DNA secondary structures: stability and function of G-quadruplex structures

Matthew L. Bochman[1]*, Katrin Paeschke[2]* and Virginia A. Zakian[1]

Abstract | In addition to the canonical double helix, DNA can fold into various other inter- and intramolecular secondary structures. Although many such structures were long thought to be *in vitro* artefacts, bioinformatics demonstrates that DNA sequences capable of forming these structures are conserved throughout evolution, suggesting the existence of non-B-form DNA *in vivo*. In addition, genes whose products promote formation or resolution of these structures are found in diverse organisms, and a growing body of work suggests that the resolution of DNA secondary structures is critical for genome integrity. This Review focuses on emerging evidence relating to the characteristics of G-quadruplex structures and the possible influence of such structures on genomic stability and cellular processes, such as transcription.

B-form DNA
(B-DNA). The canonical right-handed double helical secondary structure assumed by bulk DNA *in vivo*.

Non-B-form secondary structures
Any DNA secondary structure that differs from B-form DNA. Such structures are likely to arise at defined sequence motifs owing to local factors acting on the B-form DNA.

[1]*Department of Molecular Biology, Princeton University, 101 Lewis Thomas Laboratory, Washington Rd., Princeton, New Jersey 08544, USA.*
[2]*Department of Biochemistry, University of Würzburg, Am Hubland, 97074 Würzburg, Germany.*
*These authors contributed equally to this work.
Correspondence to V.A.Z.
e-mail: vzakian@princeton.edu

The right-handed double helical structure of B-form DNA (B-DNA) has been known since 1953 (REF. 1). However, it has become increasingly clear that DNA can adopt a variety of alternative conformations based on particular sequence motifs and interactions with various proteins. These non-B-form secondary structures, which include G-quadruplex structures (G4 structures) (FIG. 1) as well as Z-DNA, cruciforms and triplexes (BOX 1), were originally characterized *in vitro* using biophysical techniques (for example, circular dichroism[2]). Accumulating evidence now points towards the existence of these structures under physiologically relevant conditions, and all of them are hypothesized, or even known, to have functional roles *in vivo*. The current wealth of genomic data — which is enabling the evolutionary comparison of motifs that can adopt non-B-form secondary structures *in vitro* — and the use of structure-specific antibodies, structure-binding ligands and clever experimental techniques are driving progress in this field.

Although the high thermal stability of G4 structures — potentially an impediment to DNA trans-actions — has led to some scepticism concerning their *in vivo* relevance, interest in G4 structures has increased enormously in recent years owing to their unique physical properties and the presence of G-rich sequences in biologically functional regions of many genomes. For example, G-rich regions with the potential to form G4 structures (hereafter called G4 motifs)

are over-represented in telomeres, mitotic and meiotic double-strand break (DSB) sites, and transcriptional start sites (TSSs; often near promoters). These findings suggest multiple roles for G4 structures. Moreover, recent work suggests that failure to resolve non-canonical DNA structures makes the sequence motifs capable of forming structural hotspots for genomic instability.

We begin this Review with an overview of G4 DNA structures, including their *in vitro* characterization and chromosomal locations in diverse organisms. Next, we discuss the putative roles of G4 structures at telomeres, during DNA replication, in gene regulation and in various other biological processes. Finally, we conclude by summarizing outstanding questions in the field and suggesting possible ways to address these issues.

Overview of G4 DNA

Biochemical characteristics and **in vitro** *analyses.* G4 structures are stacked nucleic acid structures that can form within specific repetitive G-rich DNA or RNA sequences (reviewed in REF. 3). In 1910, Bang[4] was the first to report the fact that guanylic acid forms a gel at high concentrations, which suggested that G-rich sequences in DNA may form higher-order structures. Fifty years later, Gellert and colleagues[5] used X-ray diffraction to demonstrate that guanylic acids can assemble into tetrameric structures. In these tetramers, four guanine molecules form a square planar arrangement
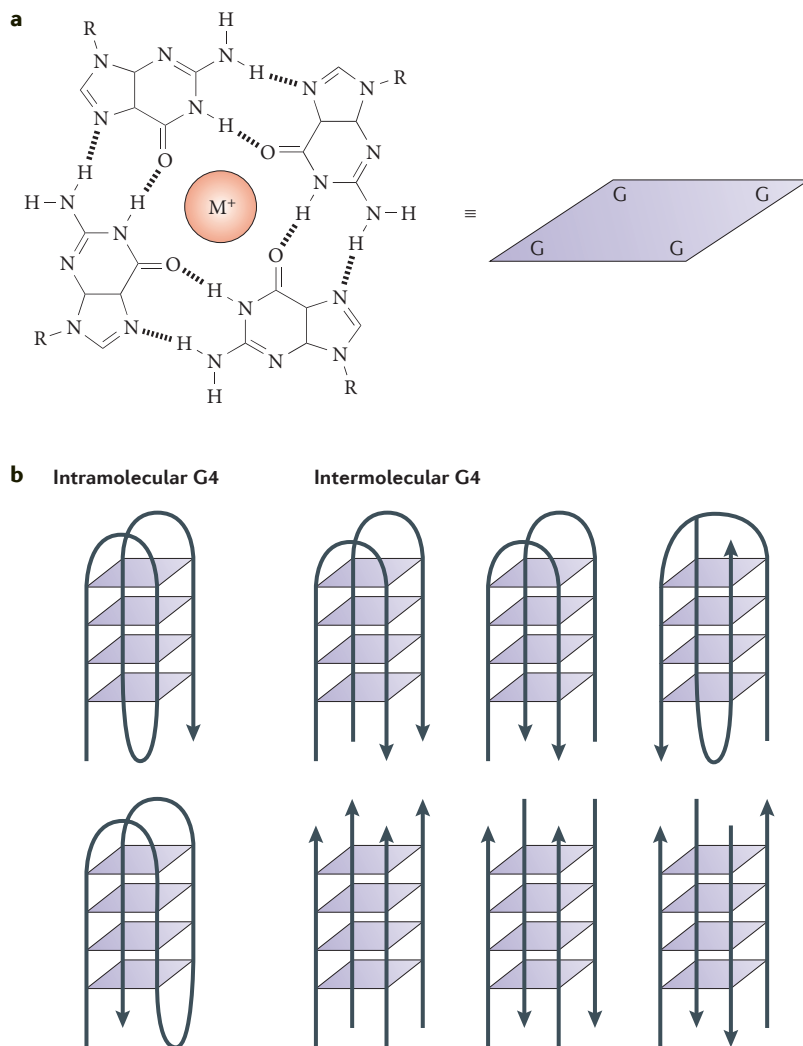
Figure 1 | **G-quadruplex DNA. a** | An illustration of the interactions in a G-quartet. This quartet is represented schematically as a square in the other panels of this figure. M[+] denotes a monovalent cation. **b** | Schematic diagrams of intramolecular (left) and intermolecular (right) G-quadruplex (G4) DNA structures. The arrowheads indicate the direction of the DNA strands. The intermolecular structures shown have two (upper) or four (lower) strands.

G-quadruplex structures
(G4 structures). Stable DNA secondary structures that can form from motifs containing tracts of tandem guanines. The guanines hydrogen bond in a planar arrangement, forming stacks connected by single-stranded DNA loops. The DNA strands can be parallel or antiparallel, and the G4 structures can form intra- or intermolecularly.

Z-DNA
Left-handed helical DNA that can form from tracts of alternating purines and pyrimidines.

in which each guanine is hydrogen bonded to the two adjacent guanines (that is, a G-quartet (FIG. 1a)). Stacked G-quartets form a G4 structure, and the intervening sequences are extruded as single-strand loops (although tetramolecular G4 structures may also lack loops). The sequence and size of the loop regions varies. However, loops are usually small (1–7 nucleotides (nt)), and smaller loops result in more stable G4 structures, as do longer G-tracts[3]. This structure is stabilized by monovalent cations that occupy the central cavities between the stacks, neutralizing the electrostatic repulsion of inwardly pointing guanine oxygens[6–8].

G4 structures adopt a variety of topologies and can be classified into various groups depending on the orientation of the DNA strands (FIG. 1b). Thus, G4 structures can be parallel, antiparallel or hybrids thereof. Furthermore, they can form within one strand (intramolecular) or

from multiple strands (intermolecular), and various loop structures are also possible[9,10]. G4 structures can be extremely stable, although the topology and stability of the G4 structure depends on many factors, including the length and sequence composition of the total G4 motif, the size of the loops between the guanines, strand stoichiometry and alignment[11–13], and the nature of the binding cations[14].

Chromosomal location of G4 motifs. Intramolecular G4 structures are predicted to form at specific G-rich regions *in vivo* that have in common a sequence motif with at least four runs of guanines (G-tracts), in which each G-tract most often contains at least three guanines ($G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$). G4 structures with only two stacks of guanines are possible but have low stability; here, when we refer to G4 motifs, we refer to motifs in which each G-tract contains three or more guanines. Computational analyses reveal that there are >375,000 G4 motifs in the human genome, whereas there are >1,400 G4 motifs in the *Saccharomyces cerevisiae* nuclear genome, including those in ribosomal and telomeric DNA, which are both particularly G4-rich[15–18]. Thus far, it is unclear how many of these motifs form stable G4 structures *in vivo* and, if they do, when they form.

Computational studies in various organisms have revealed that G4 motifs are not randomly located within genomes, but rather they tend to cluster in particular genomic regions (reviewed in REF. 19). In human, yeast and bacterial genomes, G4 motifs are similarly distributed and are over-represented in certain functional regions, such as promoters[15–18,20]. Furthermore, the locations and nucleotide compositions of G4 motifs are conserved in human populations and among related yeast species[15,21]. The nonrandom distribution of G4 motifs and the evolutionary conservation of their positions in genomes suggest that G4 motifs have one or more positive functions in the cell. In many organisms, telomeres contain a high concentration of G4 motifs owing to their high GC content and the single-stranded nature of the telomeric overhang. In diverse organisms, G4 DNA motifs are also common in G-rich micro- and minisatellites, up- and downstream of TSSs (often near promoters), within the ribosomal DNA, near transcription factor binding sites, and at preferred mitotic and meiotic DSB sites[15,17,18,21,22].

G4 structures at telomeres
Telomeres are nucleoprotein complexes at the ends of linear chromosomes. They are composed of a double-stranded region and a single-stranded G-rich 3′ overhang. Telomeres are essential to protect chromosomes from degradation, end-to-end fusions, and being recognized as DSBs[23]. In most telomeric DNAs, guanines and cytosines are distributed asymmetrically between the two DNA strands, with the G-rich strand running 5′ to 3′ from the centromere to the telomere. For example, vertebrate telomeric DNA consists of 5′-T_2AG_3-3′ repeats, whereas certain ciliated protozoans such as *Stylonychia lemnae* have 5′-T_4G_4-3′ repeats. Moreover, the G-rich strand is longer than its complement,

Box 1 | Other non-B-form DNA secondary structures

a

B-DNA   Z-DNA

b

Cruciform

c

Triplex
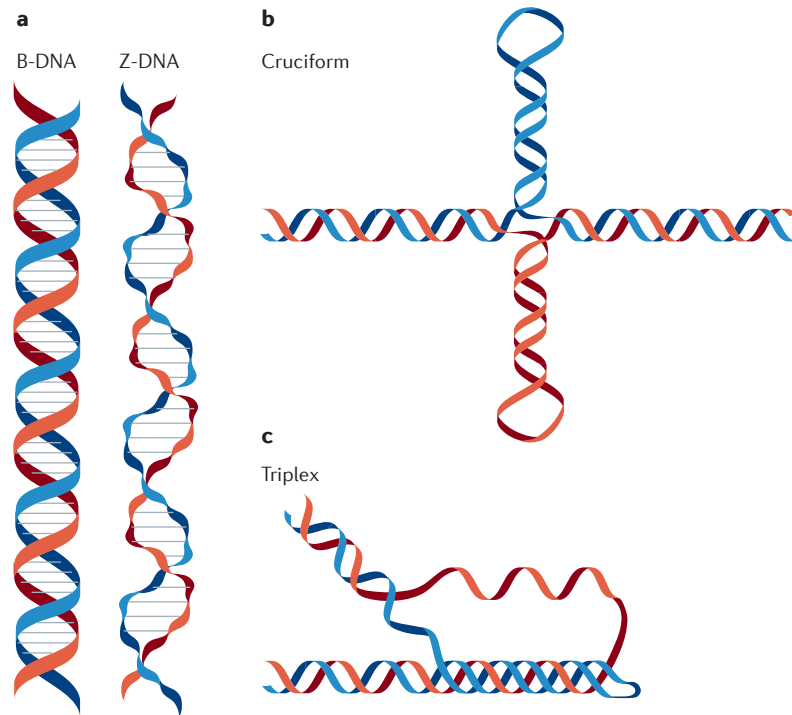
G-quadruplex (G4) structures are only one of many (ten or more) non-B-form DNA secondary structures analysed to date[127]. Brief descriptions of three well-studied structures are provided below.

Z-DNA

In contrast to standard B-form DNA (B-DNA), Z-DNA is a left-handed helix[128] (see the figure, part **a**). Z-DNA motifs (that is, sequences that form Z-DNA *in vitro*) are tracts of alternating purines and pyrimidines, which occur about once every 3,000 bp in metazoans[129]. Negative supercoiling stabilizes the formation of Z-DNA under physiological salt conditions[130], and it is hypothesized that Z-DNA relieves transcription-induced torsional stress[131]. Z-DNA motifs are tightly associated with transcriptional start sites in eukaryotic genomes[132], and these motifs can also cause genome instability, although the type of damage they cause varies from prokaryotes (dinucleotide insertions and deletions) to eukaryotes (double-strand breaks resulting in larger deletions)[120,121,133,134].

Cruciform structures

Negative supercoiling can also cause B-DNA to adopt a four-armed, cruciform secondary structure that resembles a Holliday junction[135] (see the figure, part **b**). These structures require ≥6-nucleotide inverted repeats (cruciform motif) to form, and such motifs are located near replication origins, breakpoint junctions and promoters in diverse organisms[136,137]. In metazoans, cruciform motifs are enriched near sites of gross chromosomal rearrangements[138], and deletions and translocations occur more frequently *in vivo* at sites of cruciform motifs than in B-DNA[139–141]. However, cruciforms might also serve positive roles (for example, stabilizing the human Y chromosome (reviewed in REF. 134)).

Triplex DNA

Three-stranded triplex DNA occurs when single-stranded DNA forms Hoogsteen hydrogen bonds in the major groove of purine-rich double-stranded B-DNA[142] (see the figure, part **c**). Triplexes in which the third strand is antiparallel to the DNA duplex can form at physiological pH, and these structures are stabilized by negative supercoiling[142]. Sequences capable of forming triplexes are common in eukaryotes but much rarer in prokaryotes[143]. In mammals, triplex-forming motifs are enriched in the introns of a variety of essential genes, including those involved in development and signalling[144]. Additionally, triplexes are hypothesized to cause genomic instability by causing double-strand breaks that result in translocations[145]. However, the formation of a triplex structure in a trinucleotide repeat sequence (for example, (CAG)n) can prevent the expansion of the repeat[138,139]; repeat expansion is related to human genetic disorders[146,147].

resulting in single-strand 'G-tails' at the very termini of chromosomes. Regardless of the precise sequence of the telomere, the G-rich strand of various telomeric sequences can usually form stable G4 structures *in vitro* (FIG. 2); for example, in non-denaturing polyacrylamide gels, oligonucleotides corresponding to the telomeric G-rich strand display unexpected banding patterns that are due to the formation of G4 structures[6,24–26].

Evidence for G4 structures at telomeres. The possibility that G4 structures might form *in vivo* is demonstrated by *in vitro* experiments showing that telomere structural proteins, such as TEBPα and TEBPβ in ciliates and Rap1 in *S. cerevisiae*, can promote the formation of G4 DNA[25,27–29]. By contrast, the human telomeric G-strand binding protein protection of telomeres protein 1 (POT1) promotes the unfolding of G4 structures *in vitro*[30,31]. Thus far, the most direct evidence that G4 structures exist at telomeres comes from studies in ciliates that exploit antibodies raised by ribosome display against parallel and antiparallel telomeric G_4T_4 structures. With these antibodies, it is possible to show that G4 structures exist *in vivo* at *Stylonychia lemnae* telomeres and to determine proteins that are required for their formation and unfolding[28,32,33]. Only the antibodies raised against antiparallel G4 structures bind to *S. lemnae* telomeres, indicating that antiparallel, and not parallel, G4 DNA is present *in vivo*[32]. In addition, several *in vivo* control experiments demonstrated that the anti-G4 antibodies do not induce the formation of G4 structures. The visualization of the regulation of G4 structures is an important observation because unresolved G4 structures are likely to be an obstacle for DNA replication and telomere elongation. Accordingly, telomeric G4 structures, which are present during most of the *S. lemnae* cell cycle, are resolved during DNA replication[32]. Further analysis using RNAi to silence gene expression indicates that the formation of telomeric G4 structures is dependent on two telomere binding proteins: TEBPα and TEBPβ. TEBPα binds to the telomeric overhang and recruits TEBPβ, which is able to promote the formation of G4 structures with its highly charged carboxyl terminus, as shown *in vitro*[27,28].

As stated above, G4 structures are not present at *S. lemnae* telomeres during S phase. *In vitro* and *in vivo* studies demonstrate that G4 unfolding is dependent on at least three conditions. First, TEBPβ, which is essential for the formation of G4 structures, must be removed from the telomeres. This removal happens during DNA replication and requires phosphorylation of TEBPβ. Second and third, immunofluorescence and gene knockdown analyses show that two enzymes, the telomerase holoenzyme and a RecQ family helicase, are recruited to telomeric G4 structures at the end of S phase and are essential for the unfolding of telomeric G4 structures[28,33–35]. Currently, it is not clear how or why telomerase is needed to unwind G4 structures during DNA replication nor whether this regulation is conserved among other organisms. However, RecQ helicases in other organisms, such as Sgs1 in *S. cerevisiae* and WRN and BLM in humans, also act on telomeres

and can unwind G4 structures *in vitro* (reviewed in REF. 36). To date, no-one has isolated antibodies against the human telomeric G4 structure, but the fact that TEBP homologues exist in vertebrates suggests that similar mechanisms might exist in higher eukaryotes.

There is also evidence for G4 DNA at telomeres in human cultured cells: BMVC (3,6-bis(1-methyl-4-vinylpyridinium) carbazole diiodide) is a fluorescent biomarker that binds and stabilizes G4 structures *in vitro*, and *in vivo* staining with BMVC marks the distal ends of metaphase chromosomes in human lung adenocarcinoma cells[37,38], suggesting telomeric binding. However, it is not clear whether this ligand detects G4 structures formed *in vivo* or whether it induces G4 DNA formation. Additional *in vivo* experiments are required to prove the specificity of such ligands.
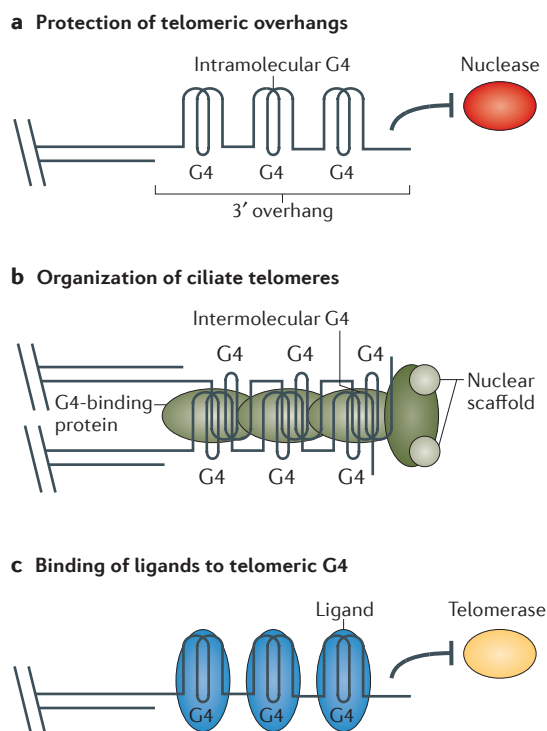
Possible consequences of G4 structures at telomeres. Owing to the biochemical properties of DNA polymerases, they cannot replicate the very ends of linear chromosomes. In most organisms, telomerase, a telomere-dedicated reverse transcriptase, uses its RNA subunit as a template to lengthen the G-strand of the telomere. Human telomerase is inactive in most somatic cells but is upregulated in most cancers, in which it is thought to promote the lifespan of malignant cells[39]. G4 structures influence telomerase activity: intramolecular antiparallel G4 structures block telomerase activity, whereas intermolecular parallel G4 DNA is permissive for extension by telomerase[40–42].

Because telomerase is active in most human cancers and this activity can be influenced by G4 structures, a variety of small molecule ligands with different specificities and target regions that bind and stabilize G4 structures are being tested in various assays[43]. The hope is that ligands that promote the formation of certain types of telomeric G4 structures might inhibit telomerase by preventing annealing of telomerase RNA to G-strand overhangs. For example, telomestatin has nanomolar affinity for telomeric G4 structures (which is nearly two orders of magnitude lower than its affinity for double-stranded DNA) and stabilizes intramolecular antiparallel G4 structures *in vitro*[44,45]. Moreover, telomestatin inhibits telomerase[46] and causes gradual telomere shortening and growth arrest or apoptosis in human tissue culture cancer cells[47–52]. However, telomeric DNA damage also increases in telomestatin-treated cells[50,53,54]. Thus, telomere shortening in telomestatin-treated cells might also be due to capping defects, especially as telomere repeat binding factor 2 (TRF2) and POT1 telomere binding are lost in these cells. Indeed, in *S. cerevisiae*, G4 structures are thought to contribute to telomere capping when natural capping is impaired[55]. Further research is required to determine whether G4 ligands are effective *in vivo*, whether they are specific for telomeric DNA and whether their presence has deleterious effects on non-telomeric G4 structures.

Effects of G4 structures on DNA replication

During DNA replication, the two strands of the DNA double helix are separated by the replicative helicase:

Cruciforms
Four-armed DNA secondary structures, similar to Holliday junctions, that can form at inverted repeat sequences and are stabilized by DNA supercoiling.
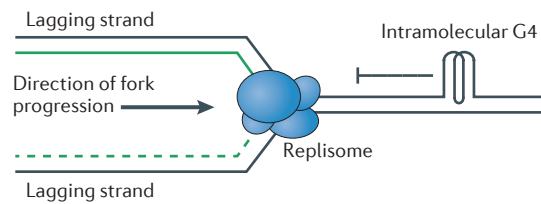
Triplexes
Three-stranded DNA in which single-stranded DNA hydrogen bonds into the major groove of purine-rich standard B-form DNA.

Telomeres
The ends of linear chromosomes, usually consisting of GC-rich repeated DNA, with guanines clustered in the strand that forms the 3′ end of the chromosome. The G-rich strand is longer than the C-rich strand so that telomeres contain both double- and single-stranded DNA. Sequence-specific binding proteins protect both duplex and single-stranded telomeric DNA from degradation, fusions and checkpoints.

Helicase
A class of enzymes that function as molecular motors, using the energy of ATP hydrolysis to unwind base-paired DNA or RNA. Helicases can also translocate along and displace proteins from nucleic acids.

a | Protection of telomeric overhangs

b | Organization of ciliate telomeres

c | Binding of ligands to telomeric G4

Figure 2 | **Putative functional roles of G-quadruplex structures at telomeres.** Telomeric sequences can fold into G-quadruplex (G4) structures *in vitro*. Currently, many groups are investigating the physiological relevance of this phenomenon. **a** | G4 structures may form at the telomeric 3′ overhang and have a role in protecting telomeres from degradation by nucleases (red) or other events. **b** | Work in ciliates shows that G4 structures do form at telomeres and have a role in telomere protection and tethering to the nuclear scaffold. Formation, stabilization and tethering is faciliated by G4-binding proteins (green). **c** | Ligands (blue) that bind to telomeric G4 structures are currently being analysed for their ability to influence telomere length by altering telomerase (yellow) activity.

one strand serves as the template for leading strand synthesis and the other for lagging strand synthesis. Although leading strand DNA replication can be continuous, the lagging strand is replicated discontinuously, making it transiently single-stranded; this is a conformation that provides opportunities for G4 structure formation. Thus, during DNA replication, G4 structures may form inappropriately, especially on the lagging strand template (FIG. 3), and this formation is more likely to occur when DNA replication is slowed. In addition, some G4 structures could be present during DNA replication because they have roles in transcriptional regulation (see below). Whether G4 structures are pre-existing or form during DNA replication, they must be resolved for completion of DNA replication because the sequence comprising the G4 structure cannot serve as a template until it is unfolded. Thus, helicases are likely to be necessary to unwind G4 structures.

Figure 3 | **Putative functional roles of G-quadruplex structures during DNA replication.** Computational studies show that in all tested organisms, many regions in the genome have the ability to form G-quadruplex (G4) structures. *In vitro* and *in vivo* studies indicate that unresolved G4 structures may influence DNA replication by slowing or stalling the replication fork machinery (replisome; blue).

We surveyed the literature and found that 22 different helicases have been tested for their ability to bind and/or unwind G4 structures *in vitro*, and all but one, the *Escherichia coli* RecBCD helicase, was positive (K.P., M.L.B. and V.A.Z., unpublished observations; summarized in Supplementary information S1 (table)). These data suggest that G4 unwinding is a non-specific activity of many DNA helicases. However, most of these unwinding studies are qualitative, and it is difficult to ascertain from them whether a given helicase is particularly effective at unwinding G4 structures and/or whether G4 structures are a preferred substrate for that helicase. Most of the human helicases that unwind G4 structures *in vitro*[56–60] are associated with human diseases that cause genomic instability, including the RecQ helicases WRN (associated with premature ageing) and BLM (associated with increased cancer risk) as well as FANCJ (associated with increased cancer risk) and PIF1 (associated with increased cancer risk). The best evidence that human disease is associated with loss of G4 unwinding comes from the finding that cell lines from human patients with Fanconi anaemia carrying FANCJ mutations display deletions that overlap G-rich regions with the potential to form G4 structures[56]. In addition, telomestatin, a chemical ligand that is able to stabilize G4 structures *in vitro*[53,61,62], causes impaired proliferation and increased apoptosis and DNA damage in FANCJ-deficient cells[63]. The association of these helicases with inherited genome instability has heightened interest in the possibility that G4 unwinding might suppress both premature ageing and cancer by regulating G4 structures.

Some enzymes are far more active on G4 structures than others. The *S. cerevisiae* Pif1 helicase acts at G4 motifs[64], and members of the Pif1 DNA helicase family are particularly efficient *in vitro* unwinders of parallel intramolecular G4 substrates[59]. Pif1 is a multifunctional DNA helicase that binds >1,000 sites in the genome of mitotic cells, of which ~10% overlap G4 motifs, which represents ~25% of the G4 motifs in this organism. Twenty-five per cent is likely to be an underestimate as, for technical reasons, this number excludes the large number of G4 motifs in ribosomal and telomeric DNA, both of which are strong Pif1 binding sites[64].

Several genetic assays show that in the absence of Pif1, DNA replication slows and DSBs occur at many of the G4 motifs that are normally bound by Pif1. G4 motifs also show a high mutation rate in Pif1-deficient cells, and these mutations eliminate the ability of the motif to form a G4 structure without necessarily reducing the high GC content of the motif. When these mutated motifs are put back in the genome, they no longer bind Pif1, slow DNA replication or cause DSBs. Together, these data make a strong argument that G4 structures form *in vivo* and that their resolution by Pif1 suppresses genome instability[64]. Other studies also found instability of G4 motifs in *pif1* cells[59,65]. This instability was particularly pronounced when the G4 motifs were on the template for leading strand synthesis, but this result may reflect the repetitive nature of the G4 substrate used in this analysis. The frequent mutation of G4 motifs in *pif1* mutant cells suggests the involvement of error-prone processes when G4 motifs are replicated and repaired in Pif1-deficient cells[64]. Indeed, in DT40 chicken cells, REV1, a translesion polymerase, is implicated in replication fork progression past G4 motifs on the leading strand[66].

There are also suggestions that human PIF1 acts at G4 motifs. One study used chromatin immunoprecipitation followed by sequencing (ChIP–seq) in combination with *in vivo* labelling with pyridostatin, a G4 binding molecule[67]. Genome-wide, pyridostatin bound preferentially to G4 motifs, where it caused replication and transcription-dependent damage that was detected by its high γH2Ax content. Many of the γH2Ax foci overlap with GFP–PIF1 foci in the pyridostatin-treated human cells. The current hypothesis is that G4 formation or stabilization blocks transcription and/or replication, resulting in DNA damage.

Similar to what is seen in cells from patients with Fanconi anaemia whose disease is due to mutations in the FANCJ helicase, mutations in the *Caenorhabditis elegans* DOG-1 helicase, which is distantly related to FANCJ, cause genome-wide deletions in G-rich sequences with the potential to form G4 structures[68,69]. The mutation rate in *dog-1* mutants is very high (up to 4% per generation[68]) and increases with the length of the G-tract[69]. Finally, the activity of regulator of telomere elongation helicase (RTEL) family helicases is also hypothesized to be directed towards G4 structures. Recent data indicate that the human RTEL helicase helps to resolve G4 DNA at telomeres, perhaps in conjunction with BLM, to ensure telomere stability[70]. Although biochemical evidence of G4 unwinding is lacking for RTEL homologues from other organisms, current data indicate that they may function similarly to human RTEL. For instance, *C. elegans rtel-1* has high sequence similarity to *dog-1*, although G-rich sequences are not unstable in worms deficient for *rtel-1* (REF. 71) as they are in *dog-1* mutant animals. However, mutation of *rtel-1* and *him-6* (a *BLM* homologue) is synthetically lethal in *C. elegans*, suggesting that RTEL-1 may function in concert with one or more additional helicases (DOG-1 and/or HIM-6) to resolve G4 structures.

γH2Ax
A phosphorylated histone H2A variant that accumulates at regions of DNA damage.
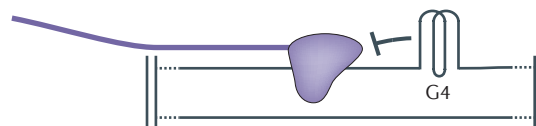
G4 structures in transcription

The high concentration of G4 motifs near promoter regions suggests a potential function of G4 structures in gene regulation. Indeed, one or more G4 motifs are found within 1,000 nt upstream of the TSS of 50% of human genes[72]. Intriguingly, bioinformatics show that the promoters of human oncogenes and regulatory genes (for example, transcription factors) are more likely than the average gene to contain G4 motifs, whereas G4 motifs are under-represented in the promoters of housekeeping and tumour suppressor genes[22,72]. A similar enrichment of G4 motifs in promoter regions is found in other organisms, including yeast, plants and bacteria[15,17,20,73,74]. Additionally, in humans, G4 motifs are less often found in the template strand than in the non-template strand. Those that are on the template strand tend to cluster at the 5′ end of the 5′UTR[75]. In yeast, there is no distinct asymmetry in G4 motif location between the non-template and template strands, but there is a correlation between nucleosome-free regions and G4 motifs in promoters[15], a finding that supports the prediction that G4 structures will form more easily in nucleosome-free regions[17]. Experiments in bacteria using a G4 motif on the non-template strand of a plasmid-borne transcribed gene demonstrate loop formation on the opposite strand of the G4 motif, suggesting the existence of G4 structures that form upon transcription in living cells[76]. Such structures may help to keep the transcribed template accessible for transcription by preventing it annealing to its complementary strand. In this way, G4 structures could contribute to high transcription levels of certain genes (FIG. 4).

Possible consequences of G4 structures formed during transcription. It is well known that supercoiling has both positive and negative effects on transcription[77], and G4 structures are thought to form as a result of supercoiling-induced stress during transcription[78]. *In vitro* studies show that the formation of G4 structures can compensate for the negative supercoiling[78,79]. These findings suggest that G4 structures in or near promoter regions may influence transcription in both positive and negative ways (FIG. 4). First, depending on which DNA strand encodes the G4 motif, the structure could either inhibit transcription (if the motif is on the template strand, blocking the transcription machinery) or enhance transcription (if the motif is on the non-template strand, maintaining the transcribed strand in a single-stranded conformation). Second, proteins bound to the G4 structures (for example, transcriptional enhancers versus repressors) could also affect transcription (reviewed in REF. 80).

One of the best-studied systems for a role of G4 structures in transcription involves the mammalian *MYC* (also known as *c-MYC*) locus (reviewed in REFS 3,79), although findings similar to those discussed below have been reported for multiple loci[80–84]. MYC is a transcription factor whose expression is associated with cell proliferation. Increased levels of *MYC* expression are observed in 80% of human cancer cells, and this increase promotes tumorigenesis[85–90]. Nuclease hypersensitive element III$_1$ (NHE III$_1$), which is downstream of the *MYC* promoter, controls >80% of the *MYC* transcription. This element

contains a G4 motif that forms a G4 structure *in vitro*[91]. Footprinting studies and luciferase reporter assays comparing the expression of a gene with a wild-type NHE III$_1$ versus one with a mutated NHE III$_1$ that cannot form a G4 structure demonstrate that the G4 motif in NHE III$_1$ represses transcription[92]. In another study, TMPyP4, a compound that binds to and stabilizes G4 structures (but also binds duplex DNA)[93,94], reduced *MYC* transcription in lymphoma cell lines and showed antitumour activity in mice[92,95]. This reduction is speculated to be mediated by TMPyP4 binding to the G4 structure in
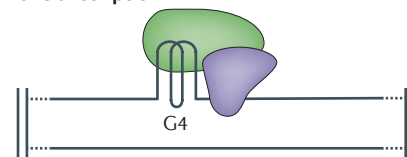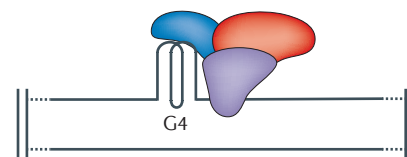

a Polymerase stalling, inhibition of transcription

b Facilitating transcription
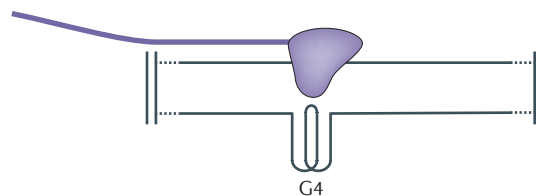
c Stimulation of transcription

d Transcription blocking (recruitment of repressors)

Figure 4 | **Putative functional roles of G-quadruplex structures during transcription.** Genome-wide bioinformatic analyses identified loci with high potential to form G-quadruplex (G4) structures. Among these loci, the promoters, transcription factor binding sites and 5′UTR regions of mRNAs are highly enriched for G4 motifs. These analyses, together with protein–G4 interaction studies, provide insights into predicted functions of G4 structures during transcription. **a** | G4 structures are postulated to block transcription by inhibiting polymerase (purple). **b** | G4 structures are postulated to facilitate transcription by keeping the transcribed strand in the single-stranded conformation. **c** | G4 DNA may stimulate transcription by recruiting proteins (green) that recruit or stimulate polymerase. **d** | G4 structures are suggested to block transcription via the recruitment of G4 binding proteins (blue), which directly or indirectly (red) repress transcription.

NHE III$_1$ of *MYC*. However, given that TMPyP4 binding is not limited to G4 structures and the many G4 motifs in the genome, more analysis is required to determine its mechanism of action. GQC-05, an analogue of ellipticine (an antineoplastic drug), is another promising therapeutic ligand. GQC-05 binds the G4 structure in the NHE III$_1$ region of *MYC in vitro* with high affinity and selectivity, and when added to Burkitt's lymphoma cell lines, GQC-05 results in reduced levels of transcribed MYC mRNA[96]. However, a recent publication found that 11 known G4 DNA ligands that affect *MYC* expression in cell-free assays do not interact directly with the MYC G4 structure in certain Burkitt's lymphoma cell lines[97], clouding the interpretation of the GQC-05 results.

Nucleolin, the most abundant nucleolar phosphoprotein in eukaryotic cells, is also proposed to regulate *MYC* transcription via its interaction with NHE III$_1$. This hypothesis is based on the *in vivo* binding of nucleolin to the *MYC* promoter in HeLa cells and the dose-dependent reduction in *MYC* transcription that occurs in nucleolin-treated cells[98]. One hypothesis is that nucleolin-mediated G4 formation in NHE III$_1$ inhibits *MYC* transcription by masking binding sites for *MYC* transcriptional activators, such as the transcripton factor SP1 and cellular nucleic acid-binding protein (CNBP)[99]. However, human nucleolin binds many G4 structures and can induce the formation of G4 DNA *in vitro*[98,100–103]. Thus, more work is needed to establish that nucleolin-associated changes in *MYC* transcription are a direct result of its effects on G4 structure formation within the NHE III$_1$ element.

Regulation through proteins binding to G4 structures. Transcription may also be altered by G4 binding proteins that affect the formation and unfolding of G4 structures. For example, myosin D (MyoD) family proteins are transcription factors that bind to E-boxes in the promoters of several muscle-specific genes to regulate muscle development[104]. *In vitro*, MyoD homodimers bind preferentially to G4 structures that are derived from the promoter sequences of muscle specific genes[105]. One hypothesis is that when G4 structures form in the promoters of E-box driven genes, MyoD homodimers preferentially bind to the G4 structure and not the E-box. Consequently, MyoD–MyoE heterodimers, which cannot bind G4 structures, bind to the E-box instead and enhance gene transcription[106]. However, like the *MYC* experiments, additional work is needed to prove this hypothesis.

In addition to gene-specific approaches, results from genome-wide studies analysing the effects of drugs that stabilize and/or induce G4 formation have been used to argue that G4 structures affect transcription[79,107]. Indeed, expression levels of many genes are influenced by treating cells with G4 ligands. Similar studies have investigated the effects of mutations in helicases known to unwind G4 DNA on transcription genome wide[17,108]. For instance, in human fibroblasts deficient for the WRN or BLM RecQ helicases, the transcription of genes that are predicted to form intramolecular G4 structures is significantly upregulated ($P < 0.0001$), and this upregulation correlates with the G4 motifs, not simple G-richness[108]. The genes associated with G4 motifs account for 20–30% of all transcripts that are upregulated in WRN and BML mutant cells.

Although such studies support a role for G4 structures in transcription, when interpreting genome-wide studies the possibility must be considered that many of the observed changes in gene expression may be indirect. However, in diverse organisms, genes whose expression is affected by G4 ligands are statistically associated with the presence of nearby G4 motifs, which provides some of the best evidence for widespread effects of G4 structures on transcription.

A general criticism of models in which G4 structures affect transcription is that G4 formation is too slow and the stability of G4 structures is too high for them to be used as regulatory elements. This criticism can also be raised against hypotheses suggesting that G4 structures affect telomeres or DNA replication. Indeed, it is well documented that intermolecular G4 DNA structures form and resolve slowly under physiological conditions[109,110]. However, the existence of chaperones (for example, TEBPβ and Rap1) that promote the formation of G4 DNA[27–29] suggests that nature has evolved mechanisms to overcome this slow formation. A recent thermodynamic and kinetic measurement of G4 structure formation indicates that G4 structures can form cooperatively[111]. Rates of formation for intramolecular G4 structures have also been reported for human telomeric G4 DNA (millisecond timescale)[112], and it is possible that other intramolecular G4 structures form as readily. This possibility is simple to test and should be demonstrated directly for other G4 motifs that are proposed to form intramolecular G4 structures that function *in vivo*. Unwinding of G4 structures in a timely manner can also no longer be considered a problem given the discovery of helicases that bind and unwind G4 motifs with high efficiency (see above).

Other roles for G4 DNA

Epigenetic regulation. A new hypothesis suggests that G4 structures might influence epigenetic regulation of gene expression. Maintaining epigenetic marks, such as histone methylation, is essential for stable gene expression and cell identity, and these marks must therefore be preserved after DNA replication and repair. As reported above, G4 structures are thought to cause replication fork stalling. These stalled forks might be restarted with the aid of translesion polymerases, as suggested by data from DT40 chicken cells[66], in which REV1, a Y family translesion polymerase[113], is implicated in G4 lesion bypass. In the absence of REV1, DNA synthesis is uncoupled from histone recycling mechanisms, and transcriptional activation is blocked[66]. The authors postulate that REV1 functions in replication at G4 motifs in order to preserve histone modifications[66]. A recent publication extends this work by showing by microarray analysis that lack of REV1 causes genome-wide dysregulation of G4-dependent transcription in DT40 cells (P value = 0.005), and this dysregulation is worsened by mutation of the WRN, BLM and FANCJ helicases[114].

Telomere-dependent bouquet structure
A structure formed by telomeres in early meiosis. It is associated with the nuclear scaffold.

Hoogsteen base pairing
Base pairing that differs from the normal Watson–Crick base pairing.

Origins of replication. It is well documented that chromatin can influence the timing of origin activation during DNA replication[115]. Recently, genome-wide analysis of replication origins[116] using a short nascent strand sequencing approach together with deep sequencing techniques identified a large number of new origins in different human cell types. Most of the identified peaks overlap with previously identified origins; however, many of the newly identified origins are significantly associated with G4 motifs. The authors propose that G4 structures near origins promote origin of replication complex binding and thereby influence origin activation[116], although direct proof for this model is not yet available.

Meiosis. G4 structures are also suggested to be involved in the alignment of sister chromatids during meiosis. One hypothesis is that G4 structures assist in the formation of the telomere-dependent bouquet structure during meiosis (FIG. 5a), but there is no direct evidence for this appealing possibility[26] Various G4-promoting proteins (FIG. 5a, pink) might be involved in formation of the G4 and tethering of the bouquet. G4 structures are also proposed to have a more general role in meiosis: for example, by promoting meiotic homologous recombination[76,117] (FIG. 5b). This idea is supported by genome-wide computational studies in yeast that demonstrate overlap between G4 motifs and preferred meiotic DSB sites[15], but Spo11, the enzyme that makes the DSBs, does not cleave at G4 motifs[118]. However, a role for G4 DNA in meiosis is supported by the finding that the *S. cerevisiae* Hop1 protein, which is a major component of the chromosome axial element–synaptonemal complex during meiosis, promotes G4 formation *in vitro*[119,120]. The multifunctional protein Kem1 also binds G4 structures *in vitro* and cleaves in the single stranded region 5′ of the G4 structures. Together with the fact that *kem1Δ* cells arrest during meiotic prophase, these results led to speculation that Kem1 acts on G4 structures *in vivo*[121]. In addition, the MRX complex, which is composed of Mre11, Rad50 and Xrs2 and acts during meiotic DSB formation, has a high affinity for G4 structures *in vitro*[122,123]. However, there is not yet *in vivo* evidence that Hop1, Kem1 or the MRX complex carry out their meiotic functions by acting at G4 structures.

Recombination. In several pathogenic microorganisms, recombination provides the basis for antigenic variation in which the pathogen escapes its host's immune surveillance by changing the identity of a surface antigen. There is good evidence that *Neisseria gonorrhoeae*, the bacterium that causes human gonorrhoea, uses a G4 based system to regulate expression of the genes that allow it to avoid the human immune system[124]. *N. gonorrhoeae* encodes many pilin genes, the products of which make up the hair-like projections, called pili, on the bacterium's surface. However, only the gene in the *pilE* locus is expressed, and the identity of the gene at this site switches among the different pilin genes by a recombinational mechanism. The region upstream of the *pilE* locus contains a 12 bp G-rich segment that is required for antigenic

variation and that can form a parallel intramolecular G4 structure *in vitro*. Mutations that eliminate antigenic variation *in vivo* also eliminate the ability of the segment to form a G4 structure, while mutations in the loop region of the structure affect neither antigenic variation nor G4 structure formation. Moreover, treating cells with the G4 ligand *N*-methyl mesoporphyrin IX affects *pilE* gene conversion events. The *N. gonorrhoeae* RecQ helicase is one of several proteins required in *trans* for efficient antigenic variation, providing additional evidence for a role for RecQ helicases at G4 structures *in vivo*. G4-based *N. gonorrhoeae pilE* recombination is perhaps the best evidence for a functional role of G4 DNA.

Conclusions

Although different in their three-dimensional conformation, G4 structures and the other non-B-form DNA secondary structures included in BOX 1 display some similarities. First, they can all form readily under the proper *in vitro* conditions. Second, formation of all of these secondary structures can help to relax negative DNA supercoiling, and Hoogsteen base pairing is often involved in stabilizing the structures. Third, the evolutionary conservation of the motifs capable of forming these secondary structures and the cellular machinery available to resolve them (for example, helicases and mismatch repair) argues for their existence *in vivo*. However, although they are of considerable interest

a Organization of telomeric bouquet
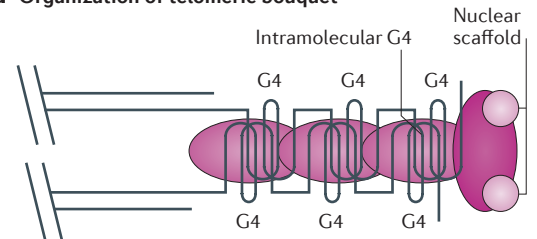
b Role in homologous recombination
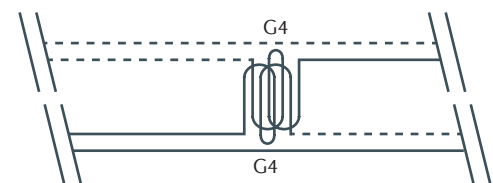
Figure 5 | **Putative roles for G-quadruplex structures in meiosis. a** | It has been proposed that G-quadruplex (G4) structures might assist in the formation of the telomere-dependent bouquet structure during meiosis[26]. G4-promoting proteins (pink) could be involved in the formation of G4 structures and tethering of the telomeric bouquet to the nuclear scaffold. **b** | It has also been suggested that G4 structures could promote meiotic homologous recombination[76,117] if there is overlap between G4 motifs and preferred meiotic double-strand break sites[15]. Sites of homologous recombination are indicated by the dashed lines.

from a chemical standpoint, some chromosome biologists remain sceptical that these secondary structures are physiologically relevant. G4 DNA provides an excellent example of the gulf between the wealth of *in vitro* data and the relative scarcity of results demonstrating formation and function of these structures *in vivo*. The findings that G4 motifs are evolutionarily conserved, over-represented in certain regions and associated with a specific subset of genomic features provides good, albeit indirect, evidence for G4 structures *in vivo*.

Direct evidence for G4 structures *in vivo* has been slow in coming. G4-specific antibodies and ligands provide support for G4 DNA *in vivo*, especially at telomeres, but it is difficult to demonstrate convincingly that the specificity of these reagents is high enough to rule out the possibility that their effects are due to association with B-DNA. Genetic experiments provide the most persuasive evidence to date for the *in vivo* existence of G4 structures during replication[64,68,69] and transcription[99]. Regardless of the process or function in question, one must test directly for positive roles of G4 structures, for instance by mutating G4 motifs in promoter regions or meiotic DSB sites and determining whether loss of the ability to form a G4 structure affects downstream processes. However, in the end, the most convincing evidence for the existence of G4 structures *in vivo* will be a direct demonstration of these structures *in vivo*. Doing so will require a creative approach to isolate the structures with sufficient purity that they can be characterized by the kinds of approaches used to analyse *in vitro*-generated G4 structures.

To summarize, G4 motifs are ubiquitous in prokaryotic and eukaryotic genomes, and their location is often conserved in closely related species. These motifs may form G4 structures *in vivo*, and the G4 structures may have functional roles, such as regulating recombination, meiotic DSB formation and/or transcription or providing a template for an RNA that forms a G4 structure that affects its post-transcriptional behaviour (see below). Alternatively (or in addition), G4 DNA formation may be pathological, occurring only occasionally owing to a problem in DNA mechanics, such as slowed DNA replication (as in the presence of hydroxyurea), which would provide more time for G4 DNA formation, especially during lagging strand replication. Pathological G4 structures could form at sites where G4 DNA has a direct or indirect function (for example, meiotic DSB sites in mitotic cells) or at sites that are complementary to an RNA containing a G4 structure that has a function in the RNA (in this case, the G4 RNA has a function but its complement in the DNA does not). Although this Review concerns DNA secondary structures, we would be remiss without noting that similar structures, especially G4 structures, can form in RNA. One possibility is that G4 motifs are encoded in the DNA but mainly function at the RNA level. G4 RNA structures are reported to affect mRNA splicing, translation and degradation (reviewed in REFS 8,125,126). It seems clear that the study of non-canonical RNA and DNA secondary structures will provide fertile ground for research for the foreseeable future.

1. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
2. Kypr, J., Kejnovska, I., Renciuk, D. & Vorlickova, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713–1725 (2009).
3. Huppert, J. L. Structure, location and interactions of G-quadruplexes. *FEBS J.* **277**, 3452–3458 (2010).
4. Bang, I. Untersuchungen über die Guanylsäure. *Biochem. Z.* **26**, 293–231 (1910) (in German).
5. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA* **48**, 2013–2018 (1962).
   This is the first observation that guanylic acids can assemble into higher-order structures.
6. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59**, 871–880 (1989).
   This study demonstrates that oligonucleotides composed of cilate telomeric repeat sequences form G-quartets in the presence of certain monovalent cations (for example, Na$^+$ and K$^+$). The authors also propose that G-quartets may form at telomeres *in vivo* and must be dealt with by the replication machinery.
7. Wilson, W. D. & Sugiyama, H. First international meeting on quadruplex DNA. *ACS Chem. Biol.* **2**, 589–594 (2007).
8. Wong, H. M., Payet, L. & Huppert, J. L. Function and targeting of G-quadruplexes. *Curr. Opin. Mol. Ther.* **11**, 146–155 (2009).
9. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415 (2006).
10. Hazel, P., Parkinson, G. N. & Neidle, S. Predictive modelling of topology and loop variations in dimeric DNA quadruplex structures. *Nucleic Acids Res.* **34**, 2117–2127 (2006).
11. Hardin, C. C., Perry, A. G. & White, K. Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* **56**, 147–194 (2000).
12. Guedin, A., Gros, J., Alberti, P. & Mergny, J. L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **38**, 7858–7868 (2010).
13. Bugaut, A. & Balasubramanian, S. 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 (2008).
14. Patel, D. J., Phan, A. T. & Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **35**, 7429–7455 (2007).
15. Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.* **6**, e1000861 (2010).
    This study reports a genome-wide computational analysis identifying the location and evolutionary conservation of G4 motifs in *S. cerevisiae* and related yeasts.
16. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901–2907 (2005).
17. Hershman, S. G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **36**, 144–156 (2008).
18. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
    The authors performed a genome-wide computational analysis that identified all regions in the human genome with a high potential to form G4 structures.
19. Huppert, J. L. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* **37**, 1375–1384 (2008).
20. Rawal, P. *et al.* Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.* **16**, 644–655 (2006).
21. Nakken, S., Rognes, T. & Hovig, E. The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic Acids Res.* **37**, 5749–5756 (2009).
22. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887–3896 (2006).
23. Zakian, V. A. Telomeres: the beginnings and ends of eukaryotic chromosomes. *Exp. Cell Res.* **318**, 1456–1460 (2012).
24. Henderson, E. *et al.* Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* **51**, 899–908 (1987).
25. Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342**, 825–829 (1989).
    This *in vitro* analysis demonstrates that telomeric DNA can fold into G4 structures.
26. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366 (1988).
27. Fang, G. & Cech, T. R. The β subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* **74**, 875–885 (1993).
28. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. & Lipps, H. J. Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nature Struct. Mol. Biol.* **12**, 847–854 (2005).
    This study includes compelling evidence for the *in vivo* existence of G4 structures at telomeres. Telomere-binding proteins are shown to regulate the formation of such structures.
29. Giraldo, R. & Rhodes, D. The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* **13**, 2411–2420 (1994).
30. Zaug, A. J., Podell, E. R. & Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension *in vitro*. *Proc. Natl Acad. Sci. USA* **102**, 10864–10869 (2005).

31. Wang, H., Nora, G. J., Ghodke, H. & Opresko, P. L. Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. *J. Biol. Chem.* **286**, 7479–7489 (2011).

32. Schaffitzel, C. *et al. In vitro* generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl Acad. Sci.* **98**, 8572–8577 (2001).

33. Paeschke, K. *et al.* Telomerase recruitment by the telomere end binding protein-β facilitates G-quadruplex DNA unfolding in ciliates. *Nature Struct. Mol. Biol.* **15**, 598–604 (2008).

34. Postberg, J., Tsytlonok, M., Sparvoli, D., Rhodes, D. & Lipps, H. J. A telomerase-associated RecQ protein-like helicase resolves telomeric G-quadruplex structures during replication. *Gene* **497**, 147–154 (2012).

35. Juranek, S. A. & Paeschke, K. Cell cycle regulation of G-quadruplex DNA structures at telomeres. *Curr. Pharm. Des.* **18**, 1867–1872 (2012).

36. Paeschke, K., McDonald, K. R. & Zakian, V. A. Telomeres: structures in need of unwinding. *FEBS Lett.* **584**, 3769–3772 (2010).

37. Yang, Q. *et al.* Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: I. recognizing mixed G-quadruplex in human telomeres. *Chem. Commun.* **9**, 1103–1105 (2009).

38. Chang, C. C. *et al.* A novel carbazole derivative, BMVC: a potential antitumor agent and fluorescence marker of cancer cells. *Chem. Biodivers.* **1**, 1377–1384 (2004).

39. Shay, J. W. & Wright, W. E. Role of telomeres and telomerase in cancer. *Seminars Cancer Biol.* **21**, 349–353 (2011).

40. 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).
    The authors report the first observation that telomerase action is influenced by G4 structures.

41. Oganesian, L., Moon, I. K., Bryan, T. M. & Jarstfer, M. B. Extension of G-quadruplex DNA by ciliate telomerase. *EMBO J.* **25**, 1148–1159 (2006).

42. Oganesian, L., Graham, M. E., Robinson, P. J. & Bryan, T. M. Telomerase recognizes G-quadruplex and linear DNA as distinct substrates. *Biochemistry* **46**, 11279–11290 (2007).

43. Neidle, S. Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **277**, 1118–1125 (2010).

44. Rezler, E. M. *et al.* Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J. Am. Chem. Soc.* **127**, 9439–9447 (2005).

45. Kim, M. Y., Vankayalapati, H., Shin-Ya, K., Wierzba, K. & Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. *J. Am. Chem. Soc.* **124**, 2098–2099 (2002).

46. De Cian, A. *et al.* Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proc. Natl Acad. Sci. USA* **104**, 17347–17352 (2007).

47. Gomez, D. *et al.* Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.* **32**, 371–379 (2004).

48. Kim, M. Y., Gleason-Guzman, M., Izbicka, E., Nishioka, D. & Hurley, L. H. The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. *Cancer Res.* **63**, 3247–3256 (2003).

49. Shammas, M. A. *et al.* Telomerase inhibition and cell growth arrest after telomestatin treatment in multiple myeloma. *Clin. Cancer Res.* **10**, 770–776 (2004).

50. Tahara, H. *et al.* G-quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3′ telomeric overhang in cancer cells. *Oncogene* **25**, 1955–1966 (2006).

51. Tauchi, T. *et al.* Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095), against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene* **22**, 5338–5347 (2003).

52. Tauchi, T. *et al.* Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: *in vitro* and *in vivo* studies in acute leukemia. *Oncogene* **25**, 5719–5725 (2006).

53. Gomez, D. *et al.* Interaction of telomestatin with the telomeric single-strand overhang. *J. Biol. Chem.* **279**, 41487–41494 (2004).

54. Gomez, D. *et al.* The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences *in vitro* and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer Res.* **66**, 6908–6912 (2006).

55. Smith, J. S. *et al.* Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nature Struct. Mol. Biol.* **18**, 478–485 (2011).

56. London, T. B. *et al.* FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.* **283**, 36132–36139 (2008).

57. Mohaghegh, P. *et al.* The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.* **29**, 2843–2849 (2001).

58. Huber, M. D., Lee, D. C. & Maizels, N. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.* **30**, 3954–3961 (2002).

59. 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).

60. Sanders, C. M. Human Pif1 helicase is a G-quadruplex DNA binding protein with G-quadruplex DNA unwinding activity. *Biochem. J.* **430**, 119–128 (2010).

61. Arola, A. & Vilar, R. Stabilisation of G-quadruplex DNA by small molecules. *Curr. Top. Med. Chem.* **8**, 1405–1415 (2008).

62. Neidle, S. The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* **19**, 239–250 (2009).

63. Wu, Y., Shin-ya, K. & Brosh, R. M. Jr. FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.* **28**, 4116–4128 (2008).

64. 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).
    This paper demonstrates that the *S. cerevisiae* Pif1 helicase binds to G4 motifs genome-wide and is important for DNA replication and genome stability at such sites.

65. Lopes, J. *et al.* G-quadruplex-induced instability during leading-strand replication. *EMBO J.* **30**, 4033–4046 (2011).

66. 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).
    The authors show that failure to properly replicate through G4 motifs affects chromatin structure.

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

68. Cheung, I., Schertzer, M., Rose, A. & Lansdorp, P. M. Disruption of *dog-1* in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nature Genet.* **31**, 405–409 (2002).
    Genetic experiments reveal that deficiencies in the DOG-1 helicase lead to genome instability at G-rich sequences in *C. elegans*.

69. Kruisselbrink, E. *et al.* Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective *C. elegans*. *Curr. Biol.* **18**, 900–905 (2008).

70. 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).
    This study indicates that G4 structures cause vertebrate telomere fragility, which can be counteracted by the RTEL1 helicase.

71. Barber, L. J. *et al.* RTEL1 maintains genomic stability by suppressing homologous recombination. *Cell* **135**, 261–271 (2008).

72. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406–413 (2007).
    Computational analysis revealed that G4 motifs are significantly enriched at promoters in human DNA.

73. Yadav, V. K., Abraham, J. K., Mani, P., Kulshrestha, R. & Chowdhury, S. QuadBase: genome-wide database of G4 DNA—occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.* **36**, D381–D385 (2008).

74. Mullen, M. A. *et al.* RNA G-quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. *Nucleic Acids Res.* **38**, 8149–8163 (2010).

75. Huppert, J. L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* **36**, 6260–6268 (2008).

76. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* **18**, 1618–1629 (2004).

77. Kouzine, F., Sanford, S., Elisha-Feil, Z. & Levens, D. The functional response of upstream DNA to dynamic supercoiling *in vivo*. *Nature Struct. Mol. Biol.* **15**, 146–154 (2008).

78. Sun, D. & Hurley, L. H. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.* **52**, 2863–2874 (2009).

79. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **277**, 3459–3469 (2010).

80. Qin, Y. & Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* **90**, 1149–1171 (2008).

81. Gunaratnam, M. *et al.* G-quadruplex compounds and cis-platin act synergistically to inhibit cancer cell growth *in vitro* and *in vivo*. *Biochem. Pharmacol.* **78**, 115–122 (2009).

82. Hsu, S. T. *et al.* A G-rich sequence within the *c-kit* oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Am. Chem. Soc.* **131**, 13399–13409 (2009).

83. Palumbo, S. L., Ebbinghaus, S. W. & Hurley, L. H. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc.* **131**, 10878–10891 (2009).

84. Bejugam, M. *et al.* Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of *c-kit* oncogene expression. *J. Am. Chem. Soc.* **129**, 12926–12927 (2007).

85. Marcu, K. B., Bossone, S. A. & Patel, A. J. *myc* function and regulation. *Annu. Rev. Biochem.* **61**, 809–860 (1992).

86. D'Cruz, C. M. *et al.* c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous *Kras2* mutations. *Nature Med.* **7**, 235–239 (2001).

87. Strieder, V. & Lutz, W. Regulation of N-*myc* expression in development and disease. *Cancer Lett.* **180**, 107–119 (2002).

88. Lutz, W., Leon, J. & Eilers, M. Contributions of Myc to tumorigenesis. *Biochim. Biophys. Acta* **1602**, 61–71 (2002).

89. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nature Rev. Cancer* **2**, 764–776 (2002).

90. Pelengaris, S., Khan, M. & Evan, G. I. Suppression of Myc-induced apoptosis in beta cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression. *Cell* **109**, 321–334 (2002).

91. Simonsson, T., Pecinka, P. & Kubista, M. DNA tetraplex formation in the control region of c-*myc*. *Nucleic Acids Res.* **26**, 1167–1172 (1998).

92. 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).

93. Han, H., Langley, D. R., Rangan, A. & Hurley, L. H. Selective interactions of cationic porphyrins with G-quadruplex structures. *J. Am. Chem. Soc.* **123**, 8902–8913 (2001).

94. Sun, D. *et al.* Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.* **40**, 2113–2116 (1997).

95. Grand, C. L. *et al.* The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth *in vivo*. *Mol. Cancer Ther.* **1**, 565–573 (2002).

96. Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H. & Brooks, T. A. Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *J. Biol. Chem.* **286**, 41018–41027 (2011).

97. Boddupally, P. V. *et al.* Anticancer activity and cellular repression of c-MYC by the G-quadruplex-stabilizing 11-piperazinylquindoline is not dependent on direct targeting of the G-quadruplex in the c-MYC promoter. *J. Med. Chem.* **55**, 6076–6086 (2012).

98. Gonzalez, V., Guo, K., Hurley, L. & Sun, D. Identification and characterization of nucleolin as a c-*myc* G-quadruplex-binding protein. *J. Biol. Chem.* **284**, 23622–23635 (2009).

99. Gonzalez, V. & Hurley, L. H. The c-*MYC* NHE III$_1$: function and regulation. *Annu. Rev. Pharmacol. Toxicol.* **50**, 111–129 (2010).

100. Bates, P. J., Kahlon, J. B., Thomas, S. D., Trent, J. O. & Miller, D. M. Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. *J. Biol. Chem.* **274**, 26369–26377 (1999).

101. Brys, A. & Maizels, N. LR1 regulates c-*myc* transcription in B-cell lymphomas. *Proc. Natl Acad. Sci. USA* **91**, 4915–4919 (1994).

102. Dempsey, L. A., Sun, H., Hanakahi, L. A. & Maizels, N. G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, a role for G-G pairing in immunoglobulin switch recombination. *J. Biol. Chem.* **274**, 1066–1071 (1999).

103. Gonzalez, V. & Hurley, L. H. The C-terminus of nucleolin promotes the formation of the c-*MYC* G-quadruplex and inhibits c-*MYC* promoter activity. *Biochemistry* **49**, 9706–9714 (2010).

104. Wei, Q. & Paterson, B. M. Regulation of MyoD function in the dividing myoblast. *FEBS Lett.* **490**, 171–178 (2001).

105. Shklover, J., Weisman-Shomer, P., Yafe, A. & Fry, M. Quadruplex structures of muscle gene promoter sequences enhance *in vivo* MyoD-dependent gene expression. *Nucleic Acids Res.* **38**, 2369–2377 (2010).

106. Yafe, A., Shklover, J., Weisman-Shomer, P., Bengal, E. & Fry, M. Differential binding of quadruplex structures of muscle-specific genes regulatory sequences by MyoD, MRF4 and myogenin. *Nucleic Acids Res.* **36**, 3916–3925 (2008).

107. Fernando, H. *et al.* Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. *Nucleic Acids Res.* **37**, 6716–6722 (2009).

108. 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 Acids Res.* **38**, 1114–1122 (2010).

109. Wyatt, J. R., Davis, P. W. & Freier, S. M. Kinetics of G-quartet-mediated tetramer formation. *Biochemistry* **35**, 8002–8008 (1996).

110. Mergny, J. L., De Cian, A., Ghelab, A., Sacca, B. & Lacroix, L. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.* **33**, 81–94 (2005).

111. Yu, Z. *et al.* Tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. *J. Am. Chem. Soc.* **134**, 5157–5164 (2012).

112. Gray, R. D. & Chaires, J. B. Kinetics and mechanism of K$^+$- and Na$^+$-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res.* **36**, 4191–4203 (2008).

113. Edmunds, C. E., Simpson, L. J. & Sale, J. E. PCNA ubiquitination and REV1 define temporally distinct mechanisms for controlling translesion synthesis in the avian cell line DT40. *Mol. Cell* **30**, 519–529 (2008).

114. Sarkies, P. *et al.* FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res.* **40**, 1485–1498 (2012).

115. Hiratani, I., Takebayashi, S., Lu, J. & Gilbert, D. M. Replication timing and transcriptional control: beyond cause and effect—part II. *Curr. Opin. Genet. Dev.* **19**, 142–149 (2009).

116. Besnard, E. *et al.* Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature Struct. Mol. Biol.* **19**, 837–844 (2012). **This computational study found that G4 consensus motifs are located near origins of replication in human cells.**

117. Lobachev, K. S. *et al.* Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics* **148**, 1507–1524 (1998).

118. Pan, J. *et al.* A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**, 719–731 (2011).

119. Muniyappa, K., Anuradha, S. & Byers, B. Yeast meiosis-specific protein Hop1 binds to G4 DNA and promotes its formation. *Mol. Cell. Biol.* **20**, 1361–1369 (2000).

120. Anuradha, S. & Muniyappa, K. Meiosis-specific yeast Hop1 protein promotes synapsis of double-stranded DNA helices via the formation of guanine quartets. *Nucleic Acids Res.* **32**, 2378–2385 (2004).

121. Liu, Z. & Gilbert, W. The yeast *KEM1* gene encodes a nuclease specific for G4 tetraplex DNA: implication of *in vivo* functions for this novel DNA structure. *Cell* **77**, 1083–1092 (1994).

122. Ghosal, G. & Muniyappa, K. *Saccharomyces cerevisiae* Mre11 is a high-affinity G4 DNA-binding protein and a G-rich DNA-specific endonuclease: implications for replication of telomeric DNA. *Nucleic Acids Res.* **33**, 4692–4703 (2005).

123. Ghosal, G. & Muniyappa, K. The characterization of *Saccharomyces cerevisiae* Mre11/Rad50/Xrs2 complex reveals that Rad50 negatively regulates Mre11 endonucleolytic but not the exonucleolytic activity. *J. Mol. Biol.* **372**, 864–882 (2007).

124. Cahoon, L. A. & Seifert, H. S. An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science* **325**, 764–767 (2009). **The genetics in this paper provide some of the best evidence for the existence of G4 structures *in vivo*.**

125. Bugaut, A. & Balasubramanian, S. 5′-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40**, 4727–4741 (2012).

126. Millevoi, S., Moine, H. & Vagner, S. G-quadruplexes in RNA biology. *Wiley Interdisciplinary Rev. RNA* **3**, 495–507 (2012).

127. Svozil, D., Kalina, J., Omelka, M. & Schneider, B. DNA conformations and their sequence preferences. *Nucleic Acids Res.* **36**, 3690–3706 (2008).

128. Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A. & Wang, A. H. The molecular structure of the left-handed Z-DNA double helix at 1.0-Å atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). *J. Biol. Chem.* **264**, 7921–7935 (1989).

129. Khuu, P., Sandor, M., DeYoung, J. & Ho, P. S. Phylogenomic analysis of the emergence of GC-rich transcription elements. *Proc. Natl Acad. Sci. USA* **104**, 16528–16533 (2007).

130. Rahmouni, A. R. & Wells, R. D. Stabilization of Z DNA *in vivo* by localized supercoiling. *Science* **246**, 358–363 (1989).

131. Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y. G. & Kim, K. K. Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* **437**, 1183–1186 (2005).

132. Schroth, G. P., Chou, P. J. & Ho, P. S. Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. *J. Biol. Chem.* **267**, 11846–11855 (1992).

133. Wang, G., Christensen, L. A. & Vasquez, K. M. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl Acad. Sci. USA* **103**, 2677–2682 (2006).

134. Zhao, J., Bacolla, A., Wang, G. & Vasquez, K. M. Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* **67**, 43–62 (2010).

135. Palecek, E. Local supercoil-stabilized DNA structures. *Crit. Rev. Biochem. Mol. Biol.* **26**, 151–226 (1991).

136. Pearson, C. E., Zorbas, H., Price, G. B. & Zannis-Hadjopoulos, M. Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. *J. Cell Biochem.* **63**, 1–22 (1996).

137. van Holde, K. & Zlatanova, J. Unusual DNA structures, chromatin and transcription. *Bioessays* **16**, 59–68 (1994).

138. Lobachev, K. S., Rattray, A. & Narayanan, V. Hairpin- and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. *Front. Biosci.* **12**, 4208–4220 (2007).

139. Glickman, B. W. & Ripley, L. S. Structural intermediates of deletion mutagenesis: a role for palindromic DNA. *Proc. Natl Acad. Sci. USA* **81**, 512–516 (1984).

140. Inagaki, H. *et al.* Chromosomal instability mediated by non-B DNA: cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. *Genome Res.* **19**, 191–198 (2009).

141. Kurahashi, H. *et al.* Palindrome-mediated chromosomal translocations in humans. *DNA Repair (Amst.)* **5**, 1136–1145 (2006).

142. Jain, A., Wang, G. & Vasquez, K. M. DNA triple helices: biological consequences and therapeutic potential. *Biochimie* **90**, 1117–1130 (2008).

143. Manor, H., Rao, B. S. & Martin, R. G. Abundance and degree of dispersion of genomic d(GA)n.d(TC)n sequences. *J. Mol. Evol.* **27**, 96–101 (1988).

144. Bacolla, A. *et al.* Long homopurine*homopyrimidine sequences are characteristic of genes expressed in brain and the pseudoautosomal region. *Nucleic Acids Res.* **34**, 2663–2675 (2006).

145. Wang, X. & Haber, J. E. Role of *Saccharomyces* single-stranded DNA-binding protein RPA in the strand invasion step of double-strand break repair. *PLoS Biol.* **2**, e21 (2004).

146. Owen, B. A. *et al.* (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nature Struct. Mol. Biol.* **12**, 663–670 (2005).

147. Rolfsmeier, M. L., Dixon, M. J. & Lahue, R. S. Mismatch repair blocks expansions of interrupted trinucleotide repeats in yeast. *Mol. Cell* **6**, 1501–1507 (2000).

FURTHER INFORMATION
Virginia A. Zakian's homepage: http://www.molbio1.princeton.edu/labs/zakian/

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF