4s. It will nonetheless be important to consider orthogonal approaches to detect G4 structures in chromatin and in cells to further consolidate these findings.

Biological Significance and Therapeutic Opportunities

Much of the early work on G4s focused on biophysical studies and functional studies on telomeres and telomerase. This section will focus primarily on some of the insights that recent imaging and G4 mapping has provided on the biology of G4s in non-telomeric regions. Such observations suggest a broader number of biological processes and the associated possibilities for therapeutic intervention.

Transcription, Replication and Intrinsic function(s). There have been a number of cellular studies describing G4 targeting ligands that alter levels of mRNA transcripts

for genes that have G4 motifs in their upstream (promoter) elements, for example the proto-oncogenes *MYC*[23] and *KRAS*[24]. Recent studies on zebrafish embryos demonstrated the use of G4-targeting small molecules or synthetic oligonucleotides to target conserved G4 motifs in promoters of developmental genes to lower transcription of the targeted genes and cause the expected phenotypic change[25]. Further work is needed to confirm and fully elucidate the mechanistic details of cause and effect in such studies. That genes physically targeted by the small molecule PDS, as judged by localised DNA damage, cause a concomitant reduction in transcript levels[13], suggests the relationship between G4 ligands and transcriptional changes at proximal genes may, at least for some cases, be more complex than a simple reversible binding mechanism. Recent data that includes G4 ChIP-seq[17], the genomic binding sites of proteins XPB, XPD[20] and SP1[17], the colocalization of G4-antibodies BG4 or 1H6 with transcriptionally active regions (marked by RNA polymerase II and H3K4me3)[17], support that G4 structures form in transcriptionally active chromatin in human cells. Dysfunctional mutations in WRN and BLM helicases cause altered regulation of genes that are enriched in predicted G4 motifs, consistent with a link between G4s and transcription[26-28]. It is noteworthy that in *D. mel.* G4 structure formation has actually been observed in the heterochromatin of polytene chromosomes during embryonic development, by immuno-gold labeling using 1H6 and microscopy[29], revealing differences in this very different genome, from mammalian systems. Overall, the weight of recent data from mammalian cells is consistent with a functional and dynamic link between G4 structures and transcription. Further studies are needed to elucidate the mechanistic details of this link, including the specific roles of proteins associated with transcription, such as SP1 or XPB/XPD, and their interaction(s) with loci where G4 structures have been observed to exhibit dynamic formation or unfolding.

DNA replication is a carefully regulated process and is initiated at many thousands of sites called DNA replication origins. Conserved DNA structures have been found at origin of replications sites in prokaryotes[30]. Recently, locations of human replication origins have been mapped via deep sequencing of short nascent strands and predicted to contain G4-motifs[31]. In addition, the human origin recognition complex (ORC) has been shown to bind G4-forming DNA and RNA sequences *in vitro*[32]. Such studies have led to the hypothesis that G4 DNA structures may somehow be involved in the

mechanism of initiating replication origins. Further studies that experimentally support G4 structure formation at origins more directly are necessary to advance our understanding of these findings.

Genomic Instability. In the absence of helicases that resolve G4 structures in DNA, stable G4 structures can pose an impediment for DNA polymerase progression, leading to replication stalling, DNA-damage and genomic instability (Figure 3a). Quantitative assays have been used to monitor G4 induced genome instability in *S. cerevisiae*, to show that Pif1 and Rrm3 helicases are essential to supress and prevent G4-induced DNA strand breakage[33,34]. The RTEL1 helicase has been shown to resolve telomeric G4 to maintain telomere integrity in mouse cells[35]. Similarly, FANCJ, BLM and WRN helicases have been shown to recognise and unwind G4 structures *in vitro*[36]. Mutations causing dysfunction in these helicases have been associated with premature ageing and predisposition to cancer development, though since the helicases also operate on duplex DNA the extent to which this is a G4-related effect must be further elucidated. Regulators of DNA synthesis, such as REV1 or PrimPol, affect gene expression by a mechanism proposed to ordinarily maintain epigenetic stability at replicated G4 DNA, which when compromised, e.g. by REV1 deficiency, leads to gene activation[37] or repression[38] depending on the location of the predicted G4-forming sequence.

Ligands that stabilise G4 structures can induce DNA breakage in humans[13] as well as insertion and deletions at predicted G4 motifs in yeast[33,34]. In *S. cerevisiae* the G4 ligand PhenDC3 triggers G4-induced[39] and G4-stability[40] dependent genomic instability, as measured by the increased genetic insertion and deletion at a level comparable to when Pif1 function is impaired. Interestingly, the yeast system provided a useful platform in which G4 stability could be systematically varied by mutagenesis to demonstrate a clear correlation between G4 stability and DNA instability at CEB-1 microsatellite, with short G4 loops (≤4 nt) causing higher levels of genome instability[40]. Given that G4-seq[18] and G4 ChIP-seq[17] both show G4s are enriched in SCNA-amplifications associated with cancer and that a similar association has been reported by computational analysis of SCNA associated breakpoints[18], it appears that G4 structures also represent vulnerable regions in the genome (Figure 3a). These associations are consistent with the immunohistochemistry observations from matched normal and cancerous gut/stomach tissues using BG4 revealed higher

apparent levels of G4 in the cancerous state[41]. This empirical data suggests, at least for some cases, that there are aspects of a cancer genome that exhibit more G4 structures, which immediately suggests G4s may have potential as both a cancer-biomarker and as a therapeutic target.

Therapeutic Opportunities. The data supporting concepts that link G4s with telomere biology, transcription regulation (of cancer genes) and trigger points for instability and DNA strand breakage, have stimulated rationales for targeting G4s with small molecules for therapeutics. Several small molecule ligands with high selectivity, as determined biophysically, for G4 relative to double-stranded DNA have been designed and evaluated for their therapeutic potential. Recent examples include the tetra-substituted naphthalene diimide MM41 caused an 80% decrease in tumour growth in MIA PaCa-2 pancreatic cancer xenograft[42]. While this may be explained in part by the accompanied strong reduction of *KRAS* and *BCL-2* gene expression, there may be other G4-related modalities that also contribute to its efficacy. The G4-ligands PDS and RHPS4 trigger an ATM-dependent DNA damage response (DDR), DNA double strand breaks (DSB) and activation of DNA repair pathways, such as Homologous Recombination (HR) and Non Homologous End Joining (NHEJ), and signalling of single-strand DNA breaks by the synthesis of poly ADP-ribose chains (PARs) by PAR protein (PARP)[13,43]. This has inspired the application of G4 ligands in combination with DNA-PK (NHEJ) or PARP inhibitors, as well as with G4 helicases (WRN) inhibitors for a greater effect (i.e. synergy) than the observed sum of the individual effects. Exposure of HeLa and U2OS cancer cells to the WRN inhibitor NSC-19630 sensitize the cells to the G4 ligand telomestatin, showing exacerbated S phase prolongation and DNA damage response[44]. Similarly, treatment of HT29 human colon cancer xenografts with the G4 ligand RHPS4 in combination with the PARP1 inhibitor GPI resulted in a 50% reduction of tumor weight and an increase of 45% of mice survival, significantly higher of what could be obtained by treating the mice with either RHPS4 or GPI individually[43]. Equally, by inhibiting NHEJ repair with the DNA-PK inhibitor NU7441, a significant sensitization to the G-quadruplex ligand PDS can be observed in human HT1080 fibrosarcoma cells (~ 45% synergy, as calculated by a Bliss independent score model)[13,45]. Furthermore, HCT116 colon cancer cells deficient in HR ($BRCA2^{-/-}$) displayed a ~10 fold increase in sensitivity against PDS compared to their isogenic counterpart that is HR-proficient

(BRCA2$^{+/+}$)[45]. DNA repair deficiencies have been further demonstrated to stimulate sensitization to the G4 ligand PDS in DLD1 human cells BRCA2$^{-/-}$, as well as human HEK-293T subjected to knock-down of the DNA repair proteins BRCA1 and RAD51[45,46]. PDS sensitization is further retained in HR deficient cells after they have acquired resistance to the drug Olaparib, highlighting the potential of G4 ligands as therapeutic agents against HR compromised tumors with acquired drug resistance (Figure 3b)[46]. These recent findings indicate clear potential for G4-ligands to be considered as cancer therapeutics especially for tumours genetically deficient in DNA-repair machinery such as HR[45,46].

Conclusions and Perspectives

Recent advances have provided a substantial body of new data to support the existence of G4 structures in the genomes of human cells. There is now more explicit experimental data to show G4s form throughout the human genome and in regulatory regions, largely aligned to previous computational predictions. These findings have provided new insights into the fundamental biology of G4 structures, suggesting roles in marking regulatory chromatin, whilst also being hotspots for genome instability particularly in cases where there are specific genetic/functional deficiencies. Fundamental insights into endogenous G4 function(s) enabled by advances in experiments with probe molecules that bind G4s suggest rationales for therapeutic strategies against cancer that may provide the window of selectivity that would be required for future clinical development. Whilst there is much more to be understood about the mechanistic details relating to DNA G4s in biology, the developments of the past five or so years would appear to have moved the field substantially closer to the realm of functional biology.

References:

1. Davis, J. T. G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed.* **43**, 668-698 (2004).

2. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acid Res*. **43**, 8627-8637 (2015).

3. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **11**, 770-780 (2012).

4. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acid Res*. **33**, 2908-2916 (2005).

5. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motif in human DNA. *Nucleic Acid Res*. **33**, 2901-2907 (2005).

6. Bedrat, A., Lacroix, L. & Mergny J-L. Re-evaluation of G-quadruplex proprensity with G4Hunter. *Nucleic Acid Res*. **44**, 1746-1759 (2016).

7. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J. & Plückthun, A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. USA.* **98**, 8572-8577 (2001).

8. Paeschke, K. Juranek, S., Simonsson, T. Hempel, A. Rhodes, D. & Lipps, H. J. Telomerase recruitment by the telomere end binding protein-β facilitates G-quadruplex DNA unfolding in ciliates. *Nat. Struct. Mol. Biol.* **15**, 598-604 (2008).

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

10. Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acid Res*. **42**, 860-869 (2014).

11. Moye, A. L. *et al.* Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **6**, 7643 (2015).

12. Zahler, A. M. Williamson, J. R. Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature.* **350**, 718-720 (1991).

13. Rodriguez, R. *et al.* Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301-310 (2012).

14. Huang W. C. *et al.* Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acid Res*. **43**, 10102-10113 (2015).

15. Shivalingam, A. *et al.* The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.* **6**, 8178 (2015).

16. Granotier, C. *et al.* Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acid Res*. **33**, 4182-4190 (2005).

17. Hänsel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267-1272 (2016).

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

19. Law, M. J. *et al.* ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell*. **143**, 335-336 (2010).

20. Gray. L.T., Vallur, A.C., Eddy, J. & Maizels N. G-quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **10**, 313-318 (2014).

21. Paeschke, K., Capra, J. A. & Zakian V. A. DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*. **145**, 678-691 (2011).

22. Kanoh, Y. *et al.* Rif1 binds to G-quadruplexes and suppresses replication over long distances. *Nat. Struct. Mol. Biol.* **22**, 889–897 (2015).

23. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. USA.* **99**, 11593-11598 (2002).

24. Cogoi, S. & Xodo, L. E. G-quadruplex formation within the promoter of the *KRAS* proto-oncogene and its effect on transcription. *Nucleic Acid Res*. **34**, 2536-2549 (2006).

25. David, A. P., Margarit, E., Domizi, P., Banchio, C., Armas, P. & Calcateraa, N. B. G-quadruplexes as novel cis-elements controlling transcription during embryonic development. *Nucleic Acid Res*. **43**, 4163-4173 (2015).

26. Johnson, J. E., Cao, K., Ryvkin, P., Wang, L. S. & Johnson F. B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acid Res*. **38**, 1114-1122 (2010).

27. Nguyen, G. H. *et al.* Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proc. Natl. Acad. Sci. USA.* **111**, 9905-9910 (2014).

28. Tang, W. *et al.* The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Hum. Mol. Genet.* **25**, 2060-2069 (2016).

29. Hoffmann, R. F. *et al.* Guanine quadruplex structures localize to heterochromatin *Nucleic Acid Res*. **43**, 152-163 (2015).

30. Eckdahl, T. T. & Anderson, J. N. Conserved DNA structures in origins of replication. *Nucleic Acid Res*. **18**, 1609-1612 (1990).

31. Besnard, E. *et al.* Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).

32. Hoshina, S. *et al.* Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J. Biol. Chem.* **288**, 30161-30171 (2013).

33. Ribeyre C. *et al.* The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet.* **5**, e1000475 (2009).

34. Paeschke, K. *et al.* Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**, 458-462 (2013).

35. Vannier, J. B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I., Ding, H. & Boulton S. J. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell*. **149**, 795-806 (2012).

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

37. Sarkies, P., Reams, C., Simpson, L. J. & Sale, J. E. Epigenetic Instability due to Defective Replication of Structured DNA. *Mol. Cell*, **40**, 703-713 (2010).

38. Schiavone, D. *et al.* Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J.* **33**, 2507-25020 (2014).

39. Piazza, A. *et al.* Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae. *Nucleic Acid Res*. **38**, 4337-4348 (2010).

40. Piazza, A. *et al.* Short loop length and high thermal stability determine genomic instability induced by G-quadruplex forming minisatellites. *EMBO J.* **34**, 1718-1734 (2015).

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

42. Ohnmacht, S.A. *et al.* A G-quadruplex-binding compound showing anti-tumour activity in an in vivo model for pancreatic cancer. *Sci. Rep.* **5**, 11385 (2015).

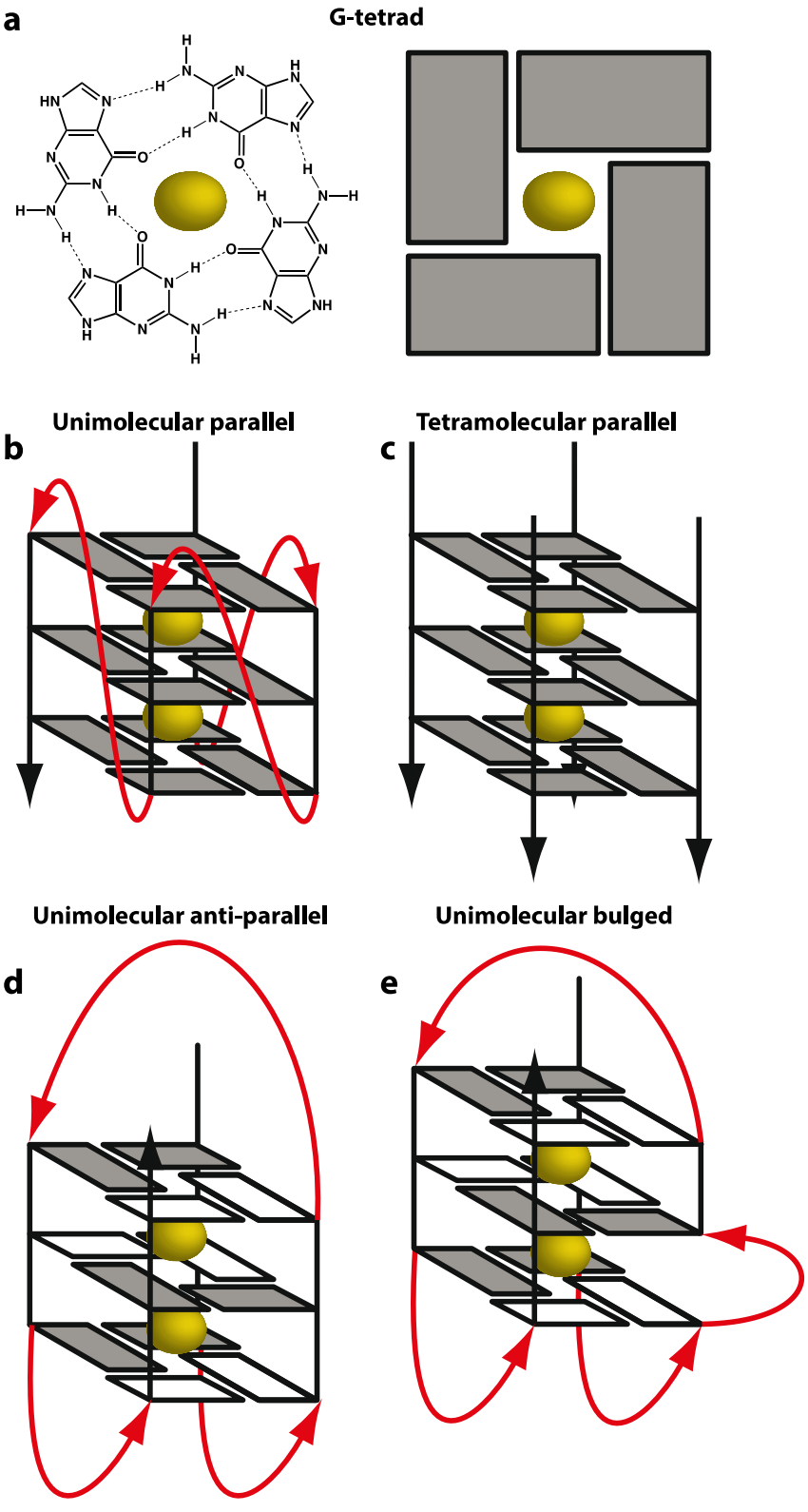
43. Salvati, E. *et al.* PARP1 is activated at telomeres upon G4 stabilization: possible target for telomere-based therapy. *Oncogene* **29,** 6280-6293 (2010).

44. Aggarwal, M., Sommers, J. A., Shoemaker, R. H. & Brosh, R. M. Jr. Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN)

function in the cellular response to DNA damage or replication stress. *Proc. Natl. Acad. Sci. USA.* **108**, 1525-1530 (2011).

45. McLuckie, K. I. *et al.* G-quadruplex DNA as a Molecular Target for Induced Synthetic Lethality in Cancer Cells. *J. Am. Chem. Soc.* **135**, 9640-9642 (2013).

46. Zimmer, J. *et al.* Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds. *Mol. Cell*, **61**, 449-460 (2016).



414

415

416 **Figure 1: G-quadruplex structures**. G-quadruplex structures can be generated from
417 one DNA strand (unimolecular) or multiple DNA strands coming together (e.g. bi- or
418 tetra molecular). G4 structures can be classified by the relative strand orientations:

parallel G4s have the same strand orientation within the structure whereas antiparallel G4s have alternating strand orientations. **a**, Structural (left) and schematic (right) representations of a G-tetrad that makes up the core of G-quadruplex structures. **b**, Schematic representation of unimolecular parallel G4s **c**, Schematic representation of a tetramolecular G4s . **d**, Schematic representation of an antiparallel intramolecular G-quadruplex structure **e**, Schematic representation of an antiparallel intramolecular G-quadruplex structure containing a bulge.

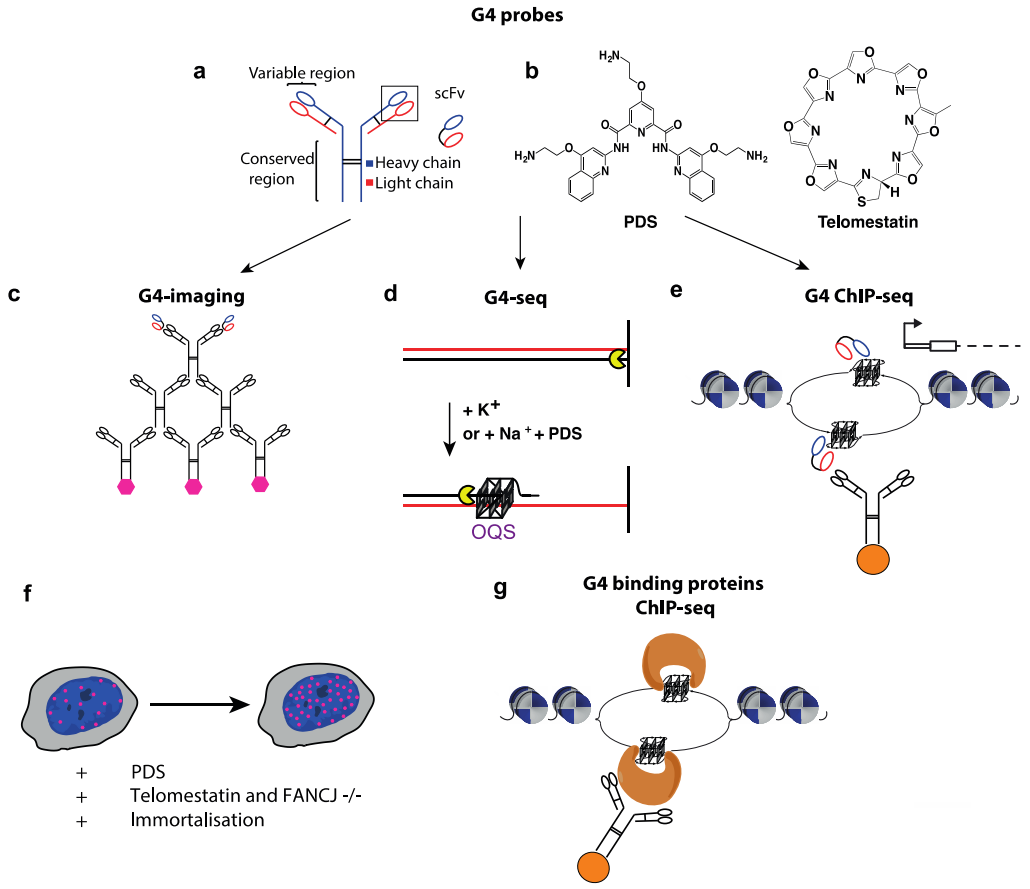


Figure 2: Visualisation and mapping of G-quadruplex structures. **a**, Schematic representation of a single chain antibody (scFv) as used to probe G4s, such as the BG4 or 1H6. **b**, Chemical structures of the selective G4 ligands pyridostatin (PDS) and telomestatin. **c**, Visualization of G4 structure can be achieved using G4 antibodies (e.g. BG4 or 1H6) together with secondary or tertiary antibodies that carry a fluorescent label. **d,** Schematic representation of the G4-seq method. DNA templates are sequenced a first time under non G4-stabilising conditions and a second time after

the addition of G4 stabilising agents (e.g. K^+ or PDS). Only DNA templates containing a G4 forming sequence will cause stalling of sequencing polymerase under G4-stabilising conditions, enabling selective detection of G4-forming genomic sequences. **e**, Schematic representation of G4 ChIP-seq: isolated chromatin is immuno-precipitated with BG4 and G4 structures detectable in chromatin are enriched and detected by sequencing. **f**, BG4 and 1H6 *foci* (red) detected in the nucleus of human cells (blue) are markedly increase in upon treatment with pyridostatin (PDS), telomestatin and FANCJ knock-down, and after immortalisation of normal human epidermal keratinocytes. **g.** Schematic representation of a typical ChIP-Seq of endogenous G4 binding proteins: isolated chromatin is immuno-precipitated using a selective antibody against the protein of interest (e.g. ATRX, PIF1, XPB/XPD). DNA sequences associated with those proteins are detected by sequencing.

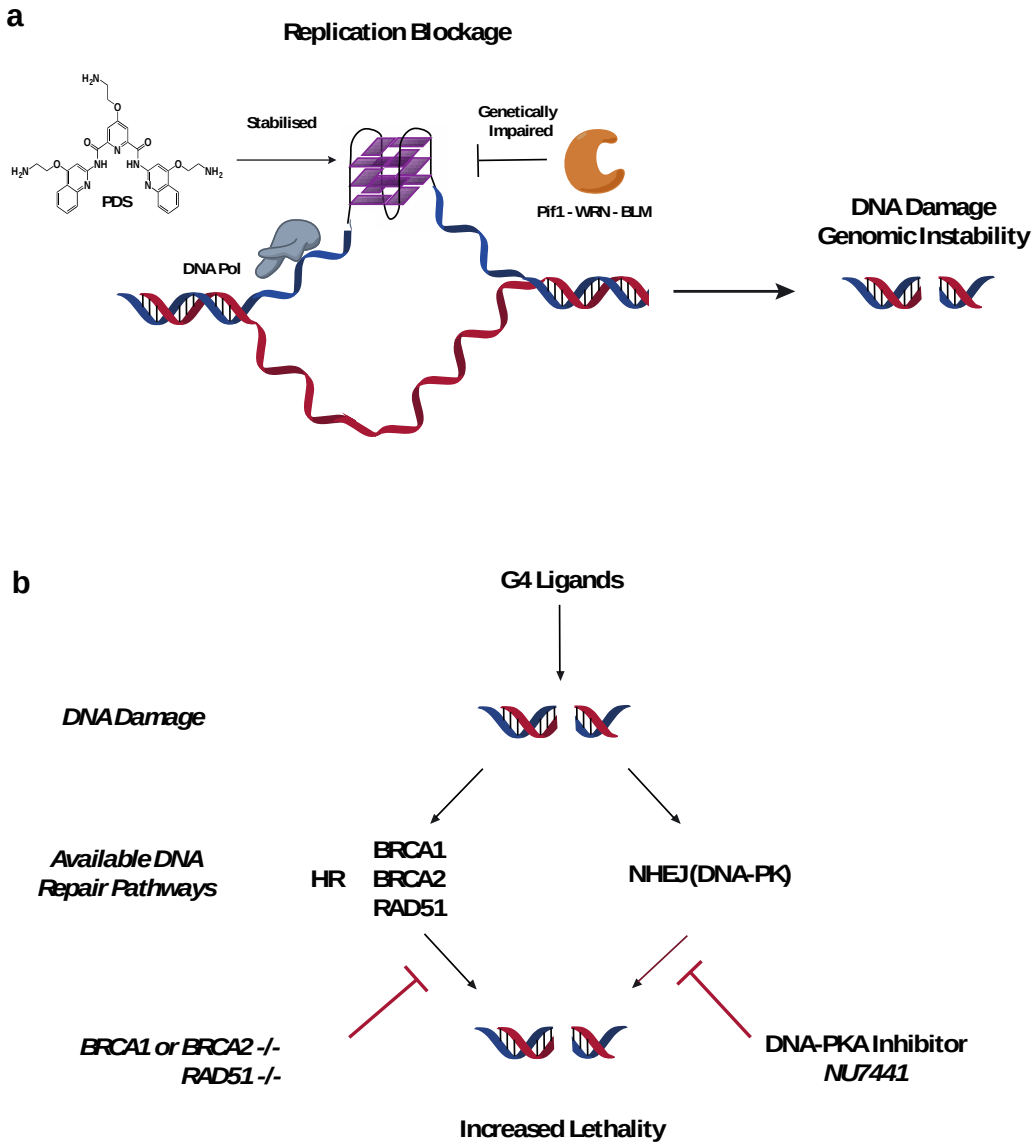


Figure 3: Therapeutic opportunities. a, Schematic representation of DNA damage response and genomic instability events that can be triggered by DNA G-quadruplexes either by stabilization with small molecules or by impairment of helicases that resolve G-quadruplexes. **b**, G-quadruplex ligands have been explored for their potential as cancer therapeutic agents. Representative scheme illustrating one the possible rationale behind the use of G4 ligands in combined therapies. DNA damage is triggered by exposure to a G4 ligand: sensitivity to ligand exposure can be obtained in cells genetically impaired in BRCA1/2 and RAD51, which regulate one of the two DNA repair pathway (HR). Selective killing of BRCA1/2 and RAD51 impaired cells can be achieved by combined treatment with G4-ligands and a

chemical inhibitor of the kinase DNA-PK that regulates the alternative DNA repair pathway (NHEJ).