Genome Biology

Open Access

High-throughput techniques enable advances in the roles of DNA and RNA secondary structures in transcriptional and post-transcriptional gene regulation

Ilias Georgakopoulos-Soares[1,2], Candace S. Y. Chan[1,2], Nadav Ahituv[1,2] and Martin Hemberg[3*]

*Correspondence:
mhemberg@bwh.harvard.edu

[1] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, USA
[2] Institute for Human Genetics, University of California San Francisco, San Francisco, California, USA
[3] Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA
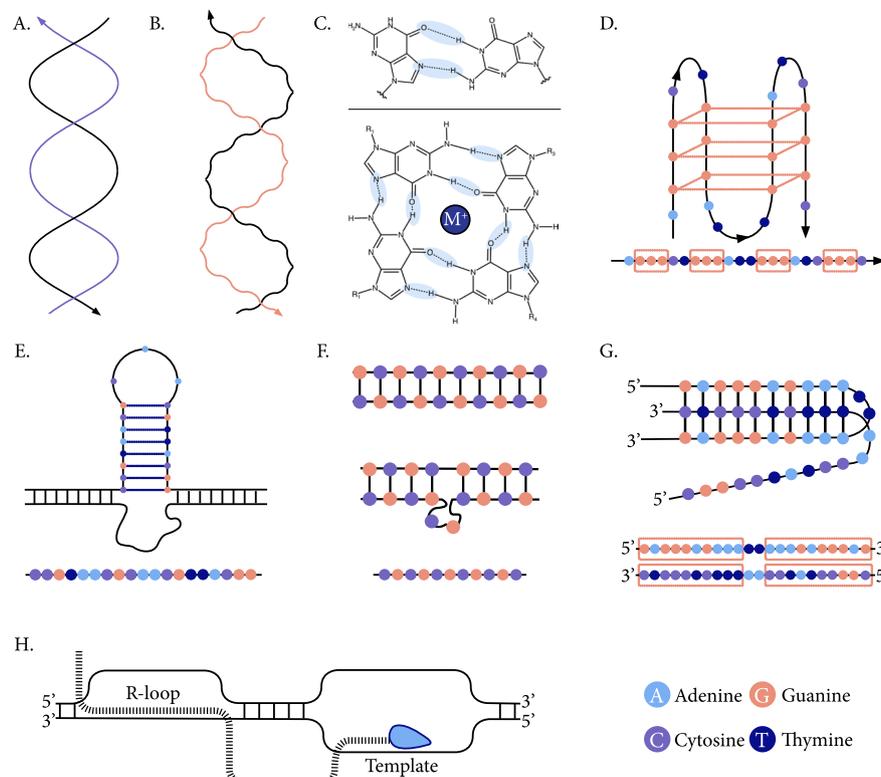
Abstract

The most stable structure of DNA is the canonical right-handed double helix termed B DNA. However, certain environments and sequence motifs favor alternative conformations, termed non-canonical secondary structures. The roles of DNA and RNA secondary structures in transcriptional regulation remain incompletely understood. However, advances in high-throughput assays have enabled genome wide characterization of some secondary structures. Here, we describe their regulatory functions in promoters and 3'UTRs, providing insights into key mechanisms through which they regulate gene expression. We discuss their implication in human disease, and how advances in molecular technologies and emerging high-throughput experimental methods could provide additional insights.

Introduction

The canonical conformation of the DNA molecule is a double helix which under physiological conditions forms a stable structure known as B DNA. However, certain environments and sequence motifs favor other DNA conformations, known as non-canonical secondary structures. Moreover, many of the structures are also present in RNA molecules. These structures include perfect and imperfect hairpins, cruciforms, slipped structures, R-loops, G-quadruplexes, i-motifs, Z-DNA, Z-RNA, triple-stranded DNA, RNA and hybrid structures. Sequences that are predisposed to secondary structure formation are enriched at regulatory regions, including open chromatin regions, promoters, 5'UTRs and 3'UTRs [1, 2] (Fig. 1). In particular, they are over-represented and positioned relative to key gene features, such as transcription start and transcription end sites, splice junctions and translation initiation regions, while their formation is associated with transcriptionally active loci [4–11]. Thus, secondary structures can have a functional impact since promoter regions

Fig. 1 Schematics of DNA and RNA structures. **A** The canonical right handed double helix, also known as B DNA secondary structure. **B** Z-DNA forms a left-handed double helix. **C** G-quadruplexes are formed by the stacking of multiple G-quartets held together by Hoogsteen hydrogen bonds (top). Four guanines establish hydrogen bonds with each other to form a G-quartet (bottom). Hoogsteen hydrogen bonds are highlighted in blue. The monovalent cation that can stabilize the G-quadruplex structure is marked with M. **D** Hairpins are formed at inverted repeats, in which the stem base pairs hybridize with hydrogen bonds, while the loop remains single-stranded. **E** Slipped-strand mispairing at tandem repeats results in slipped structure formation. **F** Depiction of a homopurine-homopyrimidine sequence with mirror symmetry. H-DNA is a triple helix secondary structure where the third strand hybridizes with Hoogsteen hydrogen bonds with the duplex DNA, while the fourth strand remains single stranded. **G** R-loops are formed co-transcriptionally at the template strand. The nascent RNA produced by the RNA-polymerase hybridizes with the template strand to form an R-loop structure, while the non-template strand remains single-stranded

control transcription initiation, while 3'UTRs have a number of functions, including impacting the stability of the transcript and its rate of degradation, providing binding sites for regulatory elements such as miRNAs and RNA binding proteins (RBPs), and containing signals for the localisation of the transcript in the cell.

Non-canonical secondary structures often impact gene expression at the DNA and RNA levels, having key roles in gene regulation. Nevertheless, the role of DNA and RNA secondary structures in transcriptional and post-transcriptional regulation remains incompletely understood. Incorporating these effects in models of gene regulation has the potential to improve our understanding of how cells fine-tune gene expression and enable the development of novel therapies. Thanks to technological advances and novel computational and experimental methods several powerful tools have recently become available that have the potential to fundamentally alter our understanding of the role of secondary structures in gene regulation. At the same

Georgakopoulos-Soares *et al. Genome Biology*   (2022) 23:159

Page 3 of 22

time, DNA and RNA secondary structure formation has been related to a number of diseases and are emerging as novel therapeutic targets [12–16]. In this review, we summarize recent advances that are changing our understanding regarding the roles of DNA and RNA secondary structures in promoters and 3'UTRs with a particular focus on transcriptional and post-transcriptional gene regulation and discuss emerging prospects, challenges and therapeutic opportunities.

Non-canonical DNA and RNA secondary structures

The characterization of the DNA secondary structure in a landmark paper [17] provided fundamental insights into how genetic information is stored and used. Double stranded DNA is composed of two helices held together by hydrogen bonds, and most often adopts the canonical right-handed double helical secondary structure, also known as B-DNA (Fig. 2a), with 10.5 residues per turn [18]. Similarly to B-DNA, A-DNA is also right-handed and double helical albeit with 11 residues per turn and forms more readily at GC-rich regions [19, 20]. Larger distortions in the DNA structure occur at sequences that are predisposed to alternative conformations and are collectively termed non-B DNA, encompassing multifarious DNA secondary structures [21]. Although several DNA and RNA secondary structures are shared, the two molecules sometimes substantially differ in their thermodynamic stability and likelihood of secondary structure formation. The single strandedness of RNA molecules enables long range base pairing interactions, while physical constraints in the DNA molecule restrict such interactions to directly adjacent sequences. Below we describe the DNA and RNA secondary structures that are discussed in this review.

Z-DNA is a left-handed double helical structure (Fig. 2b) that is formed primarily at regions with alternating purine pyrimidine tracts, particularly at GC repeats [22–24]. Z-DNA is less energetically favorable than B DNA under physiological conditions, and consequently it requires negative supercoiling. Z-DNA formation has been associated with active transcription [25], supporting the correlation between Z-DNA levels and transcription observed in the past [26].

G-quadruplexes are nucleic acid structures held together with Hoogsteen hydrogen bonds between guanines that form stacked G-tetrads (Fig. 2c). Hoogsteen base pairing refers to non-Watson–Crick base pairing as shown in Fig. 2c. These bonds connect four guanines forming a square planar arrangement which is known as a G-quartet. Stacking multiple G-quartets results in the formation of a G-quadruplex. G-quadruplex formation is driven by the inherent propensity of guanines to self-assemble in
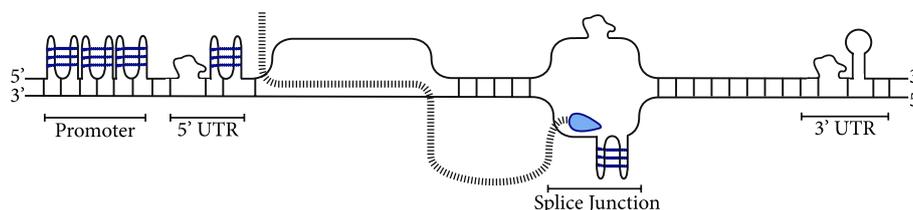
Fig. 2 Schematic overview of Non-B DNA enrichment relative to gene features. Higher density of non-B DNA structures is observed at promoter regions, 5'UTRs, regions flanking splice sites and at the 3'UTR. Formation of secondary structures is also facilitated by negative supercoiling and at actively transcribed regions relative to the direction of the transcribing RNA polymerase [2, 3]

the presence of monovalent cations into planar structures [27]. Traditionally, they are classified as parallel, antiparallel or hybrid depending on the folding topology [28]. However, recent work has shown the existence of additional, more complex arrangements [29–31].

DNA and RNA hairpins can form at perfect and imperfect inverted repeats. Inverted repeats are composed of two adjacent copies of the same sequence, one of which is found in the reverse complement orientation (Fig. 2d). A hairpin is held together by hydrogen bonds between the two complementary arms, while the spacer remains single stranded. A closely related structure is cruciforms, consisting of two hairpins and a 4-way junction [32]. It has been shown that sequence properties of inverted repeats, including spacer and arm length, interruptions and nucleotide composition can affect the folding kinetics and stability of hairpins and cruciforms. Inverted repeats with an arm length of >6 nts have been shown to form *in vivo* in *Saccharomyces cerevisiae* [33]. Hairpin formation dynamics have been studied in detail by varying the spacer lengths, arm lengths, and nucleotide composition, as well as by examining the folding and mutagenic potential [33–36]. For instance, arms with higher GC content display more stable hairpin formation [36].

Slipped structures form at consecutive repeat sequences, in which one repeat unit misaligns with the second repeat unit on the opposite strand [37] (Fig. 2e). Direct and short tandem repeats can form these structures, and due to inefficient repair they are often expanded or contracted [38]. There are over one million tandem repeats in the human genome, and many are polymorphic [16].

Intramolecular triple stranded DNA (H-DNA) forms at homopurine-homopyrimidine stretches that contain a mirror symmetry [39, 40]. One strand folds back, joins the double-stranded DNA, and is held with Hoogsteen or reverse-Hoogsteen bonds. The other strand remains single, and the result is a triple helical structure (Fig. 2f). Intermolecular triplexes can occur between DNA molecules, RNA molecules, or as a hybrid involving a DNA and a RNA molecule [41].

During transcription, dynamic hybrid structures between DNA and nascent RNA transcripts can be formed [42, 43]. One example is R-loops, where an RNA molecule invades and pairs up with one DNA strand, while displacing the other (Fig. 2g). Formation and stabilization of R loops is particularly favorable when the non-template strand is G-rich, and it can also be promoted by DNA supercoiling, the presence of DNA nicks, and the formation of G-quartets [44, 45]. The continuous activity of DNA/RNA helicases and ribonucleases H (RNAse H1 and H2) maintain R-loop formation at low levels [45]. Interestingly, R-loops and G-quadruplexes were both found to be unwound by the helicase DHX9 [46]. This helicase activity is important to avoid single-stranded DNA damage and to preserve genomic stability. It has been shown that R-loops can occupy ~5% of the human genome, and that they are depleted in intergenic regions relative to genic regions. However, R-loops are highly dynamic and it has been estimated that <10% of loops are formed at any point in time [47, 48]. Formation of R-loops shows a remarkable strand asymmetry, with more than 90% of R-loops occurring co-linearly with transcription [48].

High-throughput techniques to identify non-B DNA structures

A number of biophysical and biochemical methods as well as highly sensitive and specific assays have been developed to study the formation of DNA and RNA structures *in vitro* [49]. However, *in vivo* identification of secondary structures, even in cell cultures, has remained more challenging. Recently, permanganate footprinting was combined with genome-wide sequencing to identify transiently formed single stranded DNA regions [2], providing a readout for genome-wide non-B DNA formation. Furthermore, antibodies with high affinity and specificity have been developed targeting specific secondary structures and enabling their visualization [50–53], while other methods use the cleavage of DNA-RNA hybrids by specific nucleases to map secondary structures or nucleotide hybridization within RNA molecules to determine their folding [2, 54, 55].

Genome-wide maps of sequences that can form G-quadruplexes *in vitro* under favorable conditions have been generated using a modified sequencing method that stalls at G-quadruplexes [29, 56]. This technique has been termed G4-seq. Similarly, rG4-seq, which is a variant of G4-seq, has enabled transcriptome-wide identification of RNA G-quadruplexes [57]. Recently, a novel method called G4-miner was used for genome-wide profiling of G-quadruplexes from standard whole-genome sequencing based on deviations in sequencing quality, with drop in sequencing quality at G-quadruplex structures [58]. Antibodies with high affinity for G-quadruplexes have been used for G4 ChIP-seq experiments to demonstrate genome-wide G-quadruplex structure formation in human cells [1, 59]. Crucially, the number of loci discovered forming G-quadruplexes by ChIP-seq *in vivo* is ~10,000 while G4-seq *in vitro* identifies ~700,000 peaks [29, 56, 59]. This discrepancy could be the result of rapid resolution of G-quadruplexes by helicases in the cellular environment [60], it could reflect differences in the genomic locality, associated with epigenetic changes and chromatin accessibility [61] and could be biased by the chemical perturbation with K^+ or Pyridostatin (PDS) in G4-seq experiments. In addition, in ChIP-seq this discrepancy could also be explained by the antibody being able to recognize only certain G-quadruplexes. Antibody-associated differences are indeed observed between BG4 and D1 antibodies [1, 62]. Large variability in the number of peaks has also been observed in G4-seq experiments where the G-quadruplex stabilization method seems to influence the results with substantial differences in PDS and K^+ treatments. PDS is a small molecule compound that has been shown to bind to G-quadruplexes, leading to their stabilization [63]. Similarly, K^+ ions interact with G-quadruplexes and stabilize them. High-throughput G-quadruplex detection methods identify sites that do not conform to the consensus G-quadruplex motif. However, these high-throughput methods entail certain limitations. These include the inability to discern the G-quadruplex forming potential between different cell types or to measure temporal effects; they do not provide information about the kinetics and thermal stability of G-quadruplexes at individual loci. More recently, single molecule G-quadruplex tracking in living cells, using a G-quadruplex specific fluorescent probe, has indicated dynamic fluctuations between folded and unfolded states [64].

Z-DNA binding proteins have a plethora of biological functions, and several proteins that have a Z-DNA binding domain (Zα) have been identified [65–67]. Isolation of proteins that bind preferentially to Z-DNA over B-DNA identified ADAR1 [68] and the specificity has been used in ChIP-seq experiments to identify sites of Z-DNA

formation in the human genome [68]. Methods have also been developed to identify DNA:RNA structures from cells such as isolation of DNA-associated RNA coupled with high-throughput sequencing [69]. *In vitro* R-loop identification was initially potentiated through the development of an antibody with high specificity to its secondary structure [50, 70, 71]. Advances in molecular technologies have resulted in numerous variants of the S9.6 RNA:DNA hybrid antibody, and by combining them with nucleases that cleave DNA:RNA hybrids systematic studies of R-loops have been made possible [70, 71]. R-ChIP was the first method that mapped genome-wide R-loops using catalytically dead RNase H coupled with by amplification of immunoprecipitated DNA [72]. The development of DNA-RNA immunoprecipitation sequencing (DRIP-seq) has further enabled genome-wide profiling of R-loops and the identification of genomic sites that form R-loops with higher propensity [70, 73], while recent antibody-independent nuclease-based methods, include MapR and BisMapR, which provide high-resolution genome-wide detection of R-loops [55, 74].

Secondary structures are involved in transcriptional regulation at promoters

Non-canonical DNA secondary structures at the promoter play an important role across a range of processes. First, DNA secondary structures can act as landing pads for certain transcription factors [25, 75–80], many of which show preferential binding relative to B DNA (Fig. 3a, b). Second, the formation of non-canonical DNA secondary structures can act as a physical barrier for nucleosome formation [48, 91, 99, 100], thereby promoting accessibility (Fig. 3c). Third, secondary structures may influence genome organization and long-range DNA looping [94–97] (Fig. 3d). Fourth, regions with high propensity of forming secondary structures are associated with RNAPII pausing [72, 81, 89], a phenomenon important for several regulatory processes including promoter-proximal pausing, exon recognition, splicing and transcription termination (Fig. 3e, f).

Arguably the most studied non-canonical DNA secondary structure in promoter regions is the G-quadruplex. Estimates using the consensus G-quadruplex motif show that >50% of extended promoter regions harbor at least one G-quadruplex motif [101]. The orientation of G-quadruplexes relative to transcription direction is skewed with a higher frequency at the template strand upstream of the TSS and at the non-template strand downstream [102, 103]. In addition, TATA-less promoters have a substantially higher G-quadruplex density, both near the transcription start site and in the broader promoter region [104]. Moreover, genome-wide G4-seq experiments have demonstrated that a large proportion of G-quadruplexes do not adhere to the consensus motif, with inter-molecular G-quadruplexes as well as structures with bulges, disruptions and longer loops being over-represented in promoter regions [29]. The enrichment of G-quadruplexes in promoters relative to other sites in the genome was also recapitulated in G4 ChIP-seq experiments. Again, it was found that only 21% of G-quadruplex peaks contain a consensus motif [1], indicating a high G-quadruplex sequence diversity.

Additionally, epigenetic modifications influence the likelihood of G-quadruplex formation and stability; open chromatin regions and highly transcribed genes are enriched for G-quadruplex structures [1, 105], while cytosine methylation of G-quadruplexes increases stability [106, 107]. Conversely, G-quadruplexes are enriched at CpG sites and their formation is associated with CpG island hypomethylation through inhibition
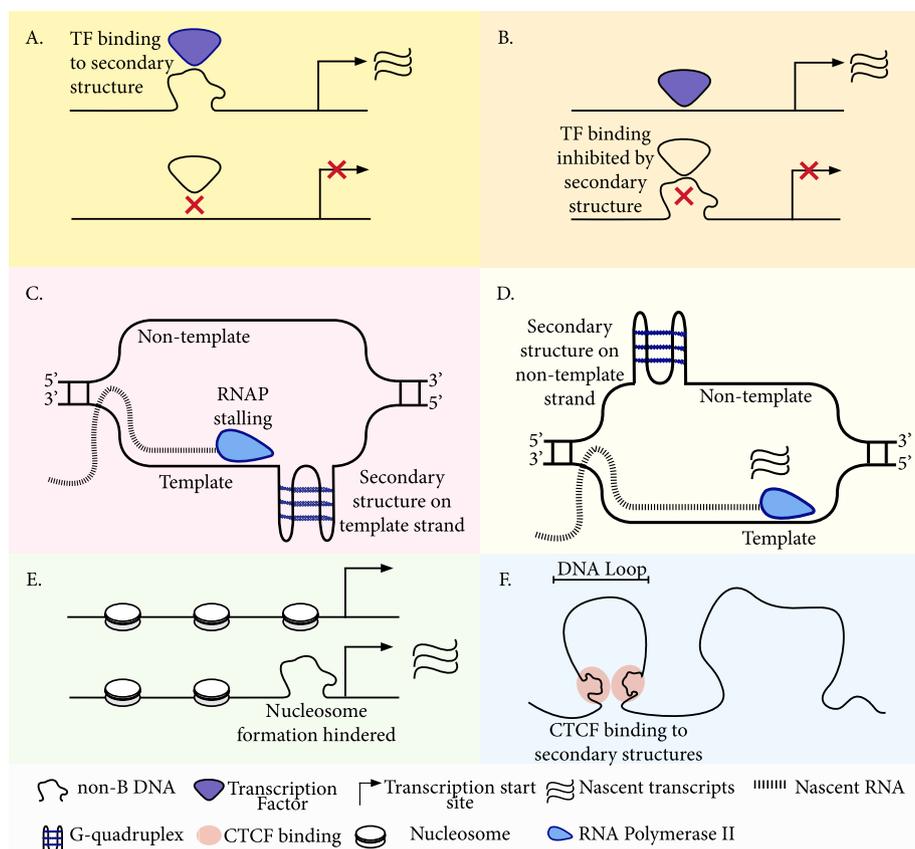
Fig. 3 Functional roles of secondary structures in promoter regulation. **a** Non-B DNA structure formation can mediate transcription factor binding and promote transcription [81–86]. **b** Non-B DNA structure formation can hinder transcription factor binding to inhibit transcription [87, 88]. **c** RNAP stalling at the G-quadruplex structure formed at the template strand [81, 89]. **d** G-quadruplex structure formation at the non-template strand enables the template strand to remain single-stranded and promotes expression [90]. **e** Non-B DNA structure formation can hinder nucleosome formation to promote transcription [1, 91–93]. **f** Non-B DNA structures can promote long range DNA looping [94–98]

of DNA methyltransferase 1 enzymatic activity [108]. The epigenetic guanine conversion to 8-oxo-7,8-dihydroguanine through DNA damage has also been associated with increased G-quadruplex formation in promoters [109]. Topologically associated domains (TADs) represent self-interacting genomic sites, and G-quadruplexes are enriched at TAD boundaries [96, 97] where they interact with both CTCF [97] and Yin and Yang 1 (YY1) [95] to facilitate long-range DNA looping (Fig. 3f). In particular, YY1 binds directly to the G-quadruplex structure and stabilization of the G-quadruplexes results in coordination of the genes in the same DNA loop [95].

The relationship between G-quadruplex formation at promoters and gene expression levels varies, and likely depends on several factors, including strand orientation, positioning relative to TSS, biophysical properties of the G-quadruplex, presence of transcription factor binding sites in the vicinity of the G-quadruplex, and epigenetic marks. Thus, it has been shown that G-quadruplexes can both promote [110–112] and inhibit expression [4, 113]. However, since G-quadruplexes are landing pads for a number of proteins (Fig. 3b, c), e.g. SP1 [77], NM23-H2, CNBP25 and Nucleolin [114], it is difficult

Georgakopoulos-Soares *et al. Genome Biology* (2022) 23:159

Page 8 of 22

to estimate their contribution towards gene expression [115, 116]. Indeed, multiple transcription factor binding sites have been found to coincide with G-quadruplexes significantly more often than expected by chance. A recent high-throughput massively parallel reporter assay (MPRA) study examined the contribution of promoter G-quadruplexes and reported a positive correlation between their presence and expression levels. However, after accounting for GC-content differences, G-quadruplexes were either not significantly associated with expression levels or were associated with reduced gene expression [117]. The study also identified a bias with G-quadruplexes at the template strand resulting in lower expression levels. These results suggest that nucleotide composition is a major confounder in understanding the contribution of G-quadruplexes to gene expression.

Z-DNA sequences are enriched in promoters upstream of the TSS [118, 119], where they can act as nucleosome boundary elements to promote open chromatin [91, 120] (Fig. 3c). However, there is conflicting evidence regarding their impact on gene expression levels. It has been reported that Z-DNA forming in the first exon of *ADAM12* acts as a repressor [8, 121]. By contrast, studies in yeast have shown that Z-DNA can serve as activators [122]. Similarly, it has been found that formation of Z-DNA at the *CSF1* gene stabilizes the open chromatin and aids the recruitment of RNA polymerase [123], while in the promoter region of *HO-1* formation of Z-DNA precedes the recruitment of RNAPII, resulting in transcriptional activation [124]. The latter examples are supported by a recent MPRA where Z-DNA was shown to be a positive regulator of gene expression [117].

R-loops are enriched at promoters [70], with a 2-fold enrichment over background levels [48], and transcription is positively correlated with their frequency at promoter regions [125, 126]. R-loops can be stabilized at G-quadruplex sites [127] and R-loop formation is more frequent at CpG-island containing promoters where they increase expression through reduced methylation levels [70]. Formation of R-loops in promoters can facilitate histone modifications and is associated with open chromatin regions and GC-skew [43, 48, 128, 129]. The DNA-RNA helicase Senataxin [130], the RNA helicase DHX9 [131], the RNases H1 and H2 [132, 133] and topoisomerases inhibit the formation or cleave R-loops, while their accumulation can cause DNA damage, genome instability and is deleterious [45]. Roles for R-loops in long non-coding RNA (lncRNA) transcription have also been shown, in which R-loop formation can induce lncRNA generation which in turn act as transcriptional inducers [134, 135]. Specifically, at promoters R-loop formation can generate antisense lncRNAs, and their removal results in the inhibition of antisense lncRNAs [135]. R-loops have been found to be important for lncRNA function, for example the HOTTIP lncRNA recruits CTCF and cohesin, interacts with R-loop associated proteins and induces the formation of R-loops at TAD boundaries, which in turn regulates and reinforces those boundaries [98]. In *Arabidopsis thaliana*, the lncRNA APOLO recognizes distal targets via R-loops to epigenetically silence them [136]. There is also accumulating evidence for roles of R-loops in promoter-proximal RNAPII pausing [72, 137–139]. Transcription perturbation experiments have provided evidence for a strong link between R-loop induction and RNAPII pausing near transcription start sites [72]. Another line of evidence comes from a study which showed that BRCA1 can resolve R-loops at promoter-proximal RNAPII pausing sites [139], while in

Georgakopoulos-Soares *et al. Genome Biology* (2022) 23:159

Page 9 of 22

BRCA1 mutant cells, R-loops accumulate at the 5′ end of genes resulting in promoter-proximal RNAPII pausing [139].

Triple helix structures (single-strand RNA hybridizing to double-strand DNA with Hoogsteen bonds) have been shown to contribute towards genome organization [94], and some of the most thoroughly studied lncRNAs, e.g. HOTAIR, MEG3 and PARTI-CLE, form triple stranded DNA-RNA hybrids to perform epigenetic modifications and to regulate gene expression [140–142]. Short tandem repeats are enriched in promoters [143, 144] and enhancers [145] and have multiple regulatory roles, e.g. enabling the formation of slipped structures, G-quadruplexes and R-loops, creating additional or destroying existing transcription factor binding sites [146–150], altering DNA methylation [151, 152], and influencing nucleosome positioning [153]. Polymorphic short tandem repeats are estimated to account for 10-15% of the variance in gene expression [146]. In particular, 10-20% of eukaryotic genes and promoters contain an unstable repeat tract [154], and changes in repeat length at short tandem repeats are causal expression quantitative trait loci (eQTLs) [155]. In yeast, up to a quarter of promoters contain a highly variable tandem repeat sequence that affects gene expression [156]. Moreover, a study of 17 human tissues identified ~28,000 short tandem repeats for which the number of repeat units was associated with the expression of nearby genes [78].

Secondary structures at the 3'UTR in transcriptional and post-transcriptional regulation

The stability of an mRNA transcript and its rate of degradation are major contributors to expression levels. Perhaps the most important determinant of mRNA half-life is the 3'UTR which contains polyadenylation signals, binding sites for RBPs, and miRNA target sites. Multiple alternative polyadenylation signals are found in the majority of 3'UTRs [157], which can alter the length of the 3'UTR, influencing mRNA structure, stability and translation efficiency [158]. Moreover, the folding of the 3'UTR can influence the maturation, localization and metabolism of the transcript [159–161]. Although highly structured mRNA molecules are less stable [159], studying the role of RNA structure is complicated by the fact that 3'UTR regions in humans are more structured *in vivo* than *in vitro* [162].

Secondary structure formation within the 3'UTR impacts expression levels in a variety of ways. For instance, although longer distance between polyadenylation signals and polyadenylation [poly(A)] sites is associated with a shorter RNA half life, secondary structures may juxtapose the poly(A) signals and the polyadenylation sequences [160] (Fig. 4a). The formation of hairpin structures impacts mRNA stability distinctly from the AU-rich sequences [168, 175]. A recent MPRA study quantified the contribution of secondary structures found in the 3'UTR, demonstrating their contribution alongside RBPs and miRNAs [176].

G-quadruplexes have a 2-fold enrichment in human 3'UTRs, and they are 1.28-fold more frequent at the template than the non-template strand [167, 177]. Overall ~15% of these 3'UTRs harbor at least one G-quadruplex motif, but the enrichment is ~30% for neuronal mRNAs that locate to dendrites. For instance, G-quadruplexes at the 3'UTR of two post-synaptic proteins, PSD-95 and CaMKIIa, are necessary and sufficient for
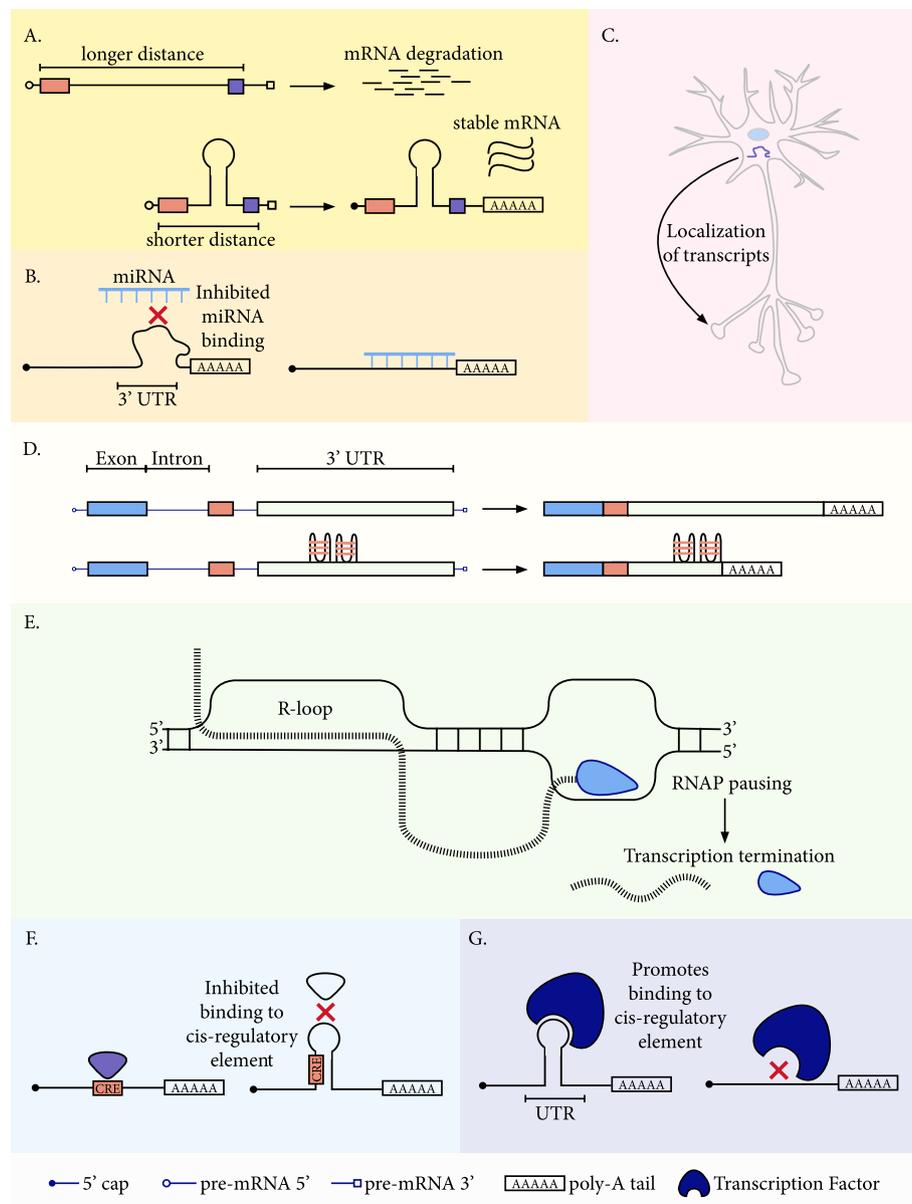
Fig. 4 Functional roles of secondary structures in transcription termination. **a** Juxtaposition of the poly(A) signal and the polyadenylation sequence [160, 163]. Formation of a hairpin structure brings the poly(A) signal in closer proximity to the polyadenylation sequence, resulting in higher expression levels. **b** Modulation of miRNA binding [164–166]. **c** Formation of the secondary structure enables transcript localization [161], **d.** Modulation of alternative polyadenylation by structure formation [167]. **e** R-loop formation at the 3'UTR promotes transcription termination. **f** Concealment of cis-regulatory elements [168, 169], **g** Exposure of cis-regulatory elements [170–174]

their localization in dendrites [161] (Fig. 4b). Studies have also found roles for G-quadruplexes in miRNA binding. For instance, in the FADS2 mRNA/mir331-3p pair, the G-quadruplex at the 3'UTR prevented the binding of the miRNA [164]. 3'UTRs most often contain alternative polyadenylation signals which can alter the regulation, stability and localization [178] (Fig. 4c). The G-quadruplexes at the 3'UTR of two genes, *LRP5*

and *FXR1*, were tested with luciferase assay experiments and were found to increase the efficiency of alternative polyadenylation site usage and act as *cis*-regulatory elements that alter miRNA regulation [167] (Fig. 4d). In the lncRNA MALAT1, three G-quadruplexes in its 3' end form stable RNA structures that can interact with proteins such as nucleolin and nucleophosmin in HeLa cells, and the G-quadruplexes are crucial for the localization of these proteins to nuclear speckles [179] (Fig. 4b, d).

Interestingly, within 100 base pairs downstream of the transcription end site there is a depletion of G-quadruplexes at the template strand as well as a profound enrichment at the non-template strand. Since the downstream enrichment is pronounced when there are neighboring genes in close proximity, it has been proposed that G-quadruplexes have a role in transcription termination [177]. In p53, a RNA G-quadruplex downstream of the gene controls pre-mRNA 3'-end processing regulation and is involved in the dynamic response to DNA damage [180]. The functional relevance of G-quadruplexes in 3'UTRs is also evidenced by the fact that they are selectively constrained and enriched for eQTLs, RBP sites and disease-associated variants. In a study of 150 RBPs, 15 RBPs were found to bind more often than expected by chance at G-quadruplex sites [181] (Fig. 4d, e).

R-loops are important for transcription termination, and they are over-represented at gene terminal regions, with a 3-fold enrichment [48]. The site of transcription termination is a major site of RNAPII pausing [182], and the R-loop formation enables the pausing and efficient transcription termination (Fig. 4d). Senataxin-deficient cells display transcription initiation, elongation and termination defects and increased RNAPII density downstream of the poly(A) signal. In particular, Senataxin resolves R-loop structures, enabling XRN2-mediated 3' transcript degradation and RNAPII termination [183]. BRCA1 mediates the recruitment of Senataxin at R-loop sites, and in its absence there is increased mutagenesis at those sites [184]. Therefore, R-loops are needed for efficient transcription termination, but once formed they need to be resolved by Senataxin and BRCA1.

Hairpin structures at the 3'UTR of transcripts can modulate expression, mRNA stability and localization. They can also act as responsive elements to environment changes [185], and can conceal or expose cis-regulatory elements [165, 186] (Fig. 4e, f). For instance, the iron-responsive element is a hairpin structure found in 5'UTRs and 3'UTRs, which interacts with two iron regulatory proteins, IRP1 and IRP2 for cellular iron homeostasis [187]. Another example is the perinuclear localization of c-myc mRNA, which is controlled by a 3'UTR hairpin structure [188]. The constitutive decay element forms a hairpin at the 3'UTR of TNF-alpha [189], which Roquin and Roquin2 proteins can bind to promote mRNA decay. In addition, Roquin binds at the 3'UTR at hairpin RNA structures to mediate mRNA deadenylation [172]. Recent MPRA work, in which the biophysical properties of the hairpin of the constitutive decay element were altered, provided insights into the hairpin's role in transcript levels and degradation rate [176, 190].

Non-canonical secondary structures in disease

Mutations in the human genome are not distributed homogeneously. Sequences that are predisposed to secondary structure formation are mutational hotspots and their instability has been associated with the development of several diseases, including multiple

Georgakopoulos-Soares *et al. Genome Biology* (2022) 23:159

Page 12 of 22

neurological disorders [191, 192] and cancer [193, 194]. Therefore, advances in our understanding of the function of these sequences could have important implications in understanding cancer development, improve the etiology of other diseases and facilitate modeling evolution with higher precision.

Longer inverted repeats are more prone to mutagenesis than their shorter counterparts [195]. Negative supercoiling during transcription aids the formation of cruciforms and hairpins which in turn impact transcription and transcription factor binding [196–198]. At promoters, recurrent mutagenesis at inverted repeats has been observed across cancer types [199–201], which has been attributed to APOBEC off-target mutagenesis [201]. However, it remains unclear whether these recurrent mutations have a role in tumor progression [194]. Analysis of an inverted repeat at the *PLEKHS1* promoter that is recurrently mutated across disparate cancer types did not find reproducible evidence for changes in its expression levels [202]. This result contrasts with the findings for the TERT promoter G-quadruplex, which coincides with driver mutations.

Developmental genes and oncogenes [203], including *CKIT*, *KRAS*, *CMYC* and *BCL2*, are enriched for the presence of G-quadruplexes in their promoters. In particular, the promoter of the oncogene *CMYC* was one of the first where a role for G-quadruplexes in expression modulation was demonstrated. A pioneering study showed that both mutations that disrupt the structure formation and a stabilizing small molecule compound substantially alter expression levels in opposite directions [4]. Similarly, the *BCL2* promoter contains a G-quadruplex, whose stabilization with quindoline derivatives results in significantly decreased expression, while mutagenesis that disrupts formation of the structure increases expression [204]. Interestingly, the *TERT* promoter harbors the most frequent recurrently mutated sites across non-coding regions found in multiple cancer types [200, 202], and in ~90% of human cancers TERT expression is upregulated [205]. It has been shown that a G-quadruplex structure can form in the commonly mutated region of the TERT promoter, and stabilization by specific chemical compounds [206–208], leads to the down-regulation of TERT expression, directly implicating mutations of the G-quadruplex locus in carcinogenesis.

Inefficient repair of tandem repeats often leads to repeat expansions or contractions. There are over one million tandem repeats in the human genome, many of which are polymorphic, and their expansion is causal for many disorders [16, 209, 210] such as Huntington disease, spinocerebellar ataxias, Friedreich ataxia and Fragile X syndrome [16, 209, 210]. There is growing evidence that persistent R-loop formation can result in genomic instability [211] and R-loops are implicated in a number of diseases including cancers, autoimmune and neurological disorders [15, 43, 212]. The gene fusion of EWS-FLI or SS18-SSX in Erwin sarcoma has been shown to cause R-loop accumulation and increased replication stress [213]. In cells derived from Friedreich ataxia patients there is an accumulation of R-loops at the expanded GAA repeats of *FXN* gene which causes transcriptional repression [214], while introduction of anti-GAA duplex RNAs interferes with R-loop formation and restores FXN protein levels [215]. In spinal muscular atrophy, Senataxin-deficiency results in accumulation of R-loops, while its over-expression reverses this effect and rescues neurodegeneration [216, 217]. In Wiskott-Aldrich syndrome, which is due to a mutation in Wiskott-Aldrich syndrome protein, there is an accumulation of R-loops leading to genomic instability [218]. In Aicardi-Goutières

syndrome, an excess of R-loops has been observed, especially at DNA hypomethylated sites [219].

Technological advances and prospects

Advances in genome-wide and transcriptome-wide structure inference and visualization methods have allowed for quantification of the abundance and topological characteristics of multiple DNA and RNA structures. While rapid progress has been made recently, we believe that we are only starting to decipher the regulatory roles of secondary structures. Importantly, many recently developed technologies have not been implemented in this field.

One example of such a technology is single cell profiling methods which have the potential to identify relevant differences between cell types. Although it has been shown that G-quadruplex structure formation can be cell-type specific and is associated with higher expression levels and open chromatin [105], the degree to which non-B DNA and RNA structure formation is influenced by the tissue and cell type remains largely unstudied [220]. Single cell technologies could make it possible to apply genome wide assays or small molecules to individual cells to investigate the role of non-B DNA motifs across cell types.

Long-read sequencing technologies are required to map highly repetitive regions of the genome and the transcriptome, and a recent study provided evidence that centromeres are highly enriched in non-B DNA sequences [221]. Moreover, long microsatellite and minisatellite sequences are also routinely excluded from analyses due to prohibiting error rates and mapping problems [222]. Estimation of the full implication of short tandem repeat variation has been limited by sequencing technologies, but with long-read sequencing their contribution will become better understood.

High-throughput experiments enable the systematic investigation of thousands of sequences in a single experiment. Multiple technologies have been developed including those based on synthetic library designs such as massively parallel reporter assays and those based on genome-fragmentation approaches such as STARR-seq [223, 224]. They have provided valuable insights regarding the roles of non-canonical secondary structures in promoters [117], 5'UTRs [225] and 3'UTRs [176].

The field has benefited from a plethora of G-quadruplex ligands, e.g PDS, cPDS, BRACO19, Phen-DC3, L2H2-6OTD, L1H1-7OTD, TMPyP4, which differ in their binding preference for DNA or RNA G-quadruplexes and can shift the equilibrium between folded and unfolded states. The modulation of specific DNA and RNA secondary structures with high specificity could allow for treatments of numerous disorders, including cancer and neurological disorders [226]. For example, Quarfloxin, which interacts with G-quadruplexes [227], reached Phase II trials for several cancer types. Unfortunately, Phase III trials are currently not proceeding due to side effects [228, 229]. CX-5461 is a G-quadruplex ligand that is currently part of a phase I clinical trial (NCT02719977) due to its cytotoxicity to cancer cells, e.g. those that are BRCA1-deficient or BRCA2-deficient [230], and it has been shown to exert its effects due to induction of G-quadruplex formation and topoisomerase II poisoning [231]. Finally, there is a number of pre-clinical studies that suggest that G-quadruplexes could have therapeutic potential, an example being the G-quadruplex ligand CM03, which binds to G-quadruplexes and

down-regulates the expression of multiple genes that are involved in pancreatic ductal adenocarcinoma survival, metastasis and drug resistance [232].

G-quadruplexes can be characterized across the genome [56] and the transcriptome [57] using methods based on high-throughput sequencing. Similarly, methods have been developed to identify R-loops genome-wide [70, 233]. The combination of permanganate footprinting with high-throughput sequencing has enabled the genome-wide detection of single-stranded DNA and the deduction of non-B DNA structures [234]. However, such methods are currently lacking for other non-B DNA structures, e.g. hairpins and slipped structures. The development of novel antibodies and probe technologies could enable the estimation of the frequency and localization of each type of non-B DNA structure globally and could provide insights into how they can switch between their unfolded and folded states.

The cellular mechanisms mediating the stabilization of DNA and RNA secondary structures and those that resolve them, e.g. RBPs, remain incompletely understood. In addition, the effect of interrupting the function of these mechanisms and the relevance to disease progression is unclear. Recent development of high throughput screens coupled with short hairpin RNAs (shRNAs) or CRISPR-based technologies have enabled systematic interrogation of the roles of diverse proteins, such as transcription factors, RBPs, helicases, and topoisomerases. Mutational analysis with CRISPR-Cas9 could also be used to study the effects of non-B DNA motif disruption *in vivo*, while variants of the technology without endonuclease activity could also be used to elucidate their functions.

Concluding Remarks

High-throughput technologies enable the systematic investigation of non-canonical secondary structures as well as the design of experiments to quantify their contribution in the regulation of gene expression and to directly testing their mechanisms of action. The discovery of new methods to dynamically identify non-B DNA and RNA structures is gradually revealing their widespread and diverse contributions in gene regulation. However, it remains difficult to capture their dynamic changes across cellular conditions and their interplay with proteins. We believe that the implementation of novel technologies will enable breakthrough discoveries for their roles with important implications in our understanding of gene regulation. Crucially, a better understanding of the mechanisms through which secondary structures impact gene expression will allow for the development of novel therapeutic strategies for a wide range of diseases, including cancer and neurodegenerative disorders.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-022-02727-6.

Additional file 1. Review history.

Georgakopoulos-Soares *et al. Genome Biology*   (2022) 23:159

Page 15 of 22

Declarations

References

1. Hänsel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, et al. G-quadruplex structures mark human regulatory chromatin. Nat Genet. 2016;48:1267–72.
2. Kouzine F, Wojtowicz D, Baranello L, Yamane A, Nelson S, Resch W, et al. Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. Cell Syst. 2017;4:344–56. e7.
3. Zhao J, Bacolla A, Wang G, Vasquez KM. Non-B DNA structure-induced genetic instability and evolution. Cell Mol Life Sci. 2010;67:43–62.
4. 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. 2002;99:11593–8.
5. Belotserkovskii BP, De Silva E, Tornaletti S, Wang G, Vasquez KM, Hanawalt PC. A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. J Biol Chem. 2007;282:32433–41.
6. Ditlevson JV, Tornaletti S, Belotserkovskii BP, Teijeiro V, Wang G, Vasquez KM, et al. Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA polymerase. Nucleic Acids Res. 2008;36 3163–70. Available from: https://doi.org/10.1093/nar/gkn136
7. 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. 2007. 3:218–221. Available from: https://doi.org/10.1038/nchembio864
8. Ray BK, Dhar S, Shakya A, Ray A. Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression. Proc Natl Acad Sci U S A. 2011;108:103–8.
9. Agarwala P, Pandey S, Mapa K, Maiti S. The G-Quadruplex Augments Translation in the 5′ Untranslated Region of Transforming Growth Factor β2 Biochemistry. 2013. 52:1528–1538. Available from: https://doi.org/10.1021/bi301365g
10. Georgakopoulos-Soares I, Parada GE, Wong HY, Miska EA, Kwok CK, Hemberg M. Alternative splicing modulation by G-quadruplexes. Available from: https://doi.org/10.1101/700575
11. Brázda V, Bartas M, Lýsek J, Coufal J, Fojta M. Global analysis of inverted repeat sequences in human gene promoters reveals their non-random distribution and association with specific biological pathways. Genomics. 2020;112:2772–7.
12. Asamitsu S, Obata S, Yu Z, Bando T, Sugiyama H. Recent Progress of Targeted G-Quadruplex-Preferred Ligands Toward Cancer Therapy. Molecules. 2019;24:429 Available from: https://doi.org/10.3390/molecules24030429
13. Asamitsu S, Takeuchi M, Ikenoshita S, Imai Y, Kashiwagi H, Shioda N. Perspectives for Applying G-Quadruplex Structures in Neurobiology and Neuropharmacology. Int J Mol Sci. 2019;20. Available from: https://doi.org/10.3390/ijms20122884
14. Hänsel-Hertsch R, Di Antonio M, Balasubramanian S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nat Rev Mol Cell Biol. 2017;18:279–84.
15. Richard P, Manley JL. R Loops and Links to Human Disease. J Mol Biol. 2017. p. 3168–3180. Available from: https://doi.org/10.1016/j.jmb.2016.08.031
16. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. Nat Rev Genet. 2018;19:286–98.
17. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953;171:737–8.
18. Belmont P, Constant JF, Demeunynck M. Nucleic acid conformation diversity: from structure to function and regulation. Chem Soc Rev. 2001. p. 70–81. Available from: https://doi.org/10.1039/a904630e
19. Arnott S, Hukins DWL. Optimised parameters for A-DNA and B-DNA [Internet]. Biochem Biophys Res Commun. 1972. p. 1504–1509. Available from: https://doi.org/10.1016/0006-291x(72)90243-4

20. Kaushik M, Kaushik S, Roy K, Singh A, Mahendru S, Kumar M, et al. A bouquet of DNA structures: Emerging diversity. Biochem Biophys Rep. 2016;5:388–95.
21. Ghosh A, Bansal M. A glossary of DNA structures from A to Z. Acta Crystallogr D Biol Crystallogr. 2003;59:620–6.
22. Peck LJ, Nordheim A, Rich A, Wang JC. Flipping of cloned d(pCpG)n.d(pCpG)n DNA sequences from right- to left-handed helical structure by salt, Co(III), or negative supercoiling. Proc Natl Acad Sci U S A. 1982;79:4560–4.
23. Wang AH, Quigley GJ, Kolpak FJ, Crawford JL, van Boom JH, van der Marel G, et al. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature. 1979;282:680–6.
24. Gessner RV, Frederick CA, Quigley GJ, Rich A, Wang AH. The molecular structure of the left-handed Z-DNA double helix at 1.0-A atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). J Biol Chem. 1989;264:7921–35.
25. Shin S-I, Ham S, Park J, Seo SH, Lim CH, Jeon H, et al. Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome. DNA Res. 2016;23:477–86.
26. Wittig B, Dorbic T, Rich A. Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei. Proc Natl Acad Sci U S A. 1991;88:2259–63.
27. Gellert M, Lipsett MN, Davies DR. Helix formation by guanylic acid. Proc Natl Acad Sci U S A. 1962;48:2013–8.
28. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 2006. p. 5402–5415. Available from: https://doi.org/10.1093/nar/gkl655
29. 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. 2015;33:877–81.
30. Bartas M, Brázda V, Karlický V, Červeň J, Pečinka P. Bioinformatics analyses and in vitro evidence for five and six stacked G-quadruplex forming sequences. Biochimie. 2018;150:70–5.
31. Lim KW, Jenjaroenpun P, Low ZJ, Khong ZJ, Ng YS, Kuznetsov VA, et al. Duplex stem-loop-containing quadruplex motifs in the human genome: a combined genomic and structural study. Nucleic Acids Res. 2015;43:5630–46.
32. Watson J, Hays FA, Ho PS. Definitions and analysis of DNA Holliday junction geometry. Nucleic Acids Res. 2004;32:3017–27.
33. Nag DK, Petes TD. Seven-base-pair inverted repeats in DNA form stable hairpins in vivo in Saccharomyces cerevisiae. Genetics. 1991;129:669–73.
34. Varani G. Exceptionally stable nucleic acid hairpins. Annu Rev Biophys Biomol Struct. 1995;24:379–404.
35. Lobachev KS, Shor BM, Tran HT, Taylor W, Keen JD, Resnick MA, et al. Factors affecting inverted repeat stimulation of recombination and deletion in Saccharomyces cerevisiae. Genetics. 1998;148:1507–24.
36. Woodside MT, Anthony PC, Behnke-Parks WM, Larizadeh K, Herschlag D, Block SM. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. Science. 2006;314:1001–4.
37. Sinden RR, Pytlos-Sinden MJ, Potaman VN. Slipped strand DNA structures. Front Biosci. 2007;12:4788–99.
38. Moore H, Greenwell PW, Liu CP, Arnheim N, Petes TD. Triplet repeats form secondary structures that escape DNA repair in yeast. Proc Natl Acad Sci U S A. 1999;96:1504–9.
39. Voloshin ON, Mirkin SM, Lyamichev VI, Belotserkovskii BP, Frank-Kamenetskii MD. Chemical probing of homopurine-homopyrimidine mirror repeats in supercoiled DNA. Nature. 1988;333:475–6.
40. Broitman SL. H-DNA:DNA triplex formation within topologically closed plasmids. Prog Biophys Mol Biol. 1995;63:119–29.
41. Brown JA. Unraveling the structure and biological functions of RNA triple helices. Wiley Interdiscip Rev RNA. 2020;11:e1598.
42. Crossley MP, Bocek M, Cimprich KA. R-Loops as Cellular Regulators and Genomic Threats. Mol Cell. 2019;73:398–411.
43. Niehrs C, Luke B. Regulatory R-loops as facilitators of gene expression and genome stability. Nat Rev Mol Cell Biol. 2020;21:167–78.
44. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. 2004;18:1618–29.
45. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. Nat Rev Genet. 2015;16:583–97.
46. Chakraborty P, Grosse F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. DNA Repair. 2011:654–65. Available from:. https://doi.org/10.1016/j.dnarep.2011.04.013.
47. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature. 2014;516:436–9.
48. Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, et al. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell. 2016;63:167–78.
49. Sinden RR. DNA Structure and Function. San Diego: Gulf Professional Publishing; 1994.
50. Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, et al. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods. 1986;89:123–30.
51. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. Nat Chem. 2013;5:182–6.
52. Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, et al. Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. 2014;42:860–9.
53. Hartono SR, Malapert A, Legros P, Bernard P, Chédin F, Vanoosthuyse V. The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. J Mol Biol. 2018;430:272–84.
54. Chen L, Chen J-Y, Zhang X, Gu Y, Xiao R, Shao C, et al. R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol Cell. 2017;68:745–57.e5.
55. Yan Q, Shields EJ, Bonasio R, Sarma K. Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach. Cell Rep. 2019;29:1369–80.e5.

56. Marsico G, Chambers VS, Sahakyan AB, McCauley P, Boutell JM, Antonio MD, et al. Whole genome experimental maps of DNA G-quadruplexes in multiple species. Nucleic Acids Res. 2019;47:3862–74.

57. Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. Nat Methods. 2016;13:841–4.

58. Tu J, Duan M, Liu W, Lu N, Zhou Y, Sun X, et al. Direct genome-wide identification of G-quadruplex structures by whole-genome resequencing. Nat Commun. 2021;12:6014.

59. Hänsel-Hertsch R, Spiegel J, Marsico G, Tannahill D, Balasubramanian S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat Protoc. 2018;13:551–64.

60. Mendoza O, Bourdoncle A, Boulé J-B, Brosh RM Jr, Mergny J-L. G-quadruplexes and helicases. Nucleic Acids Res. 2016;44:1989–2006.

61. Hänsel-Hertsch R, Spiegel J, Marsico G, Tannahill D, Balasubramanian S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat Protoc. 2018;13:551–64.

62. Liu H-Y, Zhao Q, Zhang T-P, Wu Y, Xiong Y-X, Wang S-K, et al. Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. Cell Chem Biol. 2016;23:1261–70.

63. Müller S, Kumari S, Rodriguez R, Balasubramanian S. Small-molecule-mediated G-quadruplex isolation from human cells. Nat Chem. 2010;2:1095–8.

64. Di Antonio M, Ponjavic A, Radzevičius A, Ranasinghe RT, Catalano M, Zhang X, et al. Single-molecule visualization of DNA G-quadruplex formation in live cells. Nat Chem. 2020;12:832–7.

65. Herbert A, Alfken J, Kim YG, Mian IS, Nishikura K, Rich A. A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. Proc Natl Acad Sci U S A. 1997;94:8421–6.

66. Kim YG, Lowenhaupt K, Schwartz T, Rich A. The interaction between Z-DNA and the Zab domain of double-stranded RNA adenosine deaminase characterized using fusion nucleases. J Biol Chem. 1999;274:19081–6.

67. Schwartz T, Rould MA, Lowenhaupt K, Herbert A, Rich A. Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. Science. 1999;284:1841–5.

68. Herbert A, Lowenhaupt K, Spitzner J, Rich A. Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. Proc Natl Acad Sci U S A. 1995;92:7550–4.

69. Cetin NS, Kuo C-C, Ribarska T, Li R, Costa IG, Grummt I. Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. Nucleic Acids Res. 2019;47:2306.

70. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell. 2012;45:814–25.

71. Chen L, Chen J-Y, Zhang X, Gu Y, Xiao R, Shao C, et al. R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol Cell. 2017;68:745–57.e5.

72. Chen L, Chen J-Y, Zhang X, Gu Y, Xiao R, Shao C, et al. R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol Cell. 2017;68:745–57.e5.

73. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell. 2012;45:814–25.

74. Wulfridge P, Sarma K. BisMapR: a strand-specific, nuclease-based method for genome-wide R-loop detection [Internet]. Available from: https://doi.org/10.1101/2021.01.22.427764

75. Zuber PK, Artsimovitch I, NandyMazumdar M, Liu Z, Nedialkov Y, Schweimer K, et al. The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. Elife. 2018;7. Available from: https://doi.org/10.7554/eLife.36349

76. Spiro C, McMurray CT. Switching of DNA secondary structure in proenkephalin transcriptional regulation. J Biol Chem. 1997;272:33145–52.

77. Raiber E-A, Kranaster R, Lam E, Nikan M, Balasubramanian S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. Nucleic Acids Res. 2012;40:1499–508.

78. Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, et al. The impact of short tandem repeat variation on gene expression. Nat Genet. 2019;51:1652–9.

79. Spiegel J, Cuesta SM, Adhikari S, Hänsel-Hertsch R, Tannahill D, Balasubramanian S. G-quadruplexes are transcription factor binding hubs in human chromatin. Genome Biol. 2021;22:117.

80. Gao J, Zybailov BL, Byrd AK, Griffin WC, Chib S, Mackintosh SG, et al. Yeast transcription co-activator Sub1 and its human homolog PC4 preferentially bind to G-quadruplex DNA. Chem Commun. 2015;51:7242–4.

81. Szlachta K, Thys RG, Atkin ND, Pierce LCT, Bekiranov S, Wang Y-H. Alternative DNA secondary structure formation affects RNA polymerase II promoter-proximal pausing in human. Genome Biol. 2018;19:89.

82. Farokhzad OC, Teodoridis JM, Park H, Arnaout MA, Shelley CS. CD43 gene expression is mediated by a nuclear factor which binds pyrimidine-rich single-stranded DNA. Nucleic Acids Res. 2000;28:2256–67.

83. Qin Y, Fortin JS, Tye D, Gleason-Guzman M, Brooks TA, Hurley LH. Molecular cloning of the human platelet-derived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-beta expression. Biochemistry. 2010;49:4208–19.

84. Cogoi S, Paramasivam M, Membrino A, Yokoyama KK, Xodo LE. The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. J Biol Chem. 2010;285:22003–16.

85. Uribe DJ, Guo K, Shin Y-J, Sun D. Heterogeneous nuclear ribonucleoprotein K and nucleolin as transcriptional activators of the vascular endothelial growth factor promoter through interaction with secondary DNA structures. Biochemistry. 2011;50:3796–806.

86. David AP, Margarit E, Domizi P, Banchio C, Armas P, Calcaterra NB. G-quadruplexes as novel cis-elements controlling transcription during embryonic development. Nucleic Acids Res. 2016;44:4163–73.

87. Saha D, Singh A, Hussain T, Srivastava V, Sengupta S, Kar A, et al. Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex–dependent fashion. J Biol Chem. 2017. p. 15205–15. Available from: https://doi.org/10.1074/jbc.m117.792077

88. Palumbo SL, Memmott RM, Uribe DJ, Krotova-Khan Y, Hurley LH, Ebbinghaus SW. A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. Nucleic Acids Res. 2008;36:1755–69.

89. Eddy J, Vallur AC, Varma S, Liu H, Reinhold WC, Pommier Y, et al. G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. Nucleic Acids Res. 2011;39:4975–83.

90. Tan J, Wang X, Phoon L, Yang H, Lan L. Resolution of ROS-induced G-quadruplexes and R-loops at transcriptionally active sites is dependent on BLM helicase. FEBS Lett. 2020;594:1359–67.

91. Wong B, Chen S, Kwon J-A, Rich A. Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 2007;104:2229–34.

92. Wong HM, Huppert JL. Stable G-quadruplexes are found outside nucleosome-bound regions. Mol Biosyst. 2009;5:1713–9.

93. Halder K, Halder R, Chowdhury S. Genome-wide analysis predicts DNA structural motifs as nucleosome exclusion signals. Mol Biosyst. 2009;5:1703–12.

94. Farabella I, Di Stefano M, Soler-Vila P, Marti-Marimon M, Marti-Renom MA. Three-dimensional genome organization via triplex-forming RNAs. Nat Struct Mol Biol. 2021;28:945–54.

95. Li L, Williams P, Ren W, Wang MY, Gao Z, Miao W, et al. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. Nat Chem Biol. 2020;17:161–8.

96. Tikhonova P, Pavlova I, Isaakova E, Tsvetkov V, Bogomazova A, Vedekhina T, et al. DNA G-Quadruplexes Contribute to CTCF Recruitment. Int J Mol Sci. 2021;22. Available from: https://doi.org/10.3390/ijms22137090

97. Hou Y, Li F, Zhang R, Li S, Liu H, Qin ZS, et al. Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. Epigenetics. 2019;14:894–911.

98. Luo H, Zhu G, Eshelman MA, Fung TK, Lai Q, Wang F, et al. HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia. Mol Cell. 2022;82:833–51.e11.

99. Westin L, Blomquist P, Milligan JF, Wrange O. Triple helix DNA alters nucleosomal histone-DNA interactions and acts as a nucleosome barrier. Nucleic Acids Res. 1995;23:2184–91.

100. Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang L-S, et al. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. Nucleic Acids Res. 2008;36:144–56.

101. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. 2007;35:406–13.

102. Du Z, Zhao Y, Li N. Genome-wide colonization of gene regulatory elements by G4 DNA motifs. Nucleic Acids Res. 2009;37:6784–98.

103. Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. Nucleic Acids Res. 2008;36:1321–33.

104. Yella VR, Bansal M. DNA structural features of eukaryotic TATA-containing and TATA-less promoters. FEBS Open Bio. 2017;7:324–34.

105. Lago S, Nadai M, Cernilogar FM, Kazerani M, Domíniguez Moreno H, Schotta G, et al. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. Nat Commun. 2021;12:3885.

106. Lin J, Hou J-Q, Xiang H-D, Yan Y-Y, Gu Y-C, Tan J-H, et al. Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation. Biochem Biophys Res Commun. 2013;433:368–73.

107. Hardin CC, Corregan M, Brown BA 2nd, Frederick LN. Cytosine-cytosine+ base pairing stabilizes DNA quadruplexes and cytosine methylation greatly enhances the effect. Biochemistry. 1993;32:5870–80.

108. Mao S-Q, Ghanbarian AT, Spiegel J, Martínez Cuesta S, Beraldi D, Di Antonio M, et al. DNA G-quadruplex structures mold the DNA methylome. Nat Struct Mol Biol. 2018;25:951–7.

109. Fleming AM, Burrows CJ. Interplay of Guanine Oxidation and G-Quadruplex Folding in Gene Promoters. J Am Chem Soc. 2020. p. 1115–1136. Available from: https://doi.org/10.1021/jacs.9b11050

110. Huang Z-L, Dai J, Luo W-H, Wang X-G, Tan J-H, Chen S-B, et al. Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. J Am Chem Soc. 2018;140:17945–55.

111. Baral A, Kumar P, Halder R, Mani P, Yadav VK, Singh A, et al. Quadruplex-single nucleotide polymorphisms (Quad-SNP) influence gene expression difference among individuals. Nucleic Acids Res. 2012;40:3800–11.

112. Renčiuk D, Ryneš J, Kejnovská I, Foldynová-Trantírková S, Andäng M, Trantírek L, et al. G-quadruplex formation in the Oct4 promoter positively regulates Oct4 expression. Biochim Biophys Acta Gene Regul Mech. 2017;1860:175–83.

113. Cogoi S, Xodo LE. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. Nucleic Acids Res. 2006;34:2536–49.

114. Dempsey LA, Sun H, Hanakahi LA, Maizels N. G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. J Biol Chem. 1999;274:1066–71.

115. González V, Guo K, Hurley L, Sun D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. J Biol Chem. 2009;284:23622–35.

116. Thakur RK, Kumar P, Halder K, Verma A, Kar A, Parent J-L, et al. Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. Nucleic Acids Res. 2009;37:172–83.

117. Georgakopoulos-Soares I, Victorino J, Parada GE, Agarwal V, Zhao J, Wong HY, et al. High-throughput characterization of the role of non-B DNA motifs on promoter function. Cell Genomics. 2022. p. 100111. Available from: https://doi.org/10.1016/j.xgen.2022.100111

118. Schroth GP, Chou PJ, Ho PS. Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes. J Biol Chem. 1992;267:11846–55.

119. Khuu P, Sandor M, DeYoung J, Ho PS. Phylogenomic analysis of the emergence of GC-rich transcription elements. Proc Natl Acad Sci U S A. 2007;104:16528–33.

120. Garner MM, Felsenfeld G. Effect of Z-DNA on nucleosome placement. J Mol Biol. 1987. p. 581–590. Available from: https://doi.org/10.1016/0022-2836(87)90034-9

121. Ray BK, Dhar S, Henry C, Rich A, Ray A. Epigenetic regulation by Z-DNA silencer function controls cancer-associated ADAM-12 expression in breast cancer: cross-talk between MeCP2 and NF1 transcription factor family. Cancer Res. 2013;73:736–44.

122. Oh D-B, Kim Y-G, Rich A. Z-DNA-binding proteins can act as potent effectors of gene expression in vivo. Proc Natl Acad Sci U S A. 2002;99:16666–71.

123. Liu R, Liu H, Chen X, Kirby M, Brown PO, Zhao K. Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. Cell. 2001;106:309–18.

124. Maruyama A, Mimura J, Harada N, Itoh K. Nrf2 activation is associated with Z-DNA formation in the human HO-1 promoter. Nucleic Acids Res. 2013;41:5223–34.

125. Dumelie JG, Jaffrey SR. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. Elife. 2017;6. Available from: https://doi.org/10.7554/eLife.28306

126. Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev. 2016;30:1327–38.

127. De Magis A, Manzo SG, Russo M, Marinello J, Morigi R, Sordet O, et al. 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. 2019;116:816–25.

128. Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat Struct Mol Biol. 2015;22:999–1007.

129. Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, et al. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc Natl Acad Sci U S A. 2013;110:13938–43.

130. Cohen S, Puget N, Lin Y-L, Clouaire T, Aguirrebengoa M, Rocher V, et al. Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. Nat Commun. 2018;9:533.

131. Cristini A, Groh M, Kristiansen MS, Gromak N. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Rep. 2018;23:1891–905.

132. Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. FEBS J. 2009;276:1494–505.

133. Zimmer AD, Koshland D. Differential roles of the RNases H in preventing chromosome instability. Proc Natl Acad Sci U S A. 2016;113:12220–5.

134. Guh C-Y, Hsieh Y-H, Chu H-P. Functions and properties of nuclear lncRNAs—from systematically mapping the interactomes of lncRNAs [Internet]. J Biomed Sci 2020. Available from: https://doi.org/10.1186/s12929-020-00640-3

135. Tan-Wong SM, Dhir S, Proudfoot NJ. R-Loops Promote Antisense Transcription across the Mammalian Genome. Mol Cell. 2019;76:600–16.e6.

136. Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, Veluchamy A, et al. R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. Mol Cell. 2020;77:1055–65.e4.

137. Belotserkovskii BP, Neil AJ, Saleh SS, Shin JHS, Mirkin SM, Hanawalt PC. Transcription blockage by homopurine DNA sequences: role of sequence composition and single-strand breaks. Nucleic Acids Res. 2013;41:1817–28.

138. Kellner WA, Bell JSK, Vertino PM. GC skew defines distinct RNA polymerase pause sites in CpG island promoters. Genome Res. 2015;25:1600–9.

139. Zhang X, Chiang H-C, Wang Y, Zhang C, Smith S, Zhao X, et al. Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. Nat Commun. 2017;8:15908.

140. Kalwa M, Hänzelmann S, Otto S, Kuo C-C, Franzen J, Joussen S, et al. The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. Nucleic Acids Res. 2016;44:10631–43.

141. Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, et al. MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. Nat Commun. 2015;6. Available from: https://pubmed.ncbi.nlm.nih.gov/26205790/

142. O'Leary VB, Ovsepian SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, et al. PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation. Cell Rep. 2015;11:474–85.

143. Sawaya S, Bagshaw A, Buschiazzo E, Kumar P, Chowdhury S, Black MA, et al. Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. PLoS One. 2013;8:e54710.

144. Sawaya SM, Lennon D, Buschiazzo E, Gemmell N, Minin VN. Measuring Microsatellite Conservation in Mammalian Evolution with a Phylogenetic Birth–Death Model. Genome Biol Evol. 2012. p. 636–647. Available from: https://doi.org/10.1093/gbe/evs050

145. Yáñez-Cuna JO, Arnold CD, Stampfel G, Boryń LM, Gerlach D, Rath M, et al. Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. Genome Res. 2014;24:1147–56.

146. Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, et al. Abundant contribution of short tandem repeats to gene expression variation in humans. Nat Genet. 2016;48:22–9.

147. Quilez J, Guilmatre A, Garg P, Highnam G, Gymrek M, Erlich Y, et al. Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. Nucleic Acids Res. 2016;44:3750–62.

148. Shimajiri S, Arima N, Tanimoto A, Murata Y, Hamada T, Wang KY, et al. Shortened microsatellite d(CA)21 sequence down-regulates promoter activity of matrix metalloproteinase 9 gene. FEBS Lett. 1999;455:70–4.

149. Gebhardt F, Zänker KS, Brandt B. Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. J Biol Chem. 1999;274:13176–80.

150. Kennedy GC, German MS, Rutter WJ. The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. Nat Genet. 1995;9:293–8.

151. Zhang C, Chen L, Peng D, Jiang A, He Y, Zeng Y, et al. METTL3 and N6-Methyladenosine Promote Homologous Recombination-Mediated Repair of DSBs by Modulating DNA-RNA Hybrid Accumulation. Mol Cell. 2020;79:425–42.e7.

152. Liu N, Zhou KI, Parisien M, Dai Q, Diatchenko L, Pan T. N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. Nucleic Acids Res. 2017;45:6051–63.

153. Raveh-Sadka T, Levo M, Shabi U, Shany B, Keren L, Lotan-Pompan M, et al. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat Genet. 2012;44:743–50.

154. Gemayel R, Vinces MD, Legendre M, Verstrepen KJ. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu Rev Genet. 2010;44:445–77.

155. Borel C, Migliavacca E, Letourneau A, Gagnebin M, Béna F, Sailani MR, et al. Tandem repeat sequence variation as causative cis-eQTLs for protein-coding gene expression variation: the case of CSTB. Hum Mutat. 2012;33:1302–9.

156. Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ. Unstable tandem repeats in promoters confer transcriptional evolvability. Science. 2009;324:1213–6.

157. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. 2005;33:201–12.

158. Moore MJ. From birth to death: the complex lives of eukaryotic mRNAs. Science. 2005;309:1514–8.

159. Litterman AJ, Kageyama R, Le Tonqueze O, Zhao W, Gagnon JD, Goodarzi H, et al. A massively parallel 3′ UTR reporter assay reveals relationships between nucleotide content, sequence conservation, and mRNA destabilization. Genome Res. 2019;29:896–906.

160. Wu X, Bartel DP. Widespread Influence of 3′-End Structures on Mammalian mRNA Processing and Stability. Cell. 2017. p. 905–17.e11. Available from: https://doi.org/10.1016/j.cell.2017.04.036

161. Subramanian M, Rage F, Tabet R, Flatter E, Mandel J-L, Moine H. G-quadruplex RNA structure as a signal for neurite mRNA targeting. EMBO Rep. 2011;12:697–704.

162. Beaudoin J-D, Novoa EM, Vejnar CE, Yartseva V, Takacs CM, Kellis M, et al. Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. Nat Struct Mol Biol. 2018;25:677–86.

163. Brown PH, Tiley LS, Cullen BR. Effect of RNA secondary structure on polyadenylation site selection. Genes Dev. 1991;5:1277–84.

164. Rouleau S, Glouzon J-PS, Brumwell A, Bisaillon M, Perreault J-P. 3′ UTR G-quadruplexes regulate miRNA binding. RNA. 2017;23:1172–9.

165. Kedde M, van Kouwenhove M, Zwart W, Oude Vrielink JAF, Elkon R, Agami R. A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility. Nat Cell Biol. 2010;12:1014–20.

166. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007;39:1278–84.

167. Beaudoin J-D, Perreault J-P. Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. Nucleic Acids Res. 2013;41:5898–911.

168. Fialcowitz EJ, Brewer BY, Keenan BP, Wilson GM. A hairpin-like structure within an AU-rich mRNA-destabilizing element regulates trans-factor binding selectivity and mRNA decay kinetics. J Biol Chem. 2005;280:22406–17.

169. Li X, Quon G, Lipshitz HD, Morris Q. Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. RNA. 2010;16:1096–107.

170. Liu X, Jiang Y, Russell JE. A potential regulatory role for mRNA secondary structures within the prothrombin 3′UTR. Thromb Res. 2010;126:130–6.

171. Erlitzki R, Long JC, Theil EC. Multiple, Conserved Iron-responsive Elements in the 3′-Untranslated Region of Transferrin Receptor mRNA Enhance Binding of Iron Regulatory Protein 2. J Biol Chem. 2002:42579–87. Available from:. https://doi.org/10.1074/jbc.m207918200.

172. Leppek K, Schott J, Reitter S, Poetz F, Hammond MC, Stoecklin G. Roquin Promotes Constitutive mRNA Decay via a Conserved Class of Stem-Loop Recognition Motifs. Cell. 2013:869–81. Available from:. https://doi.org/10.1016/j.cell.2013.04.016.

173. Sarnowska E, Grzybowska EA, Sobczak K, Konopinski R, Wilczynska A, Szwarc M, et al. Hairpin structure within the 3′UTR of DNA polymerase beta mRNA acts as a post-transcriptional regulatory element and interacts with Hax-1. Nucleic Acids Res. 2007;35:5499–510.

174. Meehan HA, Connell GJ. The hairpin loop but not the bulged C of the iron responsive element is essential for high affinity binding to iron regulatory protein-1. J Biol Chem. 2001;276:14791–6.

175. Brown CY, Lagnado CA, Goodall GJ. A cytokine mRNA-destabilizing element that is structurally and functionally distinct from A+U-rich elements. Proc Natl Acad Sci U S A. 1996;93:13721–5.

176. Bogard N, Linder J, Rosenberg AB, Seelig G. A Deep Neural Network for Predicting and Engineering Alternative Polyadenylation. Cell. 2019;178:91–106.e23.

177. Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res. 2008;36:6260–8.

178. Proudfoot NJ. Ending the message: poly(A) signals then and now. Genes Dev. 2011;25:1770–82.

179. Ghosh A, Pandey SP, Ansari AH, Sundar JS, Singh P, Khan Y, et al. Alternative splicing modulation mediated by G-quadruplex structures in MALAT1 lncRNA. Nucleic Acids Res [Internet]. Oxford University Press; 2021 [cited 2021 Nov 15]; Available from: https://academic.oup.com/nar/advance-article-pdf/doi/10.1093/nar/gkab1066/41131450/gkab1066.pdf

180. Decorsière A, Cayrel A, Vagner S, Millevoi S. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3′-end processing and function during DNA damage. Genes Dev. 2011;25:220–5.

181. Lee DSM, Ghanem LR, Barash Y. Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. Nat Commun. 2020;11:527.

182. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol. 2015;16:167–77.

183. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell. 2011;42:794–805.

184. Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, et al. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. Mol Cell. 2015;57:636–47.

185. Chabanon H, Mickleburgh I, Hesketh J. Zipcodes and postage stamps: mRNA localisation signals and their transacting binding proteins. Brief Funct Genomic Proteomic. 2004;3:240–56.

Georgakopoulos-Soares *et al. Genome Biology*   (2022) 23:159

Page 21 of 22

186. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. Understanding the transcriptome through RNA structure. Nat Rev Genet. 2011;12:641–55.

187. Kato J, Niitsu Y. Recent advance in molecular iron metabolism: translational disorders of ferritin. Int J Hematol. 2002;76:208–12.

188. Svoboda P, Di Cara A. Hairpin RNA: a secondary structure of primary importance. Cell Mol Life Sci. 2006;63:901–8.

189. Caput D, Beutler B, Hartog K, Thayer R, Brown-Shimer S, Cerami A. Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. Proc Natl Acad Sci U S A. 1986;83:1670–4.

190. Siegel DA, Tonqueze OL, Biton A, Zaitlen N, Erle DJ. Massively parallel analysis of human 3'UTRs reveals that AU-rich element length and registration predict mRNA destabilization. G3. 2021; Available from: https://doi.org/10.1093/g3journal/jkab404

191. Budworth H, McMurray CT. Bidirectional transcription of trinucleotide repeats: Roles for excision repair. DNA Repair. 2013:672–84. Available from:. https://doi.org/10.1016/j.dnarep.2013.04.019.

192. Thys RG, Lehman CE, Pierce LCT, Wang Y-H. DNA Secondary Structure at Chromosomal Fragile Sites in Human Disease. Curr Genomics. 2015. p. 60–70. Available from: https://doi.org/10.2174/1389202916666150114223205

193. Bacolla A, Tainer JA, Vasquez KM, Cooper DN. Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. Nucleic Acids Res. 2016. p. 5673–5688. Available from: https://doi.org/10.1093/nar/gkw261

194. Georgakopoulos-Soares I, Morganella S, Jain N, Hemberg M, Nik-Zainal S. Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. Genome Res. 2018;28:1264–71.

195. Sinden RR, Zheng GX, Brankamp RG, Allen KN. On the deletion of inverted repeated DNA in Escherichia coli: effects of length, thermal stability, and cruciform formation in vivo. Genetics. 1991;129:991–1005.

196. Dayn A, Malkhosyan S, Mirkin SM. Transcriptionally driven cruciform formation in vivo. Nucleic Acids Res. 1992;20:5991–7.

197. Krasilnikov AS, Podtelezhnikov A, Vologodskii A, Mirkin SM. Large-scale effects of transcriptional DNA supercoiling in Vivo 1 1Edited by I. Tinoco [Internet]. J Mol Biol. 1999. p. 1149–1160. Available from: https://doi.org/10.1006/jmbi.1999.3117

198. Branzei D, Foiani M. Maintaining genome stability at the replication fork. Nat Rev Mol Cell Biol. 2010;11:208–19.

199. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534:47–54.

200. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W. Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet. 2014;46:1160–5.

201. Buisson R, Langenbucher A, Bowen D, Kwan EE, Benes CH, Zou L, et al. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. Science. 2019;364. Available from: https://doi.org/10.1126/science.aaw2872

202. Fredriksson NJ, Ny L, Nilsson JA, Larsson E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. Nat Genet. 2014;46:1258–63.

203. Brooks TA, Kendrick S, Hurley L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J. 2010;277:3459–69.

204. Wang X-D, Ou T-M, Lu Y-J, Li Z, Xu Z, Xi C, et al. Turning off transcription of the bcl-2 gene by stabilizing the bcl-2 promoter quadruplex with quindoline derivatives. J Med Chem. 2010;53:4390–8.

205. Blasco MA, Hahn WC. Evolving views of telomerase and cancer. Trends Cell Biol. 2003;13:289–94.

206. Palumbo SL, Ebbinghaus SW, Hurley LH. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. J Am Chem Soc. 2009;131:10878–91.

207. Lim KW, Lacroix L, Yue DJE, Lim JKC, Lim JMW, Phan AT. Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. J Am Chem Soc. 2010;132:12331–42.

208. Song JH, Kang H-J, Luevano LA, Gokhale V, Wu K, Pandey R, et al. Small-Molecule-Targeting Hairpin Loop of hTERT Promoter G-Quadruplex Induces Cancer Cell Death. Cell Chem Biol. 2019;26:1110–21.e4.

209. Budworth H, McMurray CT. A brief history of triplet repeat diseases. Methods Mol Biol. 2013;1010:3–17.

210. Nakamori M, Mochizuki H. Targeting Expanded Repeats by Small Molecules in Repeat Expansion Disorders. Mov Disord. 2021;36:298–305.

211. Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 2014. p. 1384–1396. Available from: https://doi.org/10.1101/gad.242990.114

212. Perego MGL, Taiana M, Bresolin N, Comi GP, Corti S. R-Loops in Motor Neuron Diseases. Mol Neurobiol. 2019;56:2579–89.

213. Gorthi A, Romero JC, Loranc E, Cao L, Lawrence LA, Goodale E, et al. EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. Nature. 2018;555:387–91.

214. Groh M, Lufino MMP, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. 2014;10:e1004318.

215. Li L, Matsui M, Corey DR. Activating frataxin expression by repeat-targeted nucleic acids. Nat Commun. 2016;7:10606.

216. Kannan A, Bhatia K, Branzei D, Gangwani L. Combined deficiency of Senataxin and DNA-PKcs causes DNA damage accumulation and neurodegeneration in spinal muscular atrophy. Nucleic Acids Res. 2018;46:8326–46.

217. Kannan A, Cuartas J, Gangwani P, Branzei D, Gangwani L. Mutation in senataxin alters the mechanism of R-loop resolution in amyotrophic lateral sclerosis 4. Brain. 2022; Available from: https://doi.org/10.1093/brain/awab464

218. Sarkar K, Han S-S, Wen K-K, Ochs HD, Dupré L, Seidman MM, et al. R-loops cause genomic instability in T helper lymphocytes from patients with Wiskott-Aldrich syndrome. J Allergy Clin Immunol. 2018. p. 219–34. Available from: https://doi.org/10.1016/j.jaci.2017.11.023

Georgakopoulos-Soares *et al. Genome Biology* 2022, **23**(1):159

Page 22 of 22

219. Lim YW, Sanz LA, Xu X, Hartono SR, Chédin F. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutières syndrome. Elife. 2015;4. Available from: https://doi.org/10.7554/eLife.08007

220. Hui WWI, Simeone A, Zyner KG, Tannahill D, Balasubramanian S. Single-cell mapping of DNA G-quadruplex structures in human cancer cells. Sci Rep. 2021;11:23641.

221. Kasinathan S, Henikoff S. Non-B-Form DNA Is Enriched at Centromeres. Mol Biol Evol. 2018;35:949–62.

222. Tørresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarnot P, et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. Nucleic Acids Res. 2019;47:10994–1006.

223. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013;339:1074–7.

224. Muerdter F, Boryń ŁM, Arnold CD. STARR-seq — Principles and applications. Genomics. 2015. p. 145–150. Available from: https://doi.org/10.1016/j.ygeno.2015.06.001

225. Jia L, Mao Y, Ji Q, Dersh D, Yewdell JW, Qian S-B. Decoding mRNA translatability and stability from the 5′ UTR. Nat Struct Mol Biol. 2020;27:814–21.

226. Asamitsu S, Yabuki Y, Ikenoshita S, Wada T, Shioda N. Pharmacological prospects of G-quadruplexes for neurological diseases using porphyrins. Biochem Biophys Res Commun. 2020;531:51–5.

227. Drygin D, Siddiqui-Jain A, O'Brien S, Schwaebe M, Lin A, Bliesath J, et al. Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. Cancer Res. 2009;69:7653–61.

228. Rosenberg JE, Bambury RM, Van Allen EM, Drabkin HA, Lara PN Jr, Harzstark AL, et al. A phase II trial of AS1411 (a novel nucleolin-targeted DNA aptamer) in metastatic renal cell carcinoma. Invest New Drugs. 2014;32:178–87.

229. Morgan RK, Batra H, Gaerig VC, Hockings J, Brooks TA. Identification and characterization of a new G-quadruplex forming region within the kRAS promoter as a transcriptional regulator. Biochim Biophys Acta. 2016;1859:235–45.

230. Xu H, Di Antonio M, McKinney S, Mathew V, Ho B, O'Neil NJ, et al. CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. Nat Commun. 2017;8:14432.

231. Bruno PM, Lu M, Dennis KA, Inam H, Moore CJ, Sheehe J, et al. The primary mechanism of cytotoxicity of the chemotherapeutic agent CX-5461 is topoisomerase II poisoning. Proc Natl Acad Sci U S A. 2020;117:4053–60.

232. Marchetti C, Zyner KG, Ohnmacht SA, Robson M, Haider SM, Morton JP, et al. Targeting Multiple Effector Pathways in Pancreatic Ductal Adenocarcinoma with a G-Quadruplex-Binding Small Molecule. J Med Chem. 2018;61:2500–17.

233. Chen J-Y, Zhang X, Fu X-D, Chen L. R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1. Nat Protoc. 2019;14:1661–85.

234. Cai Z, Cao C, Ji L, Ye R, Wang D, Xia C, et al. RIC-seq for global in situ profiling of RNA–RNA spatial interactions. Nature. 2020. p. 432–437. Available from: https://doi.org/10.1038/s41586-020-2249-1

Publisher's Note