

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

3
4 Robert Hänsel-Hertsch¹, Marco Di Antonio^{1,2} Shankar Balasubramanian^{1,2,3,*}

5
6 1. Cancer Research UK, Cambridge Institute, Li Ka Shing Centre, Cambridge CB2
7 0RE, UK.

8 2. Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.

9 3. School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK.

10
11 ***correspondence should be addressed to S.B. (sb10031@cam.ac.uk).**

12
13 **Abstract**

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

22
23
24 **Introduction**

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

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

39 **Imaging and Mapping G4s**

40
41
42 Computational G4 predictions using simple algorithms have suggested that over
43 300,000 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
44 genome have the potential to form a G4 structure^{4,5}. A more recent algorithm has
45 predicted the number of potential G4 sequences to be substantially higher⁶. These
46 computational studies showed that G4 motifs are enriched in telomeres, promoters
47 and the first intron of genes, but have highlighted the need to generate explicit
48 experimental data about the existence and function(s) of G4s in biologically relevant
49 contexts. G4-selective probes have been developed and employed to capture G4s in
50 cells by fluorescence microscopy and also by DNA chromatin immuno-precipitation
51 followed by sequencing.

52
53 **Cellular visualisation.** One approach to visualise particular DNA structures in cells is
54 to employ structure-selective molecular probes. Antibody proteins can have exquisite
55 specificity in their recognition of molecular structures and are widely used to bind to
56 and visualise proteins within cells or map their binding sites in DNA or RNA.
57 Antibodies can be generated by immunisation or *via in vitro* affinity selection to
58 recognise a particular DNA structure or chemical feature. The first reported
59 visualisation of G4 formation in a biologically relevant context used a G4-selective
60 single-chain antibody (scFV) probe (Sty3)⁷ to show G4 formation at telomeres in the
61 macronuclei of the ciliate *Stylonychia lemnae*⁷. The same antibody was used to
62 elucidate the dynamic, formation and loss of telomeric G4 under the cooperative
63 control of telomere-end-binding proteins, and a cell-cycle dependent phosphorylation
64 of one of them⁸. Recently G4s have been visualised in human cells by
65 immunofluorescence microscopy using two G4-specific antibodies BG4⁹ and 1H6¹⁰,
66 each generated from separate labs, with the use of secondary and tertiary antibodies
67 for signal sensitivity. These independent and complementary studies were each
68 performed on *in situ* fixed nuclei and showed punctate staining of G4 in genomic
69 DNA in the nuclei of a range of human cell lines. Cell synchronisation experiments

70 revealed cell cycle dependent G4 dynamics with the quantity of G4s reaching a
71 maximum during the S-phase⁹. Immunofluorescence staining of metaphase
72 chromosome spreads revealed G4s at telomeres with the majority of foci occurring
73 away from the telomeres⁹. The number of detected G4-antibody *foci* increased after
74 exposure of live human cells to G4 ligands, that include PDS, PhenDC3 and TMPyP4,
75 demonstrating that such ligands do indeed trap out G4 structures once they form in
76 cells (Figure 2b)^{9,10}. The number of G4 foci in the presence of G4-stabilising ligand
77 telomestatin in DT40 chicken cells was higher when FANCI, a G4-specific helicase,
78 was deficient, consistent with FANCI controlling the susceptibility of G4s as
79 molecular targets for ligands (Figure 2f)¹⁰. BG4 Immunofluorescence has shown
80 colocalisation of human telomerase with a subset of endogenous telomeric G4
81 structures in cells, suggesting telomerase might be recruited to G4 to extend telomeres
82 during meiosis¹¹ and presenting an alternative perspective to the earlier views that
83 telomeric G4 structure may preclude telomerase recognition and action¹². Synthetic
84 small molecules that recognise G4s have also been employed to detect these DNA
85 structures. A derivative of the G4-ligand Pyridostatin (PDS) called PDS- α enabled
86 nuclear detection of G4s by bio-orthogonal ligation of a fluorophore to the ligand after
87 cellular incubation and formaldehyde fixation¹³. PDS- α staining was significantly
88 colocalised with the G4-helicase Pif1 in osteosarcoma (U2OS) cells, by super high-
89 resolution spectroscopy, consistent with Pif1 processing of G4 structures in human
90 cells¹³. The intrinsically fluorescent G4 ligands BMVC and DAOTA-M2, have also
91 been used to visualise G4s suggesting higher G4 prevalence in some cancer cell lines
92 compared to normal cells^{14,15}. These studies have complemented earlier work that
93 visualised accumulation of a radiolabeled small molecule G4 ligand at telomeres in
94 human cells¹⁶. Be they antibodies or small molecules, it is a fundamental consequence
95 that probes that bind to particular DNA structures can alter the intrinsic stability of
96 those structures by the very act of binding. Thus, molecular probes can alter the
97 apparent lifetimes of these dynamic structures, from their natural states. Probe-based
98 observations of natural biological dynamics (e.g. during the cell cycle)^{8,9} or
99 perturbation experiments (by ligands or manipulation of key enzymes)^{9,10,17}, are
100 helpful to visualise changes that are unlikely to be attributed to the binding effect of
101 the probes.

102

103 **Genome mapping.** A method for combining G4-dependent DNA polymerase stalling
104 and next-generation sequencing, called G4-seq, has been developed to map G4
105 structures in purified, single-stranded DNA on a human genome scale¹⁸. Typically,
106 genomic DNA isolated from cells, is sequenced first under conditions that do not
107 favour G4 structure formation, then the same DNA fragments are re-sequenced under
108 conditions that stabilize G4 structure formation, either by addition of K⁺ or the G4-
109 ligand PDS. G4-specific polymerase stalling is detected at specific sites during the
110 second sequencing run by a precipitous loss of sequencing data quality, as compared
111 to the first sequencing run. G4-seq identified over 700,000 G4s in the human genome,
112 the majority (70%) of which comprised extra-long loops and/or bulges in their G-
113 tracts, which precluded their prediction by earlier algorithms, e.g. (G_{≥3}N₁₋₇G_{≥3}N₁₋₇
114 G_{≥3}N₁₋₇G_{≥3})⁴. Together with other studies⁶, this suggests the breadth and number of
115 potential G4s is greater than originally envisaged. The G4s were enriched in
116 regulatory regions that included promoter, 5'UTR, splicing sites and were also
117 overrepresented in cancer-related genes and in somatic copy number alterations
118 (SCNAs) amplified in cancer genomes¹⁸. There are now a number of G4 predictor
119 algorithms available that vary considerably in the details and the types of G4s that are
120 captured. While, computational predictors and G4-seq provide a framework for
121 understanding the potential for G4 structures to form in genomes, it is important to
122 experimentally consider the profile and genome dynamics of G4 DNA in a biological
123 context. A step towards this is to employ probes that bind to and enrich G4s from
124 chromatin, followed by sequencing (e.g. Chromatin Immuno Precipitation
125 Sequencing; ChIP-seq). An early attempt was to map the sites of DNA double-strand
126 breaks (DSBs) induced by the G4-targeting ligand PDS in human immortalized
127 fibroblast (MRC5-SV40) cells by ChIP-seq using an antibody for the DSB marker
128 γH2AX in fixed chromatin¹³. A significant enrichment of DSBs was observed in
129 particular genomic regions rich in computationally predicted G4 motifs consistent
130 with the G4-ligand binding to G4 target structures and causing DSBs at those sites¹³.
131 The binding sites of endogenous cellular proteins that can bind or resolve G4s *in vitro*,
132 such as human ATRX¹⁹ and XPB/ XPD²⁰ and yeast PIF-1 (an inactive mutant form)²¹
133 and RIF-1²², have been mapped by ChIP-seq to regions that comprise predicted G4
134 motifs that occur for example at telomeres and gene promoters (Figure 2g). Whilst the
135 proteins that feature in such studies may also be capable of binding to other sequences

136 or structures, the data are consistent with hypotheses linking their biological functions
137 to G4 structures or genomic regions that are enriched in G4 motifs. The mapping of
138 G4 structures in chromatin was recently achieved using the G4 antibody BG4 as a G4
139 structure-specific ChIP-seq probe (G4 ChIP-seq) to map endogenous G4 structures in
140 fixed chromatin in normal (NHEK) and spontaneously immortalized pre-cancerous
141 (HaCaT) human epidermal keratinocyte cells (Figure 2e)¹⁷. In this study, about 10,000
142 and 1,000 G4s were detected in HaCaT and NHEKs, respectively, which is only ~1%
143 of those identified by G4-seq, and by G4 predictors. This suggests G4 structure
144 formation is largely suppressed in the context of chromatin, possibly due to
145 chromatin-associated proteins and other proteins that control the duplex vs. non-
146 duplex folded states of DNA. Most G4s observed were in regulatory, nucleosome-
147 depleted chromatin regions that that were on average highly transcribed¹⁷ and also
148 significantly overlapped with G4 predicted sequences enriched in earlier ChIP-Seq
149 mapping of the transcriptional helicases XPB/XPD²⁰. Furthermore, endogenous G4s
150 are enriched in promoter and 5'UTR regions of cancer-related genes and genes
151 strongly associated with somatic copy number aberrations in cancer, such as *MYC*²³.
152 A perturbation experiment using the histone deacetylase inhibitor Entinostat, caused
153 dynamic reprogramming of the G4 landscape by G4 ChIP-seq with the loss and
154 emergence of G4s showing, on average, a coupling to transcriptionally active
155 chromatin¹⁷. The observation of G4 dynamics goes some way towards addressing the
156 possibility that the G4s may be an artifact of antibody stabilization, which was
157 discussed earlier in relation to antibody imaging of G4s. It will nonetheless be
158 important to consider orthogonal approaches to detect G4 structures in chromatin and
159 in cells to further consolidate these findings.

160

161 **Biological Significance and Therapeutic Opportunities**

162

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

168

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

171 for genes that have G4 motifs in their upstream (promoter) elements, for example the
172 proto-oncogenes *MYC*²³ and *KRAS*²⁴. Recent studies on zebrafish embryos
173 demonstrated the use of G4-targeting small molecules or synthetic oligonucleotides to
174 target conserved G4 motifs in promoters of developmental genes to lower
175 transcription of the targeted genes and cause the expected phenotypic change²⁵.
176 Further work is needed to confirm and fully elucidate the mechanistic details of cause
177 and effect in such studies. That genes physically targeted by the small molecule PDS,
178 as judged by localised DNA damage, cause a concomitant reduction in transcript
179 levels¹³, suggests the relationship between G4 ligands and transcriptional changes at
180 proximal genes may, at least for some cases, be more complex than a simple
181 reversible binding mechanism. Recent data that includes G4 ChIP-seq¹⁷, the genomic
182 binding sites of proteins XPB, XPD²⁰ and SP1¹⁷, the colocalization of G4-antibodies
183 BG4 or 1H6 with transcriptionally active regions (marked by RNA polymerase II and
184 H3K4me3)¹⁷, support that G4 structures form in transcriptionally active chromatin in
185 human cells. Dysfunctional mutations in WRN and BLM helicases cause altered
186 regulation of genes that are enriched in predicted G4 motifs, consistent with a link
187 between G4s and transcription²⁶⁻²⁸. It is noteworthy that in *D. mel.* G4 structure
188 formation has actually been observed in the heterochromatin of polytene
189 chromosomes during embryonic development, by immuno-gold labeling using 1H6
190 and microscopy²⁹, revealing differences in this very different genome, from
191 mammalian systems. Overall, the weight of recent data from mammalian cells is
192 consistent with a functional and dynamic link between G4 structures and
193 transcription. Further studies are needed to elucidate the mechanistic details of this
194 link, including the specific roles of proteins associated with transcription, such as SP1
195 or XPB/XPD, and their interaction(s) with loci where G4 structures have been
196 observed to exhibit dynamic formation or unfolding.

197

198 DNA replication is a carefully regulated process and is initiated at many thousands of
199 sites called DNA replication origins. Conserved DNA structures have been found at
200 origin of replications sites in prokaryotes³⁰. Recently, locations of human replication
201 origins have been mapped via deep sequencing of short nascent strands and predicted
202 to contain G4-motifs³¹. In addition, the human origin recognition complex (ORC) has
203 been shown to bind G4-forming DNA and RNA sequences *in vitro*³². Such studies
204 have led to the hypothesis that G4 DNA structures may somehow be involved in the

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

208

209 **Genomic Instability.** In the absence of helicases that resolve G4 structures in DNA,
210 stable G4 structures can pose an impediment for DNA polymerase progression,
211 leading to replication stalling, DNA-damage and genomic instability (Figure 3a).
212 Quantitative assays have been used to monitor G4 induced genome instability in *S.*
213 *cerevisiae*, to show that Pif1 and Rrm3 helicases are essential to suppress and prevent
214 G4-induced DNA strand breakage^{33,34}. The RTEL1 helicase has been shown to
215 resolve telomeric G4 to maintain telomere integrity in mouse cells³⁵. Similarly,
216 FANCI, BLM and WRN helicases have been shown to recognise and unwind G4
217 structures *in vitro*³⁶. Mutations causing dysfunction in these helicases have been
218 associated with premature ageing and predisposition to cancer development, though
219 since the helicases also operate on duplex DNA the extent to which this is a G4-
220 related effect must be further elucidated. Regulators of DNA synthesis, such as REV1
221 or PrimPol, affect gene expression by a mechanism proposed to ordinarily maintain
222 epigenetic stability at replicated G4 DNA, which when compromised, e.g. by REV1
223 deficiency, leads to gene activation³⁷ or repression³⁸ depending on the location of the
224 predicted G4-forming sequence.

225 Ligands that stabilise G4 structures can induce DNA breakage in humans¹³ as well as
226 insertion and deletions at predicted G4 motifs in yeast^{33,34}. In *S. cerevisiae* the G4
227 ligand PhenDC3 triggers G4-induced³⁹ and G4-stability⁴⁰ dependent genomic
228 instability, as measured by the increased genetic insertion and deletion at a level
229 comparable to when Pif1 function is impaired. Interestingly, the yeast system
230 provided a useful platform in which G4 stability could be systematically varied by
231 mutagenesis to demonstrate a clear correlation between G4 stability and DNA
232 instability at CEB-1 microsatellite, with short G4 loops (≤ 4 nt) causing higher levels
233 of genome instability⁴⁰. Given that G4-seq¹⁸ and G4 ChIP-seq¹⁷ both show G4s are
234 enriched in SCNA-amplifications associated with cancer and that a similar association
235 has been reported by computational analysis of SCNA associated breakpoints¹⁸, it
236 appears that G4 structures also represent vulnerable regions in the genome (Figure
237 3a). These associations are consistent with the immunohistochemistry observations
238 from matched normal and cancerous gut/stomach tissues using BG4 revealed higher

239 apparent levels of G4 in the cancerous state⁴¹. This empirical data suggests, at least
240 for some cases, that there are aspects of a cancer genome that exhibit more G4
241 structures, which immediately suggests G4s may have potential as both a cancer-
242 biomarker and as a therapeutic target.

243

244 ***Therapeutic Opportunities.*** The data supporting concepts that link G4s with telomere
245 biology, transcription regulation (of cancer genes) and trigger points for instability
246 and DNA strand breakage, have stimulated rationales for targeting G4s with small
247 molecules for therapeutics. Several small molecule ligands with high selectivity, as
248 determined biophysically, for G4 relative to double-stranded DNA have been
249 designed and evaluated for their therapeutic potential. Recent examples include the
250 tetra-substituted naphthalene diimide MM41 caused an 80% decrease in tumour
251 growth in MIA PaCa-2 pancreatic cancer xenograft⁴². While this may be explained in
252 part by the accompanied strong reduction of *KRAS* and *BCL-2* gene expression, there
253 may be other G4-related modalities that also contribute to its efficacy. The G4-ligands
254 PDS and RHPS4 trigger an ATM-dependent DNA damage response (DDR), DNA
255 double strand breaks (DSB) and activation of DNA repair pathways, such as
256 Homologous Recombination (HR) and Non Homologous End Joining (NHEJ), and
257 signalling of single-strand DNA breaks by the synthesis of poly ADP-ribose chains
258 (PARs) by PAR protein (PARP)^{13,43}. This has inspired the application of G4 ligands
259 in combination with DNA-PK (NHEJ) or PARP inhibitors, as well as with G4
260 helicases (WRN) inhibitors for a greater effect (i.e. synergy) than the observed sum of
261 the individual effects. Exposure of HeLa and U2OS cancer cells to the WRN inhibitor
262 NSC-19630 sensitize the cells to the G4 ligand telomestatin, showing exacerbated S
263 phase prolongation and DNA damage response⁴⁴. Similarly, treatment of HT29
264 human colon cancer xenografts with the G4 ligand RHPS4 in combination with the
265 PARP1 inhibitor GPI resulted in a 50% reduction of tumor weight and an increase of
266 45% of mice survival, significantly higher of what could be obtained by treating the
267 mice with either RHPS4 or GPI individually⁴³. Equally, by inhibiting NHEJ repair
268 with the DNA-PK inhibitor NU7441, a significant sensitization to the G-quadruplex
269 ligand PDS can be observed in human HT1080 fibrosarcoma cells (~ 45% synergy, as
270 calculated by a Bliss independent score model)^{13,45}. Furthermore, HCT116 colon
271 cancer cells deficient in HR (*BRCA2*^{-/-}) displayed a ~10 fold increase in sensitivity
272 against PDS compared to their isogenic counterpart that is HR-proficient

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

282

283

284 **Conclusions and Perspectives**

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

299

300 **References:**

- 301 1. Davis, J. T. G-quartets 40 years later: from 5'-GMP to molecular biology and
302 supramolecular chemistry. *Angew. Chem. Int. Ed.* **43**, 668-698 (2004).
- 303 2. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology.
304 *Nucleic Acid Res.* **43**, 8627-8637 (2015).
- 305 3. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability
306 and function of G-quadruplex structures. *Nat. Rev. Genet.* **11**, 770-780 (2012).

- 307 4. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human
308 genome. *Nucleic Acid Res.* **33**, 2908-2916 (2005).
- 309 5. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex
310 sequence motif in human DNA. *Nucleic Acid Res.* **33**, 2901-2907 (2005).
- 311 6. Bedrat, A., Lacroix, L. & Mergny J-L. Re-evaluation of G-quadruplex propensity
312 with G4Hunter. *Nucleic Acid Res.* **44**, 1746-1759 (2016).
- 313 7. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J. & Plückthun, A. In
314 vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with
315 *Stylonychia lemnae* macronuclei. *Proc. Natl. Acad. Sci. USA.* **98**, 8572-8577 (2001).
- 316 8. Paeschke, K. Juranek, S., Simonsson, T. Hempel, A. Rhodes, D. & Lipps, H. J.
317 Telomerase recruitment by the telomere end binding protein- β facilitates G-
318 quadruplex DNA unfolding in ciliates. *Nat. Struct. Mol. Biol.* **15**, 598-604 (2008).
- 319 9. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative
320 visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182-186
321 (2013).
- 322 10. Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic*
323 *Acid Res.* **42**, 860-869 (2014).
- 324 11. Moye, A. L. *et al.* Telomeric G-quadruplexes are a substrate and site of
325 localization for human telomerase. *Nat. Commun.* **6**, 7643 (2015).
- 326 12. Zahler, A. M. Williamson, J. R. Cech, T. R. & Prescott, D. M. Inhibition of
327 telomerase by G-quartet DNA structures. *Nature.* **350**, 718-720 (1991).
- 328 13. Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative
329 DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301-310 (2012).
- 330 14. Huang W. C. *et al.* Direct evidence of mitochondrial G-quadruplex DNA by using
331 fluorescent anti-cancer agents. *Nucleic Acid Res.* **43**, 10102-10113 (2015).
- 332 15. Shivalingam, A. *et al.* The interactions between a small molecule and G-
333 quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat.*
334 *Commun.* **6**, 8178 (2015).
- 335 16. Granotier, C. *et al.* Preferential binding of a G-quadruplex ligand to human
336 chromosome ends. *Nucleic Acid Res.* **33**, 4182-4190 (2005).
- 337 17. Hänsel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory
338 chromatin. *Nat. Genet.* **48**, 1267-1272 (2016).

- 339 18. Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P. &
340 Balasubramanian, S. High-throughput sequencing of DNA G-quadruplex structures in
341 the human genome. *Nat. Biotechnol.* **33**, 877-881 (2015).
- 342 19. Law, M. J. *et al.* ATR-X syndrome protein targets tandem repeats and influences
343 allele-specific expression in a size-dependent manner. *Cell.* **143**, 335-336 (2010).
- 344 20. Gray, L.T., Vallur, A.C., Eddy, J. & Maizels N. G-quadruplexes are genomewide
345 targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **10**, 313-318
346 (2014).
- 347 21. Paeschke, K., Capra, J. A. & Zakian V. A. DNA replication through G-quadruplex
348 motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell.* **145**,
349 678-691 (2011).
- 350 22. Kanoh, Y. *et al.* Rif1 binds to G-quadruplexes and suppresses replication over
351 long distances. *Nat. Struct. Mol. Biol.* **22**, 889–897 (2015).
- 352 23. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for
353 a G-quadruplex in a promoter region and its targeting with a small molecule to repress
354 c-MYC transcription. *Proc. Natl. Acad. Sci. USA.* **99**, 11593-11598 (2002).
- 355 24. Cogoi, S. & Xodo, L. E. G-quadruplex formation within the promoter of the
356 *KRAS* proto-oncogene and its effect on transcription. *Nucleic Acid Res.* **34**, 2536-2549
357 (2006).
- 358 25. David, A. P., Margarit, E., Domizi, P., Banchio, C., Armas, P. & Calcaterra, N. B.
359 G-quadruplexes as novel cis-elements controlling transcription during embryonic
360 development. *Nucleic Acid Res.* **43**, 4163-4173 (2015).
- 361 26. Johnson, J. E., Cao, K., Ryvkin, P., Wang, L. S. & Johnson F. B. Altered gene
362 expression in the Werner and Bloom syndromes is associated with sequences having
363 G-quadruplex forming potential. *Nucleic Acid Res.* **38**, 1114-1122 (2010).
- 364 27. Nguyen, G. H. *et al.* Regulation of gene expression by the BLM helicase
365 correlates with the presence of G-quadruplex DNA motifs. *Proc. Natl. Acad. Sci.*
366 *USA.* **111**, 9905-9910 (2014).
- 367 28. Tang, W. *et al.* The Werner syndrome RECQ helicase targets G4 DNA in human
368 cells to modulate transcription. *Hum. Mol. Genet.* **25**, 2060-2069 (2016).
- 369 29. Hoffmann, R. F. *et al.* Guanine quadruplex structures localize to heterochromatin
370 *Nucleic Acid Res.* **43**, 152-163 (2015).
- 371 30. Eckdahl, T. T. & Anderson, J. N. Conserved DNA structures in origins of
372 replication. *Nucleic Acid Res.* **18**, 1609-1612 (1990).

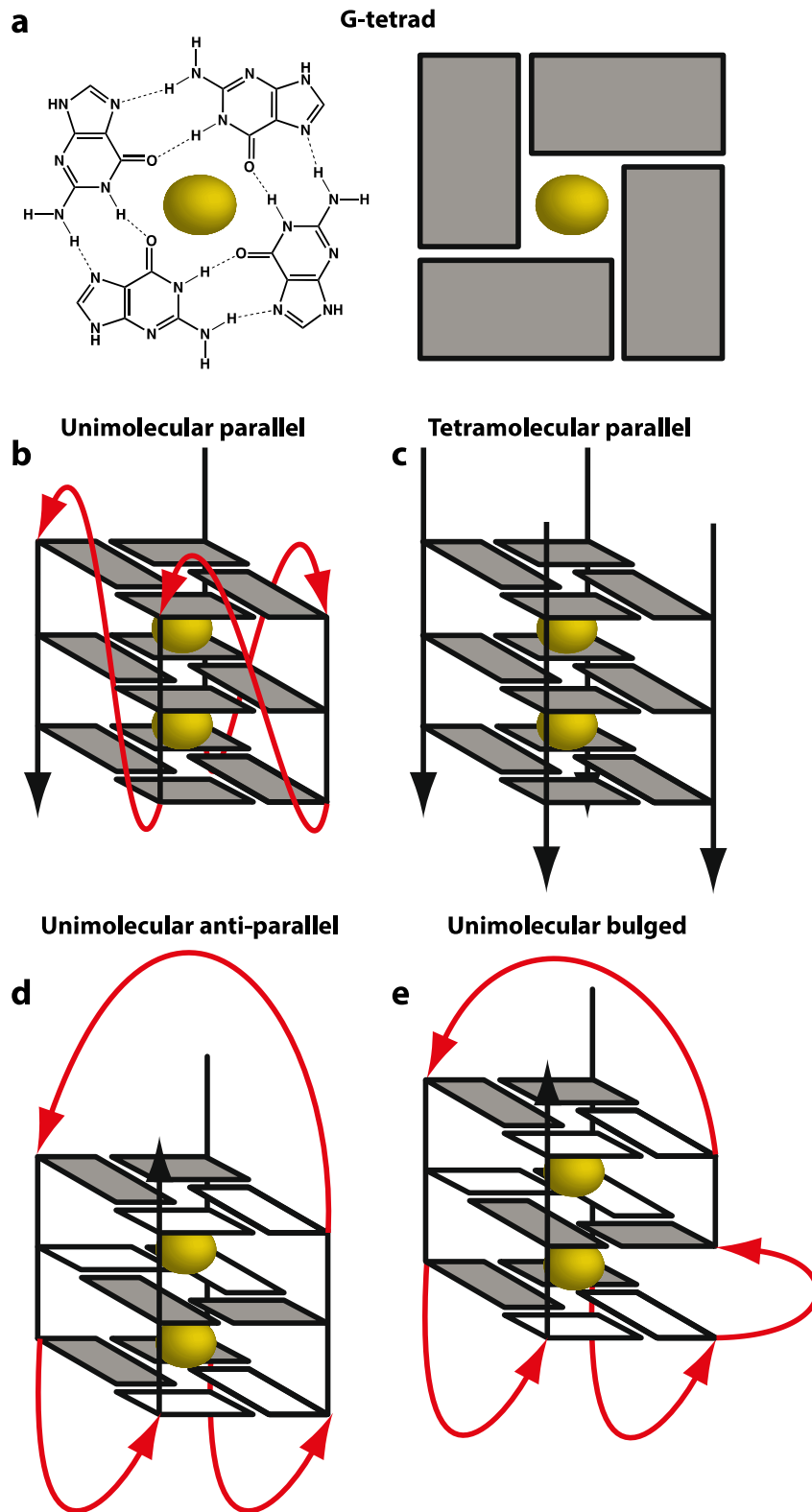
- 373 31. Besnard, E. *et al.* Unraveling cell type-specific and reprogrammable human
374 replication origin signatures associated with G-quadruplex consensus motifs. *Nat.*
375 *Struct. Mol. Biol.* **19**, 837–844 (2012).
- 376 32. Hoshina, S. *et al.* Human origin recognition complex binds preferentially to G-
377 quadruplex-preferable RNA and single-stranded DNA. *J. Biol. Chem.* **288**, 30161-
378 30171 (2013).
- 379 33. Ribeyre C. *et al.* The yeast Pif1 helicase prevents genomic instability caused by
380 G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet.* **5**, e1000475 (2009).
- 381 34. Paeschke, K. *et al.* Pif1 family helicases suppress genome instability at G-
382 quadruplex motifs. *Nature* **497**, 458-462 (2013).
- 383 35. Vannier, J. B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I., Ding, H. & Boulton S.
384 J. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain
385 telomere integrity. *Cell.* **149**, 795-806 (2012).
- 386 36. Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M. Jr. & Mergny, J-L. G-
387 quadruplexes and helicases. *Nucleic Acid Res.* **44**, 1989-2006 (2016).
- 388 37. Sarkies, P., Reams, C., Simpson, L. J. & Sale, J. E. Epigenetic Instability due to
389 Defective Replication of Structured DNA. *Mol. Cell*, **40**, 703-713 (2010).
- 390 38. Schiavone, D. *et al.* Determinants of G quadruplex-induced epigenetic instability
391 in REV1-deficient cells. *EMBO J.* **33**, 2507-25020 (2014).
- 392 39. Piazza, A. *et al.* Genetic instability triggered by G-quadruplex interacting Phen-
393 DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acid Res.* **38**, 4337-4348
394 (2010).
- 395 40. Piazza, A. *et al.* Short loop length and high thermal stability determine genomic
396 instability induced by G-quadruplex forming minisatellites. *EMBO J.* **34**, 1718-1734
397 (2015).
- 398 41. Biffi, G., Tannahill, D., Miller, J., Howat, W. J. & Balasubramanian, S. Elevated
399 levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLoS*
400 *One* **9**, e102711 (2014).
- 401 42. Ohnmacht, S.A. *et al.* A G-quadruplex-binding compound showing anti-tumour
402 activity in an in vivo model for pancreatic cancer. *Sci. Rep.* **5**, 11385 (2015).
- 403 43. Salvati, E. *et al.* PARP1 is activated at telomeres upon G4 stabilization: possible
404 target for telomere-based therapy. *Oncogene* **29**, 6280-6293 (2010).
- 405 44. Aggarwal, M., Sommers, J. A., Shoemaker, R. H. & Brosh, R. M. Jr. Inhibition of
406 helicase activity by a small molecule impairs Werner syndrome helicase (WRN)

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

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

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

413

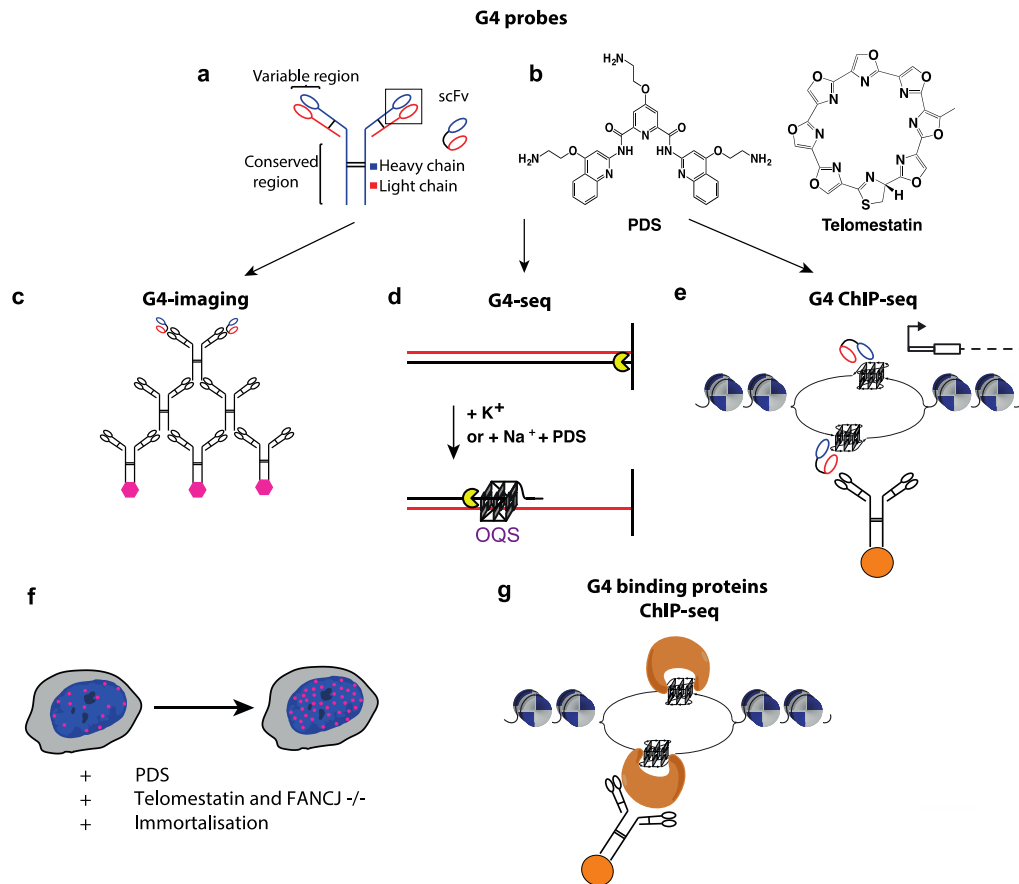


414

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:

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



426

427

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

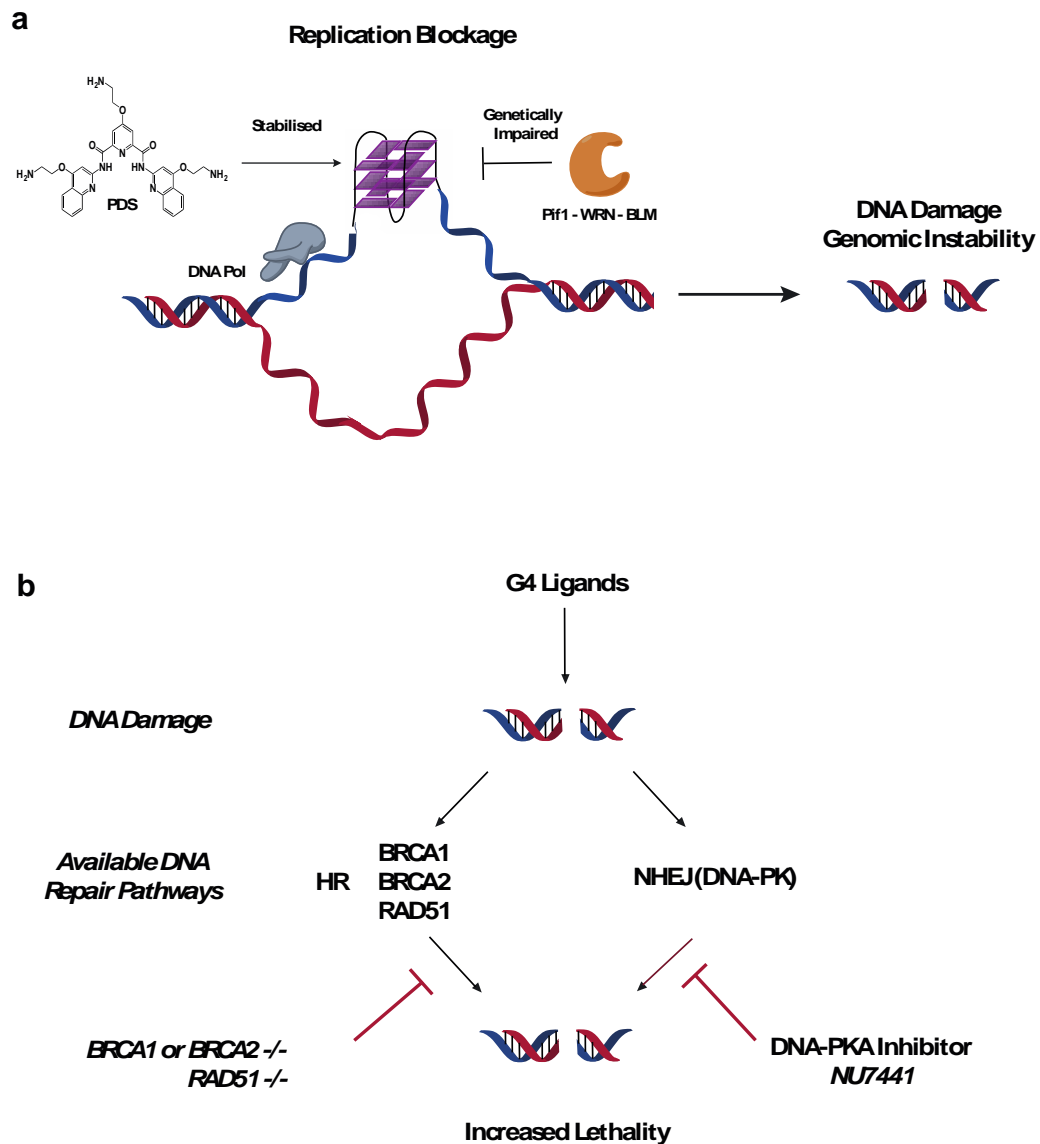
435 the addition of G4 stabilising agents (e.g. K^+ or PDS). Only DNA templates
436 containing a G4 forming sequence will cause stalling of sequencing polymerase under
437 G4-stabilising conditions, enabling selective detection of G4-forming genomic
438 sequences. **e.** Schematic representation of G4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

448

449

450

451



452

453

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

464 chemical inhibitor of the kinase DNA-PK that regulates the alternative DNA repair
465 pathway (NHEJ).
466