

Review

Not peer-reviewed version

A Compendium of G-flipon Biological Functions that have Experimental Validation

[Alan Herbert](#) *

Posted Date: 23 August 2024

doi: 10.20944/preprints202408.1688.v1

Keywords: Flipons; G-Quadruplex; Triplex; Z-DNA; Proliferation; Transcription; Translation; Repair; Repression; Sister chromatids; CTCG; Chromatin Loops; Class Switch Recombination



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

A Compendium of G-flipon Biological Functions that have Experimental Validation

Alan Herbert

Discovery, InsideOutBio, 42 8th Street, Unit 3412, Charlestown and MA 02129 alan.herbert@insideoutbio.com

Abstract: As with all new fields of discovery, work on the biological role of G-quadruplexes (GQ) has produced a number of results that at first glance are quite baffling, sometimes because they do not fit well together, but mostly because they are different from commonly held expectations. Like other classes of flipons, those that form G-quadruplexes have a repeat sequence motif that enables the fold. The canonical DNA motif $(G_3N_{1-7})_3G_3$, where N is any nucleotide and G is guanine, a feature that is under active selection in avian and mammalian genomes. The involvement of G-flipons in genome maintenance traces back to the invertebrate *C. elegans* and to ancient DNA repair pathways. A role for GQ in transcription is supported by the observation that yeast Rap1 protein binds both B-DNA, in a sequence-specific manner, and GQ, in a structure-specific manner, through the same helix. Other sequence-specific transcription factors (TF) also engage both conformations. RNAs can also modulate GQ formation in a sequence-specific manner and engage the same cellular machinery as localized by TF, linking the ancient RNA world with the modern protein world. The coevolution of both sequence-specific RNA and proteins is supported by studies of early embryonic development, where the transient formation of G-quadruplexes potentially coordinates the epigenetic specification of cell fate.

Keywords: Flipons; G-Quadruplex; Triplex; Z-DNA; Proliferation; Transcription; Translation; Repair; Repression; Sister chromatids; CTCF; Chromatin Loops; Class Switch Recombination

1. Introduction

GQ forming sequences are defined by the canonical DNA motif $(G_3N_{1-7})_3G_3$, where G is guanine and N is any nucleotide. Four G-bases hydrogen bond to each other to form a tetrad that then folds into a four stranded structure. In addition to the Watson-Crick bonding scheme, the rather unconventional Hoogsteen hydrogen bonds stabilize the interaction (Figure 1). The G-tetrad was first observed in X-ray diffraction studies of 5'-GMP and 3'-GMP gels, each with a different arrangement for the staking of the tetrads, one on top of another [1]. The preferred helical arrangement of GQ crystalline fibers was later revealed by structural studies of polyinosinic and polyguanylic RNAs [2].

It was once widely believed that GQ did not exist in cells. If present, then the GQ formed predisposed to genetic instability and to disease [3]. There was much excitement when the Tetrahymena telomere sequence repeats [4] were shown to form GQ [5]. In contrast, later work revealed that telomeres in vivo were more likely to form a different type of structure called a T-loop [6], seemingly ruling out a role for GQ in telomere maintenance (but see below). The prevailing view that GQ were bad was reinforced by the many loss-of-function (LOF) helicase variants that were associated with human mendelian diseases. Their failure to resolve GQ was considered causal, even though they also are required to resolve other non-B structures, such as cruciforms and the Holliday junctions formed during recombination [7]. Further, repeat expansion diseases involved sequences predicted to form extended GQ that were shown to interfere with a variety of cellular functions, including DNA replication, transcription and RNA processing [8].

However, there was evidence that GQ played essential biological roles. The GQ were associated with class switch recombination of immunoglobulin heavy chain (IgH) class switch recombination. Of interest were the noncoding switch (S) regions in the IgH gene that underwent transcription to

produce R-loops. The non-template strand was G-rich, with repeats from 2 to 10 kb in length and they formed GQ [9]. The targeting of the structure by the AID cytosine deaminase protein that was essential for both class switching and immunoglobulin somatic hypermutation is mediated by the interaction of the helicase DDX1 with GQ that is stabilized by the K52A mutation to the Walker A-box that inhibit ATPase activity [9–12]. In other situations, R-loop formation was proposed as pathogenic [13].

Other experimental approaches to unraveling the biology of GQ were complicated by the equilibrium that exists between different flipon conformations, with the transition occurring in unmodified DNA and without requiring any strand cleavage [14]. Early experiments using dimethyl sulfoxide footprinting of RNA failed to show the protection of guanine bases expected if a GQ had formed inside a cell [15], overlooking the fact that if the G-quadruplexes had unfolded during the time course of the experiment, then the chemical modification would prevent GQ from reforming [16]. Nevertheless, the chemical footprinting study highlighted a potential role for helicases in resolving the GQ that form in cells. Other tools designed to detect the non-B-DNA structures like GQ also can induce their formation. This risk of artefact increases when assays are performed on cell extracts. Here various factors come into play, such as the buffers used and the loss of proteins that might otherwise restrain the flip to GQ. Even well accepted ChIP-seq protocols to map protein interactions potentially mislead, as recently shown by a stringent analysis of the GQ binding substrates of PRC2 (Polycomb Repressor Complex 2) interactions [17]. Combined, these uncertainties limited the widespread acceptance of G-flipons as important components of the genetic repertoire.

The intent of this review is to integrate information from a wide range of research papers, including some whose significance has been long overlooked and are not mentioned in many recent GQ reviews [18–26]. The initial focus is on the genetic evidence that speaks to an early evolutionary role for G-flipons in maintaining genomic stability and on the proteins that localize the machinery involved in nucleotide and base excision repair (NER, BER) by inducing GQ formation. Helicases then power the resolution of GQ to reform B-DNA. The responses can also couple environmental events to the expression of genes that offset their effects.

We will then discuss the more complex roles for G-flipons in transcription that emerged later in evolution. This feature reflects a change in how GQ recognition occurs, from interactions involving base extrusion or dependent on modified bases in loops, to those mediated by proteins that bind both B-DNA and G-quadruplexes through a different face of the same helix.

2. Biophysical and Computational Studies of the G-Quadruplex

The basic building block for a G-quadruplex is a guanine tetrad formed by hydrogen bonding, both Watson and Crick as well as Hoogsteen variants [2], between bases that is stabilized by a central metal, with a potassium ion preferred over the smaller sodium and lithium ions [27], but with the dependence varying with the GQ fold and whether it is composed of RNA or DNA [28]. A recent review describes 48 different possible folds reflecting whether the four strands are parallel, anti-parallel or a mix, made from one to 4 different strands with lateral, diagonal or propeller loop topology [29] (Figure 1). In cases of GGA repeats that form tetrads, the adenine bases can interact with the tetrads to produce heptads [30,31]. The stability of GQ can decrease with the loop length and vary with their base composition [32]. With long repeats, the loops can also basepair to give even higher order structures, with 67% located within 6M bp (basepairs) of telomeres. [33]. Interestingly, an increase in loop length and composition has occurred during evolution, especially in mammals, along with increased variation of GQ length, number and density in the genome [34]. There is also an enrichment of G-flipons on the non-template strand of coding genes [35,36].

The intermolecular GQ can form by stacking GQ, one on top of the other or by pairing bases from unconnected strands to form a G-wire [37,38], as originally proposed for the alignment of homologous chromosomes during meiosis [39]. Pairs of G3 repeats in enhancers have been proposed form GQ by forming tetrads with a pair of G3 repeats in promoters [40]. Similarly, pairs of DNA G3 repeats on the non-template strand may form a hybrid GQ with a matching RNA G3 pair in the

nascent transcript [41]. Tetrads missing the fourth base can incorporate into the vacant space a guanine provided *in trans* [42].

RNA tetrads only form parallel GQ [28]. GQ composed of only 2 RNA tetrads have been reported [43] and are stabilized by the 2'-hydroxyl group present in RNA [28]. In contrast, there are many variations of DNA GQ of 3 or more tetrads possible, making it difficult to computationally predict from sequence alone those flipons able to form GQ *in vivo*. A database that combines results from a variety of experimental methods now overcomes this problem by providing a set of well validated G-flipons detected in many different studies using a variety of approaches [44]. The mappings show that In the human genome, GQ forming sequences are enriched in transcription start sites (TSS), in introns and at transcription termination sites (TTS) [34].

The plethora of different GQ DNA topologies allows for different modes of protein recognition. Strategies to confirm these interactions and the specificity of binding to GQ include those that synthesize control oligonucleotides containing the 8-aza-7-deazaguanosine base (Figure 1C) that will not form GQ despite having the same chemical composition as guanine [45]. In these studies, different modes of docking to GQ have been identified, including binding to loop sequences or to 5' and 3' single-strand extensions that give the helicases something to pull on so that they can unwind the structure. Proteins can bind to loops formed when adducted bases such as 8-oxo-G prevent the incorporation of a DNA strand into a GQ, or to the everted bases across from an apurinic/aprimidinic (AP) site. Proteins also dock to the planar tetrad surfaces that form the GQ endplate. Specific binding to RNA rather than DNA GQ is favored by intrinsically disordered regions (IDR) enriched in arginine, glycine repeats, as recently reviewed [46], and visualized in the FMR crystal structure of the Fragile X Mental Retardation Protein bound to an RNA GQ [47]. In principle, the preformed GQ site for docking IDR lowers the entropic cost of binding.

The strength of loop interactions can vary with their length and sequence, as shown for nucleolin and the 2E4 Darpin [48,49]. Further, the latching of a single base by the REV1 polymerase [50], and the docking to an AP site by the APE1 endonuclease [51], can create a surface that induces GQ folding. As we will discuss, the use of SANT domains to recognize parallel-strand GQ is of particular interest as the domains can use the same helix to bind B-DNA in a sequence-specific manner (Figure 2). In total, 50 GQ-peptide structures are present in the Protein Data Base showing a variety of interactions [24,49].

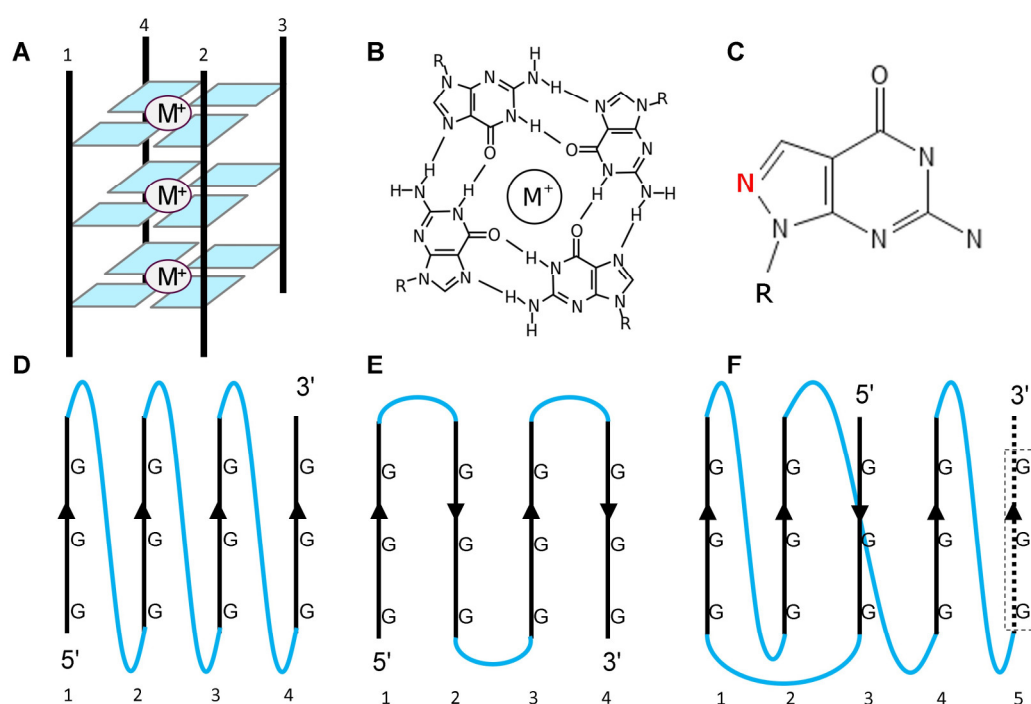


Figure 1. G-quadruples form in many ways. A. The core four-stranded structure formed by stacking the guanine tetrads shown in B. The four strands may form from G repeats on the same molecule or

arise from different molecules or from either RNA or DNA. C, The base 8-aza-7-deazaguanosine retains the same molecular composition as guanosine, but with the red ring nitrogen in a different position. Control oligonucleotides incorporating this nucleotide will not form GQ. In intramolecular GQ, the stands may be parallel (D), anti-parallel (E), or hybrid. The topology of the connecting loops is shown in blue and can be propeller (D), lateral (E) or diagonal (F). M⁺ indicates a metal ion located at the core of the tetrad. K⁺ promotes GQ formation while Li⁺ does not. The dotted strand in F labeled 5 indicates that many sequences capable of forming GQ contain a “spare” tire that can maintain the fold when one of the other repeats is damaged [52].

3. The Accumulating Evidence for the Biological Importance of G-Quadruplexes

G-flipon maintenance. Interestingly, the first evidence hinting at a biological role for GQ came from *C. elegans*. Sequences with the G-quadruplex motif were deleted in strains with *dog-1* (deletions of guanine-rich DNA) LOF variants, but not those with only 3 of the 4 G₃ sequence repeats [53]. Mutant strains of *dog-1* lacking the trans-lesion polymerases (TLS) polymerases, POL eta and POL kappa had significantly more G-tract deletions than *dog-1* by itself [54]. Interestingly, the combined deletion of *dog-1* and the spindle-checkpoint component *mdf-1* enabled long term survival [55], even though a high incidence of lethal mutations in this variant was revealed by the use of balancer chromosomes. In total, 126 (13%) of the 954 mono-G/C tracts, larger than 14 bp, were deleted over 470 generations in a strain deficient in both genes. The effects of *dog-1* LOF on the spindle checkpoint suggest an effect of GQ on sister chromatid alignment by the cohesin proteins during mitosis. The results also supported the consensus that GQ has only a limited role in determining phenotype not only in *C. elegans*, but also other organisms.

Epigenetic Modification. The *dog-1* homolog in the DT40 chicken lymphoblastoid cell line, the 5' FANCI (Fanconi Anemia Complementation Group J) helicase [56] also was found to prevent deletion of G-tracts that also had the potential to form G-quadruplexes. Effects were enhanced by loss of the REV1 polymerase that localizes TLS to sites of polymerase stalling. Interestingly, REV1 catalytic activity was not necessary to prevent deletion, although the LOF variant did enhance the rate of guanine rich DNA loss. Also, in the FANCI model, the combined deletion of the superfamily 2 RecQ (SF2) Werner and Bloom Syndrome 3' helicases [57] also increased G-repeat loss, likely because of reduced GQ resolution [56].

Of interest was that the TLS pathway was required to maintain the epigenetic state of dividing cells, as monitored by cell-surface expression of a protein with an intronic G-flipon that regulated gene expression. Whereas in the wildtype cell, the histone modifications associated with this G-flipon were maintained, they were lost following *rev1* deletion. Instead, resolution of the GQ formed during DNA replication was through the gap-fill repair pathway. The subsequent incorporation of unmodified histones led to diminished gene transcription and surface marker expression. This *rev1*-dependent phenotype could be reverted by re-expression of human SF2 Fe-S FANCI helicase [56]. The opposite effect was observed when a G-flipon was experimentally inserted into a repressed locus. In this case, *rev1* deletion led to depression of the locus, consistent with the replacement of repressive histone with unmodified histones that were permissive to gene expression [58]. These results support a model where the formation of GQ by G-flipons during periods of cell proliferation help transmit the current epigenetic state to progeny, an important biological outcome.

DNA replication and Sister Chromatid Conformation. The involvement of GQ in cell proliferation is further supported by other evidence. During assembly of the DNA polymerase complex at the origin of replication (OR), the MTBP protein assists in the loading of CDC45 into the replicative helicase. The C-terminal domain of MTBP binds GQ in vitro [59]. Notably, G-flipons are enriched in OR. Indeed, in chicken DT20 cells a minimal, functional OR consists of a 90 bp fragments that has two G-flipons on the same strand [60]. These constructs establish the nucleosome depleted region (NDR) bounded by histone H2A.Z that is typical of OR. Collectively, the results suggest a model in which the MTBP binds GQ at the OR to initiate the assembly of the replication complex.

Another potential role for GQ during proliferation and transmission of epigenetic state is to align sister chromatids, as mapping of intra- and inter chromatin interactions between homologous

chromosomes reveals a high degree of symmetry in the architecture of topologically associated domains (TADS), and in the loops formed within TADs [61]. In this regard, a recent report suggests that G-flipons are enriched near sites bound by the CTCF (CCCTC-binding factor), a protein associated with loop formation. Interestingly, the strand orientation of the G-flipons mirrors the inverse orientation of the two CTCF sites that associate with each other to form the base of the loop [62]. CTCF however is not known to bind GQ [45].

G-flipons in NER. The REV1 pathway also plays a role in nucleotide excision repair that is triggered by UV irradiation and the formation of DNA crosslinks. In this situation, loading of the repair pathway proteins such as XPCC and RAD23 is triggered by the protein ZRF1 and its yeast homolog Zuo1 that recognizes the lesion and induces GQ formation [63]. Triggering of this pathway by cytosine deaminases can result in single base substitutions, with a C to G transversion occurring from the preferential insertion of cytidine into the lesion by REV1 [64]. This mutation (STS13) is prevalent in cancers [65]. The Cockayne Syndrome B helicase (encoded by ERCC6 that also binds GQ and performs NER in the transcription coupled repair pathway (TCR) [66]. On sensing a lesion, CSB displaces DSIF from the RNA polymerase 2 (RNAP2) complex, inducing a conformational switch that halts transcriptional elongation and initiates TCR [67]. LOF variants of this enzyme are associated with premature aging phenotypes [66].

G-flipons in BER. APE1 plays a similar role in stabilizing GQ formed by AP DNA, but not unmodified DNA, to initiate the base excision repair (BER) pathway [51]. The pathway removes oxidized bases, such as 8-oxo-G. It is proposed that regulation of the APE1 by acetylation coordinates the expression of genes involved in pathways that respond to oxidative damage to the cell. Interestingly, the GQ involved are formed from G-flipons with a “spare tire”. The extra runs of G-repeats allow formation of a GQ despite damage to one of the other repeats [68]. The flavin-dependent LSD1 demethylation of H3K9me2 also produces 8-oxoG through the hydrogen peroxide generated by the enzyme. The reaction is thought to explain the estrogen-induced activation of BCL-2 [51].

GQ and Hemin. GQ may also play a different role in protecting the genome against oxidative damage caused by intracellular hemin [69]. The high affinity ($K_d \sim 10$ nM) interaction of hemin with GQ was initially highlighted for its ability to generate oxidative species [70], but it appears that this reaction in cells is squelched, presumably by proteins that bind to GQ [69]. Instead, GQ sequesters hemin to prevent damage. Indeed, when released by small molecules that compete for binding to GQ, hemin induces the expression of genes, such as heme oxidase I, that modulate cell death by ferroptosis [71].

GQ and telomere protection. The formation of T-loops by telomeres described above does not rule out a role for GQ formation in telomere protection. Indeed, the TRF2/RAP1 complex protects telomeres from homologous recombination by repressing PARP1 localization to telomeres and by inhibiting the SLX4 Holliday junction (HJ) resolvase. Loss of TRF2 and RAP1 in both humans and mouse leads to rapid telomere attrition, with increased rates of telomere deletion and fusion [72]. TRF2 preferentially docks to GQ formed by RNA, rather than DNA. The protein binds GQ formed by the noncoding Telomeric Repeat-Containing RNA (TERRA) telomere transcript through an RG rich domain [73].

GQ, the SANT domain, and gene expression. Prior to its role in telomere protection, Rap1 was characterized as a sequence-specific transcription factor that bound to an upstream activating site in yeast [74]. The base-specific interaction with B-DNA was confirmed by crystallographic study of a telomeric sequence [75]. Only later did crystal structures show that Rap1 also docked to GQ. Surprisingly, both DNA interactions involved the same helix, but a different face [76] (Figure 2). The GQ contacts were hydrophobic, with the helix lying on the planar surface of the terminal tetrad.

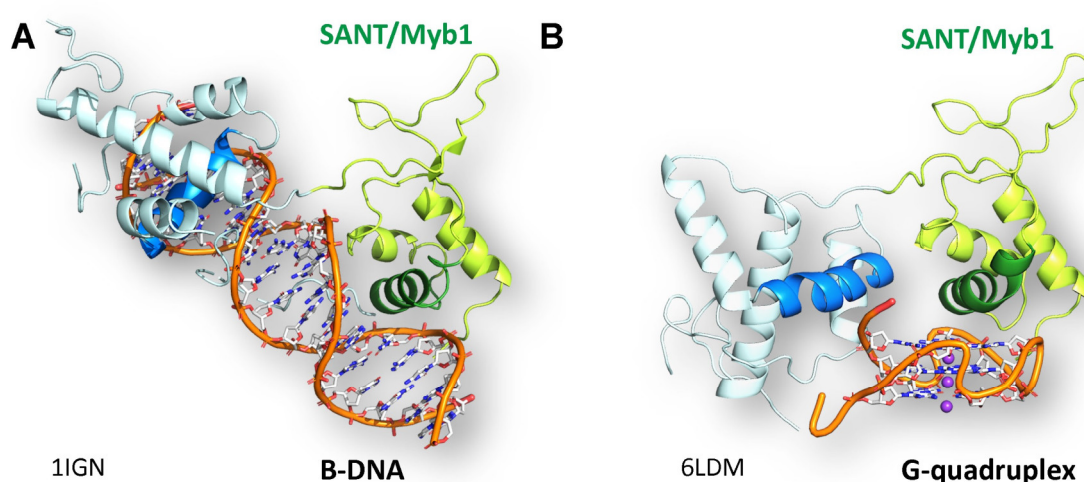


Figure 2. The yeast Rap1 protein binds both B-DNA in a sequence-specific manner (A) [75] and to GQ in a structure-specific manner (B). [76] through different faces of the same helix in the SANT/Myb1 domain Images by Daniela Rhodes.

While this finding might seem anomalous, many subsequent studies have demonstrated the ability of proteins to bind specifically to their cognate B-DNA site, and also to a GQ. This finding is true for binding of the SP1 transcription factor to the c-MYC parallel GQ [77] and for a range of other proteins that bind GQ with nanomolar affinity [19]. Interestingly, like Rap1, many of the GQ binding proteins include a SANT/Myb domain such as ZRF1 [78] and TRF2 [79,80]. Interestingly, the yeast Zuo1 protein has replaced the SANT domain with a highly hydrophobic helix [78]. SANT domain proteins are found in multiple chromatin-modifying and remodeling complexes, although their binding to GQ has not been characterized for each [81].

Helicases and Resolution of GQ. The DEAH box SF2 helicase DDX36 (RHAU), a highly specific GQ resolvase, also makes helical contacts with the GQ end plate [82,83]. However, the helix alone has a K_d of 1 μ M. The additional engagement of a 3' single-stranded DNA tail contributes to the nM affinity for the substrate and the ATP-dependent unwinding activity. Using a ratchet mechanism, the helicase disassembles the GQ, one guanine at a time. The chemical energy derived from ATP is converted into a pulling force by the C-terminal domain rotation that is associated with an opening of the helicase core [83]. In the absence of nucleotide, or in the ADP bound state, *D. melanogaster* DDX36 stabilizes the GQ [84]. The crystal structure of the SF1 *Thermus oshimai* 5'-3' DNA Pif1 helicase shows the enzyme in an unwinding state with engagement of single-stranded thymine repeats [85]. The related yeast helicases *Pif1* and *Rrm3* cooperate to unfold GQ formed not only by telomeres, but also by centromeres and tRNA repeat sequences that form GQ [86,87]. The interaction of the Pif1 with the parallel stranded GQ differs from that of DDX36. The contact is mediated by a cluster of amino acids, including two arginine/lysine cation- π interactions at either end of the GQ, plus ionic contacts with the phosphate backbone. The enzyme unfolds G-quadruplexes in an ATP-dependent manner, unwinding both parallel and antiparallel GQ [85]. The SF2 RecQ BLM helicase also can unfold a range of GQ folds through a variety of different mechanisms [88]. Collectively, the helicases play key but distinct roles in flipping GQ back to the B-DNA conformation.

GQ and transcription complexes. Given the enrichment of G-flipons in promoters, a key question was how do GQ stabilizing and resolving proteins impact transcription. One of the surprises of the ENCODE project was the identification of HOT (high occupancy target) loci where upwards of a 100 TF bound, even to sites lacking their sequence-specific binding motif. The findings were initially dismissed as methodological artefacts [89], but were later shown not to be [90,91]. The work focused on the sequence-specificity of TFs, not the GQs that were also formed at promoters. The ability of TF to bind both B-DNA and GQ offered a resolution to this dilemma presented by HOT [45], especially given the possibility that the helicases engaged by each TF through various domains might recognize

a specific GQ fold, a GQ loop of particular length or composition, or display a preference for a 5' or 3' single-stranded flanking sequence. Indeed, GQ formation is associated with increased TF binding [92].

GQ and Promoter Maintenance. One possibility is that the promiscuous binding of TF at HOT is a mechanism to protect the genome from the oxidative damage due to the sequestration of hemin by promoter GQ [69]. The diversity of associated helicase with TF would allow recognition of a wide variety of GQ folds, including those generated by various replication and repair errors occurring during a cell's lifespan. Another explanation, one that is not mutually exclusive, is that once transcription is initiated by sequence-specific interactions, it could be maintained by docking of TF to GQ without the need for any further sequence-specific interaction. However, this possibility is not consistent with the observed rapid reset of promoters that occurs each round of transcription [93,94]. This disassembly of the transcriptional complexes mirrors the dissolution of promoter condensates that is triggered by the high levels of nascent RNAs produced during transcriptional bursting [95].

GQ and Promoter resets. An alternative explanation for the engagement of GQ by so many TF accounts for this promoter reset. By increasing the local concentration of helicases through the complexes they localize, the TF promote the rapid resolution of both GQ and R-loops structures. The reset then allows the DNA duplex to reform and B-DNA sequence-specific proteins to seed formation of a new pre-initiation complex and so begin a new cycle.

Earlier experiments based on single molecule FISH suggest that the reset of promoters determine burst frequency but not burst size [93]. One contribution to burst size is the frequency with which sister chromatids are transcribed. Curiously, only one allele is active at a time rather than both undergoing simultaneous transcription [93]. The localization of many different helicases to the locus may allow one allele to reload sequence-specific TF to reform an initiation complex while the other one fires. Such coordinated activity is consistent with the symmetrical chromatin architecture observed for sister chromatids, as described above [61]. Many other factors affect promoter reset. For example, the capture by Z-flipons of the energy released by an elongation RNAP2 is thought to power the re-engagement of the RNAP2 complex [96,97]. Of note is that some flipons may experimentally form either GQ or Z-DNA [98]. For example, a GQ loop may contain an embedded Z-DNA forming sequence [99]. The alternative structure adopted may depend on which alternative promoter is used, with the formation of Z-DNA downstream driven by transcription, while GQ upstream of the TSS localizes TF to the locus.

GQ and Gene Repression. The promoter reset occurs in competition with complexes that suppress gene expression. These competitors include the PRC2 complex that engages the GQ formed at promoters through the SANT (Swi3, Ada2, N-Cor, and TFIIB) domain of the EZH2 (enhancer of zeste 2) component. For active genes, binding of PRC2 the GQ formed by a nascent RNA likely prevents engagement of promoter DNA [100]. However, in other situations, binding of a small RNA *in trans* to the coding strand would promote GQ formation by the promoter DNA without the transcription of a GQ RNA competitor. In this situation, proteins, such as PRC2, that are localized to the site by the small RNA, would enhance formation of a repressive complex at the promoter [99]. Indeed, the small RNAs direct the *piwi* mediated repression of endogenous retroelements in early development [101]. By localizing a different set of proteins to the site, small RNAs acting *in trans* could also promote transcriptional activation. Such a role has been proposed for the other *piwi*-related agonaute family member complexes [102,103].

GQ and enhancers. Other mechanisms exist for the induction of alternative flipon conformations. Sequence-specific pioneering transcription factors, such as HNF4 and GATA4, can dock to their motifs on nucleosome bound DNA. The master regulators then localize complexes that evict histone octamers from the locus, generating a negatively supercoiled NDR at the site [104,105]. The energy released by removal of a nucleosome is sufficient to induce a number of different alternative DNA conformations [106]. The relaxation of these structures to B-DNA is sufficient to power the assembly of the different biological machines that actuate cellular responses (Figure 3).

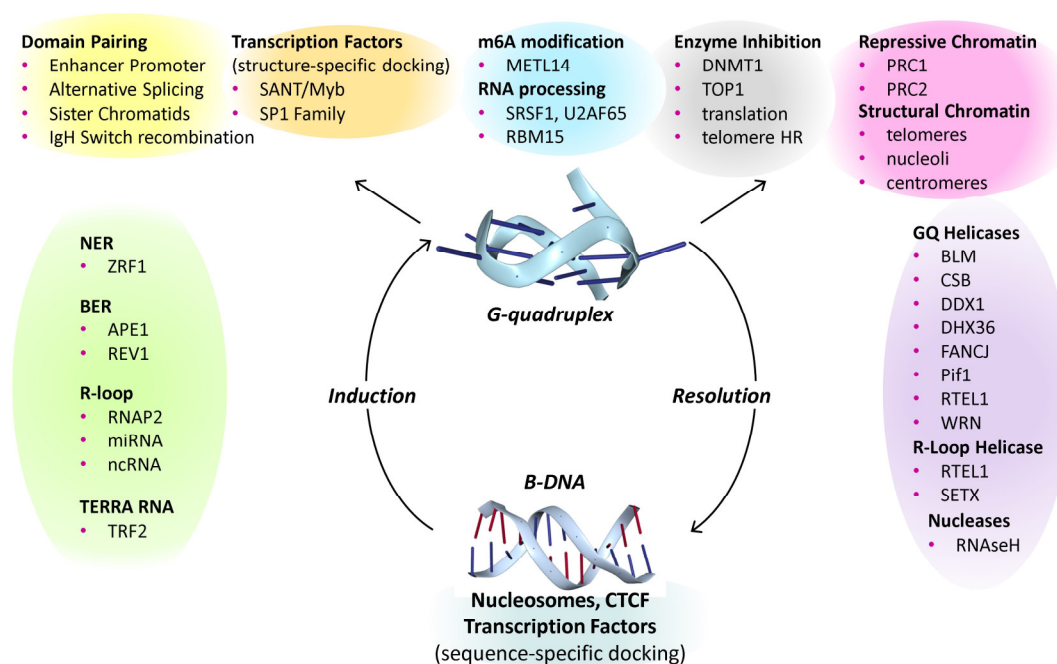


Figure 3. G-flipons cycle between conformations, with specific factors that induce and resolve them, and others that prevent their formation such as histones and enzymes that rapidly remove R-loops. Transcription factors can dock to G-flipons in either a sequence-specific manner to their right-handed conformation or to a GQ, with the affinity dependent on their interactions with loop residues and modified by other domains present in the protein [48,49]. Once formed, the resolution of GQ can be coupled to the different outcomes shown, with both activation and inhibition of gene expression. The inhibition of enzymes like DNMT1 and TOP1 favors maintenance of an unmethylated, nucleosome depleted state that is necessary to rapidly reprogram cellular responses to environmental inputs. (BER: base excision repair; NER: nucleotide excision repair). References for each gene are given in the text.

Indeed, the formation of GQ at promoters often is detectable before the initiation of transcription [107]. In such situations, there is no preference for which strand forms a GQ, further suggesting that gene transcription is not required to induce the flip from B-DNA [77]. The formation GQ formation in promoter regions is not modulated by human topoisomerase I (TOP1), even though the enzyme is enriched at these sites [108]. Instead, TOP1 is inhibited by GQ, with an IC₅₀ ~ 100 nM [109]. TOP1 activation is associated with RNAP2 release from the promoter and occurs 3' to the polymerase [108].

GQ, RNA splicing and RNAP2 pausing. GQ can also form in the gene body and in the RNA transcript produced. These structures have the potential to alter elongation rate and RNA processing [35,36]. Indeed, the small molecule cephaeline and the related compound emetine are both reported to disrupt GQ structure and globally disrupt alternative splicing [110]. The GQ formed by RNA also affects alternative splicing. For example, the splicing factors U2AF65 and SRSF1 bind to GQ RNA with nanomolar affinity, showing specificity for different GQ substrates [111]. The role of GQ in the co-transcriptional N6-methyladenosine (m6A) modification is more controversial. Interestingly, the methyltransferase METTL3/METTL14 heterodimer that writes m6A within the consensus DRACH motif (D = A, G, or U; R = A or G; H = A, C, or U) preferentially binds to rG4 structures through its RGG domain [112,113]. Also, the RBM15 protein that also binds G4 localizes METL3 to certain transcripts and to H3K36me3 marks [45,114,115]. The mapping of GQ and m6A to splice junctions is dependent on the methods used. Over 81% of GQ that map in HeLa cells are formed from only 2 tetrads that will only form GQ in RNA, but the loops are not enriched for the DRACH motif [116]. The mapping frequency likely depends on the m6A detection protocol employed and the cell line studied, varying from 14% in HeLa cells to 40% in HEK cells [116]. More recent methods are even more sensitive than those used in this analysis, but reproducibility across studies remains a problem [117]. The GQ formed by GGA repeats are destabilized by m6A modification [118].

The m6A bases are read by heterogeneous ribonucleic acid proteins (hnRNPs) involved in alternative splicing, such as hnRNP C and hnRNP A2B1 [Ye, 2024 #3182] [119]. However, expression of a hypomorphic METTL3 allele in mouse embryonic stem cells does not appear to change splicing patterns, although turnover of many m6A modified RNAs was slower [120]. In wildtype cells, the distribution of m6A in processed nuclear was similar to that found in cytoplasmic RNAs. Around 70% of the observed m6A sites were in terminal exons, with ~70% in the 3' UTR. In chromatin, ~93% of the m6As in the partially spliced CA-RNAs were in exons and only about 10% of m6As are within 50 nucleotides of 5' or 3' splice sites. Methylation was mostly performed before splicing [121]. Rather than genomic knockout, another group examined the immediate effects of acute depletion of METTL3 protein in order to minimize downstream effects. Around 6%–10% of high-confidence m6A regions were mapped to introns, mainly to those of protein coding genes, either around stop-codon regions or at the beginning of the 3' UTR. The loss of METTL3 disrupted inclusion of alternative introns/exons in the nascent transcriptome, particularly at those 5' splice sites proximal to m6A peaks, suggesting that the sites were occluded or the isoforms were protected by proteins bound to m6A. Above those genes showing altered splicing were those encoding proteins for m6A modification (*Wtap*, *Ythdc1*, *Ythdf1*, and *Spen*), suggesting a negative feedback regulatory mechanism that would be absent in cells with METTL3 deleted from the germline [115]. Overall, the different results for GQ RNA formation at splice sites and METTL3 deficiency are consistent with a model where GQ folding in introns can promote m6A modification of exons, with rapid turnover of splicing isoforms that retain introns marked by m6A.

How GQ formation by DNA affects splicing is therefore of considerable interest as this may impact the timing and location of the events involved. One particular model proposes that GQ formation alters RNAP2 elongation [35]. Thus, sites of RNAP2 pausing in the gene body are of considerable interest and have been investigated at nucleotide resolution. Careful *in vivo* measurement shows dependence of pause sites on the structure of the RNA:DNA hybrid produced, but not with GQ formation [122]. The outcome may reflect the action of the FACT (Facilitates Chromatin Transcription) complex. FACT plays an important role in removing nucleosomes in front of the RNAP2 and replacing them behind the enzyme, preventing the net accumulation of local DNA supercoiling, while maintaining the existing epigenetic state [123].

In contrast, a number of *in cellulo* studies reveal that delays in RNAP2 transcript elongation occur at the CTCF binding sites involved in chromatin loop formation. CTCF binds to the large subunit of RNAP2 and the interaction is also associated with cohesin recruitment [124–126]. Pausing of RNAP2 is associated with alternative splicing (reviewed in [127]). Indeed, the complement of the degenerate pause motif given by Gajos et al., is consistent with a CTCF motif (i.e., the orientation is inverted relative to those enriched at TSS). Interestingly, the binding of CTCF to DNA is inhibited by methylation. The increased engagement of CTCF binding with DNA following deletion of the DNA methylase DNMT1 is associated with an increase in alternative splicing of the CD45 transcript. A similar outcome occurs when DNA demethylation by TET1 (tet methylcytosine dioxygenase 1) and TET2 enzymes [128,129].

These findings are consistent with a model where CTCF induced stalling of the polymerase produces an R-loop that induces GQ formation at the site. The GQ structure then inhibits DNMT1 to prevent DNA methylation of the locus. The binding affinity of DNMT1 is higher for GQ than to duplex, hemi-methylated or single-stranded DNA [130]. The resolution of the GQ by helicases then allows redocking of B-DNA binding CTCF to the original site and the reinstatement of the chromatin loop formed with the promoter. The CTCF binding sites facilitate the interaction because they are in reverse orientation to each other. The interaction between CTCF, RNAP2 and cohesin then sets the scene for the splicing of the subsequent transcript.

The loops formed between promoter and the spliceosome are associated with the transfer of various factors that accumulate in promoter regions to the site where they modulate splicing [131]. Similar outcomes are also reported for the polyadenylation of transcripts [132,133]. The CTCF mechanism of connecting promoters with RNA processing condensates is quite flexible. For example, the multiple alternative splices of the protocadherin *Pcdh* gene family connect the production of each

isoform with a different active promoter [134,135]. The outcomes are aided by the formation of GQ that prevent the loss of CTCF binding sites through methylation and that promote the subsequent rebinding of CTCF by localizing helicases to the region.

Modulation of R-loops. Other mechanisms exist to regulate GQ formation by R-loops. For example, helicases such as SETX, and RTEL1 can facilitate the flip of GQ back to B-DNA through the resolution of RNA:DNA hybrids [136,137]. Nucleases that digest the RNA strand of hybrids, such as RNaseH1, play an important role in their removal [138]. Other proteins such as ATRX prevent R-loop formation at telomeres by sequestering RNA to prevent formation of R-loops. Deletion of ATRX leads to increased formation of GQ at telomeres [139].

GQ and ribosome assembly. GQ appear to play an important role in ribosome structure and maturation, with ribosomal RNAs enriched for G-flipons [140]. A number of ribosomal proteins have been identified as GQ ligands in different screens [111,141]. Further, GQ binding and resolving proteins such as nucleolin and nucleophosmin help structure the nucleolar condensates that guide ribosome assembly [48,142–144].

GQ and translation GQ formation by mRNA is the subject of much interest, especially in the untranslated regions that regulate translation. These exons contain alternative translation initiation sites and microRNA (miR) binding sites that affect the production of different protein isoforms. The complexities involved are described in a number of recent reviews [145–148]. Analysis of G-flipons in 5' and 3' untranslated regions provides evidence of positive selection. These regions also undergo alternative splicing and have single nucleotide variants that are associated with quantitative trait loci [149].

By modulating mRNA translation RNAs, GQ contribute in many ways to phenotypic pliability [26]. Here helicases such as DHX36 and CCHC-type zinc finger nucleic acid-binding protein (CNBP/ZNF9) play a central role in promoting mRNA translation by resolving GQ [150,151]. The m6A modifications of RNA that are associated with GQ formation during transcription (as described above) also impact translation. The removal of these marks from the 5' UTR near the start codon by the m6A erasers AlkB homolog H5 (*ALKBH5*) and fat mass and obesity (*FTO*) decreases ribosome translational pausing, increasing protein synthesis [152]. Such m6A modifications also dynamically regulate heat shock responses by enhancing N7-methylguanosine cap-independent translation [153]. Further, the class I cytoplasmic m6A readers, YTHDF1 and YTHDF3, promote the degradation of target transcripts [154], potentially eliminating partially processed transcripts with retained introns. The endogenous repeat elements present in these introns, such as ALU SINE inverted repeats, might otherwise activate dsRNA and Z-RNA dependent immune responses [97]. The potential of GQ to enhance m6A modifications provides additional mechanistic insight into how G-flipons increase phenotype pliability by regulating RNA dependent epigenetic outcomes.

GQ and development. GQ are able to facilitate a number of different processes in the cell that are directed by sequence-specific TF. Small RNAs, such as those used in the *piwi* system to regulate endogenous retroelements [155], provide another means by which GQ formation can be regulated in a sequence-specific manner. In both cases, the alternative flipon conformations engage the same structure-specific cellular machinery. The question arises as to two these two different systems for sequence-specific regulation of gene expression and RNA translation are used to coordinate development, especially during early embryogenesis. To explore the role of small RNAs in this process, the sequence-specific match between experimentally confirmed flipons and miR highly conserved in eutherian mammals was explored. Intriguingly, promoters with miR matches to G- and Z-flipons were highly enriched in developmental genes ($FDR > 10^{-100}$), consistent with a role in early development [99].

Indeed, GQ are enriched in human embryonic stem cells (hESC). About 18,000 GQ were mapped to NDR as defined by ATAC seq. Following differentiation into neural stem cells and cranial neural crest cells, the number of detectable GQ was reduced by 25-50%, differing by lineage [156]. In hESC, GQ were mapped to ~50% of bivalent promoters that contain both active H3K4me1 and repressive H3K27me3 marks and are lowly transcribed. The GQ in hESC overlapped sites bound by CTCF (~36 %), the cohesin component RAD51 (~50%) and RING1B that mediates repression by recruiting PRC1

to R-loops (~55%) [157]. Differentiation was associated with the loss of bivalent promoters reflecting the potential of GQ to localize either activating or repressive protein complexes during lineage specification. Collectively the results are consistent with a model where small RNAs bootstrap development, much in the same way a computer loads an initial program to specify the inputs and outputs that are necessary for an operating system to run. Here, the programming of flipon conformation by small RNAs would establish epigenetic marks to template tissue differentiation by sequence-specific B-DNA binding proteins. The bootstrapping by small RNAs that occurs after the erasure of existing parental epigenetic marks early in development could potentially involve miR transmitted by either maternal or paternal gametes [158–161]. Further research is needed to address such mechanisms.

4. Summary and Outlook

Flipons are genetically encoded elements that dynamically change their conformation under physiological conditions without requiring strand cleavage. They vary by the non-B-DNA structure they form. Z-flipons flip rapidly, with an in vitro relaxation time of 100 ms and have ancient, well documented roles in self-recognition and immunity through the structure-specific interaction with Z α domain [97]. G-flipons are much more stable, with higher melting temperatures than their B-DNA structure. Yet, like Z-flipons, GQ are formed and resolved dynamically to perform a number of important biological roles (Figure 3). Flipons that form triplexes are also likely to influence gene expression and development [162,163], with examples related to the hemoglobin locus [164], stabilization of by histone H3 tails [165,166], and by binding of the *Drosophila* GAGA protein triplex-DNA through the same domain that engages B-DNA in a sequence-specific manner [167]. Triplex forming sequences are also enriched in repeat elements, such as ALU SINEs (short interspersed nuclear element) that form part of the repetitive genome [168]. Their biology may reflect the RNA motifs they deliver to a locus that engage both sequence- and structure-specific proteins that scaffold formation of various chromatin modifying complexes [169].

Based on a dynamic form of encoding, flipon biology can be best visualized as a cycle that exchanges energy for information. The flip to an alternative conformation is regulated both genetically and by environmental events, by base modifications that enhance or suppress the transition and depend upon proteins and noncoding RNAs that modulate formation or resolution of the alternative conformation. These modulators are also subject to modification to tune the cycle. Other factors also affect the equilibrium by binding in a sequence-specific manner to the right-handed B-DNA conformations or to single-stranded RNA. While it has been usual to consider the effects of evolution on the individual components involved in cellular processes, the optimization of so many different parameters represents a combinatorically challenging calculation full of cascading complexity, similar in logic to the epicycles once used to predict planetary orbits. Instead, flipons offer a simpler alternative to optimize context-specific responses that allows rapid adjustments of cellular state. By programming and refreshing epigenetic state, flipons facilitate the formation and maintenance of cellular memory [170]. Here, the various ways in which G-flipons impact a wide variety of biological processes is described, with a focus on recent experimental studies of GQ and descriptions of the current unknowns.

References

1. Gellert, M.; Lipsett, M. N.; Davies, D. R., Helix Formation by Guanylic Acid. *Proceedings of the National Academy of Sciences* **1962**, 48, (12), 2013-2018.
2. Arnott, S.; Chandrasekaran, R.; Marttila, C. M., Structures for polyinosinic acid and polyguanylic acid. *Biochem J* **1974**, 141, (2), 537-43.
3. Sauer, M.; Paeschke, K., G-quadruplex unwinding helicases and their function in vivo. *Biochem Soc Trans* **2017**, 45, (5), 1173-1182.
4. Blackburn, E. H.; Gall, J. G., A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J Mol Biol* **1978**, 120, (1), 33-53.
5. Sundquist, W. I.; Klug, A., Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **1989**, 342, (6251), 825-9.

6. Griffith, J. D.; Comeau, L.; Rosenfield, S.; Stansel, R. M.; Bianchi, A.; Moss, H.; de Lange, T., Mammalian telomeres end in a large duplex loop. *Cell* **1999**, 97, (4), 503-14.
7. Huber, M. D.; Duquette, M. L.; Shiels, J. C.; Maizels, N., A conserved G4 DNA binding domain in RecQ family helicases. *J Mol Biol* **2006**, 358, (4), 1071-80.
8. Maizels, N., G4-associated human diseases. *EMBO Rep* **2015**, 16, (8), 910-22.
9. Duquette, M. L.; Handa, P.; Vincent, J. A.; Taylor, A. F.; Maizels, N., Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* **2004**, 18, (13), 1618-29.
10. Okazaki, I. M.; Kinoshita, K.; Muramatsu, M.; Yoshikawa, K.; Honjo, T., The AID enzyme induces class switch recombination in fibroblasts. *Nature* **2002**, 416, (6878), 340-5.
11. Qiao, Q.; Wang, L.; Meng, F. L.; Hwang, J. K.; Alt, F. W.; Wu, H., AID Recognizes Structured DNA for Class Switch Recombination. *Mol Cell* **2017**, 67, (3), 361-373 e4.
12. Ribeiro de Almeida, C.; Dhir, S.; Dhir, A.; Moghaddam, A. E.; Sattentau, Q.; Meinhart, A.; Proudfoot, N. J., RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Mol Cell* **2018**, 70, (4), 650-662 e8.
13. Richard, P.; Manley, J. L., R Loops and Links to Human Disease. *J Mol Biol* **2017**, 429, (21), 3168-3180.
14. Herbert, A., A Genetic Instruction Code Based on DNA Conformation. *Trends Genet* **2019**, 35, 887-890.
15. Guo, J. U.; Bartel, D. P., RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **2016**, 353, (6306).
16. Di Antonio, M.; Ponjavic, A.; Radzevicius, A.; Ranasinghe, R. T.; Catalano, M.; Zhang, X.; Shen, J.; Needham, L. M.; Lee, S. F.; Klenerman, D.; Balasubramanian, S., Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat Chem* **2020**, 12, (9), 832-837.
17. Guo, J. K.; Blanco, M. R.; Walkup, W. G. t.; Bonesteele, G.; Urbinati, C. R.; Banerjee, A. K.; Chow, A.; Ettlin, O.; Strehle, M.; Peyda, P.; Amaya, E.; Trinh, V.; Guttman, M., Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo. *Mol Cell* **2024**, 84, (7), 1271-1289 e12.
18. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S., The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol* **2020**, 21, 459-474.
19. Spiegel, J.; Adhikari, S.; Balasubramanian, S., The Structure and Function of DNA G-Quadruplexes. *Trends Chem* **2020**, 2, (2), 123-136.
20. Yadav, P.; Kim, N.; Kumari, M.; Verma, S.; Sharma, T. K.; Yadav, V.; Kumar, A., G-Quadruplex Structures in Bacteria: Biological Relevance and Potential as an Antimicrobial Target. *J Bacteriol* **2021**, 203, (13), e0057720.
21. Wang, E.; Thombre, R.; Shah, Y.; Latanich, R.; Wang, J., G-Quadruplexes as pathogenic drivers in neurodegenerative disorders. *Nucleic Acids Res* **2021**, 49, (9), 4816-4830.
22. Lejault, P.; Mitteaux, J.; Sperti, F. R.; Monchaud, D., How to untie G-quadruplex knots and why? *Cell Chem Biol* **2021**, 28, (4), 436-455.
23. Sato, K.; Knipscheer, P., G-quadruplex resolution: From molecular mechanisms to physiological relevance. *DNA Repair (Amst)* **2023**, 130, 103552.
24. Troisi, R.; Sica, F., Structural overview of DNA and RNA G-quadruplexes in their interaction with proteins. *Curr Opin Struct Biol* **2024**, 87, 102846.
25. Sahayasheela, V. J.; Sugiyama, H., RNA G-quadruplex in functional regulation of noncoding RNA: Challenges and emerging opportunities. *Cell Chem Biol* **2024**, 31, (1), 53-70.
26. Cammas, A.; Desprairies, A.; Dassi, E.; Millevoi, S., The shaping of mRNA translation plasticity by RNA G-quadruplexes in cancer progression and therapy resistance. *NAR Cancer* **2024**, 6, (2), zcae025.
27. Sen, D.; Gilbert, W., A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **1990**, 344, (6265), 410-4.
28. Fay, M. M.; Lyons, S. M.; Ivanov, P., RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J Mol Biol* **2017**, 429, (14), 2127-2147.
29. Sundaresan, S.; Uttamrao, P. P.; Kovuri, P.; Rathinavelan, T., The entangled world of DNA quadruplex folds. *BioRxiv* **2024**, 2024.
30. Matsugami, A.; Okuizumi, T.; Uesugi, S.; Katahira, M., Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J Biol Chem* **2003**, 278, (30), 28147-53.
31. Palumbo, S. L.; Memmott, R. M.; Uribe, D. J.; Krotova-Khan, Y.; Hurley, L. H.; Ebbinghaus, S. W., A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic Acids Res* **2008**, 36, (6), 1755-69.
32. Piazza, A.; Adrian, M.; Samazan, F.; Heddi, B.; Hamon, F.; Serero, A.; Lopes, J.; Teulade-Fichou, M. P.; Phan, A. T.; Nicolas, A., Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *Embo J* **2015**, 34, (12), 1718-34.

33. Williams, J. D.; Houserova, D.; Johnson, B. R.; Dyniewski, B.; Berroyer, A.; French, H.; Barchie, A. A.; Bilbrey, D. D.; Demeis, J. D.; Ghee, K. R.; Hughes, A. G.; Kreitz, N. W.; McInnis, C. H.; Pudner, S. C.; Reeves, M. N.; Stahly, A. N.; Turcu, A.; Watters, B. C.; Daly, G. T.; Langley, R. J.; Gillespie, M. N.; Prakash, A.; Larson, E. D.; Kasukurthi, M. V.; Huang, J.; Jinks-Robertson, S.; Borchert, G. M., Characterization of long G4-rich enhancer-associated genomic regions engaging in a novel loop:loop 'G4 Kissing' interaction. *Nucleic Acids Res* **2020**, *48*, (11), 5907-5925.
34. Wu, F.; Niu, K.; Cui, Y.; Li, C.; Lyu, M.; Ren, Y.; Chen, Y.; Deng, H.; Huang, L.; Zheng, S.; Liu, L.; Wang, J.; Song, Q.; Xiang, H.; Feng, Q., Genome-wide analysis of DNA G-quadruplex motifs across 37 species provides insights into G4 evolution. *Commun Biol* **2021**, *4*, (1), 98.
35. Lee, C. Y.; McNerney, C.; Ma, K.; Zhao, W.; Wang, A.; Myong, S., R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nat Commun* **2020**, *11*, (1), 3392.
36. Georgakopoulos-Soares, I.; Parada, G. E.; Wong, H. Y.; Medhi, R.; Furlan, G.; Munita, R.; Miska, E. A.; Kwok, C. K.; Hemberg, M., Alternative splicing modulation by G-quadruplexes. *Nat Commun* **2022**, *13*, (1), 2404.
37. Varizhuk, A. M.; Protopopova, A. D.; Tsvetkov, V. B.; Barinov, N. A.; Podgorsky, V. V.; Tankevich, M. V.; Vlasenok, M. A.; Severov, V. V.; Smirnov, I. P.; Dubrovin, E. V.; Klinov, D. V.; Pozmogova, G. E., Polymorphism of G4 associates: from stacks to wires via interlocks. *Nucleic Acids Res* **2018**, *46*, (17), 8978-8992.
38. Kolesnikova, S.; Curtis, E. A., Structure and Function of Multimeric G-Quadruplexes. *Molecules* **2019**, *24*, (17).
39. Sen, D.; Gilbert, W., Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988**, *334*, (6180), 364-6.
40. Hegyi, H., Enhancer-promoter interaction facilitated by transiently forming G-quadruplexes. *Sci Rep* **2015**, *5*, 9165.
41. Zheng, K. W.; Xiao, S.; Liu, J. Q.; Zhang, J. Y.; Hao, Y. H.; Tan, Z., Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res* **2013**, *41*, (10), 5533-41.
42. Li, X.-m.; Zheng, K.-w.; Zhang, J.-y.; Liu, H.-h.; He, Y.-d.; Yuan, B.-f.; Hao, Y.-h.; Tan, Z., Guanine-vacancy-bearing G-quadruplexes responsive to guanine derivatives. *Proceedings of the National Academy of Sciences* **2015**, *112*, (47), 14581-14586.
43. Lavezzo, E.; Berselli, M.; Frasson, I.; Perrone, R.; Palu, G.; Brazzale, A. R.; Richter, S. N.; Toppo, S., G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide. *PLoS Comput Biol* **2018**, *14*, (12), e1006675.
44. Qian, S. H.; Shi, M. W.; Xiong, Y. L.; Zhang, Y.; Zhang, Z. H.; Song, X. M.; Deng, X. Y.; Chen, Z. X., EndoQuad: a comprehensive genome-wide experimentally validated endogenous G-quadruplex database. *Nucleic Acids Res* **2024**, *52*, (D1), D72-D80.
45. Spiegel, J.; Cuesta, S. M.; Adhikari, S.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S., G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol* **2021**, *22*, (1), 117.
46. Brazda, V.; Cerven, J.; Bartas, M.; Mikyskova, N.; Coufal, J.; Pecinka, P., The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* **2018**, *23*, (9).
47. Vasilyev, N.; Polonskaia, A.; Darnell, J. C.; Darnell, R. B.; Patel, D. J.; Serganov, A., Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proceedings of the National Academy of Sciences* **2015**, *112*, (39), E5391-E5400.
48. Saha, A.; Duchambon, P.; Masson, V.; Loew, D.; Bombard, S.; Teulade-Fichou, M. P., Nucleolin Discriminates Drastically between Long-Loop and Short-Loop Quadruplexes. *Biochemistry* **2020**, *59*, (12), 1261-1272.
49. Ngo, K. H.; Liew, C. W.; Heddi, B.; Phan, A. T., Structural Basis for Parallel G-Quadruplex Recognition by an Ankyrin Protein. *J Am Chem Soc* **2024**, *146*, (20), 13709-13713.
50. Weaver, T. M.; Cortez, L. M.; Khoang, T. H.; Washington, M. T.; Agarwal, P. K.; Freudenthal, B. D., Visualizing Rev1 catalyze protein-template DNA synthesis. *Proc Natl Acad Sci U S A* **2020**, *117*, (41), 25494-25504.
51. Roychoudhury, S.; Pramanik, S.; Harris, H. L.; Tarpley, M.; Sarkar, A.; Spagnol, G.; Sorgen, P. L.; Chowdhury, D.; Band, V.; Klinkebiel, D.; Bhakat, K. K., Endogenous oxidized DNA bases and APE1 regulate the formation of G-quadruplex structures in the genome. *Proc Natl Acad Sci U S A* **2020**, *117*, (21), 11409-11420.
52. Fleming, A. M.; Zhou, J.; Wallace, S. S.; Burrows, C. J., A Role for the Fifth G-Track in G-Quadruplex Forming Oncogene Promoter Sequences during Oxidative Stress: Do These "Spare Tires" Have an Evolved Function? *ACS Cent Sci* **2015**, *1*, (5), 226-233.

53. Kruisselbrink, E.; Guryev, V.; Brouwer, K.; Pontier, D. B.; Cuppen, E.; Tijsterman, M., Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Curr Biol* **2008**, *18*, (12), 900-5.
54. Jones, M.; Rose, A., A DOG's View of Fanconi Anemia: Insights from *C. elegans*. *Anemia* **2012**, *2012*, 323721.
55. Tarailo-Graovac, M.; Wong, T.; Qin, Z.; Flibotte, S.; Taylor, J.; Moerman, D. G.; Rose, A. M.; Chen, N., Spectrum of variations in dog-1/FANCD1 and mdf-1/MAD1 defective *Caenorhabditis elegans* strains after long-term propagation. *BMC Genomics* **2015**, *16*, (1), 210.
56. Sarkies, P.; Murat, P.; Phillips, L. G.; Patel, K. J.; Balasubramanian, S.; Sale, J. E., FANCD1 coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res* **2012**, *40*, (4), 1485-98.
57. Liu, Y.; Zhu, X.; Wang, K.; Zhang, B.; Qiu, S., The Cellular Functions and Molecular Mechanisms of G-Quadruplex Unwinding Helicases in Humans. *Front Mol Biosci* **2021**, *8*, 783889.
58. Sarkies, P.; Reams, C.; Simpson, L. J.; Sale, J. E., Epigenetic instability due to defective replication of structured DNA. *Mol Cell* **2010**, *40*, (5), 703-13.
59. Kumagai, A.; Dunphy, W. G., MTBP, the partner of Treslin, contains a novel DNA-binding domain that is essential for proper initiation of DNA replication. *Mol Biol Cell* **2017**, *28*, (22), 2998-3012.
60. Poulet-Benedetti, J.; Tonnerre-Doncarli, C.; Valton, A. L.; Laurent, M.; Gerard, M.; Barinova, N.; Parisi, N.; Massip, F.; Picard, F.; Prioleau, M. N., Dimeric G-quadruplex motifs-induced NFRs determine strong replication origins in vertebrates. *Nat Commun* **2023**, *14*, (1), 4843.
61. Mitter, M.; Gasser, C.; Takacs, Z.; Langer, C. C. H.; Tang, W.; Jessberger, G.; Beales, C. T.; Neuner, E.; Ameres, S. L.; Peters, J. M.; Goloborodko, A.; Micura, R.; Gerlich, D. W., Conformation of sister chromatids in the replicated human genome. *Nature* **2020**, *586*, (7827), 139-144.
62. Hou, Y.; Li, F.; Zhang, R.; Li, S.; Liu, H.; Qin, Z. S.; Sun, X., Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics* **2019**, *14*, (9), 894-911.
63. De Magis, A.; Gotz, S.; Hajikazemi, M.; Fekete-Szucs, E.; Caterino, M.; Juranek, S.; Paeschke, K., Zuo1 supports G4 structure formation and directs repair toward nucleotide excision repair. *Nat Commun* **2020**, *11*, (1), 3907.
64. Ketkar, A.; Smith, L.; Johnson, C.; Richey, A.; Berry, M.; Hartman, J. H.; Maddukuri, L.; Reed, M. R.; Gunderson, J. E. C.; Leung, J. W. C.; Eoff, R. L., Human Rev1 relies on insert-2 to promote selective binding and accurate replication of stabilized G-quadruplex motifs. *Nucleic Acids Res* **2021**, *49*, (4), 2065-2084.
65. Sondka, Z.; Dhir, N. B.; Carvalho-Silva, D.; Jupe, S.; Madhumita; McLaren, K.; Starkey, M.; Ward, S.; Wilding, J.; Ahmed, M.; Argasinska, J.; Beare, D.; Chawla, M. S.; Duke, S.; Fasanella, I.; Neogi, A. G.; Haller, S.; Hetenyi, B.; Hodges, L.; Holmes, A.; Lyne, R.; Maurel, T.; Nair, S.; Pedro, H.; Sangrador-Vegas, A.; Schuilenburg, H.; Sheard, Z.; Yong, S. Y.; Teague, J., COSMIC: a curated database of somatic variants and clinical data for cancer. *Nucleic Acids Res* **2024**, *52*, (D1), D1210-D1217.
66. Liano, D.; Chowdhury, S.; Di Antonio, M., Cockayne Syndrome B Protein Selectively Resolves and Interacts with Intermolecular DNA G-Quadruplex Structures. *J Am Chem Soc* **2021**, *143*, (49), 20988-21002.
67. Kokic, G.; Wagner, F. R.; Chernev, A.; Urlaub, H.; Cramer, P., Structural basis of human transcription-DNA repair coupling. *Nature* **2021**, *598*, (7880), 368-372.
68. Fleming, A. M.; Zhu, J.; Ding, Y.; Esders, S.; Burrows, C. J., Oxidative Modification of Guanine in a Potential Z-DNA-Forming Sequence of a Gene Promoter Impacts Gene Expression. *Chem Res Toxicol* **2019**, *32*, (5), 899-909.
69. Gray, L. T.; Puig Lombardi, E.; Verga, D.; Nicolas, A.; Teulade-Fichou, M. P.; Londono-Vallejo, A.; Maizels, N., G-quadruplexes Sequester Free Heme in Living Cells. *Cell Chem Biol* **2019**, *26*, (12), 1681-1691 e5.
70. Li, Y.; Geyer, C. R.; Sen, D., Recognition of anionic porphyrins by DNA aptamers. *Biochemistry* **1996**, *35*, (21), 6911-22.
71. Tang, D.; Chen, X.; Kang, R.; Kroemer, G., Ferroptosis: molecular mechanisms and health implications. *Cell Res* **2021**, *31*, (2), 107-125.
72. Rai, R.; Chen, Y.; Lei, M.; Chang, S., TRF2-RAP1 is required to protect telomeres from engaging in homologous recombination-mediated deletions and fusions. *Nat Commun* **2016**, *7*, 10881.
73. Mei, Y.; Deng, Z.; Vladimirova, O.; Gulve, N.; Johnson, F. B.; Drosopoulos, W. C.; Schildkraut, C. L.; Lieberman, P. M., TERRA G-quadruplex RNA interaction with TRF2 GAR domain is required for telomere integrity. *Sci Rep* **2021**, *11*, (1), 3509.
74. Huet, J.; Cottrelle, P.; Cool, M.; Vignais, M. L.; Thiele, D.; Marck, C.; Buhler, J. M.; Sentenac, A.; Fromageot, P., A general upstream binding factor for genes of the yeast translational apparatus. *Embo J* **1985**, *4*, (13A), 3539-3547.
75. König, P.; Giraldo, R.; Chapman, L.; Rhodes, D., The Crystal Structure of the DNA-Binding Domain of Yeast RAP1 in Complex with Telomeric DNA. *Cell* **1996**, *85*, (1), 125-136.
76. Traczyk, A.; Liew, C. W.; Gill, D. J.; Rhodes, D., Structural basis of G-quadruplex DNA recognition by the yeast telomeric protein Rap1. *Nucleic Acids Res* **2020**, *48*, (8), 4562-4571.

77. Esain-Garcia, I.; Kirchner, A.; Melidis, L.; Tavares, R. C. A.; Dhir, S.; Simeone, A.; Yu, Z.; Madden, S. K.; Hermann, R.; Tannahill, D.; Balasubramanian, S., G-quadruplex DNA structure is a positive regulator of MYC transcription. *Proc Natl Acad Sci U S A* **2024**, 121, (7), e2320240121.
78. Shrestha, O. K.; Sharma, R.; Tomiczek, B.; Lee, W.; Tonelli, M.; Cornilescu, G.; Stolarska, M.; Nierzwicki, L.; Czub, J.; Markley, J. L.; Marszalek, J.; Ciesielski, S. J.; Craig, E. A., Structure and evolution of the 4-helix bundle domain of Zuotin, a J-domain protein co-chaperone of Hsp70. *PLoS One* **2019**, 14, (5), e0217098.
79. Biffi, G.; Tannahill, D.; Balasubramanian, S., An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J Am Chem Soc* **2012**, 134, (29), 11974-6.
80. Sharma, S.; Mukherjee, A. K.; Roy, S. S.; Bagri, S.; Lier, S.; Verma, M.; Sengupta, A.; Kumar, M.; Nesse, G.; Pandey, D. P.; Chowdhury, S., Human telomerase is directly regulated by non-telomeric TRF2-G-quadruplex interaction. *Cell Rep* **2021**, 35, (7), 109154.
81. Boyer, L. A.; Latek, R. R.; Peterson, C. L., The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol* **2004**, 5, (2), 158-63.
82. Heddi, B.; Cheong, V. V.; Martadinata, H.; Phan, A. T., 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* **2015**, 112, (31), 9608-13.
83. Chen, M. C.; Tippiana, R.; Demeshkina, N. A.; Murat, P.; Balasubramanian, S.; Myong, S.; Ferre-D'Amare, A. R., Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **2018**, 558, (7710), 465-469.
84. You, H.; Lattmann, S.; Rhodes, D.; Yan, J., RHAU helicase stabilizes G4 in its nucleotide-free state and destabilizes G4 upon ATP hydrolysis. *Nucleic Acids Res* **2017**, 45, (1), 206-214.
85. Dai, Y. X.; Guo, H. L.; Liu, N. N.; Chen, W. F.; Ai, X.; Li, H. H.; Sun, B.; Hou, X. M.; Rety, S.; Xi, X. G., Structural mechanism underpinning Thermus oshimai Pif1-mediated G-quadruplex unfolding. *EMBO Rep* **2022**, 23, (7), e53874.
86. Muellner, J.; Schmidt, K. H., Yeast Genome Maintenance by the Multifunctional PIF1 DNA Helicase Family. *Genes (Basel)* **2020**, 11, (2).
87. Varon, M.; Dovrat, D.; Heuze, J.; Tsirkas, I.; Singh, S. P.; Pasero, P.; Galletto, R.; Aharoni, A., Rrm3 and Pif1 division of labor during replication through leading and lagging strand G-quadruplex. *Nucleic Acids Res* **2024**, 52, (4), 1753-1762.
88. Wu, W. Q.; Hou, X. M.; Li, M.; Dou, S. X.; Xi, X. G., BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic Acids Res* **2015**, 43, (9), 4614-26.
89. Wreczycka, K.; Franke, V.; Uyar, B.; Wurmus, R.; Bulut, S.; Tursun, B.; Akalin, A., HOT or not: examining the basis of high-occupancy target regions. *Nucleic Acids Res* **2019**, 47, (11), 5735-5745.
90. Ramaker, R. C.; Hardigan, A. A.; Goh, S. T.; Partridge, E. C.; Wold, B.; Cooper, S. J.; Myers, R. M., Dissecting the regulatory activity and sequence content of loci with exceptional numbers of transcription factor associations. *Genome Res* **2020**, 30, (7), 939-950.
91. Partridge, E. C.; Chhetri, S. B.; Prokop, J. W.; Ramaker, R. C.; Jansen, C. S.; Goh, S. T.; Mackiewicz, M.; Newberry, K. M.; Brandsmeier, L. A.; Meadows, S. K.; Messer, C. L.; Hardigan, A. A.; Coppola, C. J.; Dean, E. C.; Jiang, S.; Savic, D.; Mortazavi, A.; Wold, B. J.; Myers, R. M.; Mendenhall, E. M., Occupancy maps of 208 chromatin-associated proteins in one human cell type. *Nature* **2020**, 583, (7818), 720-728.
92. Lago, S.; Nadai, M.; Cernilogar, F. M.; Kazerani, M.; Dominiguez Moreno, H.; Schotta, G.; Richter, S. N., Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat Commun* **2021**, 12, (1), 3885.
93. Bartman, C. R.; Hsu, S. C.; Hsiung, C. C.; Raj, A.; Blobel, G. A., Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell* **2016**, 62, (2), 237-247.
94. Hasegawa, Y.; Struhl, K., Promoter-specific dynamics of TATA-binding protein association with the human genome. *Genome Res* **2019**, 29, (12), 1939-1950.
95. Henninger, J. E.; Oksuz, O.; Shrinivas, K.; Sagi, I.; LeRoy, G.; Zheng, M. M.; Andrews, J. O.; Zamudio, A. V.; Lazaris, C.; Hannett, N. M.; Lee, T. I.; Sharp, P. A.; Cisse, I.; Chakraborty, A. K.; Young, R. A., RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **2021**, 184, (1), 207-225 e24.
96. Herbert, A., Flipons and small RNAs accentuate the asymmetries of pervasive transcription by the reset and sequence-specific microcoding of promoter conformation. *J Biol Chem* **2023**, 299, (9), 105140.
97. Herbert, A., The ancient Z-DNA and Z-RNA specific α fold has evolved modern roles in immunity and transcription through the natural selection of flipons. *Royal Society Open Science* **2024**, 11, (6).
98. Kouzine, F.; Wojtowicz, D.; Baranello, L.; Yamane, A.; Nelson, S.; Resch, W.; Kieffer-Kwon, K. R.; Benham, C. J.; Casellas, R.; Przytycka, T. M.; Levens, D., Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst* **2017**, 4, (3), 344-356.
99. Herbert, A.; Pavlov, F.; Konovalov, D.; Poptsova, M., Conserved microRNAs and Flipons Shape Gene Expression during Development by Altering Promoter Conformations. *Int J Mol Sci* **2023**, 24, (5).

100. Song, J.; Gooding, A. R.; Hemphill, W. O.; Love, B. D.; Robertson, A.; Yao, L.; Zon, L. I.; North, T. E.; Kasinath, V.; Cech, T. R., Structural basis for inactivation of PRC2 by G-quadruplex RNA. *Science* **2023**, 381, (6664), 1331-1337.
101. Watanabe, T.; Totoki, Y.; Toyoda, A.; Kaneda, M.; Kuramochi-Miyagawa, S.; Obata, Y.; Chiba, H.; Kohara, Y.; Kono, T.; Nakano, T.; Surani, M. A.; Sakaki, Y.; Sasaki, H., Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **2008**, 453, (7194), 539-43.
102. Li, L. C.; Okino, S. T.; Zhao, H.; Pookot, D.; Place, R. F.; Urakami, S.; Enokida, H.; Dahiya, R., Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci U S A* **2006**, 103, (46), 17337-42.
103. Matsui, M.; Chu, Y.; Zhang, H.; Gagnon, K. T.; Shaikh, S.; Kuchimanchi, S.; Manoharan, M.; Corey, D. R.; Janowski, B. A., Promoter RNA links transcriptional regulation of inflammatory pathway genes. *Nucleic Acids Res* **2013**, 41, (22), 10086-109.
104. Cirillo, L. A.; Lin, F. R.; Cuesta, I.; Friedman, D.; Jarnik, M.; Zaret, K. S., Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **2002**, 9, (2), 279-89.
105. Zaret, K. S., Pioneer Transcription Factors Initiating Gene Network Changes. *Annu Rev Genet* **2020**, 54, 367-385.
106. Herbert, A., Nucleosomes and flipons exchange energy to alter chromatin conformation, the readout of genomic information, and cell fate. *Bioessays* **2022**, 44, (12), e2200166.
107. Shen, J.; Varshney, D.; Simeone, A.; Zhang, X.; Adhikari, S.; Tannahill, D.; Balasubramanian, S., Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome Biol* **2021**, 22, (1), 143.
108. Baranello, L.; Wojtowicz, D.; Cui, K.; Devaiah, B. N.; Chung, H. J.; Chan-Salis, K. Y.; Guha, R.; Wilson, K.; Zhang, X.; Zhang, H.; Piotrowski, J.; Thomas, C. J.; Singer, D. S.; Pugh, B. F.; Pommier, Y.; Przytycka, T. M.; Kouzine, F.; Lewis, B. A.; Zhao, K.; Levens, D., RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **2016**, 165, (2), 357-71.
109. Marchand, C.; Pourquier, P.; Laco, G. S.; Jing, N.; Pommier, Y., Interaction of Human Nuclear Topoisomerase I with Guanosine Quartet-forming and Guanosine-rich Single-stranded DNA and RNA Oligonucleotides. *Journal of Biological Chemistry* **2002**, 277, (11), 8906-8911.
110. Zhang, J.; Harvey, S. E.; Cheng, C., A high-throughput screen identifies small molecule modulators of alternative splicing by targeting RNA G-quadruplexes. *Nucleic Acids Res* **2019**, 47, (7), 3667-3679.
111. von Hacht, A.; Seifert, O.; Menger, M.; Schutze, T.; Arora, A.; Konthur, Z.; Neubauer, P.; Wagner, A.; Weise, C.; Kurreck, J., Identification and characterization of RNA guanine-quadruplex binding proteins. *Nucleic Acids Res* **2014**, 42, (10), 6630-44.
112. Fleming, A. M.; Nguyen, N. L. B.; Burrows, C. J., Colocalization of m(6)A and G-Quadruplex-Forming Sequences in Viral RNA (HIV, Zika, Hepatitis B, and SV40) Suggests Topological Control of Adenosine N (6)-Methylation. *ACS Cent Sci* **2019**, 5, (2), 218-228.
113. Yoshida, A.; Oyoshi, T.; Suda, A.; Futaki, S.; Imanishi, M., Recognition of G-quadruplex RNA by a crucial RNA methyltransferase component, METTL14. *Nucleic Acids Res* **2022**, 50, (1), 449-457.
114. Patil, D. P.; Chen, C. K.; Pickering, B. F.; Chow, A.; Jackson, C.; Guttman, M.; Jaffrey, S. R., m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **2016**, 537, (7620), 369-373.
115. Wei, G.; Almeida, M.; Pintacuda, G.; Coker, H.; Bowness, J. S.; Ule, J.; Brockdorff, N., Acute depletion of METTL3 implicates N (6)-methyladenosine in alternative intron/exon inclusion in the nascent transcriptome. *Genome Res* **2021**, 31, (8), 1395-1408.
116. Jara-Espejo, M.; Fleming, A. M.; Burrows, C. J., Potential G-Quadruplex Forming Sequences and N(6)-Methyladenosine Colocalize at Human Pre-mRNA Intron Splice Sites. *ACS Chem Biol* **2020**, 15, (6), 1292-1300.
117. Ye, H.; Li, T.; Rigden, D. J.; Wei, Z., m6ACali: machine learning-powered calibration for accurate m6A detection in MeRIP-Seq. *Nucleic Acids Res* **2024**, 52, (9), 4830-4842.
118. Iwasaki, Y.; Ookuro, Y.; Iida, K.; Nagasawa, K.; Yoshida, W., Destabilization of DNA and RNA G-quadruplex structures formed by GGA repeat due to N(6)-methyladenine modification. *Biochem Biophys Res Commun* **2022**, 597, 134-139.
119. Shi, H.; Wei, J.; He, C., Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Mol Cell* **2019**, 74, (4), 640-650.
120. Darnell, R. B.; Ke, S.; Darnell, J. E., Jr., Pre-mRNA processing includes N(6) methylation of adenosine residues that are retained in mRNA exons and the fallacy of "RNA epigenetics". *Rna* **2018**, 24, (3), 262-267.
121. Ke, S.; Pandya-Jones, A.; Saito, Y.; Fak, J. J.; Vagbo, C. B.; Geula, S.; Hanna, J. H.; Black, D. L.; Darnell, J. E., Jr.; Darnell, R. B., m(6)A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev* **2017**, 31, (10), 990-1006.
122. Gajos, M.; Jasnovidova, O.; van Bommel, A.; Freier, S.; Vingron, M.; Mayer, A., Conserved DNA sequence features underlie pervasive RNA polymerase pausing. *Nucleic Acids Res* **2021**, 49, (8), 4402-4420.
123. Ehara, H.; Kujirai, T.; Shirouzu, M.; Kurumizaka, H.; Sekine, S. I., Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT. *Science* **2022**, 377, (6611), eabp9466.

124. Chernukhin, I.; Shamsuddin, S.; Kang, S. Y.; Bergstrom, R.; Kwon, Y. W.; Yu, W.; Whitehead, J.; Mukhopadhyay, R.; Docquier, F.; Farrar, D.; Morrison, I.; Vigneron, M.; Wu, S. Y.; Chiang, C. M.; Loukinov, D.; Lobanenko, V.; Ohlsson, R.; Klenova, E., CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol Cell Biol* **2007**, *27*, (5), 1631-48.
125. Gomes, N. P.; Espinosa, J. M., Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding. *Genes Dev* **2010**, *24*, (10), 1022-34.
126. Nanavaty, V.; Abrash, E. W.; Hong, C.; Park, S.; Fink, E. E.; Li, Z.; Sweet, T. J.; Bhasin, J. M.; Singuri, S.; Lee, B. H.; Hwang, T. H.; Ting, A. H., DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex. *Mol Cell* **2020**, *78*, (4), 752-764 e6.
127. Alharbi, A. B.; Schmitz, U.; Bailey, C. G.; Rasko, J. E. J., CTCF as a regulator of alternative splicing: new tricks for an old player. *Nucleic Acids Res* **2021**, *49*, (14), 7825-7838.
128. Shukla, S.; Kavak, E.; Gregory, M.; Imashimizu, M.; Shutinoski, B.; Kashlev, M.; Oberdoerffer, P.; Sandberg, R.; Oberdoerffer, S., CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **2011**, *479*, (7371), 74-9.
129. Marina, R. J.; Sturgill, D.; Bailly, M. A.; Thenoz, M.; Varma, G.; Prigge, M. F.; Nanan, K. K.; Shukla, S.; Haque, N.; Oberdoerffer, S., TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. *Embo J* **2016**, *35*, (3), 335-55.
130. Mao, S. Q.; Ghanbarian, A. T.; Spiegel, J.; Martinez Cuesta, S.; Beraldi, D.; Di Antonio, M.; Marsico, G.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S., DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol* **2018**, *25*, (10), 951-957.
131. He, X.; Yuan, J.; Gao, Z.; Wang, Y., Promoter R-Loops Recruit U2AF1 to Modulate Its Phase Separation and RNA Splicing. *J Am Chem Soc* **2023**, *145*, (39), 21646-21660.
132. Lamas-Maceiras, M.; Singh, B. N.; Hampsey, M.; Freire-Picos, M. A., Promoter-Terminator Gene Loops Affect Alternative 3'-End Processing in Yeast. *J Biol Chem* **2016**, *291*, (17), 8960-8.
133. Tan-Wong, S. M.; Zaugg, J. B.; Camblong, J.; Xu, Z.; Zhang, D. W.; Mischo, H. E.; Ansari, A. Z.; Luscombe, N. M.; Steinmetz, L. M.; Proudfoot, N. J., Gene loops enhance transcriptional directionality. *Science* **2012**, *338*, (6107), 671-5.
134. Guo, Y.; Monahan, K.; Wu, H.; Gertz, J.; Varley, K. E.; Li, W.; Myers, R. M.; Maniatis, T.; Wu, Q., CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. *Proc Natl Acad Sci U S A* **2012**, *109*, (51), 21081-6.
135. Monahan, K.; Rudnick, N. D.; Kehayova, P. D.; Pauli, F.; Newberry, K. M.; Myers, R. M.; Maniatis, T., Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of protocadherin-alpha gene expression. *Proc Natl Acad Sci U S A* **2012**, *109*, (23), 9125-30.
136. Leonaite, B.; Han, Z.; Basquin, J.; Bonneau, F.; Libri, D.; Porrua, O.; Conti, E., Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. *Embo J* **2017**, *36*, (11), 1590-1604.
137. Lansdorp, P.; van Wietmarschen, N., Helicases FANCI, RTEL1 and BLM Act on Guanine Quadruplex DNA in Vivo. *Genes (Basel)* **2019**, *10*, (11).
138. Nguyen, H. D.; Yadav, T.; Giri, S.; Saez, B.; Graubert, T. A.; Zou, L., Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. *Mol Cell* **2017**, *65*, (5), 832-847 e4.
139. Yan, Q.; Wulfridge, P.; Doherty, J.; Fernandez-Luna, J. L.; Real, P. J.; Tang, H. Y.; Sarma, K., Proximity labeling identifies a repertoire of site-specific R-loop modulators. *Nat Commun* **2022**, *13*, (1), 53.
140. Mestre-Fos, S.; Penev, P. I.; Suttapitugsakul, S.; Hu, M.; Ito, C.; Petrov, A. S.; Wartell, R. M.; Wu, R.; Williams, L. D., G-Quadruplexes in Human Ribosomal RNA. *J Mol Biol* **2019**, *431*, (10), 1940-1955.
141. Bourdon, S.; Herviou, P.; Dumas, L.; Destefanis, E.; Zen, A.; Cammas, A.; Millevoi, S.; Dassi, E., QUADRAtlas: the RNA G-quadruplex and RG4-binding proteins database. *Nucleic Acids Res* **2023**, *51*, (D1), D240-D247.
142. Scognamiglio, P. L.; Di Natale, C.; Leone, M.; Poletto, M.; Vitagliano, L.; Tell, G.; Marasco, D., G-quadruplex DNA recognition by nucleophosmin: new insights from protein dissection. *Biochim Biophys Acta* **2014**, *1840*, (6), 2050-9.
143. Okuwaki, M.; Saotome-Nakamura, A.; Yoshimura, M.; Saito, S.; Hirawake-Mogi, H.; Sekiya, T.; Nagata, K., RNA-recognition motifs and glycine and arginine-rich region cooperatively regulate the nucleolar localization of nucleolin. *J Biochem* **2021**, *169*, (1), 87-100.
144. Santos, T.; Salgado, G. F.; Cabrita, E. J.; Cruz, C., Nucleolin: a binding partner of G-quadruplex structures. *Trends Cell Biol* **2022**, *32*, (7), 561-564.
145. Tian, B.; Manley, J. L., Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **2017**, *18*, (1), 18-30.
146. Leppek, K.; Das, R.; Barna, M., Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* **2018**, *19*, (3), 158-174.
147. Schuster, S. L.; Hsieh, A. C., The Untranslated Regions of mRNAs in Cancer. *Trends Cancer* **2019**, *5*, (4), 245-262.
148. Mayr, C., What Are 3' UTRs Doing? *Cold Spring Harb Perspect Biol* **2019**, *11*, (10).

149. Lee, D. S. M.; Ghanem, L. R.; Barash, Y., Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat Commun* **2020**, 11, (1), 527.
150. Sauer, M.; Juranek, S. A.; Marks, J.; De Magis, A.; Kazemier, H. G.; 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* **2019**, 10, (1), 2421.
151. Benhalevy, D.; Gupta, S. K.; Danan, C. H.; Ghosal, S.; Sun, H. W.; Kazemier, H. G.; Paeschke, K.; Hafner, M.; Juranek, S. A., The Human CCHC-type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation. *Cell Rep* **2017**, 18, (12), 2979-2990.
152. Dong, L.; Mao, Y.; Zhou, A.; Liu, X. M.; Zhou, J.; Wan, J.; Qian, S. B., Relaxed initiation pausing of ribosomes drives oncogenic translation. *Sci Adv* **2021**, 7, (8).
153. Zhou, J.; Wan, J.; Gao, X.; Zhang, X.; Jaffrey, S. R.; Qian, S. B., Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **2015**, 526, (7574), 591-4.
154. Zaccara, S.; Jaffrey, S. R., A Unified Model for the Function of YTHDF Proteins in Regulating m(6)A-Modified mRNA. *Cell* **2020**, 181, (7), 1582-1595 e18.
155. Czech, B.; Munafo, M.; Ciabrelli, F.; Eastwood, E. L.; Fabry, M. H.; Kneuss, E.; Hannon, G. J., piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet* **2018**, 52, 131-157.
156. Zyner, K. G.; Simeone, A.; Flynn, S. M.; Doyle, C.; Marsico, G.; Adhikari, S.; Portella, G.; Tannahill, D.; Balasubramanian, S., G-quadruplex DNA structures in human stem cells and differentiation. *Nat Commun* **2022**, 13, (1), 142.
157. Skourti-Stathaki, K.; Torlai Triglia, E.; Warburton, M.; Voigt, P.; Bird, A.; Pombo, A., R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Mol Cell* **2019**, 73, (5), 930-945 e4.
158. Yang, Q.; Lin, J.; Liu, M.; Li, R.; Tian, B.; Zhang, X.; Xu, B.; Liu, M.; Zhang, X.; Li, Y.; Shi, H.; Wu, L., Highly sensitive sequencing reveals dynamic modifications and activities of small RNAs in mouse oocytes and early embryos. *Sci Adv* **2016**, 2, (6), e1501482.
159. Zhang, Y.; Zhang, X.; Shi, J.; Tuorto, F.; Li, X.; Liu, Y.; Liebers, R.; Zhang, L.; Qu, Y.; Qian, J.; Pahima, M.; Liu, Y.; Yan, M.; Cao, Z.; Lei, X.; Cao, Y.; Peng, H.; Liu, S.; Wang, Y.; Zheng, H.; Woolsey, R.; Quilici, D.; Zhai, Q.; Li, L.; Zhou, T.; Yan, W.; Lyko, F.; Zhang, Y.; Zhou, Q.; Duan, E.; Chen, Q., Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat Cell Biol* **2018**, 20, (5), 535-540.
160. Paloviita, P.; Hyden-Granskog, C.; Yohannes, D. A.; Paluoja, P.; Kere, J.; Tapanainen, J. S.; Krjutskov, K.; Tuuri, T.; Vosa, U.; Vuoristo, S., Small RNA expression and miRNA modification dynamics in human oocytes and early embryos. *Genome Res* **2021**, 31, (8), 1474-1485.
161. Tomar, A.; Gomez-Velazquez, M.; Gerlini, R.; Comas-Armangué, G.; Makharadze, L.; Kolbe, T.; Boersma, A.; Dahlhoff, M.; Burgstaller, J. P.; Lassi, M.; Darr, J.; Toppari, J.; Virtanen, H.; Kuhnappel, A.; Scholz, M.; Landgraf, K.; Kiess, W.; Vogel, M.; Gailus-Durner, V.; Fuchs, H.; Marschall, S.; Hrabe de Angelis, M.; Kotaja, N.; Korner, A.; Teperino, R., Epigenetic inheritance of diet-induced and sperm-borne mitochondrial RNAs. *Nature* **2024**, 630, (8017), 720-727.
162. Maldonado, R.; Langst, G., The chromatin - triple helix connection. *Biol Chem* **2023**, 404, (11-12), 1037-1049.
163. Leisegang, M. S.; Warwick, T.; Stotzel, J.; Brandes, R. P., RNA-DNA triplexes: molecular mechanisms and functional relevance. *Trends Biochem Sci* **2024**, 49, (6), 532-544.
164. Zhou, Z.; Giles, K. E.; Felsenfeld, G., DNA:RNA triple helix formation can function as a cis-acting regulatory mechanism at the human beta-globin locus. *Proc Natl Acad Sci U S A* **2019**, 116, (13), 6130-6139.
165. Maldonado, R.; Schwartz, U.; Silberhorn, E.; Langst, G., Nucleosomes Stabilize ssRNA-dsDNA Triple Helices in Human Cells. *Mol Cell* **2019**, 73, (6), 1243-1254 e6.
166. Kohestani, H.; Wereszczynski, J., The effects of RNA:DNA-DNA triple helices on nucleosome structures and dynamics. *Biophys J* **2023**, 122, (7), 1229-1239.
167. Jimenez-Garcia, E.; Vaquero, A.; Espinas, M. L.; Soliva, R.; Orozco, M.; Bernues, J.; Azorin, F., The GAGA factor of Drosophila binds triple-stranded DNA. *J Biol Chem* **1998**, 273, (38), 24640-8.
168. Herbert, A., ALU non-B-DNA conformations, flipons, binary codes and evolution. *Royal Society Open Science* **2020**, 7, (6), 200222.
169. Leisegang, M. S.; Bains, J. K.; Serebinski, S.; Oo, J. A.; Krause, N. M.; Kuo, C. C.; Gunther, S.; Senturk Cetin, N.; Warwick, T.; Cao, C.; Boos, F.; Izquierdo Ponce, J.; Haydar, S.; Bednarz, R.; Valasarajan, C.; Fuhrmann, D. C.; Preussner, J.; Looso, M.; Pullamsetti, S. S.; Schulz, M. H.; Jonker, H. R. A.; Richter, C.; Rezende, F.; Gilsbach, R.; Pfluger-Muller, B.; Wittig, I.; Grummt, I.; Ribarska, T.; Costa, I. G.; Schwalbe, H.; Brandes, R. P., HIF1alpha-AS1 is a DNA:DNA:RNA triplex-forming lncRNA interacting with the HUSH complex. *Nat Commun* **2022**, 13, (1), 6563.
170. Herbert, A., *Flipons and the logic of soft-wired genomes*. 1st ed.; CRC Press: Boca Raton, 2024.

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.