G-quadruplex DNA and RNA: their roles in regulation of DNA replication and other biological functions

Hisao Masai and Taku Tanaka

From Department of Genome Medicine, Tokyo Metropolitan Institute of Medical Science, 2-1-6 Kamikitazawa, Setagaya-ku, Tokyo 156-8506, Japan

Correspondence to
Hisao Masai
Tel: +81-3-5316-3220
Fax: +81-3-5316-3145
E-mail: masai-hs@igakuken.or.jp

Key words: G-quadruplex; DNA replication initiation; replication timing; transcription; epigenome; chromatin regulation

Abstract

G-quadruplex is one of the best-studied non-B type DNA that is now known to be prevalently present in the genomes of almost all the biological species. Recent studies reveal roles of G-quadruplex (G4) structures in various nucleic acids and chromosome transactions. In this short article, we will first describe recent findings on the roles of G4 in regulation of DNA replication. G4 is involved in regulation of spatio-temporal regulation of DNA replication through interaction with a specific binding protein, Rif1. This regulation is at least partially mediated by generation of specific chromatin architecture through Rif1-G4 interactions. We will also describe recent studies showing the potential roles of G4 in initiation of DNA replication. Next, we will present showcases of highly diversified roles of DNA G4 and RNA G4 in regulation of nucleic acid and chromosome functions. Finally, we will discuss how the formation of cellular G4 could be regulated.

Contents

1 Introduction

G-quadruplex is one of the non-B type nucleic acid structures, which is now known to be widely present on the genomes of almost all the species. It is composed of stacked planar structures made of four guanines (called G-quartet). G4 is formed on single-stranded DNA with a sequence context "$G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$(x=1~7)". This algorithm predicted the presence of 375,000 potential G4 on the human genome [1-4], but search with "G4-seq" that utilizes the nature of G4 to block DNA polymerase led to identification of 716,310 putative G4 on the human genome [5]. Analyses of RNA-G4 with a similar approach using reverse transcriptase led to identification of a number of potential G4 on mRNA [6]. G4 can adopt different topologies, depending on how the four strands run with each other, and its three-dimensional structures also vary. A number of G4 proteins that selectively bind to G4 have been described [7]. Among them, helicases that are capable of unfolding the G4 structure have been shown to play crucial roles in maintenance of chromosome stability or promotion of translation or splicing of mRNA [8]. Various G4s probably interact with selective interactors to serve for specific biological functions.

In the first part, we will discuss the roles of G4 in regulation of DNA replication including those in chromatin-mediated regulation. DNA replication is one of the most fundamental processes for cell growth. Recent genome-wide mapping of initiation sites of DNA replication in higher eukaryotes including human indicates close association of G4-forming sequences with mapped origins [9-11]. Physical evidence suggests that G4 structure is required for initiation in cells [12]. In *Escherichia coli*, an alternative mode of chromosome DNA replication dependent on RNA-DNA hybrid is known to exist [13]. The potential role of G4-forming sequences and G4 structure in this mode of DNA replication would be of interest, since transcription of G4-forming sequences facilitates formation of RNA-DNA hybrid in conjunction with G4 structure composed most likely of non-template DNA strand and transcribed RNA ([14-18]).

Recent studies on chromatin structures in nuclei indicate the existence of ordered chromatin architecture, topologically associated domain (TAD) determined by the Chromatin Conformation Capture method, that may be intimately related to transcriptional pattern and changes during development or differentiation [19]. Accumulating evidence indicates that temporal and spatial regulation of DNA replication is determined by replication timing domain, which may be related to TAD

[20,21]. Rif1, an evolutionally conserved nuclear factor, was identified as a key regulator for regulation of DNA replication timing through altering the chromatin architecture [22]. Rif1 binds specifically G4 and may hold together the chromatin fiber through its G4 binding and multimization abilities, generating chromatin compartment that may be inhibitory for initiation. These results point to important roles of G4 in both positive and negative regulation of DNA replication.

G4, appearing in both DNA and RNA, plays pleiotropic roles in various DNA and RNA transactions. These include telomere regulation, epigenome regulation, transcription, translation, splicing, viral maintenance and functions, transpositions, repair and recombination (**Figure 1**). The G4s and their binding proteins regulate important biological responses including maintenance of genome stability, stress responses, immunogulobulin class switch recombination. In the second part, we will summarize recent findings on the roles of G4 in various biological reactions. Finally, we will discuss how dynamic formation and disappearance (unfolding) of G4 may be regulated in cells.

2 G4 and DNA replication

2-1 Regulation of DNA replication

DNA replication is one of the most fundamental processes of cell growth, which takes 6~8 hrs during cell cycle. One of the most important rules for regulation of DNA replication is that it occurs once and only once during one cell cycle [23]. This is achieved by multiple layers of mechanisms that ensure the assembly of pre-RC (pre-Replicative Complex) only during G1 phase, but not after the onset of DNA synthesis. pre-RC is generated on chromatin by sequential assembly of ORC, Cdc6 and Cdt1-MCM. Cell cycle-specific degradation of Cdt1 and stabilization of its inhibitor Geminin, as well as phosphorylation dependent-degradation and chromatin dissociation of pre-RC factors after S phase onset, is a part of the mechanisms for inhibition of rereplication. At the G1-S transition, Cdc7-Dbf4 kinase phosphorylates components of pre-RC, most notably Mcm, and triggers assembly of an active replicative helicase complex, CMG [see reviews in Masai, H. and M. Foiani, DNA Replication: From Old Principles to New Discoveries. 2018: Springer Singapore.].

During S phase, tens of thousands of replication origins are activated to ensure the replication of the entire genome within the allotted time. For this, replication

is compartmentalized both spatially and temporally [23]. In the early S phase, replication occurs mainly in the interior of the nuclei and at minute uniformly scattered foci. During mid-late S phase, replication occurs at larger fewer foci located at the nuclear periphery and around nucleoli. In late S phase, replication occurs at large, limited numbers of foci that colocalize with heterochromatin.

The genome segments containing the origins that fire during the same timing of S phase are often called “replication domains”, and their lengths range from several hundred kb to Mb [20]. The origins to be fired may be selected in a stochastic manner within the same replication domain. Pre-RCs are generated on chromosomes in excess over those required for replicating the entire genome, and those that are not used for firing could serve as “back-ups” in case of unscheduled fork arrest [24].

Many factors affect the replication timing. Generally, local chromatin structures affect the timing; open chromatin (nucleosome-free chromatin) is associated with early replication, and closed chromatin (heterochromatin) with late replication. Manipulation of chromatin structures (e.g. induction of histone acetylation) can locally induce early firing of origins. A transcription factor may induce clustering of replication origins for coordinated early replication in yeast [25]. "Early replication control element" (ERCEs) was identified in mouse ES cells, which facilitates early replication and active transcription within a compartmentalized domain of the chromosome [26]. On the other hand, a factor negatively regulating early firing was identified. In a mutant lacking Rif1 protein both in yeast and mammals, genome-wide and large scale alteration of replication timing was observed [22,27-29], as described in detail below.

2-2 G-quadruplex as an architectural basis for chromatin organization that negatively regulates replication initiation (Figure 2)

2-2-1 Identification of Rif1 as a conserved regulator of replication timing

Rif1, originally identified as a telomere binding factor in yeast [30,31] was rediscovered in a screen for bypass suppressors of the fission yeast hsk1 null mutant [25]. *hsk1*$^{+}$ encodes a fission yeast kinase related to Cdc7 [32]. Earlier studies suggested that mutants bypassing the requirement of *hsk1* for DNA replication encode factors that affect the replication timing [33,34]. In accordance with this expectation, *rif1*Δ cells, that could vigorously support the growth of *hsk1*Δ cells, exhibited striking change of origin firing pattern; 30% of the late-firing origins fired early, and similarly, 30% of

early-firing origins became late-firing in *rif1*Δ cells. The ChIP analyses indicated the binding of Rif1 to selected segments on the fission yeast chromosome; interestingly, Rif1 bound preferentially to the segments in the vicinity of the late-firing origins that are affected by Rif1 mutation. This suggested a possibility that Rif1 binding to chromosome inhibits firing of nearby origins [27]. Rif1 homologs in higher eukaryotes were not associated with telomeres, and its depletion did not affect the telomere maintenance [35,36]. In mammalian cells, Rif1 depletion caused hyperphosphorylation of Mcm2 and Mcm4, targets of phosphorylation by Cdc7 kinase, and hypereplication in early S phase [28,37].

2-2-2 Mechanisms of replication timing regulation by Rif1protein

Subsequent studies indicated two mechanisms for this regulation. One is through its ability to recruit a phosphatase PP1 [37-41]. All the Rif1 orthologs except for those from plants carry a conserved motif, to which PP1 binds. Mutations of this motif results in loss of replication timing regulation and hyperphosphorylation of Mcms, the targets of Cdc7 kinase. The recruited PP1 inhibits initiation by dephosphorylating Mcm and others that are crucial for initiation. Rif1's ability to generate specific chromatin structures contributes to a long-range effect of Rif1 binding on origin activity. This is exemplified by deregulation of origin firing over ~100 kb segment by a mutation of a single Rif1BS (Rif1 binding site).

Depletion of Rif1 in mammalian cells resulted in specific disappearance of mid-S type replication foci pattern and persistent display of early-S phase replication foci pattern until very late S phase, when it is replaced by the late heterochromatin replication foci pattern, indicating that Rif1 is specifically required for generating mid-S replication domains [26]. Rif1 is preferentially localized at nuclear periphery as well as around nucleoli. This localization is reminiscent to the replication foci observed during mid-to-late S phase. Rif1 is biochemically fractionated into detergent- and nuclease-insoluble fractions (nuclear matrix), cofractionated with LaminB1 protein. Rif1 is also localized in a close proximity of Lamin B1 protein at the nuclear periphery [26].

2-2-3 Chromatin binding of Rif1: recognition and binding of G4

Initial ChIP-chip analyses of Rif1 binding sites in fission yeast revealed binding sites along the three chromosomes (much less on the chromosome 3 the most of which is replicated early S phase), and the binding is maximum during G1 and gradually

decreased as cells move from S to G2. More accurate analyses with ChIP-seq identified 155 binding sites in G1/S phase [42] The survey of the sequences of Rif1-bound regions led to identification of a conserved motif sequence, GNNCNANG(T/A)GGGGG (designated Rif1CS). Further analyses indicated a set of 35 strong Rif1BS, normally containing two or more Rif1CS. Rif1CS contains a G-tract and appears twice or more with an average of 100 bp interval. In 75% of the case, Rif1CSs appear in a head-to-tail orientation, generating multiple G-tracts on one strand. Analyses of these sequences on an acrylamide gel indicated that they generate slow-migrating forms upon prior heat denaturation, and they disappeared after mutagenesis of the G-tracts, suggesting the formation of specific nucleic acid structures in a G-tract-dependent manner. Furthermore, they were stabilized in the presence of chemical ligands that are known to stabilize G4 (G4 ligands; see Nagasawa's article in this special issue). These results strongly suggested a possibility that Rif1BS can adopt G4 structures, and Rif1 may recognize the structured DNA. This possibility was examined by the DNA binding assay with purified fission yeast Rif1 protein. Gel shift assays with various forms of G4 DNA indeed have shown that Rif1 binds to parallel-type G4 DNA with high affinity (Kd=~0.3nM) [42-44].

ChIP-seq and 4C analyses of Rif1 in mouse ES cells revealed that it binds to late-replicating domains, and restricts the interactions between different replication timing domains during G1 phase. Rif1 generally makes extensive and rather weak contacts with the chromosomes [45]. We detected strong bindings at selected locations on the chromosomes, and these binding sites contain G-tracts that are capable of forming G4 (Yoshizawa *et al*. unpublished data). Purified human and mouse Rif1 proteins also specifically bind to G4 [46].

2-2-4 Biochemical characterization of Rif1 protein

Rif1 is a highly oligomeric protein, and biochemical evidence indicates it can bind to multiple DNAs simultaneously. Both fission yeast and mammalian Rif1 proteins carry dual G4 binding domains, one in the C-terminal segment and the other within the HEAT repeat. Each segment can bind to G4 but with reduced affinity compared to the full-length polypeptide [43,46]. The C-terminal segment also contain oligomerization activity. In budding yeast Rif1, a C-terminal segment can form a tetramer [47]. Origin suppression activity, as measured by the ability to suppress a *hsk1* mutant, is compromised in a *rif1* mutant deficient in chromatin binding. In fission yeast Rif1,

C-terminal 90 aa is sufficient to generate an oligomer, A model has been presented on how Rif1 may generate chromatin architecture through formation of chromatin loops [41,42,44]. Localization of Rif1 at the nuclear periphery suggests that it generates chromatin compartments along the nuclear membrane. Indeed, budding yeast Rif1 was reported to be palmitoylated by Pfa4 palmitoyl transferase, and this lipid modification facilitated DSB repair by localizing the DSB repair proximal to the inner nuclear membrane [48,49]. Similarly, mammalian Rif1 may be modified by lipid for its localization at the inner nuclear membrane. Identification of Rif1 mutants defective in a specific function of Rif1 (oligomerization, inner nuclear membrane association, and G4 binding) and analyses of their functions will clarify the mechanisms of Rif1-mediated chromatin reorganization and its role in defining replication timing domains.

2-3 G-quadruplex on RNA-DNA hybrid as a potential novel regulator of replication initiation (Figure 2)

Genome-wide analyses of replication origins in higher eukaryotes have been conducted with a hope that the results would answer long-standing questions of "Where are replication origins (*ori*s) in the human genome, and how are they selected?"

Genome-wide analysis of replication origins by nascent strand (NS) purification method in Drosophila and mouse cells indicate that most CpG islands (CGI) contain *ori*s [9,10,50]. DNA synthesis starts at the borders of CGI, suggestive of a dual initiation event. *Ori*s contain a unique nucleotide skew around NS peaks, characterized by G/T and C/A overrepresentation at the 5' and 3' of *ori* sites, respectively. Generally, regions rich in *ori*s are early replicating, whereas *ori*-poor regions are late replicating [9,10,50]. The same group reported the presence of Origin G-rich Repeated Elements (OGRE) potentially forming G4 at most origins. These coincide with nucleosome-depleted regions located upstream of the initiation sites, which are associated with a labile nucleosome containing H3K64ac [51]. Similar enrichment of G4 forming sequences near the origins was observed with origin mapping by initiation site sequencing (ini-seq), in which newly synthesized DNA is directly labelled with digoxigenin-dUTP in a cell-free system [11]. This result circumvents the criticism that the potential G4 enrichment is an artifact of NS analyses caused by the secondary structure of G4.

Furthermore, mutations affecting the G4 formation decreased the origin

efficiency measured by short nascent DNA analyses in DT40 as well as in mammalian cells [12]. In this report, authors discovered that the orientation of G4 determines the precise position of the replication start site. The results also indicate that a single G4 is not sufficient for origin activity and must cooperate with a 200-bp cis-regulatory element, which may be binding sites of transcription factors [12]. More recently, multiple functional assays using the transient replication of EBV-based episome as well as the replication in the *Xenopus* egg extracts suggested the requirement of G4 formation for efficient origin activity. The latter suggested the role of G4 in in the activation of pre-RC, but not in its formation [52]. Although these results suggest a direct role of G4 in origin firing, the precise mechanism is totally unknown. The purified human ORC protein was shown to bind to G4 *in vitro* [53]. However, its significance in origin recognition and firing is currently unknown.

Bacterial replicons are replicated through specific recognition of an origin (replicator) by an initiator protein, which permits highly efficient initiation event. This is a very robust system and occurs with nearly 100% efficiency. On the other hand, in *E. coli*, another replication system that does not depend on the major initiator-*ori* system is known to exist. This mode of replication known as stable DNA replication (SDR) occurs constitutively in RNaseH deficient mutant cells (*rnhA*) [13]. This replication depends on transcription, RecA and a replication factor, PriA, a conserved DEHX-type helicase [54].

The SDR mode of replication does not depend on DnaA-*oriC*, and presumably is initiated at multiple locations on the genome. Indeed, the genetic search for specific origins for SDR suggested the absence of specific segments required for SDR. The efficient occurrence of SDR in *rnhA* mutant suggests that RNA-DNA hybrids are important for this mode of replication. The search for cellular RNA-DNA hybrids by using a specific probe (human RNaseH1 lacking the catalytic activity) led to identification of ~100 sites under RNaseH-deficient condition. Comparison of RNA-DNA hybrid loci with the predicted loci for candidate G4 showed 66% of the putative G4 overlaps with RNA-DNA hybrids (Sagi *et al*. unpublished data). Indeed, it has been reported that formation of RNA-DNA hybrid is strongly coupled with the presence of G-tracts on the template DNA (on the non-template strand). It was proposed that both transcribed RNA and non-template DNA strand participate in forming “hybrid G4” [55]. Replication of ColE1-type plasmids can occur both in RNaseH$^+$ and RNaseH$^-$

cells [56]. In the latter condition, RNA-DNA hybrid may persist in the absence of RNaseH, where DNA replication is initiated. The formation of this RNA-DNA hybrid may also depend on G4, since mutagenesis of G-tracts within the transcribed segment results in loss of replication activity (Fukatsu *et al*. unpublished).

Mechanisms of replication initiation on RNA/DNA hybrid-G4 are still not clear. Replication initiation on pBR322 with purified proteins indicated the persistent RNA-DNA hybrid at origin provides a loading signal for primosome on the displaced lagging strand template, which may promote the synthesis of the lagging strand [57]. However, it is not clear how exactly an RNA-DNA hybrid structure is generated, how it is recognized by PriA to generate a replication fork and how a potential G4 structure contributes to this process. An alternative mode of genome replication was reported also in arhcae, Haloferax volcanii, in the absence of normal origins. Interestingly, this mode of replication initiates at multiple genomic loci and depends on the recombinase [58], similar to the SDR mode of DNA replication in *E. coli.*

Formation of G4 is expected to be dynamic, influenced by many factors within cells (see the last chapter). Thus, we expect that biological events regulated by G4 may undergo very dynamic alterations. This sort of "dynamics" may suit the needs of cells to be flexible, adaptive and robust in response to various external perturbations. It could possibly provide molecular basis for the stochastic nature of origin selection as well as stochastic activation/ inactivation of transcription in various cells.

3 G4 and chromosome transactions (see Supplementary Table S1 for summary)

3-1 G4 and telomere

G4 is most enriched at telomeres and may be protecting the chromosome ends. At the same time, replication fork blocks expected to be induced by G4 at the telomeres may be the sources of genomic instability.

Cancer cells mutated in ATRX are known to maintain telomere length by the homologous recombination (HR)-associated alternative lengthening of telomeres (ALT) pathway. ATRX suppresses the ALT pathway and shortens telomeres. Importantly ATRX-mediated ALT suppression depends on the histone chaperone DAXX. ATRX suppresses replication fork stalling, a known trigger for HR at telomeres. A G4 stabilizer partially reverses the effect of ATRX, suggesting that stabilization of G4 promotes ALT. It was proposed that loss of ATRX causes telomere chromatinization which results in

the persistence of G4 structures, leading to replication fork stalling, collapse, HR and subsequent recombination-mediated telomere synthesis in ALT cancers (**Figure 3**). It is known that G4s are generally present in the nucleosome free regions [59]. It is an interesting possibility that ATRX may serve as an anti-G4 factor through its ability to modify chromatin structures [60].

Replication protein A (RPA) prevents G4 formation at lagging-strand telomeres to maintain the telomere length. In fission yeast, Rpa1-D223Y mutation induces telomere shortening. This mutation impairs lagging-strand telomere replication and leads to the accumulation of secondary structures and recruitment of the homologous recombination factor Rad52 with concomitant reduction of Pot1 and Ccq1 association at telomeres. Expression of the budding yeast Pif1 that can unwind G4 rescues all the telomeric defects of the D223Y mutant. human RPA with the identical mutation is deficient in G4 binding, Thus, RPA prevents the formation of G4 at lagging-strand telomeres to promote shelterin association and maintain telomeres [61]. Similarly, CST (CTC1-STN1-TEN1), an RPA-like complex, that can bind and disrupts G4 structure, associates with G-rich single-strand DNA and helps resolve replication problems both at telomeres and chromosome arms [62] (**Figure 3**).

The helicase RTEL1 plays a role in the removal of telomeric DNA secondary structures, which is essential for preventing telomere fragility and loss (**Figure 3**). In the absence of RTEL1, T loops are inappropriately resolved by the SLX4 nuclease complex, resulting in loss of the telomere as a circle. Depleting SLX4 or blocking DNA replication abolished telomere circles (TCs) and rescued telomere loss in RTEL1(-/-) cells but failed to suppress telomere fragility. Conversely, stabilization of telomeric G4-DNA or loss of BLM dramatically enhanced telomere fragility in RTEL1-deficient cells but had no impact on TC formation or telomere loss. It was proposed that RTEL1 performs two distinct functions at telomeres: it disassembles T loops and also counteracts telomeric G4-DNA structures, which together ensure the dynamics and stability of the telomere. RTEL1 also facilitates the replication in other heterochromatin regions (see below) [63]

It is known that the common oxidative lesion 8-oxo-7,8-dihydro-2'-deoxyguanine (8-oxoG) regulates telomere elongation by human telomerase. When 8-oxoG is present in the dNTP pool as 8-oxodGTP, telomerase utilization of the oxidized nucleotide during telomere extension is mutagenic and

terminates further elongation. In contrast, a preexisting 8-oxoG within the telomeric DNA sequence promotes telomerase activity by destabilizing the G4 DNA structure. Thus, telomerase is inhibited or stimulated depending on how 8-oxoG arises in telomeres; inhibited by insertion of oxidized nucleotides and activated by direct reaction with free radicals. [64]

Although G4 at telomeres is inhibitory for telomere extension by telomerase, and needs to be removed for telomere elongation as described above, telomerase appears to be able to cope with this on its own. The core telomerase enzyme complex is sufficient for partial G4 resolution and extension. Human telomerase can extend purified conformationally homogenous human telomeric G4, which is in a parallel, intermolecular conformation *in vitro*. Thus, telomere G4 in a specific conformation may have some functional significance. [65]

Other telomere binding proteins potentially recognizing G4 includes HMGB1, a ubiquitous non-histone protein, depletion of which induces telomere damages [66] (**Figure 3**)

3-2 G4 and genome instability

G4 can impede the progression of replication fork. Extensive analyses of the ability of G4 structures using yeast assays indicated short loop length with pyrimidine bases and high thermal stability as critical determinants for the genomic instability induced by G4-forming minisatellites [67].

Specific helicases can unfold G4 structures, and facilitate the replication fork movement at these "hard-to-replicate" segments. Pif1 is one of the best-studied G4 resolvases. A significant subset of the Pif1 binding sites, as determined by genome-wide chromatin immunoprecipitation, overlaps with G4 motifs. Replication slowed in the vicinity of these motifs, and they were prone to undergo breakage in Pif1-deficient cells. These data suggest that G4 structures form *in vivo* and that they are resolved by Pif1 to prevent replication fork stalling and DNA breakage [68] (**Figure 4**).

Binding of Pif1 to cellular G4 may be assisted by other proteins. A co-transcriptional activator Sub1 of *Saccharomyces cerevisiae* is a G4-binding factor, and contributes to genome maintenance at G4-forming sequences. Genome instability linked to co-transcriptionally formed G4 DNA in Top1-deficient cells is significantly augmented in the absence of Sub1. This can be suppressed by its DNA binding domain

or the human homolog PC4. Sub1 interacts specifically with co-transcriptionally formed G4 DNA *in vivo*. Furthermore, Sub1 interacts with the G4-resolving helicase Pif1, suggesting a possible mechanism by which Sub1 suppresses instability at G4 DNA [69]. Mms1, an E3 ubiquitin ligase complex protein, was identified as a novel G4-DNA binding protein and promotes Pif1 helicase binding to these regions to maintain genome stability. [70]

RecQ helicases are known to target G4 and resolve its structure. BLM helicase belongs to this group of helicase and is mutated in Bloom syndrome, a cancer predisposition disorder. Cells from Bloom syndrome patients exhibit striking genomic instability characterized by excessive sister chromatid exchange events (SCEs). The mapping of the genomic locations of SCEs in the absence of BLM indicates striking enrichment at coding regions, specifically at sites of G4 motifs in transcribed genes (**Figure 5**). Thus, BLM protects against genome instability by suppressing recombination at sites of G4 structures, particularly in transcribed regions of the genome [71].

Using a zinc-finger nuclease translocation reporter assays, poly(ADP)ribose polymerase 3 (PARP3) was identified as a promoter of chromosomal rearrangements across human cell types. PARP3 regulates G4 DNA that inhibits repair by nonhomologous end-joining and homologous recombination (**Figure 5**). Chemical stabilization of G4 DNA in PARP3$^{-/-}$ cells leads to widespread DNA double-strand breaks and lethality. Thus, PARP3 suppresses G4 DNA and facilitates DNA repair by multiple pathways [72].

Recent *in vitro* experiments indicate strong correlation between G4 and R-loops (RNA-DNA hybrid). It was shown that transcription of G-rich DNA generates G4 structures composed of both non-sense DNA and transcribed RNA (hybrid G4) on RNA-DNA hybrid [15,16]. Specific G4 ligands stabilize G4s and simultaneously increase R-loop levels in human cancer cells. Genome-wide mapping of R loops showed that the studied G4 ligands likely cause the spreading of R loops to adjacent regions containing G4, preferentially at 3'-end regions of expressed genes. Overexpression of an exogenous human RNaseH1 rescued DNA damage induced by G4 ligands in BRCA2-proficient and BRCA2-silenced cancer cells. Thus, G4 ligands can induce DNA damage by an R loop-dependent mechanism [73].

3-3 Mechanisms of G4 recognition and unwinding by G4 helicases

A number of proteins that interact with G4 have been reported. Some of them are helicases that are capable of unfolding the G4 structures. There has been progress in understanding the structural basis of the G4 recognition and unwinding by these proteins (see also the article by Nakanishi and Seimiya in this issue).

DHX36, a DEAH-box helicase, unwinds G4 DNA in an ATP-dependent manner, but unfolds G4 RNA in an ATP-independent manner, which is followed by ATP-dependent refolding. The ATP-dependent activity of DHX36 arises from the RNA tail rather than the G4 [74]

The structure of a bacterial RecQ DNA helicase bound to unwound G4 indicates that a guanine base from the unwound G4 is sequestered within a guanine-specific binding pocket. Disruption of the pocket in RecQ blocks G4 unwinding, but not G4 binding or duplex DNA unwinding, indicating its essential role in structure-specific G4 resolution. [75]

The three representative G4 resolvases, DHX36, Bloom helicase (BLM), and Werner helicase (WRN), exhibit distinct G4 conformation specificity, but use a common mechanism of repetitive unfolding that leads to disrupting G4 structure multiple times in succession. The same resolving activity is sufficient to dislodge a stably bound G4 ligand, including BRACO-19, NMM, and Phen-DC3 [76].

Fragile X Mental Retardation Protein (FMRP) plays a central role in the development of several human disorders including Fragile X Syndrome (FXS) and autism. FMRP specifically interacts with G4 on mRNAs through an arginine-glycine-rich (RGG) motif. It recognizes a G4 RNA by a β-turn in the RGG motif of FMRP [77].

A 18-amino acid G4-binding domain of DHX36 binds to a parallel DNA G4 through the terminal guanine base tetrad (G-tetrad), and clamps the G4 using three-anchor-point electrostatic interactions between three positively charged amino acids and negatively charged phosphate groups. This mode of binding is strikingly similar to that of most ligands interacting with specific G4 [78].

More recently, co-crystal structure of DHX36 bound to a DNA with a G4 and a 3' single-stranded DNA segment was reported. The N-terminal DHX36-specific motif folds into a DNA-binding-induced α-helix that, together with the OB-fold-like subdomain, selectively binds parallel G4. G4 binding alone induces rearrangements of

the helicase core; by pulling on the single-stranded DNA tail, these rearrangements drive G4 unfolding one residue at a time [79].

Pif1 unwinds double-stranded DNA by a branched mechanism with two modes of activity. In the dominant mode, only short stretches of DNA can be processively and repetitively opened, with reclosure of the DNA occurring by mechanisms other than strand-switching. In the other less frequent mode, longer stretches of DNA are unwound via a path that is separate from the one leading to repetitive unwinding [80].

Three well-known G4 ligands (an oxazole telomestatin derivative, pyridostatin, and PhenDC3), which thermally stabilize G4 at different levels, were shown to inhibit BLM helicase activity at similar levels (2-3 fold). Thus, G4 ligands can inhibit a G4 resolvase activity, although the precise mechanism for this inhibition is not known [81].

3-4 G4 and transcription

Potential G4 forming sequences are known to be enriched in the promoter regions of actively transcribed genes, most notably a number of oncogenes, implicating the crucial roles of G4 in regulation of gene expression [1,82].

MYC, one of the most critical oncogenes, forms a DNA G4 in its proximal promoter region (MycG4) that functions as a transcriptional silencer. However, MycG4 is highly stable *in vitro* and its regulatory role would require active unfolding. DDX5, a DEAD-box RNA helicase, is extremely proficient at unfolding MycG4-DNA. DDX5 efficiently resolves MycG4-DNA in the absence of a single-stranded overhang and ATP hydrolysis is not directly coupled to G4-unfolding by DDX5. G-rich sequences are bound by DDX5 *in vivo*, and DDX5 is enriched at the MYC promoter and activates MYC transcription by resolving DNA and RNA G4s. On the other hand, G4 ligands inhibit DDX5-mediated MYC activation. Thus, DDX5 is a therapeutic candidate of Myc suppression for cancer intervention [83].

A small molecule, DC-34, was developed which downergulates Myc expression by a G4-dependent mechanism. Inhibition by DC-34 is significantly greater for MYC than for other G4-driven genes. The design of G4-interacting small molecules that can selectively control gene expression will have therapeutic merit [84].

3-5 G4 and epigenome regulation

ATRX (α-thalassemia mental retardation X-linked), an X-linked gene of the SWI/SNF family, is required for α-globin expression, and mutations in ATRX cause syndromal mental retardation. ATRX binds to tandem repeat (TR) sequences in both telomeres and euchromatin. Many of the TRs are G rich and predicted to form non-B DNA structures (including G4) *in vivo*. Indeed, ATRX binds G4 structures *in vitro*. Genes associated with these TRs can be dysregulated when ATRX is mutated, and the change in expression is determined by the size of the TR, producing skewed allelic expression [85].

Polycomb repressive complex 2 (PRC2) preferentially binds G tracts within nascent precursor mRNA (pre-mRNA), especially within predicted G4 structures. In cells, chromatin-associated G-tract RNA removes PRC2 from chromatin, leading to H3K27me3 depletion from genes. Targeting G-tract RNA to the tumor suppressor gene CDKN2A in malignant rhabdoid tumor cells reactivates the gene and induces senescence. These data indicate a novel role of G4 RNA in removal of an epigenome regulator from the chromatin [86] (**Figure 6**).

The presence of G4 structure is tightly associated with CpG islands (CGI) hypomethylation in the human genome. These G4 sites are enriched for DNA methyltransferase 1 (DNMT1) occupancy. Indeed, DNMT1 exhibits higher binding affinity for G4s as compared to duplex, hemi-methylated, or single-stranded DNA. G4 structure itself, rather than sequence, inhibits DNMT1 enzymatic activity. It was proposed that G4 formation sequesters DNMT1 thereby protecting certain CGIs from methylation and inhibiting local methylation. [87]

Heterochromatin (e.g., pericentromeres, centromeres, and telomeres) impedes replication fork progression, eventually leading, in the event of replication stress, to chromosome fragility, aging, and cancer. The shelterin subunit TRF2 ensures progression of the replication fork through pericentromeric heterochromatin, but not centromeric chromatin. TRF2 binds to pericentromeric Satellite III sequences during S phase, and recruits the G4-resolving helicase RTEL1 to facilitate fork progression [88].

Genome-wide ChIP-Seq analyses of TRF2-bound chromatin led to identification of thousands of TRF2-binding sites within the extra-telomeric genome. These sites are enriched in potential G4-forming DNA sequences. TRF2 binding altered expression and the epigenetic state of several target promoters, indicated by histone

modifications. Thus, TRF2 regulates expression of genes outside telomeres in a G4-dependent fashion [89].

The unstructured hinge domain, necessary for the targeting of HP1α to constitutive heterochromatin, recognizes parallel G4 formed by the Telomeric Repeat-containing RNA (TERRA) transcribed from the telomere. Through this mechanism, TERRA can enrich HP1α at telomeres to maintain heterochromatin [90].

3-6 Relevance of G4 to diseases

G4 and other non-B type DNA structures are formed on sequences with characteristic nucleotide repeats. Expansion of simple triplet repeats (TNR) underlies more than 30 severe degenerative diseases. These and other "repeat" syndromes may be associated with formation of aberrant G4 or other unusual structures on DNA and/ or transcribed RNA [91]. Cellular G4s profoundly affect the genomic stability through impeding replication fork progression at telomeres and at other chromosome arms, eventually leading to DSB, cell death, aging, genomic alterations, and cancer formation [92]. The specific G4 ligands can be exploited to suppress G4-induced genome instability or specific cancer cell killing.

CX-5461, a G-quadruplex stabilizer, exhibits specific toxicity against BRCA deficient cancer cells and polyclonal patient-derived xenograft models, including tumors resistant to PARP inhibition. CX-5461 blocks replication forks and induces ssDNA gaps or breaks. The BRCA and NHEJ pathways are required for the repair of G4 ligand-induced DNA damage and failure to do so leads to lethality. These data strongly support the idea that G4 targeting in HR and NHEJ deficient cancer cells can be an efficient therapeutic choice [93].

Cockayne syndrome is a neurodegenerative accelerated aging disorder caused by mutations in the CSA or CSB genes. Loss of CSA or CSB leads to polymerase stalling at non-B DNA in a neuroblastoma cell line, in particular at G4 structures, and recombinant CSB can melt G4 structures (**Figure 4**). Stabilization of G4 structures activates the DNA damage sensor poly-ADP ribose polymerase 1 (PARP1) and leads to accelerated aging in Caenorhabditis elegans. Thus, transcription-coupled resolution of secondary structures may be a mechanism to repress spurious activation of a DNA damage response. [94]

Replication efficiency of guanine-rich (G-rich) telomeric repeats is decreased

significantly in cells lacking homologous recombination (HR). Treatment with the G4 ligand, pyridostatin (PDS), increases telomere fragility in BRCA2-deficient cells, suggesting that G4 formation drives telomere instability. PDS induces DSB accumulation, checkpoint activation, and deregulated G2/M progression and enhances the replication defect intrinsic to HR deficiency, reducing cell proliferation of HR-defective cells. PDS toxicity extends to HR-defective cells that have acquired olaparib resistance through loss of 53BP1 or REV7. These results highlight the therapeutic potential of G4-stabilizing ligand to selectively eliminate HR-compromised cells and tumors, including those resistant to PARP inhibition. [95]

Inactivation of ATRX (α-thalassemia mental retardation X-linked) defines the mutation in large subsets of malignant glioma. Loss of ATRX induces replication stress and DNA damage by way of G4 DNA. ATRX deficiency in conjunction with chemical G4 stabilization enhances DNA damage and cell death. Other DNA-damaging therapies, including ionizing radiation, synergize with ATRX-deficiency, suggesting a potential new strategies for drug development [96].

ATRX normally binds to G4 in CpG islands of the imprinted Xlr3b gene, inhibiting its expression by recruiting DNA methyltransferases. Atrx mutation upregulates Xlr3b expression in the mouse brain. Xlr3b binds to dendritic mRNAs, and its overexpression inhibits dendritic transport of the mRNA encoding CaMKII-α, promoting synaptic dysfunction. Treatment with 5-ALA, which is converted into G-quadruplex-binding metabolites in a body, reduces RNA polymerase II recruitment, represses Xlr3b transcription and restores synaptic plasticity and cognitive proficiency in ATRX model mice. This suggests a potential therapeutic strategy to target G4 and decrease cognitive impairment associated with ATRX syndrome [97].

3-7 Novel G4 binding proteins

Guanine-rich sequences are prevalent on both RNA and DNA, and they have high propensity to form G4 and frequently play important roles in regulation of various chromatin events, as stated above. Recent reports continue to add new members to the list of proteins that can bind to RNA/ DNA structures generated by these guanine-rich sequences.

Polycomb repressive complex 2 (PRC2) is a histone methyltransferase that trimethylates H3K27, a mark of repressed chromatin. The purified human PRC2 has a

high affinity for folded G4 structures but shows little binding to duplex RNAs. G-tract motifs are significantly enriched among PRC2-binding transcripts *in vivo*. DNA sequences coding for PRC2-binding RNA motifs are enriched at PRC2-binding sites on chromatin and H3K27me3-modified nucleosomes. The results indicate the significance of G4-like structures for epigenome regulation by PRC2 [98].

Activation-induced cytidine deaminase (AID) initiates both class switch recombination (CSR) and somatic hypermutation (SHM) in antibody diversification. *In vitro* experiments indicated that G4-containing substrates mimicking the mammalian immunoglobulin switch regions are particularly good AID substrates. Crystal structures of AID alone and that in complex with deoxycytidine monophosphate showed a bifurcated substrate-binding surface that explains structured substrate recognition by capturing two adjacent single-stranded overhangs simultaneously. Moreover, G4 substrates induce cooperative AID oligomerization. Mutations disrupting bifurcated substrate recognition or oligomerization compromise CSR in splenic B cells. Thus, G4 recognition of AID and its oligomerization plays a crucial role in CSR [99]. For CSR, see also section 3-10.

3-8 G4 nucleic acids and viruses

Viruses carry dynamic genomes which appear as duplex or single-stranded DNA/RNAs as well as in other special forms of nucleic acids. Therefore, the chances of formation of G4 on virus genomes and its transcripts are high. The oncogenic Epstein-Barr virus (EBV) evades the immune system by suppressing the essential EBNA1 protein to the minimal level necessary for function while minimizing immune recognition. EBNA1 contains the glycine-alanine repeats (GAr) which forms a G4 in its mRNA. Nucleolin (NCL) binds to the G4 RNA, and inhibits the EBNA1 expression, whereas its depletion relieves the suppression. Moreover, the G-quadruplex ligand PhenDC3 prevents NCL binding to EBNA1 mRNA and reverses GAr-mediated repression of EBNA1 expression. Thus, the NCL-EBNA1 mRNA interaction is a promising therapeutic target to trigger an immune response against EBV-carrying cancers [100].

The Herpes Simplex Virus-1 (HSV-1) genome contains multiple clusters of repeated G-quadruplexes [101]. Treatment of HSV-1 infected cells with a core extended aphtalene diimide (c-exNDI) G4 ligand induced significant virus inhibition with no cytotoxicity. This was due to G4-mediated inhibition of viral DNA replication. c-exNDI

preferentially recognizes HSV-1 G4s over cellular telomeric G4s [102].

Adeno-associated viruses (AAV) associate with the DNA binding protein nucleophosmin (NPM1). Nucleophosmin was shown to interact with G4-forming sequences in the AAV genome, facilitating AAV production [103]

The long terminal repeat (LTR) of the proviral human immunodeficiency virus (HIV)-1 genome is integral to virus transcription and host cell infection. The guanine-rich U3 region within the LTR promoter forms a parallel-stranded G4, which stacks with neighboring G4 through its single-nucleotide bulges [104].

3-9 G4 and phase separation

A great deal of attention is being drawn to the membrane-less organelles observed in cells, which are now known to be formed by liquid-liquid phase separation (LLPS) caused by proteins, nucleic acids and their interactions [105]. RNA associated with repeat expansions can phase separate *in vitro* and in cells. A hexanucleotide expansion GGGGCC (G4C2), found in the C9ORF72 gene is the most common mutation associated with amyotrophic lateral sclerosis and frontotemporal dementia (C9-ALS/FTD). ALS/FTD GGGGCC repeat RNA (rG4C2) drives phase separations. The phase transition occurs at a similar critical repeat number as observed in the diseases [105]. Using rG4C2 as a model RNA, phase separation was observed *in vitro* and in cells in length- and structure-dependent manner. It was also shown that G-quadruplex structures are required for rG4C2-mediated phase separation. Recent report indicates that G4 RNA with more G-quartets and longer loops are more likely to form phase separation. These results suggest a possibility that G4 RNA can induce liquid phase separation in cells, although its physiological significance needs to wait for further analyses [106-108].

3-10 Other important biological roles of RNA G4/ DNA G4 and their binding proteins

Long interspersed nuclear elements (LINEs) are ubiquitous transposable elements in higher eukaryotes that constitute a large portion of the genomes. Guanine-rich sequences in the 3' untranslated regions (UTRs) of hominoid-specific LINE-1 elements are coupled with retrotransposon speciation and contribute to retrotransposition through the formation of G4 structures. Stabilization of the G4 motif of a human-specific

LINE-1 element by small-molecule ligands stimulates retrotransposition [109].

The targets of DHX36, one of the most potent G4 resolvases, are preferentially localized in stress granules. DHX36 binds to G4-forming sequences on mRNAs and reduces their abundance. Knockout of DHX36 results in increase of stress granule formation. Thus, DHX36 may facilitate the resolution of G4 RNA induced by cellular stress [110].

RNA elements with G4-forming capacity promote exon inclusion. Disruption of G4-forming capacity abrogates exon inclusion. G4 are enriched in heterogeneous nuclear ribonucleoprotein F (hnRNPF)-binding sites and near hnRNPF-regulated alternatively spliced exons in the human transcriptome. Moreover, hnRNPF regulates an epithelial-mesenchymal transition (EMT)-associated CD44 isoform switch in a G4-dependent manner, resulting in inhibition of EMT. These data suggest a critical role for RNA G4 in regulating alternative splicing [111]. Small molecules were identified that disrupt RNA G4 and inhibit G4-dependent alternative splicing. [112].

An anionic phthalocyanine, ZnAPC, binds to a G4-forming oligonucleotide derived from the 5'-untranslated region of N-RAS mRNA, resulting in selective cleavage of the target RNA's G4 upon photo-irradiation. Upon photo-irradiation, ZnAPC decreases N-RAS mRNA and NRAS expression and thus viability of cancer cells [113].

The formation of R-loop structures during CSR can occur as part of a post-transcriptional mechanism targeting AID to IgH S-regions. This depends on the RNA helicase DDX1, which binds to G4 structures present in intronic switch transcripts and converts them into S-region R-loops. This recruits the cytidine deaminase enzyme AID to S-regions, promoting CSR. Chemical stabilization of G4 RNA or the expression of a DDX1 ATPase-deficient mutant reduces R-loop levels over S-regions. The DDX 1 mutant acts as a dominant-negative protein and reduce CSR efficiency. Thus, S-region transcripts interconvert between G4 and R-loop structures to promote CSR in the IgH locus by the action of DDX1 [114].

The mitochondrial transcription factor A (TFAM), a high-mobility group (HMG)-box protein, plays a critical role in its expression and maintenance. TFAM binds to DNA G4 or RNA G4. TFAM multimerizes in a manner dependent on its G4 binding. Thus, TFAM-G4 interaction is important for mitochondrial DNA replication and transcription [115].

Heme is an essential cofactor for many enzymes, but free heme is toxic and its levels are tightly regulated. G4 binds heme avidly *in vitro*, and presumably in cells as well. Addition PhenDC3, a G4 ligand, displaces G4-bound heme, increasing the expression of heme-regulated genes. It was proposed that G4 sequesters heme to protect cells from the pathophysiological consequences of free heme [116].

ROS-mediated oxidation of DNA (yielding 8-oxo-7,8-dihydroguanine [OG] in gene promoters) is a signaling agent for gene activation. When OG is formed in guanine-rich, potential G4-forming sequences in promoter-coding strands, base excision repair (BER) by 8-oxoguanine DNA glycosylase (OGG1) is initiated, yielding an abasic site (AP). The AP enables melting of the duplex to unmask the G4-forming sequences, adopting a G4 fold in which apurinic/apyrimidinic endonuclease 1 (APE1) binds, but inefficiently cleaves the AP for activation of vascular endothelial growth factor (VEGF) or endonuclease III-like protein 1 (NTHL1) genes. Thus, G4 plays a role for epigenetic regulation of transcription [117].

In contrast to DNA G4, RNA G4 may be more readily formed on G-rich sequences, once it is transcribed. However, in some cases, additional step may be required for forming RNA G4. It was reported that RNA G4 can be generated from cleaved tRNA fragments, which augments the formation of stress granules by suppressing translation [118].

4 Dynamic regulation of formation and disruption of cellular G-quadruplex

Initial bioinformatics analyses indicated the presence of 375,000 potential G4 forming sequences on the human genome [1,2,119]. On the other hand, analyses with "G4-seq" method, that utilizes the polymerase stop reaction by the presence of G4, revealed 716,310 G4 structures, and 451,646 of them were not predicted by computational methods [5]. In contrast, genome wide search of cellular G4 by G4-specific antibody, BG4, revealed the presence of 1,496 and 10,560 G4 in non-cancer cells (NHEK, normal human epidermal keratinocytes) and in cancer cells (HaCaT, spontaneously immortalized counterparts of NHEK), respectively [59]. The increase of G4 in immortalized cells may reflect the active transcription and relaxed chromatin compared to the non-immortalized cells. These numbers are significantly smaller than that predicted by sequences and G4-seq, suggesting that most of the predicted G4 are not formed in the cells. However, this could be due to failure to maintain the structure during the detection

process and/or to the limitation of the BG4, which may recognize only a subset of G4 structures.

Among the 35 strong Rif1BS which form G4 *in vitro* and are very likely to form G4 in cells as well, only two overlaps with 446 potential G4-forming sequences predicted from the computational analyses [42,120]. This strongly suggests 1) it is difficult to precisely predict the G4 functional in cells simply by analyzing the sequences, and 2) there are different classes of G4 that may play differential roles in distinct chromosome reactions presumably through interacting with specific G4 binding proteins. The sequences of Rif1BS, which span close to 100 bp in some cases, are characterized by the presence of multiple stretches of long G-tract (5 or 6 guanines) and unusually long loop sequences (more that 20 nt). The detailed nuclease mapping demonstrated the presence of long loops in the G4 structure formed on the duplex DNA containing Rif1BS [121]. The formation of the higher order structure depends on the G-tracts on the G-rich strand and there is no indication for the formation of an intercalated motif (i-motif) on the complementary C-rich strand. The presence of multiple G-tracts within Rif1BS permits generation of an alternative G4 structure when some of the G-tracts are mutated [121]. This suggests a potential of G-rich G4-forming sequences to adopt different configurations of DNA structures within cells in response to changes of chromatin state or microenvironment. The formation of G4 structure on the G-rich sequences with potentially long loop segments was suggested also from the genome-wide "G4-seq" analyses on the human genome [5].

The formation of G4 depends on the presence of monovalent cation such as K^+ or Na^+ in the center of the G-quartet that has an anionic space formed by O6 of the guanine residues. Li^+ does not support the G4 formation and rather destabilizes the G4 [122,123]. G4 formation depends on Hoogsteen hydrogen bonding in stead of the Watson-Crick-type hydrogen bonding that is required for B-type double helix structure [124].

In cells, G4 needs to be formed in competition with duplex DNA form. G4 could be formed during the process that can transiently generate single-stranded DNA (**Figure 7**). DNA replication is likely to provide such opportunity. Indeed, many reports support the notion that G4 formed on the lagging strand template strand during the course of DNA replication can cause replication fork barrier that arrests the fork progression. A group of helicases are known that can unwind the G4 structure

[68,86,125] (**Figure 4**). These G4 helicases can counteract the G4 fork barrier, enabling the replication to continue (see other sections). It was reported that depletion of a helicase resulted in generation of double-stranded DNA breaks and sensitivity of growth to G4 ligands.

Transcription likewise provides an opportunity of forming single-strand. Transcription of duplex DNA containing G-clusters on the non-template strand generates slow-migrating form on a polyacrylamide gel. This structure is sensitive to RNaseH, suggesting the presence of RNA-DNA hybrid. Mutations of the guanines on the non-template strand or transcribed RNA lead to loss of the structure. Replacement of the guanine on the non-template DNA or on the transcribed RNA with deaza-guanine leads to loss of this form. 7^{th} nitrogen atom is crucial for Hoogsteen hydrogen bonding, which is a basis of G-quartet structure. All these data are consistent with the idea that RNA-DNA hybrid generated by transcription of guanine-rich double-stranded DNA carries G4 composed of both RNA and DNA [13-17, 126-129]

As stated above, the formation of G4 depends on the presence of various coordinating metals. Its stability may be affected by temperature, pH, molecular crowding in the cellular microenvironment (**Figure 7**). Epigenomic state, such as CpG methylation affects the stability of G4 (130-132). The presence of G4 was reported to be associated with CpG island hypomethylation in the human genome. This appears to be caused by the ability of G4 to inhibit DNA methyltransferase 1 enzymatic activity [87]. Thus, G4 and CpG methylation affect with each other. Other factors influencing the G4 formation/ stability include local topology (superhelicity) or chromatin architecture [133]. Stability of G4 *in vitro* varies greatly depending on the sequence or the solution condition. In cells where nucleic acids exist with many complex factors and under varied microenvironmental conditions, some of them may be stably formed, whereas others may be unstable, undergoing more dynamic transition between formation and disruption.

5 Conclusions

G4 appears both in DNA and RNA, and potentially in mixed RNA/DNA on RNA-DNA hybrids. Recent studies show a varieties of biological roles associated with nucleic acids forming G4 or its related structures. In DNA replication, replication timing during S phase is regulated by an evolutionally conserved Rif1 protein which specifically binds

to G4. Rif1 may regulate chromatin architecture by tethering chromatin fiber through binding to multiple G4. G4 has been reported to be associated with replication origins in higher eukaryotes and evidence for its potential functional importance in initiation of bacterial genome replication has been presented. Thus, G4 may serve as both genomic marks for initiation of DNA replication and platforms for generating chromatin compartments that are inhibitory for initiation of DNA replication.

Diverse roles of DNA G4 and RNA G4 have been reported for many genome and epigenome regulations as well as in RNA metabolism. Proteins that bind to G4 exert various effects on chromosome events. Some can stabilize G4 and others can disrupt it by its dissolvase activity. Some recruits other proteins to regulate the process. Some change the chromatin structures.

A low molecular weight chemicals that specifically binds to G4 can be exploited to regulate these processes. Various chemicals are generated which can even discriminate different G4 at different locations or G4 with different topology (see Nagasawa's article in this special issue). These highly selective G4 ligands will be very useful reagents that could be developed for efficient anti-cancer drugs (see Nakanishi-Seimiya's article in this special issue).

Acknowledgements

I would like to thank Naoko Kakusho for help in the preparation of this manuscript. This work was supported by JSPS KAKENHI (Grant-in-Aid for Scientific Research (A) [Grant Numbers 23247031 and 26251004] to HM. We also would like to thank the Uehara Memorial Foundation and the Takeda Science Foundation for the financial support of this study. We would like to thank all the members of the laboratory for helpful discussion.

References

[1] Huppert JL, Balasubramanian S, G-quadruplexes in promoters throughout the human genome, Nucleic Acids Res 35 (2007) 406-413.

[2] Huppert JL, Bugaut A, Kumari S, Balasubramanian S, G-quadruplexes: the beginning and end of UTRs, Nucleic Acids Res 36 (2008) 6260-6268.

[3] Todd AK, Johnston M, Neidle S, Highly prevalent putative quadruplex sequence motifs in human DNA, Nucleic Acids Res 33 (2005) 2901-2907.

[4] Bedrat, A., Lacroix, L. & Mergny, J.-L, Re-evaluation of G-quadruplex propensity with G4Hunter, Nucleic Acid. Res. 44 (2016) 1746–1759.

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

[6] Kwok CK, Marsico G, Sahakyan AB, Chambers VS, BalasubramanianS, rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome, Nat Methods 13 (2016) 841-844.

[7] Brázda V, Hároníková L, Liao JC, Fojta M, DNA and RNA quadruplex-binding proteins, Int. J. Mol. Sci. 15 (2014) 17493-17517.

[8] Sauer M, Paeschke K, G-quadruplex unwinding helicases and their function in vivo, Biochem Soc Trans 45 (2017) 1173-1182.

[9] Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, Méchali M, Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features, Genome Res. 21 (2011) 1438-1449.

[10] Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, Marin JM, Lemaitre JM, Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs, Nat Struct Mol Biol. 19 (2012) 837-844.

[11] Langley AR, Gräf S, Smith JC, Krude T, Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq), Nucleic Acids Res. 44 (2016) 10230-10247.

[12] Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintomé C, Riou JF, Prioleau MN, G4 motifs affect origin positioning and efficiency in two vertebrate replicators, EMBO J. 33 (2014) 732-746.

[13] Kogoma T, Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription, Microbiol Mol Biol Rev. 61 (1997) 212-238.

[14] Shrestha P, Xiao S, Dhakal S, Tan Z, Mao H, Nascent RNA transcripts facilitate the formation of G-quadruplexes, Nucleic Acids Res. 42 (2014) 7236-7346.

[15] Zhang JY, Zheng KW, Xiao S, Hao YH, Tan Z, Mechanism and manipulation of DNA:RNA hybrid G-quadruplex formation in transcription of G-rich DNA, J Am Chem Soc. 136 (2014) 1381-1390.

[16] Xiao S, Zhang JY, Zheng KW, Hao YH, Tan Z, Bioinformatic analysis reveals an evolutional selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals, Nucleic Acids Res. 41 (2013) 10379-10390.

[17] Zheng KW, Xiao S, Liu JQ, Zhang JY, Hao YH, Tan Z, Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control, Nucleic Acids Res. 41 (2013) 5533-5541.

[18] Zhang C, Liu HH, Zheng KW, Hao YH, Tan Z, DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix, Nucleic Acids Res. 41 (2013) 7144-7152.

[19] Rowley MJ, Corces VG. Organizational principles of 3D genome architecture, Nat Rev Genet. 19 (2018) 789-800.

[20] Ryba, T., et al., Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types, Genome Res. **20**(2010) 761-770.

[21] Rhind, N. and D.M. Gilbert, DNA replication timing, Cold Spring Harb Perspect Biol, **5** (2013) a010132.

[22] Yamazaki S, Hayano M, Masai H, Replication timing regulation of eukaryotic replicons: Rif1 as a global regulator of replication timing, Trends Genet. 29 (2013) 449-460.

[23] Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M, Eukaryotic chromosome DNA replication: where, when, and how? , Annu Rev Biochem.79 (2010) 89-130.

[24] Ge, X.Q., D.A. Jackson, and J.J. Blow, Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress, Genes Dev, 21 (2007) 3331-3341.

[25] Knott SR, Peace JM, Ostrow AZ, Gan Y, Rex AE, Viggiani CJ, Tavaré S, Aparicio OM. Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae, Cell. 148 (2012) 99-111.

[26] Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen MT, Rivera-Mulia JC, Trevilla-Garcia C, Bartlett DA, Zhao PA, Washburn BK, Nora EP, Kraft K, Mundlos S, Bruneau BG, Ljungman M, Fraser P, Ay F, Gilbert DM. Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication, Cell. 176 (2019) 816-830.

[27] Hayano M, Kanoh Y, Matsumoto S, Renard-Guillet C, Shirahige K, Masai H, Rif1 is a global regulator of timing of replication origin firing in fission yeast, Genes Dev. 26 (2012) 137-150.

[28] Yamazaki S, Ishii A, Kanoh Y, Oda M, Nishito Y, Masai H, Rif1 regulates the replication timing domains on the human genome, EMBO J. 31 (2012) 3667-3677.

[29] Cornacchia D, Dileep V, Quivy JP, Foti R, Tili F, Santarella-Mellwig R, Antony C, Almouzni G, Gilbert DM, Buonomo SB Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells, EMBO J. 31 (2012) 3678-3690.

[30] Hardy CF, Sussel L, Shore D, A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation, Genes Dev. 6 (1992) 801-814.

[31] Kanoh J, Ishikawa F, spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast, Curr Biol. 11 (2001) 1624-1630.

[32] Masai H, Miyake T, Arai K, hsk1+, a Schizosaccharomyces pombe gene related to Saccharomyces cerevisiae CDC7, is required for chromosomal replication, EMBO J. 14 (1995) 3094-3104.

[33] Matsumoto S, Hayano M, Kanoh Y, Masai H, Multiple pathways can bypass the essential role of fission yeast Hsk1 kinase in DNA replication initiation, J Cell Biol. 195 (2011) 387-401.

[34] Hayano M, Kanoh Y, Matsumoto S, Masai H, Mrc1 marks early-firing origins and coordinates timing and efficiency of initiation in fission yeast, Mol Cell Biol. 31 (2011) 2380-2391.

[35] Xu L, Blackburn EH　Human Rif1 protein binds aberrant telomeres and aligns along anaphase midzone microtubules, J Cell Biol. 167 (2004) 819–830.

[36] Silverman J, Takai H, Buonomo SB, Eisenhaber F, de Lange T, Human Rif1, ortholog of a yeast telomeric protein, is regulated by ATM and 53BP1 and functions in the S-phase checkpoint, Genes Dev. 18 (2004) 2108-2119.

[37] Alver RC, Chadha GS, Gillespie PJ, Blow JJ, Reversal of DDK-Mediated MCM Phosphorylation by Rif1-PP1 Regulates Replication Initiation and Replisome Stability Independently of ATR/Chk1, Cell Rep. 18 (2017) 2508-2520.

[38] Sreesankar E, Senthilkumar R, Bharathi V, Mishra RK, Mishra K, Functional diversification of yeast telomere associated protein, Rif1, in higher eukaryotes, 13 (2012) 255.

[39] Davé A, Cooley C, Garg M, Bianchi A, Protein phosphatase 1 recruitment by Rif1 regulates DNA replication origin firing by counteracting DDK activity, Cell Rep. 7 (2014) 53-61.

[40] Hiraga S, Alvino GM, Chang F, Lian HY, Sridhar A, Kubota T, Brewer BJ, Weinreich M, Raghuraman MK, Donaldson AD, Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex, Genes Dev. 28 (2014) 372-383.

[41] Hiraga SI, Ly T, Garzón J, Hořejší Z, Ohkubo YN, Endo A, Obuse 5, Boulton SJ, Lamond AI, Donaldson AD, Human RIF1 and protein phosphatase 1 stimulate DNA replication origin licensing but suppress origin activation, EMBO Rep. 18 (2017) 403-419.

[42] Kanoh Y, Matsumoto S, Fukatsu R, Kakusho N, Kono N, Renard-Guillet C, Masuda K, Iida K, Nagasawa K, Shirahige K, Masai H, Rif1 binds to G quadruplexes and suppresses replication over long distances, Nat Struct Mol Biol. 22 (2015) 889-897.

[43] Kobayashi S, Fukatsu R, Kanoh Y, Kakusho N, Matsumoto S, Chaen S, Masai H, Both a Unique Motif at the C Terminus and an N-Terminal HEAT Repeat Contribute to G-Quadruplex Binding and Origin Regulation by the Rif1 Protein, Mol Cell Biol. 39 (2019) e00364-18.

[44] Masai H, Fukatsu R, Kakusho N, Kanoh Y, Moriyama K, Ma Y, Iida K, Nagasawa K, Rif1 promotes association of G-quadruplex (G4) by its specific G4 binding and oligomerization activities, Sci Rep. 9 (2019) 8618.

[45] Foti R, Gnan S, Cornacchia D, Dileep V, Bulut-Karslioglu A, Diehl S, Buness A, Klein FA, Huber W, Johnstone E, Loos R, Bertone P, Gilbert DM, Manke T, Jenuwein T, Buonomo SC, Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program, Mol Cell. 61 (2016) 260-273.

[46] Moriyama K, Yoshizawa-Sugata N, Masai H, Oligomer formation and G-quadruplex binding by purified murine Rif1 protein, a key organizer of higher-order chromatin architecture, J Biol Chem. 293 (2018) 3607-3624

[47] Shi T, Bunker RD, Mattarocci S, Ribeyre C, Faty M, Gut H, Scrima A, Rass U, Rubin SM, Shore D, Thomä NH, Rif1 and Rif2 shape telomere function and architecture through multivalent Rap1 interactions, Cell 153 (2013) 1340-1353.

[48] Park S, Patterson EE, Cobb J, Audhya A, Gartenberg MR, Fox CA, Palmitoylation controls the dynamics of budding-yeast heterochromatin via the telomere-binding protein Rif1, Proc Natl Acad Sci U S A. 108 (2011) 14572-14577.

[49] Fontana GA, Hess D, Reinert JK, Mattarocci S, Falquet B, Klein D, Shore D, Thomä NH, Rass U, Rif1 S-acylation mediates DNA double-strand break repair at the inner nuclear membrane. Nat Commun. 10 (2019) 2535.

[50] Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, Segal E, Méchali M. New insights into replication origin characteristics in metazoans. Cell Cycle 11(2012) 658-667.

[51] Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, Méchali M, The chromatin environment shapes DNA replication origin organization and defines origin classes, Genome Res. 25 (2015) 1873-1885

[52] Prorok P, Artufel M, Aze A, Coulombe P, Peiffer I, Lacroix L, Guédin A, Mergny JL4, Damaschke J, Schepers A, Ballester B, Méchali M, Involvement of G-quadruplex regions in mammalian replication origin activity, Nat Commun. 10 (2019) 3274.

[53] Hoshina S, Yura K, Teranishi H, Kiyasu N, Tominaga A, Kadoma H, Nakatsuka A, Kunichika T, Obuse C, Waga S, Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA, J Biol Chem. 288 (2013) 30161-30171.

[54] Masai H, Asai T, Kubota Y, Arai K, Kogoma T, Escherichia coli PriA protein is essential for inducible and constitutive stable DNA replication, EMBO J. 13 (1994) 5338-5345.

[55] Zheng KW, Wu RY, He YD, Xiao S, Zhang JY, Liu JQ, Hao YH, Tan Z, A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site, Nucleic Acids Res. 42 (2014) 10832-10844.

[56] Kogoma T, Absence of RNase H allows replication of pBR322 in Escherichia coli mutants lacking DNA polymerase I, Proc Natl Acad Sci U S A. 81 (1984) 7845-7849.

[57] Parada CA, Marians KJ, Transcriptional activation of pBR322 DNA can lead to duplex DNA unwinding catalyzed by the Escherichia coli preprimosome, J Biol Chem. 264 (1989) 15120-15129.

[58] Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, Allers T, Accelerated growth in the absence of DNA replication origins. Nature. 503 (2013) 544-547.

[59] Hansel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, Narita M, et al: G-quadruplex structures mark human regulatory chromatin. Nat Genet 2016, 48:1267-1272.

[60] Clynes D, Jelinska C, Xella B, Ayyub H, Scott C, Mitson M, Taylor S, Higgs DR, Gibbons RJ, Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX, Nat Commun. 6 (2015) 7738.

[61] Audry J, Maestroni L, Delagoutte E, Gauthier T, Nakamura TM, Gachet Y, Saintomé C, Géli V, Coulon S, RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends, EMBO J. 34 (2015) 1942-1958.

[62] Zhang M, Wang B, Li T, Liu R, Xiao Y, Geng X, Li G, Liu Q, Price CM, Liu Y, Wang F, Mammalian CST averts replication failure by preventing G-quadruplex accumulation, Nucleic Acids Res 47 (2019) 5243-5259.

[63] Vannier JB, Pavicic-Kaltenbrunner V, Petalcorin MI, Ding H, Boulton SJ, Cell. 149 (2012) 795-806.

[64] Fouquerel E, Lormand J, Bose A, Lee HT, Kim GS, Li J, Sobol RW, Freudenthal BD, Myong S, Opresko PL, Oxidative guanine base damage regulates human telomerase activity, Nat Struct Mol Biol. 23 (2016) 1092-1100.

[65] Moye AL, Porter KC, Cohen SB, Phan T, Zyner KG, Sasaki N, Lovrecz GO, Beck JL, Bryan TM, Telomeric G-quadruplexes are a substrate and site of localization for human telomerase, Nat Commun. 6 (2015) 7643.

[66] Amato J, Cerofolini L, Brancaccio D, Giuntini S, Iaccarino N, Zizza P, Iachettini S, Biroccio A, Novellino E, Rosato A, Fragai M, Luchinat C, Randazzo A, Pagano B, Insights into telomeric G-quadruplex DNA recognition by HMGB1 protein, Nucleic Acids Res. 47 (2019) 9950-9966.

[67] Piazza A, Adrian M, Samazan F, Heddi B, Hamon F, Serero A, Lopes J, Teulade-Fichou MP, Phan AT, Nicolas A, Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites, EMBO J. 34 (2015) 1718-1734.

[68] Paeschke K, Capra JA, Zakian VA, DNAreplication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase, Cell 145 (2011) 678-691.

[69] Lopez CR, Singh S, Hambarde S, Griffin WC, Gao J, Chib S, Yu Y, Ira G, Raney KD, Kim N, Yeast Sub1 and human PC4 are G-quadruplex binding proteins that suppress genome instability at co-transcriptionally formed G4 DNA, Nucleic Acids Res. 45 (2017) 5850-5862.

[70] Wanzek K, Schwindt E, Capra JA, Paeschke K, Mms1 binds to G-rich regions in Saccharomyces cerevisiae and influences replication and genome stability, Nucleic Acids Res.45 (2017) 7796-7806.

[71] van Wietmarschen N, Merzouk S, Halsema N, Spierings DCJ, Guryev V, Lansdorp PM, BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes, Nat Commun. 9 (2018) 271.

[72] Day TA, Layer JV, Cleary JP, Guha S, Stevenson KE, Tivey T, Kim S, Schinzel AC, Izzo F, Doench J, Root DE, Hahn WC, Price BD, Weinstock DM, PARP3 is a promoter of chromosomal rearrangements and limits G4 DNA, Nat Commun. 8 (2017) 15110.

[73] De Magis A, Manzo SG, Russo M, Marinello J, Morigi R, Sordet O, Capranico G, DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells, Proc Natl Acad Sci U S A. 116 (2019) 816-825.

[74] Tippana R, Chen MC, Demeshkina NA, Ferré-D'Amaré AR, Myong S, RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36, Nat Commun. 23 (2019) 1855.

[75] Voter AF, Qiu Y, Tippana R, Myong S, Keck JL, A guanine-flipping and sequestration mechanism for G-quadruplex unwinding by RecQ helicases, Nat Commun. 9 (2018) 4201.

[76] Tippana R, Hwang H, Opresko PL, Bohr VA, Myong S, Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases, Proc Natl Acad Sci U S A. 113 (2016) 8448-8453.

[77] Vasilyev N, Polonskaia A, Darnell JC, Darnell RB, Patel DJ, Serganov A, Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP, Proc Natl Acad Sci U S A. 112 (2015) E5391-400.

[78] Heddi B, Cheong VV, Martadinata H, Phan AT, Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex, Proc Natl Acad Sci U S A. 112 (2015) 9608-9613.

[79] Chen MC, Tippana R, Demeshkina NA, Murat P, Balasubramanian S, Myong S, Ferré-D'Amaré AR, Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36, Nature 558 (2018) 465-469.

[80] Singh SP, Soranno A, Sparks MA, Galletto R3, Branched unwinding mechanism of the Pif1 family of DNA helicases, Proc Natl Acad Sci U S A. 116 (2019) 24533-24541.

[81] Maleki P, Mustafa G, Gyawali P, Budhathoki JB, Ma Y, Nagasawa K, Balci H, Quantifying the impact of small molecule ligands on G-quadruplex stability against Bloom helicase, Nucleic Acids Res. 47 (2019) 10744-10753.

[82] Eddy J, Maizels N, Gene function correlates with potential for G4 DNA formation in the human genome, Nucleic Acids Res. 34 (2006) 3887-3896.

[83] Wu G, Xing Z, Tran EJ, Yang D, DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation, Proc Natl Acad Sci U S A. 116 (2019) 20453-20461.

[84] Calabrese DR, Chen X, Leon EC, Gaikwad SM, Phyo Z, Hewitt WM, Alden S, Hilimire TA, He F, Michalowski AM, Simmons JK, Saunders LB, Zhang S, Connors D, Walters KJ, Mock BA, Schneekloth JS Jr, Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex, Nat Commun. 9 (2018) 4229.

[85] Law MJ, Lower KM, Voon HP, Hughes JR, Garrick D, Viprakasit V, Mitson M, De Gobbi M, Marra M, Morris A, Abbott A, Wilder SP, Taylor S, Santos GM, Cross J, Ayyub H, Jones S, Ragoussis J, Rhodes D, Dunham I, Higgs DR, Gibbons RJ, ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner, Cell. 143 (2010) 367-378.

[86] Beltran M, Tavares M, Justin N, Khandelwal G, Ambrose J, Foster BM, Worlock KB, Tvardovskiy A, Kunzelmann S, Herrero J, Bartke T, Gamblin SJ, Wilson JR, Jenner RG, G-tract RNA removes Polycomb repressive complex 2 from genes, Nat Struct Mol Biol. 26 (2019) 899-909.

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

[88] Mendez-Bermudez A, Lototska L, Bauwens S, Giraud-Panis MJ, Croce O, Jamet K, Irizar A, Mowinckel M, Koundrioukoff S, Nottet N, Almouzni G, Teulade-Fichou MP, Schertzer M, Perderiset M, Londoño-Vallejo A, Debatisse M, Gilson E, Ye J, Genome-wide Control of Heterochromatin Replication by the Telomere Capping Protein TRF2, Mol Cell.70 (2018) 449-461.

[89] Mukherjee AK, Sharma S, Bagri S, Kutum R, Kumar P, Hussain A, Singh P, Saha D, Kar A, Dash D, Chowdhury S, Telomere repeat-binding factor 2 binds extensively to extra-telomeric G-quadruplexes and regulates the epigenetic status of several gene promoters, J Biol Chem. 294 (2019) 17709-17722.

[90] Roach RJ, Garavís M, González C, Jameson GB, Filichev VV, Hale TK, Heterochromatin protein 1α interacts with parallel RNA and DNA G-quadruplexes, Nucleic Acids Res. 48 (2020) 682-693.

[91] Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ,Troncoso JC, Pandey A, Sattler R, Rothstein JD, Wang J, C9orf72 nucleotide repeat structures initiatemolecular cascades of disease, Nature 507 (2014) 195-200.

[92] Bochman ML, Paeschke K, Zakian VA, DNA secondary structures: stability and function of G-quadruplex structures, Nat Rev Genet. 13 (2012) 770-780.

[93] Xu H, Di Antonio M, McKinney S, Mathew V, Ho B, O'Neil NJ, Santos ND, Silvester J, Wei V, Garcia J, Kabeer F, Lai D, Soriano P, Banáth J, Chiu DS, Yap D, Le DD, Ye FB, Zhang A, Thu K, Soong J, Lin SC, Tsai AH, Osako T, Algara T, Saunders DN, Wong J, Xian J, Bally MB, Brenton JD, Brown GW, Shah SP, Cescon D, Mak TW, Caldas C, Stirling PC, Hieter P, Balasubramanian S, Aparicio S, CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours, Nat Commun. 8 (2017) 14432.

[94] Scheibye-Knudsen M, Tseng A, Borch Jensen M, Scheibye-Alsing K, Fang EF, Iyama T, Bharti SK, Marosi K, Froetscher L, Kassahun H, Eckley DM, Maul RW, Bastian P, De S, Ghosh S, Nilsen H, Goldberg IG, Mattson MP, Wilson DM 3rd, Brosh RM Jr, Gorospe M, Bohr VA, Cockayne syndrome group A and B proteins converge on transcription-linked resolution of non-B DNA, Proc Natl Acad Sci U S A. 113 (2016) 12502-12507.

[95] Zimmer J, Tacconi EMC, Folio C, Badie S, Porru M, Klare K, Tumiati M, Markkanen E, Halder S, Ryan A, Jackson SP, Ramadan K, Kuznetsov SG, Biroccio A, Sale JE, Tarsounas M, Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds, Mol Cell. 61 (2016) 449-460.

[96] Wang Y, Yang J, Wild AT, Wu WH, Shah R, Danussi C, Riggins GJ, Kannan K, Sulman EP, Chan TA, Huse JT, G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma, Nat Commun. 10 (2019) 943.

[97] Shioda N, Yabuki Y, Yamaguchi K, Onozato M, Li Y, Kurosawa K, Tanabe H, Okamoto N, Era T, Sugiyama H, Wada T, Fukunaga K, Nat Med. 24 (2018) 802-813.

[98] Wang X, Goodrich KJ, Gooding AR, Naeem H, Archer S, Paucek RD, Youmans DT, Cech TR, Davidovich C, Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines, Mol Cell. 65 (2017) 1056-1067.

[99] Qiao Q, Wang L, Meng FL, Hwang JK, Alt FW, Wu H, AID Recognizes Structured DNA for Class Switch Recombination, Mol Cell.67 (2017) 361-373.

[100] Lista MJ, Martins RP, Billant O, Contesse MA, Findakly S, Pochard P, Daskalogianni C, Beauvineau C, Guetta C, Jamin C, Teulade-Fichou MP, Fåhraeus R, Voisset C, Blondel M, Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA, Nat Commun. 8 (2017) 16043.

[101] Artusi S, Nadai M, Perrone R, Biasolo MA, Palù G, Flamand L, Calistri A, Richter SN, The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: Implications for the antiviral activity of a G-quadruplex ligand, Antiviral Res. 118 (2015) 123-131.

[102] Callegaro S, Perrone R, Scalabrin M, Doria F, Palù G, Richter SN, A core extended naphtalene diimide G-quadruplex ligand potently inhibits herpes simplex virus 1 replication, Sci Rep. 7 (2017) 2341.

[103] Satkunanathan S, Thorpe R, Zhao Y, The function of DNA binding protein nucleophosmin in AAV replication, Virology. 510 (2017) 46-54.

[104] De Nicola B, Lech CJ, Heddi B, Regmi S, Frasson I, Perrone R, Richter SN, Phan AT, Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome, Nucleic Acids Res. 44 (2016) 6442-6451.

[105] Yoshizawa T, Nozawa RS, Jia TZ, Saio T, Mori E. Biological phase separation: cell biology meets biophysics, Biophys Rev. (2020) doi: 10.1007/s12551-020-00680-x.

[106] Jain A, Vale RD, RNA phase transitions in repeat expansion disorders, Nature. 546 (2017) 243-247.

[107] Fay MM, Anderson PJ, Ivanov P, ALS/FTD-Associated C9ORF72 Repeat RNA Promotes Phase Transitions In Vitro and in Cells, Cell Rep. 21 (2017) 3573-3584.

[108] Zhang Y, Yang M, Duncan S, Yang X, Abdelhamid MAS, Huang L, Zhang H, Benfey PN, Waller ZAE, Ding Y, G-quadruplex structures trigger RNA phase separation, Nucleic Acids Res.47 (2019) 11746-11754.

[109] Sahakyan AB, Murat P, Mayer C, Balasubramanian S, G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition, Nat Struct Mol Biol. 24 (2017) 243-247.

[110] Sauer M, Juranek SA, Marks J, De Magis A, Kazemier HG, Hilbig D, Benhalevy D, Wang X, Hafner M, Paeschke K, DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions, Nat Commun.10 (2019) 2421.

[111] Huang H, Zhang J, Harvey SE, Hu X, Cheng C, RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF, Genes Dev. 31 (2017) 2296-2309.

[112] Zhang J, Harvey SE, Cheng C, A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes, Nucleic Acids Res 47 (2019) 3667-3679.

[113] Kawauchi K, Sugimoto W, Yasui T, Murata K, Itoh K, Takagi K, Tsuruoka T, Akamatsu K, Tateishi-Karimata H, Sugimoto N, Miyoshi D, An anionic phthalocyanine decreases NRAS expression by breaking down its RNA G-quadruplex, Nat Commun. 9 (2018) 2271.

[114] Ribeiro de Almeida C, Dhir S, Dhir A, Moghaddam AE, Sattentau Q, Meinhart A, Proudfoot NJ, RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination, Mol Cell. 70 (2018) 650-662.

[115] Lyonnais S, Tarrés-Solé A, Rubio-Cosials A, Cuppari A, Brito R, Jaumot J, Gargallo R, Vilaseca M, Silva C, Granzhan A, Teulade-Fichou MP, Eritja R, Solà M, The human mitochondrial transcription factor A is a versatile G-quadruplex binding protein, Sci Rep. 7 (2017) 43992.

[116] Gray LT, Puig Lombardi E, Verga D, Nicolas A, Teulade-Fichou MP, Londoño-Vallejo A, Maizels N, G-quadruplexes Sequester Free Heme in Living Cells, Cell Chem Biol. 26 (2019) 1681-1691.

[117] Fleming AM, Ding Y, Burrows CJ, Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair, Proc Natl Acad Sci U S A. 114 (2017) 2604-2609.

[118] Lyons SM, Gudanis D, Coyne SM, Gdaniec Z, Ivanov P, Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs, Nat Commun. 8 (2017) 1127.

[119] Todd AK, Johnston M, Neidle S: Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res 33 (2005) 2901-2907.

[120] Sabouri N, Capra JA, Zakian VA, The essential Schizosaccharomyces pombe Pfh1 DNA helicase promotes fork movement past G-quadruplex motifs to prevent DNA damage, BMC Biol. 12 (2014) 101.

[121] Masai H, Kakusho N, Fukatsu R, Ma Y, Iida K, Kanoh Y, Nagasawa K. Molecular architecture of G-quadruplex structures generated on duplex Rif1-binding sequences. J Biol Chem. 293 (2018) 17033-17049.

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

[123] Sen D, Gilbert W, Guanine quartet structures, Methods Enzymol. 211 (1992) 191-199.

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

[125] Wu,Y., Shin-ya,K. and 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 (2008) 4116–4128.

[126] Xiao S, Zhang JY, Wu J, Wu RY, Xia Y, Zheng KW, Hao YH, Zhou X, Tan Z, Formation of DNA:RNA hybrid G-quadruplexes of two G-quartet layers in transcription: expansion of the

prevalence and diversity of G-quadruplexes in genomes, Angew Chem Int Ed Engl. 53 (2014) 13110 –13114.

[127] Liu JQ, Xiao S, Hao YH, Tan Z, Strand-Biased Formation of G-Quadruplexes in DNA Duplexes Transcribed with T7 RNA Polymerase, Angew Chem Int Ed Engl. 54 (2015) 8992-8996.

[128] Wu RY, Zheng KW, Zhang JY, Hao YH, Tan Z, Formation of DNA:RNA hybrid G-quadruplex in bacterial cells and its dominance over the intramolecular DNA G-quadruplex in mediating transcription termination, Angew Chem Int Ed Engl. 54 (2015) 2447-2451.

[129] Zheng KW, He YD, Liu HH, Li XM, Hao YH, Tan Z, Superhelicity Constrains a Localized and R-Loop-Dependent Formation of G-Quadruplexes at the Upstream Region of Transcription, ACS Chem. Biol. 12 (2017) 2609-2618.

[130] C.C. Hardin, M. Corregan, B.A. Brown, 2nd, L.N. Frederick, Cytosine-cytosine+ base pairing stabilizes DNA quadruplexes and cytosine methylation greatly enhances the effect, Biochemistry 32 (1993) 5870-5880.

[131] B. Zamiri, M. Mirceta, K. Bomsztyk, R.B. Macgregor, Jr., C.E. Pearson, Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)8*(GGCCCC)8 repeat: effect of CpG methylation, Nucleic acids research 43 (2015) 10055-10064.

[132] J. Lin, J.Q. Hou, H.D. Xiang, Y.Y. Yan, Y.C. Gu, J.H. Tan, D. Li, L.Q. Gu, T.M. Ou, Z.S. Huang, Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation, Biochemical and biophysical research communications 433 (2013) 368-373.

[133] Sekibo,D.A.T. and Fox,K.R. The effects of DNA supercoiling on G-quadruplex formation. Nucleic Acids Res. 45 (2017) 12069–12079.

[134] Mohaghegh P, Karow JK, Brosh RM, Jr., Bohr VA, Hickson ID, The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases, Nucleic Acids Res 29 (2001) 2843-2849.

[135] Johnson JE, Cao K, Ryvkin P, Wang LS, Johnson FB, Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential, Nucleic Acids Res 38 (2010) 1114-1122.

[136] Gray LT, Vallur AC, Eddy J, Maizels N, G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD, Nat Chem Biol 10 (2014) 313-318.

[137] Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH, 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 U S A 99 (2002) 11593-11598.

[138] Fernando H, ReszkaAP, Huppert J, Ladame S, Rankin S, Venkitaraman AR, Neidle S, Balasubramanian S, A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene, Biochemistry 45 (2006) 7854-7860.

[139] Cogoi S, Paramasivam M, Spolaore B, Xodo LE, Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins, Nucleic Acids Res 36 (2008) 3765-3780.

[140] Kang HJ, Cui Y, Yin H, Scheid A, Hendricks WPD, Schmidt J, SekulicA, Kong D, Trent JM, Gokhale V, Mao H, Hurley LH, A Pharmacological Chaperone Molecule Induces Cancer Cell Death by Restoring Tertiary DNA Structures in Mutant hTERT Promoters, J Am Chem Soc 138 (2016) 13673-13692.

[141] Tauchi T, Shin-ya K, Sashida G, Sumi M, Okabe S, Ohyashiki JH, Ohyashiki K, Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia, Oncogene 25 (2006) 5719-5725.

[142] Miyazaki T, Pan Y, Joshi K, Purohit D, Hu B, Demir H, Mazumder S, Okabe S, Yamori T, Viapiano M, Shin-ya K, Seimiya H, Nakano I, Telomestatin impairs glioma stem cell survival and growth through the disruption of telomeric G-quadruplex and inhibition of the proto-oncogene, c-Myb, Clin Cancer Res 18 (2012) 1268-1280.

[143] Kumari S, Bugaut A, Huppert JL, Balasubramanian S, An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation, Nat Chem Biol 3 (2007) 218-221.

[144] Arora A, Dutkiewicz M, Scaria V, Hariharan M, Maiti S, Kurreck J, Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif, RNA. 14 (2008) 1290-1296.

[145] Wolfe AL, Singh K, Zhong Y, Drewe P, Rajasekhar VK, Sanghvi VR, Mavrakis KJ, Jiang M, Roderick JE, Van der Meulen J, Schatz JH, Rodrigo CM, Zhao C, Rondou P, de Stanchina E, Teruya-Feldstein J, Kelliher MA, Speleman F, Porco JA Jr, Pelletier J, Rätsch G, Wendel HG, RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer, Nature 513 (2014) 65-70.

[146] Weldon C, Dacanay JG, Gokhale V, Boddupally PVL, Behm-Ansmant I, Burley GA, Branlant C, Hurley LH, Dominguez C, Eperon IC, Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X, Nucleic Acids Res 46 (2018) 886-896.

[147] BeaudoinJD, Perreault JP, Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening, Nucleic Acids Res 41 (2013) 5898-5911.

[148] Subramanian M, Rage F, Tabet R, Flatter E, Mandel JL, Moine H, G-quadruplex RNA structure as a signal for neurite mRNA targeting, EMBO Rep 12 (2011) 697-704.

[149] Guo JU, Bartel DP, RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria, Science 353 (2016) pii: aaf5371

[150] Cahoon LA, Seifert HS, An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae, Science 325 (2009) 764-767.

[151] Cahoon LA, Seifert HS, Transcription of a cis-acting, noncoding, small RNA is required for pilin antigenic variation in Neisseria gonorrhoeae, PLoS Pathog, 9 (2013) e1003074.

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

Legends to figures

Figure 1. Diverse biological functions of G-quadruplex (G4).
DNA G4, RNA G4 and G4 formed on RNA-DNA hybrid serve as important nucleic acid signals for regulation of varieties of chromosome functions that involve DNA, RNA and complexes of RNA and DNA. See text and also **Supplementary Table S1** for details.

Figure 2 G4 regulates DNA replication both negatively and positively.
Left: Rif1 negatively regulates DNA replication by binding to G4, which may generate chromatin architecture inhibitory for firing of origins. Right: In *E.coli*, RNA-DNA hybrids containing G4 may serve as a signal for initiation of DNA replication through the actions of proteins recognizing G4. Proteins such as PriA, RecA or an unknown G4 binding factor may play roles in initiation of DNA replication at the G4. In higher eukaryotes, G4-forming sequences are frequently associated with replication origins.

Figure 3 Regulation of telomere by G4 and its binding proteins.
Formation of G4 in telomere inhibits replication of telomere, causing genomic instability. Various proteins can counteract the formation of G4 in telomere. RTEL1 helicase disrupts secondary structures at telomere and also disassembles T loops (not shown in the figure). These two functions contribute to telomere stability. Persistent formation of G4 could result in ALT (alternative lengthening of telomeres). This is especially apparent in cells lacking ATRX. The ATRX induces loading of Histone H3.3 onto chromatin, facilitating the chromatinization of telomere which inhibits ALT.

Figure 4. Interference of transcription and replication by G4.
Formation of G4 on a template strand can impede the movement of DNA polymerases (upper, red line) or RNA polymerases (lower, blue line). Pif1 helicase disrupt the G4 structure on the DNA template strand, and relieves the inhibition of DNA chain elongation. Sub1 or Mms1 facilitates the binding of Pif1 to G4. On the other hand, CSB is able to antagonize the block of transcription by G4.

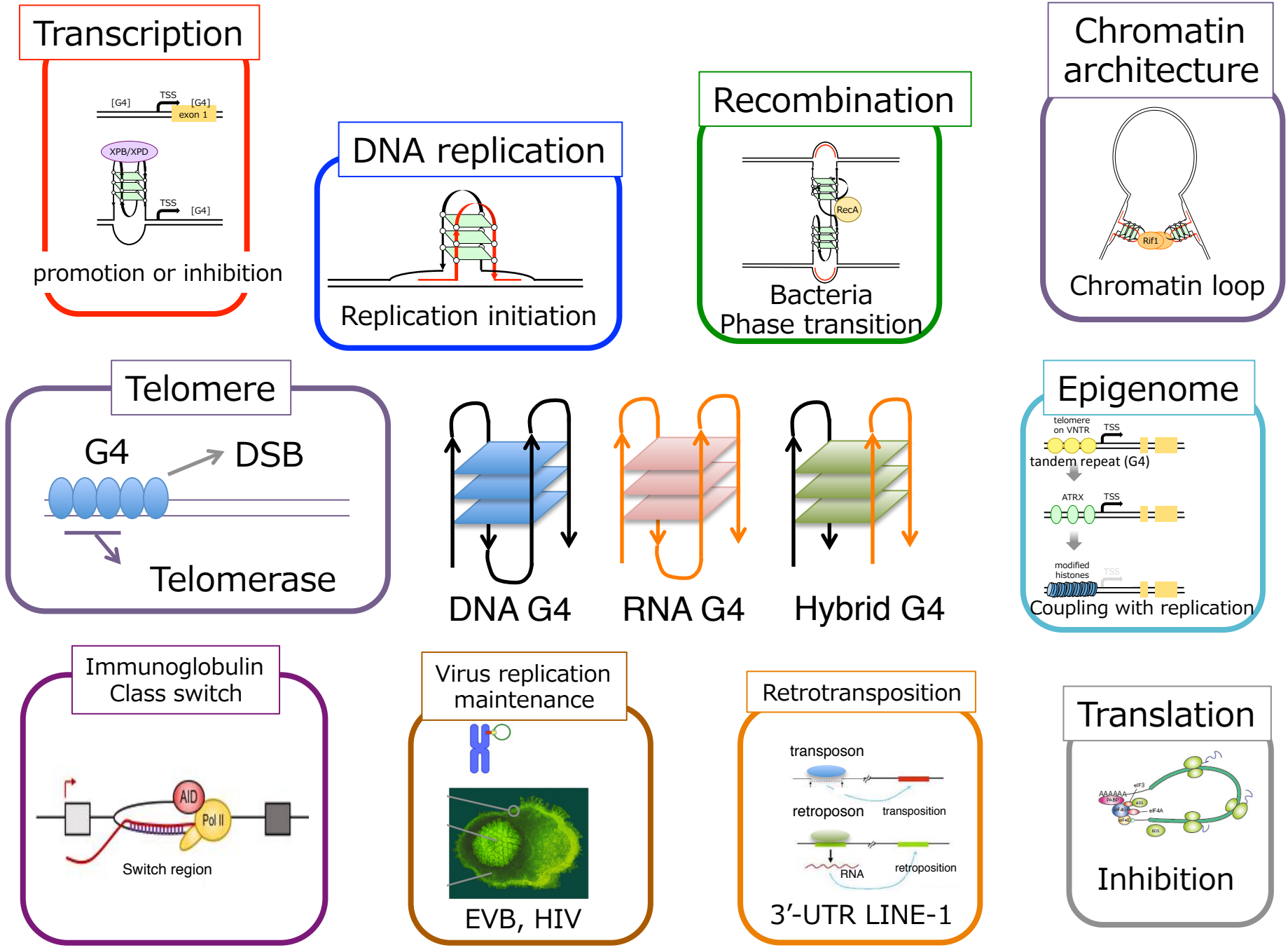
Figure 5. Roles of PARP3 and Bloom in regulation of G4 on chromosomes.
PARP3 destabilizes G4 and promotes repair of DSB through HR and NHEJ. Bloom helicase also disrupts G4 on transcribed coding segments, preventing SCE. which occurs extensively in cells from Bloom Syndrome patients.

Figure 6. Interaction of PRC2 with G4 in epigenome regulation.
Polycomb repressive complex 2 (PRC2) induces 3K27me3, but is removed by G4 on transcribed mRNA, resulting in more active histone signature.

Figure 7. Factors affecting the stability of G4 in cells.
Various factors may affect the formation of cellular G4. Thus, the formation and disruption of G4 in cells may occur in a dynamic manner.

Figure

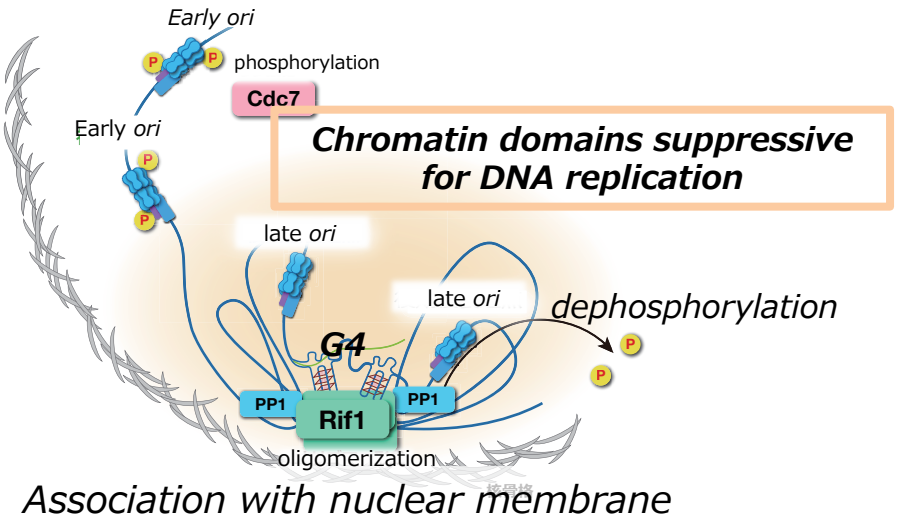
Masai & Tanaka Figure 1



Masai & Tanaka Figure 2

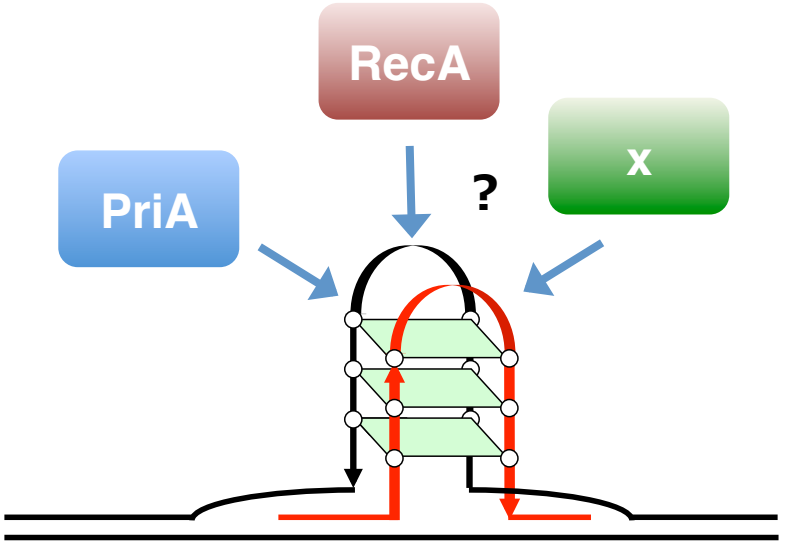
Negative

Rif1 negatively regulates DNA replication by binding to G4

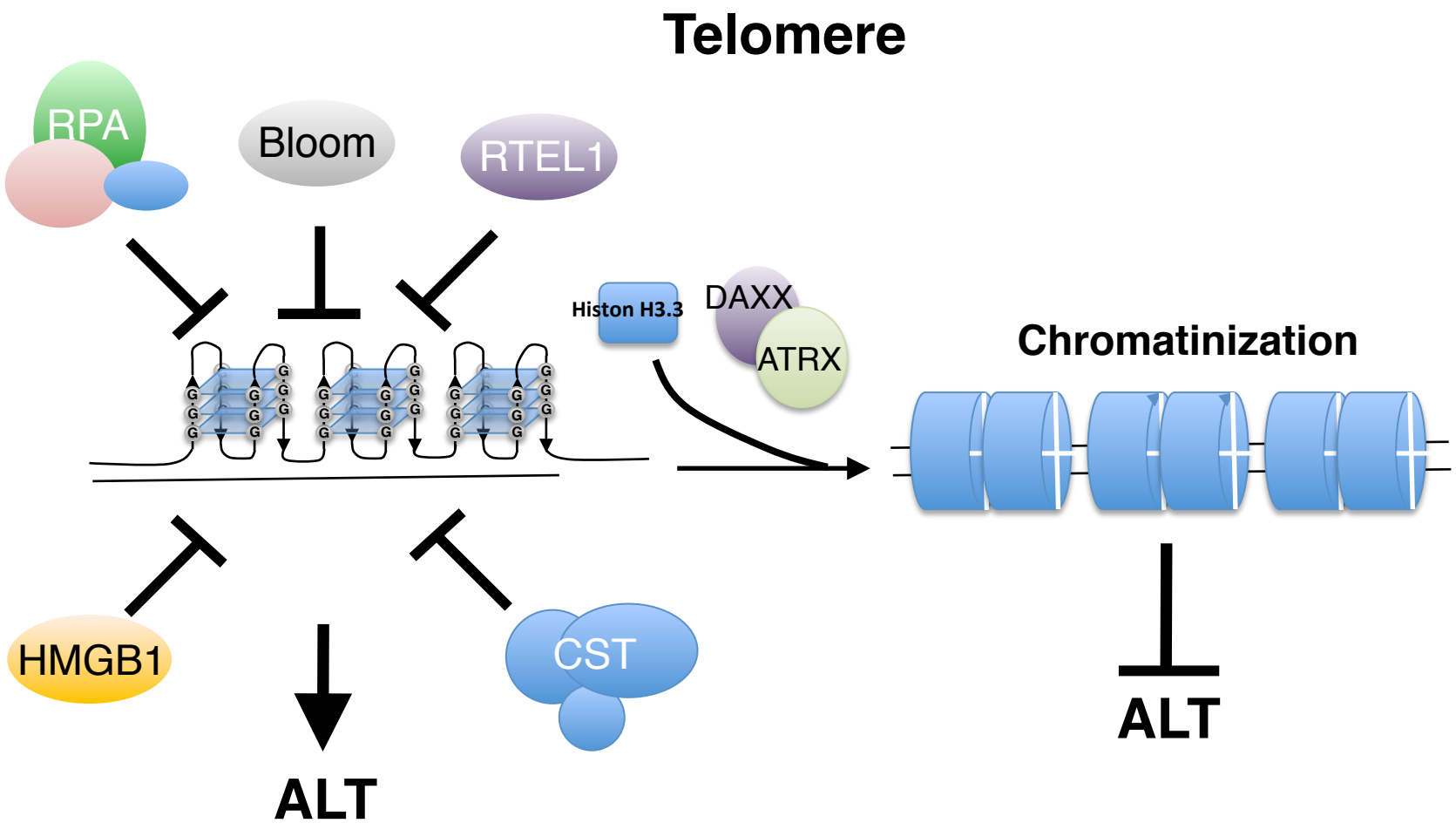


Positive

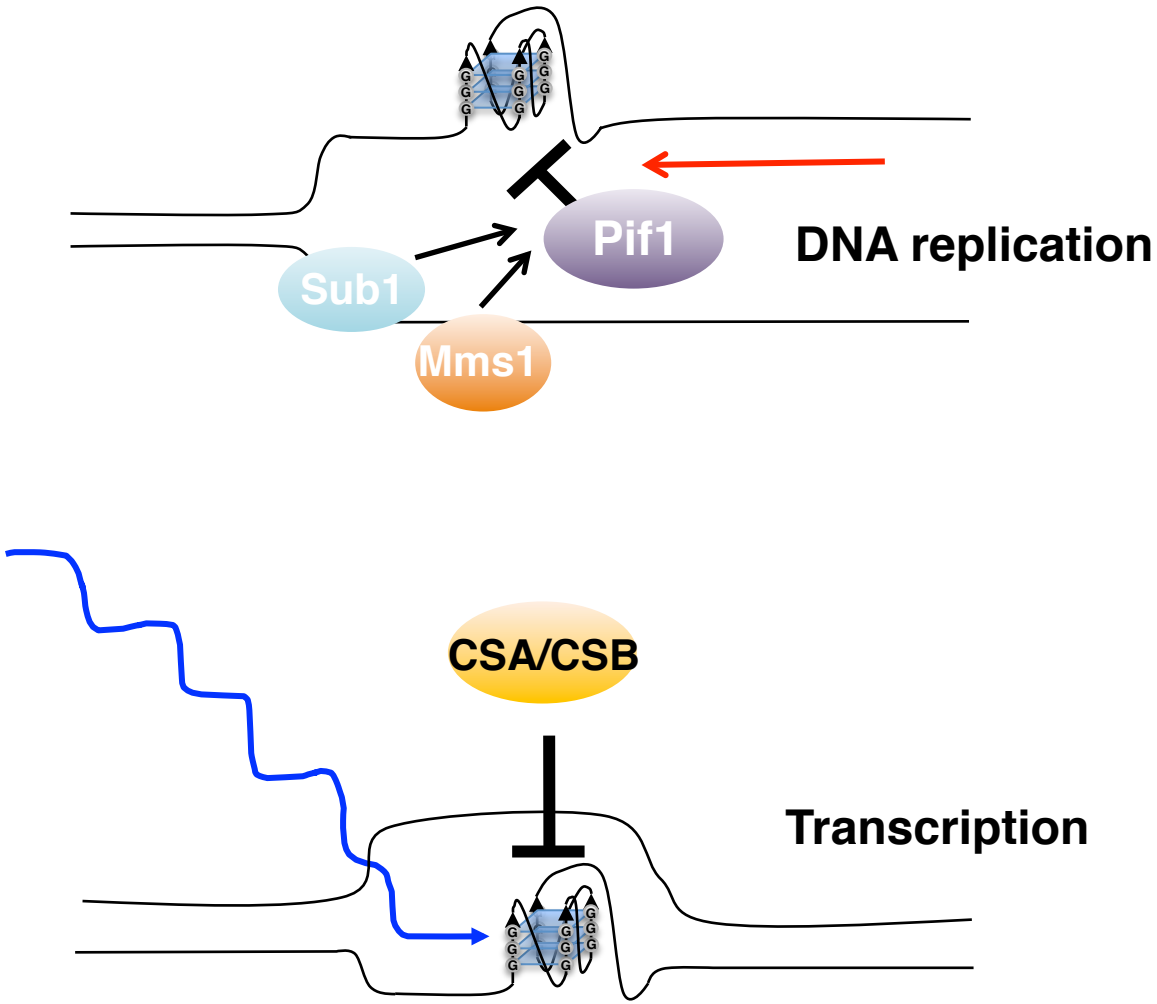
G4/RNA-DNA hybrid could be a signal for initiation of DNA replication



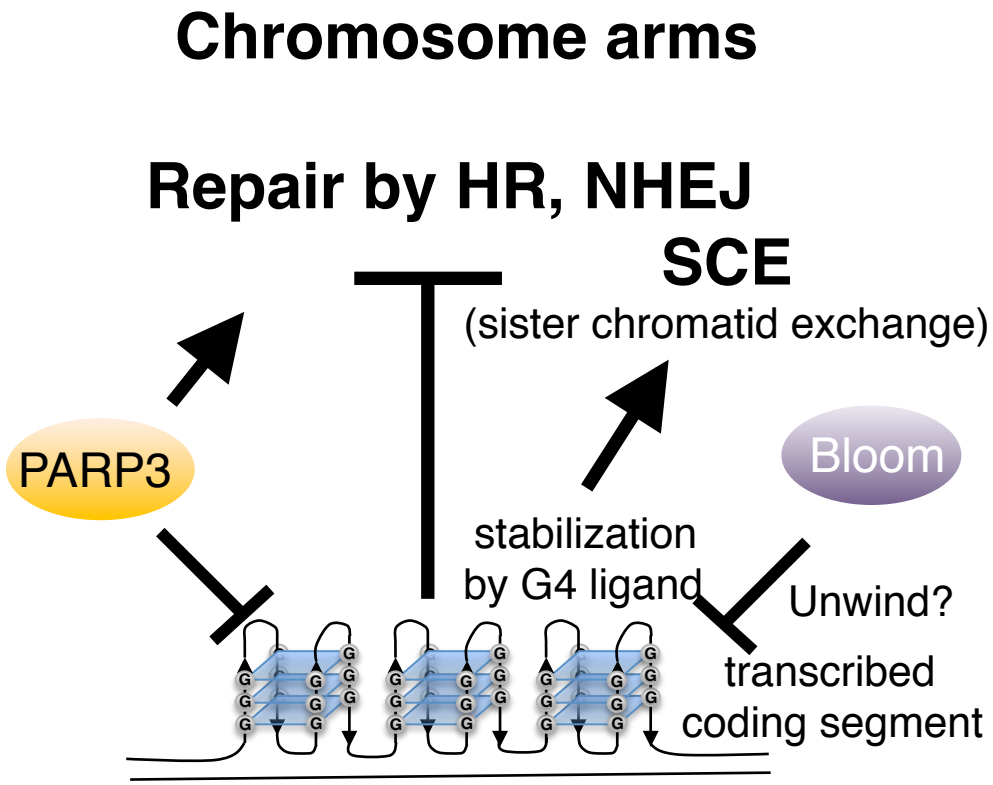
Masai & Tanaka Figure 3



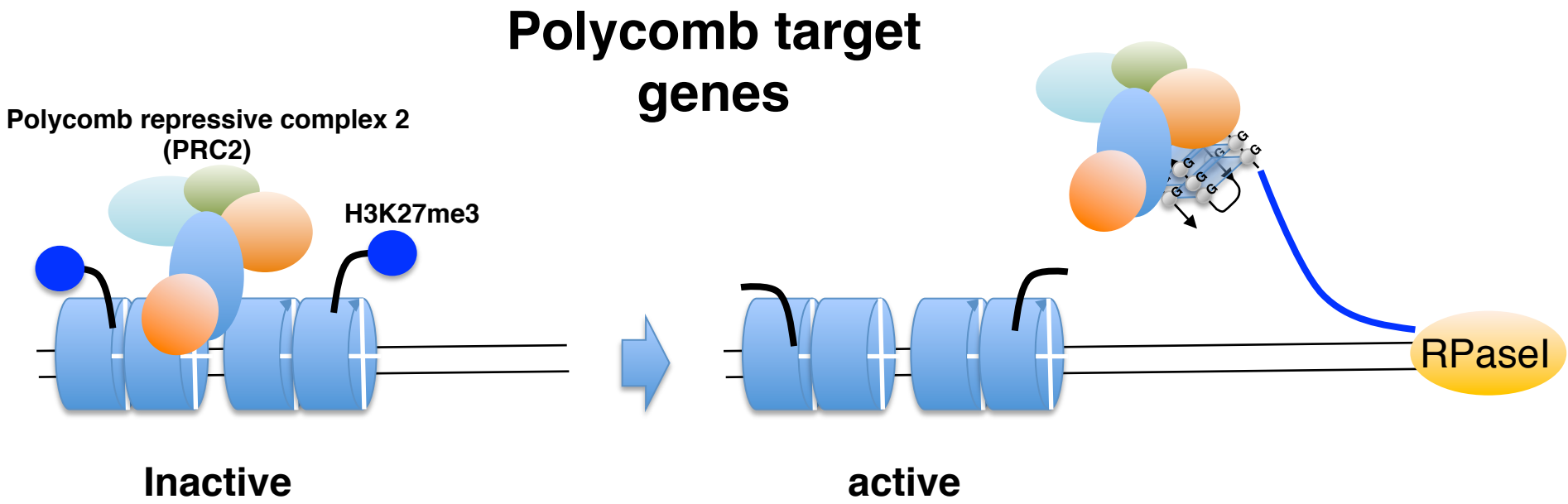
Masai & Tanaka Figure 4



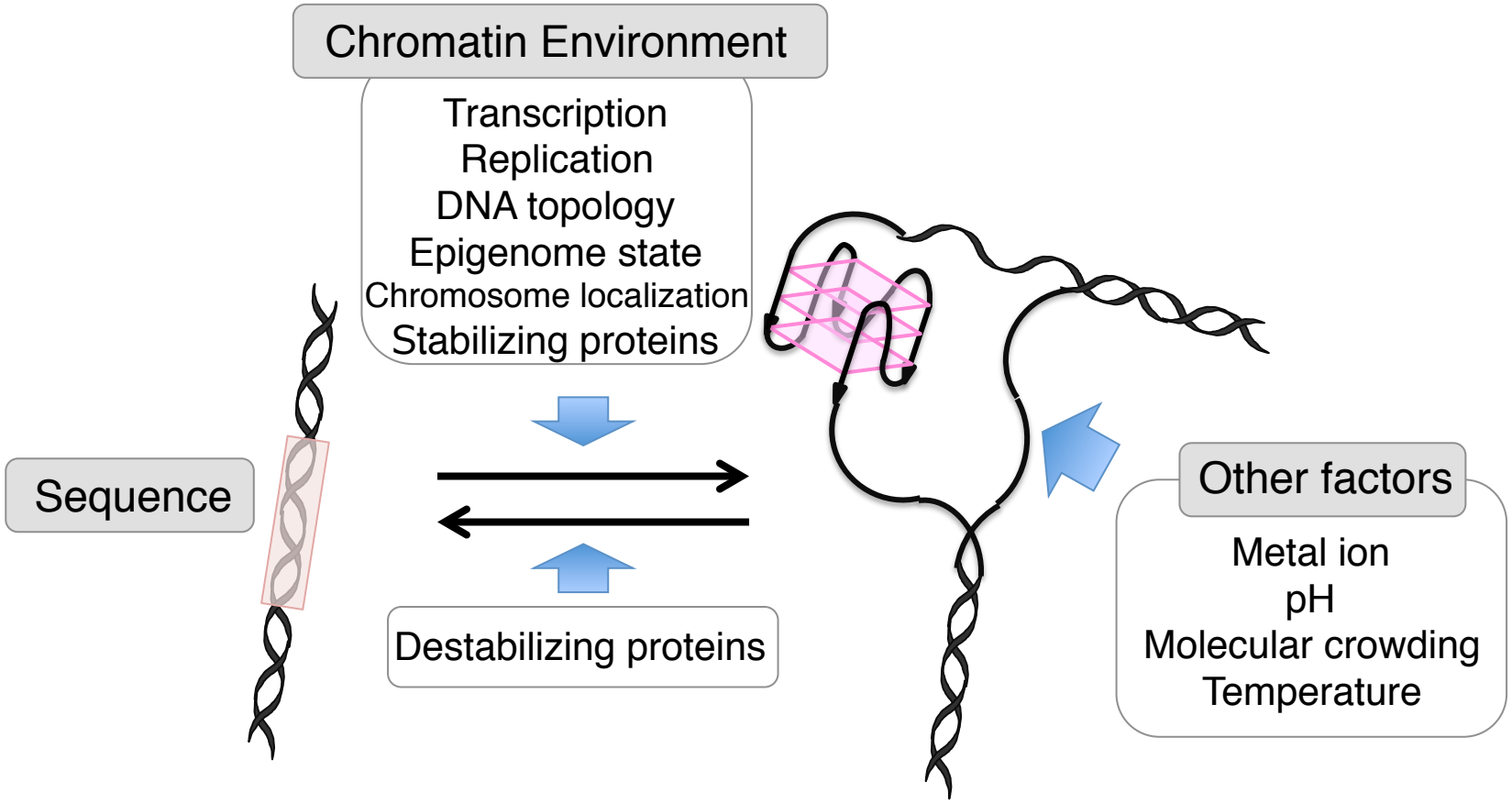
Masai & Tanaka Figure 5



Masai & Tanaka Figure 6



Masai & Tanaka Figure 7



Supplementary Table 1 Summary of various functions of DNA G4 and RNA G4 as well their binding proteins, and the effects of G4-specific ligands

DNA G4 or RNA G4	Species (organella)	Category of biological reactions	Proteins involved (or others)	G4 ligands involved	Mechanisms	Disease involved	Literature
G4 DNA	Human	Telomere regulation	ATRX (dependent on histone chaperone DAXX)	Pyridostatin (PDS)	ATRX suppresses the ALT pathway and shortens telomeres. ATRX loss causes telomere chromatinization with persistent G4.		[60]
G4 DNA	Fission yeast	Telomere regulation	RPA, Pif1		RPA prevents the formation of G4 at lagging-strand telomeres. Pif1 can unwind G4 and rescues all the telomeric defects of the RPA D223Y mutant.		[61]
G4 DNA	Human	Telomere regulation	RTEL1, SLX4		RTEL1 removes telomeric DNA secondary structures, and ensures the dynamics and stability of the telomere.		[63]
G4 DNA	Human	Telomere regulation	8-oxoG		8-oxoG present in the dNTP pool: terminates telomere elongation a preexisting 8-oxoG within telomere DNA: destabilizes G4 and stimulate telomerase		[64]
DNAG4 (a parallel, intermolecular conformation)	Human	Telomere regulation	core telomerase enzyme complex		Sufficient for partial G4 resolution and extension.		[65]
G4 DNA	Human	Telomere regulation	HMGB1 (a ubiquitous non-histone protein)		Depletion induces telomere damages.		[60]
G4 DNA	Human	Telomere regulation	CST (CTC1-STN1-TEN1)		CST maintains genome integrity through resolution of G4 structures.		[62]
DNA G4 (short loop length with pyrimidine bases)	Budding yeast	Genome instability					[67]
DNA G4	Fission yeast	Genome instability	Pif1		Pif1 binding sites on chromatin overlap with G4 sequences. Replication slows and breaks in *pif1*Δ cells.		[68]
DNA G4 (co-transcriptionally formed G4 DNA in Top1-deficient cells)	Budding yeast	Genome instability	Sub1(human homolog PC4), Pif1		Contributes to genome maintenance at G4-forming sequences. Interacts with the G4-resolving helicase Pif1.		[69]
DNA G4	Human	Genome instability	PARP3 (poly(ADP)ribose polymerase 3)	Pyridostatin Chemical stabilization of G4 DNA in PARP3$^{-/-}$ cells leads to widespread DNA double-strand breaks and lethality.	PARP3 suppresses G4 DNA and facilitates DNA repair by multiple pathways.		[72]
DNA G4 (on RNA-DNA hybrid)	Human	Genome instability			G4 ligands spread R loops to adjacent regions containing G4, preferentially at 3'-end regions of expressed genes. G4 ligands can induce DNA damage by an R loop-dependent mechanism (in BRCA2+ and BRCA- cells).		[73]
DNA G4	Fission yeast	Genome instability	Mms1 (an E3 ubiquitin ligase complex protein)		Binds G4 and promotes Pif1 helicase binding.		[70]
DNA G4 (in the transcribed regions)	Human	Genome instability	BLM helicase		Sister chromatid exchange events (SCEs) enriched at G4 motifs in transcribed genes in BLM mutant. BLM suppressed recombination at sites of G4.		[71]
DNA G4	Human	Helicase	Bloom and Werner helicase		The Bloom's and Werner's syndrome proteins are capable of unwinding DNA with specific higher-order structures incluring G-quadruplex DNA.		[134, 152]
DNA G4	Human	Unwinding mechanism	DHX36		Unwinds G4 DNA in an ATP-dependent manner, but unfolds G4 RNA in an ATP-independent manner.		[74]
DNA G4	Bacteria (E.coli)	Unwinding mechanism	RecQ DNA helicase		A guanine base from the unwound G4 is		[75]
DNA G4	Human	Unwinding mechanism	DHX36, Bloom helicase (BLM), and Werner helicase (WRN)	BRACO-19, NMM, and Phen-DC3 (stably bound G4 ligand dislodged by the resolvases)	Distinct G4 conformation specificity, but use a common mechanism of repetitive unfolding that leads to disrupting G4 structure multiple times in succession.		[76]
DNA G4	Human	Unwinding mechanism	DHX36		The N-terminal DHX36-specific motif folds into a		[77]
DNA G4	Fisson yeast	Unwinding mechanism	Pif1		Unwinds double-stranded DNA by a branched mechanism with 2 modes of activity.		[80]
DNA G4	Human	Unwinding mechanism	BLM helicase	An oxazole telomestatin derivative, pyridostatin, and PhenDC3	Inhibits BLM helicase activity at similar levels (2-3 fold).		[81]
RNA G4	Human	Mode of recognition	FMRP (Fragile X Mental Retardation Protein)		Specifically interacts with G4 on mRNAs through an arginine-glycine-rich (RGG) motif. Recognizes a G4 RNA by a β-turn in the RGG motif of FMRP.	FXS(Fragile X Syndrome)	[77]
DNA G4	Human	Mode of recognition	DHX36		A 18-amino acid segement binds to a parallel DNA G4 through the terminal guanine base tetrad (G-tetrad), and clamps the G4 using three-anchor-point electrostatic interactions between three positively charged amino acids and negatively charged phosphate groups.		[78]
DNA G4	Human	Transcription	DDX5	Inhibits DDX5-mediated MYC activation.	DDX5 unfolds MycG4-DNA. DDX5 efficiently resolves MycG4-DNA. DDX5 is enriched at the MYC promoter and activates MYC transcription by resolving G4 DNA and RNA.		[83]
DNA G4	Human	Transcription		DC-34	Downergulates Myc expression by a G4-dependent mechanism. Inhibition by DC-34 is significantly greater for MYC than for other G4-driven genes.		[84]

DNA G4	Human	Transcription			Gene expression is considerably altered in WS and BS cells as compared with normal and RTS cells. Moreover, upregulated genes in WS and BS cells are significantly enriched for PQS, but not in RTS cells.	WS (Werner syndromes), BS (Bloom syndromes) and RTS (Rothmund–Thompson syndrome)	[135]
DNA G4	Human	Transcription	XPB and XPD		XPB and XPD overlap with G4. XPD is a robust G4 DNA helicase and XPB binds G4. XPB and XPD are enriched near the transcription start site of highly transcribed genes.	XP	[136]
DNA G4	Human	Transcription	C-myc	TMPyP4	A specific G-quadruplex structure in the c-MYC promoter functions as a transcriptional repressor element. TMPyP4 can stabilize the G4 and suppress the transcription.		[137]
DNA G4	Human	Transcription	C-kit		The c-kit promoter contains a G4 formng sequence, which appears to be conserved in mammalian species.		[138]
DNA G4	Human	Transcription	heterogeneous nuclear ribonucleoprotein A1 K-ras		The human KRAS promoter contains two G4 forming sequences at nuclease hypersensitive element (NHE), which is bound by ribonucleoprotein A1 in vitro.		[139]
DNA G4	Human	Epigenome regulation	ATRX		Binds to tandem repeat (TR) sequences in both telomeres and euchromatin. ATRX binds G4 structures		[85]
DNA G4	Human	epigenome regulation	PRC2 (Polycomb repressive complex 2)		Binds G tracts within pre-mRNA. PRC2 transfers from chromatin to pre-mRNA upon gene activation, leading to H3K27me3 depletion from genes.A novel role of G4 RNA in removal of an epigenome regulator from the chromatin.		[86]
DNA G4	Human	epigenome regulation	DNMT1 (DNA methyltransferase 1)		G4 is tightly associated with CpG islands hypomethylation. DNMT1 bind G4s which inhibit DNMT1 enzymatic activity. G4 sequesters DNMT1 and inhibits local methylation.		[87]
DNA G4	Human	Heterochromatin replication	TRF2		Binds to pericentromeric Satellite III sequences during S phase, and recruit G4-resolving helicase RTEL1 to facilitate fork progression		[88]
DNA G4	Human	Epigenome replication	TRF2		TRF2 binding altered expression and the epigenetic state of several target promoters, indicated by histone modifications. TRF2 regulates expression of genes outside telomeres in a G4-dependent fashion.		[89]
RNA G4	Human	Heterochromatin regulation	HP1α		The unstructured hinge domain of HP1α recognizes parallel G4 formed by TERRA. TERRA can enrich HP1α at telomeres to maintain heterochromatin.		[90]
RNA G4	Human	Epigenome regulation	PRC2(Polycomb repressive complex 2) Histone methyltransferase that trimethylates H3K27		PRC2 has a high affinity for G4. G-tract motifs are enriched at PRC2-binding RNA and DNA (chromatin) and H3K27me3-modified nucleosomes.		[98]
DNA G4	Human	Diseases-related (G4 therapy)		CX-5461	Blocks replication forks and induces ssDNA gaps or breaks which will be repaired by the BRCA and NHEJ pathways.	G4 targeting in HR and NHEJ deficient cancer cells can be an efficient therapeutic choice.	[93]
DNA G4	Human/ C. elegans	Diseases-related	CSA/CSB	pyridostatin and TMPYP4 Stabilization of G4 activates the DNA damage sensor PARP1 (poly-ADP ribpse polymerase 1) and leads to accelerated aging in Caenorhabditis elegans.	Loss of CSA or CSB leads to polymerase stalling at G4 in a neuroblastoma cell line, and CSB can melt G4.	Cockayne syndrome	[94]
DNA G4	Human	Diseases-related		Pyridostatin (PDS) increases telomere fragility in BRCA2-deficient cells (G4 formation drives telomere instability).	PDS induces DSB accumulation, checkpoint activation, and deregulated G2/M progression and enhances the replication defect intrinsic to HR deficiency	therapeutic potential of G4-stabilizing ligand to selectively eliminate HR-compromised cells and tumors, including those resistant to PARP inhibition	[95]
DNA G4	Human	Diseases-related	ATRX	CX-3543 (G4 stabilizer)	Loss of ATRX induces replication stress and DNA damage by way of G4 DNA. ATRX deficiency in conjunction with chemical G4 stabilization enhances DNA damage and cell death.	malignant glyoma α-thalassemia mental retardation X-linked	[96]
DNA G4	Human	Diseases-related	ATRX	5-ALA (Converted into G-quadruplex-binding metabolites, reduces RNA polymerase II recruitment, represses Xlr3b transcription and restores synaptic plasticity and cognitive proficiency in ATRX model mice)	ATRX binds to G4 in CpG islands of the imprinted Xlr3b gene, inhibiting its expression by recruiting DNA methyltransferases. Atrx mutation upregulates Xlr3b expression in the mouse brain.	ATRX syndrome	[97]

DNA G4	Human	Telomere/ diseases		GTC365 (a small drug-like pharmacological chaperone (pharmacoperone) molecule)	Promoter mutations in hTERT disrupt G4 and impairs the silencing effect. GTC365 induces the correct folding of the mutant G4, reducing hTERT expression, causing cancer cell death and senescence.		[140]
DNA G4	mouse	Telomerase/ diseases		Telomestatin	Telomesatin induced apoptose in leukemia cells in a manner dependent on p38 MAP kinase.	Acute Leukemia	[141]
DNA G4	Human	Telomerase/ diseases		Telomestatin	Telomestatin inhibits growth of GSC (Glioma stem cells) through telomere disruption and c-Myb inhibition.	GBM (glioblastoma multiforme)	[142]
DNA G4/ RNA G4	Human	Neurodegenerative diseases	nucleolin		GGGGCC repeats in C9orf72 forms DNA and RNA G4 and promotes RNA•DNA hybrids (R-loops). Nucleolin, preferentially binds to these strucures, and induce nucleolar stress.	amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)	[91]
RNA G4	Human	Epstein-Barr virus (EBV)	Nucleolin	PhenDC3 prevents NCL binding to EBNA1 mRNA and reverses GAr-mediated repression of EBNA1 expression.	EBV evades the immune system by suppressing the essential EBNA1 protein to the minimal level. The glycine-alanine repeats (GAr) in EBNA1 mRNA form a G4. Nucleolin (NCL) binds to the G4 RNA, and inhibits the EBNA1 expression, whereas its depletion relieves the suppression.		[100]
DNA G4	Human	Herpes Simplex Virus-1 (HSV-1)		A core extended aphtalene diimide (c-exNDI)	Significant virus inhibition with no cytotoxicity: G4-mediated inhibition of viral DNA replication		[102]
DNA G4	Human	Adeno-associated viruses (AAV)	Nucleophosmin (NPM)	Interact with G4-forming sequences in the AAV genome, facilitating AAV production.			[103]
G4 DNA	Human	human immunodeficiency virus (HIV)-1			The G-rich U3 region within the LTR promoter forms a parallel-stranded G4, which stacks with neighboring G4 through its single-nucleotide bulges.		[104]
RNA G4	Human	Translation			Mature cytoplasmic tRNAs are cleaved during stress response to produce tRNA fragments that function to repress translation in vivo. G4 RNA can be generated from cleaved tRNA fragments, which augments the formation of stress granules by suppressing translation.		[118]
RNA G4	Human	Translation			A highly conserved, thermodynamically stable G4 RNA in the 5'-UTR of the human NRAS mRNA suppresses its translation.	N-ras	[143]
RNA G4	Human	Translation	Zic-1 zinc-finger protein.		G4 in the 5'-UTR of mRNA inhibits translation in vivo in eukaryotic cells.		[144]
RNA G4	Human	Translation	eIF4A RNA helicase		5' UTRs of select cancer genes (oncogenes, superenhancer-associated transcription factors, and epigenetic regulators) conain G4 RNA that needs to be unfolded by eIF4A RNA helicase.		[145]
RNA G4	Human	alternative splicing	hnRNPF (heterogeneous nuclear ribonucleoprotein F)		hnRNPF regulates an EMT-associated CD44 isoform switch in a G4-dependent manner, through its recognitiono of G4, enriched in hnRNPF-binding sites and near hnRNPF-regulated alternatively		[111]
RNA G4	Human	Splicing	ribonucleoprotein F (hnRNPF)	emetine and cephaeline	G4 on mRNA may be required for exon inclusion through its interaction with hnRNPF. Thus, G4 RNA may play an important role in alternative splicing.		[112]
RNA G4	Human	mRNA regulation	HP1α	An anionic phthalocyanine, ZnAPC,	Binds to a G4-forming oligonucleotide derived from the 5'-untranslated region of NRAS mRNA, resulting in selective cleavages of the target RNAs G4 upon photo-irradiation.		[113]
RNA G4	Human	Splicing			G4 RNA veryclose to splicing site inhibits the binding of splicing factors.		[146]
RNA G4	Human	mRNA metabolism			G4 in 3'-UTR affects alternative 3'-end polyadenylation and mRNA shortening		[147]
RNA G4	Human	mRNA translocation			G4 in 3'-UTR affects mRNA translocation in cortical neurites. G4 may be a common neurite localization signal.		[148]
DNA G4	Human, Drosophila	DNA replication			67-90% of mammalian replication origins have flanking G4 motifs		[9,10,11,50,51]
DNA G4	Chicken	DNA replication			G4 motifs were shown to be necessary for origin function in two model origins in chicken DT40 cells, and G4 stability correlates with origin efficiency. G4 orientation determines the precise position of the replication start site.		[12]
DNA G4	Human, Xenopus	DNA replication			G4 is required for initiation of DNA replicaition in Xenopus egg extracts.		[12,52]

DNA G4	Fission yeast	DNA replication timing/ chromatin structure	Rif1		Rif1 binds to chromain through recognition of selective G4 sequences. Rif1 forms chromatin loops through binding to multiple G4 simultaneously. This may generate chromatin architecture that may be inhibitory for initiation of DNA replication		[42,43]
DNA G4/ RNA G4	Human	DNA replication	ORC		Human ORC binds preferentially to G4 DNA and G4 RNA.		[53]
DNA G4	Neisseria gonorrhoeae	Recombination	RecA?	NMM	G4-forming sequences play a role for formation of nicks required for recombination for pilin antigenic		[150,151]
DNA G4	Human	Long interspersed nuclear elements (LINEs); transposable elements		PDS, PhenDC3, 12459 and PDC12 Stabilization with the G4 ligands stimulates retrotransposition.	G-rich sequences in the 3' untranslated regions (UTRs) of LINE-1 elements are coupled with retrotransposon speciation and contribute to retrotransposition through the formation of G4 structures		[109]
DNA G4	Human	Stress granule formation	DHX36 helicase		DHX36 binds to G4-forming sequences on mRNAs and reduces their abundance. Knockout of DHX36 results in increase of stress granule formation. DHX36 may facilitate the resolution of G4 RNA induced cellular stress.		[110]
DNA G4/ RNA G4	Human	CSR (Class Switch Recpmbination)	DDX1	Pyridostatin reduces R-loop levels over S-regions.	During CSR, DDX1 binds to G4 structures present in intronic switch transcripts and converts them into S-region R-loops. This recruits AID to S-regions, promoting CSR. S-region transcripts interconvert between G4 and R-loop structures to promote CSR in the IgH locus by the action of DDX1.		[111]
DNA G4	Human	Class switch recombination (CSR) and somatic hypermutation (SHM)	AID(Activation-induced cytidine deaminase)		G4-containing substrates are good AID substrates in vitro. G4 substrates induce cooperative AID oligomerization. Mutations disrupting bifurcated substrate recognition or oligomerization compromise CSR in splenic B cells. G4 recognition of AID and its oligomerization plays a crucial role in CSR.		[99]
DNA G4/ RNA G4	Human	Transcription/ Replication	TFAM (The mitochondrial transcription factor A)		The mitochondrial transcription factor A (TFAM), a high-mobility group (HMG)-box protein known to play a critical role in its expression and maintenance, binds to G4 DNA or G4 RNA, which are prevalent in the mitochondrial genome.		[115]
DNA G4/ RNA G4	Human	Regulation of Heme		PhenDC3 (displaces G4-bound heme, increasing the expression of heme-regulated genes)	Heme is an essential cofactor for many enzymes, but free heme is toxic and its levels are tightly regulated. G4 binds heme avidly in vitro. PhenDC3, a G-quadruplex ligand, displaces G4-bound heme in vitro and upregulating genes that support heme degradation and iron homeostasis, including heme oxidase 1, the key enzyme in heme degradation.		[116]
DNA G4/ RNA G4	Human	Epigenetic regulation of transcription Base excision repair	8-oxo-7,8-dihydroguanine (OG) APE1 (apurinic/apyrimidinic endonuclease 1)		OG(8-oxo-7,8-dihydroguanine) formed in potential G4-forming sequences triggers base excision repair (BER) by 8-oxoguanine DNA glycosylase (OGG1), yielding an abasic site (AP). AP melts duplex DNA to facilitate G4 formation. Apurinic/apyrimidinic endonuclease 1 (APE1) binds, but inefficiently cleaves the AP for activation of vascular endothelial growth factor (VEGF) or endonuclease III-like protein 1 (NTHL1) genes.		[117]
DNA G4	Human	Phase transition			ALS/FTD GGGGCC repeat RNA (rG4C2) drives phase separations. The phase transition occurs at a similar critical repeat number as observed in the diseases G4 RNA with more G-quartets and longer loops are more likely to form phase separation.		[105-107]
RNA G4	Mouse and E.coli				Evidence is presented that show potential G4-RNA in mammalian cells is mostly unfolded. On the other hand, G4-RNA can be formed in bacteria.		[149]