CellPress

Molecular Cell

Technology review

Detection of alternative DNA structures and its implications for human disease

Gabriel Matos-Rodrigues,[1,3] Julia A. Hisey,[2,3] André Nussenzweig,[1,*] and Sergei M. Mirkin[2,*]
[1]Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA
[2]Department of Biology, Tufts University, Medford, MA, USA
[3]These authors contributed equally
*Correspondence: nussenza@exchange.nih.gov (A.N.), sergei.mirkin@tufts.edu (S.M.M.)
https://doi.org/10.1016/j.molcel.2023.08.018

SUMMARY

Around 3% of the genome consists of simple DNA repeats that are prone to forming alternative (non-B) DNA structures, such as hairpins, cruciforms, triplexes (H-DNA), four-stranded guanine quadruplexes (G4-DNA), and others, as well as composite RNA:DNA structures (e.g., R-loops, G-loops, and H-loops). These DNA structures are dynamic and favored by the unwinding of duplex DNA. For many years, the association of alternative DNA structures with genome function was limited by the lack of methods to detect them *in vivo*. Here, we review the recent advancements in the field and present state-of-the-art technologies and methods to study alternative DNA structures. We discuss the limitations of these methods as well as how they are beginning to provide insights into causal relationships between alternative DNA structures, genome function and stability, and human disease.

ALTERNATIVE DNA STRUCTURES: HISTORICAL OVERVIEW

In 1953, James Watson and Francis Crick proposed a model for the structure of DNA based on the experimental data of Rosalind Franklin and Maurice Wilkins. They determined that DNA folds into the so-called "B-form," a right-handed double helix with a helical turn of 10.5 bp built of stacked purine (Pu)-pyrimidine (Py) base pairs of adenine-thymine and guanine-cytosine[1–3] (Figure 1A). Although atypical right-handed DNA helices were observed under specific ambient conditions, such as in water-alcohol solutions (A form)[4] or in the presence of lithium ions (C form),[5] B-DNA was generally believed to be the only feasible DNA structure formed under physiological conditions.

Nevertheless, experimental evidence began to accumulate pointing to the existence of alternative nucleic acid structures that are radically different from B-DNA. Three-stranded RNA structures formed by synthetic poly-A and poly-U tracts were the first ones discovered.[6] The authors correctly hypothesized that the third poly-U strand could fit into the major groove of the A:U duplex[7] forming Hoogsteen hydrogen bonds[8,9] with adenines of the duplex. Homopolymers (dA)n·(dT)n and d(TC)n·d(GA)n were soon found to form triplexes with poly-rU and poly-r(UC) RNA strands, respectively.[10,11] Subsequently, DNA triplexes, in which a d(TC)n strand formed Hoogsteen hydrogen bonds with the d(GA)n·d(TC)n duplex were detected at lower pH.[12–14]

In an independent development, tri- and tetra-nucleotides of deoxyriboguanylic acid were found to form exceptionally stable higher-order structures.[15] Studying X-ray diffraction of gels formed by guanosine monophosphate, Gellert et al.[16] proposed that G-quartets stabilized by Hoogsteen hydrogen bonds are responsible for their formation and stability. It was further suggested that monovalent cations, particularly K^+, additionally stabilize G-quartets.[17]

Although these early studies did not initially spark general interest in shifting the accepted paradigm that B-DNA is the only thermodynamically favorable structure, a striking blow to this consensus came from the first crystal DNA structure of the $(CG)_3$ repeat, which appeared to exist in a totally different DNA conformation—a left-handed Z-DNA[18] (Figure 1A). Aside from its opposite helix sign, Z-DNA differs from B-DNA in that its dinucleotide repetitive unit results in a zig-zag sugar-phosphate backbone. Z-DNA was then shown to readily form in negatively supercoiled plasmid DNA.[19]

After the discovery of Z-DNA, additional alternative DNA structures were shown to form in supercoiled plasmid DNA. For example, two halves of an inverted repeat in the same DNA strand can pair with each other, rather than with their complementary DNA strand, generating a cruciform-shaped structure (Figure 1A). The base of the DNA cruciform is a four-way DNA junction, and there are at least 3 single-stranded DNA (ssDNA) bases at each of its tips (Figure 1A).[20,21]

Triplex H-DNA formed by natural homopurine/homopyrimidine (hPu/hPy) mirror repeats was the first multistranded DNA structure discovered.[22,23] In this structure, the DNA strand corresponding to one half of a mirror repeat (either pyrimidine or purine) unwinds and folds back to form Hoogsteen or reverse Hoogsteen hydrogen bonds with the purines of the duplex half of the repeat, whereas its complementary strand remains single stranded (Figure 1A).

Shortly after the discovery of H-DNA, another multistranded DNA structure, now called G4-DNA, was found in single-stranded G-rich sequences located at immunoglobulin class switch recombination (CSR) regions and telomeres.[24–27] It is built by stacked G-quartets paired via Hoogsteen hydrogen bonds

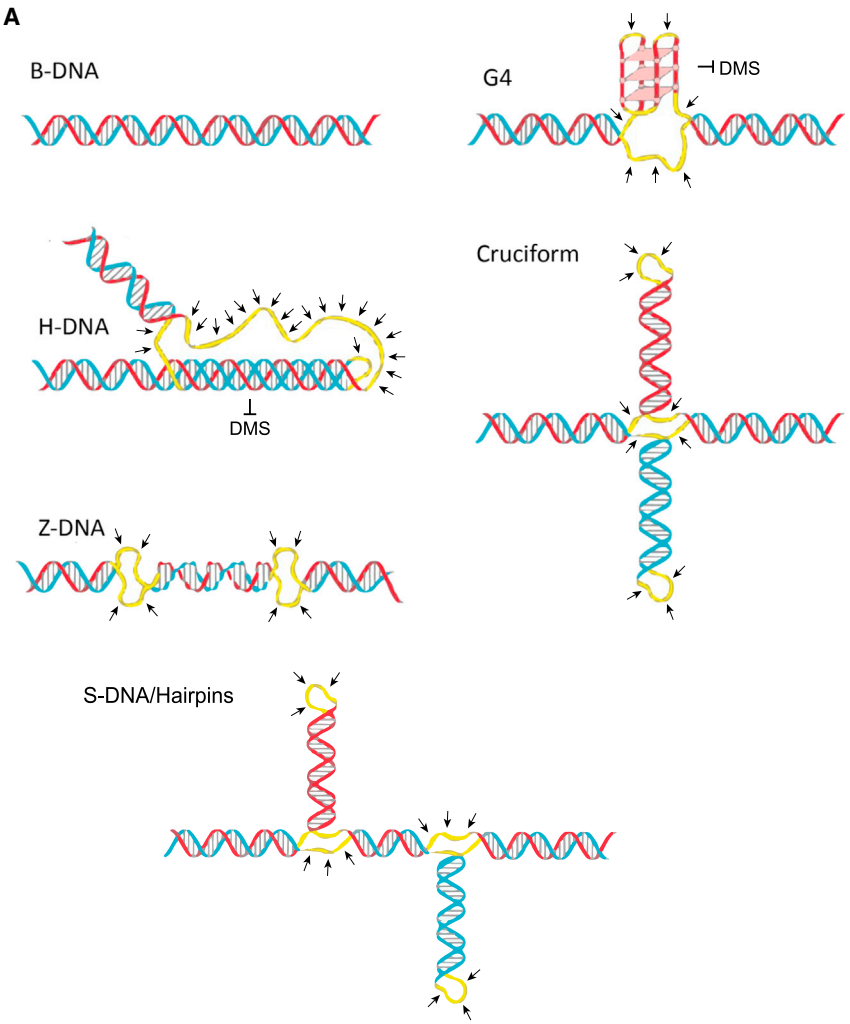

Figure 1. B-DNA and alternative (non-B) DNA structures

(A) Representative images of right-handed B-form DNA double helix. Alternative (non-B) DNA structures: G4s formed in formed in $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ consensus sequences, H-DNA formed in homopurine/homopyrimidine (hPu/hPy) mirror repeats, Z-DNA in regularly alternating $(PuPy)_n$ repeats, cruciform in inverted repeats, and hairpins/S-DNA in direct tandem repeats. Each alternative DNA structure shows its specific distribution of ssDNA stretches/unpaired bases (yellow regions). The arrows highlight regions where ssDNA processing enzymes or chemicals can act on alternative DNA structures. The block sign illustrates the inhibitory effect of multistrand DNA present in G4s and H-DNA on dimethyl sulfate (DMS)-induced guanine methylation.

(B) RNA:DNA hybrid-associated alternative DNA structures. RNA:DNA hybrids can form R-loops when the ssDNA is not folded or into other types of RNA:DNA hybrids when the non-paired ssDNA folds into alternative DNA structures, such as G4s, forming G-loops or H-DNA, forming H-loops. Each alternative DNA structure shows its specific distribution of ssDNA stretches/unpaired bases (yellow regions).

and stabilized by monovalent cations, particularly potassium (Figure 1A). Importantly, G4-DNA can be formed by one, two, or four DNA strands in various orientations relative to each other.

These studies fueled the discovery of other alternative DNA structures,[28–31] including i-motifs (iMs) composed of intercalated cytosine-cytosine base pairs,[32] DNA unwinding elements (DUEs) formed by AT-rich elements located adjacent to replication origins,[33] and S-DNA consists of slipped-strand hairpins formed by some direct tandem repeats[34,35] (Figure 1A).

The underlying structural, biophysical, and biochemical characteristics of alternative DNA structures were initially established *in vitro*. First, alternative DNA structures are formed in regions of repetitive DNA with strict sequence requirements, for example, regularly alternating $(PuPy)_n$ repeats form Z-DNA, inverted repeats form DNA cruciform, hPu/hPy mirror repeats form H-DNA, and guanine-runs belonging to the consensus sequence $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ form G4s, whereas certain direct tandem repeats form S-DNA/hairpins (reviewed in Brown and Freudenreich,[28] Wang and Vasquez,[29] and Khristich and Mirkin[30]) (Figure 1A). Second, a fundamental property of all these structures is that they are thermodynamically unfavorable in linear DNA but can be promoted by supercoiling or biological processes that unwind B-DNA, such as replication or transcription.[29,30] In transcribed regions, an RNA transcript can invade duplex DNA, forming three-stranded RNA:DNA structures called R-loops.[36–38] Furthermore, RNA transcripts can stabilize other alternative DNA structures, such as H-DNA and G4s, by binding to the free ssDNA strand to create mixed structures called H-loops (H-DNA)[39] or G-loops (G4-DNA)[40] (Figure 1B).

Tandem repeats, sequences of two or more DNA bases that are repeated numerous times in a head-to-tail fashion, account for about 3% of the human genome[41] in a total of 1,049,715 repeats (GRCh38 genome annotated using tandem repeats finder[42] last updated at UCSC: 2022-10-18). Importantly, this is still an underrepresentation of the repeats that potentially form alternative DNA structures, as some sequences, such as G4 forming repeats, are often not annotated as tandem repeats. Indeed, about 716,310 genomic DNA sequences were found to potentially form G4 alone.[43] In addition, over sixty repeat expansion diseases (REDs) caused by pathogenic expansions of tandem repeats have been reported (reviewed in Khristich and Mirkin[30] and Gall-Duncan et al.[44]). The most well-known REDs include Huntington's disease (HD), fragile X syndrome (FXS), and Friedreich's ataxia (FRDA), which are caused by the expansion of $(CAG)_n$, $(CGG)_n$, and $(GAA)_n$ repeats, respectively. Thus, DNA sequences prone to form alternative DNA structures are highly frequent in the human genome and associated with several human diseases.

The discovery of alternative DNA structures led to numerous speculations on their possible biological roles. Z-DNA was suggested to play a role in transcriptional activation.[45] It was also postulated to initiate genetic recombination, since two DNA strands of the adjacent B- and Z-DNA segments are not topologically linked.[46–48] DNA cruciforms were suggested to play a role in both site-specific[49] and homologous recombination (HR).[50] In fact, DNA cruciforms are so similar to Holliday junctions that they were used as a bait to isolate Holliday junction resolvases (e.g., MUS81).[51,52] A separate group of studies implicated the formation of DNA cruciforms in replication origins.[53–55] Since H-DNA-forming hPu/hPy mirror repeats were initially identified in the upstream promoter regions of eukaryotic genes,[23] they were thought to play a role in transcription initiation.[56–58] H-DNA-forming sequences were also proposed to serve as terminators of DNA replication[59,60] and speculated to be the hotspots of HR.[59] Localization of G4-DNA to telomeric overhangs[26,27] implied their role in chromosome end-protection. Another thought-provoking idea was that four telomeric overhangs might form a parallel-stranded G4-DNA structure during the course of homolog recognition in meiotic prophase.[24] Finally, the role of G4-DNA in immunoglobulin CSR was also discussed, given their presence at these regions.[24] S-DNA was believed to be at the heart of repeat expansions, leading to their progressive lengthening by misalignment, or "slipping," during replication or repair.[61–63]

Some of these ideas were subsequently substantiated, whereas others have yet to be substantiated (reviewed in Brown and Freudenreich,[28] Wang and Vasquez,[29] Khristich and Mirkin,[30] and Georgakopoulos-Soares et al.[31]). Notably, a major obstacle for establishing the biological role of alternative DNA structures was that their dynamic nature made it challenging to unambiguously prove their existence *in vivo*, particularly in large eukaryotic genomes, given that a specific structure may only be present in a small fraction of the cell population at any given time. Thus, the development of new methods was key for the detection of alternative DNA structures. This review concentrates on various methods to detect alternative DNA structures *in vivo*, and how the development of these technologies has changed our understanding of their biological role in health and disease.

METHODS TO DETECT ALTERNATIVE DNA STRUCTURES *IN VITRO*

New methods for the detection of alternative DNA *in vivo* are deeply grounded in earlier developed approaches for their detection *in vitro*. Unlike canonical double-stranded DNA, most alternative DNA structures contain ssDNA regions (Figures 1A and 1B) that have been harnessed in various ways for non-canonical DNA structure detection. S1 nuclease, which preferably cleaves ssDNA,[64] emerged as one of the first of these tools.[19–21] Although S1 nuclease was key in the discovery of DNA cruciforms[19–21] and H-DNA (reviewed in Mirkin and Frank-Kamenetskii[65]), it functions at acidic pH. Other ssDNA-specific nucleases, such as P1 and mung bean (MBN) nucleases, were also used (reviewed in Wang et al.[66]), since they cleave DNA under neutral pH, i.e., at near-physiological conditions.

An alternative approach utilized chemical probes that modify bases according to the type of hydrogen bonding they are involved in. Several of them modify ssDNA bases and therefore identify the absence of Watson-Crick hydrogen bonding.[67] These include osmium tetroxide (OsO_4) that modifies the unsaturated 5-6 double bond of single-stranded thymines,[68] chloroacetaldehyde (CAA) that converts single-stranded adenines, cytosines, and guanines into their etheno derivatives,[69] potassium permanganate ($KMnO_4$) that causes *cis*-dihydroxylation of the 5-6 double bond of single-stranded thymines,[70] and diethyl pyrocarbonate (DEPC) that preferably carboxylates the N6 and N7 positions of single-stranded adenines.[71] Another useful chemical probe to detect multistranded DNA structures is dimethyl sulfate (DMS), since it methylates the N7 position of guanines when they are not involved in Hoogsteen or reverse Hoogsteen hydrogen bonding.[72] Therefore, guanines are protected from DMS modification in three- and four-stranded alternative DNA structures that utilize those hydrogen bonding. All these modifications can be detected at nucleotide resolution via DNA sequencing[73–76] and were thereby used to elucidate key characteristics of alternative DNA structures. Importantly, assays that rely on protection from labeling are only suitable when alternative DNA structures are highly frequent.

For DNA cruciforms, central single-stranded loops and unwound regions at four-way junctions were modified by different single-stranded base-specific chemicals.[77–82] Z-DNA was modified at the B-to-Z junction using

single-stranded base-specific chemicals,[83,84] whereas DEPC selectively modified purines in Z-conformation.[85] H-DNA can fold in two forms: H-y and H-r. In the H-y form, the pyrimidine strand is single stranded, whereas the purine strand is single stranded for H-r. For H-DNA in the H-y form, half of the purine strand and the center of pyrimidine strand were modified by single-stranded base-specific chemicals, whereas the remaining half of the purine strand was protected from DMS modification.[73,75,76,86,87] In H-r DNA, half of the pyrimidine strand was modified by single-stranded base-specific chemicals, whereas half of the purine strand was protected from methylation by DMS.[27,88–90] In G4-DNA, all guanines were protected against DMS methylation[24,26,27] (Figure 1).

Since they generate irreversible products, one drawback of using single-strand-specific nucleases and chemicals is that they can shift the equilibrium from B-DNA to alternative DNA structures. Thus, additional methods including two-dimensional gel electrophoresis were used to detect interconversions between B-DNA and alternative structures during plasmid supercoiling.[91] This allowed detection of the Z-DNA,[92,93] DNA cruciform,[94,95] and triplex H-DNA formation.[22]

Small molecules, proteins, or antibodies that direct bind alternative DNA structures can also be used for DNA structures *in vitro* (e.g., Lam et al.,[96] Zheng et al.,[97] and Müller et al.[98]). Another feature exploited in the detection of some alternative DNA structures is their potent ability to halt polymerases. Although initially identified as an obstacle to DNA sequencing, particularly in G/C-rich sequences,[99] this feature was quickly harnessed into a tool referred to as the polymerase stop assay to detect alternative DNA structures,[100] including G4-DNA and H-DNA *in vitro*. With the advent of high-throughput sequencing methods, genome-wide modifications of the polymerase stop assay were developed. Murat et al. evaluated the kinetics and fidelity of DNA synthesis using 20,000 sequences comprising all short tandem repeats permutations in different lengths and found that polymerase stalling and pairing errors during DNA sequencing could be used to predict the formation of alternative DNA structures, such as hairpins and G4s.[101] In another method, whole-genome sequencing (WGS) was carried out with and without G-quadruplex-stabilizing ligands or ions, leading to an observable alteration in sequencing readout attributable to G4-DNA formation.[43,102]

EVIDENCE OF LOCAL ALTERNATIVE SECONDARY STRUCTURE FORMATION *IN VIVO*

Given their relevance to disease, certain loci and DNA motifs have been interrogated more closely using various tools to detect structure formation *in vivo*. One technique used is ligation-mediated PCR (LM-PCR), which allows for nucleotide-resolution mapping of DNA breaks using knowledge of only one side of the break since the second PCR primer anneals to a universal linker ligated to the DNA end.[66,103] LM-PCR was initially used to map double-strand breaks (DSBs) occurring *in vivo* to H-DNA motifs in the human c-*myc* gene,[104] then to verify that the breaks were triplex-induced by localizing the sites of breakage specifically to the loop of the triplex.[105] In another instance, DNA breaks were localized the base of a cruciform.[106] These studies pointed to the role of nucleotide excision repair (NER) components in cleaving H-DNA and DNA cruciforms.[105,106]

A clever approach to detect alternative DNA structures involves proteins either designed or found in nature that specifically bind to a sequence in its non-B-DNA form. For example, zinc-finger nucleases (ZFNs) specifically designed to cleave $(CTG)_n$ and $(CTG)_n$ hairpin structures were first characterized *in vitro*. Expression of these nucleases in human cells revealed the formation of S-DNA *in vivo* during replication of the ectopic cassette carrying expandable $(CTG)_n \cdot (CTG)_n$ repeats.[107]

An important disease-related example involves the adenosine deaminase ADAR1, whose p150 isoform contains a Z-DNA/Z-RNA-specific binding domain. In cancer, loss of ADAR1 overcomes resistance to immune checkpoint blockade[108] and subsequent activation of ZBP1-dependent necroptosis.[109] A small molecule that induces Z-DNA formation and thereby directly activates ZBP1-dependent necroptosis was able to potentiate the immune checkpoint blockade response in a melanoma mouse model.[109] On the other hand, variants of ADAR1 isoform p150 that reduce its Z-RNA binding lead to three diseases: dyschromatosis symmetrica hereditaria, Aicardi-Goutières syndrome, and bilateral striatal necrosis/dystonia, likely through dysregulation of the innate immune response.[110,111]

Another immune-related example is an interaction between activation-induced cytidine deaminase (AID) and G4-DNA in the immunoglobulin loci involved in CSR and somatic hypermutation (SHM). Transcription through the class switch region leading to G-loop formation has been known to be key to CSR.[112] More recently, G4-DNA formed by CSR regions, rather than linear B-DNA, was found to be the preferred substrate for AID.[113] Altogether, these data support the idea that formation of G-loops[40] and their subsequent cleavage by AID drive CSR. A mouse model of hyper-immunoglobulin M (IgM) syndrome bearing an orthologous AID mutation with impaired G4-DNA binding but retained catalytic activity lacks both CSR and SHM, providing evidence that this protein-structure interaction occurs and has functional significance *in vivo*.[114] Corroborating the significance of this interaction, G4-DNA stabilization using a small molecule decreases CSR in mice.[115]

The causal link between structure formation and pathological outcomes has historically been hindered by the lack of reliable methods to visualize alternative DNA structures *in vivo*. In the past few years, however, the development of new tools has enabled progress in this field.

GENOME-WIDE MAPPING OF ALTERNATIVE DNA STRUCTURES *IN VIVO*

Antibody-based and related methods

Antibody-based methods to detect alternative DNA structures can rely on direct binding of antibodies to alternative DNA structures or indirectly via expression of tagged proteins to bind alternative DNA structures and using antibodies raised against protein tags (Figure 2) (Tables 1 and S1). These antibody-based methods have been widely used for imaging[116–129] and for high-throughput sequencing studies, such as chromatin immunoprecipitation sequencing (ChIP-seq),[130–132] DRIP-seq (DNA-RNA immunoprecipitation sequencing),[36,133,134] or CUT&Tag[129,135–137] (Tables 1 and S1). Several important conclusions have come from these studies. For example, Z-DNA is preferably observed in the upstream promoter areas of human genes and is believed to be triggered by transcription.[131,138,139] G4s are enriched in cancer genomes and at telomeres,[121] making G4s potential therapeutic targets.[140] In addition, G4s associated with RNA:DNA hybrids (G-loops) have been linked to telomeres as a characteristic of telomerase-deficient tumors that use a recombination-based telomere maintenance called alternative lengthening of telomeres (ALT).[141] G4s accumulate during the S phase of the cell cycle,[121,142] and high-resolution single-molecule imaging *in vivo* allowed for direct observation of G4 formation between the helicase protein complex Cdc45, Mcm2–7, and GINS (CMG), and DNA polymerase.[143] R-loops are involved in regulatory steps during transcription initiation and termination.[36–38,133,134] Both R-loops

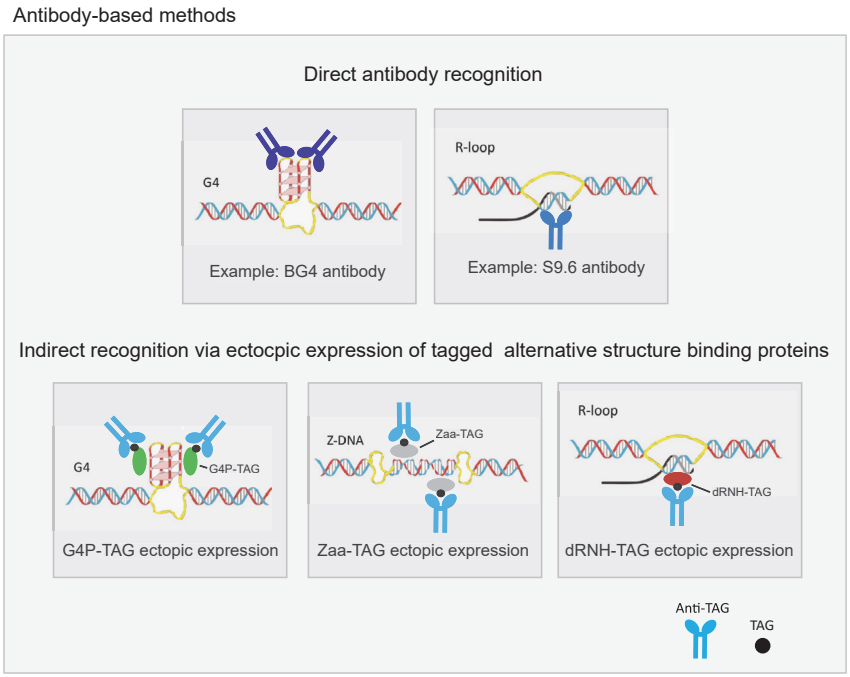

Figure 2. Antibody-based methods detection of alternative DNA structures

Top, antibodies that direct recognize alternative DNA structures. Antibodies that direct bind G4s (e.g., BG4 antibody) and RNA:DNA hybrids (e.g., S9.6 antibody) are illustrated and have been extensively used for imaging and genome-wide binding (e.g., ChIP-seq) methods. Bottom, antibody-based methods for indirect recognition of alternative DNA structures. These methods rely on ectopic expression of tagged proteins that bind alternative DNA structures. Antibody-based recognition of tagged regions allow for the score alternative DNA structures *in vivo*. G4s binding via tagged GRP, Z-DNA binding via tagged Zaa and RNA:DNA hybrid binding via tagged dRNH are exemplified and been extensively used for imaging and/or genome-wide binding (e.g., ChIP-seq) methods.

and G4s can promote DNA demethylation via their inhibitory effect on DNA methyltransferase activity at CpG island promoters.[36,133,144–146] Additionally, R-loops can threaten genome instability in specific genetic backgrounds, such as HR-deficient cells, which accumulate DNA damage due to transcription-replication conflicts associated with R-loops.[147,148]

Although antibody-based methods are used to detect alternative DNA structures, the specificity of such antibodies must be taken carefully into consideration. For example, the detection of RNA:DNA hybrids (e.g., R-loops, H-loops, or G-loops) has frequently relied on the S9.6 antibody.[165,166] However, S9.6 antibody binds to double-stranded RNA (dsRNA) *in vitro* and *in vivo*,[167,168] giving rise to non-specific signals. Furthermore, inactivation of genes associated with the resolution of RNA:DNA hybrids, *BRCA1* or *SETX*, can increase both dsRNA and RNA:DNA hybrids, making it harder to interpret results obtained with the S9.6 antibody.[167] As an alternative, several studies have used catalytically inactive RNase H1 (dRNH) that directly binds RNA:DNA hybrids (Tables 1 and S1).[169] RNase H1 binding of RNA:DNA can be achieved through ectopic expression of dRNH in cells or binding of purified enzyme and appears to be a reliable substitute to the S9.6 antibody (Figures 2 and 3).[167,169,170] A recent study used dRNH to show that deregulation of R-loop dynamics leads to the excision of RNA:DNA hybrids, which triggers the expression of pro-inflammatory immune responses.[171] Thus, alternative DNA structures not only threaten genome integrity but also trigger non-autonomous cellular responses.

Since alternative DNA structures are transient and may fold in different ways, it is critical to use orthogonal approaches to detect these structures. For example, G4s can extrude into loops of different sizes and the four-stranded guanines can bind in parallel, antiparallel, or mixed orientations.[140,173] Antibodies raised against G4s might not bind all types of DNA G4 folds and might also bind RNA G4 or RNA:DNA mixed G4s.[155,174–176] Furthermore, a G4 monoclonal antibody (1H6) was shown to cross-react with thymidine-rich regions in ssDNA.[177] As an alternative approach, small G4-binding molecules have been developed and used to visualize G4 in cells (reviewed in Varshney et al.,[140] Feng et al.,[157] Cañeque et al.,[178] and Monchaud[179]) (Tables 1 and S1). These molecules can be radiolabeled, associated with fluorophores, or detected by "click" chemistry.[140,157,178,179] Importantly, reagents that reportedly stabilize alternative structures, and are therefore utilized as positive controls for specificity or as methods to prove specific structure formation, may also have unexpected activities. For example, the G-quadruplex ligand pyridostatin was shown to poison topoisomerase 2, thereby generating DNA DSBs.[180]

Chemical and enzymatic genome-wide footprinting methods

With the advent of high-throughput sequencing, the original strategy of mapping single-stranded regions associated with alternative DNA structures has been revisited on a genomic scale (Tables 1 and S1). Kouzine et al. performed *in vivo* permanganate probing followed by S1 nuclease digestion with high-throughput sequencing to ascertain non-B-DNA presence in resting and activated mouse B cells (Figure 4).[149] The modification of bases with $KMnO_4$ *in vivo* cemented them into a single-stranded state and S1 nuclease converted them into DSBs, the ends of which were then sequenced (Figure 4).[149] Overall, high ssDNA signals were detected upstream of active genes, driven by transcriptional supercoiling. These signals coincided with non-B-DNA motifs, which included Z-DNA (~25,000 motifs), G4 (~20,000 motifs), and H-DNA (~17,000 motifs in total). This method demonstrated that ssDNA enrichment occurs in activated but not in resting B cells, therefore suggesting that they do not arise as an artifact of sample preparation.

The findings from Kouzine et al. were further validated by kethoxal-assisted ssDNA sequencing (KAS-seq) (Figure 5).[150,181] In this case, the authors used kethoxal (1,1-dihydroxy-3-ethoxy-2-butanone), which reacts with the N1 and N2 positions

Table 1. Genome-wide methods to detect alternative DNA structures *in vivo*, related to Table S1

Method	Alternative DNA structure detected	Short description	References
ssDNA-seq	Z-DNA, G4, H-DNA, and SIDD (stress-induced duplex destabilized site)	ssDNA permanganate oxidation in live cells followed by S1 nuclease cleavage of purified DNA	Kouzine et al.[149]
KAS-seq	Z-DNA, G4, H-DNA, cruciforms, and hairpins	pull-down of biotinylated guanines in single-stranded modified by N3-kethoxal	Wu et al.[150]
S1-END-seq	cruciforms and H-DNA	S1 nuclease DNA digestion from cells embedded in agarose	Matos-Rodrigues et al.[151]
P1-END-seq		P1 nuclease DNA digestion from cells embedded in agarose	Matos-Rodrigues et al.[151]
EME1-MUS81-END-seq	cruciforms	EME1-MUS81 nuclease DNA digestion of cells embedded in agarose	van Wietmarschen et al.[152]
S1-seq	H-DNA (mostly)	S1 nuclease digestion of DNA from cells embedded in agarose	Maekawa et al.[153]
G4-ChIP-seq	G4	BG4 antibody-based immunoprecipitation	Hänsel-Hertsch et al.[154]
G4P-ChIP-seq		expression of G4P (G4-binding domain of DHX36/G4R1/RHAU [G4 helicase]) in living cells, G4 capture by G4P chromatin immunoprecipitation	Zheng et al.[97]
D1-ChIP-seq		D1 antibody-based chromatin immunoprecipitation	Liu et al.[155]
G4-CUT&Tag		BG4 antibody-based CUT&Tag	Lyu et al.[135]
snG4-CUT&Tag		BG4-antibody-based single-nuclei CUT&Tag	Hui et al.[136]
SG4-CUT&Tag		G4 binding nanobody SG4-based CUT&Tag	Galli et al.[129]
G4access		MNase chromatin digestion followed by subnucleosomal fractions purification	Esnault et al.[156]
G4DP-seq		G4 binding small molecule (BioTASQ or BioCyTASQ)-based pull down	Feng et al.[157]
iM-IP-seq	i-Motif (iM)	iMab antibody-based immunoprecipitation	Ma et al.[158]
iMab-CUT&Tag		iMab antibody-based CUT&Tag	Zanin et al.[137]
DRIP-seq	DNA:RNA hybrid	S9.6 antibody-based immunoprecipitation	Ginno et al.[134]
R-ChIP		expression of tagged catalytically inactive RNaseH (dRNH), followed by tag chromatin immunoprecipatation	Chen et al.[159]
MapR		CUT&RUN-based method using catalytically inactive RNaseH1 (dRNH)-guided MNase digestion of RNA:DNA hybrids	Yan et al.[160]
R-Loop CUT&Tag		CUT&Tag-based method using N-terminal hybrid-binding domain (HBD) of RNase H1	Wang et al.[161]
S9.6-CUT&Tag-seq		CUT&Tag-based method using S9.6 antibody	Jiang et al.[162]
SMRF-seq		S9.6 antibody-based RNA:DNA hybrid enrichment followed by single-molecule footprinting using sodium bisulfite deamination of unpaired cytosines	Malig et al.[163]
spKAS-seq		strand-specific KAS-seq mapping which specifically labels only the ssDNA portion of RNA:DNA hybrids	Wu et al.[164]
Zaa-ChIP-seq	Z-DNA	expression of FLAG-tagged Z-DNA-binding protein Zaa, FLAG chromatin immunoprecipitation	Shin et al.[131]

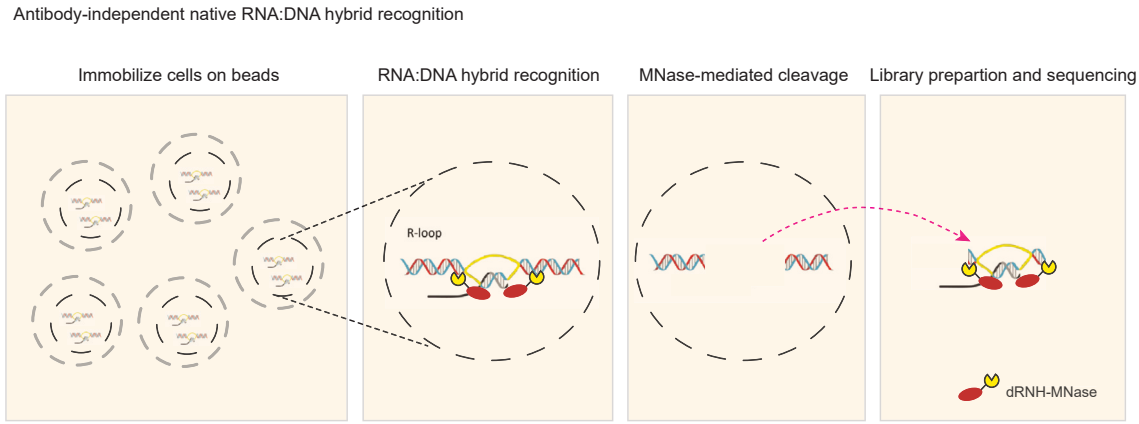

Figure 3. Antibody-independent native RNA:DNA hybrid recognition and quantification
First, cells are immobilized by magnetic beads and permeabilized. Second, a catalytic deficient mutant of RNase H fused to micrococcal nuclease (dRHN-MNase) binds to RNA-DNA hybrids in the cells. Third, MNase activation results in cleavage of DNA fragments in proximity to R-loops. Fourth, genomic regions RNA-DNA hybrid containing regions diffuse out of the cell. The dRHN-MNase cleaved DNA fragments are then recovered, sequencing libraries are prepared and submitted to high-throughput sequencing.[160,172]

of guanines in ssDNA. Kethoxal-labeled ssDNA was pulled down by biotinylation through click chemistry and enriched for high-throughput sequencing (Figure 5). Similar to what was found by Kouzine et al., this method revealed a strong correlation between ssDNA-containing regions and repetitive regions in the genome associated with alternative DNA structures, including H-DNA, cruciform, and G4s.[150] Another newly developed assay called G4access used size selection of small (<100 bp) products of micrococcal nuclease (MNase) processing of chromatin DNA to score G4s formed genome wide[156] (Figure 6).

Although much has been learned with these genome-wide mapping tools, they are generally low-resolution approaches. This makes it difficult to faithfully test models proposing how alternative structures are folded *in vivo*. In addition, as different repeat motifs and ssDNA regions may overlap, the ssDNA-labeling techniques frequently do not distinguish which alternative DNA structure is formed.

Nucleotide resolution methods for genome-wide detection of alternative DNA structures *in vivo*

In 2016, END-seq was developed to quantify DSBs genome wide at single-nucleotide resolution. In this method, cells are embedded in agarose plugs and subjected to protein and RNA degradation.[182,183] The DNA ends are then blunted by exonucleases (ExoVII/T) and then sequenced (Figure 2). END-seq has been used to quantify recurrent DSBs formed genome wide in several contexts, including endogenous/programmed DSBs during lymphocyte development and meiosis,[182,184–186] as well as to identify hotspots of DNA breakage upon treatment with genotoxic agents (e.g., etoposide and hydroxyurea).[186–188] More recently, recombinant nucleases were used to convert alternative DNA structures and endogenous sites of ssDNA breaks genome wide into DSBs, which were then mapped at single-nucleotide resolution by END-seq.[151,152,189,190]

Gene inactivation screens revealed that the WRN helicase (defective in Werner syndrome) is essential for the survival of mismatch repair (MMR)-deficient microsatellite unstable (MSI) colon cancer cells.[191–194] Using END-seq mapping, it was determined that WRN inactivation in MSI cells resulted in chromosome breakage at thousands of recurrent DSB sites.[152] Breakage was mediated by the structure-specific nuclease MUS81 targeted to $(TA)_n$ repeats that expanded specifically in MSI cells.[152] A recent study found no evidence of any generated mechanism that could confer resistance to WRN deficiency in MSI cancer cells.[195] Given that $(TA)_n$ repeats can form cruciform structures and that MUS81 can process such structures,[196–199] it was hypothesized that the formation of cruciform at expanded $(TA)_n$ repeats required WRN helicase unwinding. Indeed, both recombinant EME1-MUS81 and S1 nuclease treatment *in situ* converted expanded sites of $(TA)_n$ repeats into DSBs[152,151] (Figure 7A) (Tables 1 and S1). Notably, structure-forming expanded $(TA)_n$ repeats were highly recalcitrant to short-read sequencing and required long-read technology for their detection.[152]

In addition to revealing cruciform structures formed at expanded $(TA)_n$ repeats, S1-END-seq detected H-DNA structures in several MSI and non-MSI human cell lines[151] (Figure 7B). H-DNA peaks localized to long (averaging 200 bp) H-DNA motifs including $(GAAA)_n$, $(GGAA)_n$, and $(GAA)_n$.[151] H-DNA was formed during S phase, enhanced upon the induction of replication stress, and enriched in transformed cells and hotspots for genome instability. Accurately quantitating the absolute abundance of alternate DNA structures within the genome has been challenging due to the fact that most methods (e.g., ChIP-seq) lack adequate internal controls for normalization. In an attempt to estimate the absolute number of H-DNA structures formed per cell, pre-B cells harboring a single zinc-finger-induced DSB at the T cell receptor beta (TCRβ) enhancer[182] were used as a spike-in for S1-END-seq samples. By assuming that each spike-in mouse pre-B cell harbored a zinc-finger-induced DSB, it was possible to estimate that ~300 H-DNA structures were present at a given time within the KM12 cell line.[151] A similar normalization approach was used to provide DSB quantitation via another WGS method that scores DSBs called iBLESS.[200]

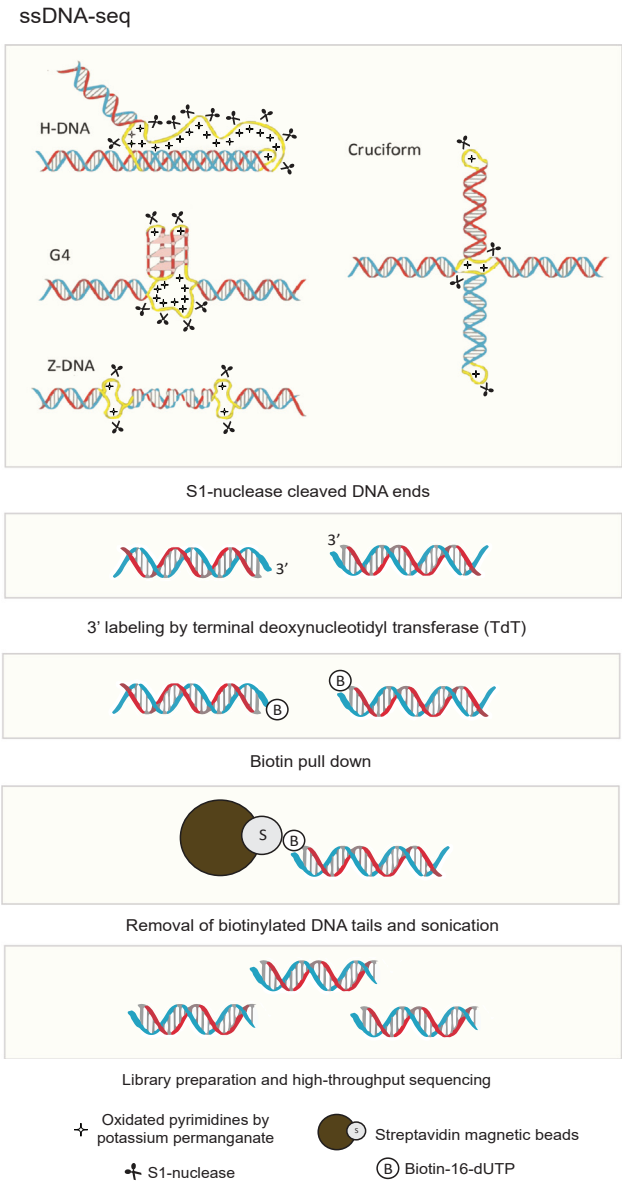

Figure 4. Genome-wide mapping of alternative DNA structures genome-wide via ssDNA-seq

Genomic regions containing ssDNA are stabilized in live cells via pyrimidine oxidation by potassium permanganate ($KMnO_4$), the DNA is then purified and treated with S1 nuclease. ssDNA-containing regions, including the ones in present in alternative DNA structures (highlighted in yellow), processed by S1 nuclease create 3′ DNA ends. Next, the 3′ DNA ends are labeled via incorporation of biotin-13-dUTP(2'-deoxyuridine, 5'-triphosphate) by terminal deoxynucleotidyl transferase (TdT). DNA ends containing biotin-13-dUTP are enriched using streptavidin magnetic beads. The biotinylated DNA tails are then removed, and the DNA submitted to sonication, followed by library preparation and high-throughput sequencing.[149]

Maekawa et al. used a similar method (S1-seq) to interrogate whether H-DNA structures are formed in mouse cells using mouse testis samples or primary B cell cultures[153] (Tables 1 and S1). This study revealed S1-induced breaks mapped to relatively short, mostly $(GA)_n$, H-DNA motifs. S1-END-seq and S1-seq have similar limitations. Both methods rely on protein and RNA degradation in plugs that can destabilize alternative DNA structures. We speculate that this might be the reason why both S1-END-seq and S1-seq do not score G4s, as these structures might be destabilized during by sample preparation. Furthermore, S1-END-seq or S1-seq requires the formation of blunted DNA ends for the ligation of sequencing adaptors. Thus, these methods are only able to visualize structures in which S1 cleavage creates substrates suitable for sequencing. As an example, we illustrate the potential processing of a S-DNA/hairpin, which would not be able to generate S1-END-seq peaks (Figure 7C). The conditions in which the enzymatic reactions are performed during S1-END-seq or S1-seq must be carefully controlled. For example, low pH, which is necessary for optimal S1 nuclease activity, can drive H-DNA formation.[23] This can be avoided using the single-strand-specific P1 nuclease that shows optimal activity in neutral pH.[151] These limitations must be taken into account when interpreting results from S1-END-seq and similar experimental procedures. The development of new methods to reveal additional structures (e.g., S-DNA/hairpins and G4s) at single-nucleotide resolution is necessary to uncover the full spectrum of alternative DNA structures formed *in vivo*.

Recently, a pipeline using enzymatic or chemical probing followed by long-read sequencing has been harnessed to decipher RNA secondary and even tertiary structures.[201–203] Additional chemical probes and bioinformatic tools may allow for the adaptation of this technology to DNA secondary structures, potentially addressing some of the drawbacks of current techniques.

Computational approaches to non-B-DNA discovery

Experimental techniques for non-B-DNA detection have been progressing simultaneously with computational technologies for non-B-DNA prediction. At first, computational tools to predict non-B-DNA formation genome wide largely relied on their motifs to determine DNA sequences that have the ability to form a certain non-B-DNA structure.[204–207] This analysis is, however, complicated by the fact that the majority of non-B-DNA sequence motifs coincide with simple-tandem repeats, which can simultaneously be inverted or mirror repeats, or both and are enormously over-represented in eukaryotic genomes. Not surprisingly, therefore, only a small percentage of those motifs actually form the structure *in vivo*, according to current *in vivo* detection techniques. For example, only about 1% of the sequences predicted to form G4-DNA were detected *in vivo* via G4-ChIP-seq,[154] and of the ~50,000 H-DNA motifs in the human genome, only 3,110 repeats form H-DNA recurrently.[151] Recently, deep learning approaches have been applied to non-B-DNA structure discovery. These methods have improved with the existence of *in vivo* structure-mapping datasets. For example, DeepG4 uses a convolutional neural network trained using *in vitro* G4-seq and *in vivo* G4-ChIP-seq data to predict G4s genome wide.[208] Similarly, DeepZ uses a recurrent neural network trained using the *in vivo* structure detection methods of Z-DNA ChIP-seq, ssDNA-seq, and KAS-seq.[209] Very recently, a non-B-DNA detection computational approach was established that harnesses the differential timing of B-DNA and non-B-DNA translocation through a nanopore using whole-genome nanopore sequencing data.[210] Importantly, future developments in prediction tools must incorporate key aspects of cell physiology, including chromatin state, transcription,

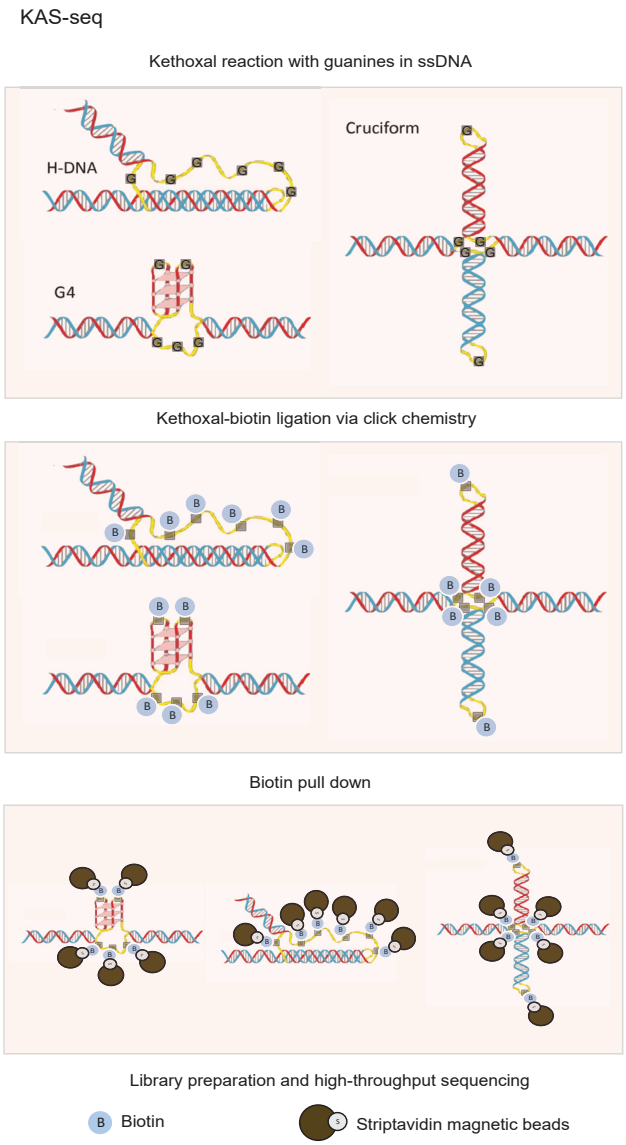

Figure 5. Genome-wide mapping of alternative DNA structures via KAS-seq

ssDNA is labeled in live cells by kethoxal reaction with exposed guanine bases (highlighted by the gray boxes). The genomic DNA is then purified and kethoxal-modified guanines are biotinylated through "click" cycloaddition with a biotin-conjugated alkyne. Next, biotinylated DNA fragments are sonicated and enriched using streptavidin beads. Libraries are prepared using the resulting purified DNA and submitted to high-throughput sequencing.[150,181]

and replication, which have been experimentally demonstrated to influence the likelihood of a given motif to fold into alternative DNA structures *in vivo*.

ALTERNATIVE DNA STRUCTURES AND GENOME INSTABILITY

Given the historical paucity of reliable methods to detect alternative DNA structures *in vivo*, researchers first focused on the biological roles of structure-prone motifs as a proxy for the structures themselves. Although these non-B motifs have been associated with several physiological processes, they have also been suggested to play pathological roles that can threaten genome integrity (reviewed in Brown and Freudenreich,[28] Wang and Vasquez,[29] and Khristich and Mirkin[30]).

It was experimentally demonstrated by us and others that structure-prone DNA repeats can cause chromosomal fragility and trigger chromosomal rearrangements, including deletions, duplications, inversions, translocations, and complex chromosomal rearrangements (reviewed in Brown and Freudenreich,[28] Wang and Vasquez,[29] and Khristich and Mirkin[30]). Strikingly, these repeats also elevate mutation rates in adjacent DNA segments, a phenomenon that we called repeat-induced mutagenesis (RIM) (reviewed in Shah and Mirkin[211]). Comparative genomics data are in line with these conclusions. For example, deletion and translocation breakpoints found in human disease and cancer genomes are associated with non-B-DNA motifs.[28,105,212,213] They also appear to have a higher density of nucleotide variants in the human genome[214,215] and somatic cancer genomes.[31] The latter conclusion may need additional verification, taking into account the high rate of sequencing errors at non-B-DNA motifs as well as overlapping and interrupted motifs.[216,217]

Recently, Erwin et al. conducted a genome-wide study of tandem repeat expansions in cancer.[218] Using WGS by short-read data from 2,622 cancer genomes encompassing 29 different cancer types, they identified 160 recurrent repeat expansions (rREs) in 7 cancer types. Notably, H-DNA-forming motifs were the most frequent type of rREs.[218]

Structure-prone repeats are hard to replicate, leading to replication fork stalling (reviewed in Khristich and Mirkin,[30] Técher et al.,[219] and Zell et al.[220]). Numerous tools have been developed to assess fork stalling at these repeats *in vivo*, including two-dimensional gel electrophoresis of replication intermediates,[221] DNA fiber analysis,[222] and fluorescent array signal monitoring.[223] Although the stalling at a structure-forming motif can be used as a surrogate for structure formation, the actual visualization of replication through the structures themselves, rather than motifs with structure-forming potential, is only beginning to come into focus.[224] Importantly, not only the formation of alternative DNA structures per se can challenge DNA stability, as the formation of protein aggregates bound to alternative DNA structures[225,226] can create "protein bumps" that block replisome progression. To counteract genome instability, multiple helicases have been shown to aid in the progression of polymerization through non-B-DNA structures *in vitro* and non-B-DNA motifs *in vivo*.[29,30] These include genes mutated in human developmental syndromes, such as Bloom syndrome (BLM), Werner syndrome (WRN), Warsaw breakage syndrome (DDX11), and Fanconi anemia (FANCJ).[29] Importantly, resolution of alternative DNA structures might become essential for cell survival. As mentioned above, WRN helicase plays an essential function in protecting MSI cells against DSBs generated by DNA cruciform formed in expanded $(TA)_n$ repeats.[152]

ALTERNATIVE DNA STRUCTURES AND HUMAN DISEASE

More than 60 REDs have been reported, and despite their differences, there are many commonalities between REDs: most are

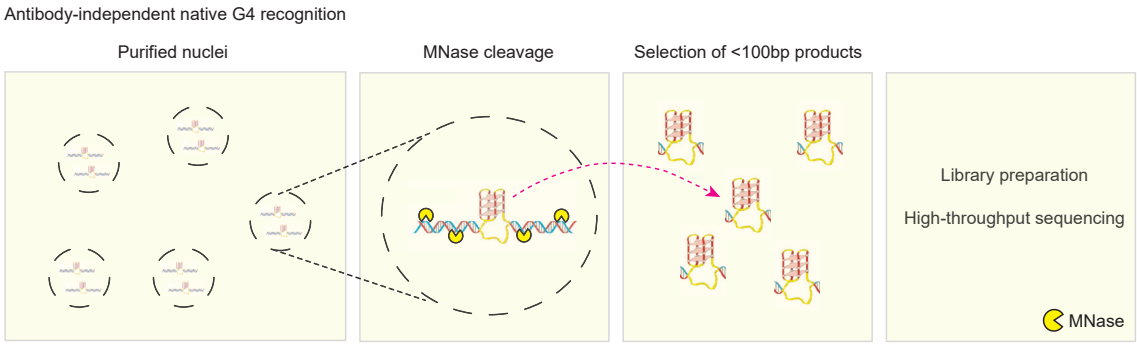

Figure 6. Antibody-independent native G4 recognition and quantification

First, cell nuclei are purified. Second, micrococcal nuclease (MNase) bind to accessible genomic regions. Third, MNase activation results in cleavage of DNA fragments and the selection of small products <100 bp allows for the enrichment of genomic regions folded in G4s. Fourth, the small (<100 bp) MNase cleaved DNA fragments are then recovered, sequencing libraries are prepared and submitted to high-throughput sequencing.[156]

rare, neurodegenerative diseases with no known cure.[30,44] REDs can be caused by expanded repeats in the noncoding or coding regions of the genome and may be dominantly or recessively inherited.[30,44] Many proposed mechanisms of repeat instability are grounded in the ability of the disease-causing repeats' to form alternative DNA structures (reviewed in Khristich and Mirkin[30] and Mirkin[227]). These models implicate imperfect hairpin and/or S-DNA formation in HD's $(CAG)_n$ and $(CTG)_n$ expansion process[228,229] and triplex/H-DNA formation in FRDA's $(GAA)_n$ repeat expansion mechanism.[230–232] The ability of repeat interruptions, hindering structure formation, to stabilize expandable repeats supports the role of structure formation in the repeat expansion process (reviewed in Khristich and Mirkin[30]).

An established model for repeat instability involves the aberrant processing of alternative DNA structures by DNA repair and/or replication machinery, where MMR and NER factors are the two most studied pathways associated with structure processing.[29,233]

Depending on the genomic location and composition of the repeat that causes disease, different pathogenic mechanisms are at play. Some include loss of function, such as in the recessively inherited FRDA, or toxic gain of function at the RNA or protein level, such as in the dominantly inherited HD.[30,44] In fact, RED patient cells show accumulation of RNA:DNA hybrids at expanded repeats, which are linked to gene silencing in FRDA and FXS.[234,235] DNA triplexes form *in vivo* in lymphoblastoid cell lines from FRDA patients who contain expanded $(GAA)_n$ repeats.[151] These triplexes are destabilized by a polyamide compound that binds to $(GAA)_n$ repeats *in vivo*, confirming similar data generated using triplex-forming plasmids.[236] Interestingly, polyamide binding to $(GAA)_n$ repeats has been shown to decrease repeat expansion in FRDA-induced pluripotent stem cells (iPSCs), thereby suggesting that H-DNA may participate in repeat expansion.[237] Alternatively, H-DNA may be linked to the downregulation of *FXN* in FRDA. Supporting this, $(GAGGA)_n$ repeat, which does not form triplexes and is more stable than $(GAA)_n$ repeat, leads to only a very mild and late-onset disease.[238,239]

Disease-causing repeat expansions in REDs can reach thousands of repeats, making these repeats impossible to sequence via short-read sequencing. The advent of long-read sequencing in the recent past has been key to expanding the tool kit to investigate repeat instability and discover new types of REDs. Indeed, long-read sequencing techniques have revealed previously unknown causes of several hereditary human diseases.[44,240] For example, CANVAS (cerebellar ataxia neuropathy vestibular areflexia syndrome), which is one of the most common recessive hereditary ataxias,[241–243] was found to be caused by the expansion of $(AAGGG)_n$ repeats in the second intron of the *RFC1* gene.[241,242] Very recently, we confirmed that $(AAGGG)_n$ motifs form H-DNA *in vitro* and *in vivo* and block DNA replication.[244] Another example is the late-onset SCA27B—one of the most frequent forms of spinocerebellar ataxias,[245,246] which is caused by the $(GAA)_n$ repeat expansion in an intron of the *FGF14* gene. Similar to FRDA, $(GAA)_n$ repeat expansions in this case cause gene silencing in neurons.[246]

Given this recent trend, it is fair to assume that the increased accessibility of long-read sequencing will lead to the progressive discovery of new REDs. Similarly, long-read sequencing of cancer genomes may open a new spectrum of repeat instabilities in cancer, which will likely amplify the number of rREs recently found by short-read sequencing.[218] In a broader sense, many pathologies with unknown genetic causes may be attributable to long, structure-prone DNA repeats. Improvement of long-read sequencing technologies to increase the sequencing depth per run, increase the speed of sequencing, and reduce costs per sample will be key for the widespread use of these methods. Understanding the etiology of these diseases is still in its infancy, but it is likely that the formation of alternative DNA structures might contribute significantly to human pathologies.

TARGETING ALTERNATIVE DNA STRUCTURES IN HUMAN DISEASE

The stabilization of alternative DNA structures has been shown to be potential targets for oncotherapy.[247–250] This has been particularly explored for G4 stabilizers but recently Z-DNA and S-DNA hairpin structures as well. Indeed, a Z-DNA-binding molecule has been shown to be a good candidate for cancer therapy associated with anti-PD1 immunotherapy[109] and a series of azacryptands that bind S-DNA/hairpin junctions are toxic to cancer cells.[251,252] G4-stabilizing ligands induce DNA

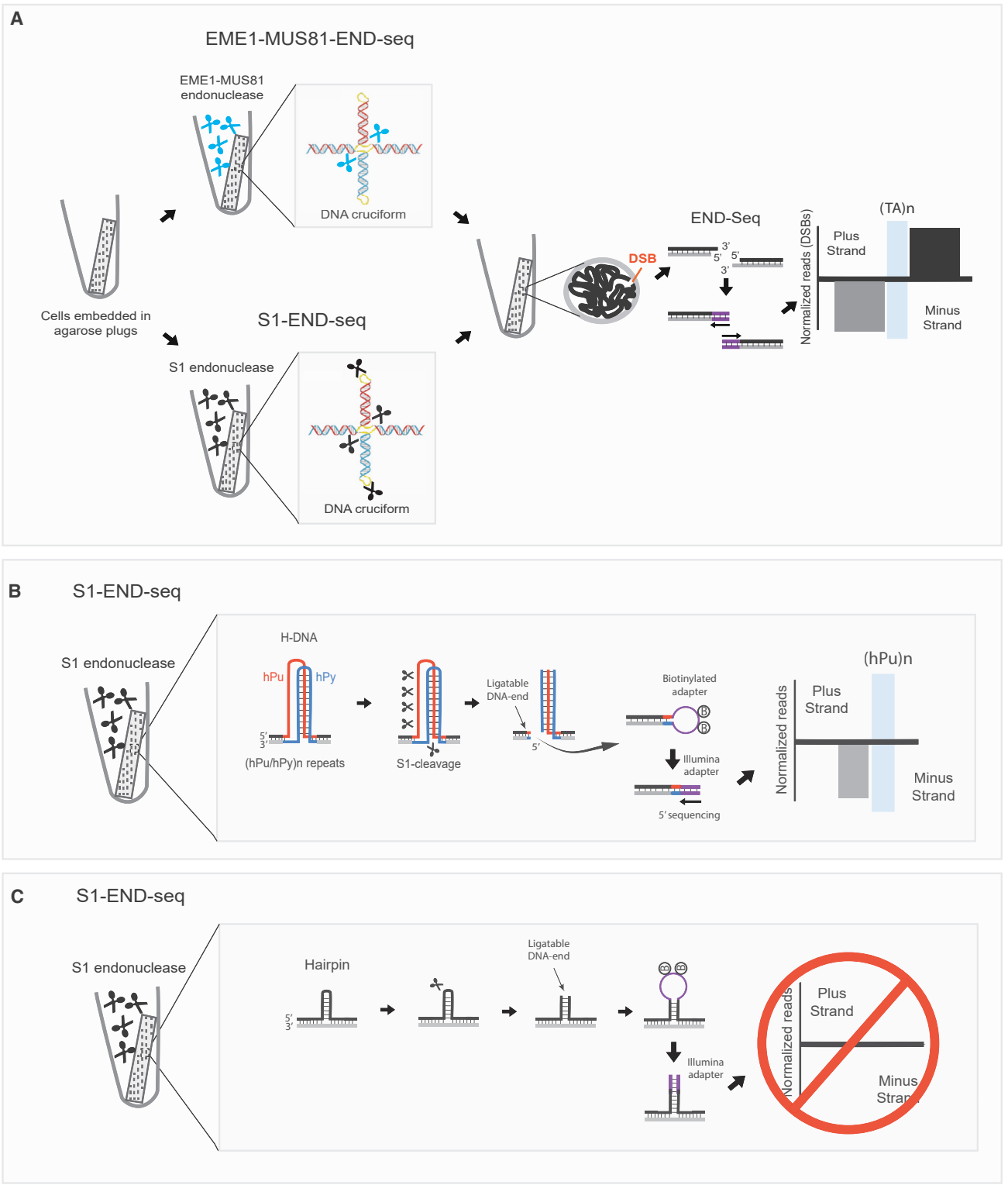

Figure 7. Genome-wide mapping of alternative DNA structures at single-nucleotide resolution using END-seq

(A) Detection of DNA cruciform in $(TA)_n$ repeats via EME1-MUS81-END-seq (top) or S1-END-seq (bottom). Cells are embedded in agarose, forming plugs, then (top) EME1-MUS81 endonuclease converts four-way junctions of DNA cruciforms present in long $(TA)_n$ repeats into DSBs or (bottom) S1 nuclease converts overall ssDNA regions including regions of unpaired bases in the loop or base of DNA cruciforms into DSBs. The ends are ligated to biotinylated adaptors. After DNA sonication, DSBs are captured by streptavidin magnetic beads, Illumina sequencing adaptors are added to the DNA ends, and the samples are subjected to

(legend continued on next page)

damage in several cell lines. This is likely caused by direct effects as ChIP-seq for the DSB marker γH2AX revealed DNA breaks at G4-DNA-forming repeats upon treatment with the G4 ligand pyridostatin.[253] Similarly, ChIP-seq for RAD51, the recombinase responsible for DSB repair via HR, identified thousands of genomic loci undergoing DSB repair after treatment with the G4 ligand CX-5461.[254] The induction of DNA damage by G4 stabilizers is often associated with cell death, and G4 stabilizers are particularly toxic to HR-deficient cancer cells independent of whether they are PARP-inhibitor sensitive or resistant.[254–256] CX-5461 is currently in phase 1 for clinical trials for patients with solid tumors and BRCA1/2, PALB2, or HR deficiency (Clinical trial ID: NCT04890613).[257,258]

The development and full characterization of small molecules targeting alternative DNA structures are promising new avenues for cancer therapy. Understanding the mechanisms of alternative DNA structure formation and processing will be key to fully depicture the potential of these molecules.

Alternative DNA structures are also promising targets for RED therapies. Using a compound called naphthyridine-azaquinolone (NA), which specifically binds and stabilizes S-DNA/hairpins formed in $(CAG)_n$ repeats, Nakamori et al. found that hairpin stabilization can induce repeat contractions in HD patient cells as well as in spiny neurons of HD mouse striatum.[259] Furthermore, NA treatment-induced contractions and improved motor function in a mouse model for dentatorubral-pallidoluysian atrophy (DRPLA), which is also caused by the expansion of a $(CAG)_n$ repeat in the *ATN1* gene.[260] Another recent study demonstrated that reactivation of *FMR1* in FXS patient cells create R-loops that fuel repeat contractions, reestablishing *FMR1* expression.[261] Very recently, it was shown that locked nucleic acid (LNA)-DNA oligonucleotides and peptide nucleic acid (PNA) oligomers targeting FRDA's expanded $(GAA)_n$ repeats interfere with triplex formation and prevent repeat expansion in an experimental human system.[262]

Overall, these studies provide an exciting new line of treatment for REDs, either by destabilizing alternative structures or by fueling alternative DNA structure formation, which allows the endogenous cellular machinery to delete the pathogenic expanded repeats.

CONCLUSIONS

Although the biological significance of alternative DNA structures was initially brought into question, recent technological advancements confirmed their presence *in vivo* and strongly correlated them with genome instability. By and large, non-B-DNA sequence motifs can be considered genomic weak links. In some instances, this instability benefits important physiological processes, as exemplified by CSR. In other instances, it causes hereditary human diseases, as exemplified by REDs, or contributes to cancer development. Notably, since the instability of structure-prone DNA repeats increases with age, REDs are generally late-onset diseases, which could explain the lack of evolutionary counter-selection against them. Future studies exploring the function of alternative DNA structures *in vivo* should aim to go beyond correlations to establish direct causal links between alternative DNA structure formation and genome instability in health and disease. Based on the recent progress described here, it is also fair to assume that targeting alternative DNA structures *in vivo* might provide new therapeutic strategies for related human diseases.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2023.08.018.

ACKNOWLEDGMENTS

We apologize to those colleagues whose work has not been cited due to space limitations. The work in the Mirkin laboratory is supported by the National Institute of General Medical Sciences (R35GM130322), the National Science Foundation (2153071), and by a generous contribution from the White family. The Nussenzweig laboratory is supported by the Intramural Research Program of the NIH funded in part with federal funds from the NCI under contract HHSN2612015000031, an Ellison Medical Foundation Senior Scholar in Aging Award (AG-SS-2633-11), the Department of Defense Awards (W81XWH-16-1-599 and W81XWH-19-1-0652), the Alex's Lemonade Stand Foundation Award, an NIH Intramural FLEX Award, and a research grant by Friedreich's Ataxia Research Alliance.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Wilkins, M.H., Stokes, A.R., and Wilson, H.R. (1953). Molecular structure of deoxypentose nucleic acids. Nature *171*, 738–740. https://doi.org/10.1038/171738a0.

2. Franklin, R.E., and Gosling, R.G. (1953). Molecular configuration in sodium thymonucleate. Nature *171*, 740–741. https://doi.org/10.1038/171740a0.

3. Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature *171*, 737–738. https://doi.org/10.1038/171737a0.

4. Fuller, W., Wilkins, W.H.F., Wilson, H.R., and Hamilton, L.D. (1965). The molecular configuration of deoxyribonucleic acid. IV. X-ray diffraction study of the A form. J. Mol. Biol. *12*, 60–76. https://doi.org/10.1016/s0022-2836(65)80282-0.

sequencing. Left end reads are aligned to the minus strand, and right end reads are aligned to the plus strand. A two-ended peak created from the processing of DNA cruciform into DSBs is illustrated.

(B) Genome-wide detection of H-DNA/triplexes via S1-END-seq. Cells are embedded in agarose, forming plugs, then S1 nuclease converts ssDNA regions, (upper right) including H-DNA structures, genome wide into DSBs, and the DSB ends are ligated to biotinylated adaptors. After DNA sonication, DSBs are captured by streptavidin magnetic beads, Illumina sequencing adaptors are added to the DNA ends, and the samples are subjected to sequencing. Left end reads are aligned to minus strand, and right end reads are aligned to the plus strand. A one-ended peak created from H-DNA/triplexes formed in homopurine/homopyrimidine (hPu/hPy) repeats which were processed into DSBs is illustrated. Triple helix-containing ends are unsuitable for adaptor ligation; thus H-DNA/triplexes generate a one-ended peak.

(C) Limitations of S1-END-seq in detecting certain alternative DNA structures, such as S-DNA/hairpins. After processing of hairpin/s-DNA ssDNA loop, the ligation of adaptors into the DSB end does not generate mappable reads.

5. Marvin, D.A., Spencer, M., Wilkins, M.H., and Hamilton, L.D. (1961). The molecular configuration of deoxyribonucleic acid III. X-ray diffraction study of the C form of the lithium salt. J. Mol. Biol. *3*, 547–565. https://doi.org/10.1016/s0022-2836(61)80021-1.

6. Felsenfeld, G., Davies, D.R., and Rich, A. (1957). Formation of a three-stranded polynucleotide molecule. J. Am. Chem. Soc. *79*, 2023–2024. https://doi.org/10.1021/ja01565a074.

7. Felsenfeld, G., and Rich, A. (1957). Studies on the formation of two- and three-stranded polyribonucleotides. Biochim. Biophys. Acta *26*, 457–468. https://doi.org/10.1016/0006-3002(57)90091-4.

8. Hoogsteen, K. (1959). The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. Acta Cryst. *12*, 822–823. https://doi.org/10.1107/S0365110X59002389.

9. Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. Acta Cryst. *16*, 907–916. https://doi.org/10.1107/S0365110X63002437.

10. Riley, M., and Maling, B. (1966). Physical and chemical characterization of two- and three-stranded adenine-thymine and adenine-uracil homopolymer complexes. J. Mol. Biol. *20*, 359–389. https://doi.org/10.1016/0022-2836(66)90069-6.

11. Morgan, A.R., and Wells, R.D. (1968). Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. J Mol Biol *37*, 63–80. https://doi.org/10.1016/0022-2836(68)90073-9.

12. Lee, J.S., Johnson, D.A., and Morgan, A.R. (1979). Complexes formed by (pyrimidine)n. (purine)n DNAs on lowering the pH are three-stranded. Nucleic Acids Res. *6*, 3073–3091. https://doi.org/10.1093/nar/6.9.3073.

13. Felsenfeld, G., and Miles, H.T. (1967). The physical and chemical properties of nucleic acids. Annu. Rev. Biochem. *36*, 407–448. https://doi.org/10.1146/annurev.bi.36.070167.002203.

14. Michelson, A.M., Massoulié, J., and Guschlbauer, W. (1967). Synthetic polynucleotides. Prog. Nucleic Acid Res. Mol. Biol. *6*, 83–141. https://doi.org/10.1016/s0079-6603(08)60525-5.

15. Ralph, R.K., Connors, W.J., and Khorana, H.G. (1962). Secondary structure and aggregation in deoxyguanosine oligonucleotides. J. Am. Chem. Soc. *84*, 2265–2266. https://doi.org/10.1021/ja00870a055.

16. Gellert, M., Lipsett, M.N., and Davies, D.R. (1962). Helix formation by guanylic acid. Proc. Natl. Acad. Sci. USA *48*, 2013–2018. https://doi.org/10.1073/pnas.48.12.2013.

17. Miles, H.T., and Frazier, J. (1978). Poly(l) helix formation. Dependence on size-specific complexing to alkali metal ions. J. Am. Chem. Soc. *100*, 8037–8038. https://doi.org/10.1021/ja00493a058.

18. Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G., and Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. Nature *282*, 680–686. https://doi.org/10.1038/282680a0.

19. Singleton, C.K., Klysik, J., Stirdivant, S.M., and Wells, R.D. (1982). Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions. Nature *299*, 312–316. https://doi.org/10.1038/299312a0.

20. Panayotatos, N., and Wells, R.D. (1981). Cruciform structures in supercoiled DNA. Nature *289*, 466–470. https://doi.org/10.1038/289466a0.

21. Lilley, D.M. (1980). The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. Proc. Natl. Acad. Sci. USA *77*, 6468–6472. https://doi.org/10.1073/pnas.77.11.6468.

22. Lyamichev, V.I., Mirkin, S.M., and Frank-Kamenetskii, M.D. (1986). Structures of homopurine-homopyrimidine tract in superhelical DNA. J. Biomol. Struct. Dyn. *3*, 667–669. https://doi.org/10.1080/07391102.1986.10508454.

23. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A., and Frank-Kamenetskii, M.D. (1987). DNA H form requires a homopurine-homopyrimidine mirror repeat. Nature *330*, 495–497. https://doi.org/10.1038/330495a0.

24. Sen, D., and Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature *334*, 364–366. https://doi.org/10.1038/334364a0.

25. Panyutin, I.G., Kovalsky, O.I., and Budowsky, E.I. (1989). Magnesium-dependent supercoiling-induced transition in (dG)n.(dC)n stretches and formation of a new G-structure by (dG)n strand. Nucleic Acids Res. *17*, 8257–8271. https://doi.org/10.1093/nar/17.20.8257.

26. Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell *59*, 871–880. https://doi.org/10.1016/0092-8674(89)90610-7.

27. Sundquist, W.I., and Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature *342*, 825–829. https://doi.org/10.1038/342825a0.

28. Brown, R.E., and Freudenreich, C.H. (2021). Structure-forming repeats and their impact on genome stability. Curr. Opin. Genet. Dev. *67*, 41–51. https://doi.org/10.1016/j.gde.2020.10.006.

29. Wang, G., and Vasquez, K.M. (2023). Dynamic alternative DNA structures in biology and disease. Nat. Rev. Genet. *24*, 211–234. https://doi.org/10.1038/s41576-022-00539-9.

30. Khristich, A.N., and Mirkin, S.M. (2020). On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. J. Biol. Chem. *295*, 4134–4170. https://doi.org/10.1074/jbc.REV119.007678.

31. Georgakopoulos-Soares, I., Morganella, S., Jain, N., Hemberg, M., and Nik-Zainal, S. (2018). Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. Genome Res. *28*, 1264–1271. https://doi.org/10.1101/gr.231688.117.

32. Gehring, K., Leroy, J.L., and Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine.cytosine base pairs. Nature *363*, 561–565. https://doi.org/10.1038/363561a0.

33. Kowalski, D., and Eddy, M.J. (1989). The DNA unwinding element: a novel, cis-acting component that facilitates opening of the Escherichia coli replication origin. EMBO J. *8*, 4335–4344. https://doi.org/10.1002/j.1460-2075.1989.tb08620.x.

34. Pearson, C.E., and Sinden, R.R. (1996). Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. Biochemistry *35*, 5041–5053. https://doi.org/10.1021/bi9601013.

35. Gacy, A.M., Goellner, G., Juranić, N., Macura, S., and McMurray, C.T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell *81*, 533–540. https://doi.org/10.1016/0092-8674(95)90074-8.

36. Niehrs, C., and Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. Nat. Rev. Mol. Cell Biol. *21*, 167–178. https://doi.org/10.1038/s41580-019-0206-3.

37. Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-loops as cellular regulators and genomic threats. Mol. Cell *73*, 398–411. https://doi.org/10.1016/j.molcel.2019.01.024.

38. García-Muse, T., and Aguilera, A. (2019). R loops: from physiological to pathological roles. Cell *179*, 604–618. https://doi.org/10.1016/j.cell.2019.08.055.

39. Neil, A.J., Liang, M.U., Khristich, A.N., Shah, K.A., and Mirkin, S.M. (2018). RNA-DNA hybrids promote the expansion of Friedreich's ataxia (GAA)n repeats via break-induced replication. Nucleic Acids Res. *46*, 3487–3497. https://doi.org/10.1093/nar/gky099.

40. Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. *18*, 1618–1629. https://doi.org/10.1101/gad.1200804.

41. 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073. https://doi.org/10.1038/nature09534.

42. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580. https://doi.org/10.1093/nar/27.2.573.

43. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P., and Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. *33*, 877–881. https://doi.org/10.1038/nbt.3295.

44. Gall-Duncan, T., Sato, N., Yuen, R.K.C., and Pearson, C.E. (2022). Advancing genomic technologies and clinical awareness accelerates discovery of disease-associated tandem repeat sequences. Genome Res. *32*, 1–27. https://doi.org/10.1101/gr.269530.120.

45. Nordheim, A., and Rich, A. (1983). Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature *303*, 674–679. https://doi.org/10.1038/303674a0.

46. Kmiec, E.B., Angelides, K.J., and Holloman, W.K. (1985). Left-handed DNA and the synaptic pairing reaction promoted by Ustilago rec1 protein. Cell *40*, 139–145. https://doi.org/10.1016/0092-8674(85)90317-4.

47. Kmiec, E.B., and Holloman, W.K. (1986). Homologous pairing of DNA molecules by Ustilago rec1 protein is promoted by sequences of Z-DNA. Cell *44*, 545–554. https://doi.org/10.1016/0092-8674(86)90264-3.

48. Blaho, J.A., and Wells, R.D. (1989). Left-handed Z-DNA and genetic recombination. Prog. Nucleic Acid Res. Mol. Biol. *37*, 107–126. https://doi.org/10.1016/s0079-6603(08)60696-0.

49. Kitts, P., Richet, E., and Nash, H.A. (1984). Lambda integrative recombination: supercoiling, synapsis, and strand exchange. Cold Spring Harb. Symp. Quant. Biol. *49*, 735–744. https://doi.org/10.1101/sqb.1984.049.01.083.

50. Gough, G.W., and Lilley, D.M. (1985). DNA bending induced by cruciform formation. Nature *313*, 154–156. https://doi.org/10.1038/313154a0.

51. Parsons, C.A., and West, S.C. (1988). Resolution of model Holliday junctions by yeast endonuclease is dependent upon homologous DNA sequences. Cell *52*, 621–629. https://doi.org/10.1016/0092-8674(88)90474-6.

52. Lilley, D.M., and Kemper, B. (1984). Cruciform-resolvase interactions in supercoiled DNA. Cell *36*, 413–422. https://doi.org/10.1016/0092-8674(84)90234-4.

53. Tenen, D.G., Haines, L.L., Hansen, U.M., Martin, R.G., and Livingston, D.M. (1985). Formation of a cruciform structure at the simian virus 40 replication origin abolishes T-antigen binding to the origin in vitro. J. Virol. *56*, 293–297. https://doi.org/10.1128/JVI.56.1.293-297.1985.

54. Frappier, L., Price, G.B., Martin, R.G., and Zannis-Hadjopoulos, M. (1989). Characterization of the binding specificity of two anticruciform DNA monoclonal antibodies. J. Biol. Chem. *264*, 334–341.

55. Nobile, C., Nickol, J., and Martin, R.G. (1986). Nucleosome phasing on a DNA fragment from the replication origin of simian virus 40 and rephasing upon cruciform formation of the DNA. Mol. Cell. Biol. *6*, 2916–2922. https://doi.org/10.1128/mcb.6.8.2916-2922.1986.

56. Hoffman, E.K., Trusko, S.P., Freeman, N., and George, D.L. (1987). Structural and functional characterization of the promoter region of the mouse c-Ki-ras gene. Mol. Cell. Biol. *7*, 2592–2596. https://doi.org/10.1128/mcb.7.7.2592-2596.1987.

57. Evans, T., DeChiara, T., and Efstratiadis, A. (1988). A promoter of the rat insulin-like growth factor II gene consists of minimal control elements. J. Mol. Biol. *199*, 61–81. https://doi.org/10.1016/0022-2836(88)90379-8.

58. Pestov, D.G., Dayn, A., Siyanova, E., George, D.L., and Mirkin, S.M. (1991). H-DNA and Z-DNA in the mouse c-Ki-ras promoter. Nucleic Acids Res. *19*, 6527–6532. https://doi.org/10.1093/nar/19.23.6527.

59. Rao, B.S., Manor, H., and Martin, R.G. (1988). Pausing in simian virus 40 DNA replication by a sequence containing (dG-dA)27.(dT-dC)27. Nucleic Acids Res. *16*, 8077–8094. https://doi.org/10.1093/nar/16.16.8077.

60. Samadashwily, G.M., Dayn, A., and Mirkin, S.M. (1993). Suicidal nucleotide sequences for DNA polymerization. EMBO J. *12*, 4975–4983. https://doi.org/10.1002/j.1460-2075.1993.tb06191.x.

61. Strand, M., Prolla, T.A., Liskay, R.M., and Petes, T.D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature *365*, 274–276. https://doi.org/10.1038/365274a0.

62. Kunkel, T.A. (1993). Nucleotide repeats. Slippery DNA and diseases. Nature *365*, 207–208. https://doi.org/10.1038/365207a0.

63. Sinden, R.R., Pytlos-Sinden, M.J., and Potaman, V.N. (2007). Slipped strand DNA structures. Front. Biosci. *12*, 4788–4799.

64. Ando, T. (1966). A nuclease specific for heat-denatured DNA isolated from a product of Aspergillus oryzae. Biochim. Biophys. Acta *114*, 158–168. https://doi.org/10.1016/0005-2787(66)90263-2.

65. Mirkin, S.M., and Frank-Kamenetskii, M.D. (1994). H-DNA and related structures. Annu. Rev. Biophys. Biomol. Struct. *23*, 541–576. https://doi.org/10.1146/annurev.bb.23.060194.002545.

66. Wang, G., Zhao, J., and Vasquez, K.M. (2009). Methods to determine DNA structural alterations and genetic instability. Methods *48*, 54–62. https://doi.org/10.1016/j.ymeth.2009.02.012.

67. Tullius, T.D. (1991). The use of chemical probes to analyse DNA and RNA structures. Curr. Opin. Struct. Biol. *1*, 428–434. https://doi.org/10.1016/0959-440X(91)90043-S.

68. Neidle, S., and Stuart, D.I. (1976). The crystal and molecular structure of an osmium bispyridine adduct of thymine. Biochim. Biophys. Acta *418*, 226–231. https://doi.org/10.1016/0005-2787(76)90072-1.

69. Kohwi-Shigematsu, T., and Kohwi, Y. (1992). Detection of non-B-DNA structures at specific sites in supercoiled plasmid DNA and chromatin with haloacetaldehyde and diethyl pyrocarbonate. Methods Enzymol. *212*, 155–180. https://doi.org/10.1016/0076-6879(92)12011-e.

70. Frenkel, K., Goldstein, M.S., Duker, N.J., and Teebor, G.W. (1981). Identification of the cis-thymine glycol moiety in oxidized deoxyribonucleic acid. Biochemistry *20*, 750–754. https://doi.org/10.1021/bi00507a014.

71. Herr, W., Corbin, V., and Gilbert, W. (1982). Nucleotide sequence of the 3′ half of AKV. Nucleic Acids Res. *10*, 6931–6944. https://doi.org/10.1093/nar/10.21.6931.

72. Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA *74*, 560–564. https://doi.org/10.1073/pnas.74.2.560.

73. Voloshin, O.N., Mirkin, S.M., Lyamichev, V.I., Belotserkovskii, B.P., and Frank-Kamenetskii, M.D. (1988). Chemical probing of homopurine-homopyrimidine mirror repeats in supercoiled DNA. Nature *333*, 475–476.

74. Hanvey, J.C., Klysik, J., and Wells, R.D. (1988). Influence of DNA sequence on the formation of non-B right-handed helices in oligopurine.oligopyrimidine inserts in plasmids. J. Biol. Chem. *263*, 7386–7396.

75. Htun, H., and Dahlberg, J.E. (1988). Single strands, triple strands, and kinks in H-DNA. Science *241*, 1791–1796. https://doi.org/10.1126/science.3175620.

76. Johnston, B.H. (1988). The S1-sensitive form of d(C-T)n.d(A-G)n: chemical evidence for a three-stranded structure in plasmids. Science *241*, 1800–1804. https://doi.org/10.1126/science.2845572.

77. Lilley, D.M. (1983). Structural perturbation in supercoiled DNA: hypersensitivity to modification by a single-strand-selective chemical reagent conferred by inverted repeat sequences. Nucleic Acids Res. *11*, 3097–3112. https://doi.org/10.1093/nar/11.10.3097.

78. Lilley, D.M., and Hallam, L.R. (1983). The interactions of enzyme and chemical probes with inverted repeats in supercoiled DNA. J. Biomol. Struct. Dyn. *1*, 169–182. https://doi.org/10.1080/07391102.1983.10507433.

79. Lilley, D.M., and Palecek, E. (1984). The supercoil-stabilised cruciform of ColE1 is hyper-reactive to osmium tetroxide. EMBO J. *3*, 1187–1192. https://doi.org/10.1002/j.1460-2075.1984.tb01949.x.

80. Greaves, D.R., Patient, R.K., and Lilley, D.M. (1985). Facile cruciform formation by an (A-T)34 sequence from a Xenopus globin gene. J. Mol. Biol. *185*, 461–478. https://doi.org/10.1016/0022-2836(85)90064-6.

81. McLean, M.J., Larson, J.E., Wohlrab, F., and Wells, R.D. (1987). Reaction conditions affect the specificity of bromoacetaldehyde as a probe for DNA cruciforms and B-Z junctions. Nucleic Acids Res. *15*, 6917–6935. https://doi.org/10.1093/nar/15.17.6917.

82. Panyutin, I., Lyamichev, V., and Mirkin, S. (1985). A structural transition in d(AT)n.d(AT)n inserts within superhelical DNA. J. Biomol. Struct. Dyn. *2*, 1221–1234. https://doi.org/10.1080/07391102.1985.10507634.

83. Johnston, B.H., and Rich, A. (1985). Chemical probes of DNA conformation: detection of Z-DNA at nucleotide resolution. Cell *42*, 713–724. https://doi.org/10.1016/0092-8674(85)90268-5.

84. Kang, D.S., and Wells, R.D. (1985). B-Z DNA junctions contain few, if any, nonpaired bases at physiological superhelical densities. J. Biol. Chem. *260*, 7783–7790.

85. Herr, W. (1985). Diethyl pyrocarbonate: a chemical probe for secondary structure in negatively supercoiled DNA. Proc. Natl. Acad. Sci. USA *82*, 8009–8013. https://doi.org/10.1073/pnas.82.23.8009.

86. Vojtísková, M., Mirkin, S., Lyamichev, V., Voloshin, O., Frank-Kamenetskii, M., and Palecek, E. (1988). Chemical probing of the homopurine.-homopyrimidine tract in supercoiled DNA at single-nucleotide resolution. FEBS Lett. *234*, 295–299. https://doi.org/10.1016/0014-5793(88)80102-9.

87. Hanvey, J.C., Shimizu, M., and Wells, R.D. (1988). Intramolecular DNA triplexes in supercoiled plasmids. Proc. Natl. Acad. Sci. USA *85*, 6292–6296. https://doi.org/10.1073/pnas.85.17.6292.

88. Kohwi, Y., and Kohwi-Shigematsu, T. (1988). Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. Proc. Natl. Acad. Sci. USA *85*, 3781–3785. https://doi.org/10.1073/pnas.85.11.3781.

89. Kohwi, Y., and Kohwi-Shigematsu, T. (1993). Structural polymorphism of homopurine-homopyrimidine sequences at neutral pH. J. Mol. Biol. *231*, 1090–1101. https://doi.org/10.1006/jmbi.1993.1354.

90. Dayn, A., Samadashwily, G.M., and Mirkin, S.M. (1992). Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. Proc. Natl. Acad. Sci. USA *89*, 11406–11410. https://doi.org/10.1073/pnas.89.23.11406.

91. Peck, L.J., and Wang, J.C. (1983). Energetics of B-to-Z transition in DNA. Proc. Natl. Acad. Sci. USA *80*, 6206–6210. https://doi.org/10.1073/pnas.80.20.6206.

92. Haniford, D.B., and Pulleyblank, D.E. (1983). Facile transition of poly [d(TG) x d(CA)] into a left-handed helix in physiological conditions. Nature *302*, 632–634. https://doi.org/10.1038/302632a0.

93. Haniford, D.B., and Pulleyblank, D.E. (1983). The in-vivo occurrence of Z DNA. J. Biomol. Struct. Dyn. *1*, 593–609. https://doi.org/10.1080/07391102.1983.10507467.

94. Courey, A.J., and Wang, J.C. (1983). Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. Cell *33*, 817–829. https://doi.org/10.1016/0092-8674(83)90024-7.

95. Lyamichev, V.I., Panyutin, I.G., and Frank-Kamenetskii, M.D. (1983). Evidence of cruciform structures in superhelical DNA provided by two-dimensional gel electrophoresis. FEBS Lett. *153*, 298–302. https://doi.org/10.1016/0014-5793(83)80628-0.

96. Lam, E.Y., Beraldi, D., Tannahill, D., and Balasubramanian, S. (2013). G-quadruplex structures are stable and detectable in human genomic DNA. Nat. Commun. *4*, 1796. https://doi.org/10.1038/ncomms2792.

97. Zheng, K.W., Zhang, J.Y., He, Y.D., Gong, J.Y., Wen, C.J., Chen, J.N., Hao, Y.H., Zhao, Y., and Tan, Z. (2020). Detection of genomic G-quadruplexes in living cells using a small artificial protein. Nucleic Acids Res. *48*, 11706–11720. https://doi.org/10.1093/nar/gkaa841.

98. Müller, S., Kumari, S., Rodriguez, R., and Balasubramanian, S. (2010). Small-molecule-mediated G-quadruplex isolation from human cells. Nat. Chem. *2*, 1095–1098. https://doi.org/10.1038/nchem.842.

99. Mizusawa, S., Nishimura, S., and Seela, F. (1986). Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res. *14*, 1319–1324. https://doi.org/10.1093/nar/14.3.1319.

100. Weitzmann, M.N., Woodford, K.J., and Usdin, K. (1996). The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. J. Biol. Chem. *271*, 20958–20964. https://doi.org/10.1074/jbc.271.34.20958.

101. Murat, P., Guilbaud, G., and Sale, J.E. (2020). DNA polymerase stalling at structured DNA constrains the expansion of short tandem repeats. Genome Biol. *21*, 209. https://doi.org/10.1186/s13059-020-02124-x.

102. Marsico, G., Chambers, V.S., Sahakyan, A.B., McCauley, P., Boutell, J.M., Antonio, M.D., and Balasubramanian, S. (2019). Whole genome experimental maps of DNA G-quadruplexes in multiple species. Nucleic Acids Res. *47*, 3862–3874. https://doi.org/10.1093/nar/gkz179.

103. Mueller, P.R., and Wold, B. (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science *246*, 780–786. https://doi.org/10.1126/science.2814500.

104. Wang, G., and Vasquez, K.M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proc. Natl. Acad. Sci. USA *101*, 13448–13453. https://doi.org/10.1073/pnas.0405116101.

105. Zhao, J., Wang, G., Del Mundo, I.M., McKinney, J.A., Lu, X., Bacolla, A., Boulware, S.B., Zhang, C., Zhang, H., Ren, P., et al. (2018). Distinct mechanisms of nuclease-directed DNA-structure-induced genetic instability in cancer genomes. Cell Rep. *22*, 1200–1210. https://doi.org/10.1016/j.celrep.2018.01.014.

106. Lu, S., Wang, G., Bacolla, A., Zhao, J., Spitser, S., and Vasquez, K.M. (2015). Short inverted repeats are hotspots for genetic instability: relevance to cancer genomes. Cell Rep. *10*, 1674–1680. https://doi.org/10.1016/j.celrep.2015.02.039.

107. Liu, G., Chen, X., Bissler, J.J., Sinden, R.R., and Leffak, M. (2010). Replication-dependent instability at (CTG) x (CAG) repeat hairpins in human cells. Nat. Chem. Biol. *6*, 652–659. https://doi.org/10.1038/nchembio.416.

108. Ishizuka, J.J., Manguso, R.T., Cheruiyot, C.K., Bi, K., Panda, A., Iracheta-Vellve, A., Miller, B.C., Du, P.P., Yates, K.B., Dubrot, J., et al. (2019). Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. Nature *565*, 43–48. https://doi.org/10.1038/s41586-018-0768-9.

109. Zhang, T., Yin, C., Fedorov, A., Qiao, L., Bao, H., Beknazarov, N., Wang, S., Gautam, A., Williams, R.M., Crawford, J.C., et al. (2022). ADAR1 masks the cancer immunotherapeutic promise of ZBP1-driven necroptosis. Nature *606*, 594–602. https://doi.org/10.1038/s41586-022-04753-7.

110. Herbert, A. (2020). Mendelian disease caused by variants affecting recognition of Z-DNA and Z-RNA by the Zalpha domain of the double-stranded RNA editing enzyme ADAR. Eur. J. Hum. Genet. *28*, 114–117. https://doi.org/10.1038/s41431-019-0458-6.

111. de Reuver, R., Dierick, E., Wiernicki, B., Staes, K., Seys, L., De Meester, E., Muyldermans, T., Botzki, A., Lambrecht, B.N., Van Nieuwerburgh, F., et al. (2021). ADAR1 interaction with Z-RNA promotes editing of endogenous double-stranded RNA and prevents MDA5-dependent immune activation. Cell Rep. *36*, 109500. https://doi.org/10.1016/j.celrep.2021.109500.

112. Dézé, O., Laffleur, B., and Cogné, M. (2023). Roles of G4-DNA and G4-RNA in class switch recombination and additional regulations in B-lymphocytes. Molecules *28*, 1159. https://doi.org/10.3390/molecules28031159.

113. Qiao, Q., Wang, L., Meng, F.L., Hwang, J.K., Alt, F.W., and Wu, H. (2017). AID recognizes structured DNA for class switch recombination. Mol. Cell *67*, 361–373.e4. https://doi.org/10.1016/j.molcel.2017.06.034.

114. Yewdell, W.T., Kim, Y., Chowdhury, P., Lau, C.M., Smolkin, R.M., Belcheva, K.T., Fernandez, K.C., Cols, M., Yen, W.F., Vaidyanathan, B., et al. (2020). A hyper-IgM syndrome mutation in activation-induced cytidine deaminase disrupts G-quadruplex binding and genome-wide chromatin localization. Immunity *53*, 952–970.e11. https://doi.org/10.1016/j.immuni.2020.10.003.

115. Dalloul, Z., Chenuet, P., Dalloul, I., Boyer, F., Aldigier, J.C., Laffleur, B., El Makhour, Y., Ryffel, B., Quesniaux, V.F.J., Togbé, D., et al. (2018). G-quadruplex DNA targeting alters class-switch recombination in B cells and attenuates allergic inflammation. J. Allergy Clin. Immunol. *142*, 1352–1355. https://doi.org/10.1016/j.jaci.2018.06.011.

116. Lee, J.S., Burkholder, G.D., Latimer, L.J., Haug, B.L., and Braun, R.P. (1987). A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. Nucleic Acids Res. *15*, 1047–1061. https://doi.org/10.1093/nar/15.3.1047.

117. Agazie, Y.M., Lee, J.S., and Burkholder, G.D. (1994). Characterization of a new monoclonal antibody to triplex DNA and immunofluorescent staining of mammalian chromosomes. J. Biol. Chem. *269*, 7019–7023.

118. Burkholder, G.D., Latimer, L.J., and Lee, J.S. (1988). Immunofluorescent staining of mammalian nuclei and chromosomes with a monoclonal antibody to triplex DNA. Chromosoma *97*, 185–192. https://doi.org/10.1007/BF00292959.

119. Brown, B.A., 2nd, Li, Y., Brown, J.C., Hardin, C.C., Roberts, J.F., Pelsue, S.C., and Shultz, L.D. (1998). Isolation and characterization of a monoclonal anti-quadruplex DNA antibody from autoimmune "viable motheaten" mice. Biochemistry *37*, 16325–16337. https://doi.org/10.1021/bi981354u.

120. Brown, J.C., Brown, B.A., 2nd, Li, Y., and Hardin, C.C. (1998). Construction and characterization of a quadruplex DNA selective single-chain autoantibody from a viable motheaten mouse hybridoma with homology to telomeric DNA binding proteins. Biochemistry *37*, 16338–16348. https://doi.org/10.1021/bi981434y.

121. Biffi, G., Tannahill, D., McCafferty, J., and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nat. Chem. *5*, 182–186. https://doi.org/10.1038/nchem.1548.

122. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Jr., Sen, D., and Lansdorp, P.M. (2014). Detection of G-quadruplex DNA in mammalian cells. Nucleic Acids Res. *42*, 860–869. https://doi.org/10.1093/nar/gkt957.

123. Arndt-Jovin, D.J., Robert-Nicoud, M., Baurschmidt, P., and Jovin, T.M. (1985). Immunofluorescence localization of Z-DNA in chromosomes: quantitation by scanning microphotometry and computer-assisted image analysis. J. Cell Biol. *101*, 1422–1433. https://doi.org/10.1083/jcb.101.4.1422.

124. Staiano-Coico, L., Stollar, B.D., Darzynkiewicz, Z., Dutkowski, R., and Weksler, M.E. (1985). Binding of anti-Z-DNA antibodies in quiescent and activated lymphocytes: relationship to cell cycle progression and chromatin changes. Mol. Cell. Biol. *5*, 3270–3273. https://doi.org/10.1128/mcb.5.11.3270-3273.1985.

125. Lipps, H.J., Nordheim, A., Lafer, E.M., Ammermann, D., Stollar, B.D., and Rich, A. (1983). Antibodies against Z DNA react with the macronucleus but not the micronucleus of the hypotrichous ciliate stylonychia Mytilus. Cell *32*, 435–441. https://doi.org/10.1016/0092-8674(83)90463-4.

126. Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J. Immunol. Methods *89*, 123–130. https://doi.org/10.1016/0022-1759(86)90040-2.

127. Skourti-Stathaki, K. (2022). Detection of R-loop structures by immunofluorescence using the S9.6 monoclonal antibody. Methods Mol. Biol. *2528*, 21–29. https://doi.org/10.1007/978-1-0716-2477-7_2.

128. Zeraati, M., Langley, D.B., Schofield, P., Moye, A.L., Rouet, R., Hughes, W.E., Bryan, T.M., Dinger, M.E., and Christ, D. (2018). I-motif DNA structures are formed in the nuclei of human cells. Nat. Chem. *10*, 631–637. https://doi.org/10.1038/s41557-018-0046-3.

129. Galli, S., Melidis, L., Flynn, S.M., Varshney, D., Simeone, A., Spiegel, J., Madden, S.K., Tannahill, D., and Balasubramanian, S. (2022). DNA G-quadruplex recognition in vitro and in live cells by a structure-specific nanobody. J. Am. Chem. Soc. *144*, 23096–23103. https://doi.org/10.1021/jacs.2c10656.

130. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D., and Balasubramanian, S. (2018). Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. Nat. Protoc. *13*, 551–564. https://doi.org/10.1038/nprot.2017.150.

131. Shin, S.I., Ham, S., Park, J., Seo, S.H., Lim, C.H., Jeon, H., Huh, J., and Roh, T.Y. (2016). Z-DNA-forming sites identified by ChIP-seq are associated with actively transcribed regions in the human genome. DNA Res. *23*, 477–486. https://doi.org/10.1093/dnares/dsw031.

132. Martinez, C.D.P., Zeraati, M., Rouet, R., Mazigi, O., Gloss, B., Chan, C.-L., Bryan, T.M., Smith, N.M., Dinger, M.E., Kummerfeld, S., and Christ, D. (2022). Human genomic DNA is widely interspersed with i-motif structures. Preprint at bioRxiv. https://doi.org/10.1101/2022.04.14.488274.

133. Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chédin, F. (2016). Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Mol. Cell *63*, 167–178. https://doi.org/10.1016/j.molcel.2016.05.032.

134. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell *45*, 814–825. https://doi.org/10.1016/j.molcel.2012.01.017.

135. Lyu, J., Shao, R., Kwong Yung, P.Y., and Elsässer, S.J. (2022). Genome-wide mapping of G-quadruplex structures with CUT&Tag. Nucleic Acids Res. *50*, e13. https://doi.org/10.1093/nar/gkab1073.

136. Hui, W.W.I., Simeone, A., Zyner, K.G., Tannahill, D., and Balasubramanian, S. (2021). Single-cell mapping of DNA G-quadruplex structures in human cancer cells. Sci. Rep. *11*, 23641. https://doi.org/10.1038/s41598-021-02943-3.

137. Zanin, I., Ruggiero, E., Nicoletto, G., Lago, S., Maurizio, I., Gallina, I., and Richter, S.N. (2023). Genome-wide mapping of i-motifs reveals their association with transcription regulation in live human cells. Nucleic Acids Res.gkad626. https://doi.org/10.1093/nar/gkad626.

138. Wittig, B., Wölfl, S., Dorbic, T., Vahrson, W., and Rich, A. (1992). Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. EMBO J. *11*, 4653–4663. https://doi.org/10.1002/j.1460-2075.1992.tb05567.x.

139. Wittig, B., Dorbic, T., and Rich, A. (1991). Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei. Proc. Natl. Acad. Sci. USA *88*, 2259–2263. https://doi.org/10.1073/pnas.88.6.2259.

140. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., and Balasubramanian, S. (2020). The regulation and functions of DNA and RNA G-quadruplexes. Nat. Rev. Mol. Cell Biol. *21*, 459–474. https://doi.org/10.1038/s41580-020-0236-x.

141. Yang, S.Y., Chang, E.Y.C., Lim, J., Kwan, H.H., Monchaud, D., Yip, S., Stirling, P.C., and Wong, J.M.Y. (2021). G-quadruplexes mark alternative lengthening of telomeres. NAR Cancer *3*, zcab031. https://doi.org/10.1093/narcan/zcab031.

142. Di Antonio, M., Ponjavic, A., Radzevičius, A., Ranasinghe, R.T., Catalano, M., Zhang, X., Shen, J., Needham, L.M., Lee, S.F., Klenerman, D., and Balasubramanian, S. (2020). Single-molecule visualization of DNA G-quadruplex formation in live cells. Nat. Chem. *12*, 832–837. https://doi.org/10.1038/s41557-020-0506-4.

143. Lee, W.T.C., Yin, Y., Morten, M.J., Tonzi, P., Gwo, P.P., Odermatt, D.C., Modesti, M., Cantor, S.B., Gari, K., Huang, T.T., and Rothenberg, E. (2021). Single-molecule imaging reveals replication fork coupled formation of G-quadruplex structures hinders local replication stress signaling. Nat. Commun. *12*, 2525. https://doi.org/10.1038/s41467-021-22830-9.

144. Mao, S.Q., Ghanbarian, A.T., Spiegel, J., Martínez Cuesta, S., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D., and Balasubramanian, S. (2018). DNA G-quadruplex structures mold the DNA methylome. Nat. Struct. Mol. Biol. *25*, 951–957. https://doi.org/10.1038/s41594-018-0131-8.

145. Halder, R., Halder, K., Sharma, P., Garg, G., SenGupta, S., and Chowdhury, S. (2010). Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. Mol. Biosyst. *6*, 2439–2447. https://doi.org/10.1039/c0mb00009d.

146. Cree, S.L., Fredericks, R., Miller, A., Pearce, F.G., Filichev, V., Fee, C., and Kennedy, M.A. (2016). DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. FEBS Lett. *590*, 2870–2883. https://doi.org/10.1002/1873-3468.12331.

147. Bhowmick, R., Lerdrup, M., Gadi, S.A., Rossetti, G.G., Singh, M.I., Liu, Y., Halazonetis, T.D., and Hickson, I.D. (2022). RAD51 protects human cells from transcription-replication conflicts. Mol. Cell *82*, 3366–3381.e9. https://doi.org/10.1016/j.molcel.2022.07.010.

148. Groelly, F.J., Dagg, R.A., Petropoulos, M., Rossetti, G.G., Prasad, B., Panagopoulos, A., Paulsen, T., Karamichali, A., Jones, S.E., Ochs, F., et al. (2022). Mitotic DNA synthesis is caused by transcription-replication conflicts in BRCA2-deficient cells. Mol. Cell *82*, 3382–3397.e7. https://doi.org/10.1016/j.molcel.2022.07.011.

149. Kouzine, F., Wojtowicz, D., Baranello, L., Yamane, A., Nelson, S., Resch, W., Kieffer-Kwon, K.R., Benham, C.J., Casellas, R., Przytycka, T.M., and Levens, D. (2017). Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. Cell Syst. *4*, 344–356.e7. https://doi.org/10.1016/j.cels.2017.01.013.

150. Wu, T., Lyu, R., You, Q., and He, C. (2020). Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. Nat. Methods *17*, 515–523. https://doi.org/10.1038/s41592-020-0797-9.

151. Matos-Rodrigues, G., van Wietmarschen, N., Wu, W., Tripathi, V., Koussa, N.C., Pavani, R., Nathan, W.J., Callen, E., Belinky, F., Mohammed, A., et al. (2022). S1-END-seq reveals DNA secondary structures in human cells. Mol. Cell *82*, 3538–3552.e5. https://doi.org/10.1016/j.molcel.2022.08.007.

152. van Wietmarschen, N., Sridharan, S., Nathan, W.J., Tubbs, A., Chan, E.M., Callen, E., Wu, W., Belinky, F., Tripathi, V., Wong, N., et al. (2020). Repeat expansions confer WRN dependence in microsatellite-unstable cancers. Nature *586*, 292–298. https://doi.org/10.1038/s41586-020-2769-8.

153. Maekawa, K., Yamada, S., Sharma, R., Chaudhuri, J., and Keeney, S. (2022). Triple-helix potential of the mouse genome. Proc. Natl. Acad. Sci. USA *119*, e2203967119. https://doi.org/10.1073/pnas.2203967119.

154. Hänsel-Hertsch, R., Beraldi, D., Lensing, S.V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M., et al. (2016). G-quadruplex structures mark human regulatory chromatin. Nat. Genet. *48*, 1267–1272. https://doi.org/10.1038/ng.3662.

155. Liu, H.Y., Zhao, Q., Zhang, T.P., Wu, Y., Xiong, Y.X., Wang, S.K., Ge, Y.L., He, J.H., Lv, P., Ou, T.M., et al. (2016). Conformation selective antibody enables genome profiling and leads to discovery of parallel G-quadruplex in human telomeres. Cell Chem. Biol. *23*, 1261–1270. https://doi.org/10.1016/j.chembiol.2016.08.013.

156. Esnault, C., Magat, T., Zine El Aabidine, A., Garcia-Oliver, E., Cucchiarini, A., Bouchouika, S., Lleres, D., Goerke, L., Luo, Y., Verga, D., et al. (2023). G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. Nat. Genet. *55*, 1359–1369. https://doi.org/10.1038/s41588-023-01437-4.

157. Feng, Y., He, Z., Luo, Z., Sperti, F.R., Valverde, I.E., Zhang, W., and Monchaud, D. (2023). Side-by-side comparison of G-quadruplex (G4) capture efficiency of the antibody BG4 versus the small-molecule ligands TASQs. iScience *26*, 106846. https://doi.org/10.1016/j.isci.2023.106846.

158. Ma, X., Feng, Y., Yang, Y., Li, X., Shi, Y., Tao, S., Cheng, X., Huang, J., Wang, X.E., Chen, C., et al. (2022). Genome-wide characterization of i-motifs and their potential roles in the stability and evolution of transposable elements in rice. Nucleic Acids Res. *50*, 3226–3238. https://doi.org/10.1093/nar/gkac121.

159. Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. Mol. Cell *68*, 745–757.e5. https://doi.org/10.1016/j.molcel.2017.10.008.

160. Yan, Q., Shields, E.J., Bonasio, R., and Sarma, K. (2019). Mapping native R-loops genome-wide using a targeted nuclease approach. Cell Rep. *29*, 1369–1380.e5. https://doi.org/10.1016/j.celrep.2019.09.052.

161. Wang, K., Wang, H., Li, C., Yin, Z., Xiao, R., Li, Q., Xiang, Y., Wang, W., Huang, J., Chen, L., et al. (2021). Genomic profiling of native R loops with a DNA-RNA hybrid recognition sensor. Sci. Adv. *7*, eabe3516. https://doi.org/10.1126/sciadv.abe3516.

162. Jiang, Y., Huang, F., Chen, L., Gu, J.H., Wu, Y.W., Jia, M.Y., Lin, Z., Zhou, Y., Li, Y.C., Yu, C., et al. (2022). Genome-wide map of R-loops reveals its interplay with transcription and genome integrity during germ cell meiosis. J. Adv. Res. https://doi.org/10.1016/j.jare.2022.10.016.

163. Malig, M., Hartono, S.R., Giafaglione, J.M., Sanz, L.A., and Chedin, F. (2020). Ultra-deep coverage single-molecule R-loop footprinting reveals principles of R-loop formation. J. Mol. Biol. *432*, 2271–2288. https://doi.org/10.1016/j.jmb.2020.02.014.

164. Wu, T., Lyu, R., and He, C. (2022). spKAS-seq reveals R-loop dynamics using low-input materials by detecting single-stranded DNA with strand specificity. Sci. Adv. *8*, eabq2166. https://doi.org/10.1126/sciadv.abq2166.

165. García-Rubio, M., Barroso, S.I., and Aguilera, A. (2018). Detection of DNA-RNA hybrids in vivo. Methods Mol. Biol. *1672*, 347–361. https://doi.org/10.1007/978-1-4939-7306-4_24.

166. Chédin, F., Hartono, S.R., Sanz, L.A., and Vanoosthuyse, V. (2021). Best practices for the visualization, mapping, and manipulation of R-loops. EMBO J. *40*, e106394. https://doi.org/10.15252/embj.2020106394.

167. Crossley, M.P., Brickner, J.R., Song, C., Zar, S.M.T., Maw, S.S., Chédin, F., Tsai, M.S., and Cimprich, K.A. (2021). Catalytically inactive, purified RNase H1: a specific and sensitive probe for RNA-DNA hybrid imaging. J. Cell Biol. *220*, e202101092. https://doi.org/10.1083/jcb.202101092.

168. Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chédin, F., and Vanoosthuyse, V. (2018). The affinity of the S9.6 antibody for double-stranded RNAs impacts the accurate mapping of R-loops in fission yeast. J. Mol. Biol. *430*, 272–284. https://doi.org/10.1016/j.jmb.2017.12.016.

169. Cerritelli, S.M., Sakhuja, K., and Crouch, R.J. (2022). RNase H1, the gold standard for R-Loop detection. Methods Mol. Biol. *2528*, 91–114. https://doi.org/10.1007/978-1-0716-2477-7_7.

170. Silva, S., Guillén-Mendoza, C., and Aguilera, A. (2022). RNase H1 hybrid-binding domain-based tools for cellular biology studies of DNA-RNA hybrids in mammalian cells. Methods Mol. Biol. *2528*, 115–125. https://doi.org/10.1007/978-1-0716-2477-7_8.

171. Crossley, M.P., Song, C., Bocek, M.J., Choi, J.H., Kousorous, J., Sathirachinda, A., Lin, C., Brickner, J.R., Bai, G., Lans, H., et al. (2023). R-loop-derived cytoplasmic RNA-DNA hybrids activate an immune response. Nature *613*, 187–194. https://doi.org/10.1038/s41586-022-05545-9.

172. Yan, Q., and Sarma, K. (2020). MapR: a method for identifying native R-loops genome wide. Curr. Protoc. Mol. Biol. *130*, e113. https://doi.org/10.1002/cpmb.113.

173. Spiegel, J., Adhikari, S., and Balasubramanian, S. (2020). The structure and function of DNA G-quadruplexes. Trends Chem. *2*, 123–136. https://doi.org/10.1016/j.trechm.2019.07.002.

174. Javadekar, S.M., Nilavar, N.M., Paranjape, A., Das, K., and Raghavan, S.C. (2020). Characterization of G-quadruplex antibody reveals differential specificity for G4 DNA forms. DNA Res. *27*, dsaa024. https://doi.org/10.1093/dnares/dsaa024.

175. Biffi, G., Di Antonio, M., Tannahill, D., and Balasubramanian, S. (2014). Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat. Chem. *6*, 75–80. https://doi.org/10.1038/nchem.1805.

176. Varshney, D., Cuesta, S.M., Herdy, B., Abdullah, U.B., Tannahill, D., and Balasubramanian, S. (2021). RNA G-quadruplex structures control ribosomal protein production. Sci. Rep. *11*, 22735. https://doi.org/10.1038/s41598-021-01847-6.

177. Kazemier, H.G., Paeschke, K., and Lansdorp, P.M. (2017). Guanine quadruplex monoclonal antibody 1H6 cross-reacts with restrained thymidine-rich single stranded DNA. Nucleic Acids Res. *45*, 5913–5919. https://doi.org/10.1093/nar/gkx245.

178. Cañeque, T., Müller, S., and Rodriguez, R. (2018). Visualizing biologically active small molecules in cells using click chemistry. Nat. Rev. Chem. *2*, 202–215. https://doi.org/10.1038/s41570-018-0030-x.

179. Monchaud, D. (2020). Quadruplex detection in human cells. Annu. Rep. Med. Chem. *54*, 133–160. https://doi.org/10.1016/bs.armc.2020.04.007.

180. Olivieri, M., Cho, T., Álvarez-Quilón, A., Li, K., Schellenberg, M.J., Zimmermann, M., Hustedt, N., Rossi, S.E., Adam, S., Melo, H., et al. (2020). A genetic map of the response to DNA damage in human cells. Cell *182*, 481–496.e21. https://doi.org/10.1016/j.cell.2020.05.040.

181. Lyu, R., Wu, T., Zhu, A.C., West-Szymanski, D.C., Weng, X., Chen, M., and He, C. (2022). KAS-seq: genome-wide sequencing of single-stranded DNA by N_3-kethoxal-assisted labeling. Nat. Protoc. *17*, 402–420. https://doi.org/10.1038/s41596-021-00647-6.

182. Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). DNA breaks and end resection measured genome-wide by end sequencing. Mol. Cell *63*, 898–911. https://doi.org/10.1016/j.molcel.2016.06.034.

183. Wong, N., John, S., Nussenzweig, A., and Canela, A. (2021). END-seq: an unbiased, high-resolution, and genome-wide approach to map DNA double-strand breaks and resection in human cells. Methods Mol. Biol. *2153*, 9–31. https://doi.org/10.1007/978-1-0716-0644-5_2.

184. Shinoda, K., Maman, Y., Canela, A., Schatz, D.G., Livak, F., and Nussenzweig, A. (2019). Intra-Vkappa cluster recombination shapes the Ig Kappa Locus repertoire. Cell Rep. *29*, 4471–4481.e6. https://doi.org/10.1016/j.celrep.2019.11.088.

185. Paiano, J., Wu, W., Yamada, S., Sciascia, N., Callen, E., Paola Cotrim, A., Deshpande, R.A., Maman, Y., Day, A., Paull, T.T., and Nussenzweig, A. (2020). ATM and PRDM9 regulate SPO11-bound recombination intermediates during meiosis. Nat. Commun. *11*, 857. https://doi.org/10.1038/s41467-020-14654-w.

186. Canela, A., Maman, Y., Huang, S.N., Wutz, G., Tang, W., Zagnoli-Vieira, G., Callen, E., Wong, N., Day, A., Peters, J.M., et al. (2019). Topoisomerase II-induced chromosome breakage and translocation is determined by chromosome architecture and transcriptional activity. Mol. Cell *75*, 252–266.e8. https://doi.org/10.1016/j.molcel.2019.04.030.

187. Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., Wu, W., Wu, X., Day, A., Wong, N., et al. (2018). Dual roles of poly(dA:dT) tracts in replication initiation and fork collapse. Cell *174*, 1127–1142.e19. https://doi.org/10.1016/j.cell.2018.07.011.

188. Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K.R., Pekowska, A., Zhang, H., Rao, S.S.P., et al. (2017). Genome organization drives chromosome fragility. Cell *170*, 507–521.e18. https://doi.org/10.1016/j.cell.2017.06.034.

189. Wu, W., Hill, S.E., Nathan, W.J., Paiano, J., Callen, E., Wang, D., Shinoda, K., van Wietmarschen, N., Colón-Mercado, J.M., Zong, D., et al. (2021). Neuronal enhancers are hotspots for DNA single-strand break repair. Nature *593*, 440–444. https://doi.org/10.1038/s41586-021-03468-5.

190. Wang, D., Wu, W., Callen, E., Pavani, R., Zolnerowich, N., Kodali, S., Zong, D., Wong, N., Noriega, S., Nathan, W.J., et al. (2022). Active DNA demethylation promotes cell fate specification and the DNA damage response. Science *378*, 983–989. https://doi.org/10.1126/science.add9838.

191. Behan, F.M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C.M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., et al. (2019). Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature *568*, 511–516. https://doi.org/10.1038/s41586-019-1103-9.

192. Chan, E.M., Shibue, T., McFarland, J.M., Gaeta, B., Ghandi, M., Dumont, N., Gonzalez, A., McPartlan, J.S., Li, T., Zhang, Y., et al. (2019). WRN helicase is a synthetic lethal target in microsatellite unstable cancers. Nature *568*, 551–556. https://doi.org/10.1038/s41586-019-1102-x.

193. Kategaya, L., Perumal, S.K., Hager, J.H., and Belmont, L.D. (2019). Werner syndrome helicase is required for the survival of cancer cells with microsatellite instability. iScience *13*, 488–497. https://doi.org/10.1016/j.isci.2019.02.006.

194. Lieb, S., Blaha-Ostermann, S., Kamper, E., Rippka, J., Schwarz, C., Ehrenhöfer-Wölfer, K., Schlattl, A., Wernitznig, A., Lipp, J.J., Nagasaka, K., et al. (2019). Werner syndrome helicase is a selective vulnerability of microsatellite instability-high tumor cells. Elife *8*, e43333. https://doi.org/10.7554/eLife.43333.

195. Zong, D., Koussa, N.C., Cornwell, J.A., Pankajam, A.V., Kruhlak, M., Wong, N., Chari, R., Cappell, S.D., and Nussenzweig, A. (2023). Comprehensive mapping of cell fates in microsatellite unstable cancer cells support dual targeting of WRN and ATR. Preprint at bioRxiv. https://doi.org/10.1101/2023.07.28.550976.

196. Kaushal, S., Wollmuth, C.E., Das, K., Hile, S.E., Regan, S.B., Barnes, R.P., Haouzi, A., Lee, S.M., House, N.C.M., Guyumdzhyan, M., et al. (2019). Sequence and nuclease requirements for breakage and healing of a structure-forming (AT)n sequence within fragile site FRA16D. Cell Rep. *27*, 1151–1164.e5. https://doi.org/10.1016/j.celrep.2019.03.103.

197. Zhang, H., and Freudenreich, C.H. (2007). An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in S. cerevisiae. Mol. Cell *27*, 367–379. https://doi.org/10.1016/j.molcel.2007.06.012.

198. Mengoli, V., Ceppi, I., Sanchez, A., Cannavo, E., Halder, S., Scaglione, S., Gaillard, P.H., McHugh, P.J., Riesen, N., Pettazzoni, P., and Cejka, P. (2023). WRN helicase and mismatch repair complexes independently and synergistically disrupt cruciform DNA structures. EMBO J. *42*, e111998. https://doi.org/10.15252/embj.2022111998.

199. Boddy, M.N., Gaillard, P.H.L., McDonald, W.H., Shanahan, P., Yates, J.R., 3rd, and Russell, P. (2001). Mus81-Eme1 are essential components of a Holliday junction resolvase. Cell *107*, 537–548. https://doi.org/10.1016/s0092-8674(01)00536-0.

200. Zhu, Y., Biernacka, A., Pardo, B., Dojer, N., Forey, R., Skrzypczak, M., Fongang, B., Nde, J., Yousefi, R., Pasero, P., et al. (2019). qDSB-Seq is a general method for genome-wide quantification of DNA double-strand breaks using sequencing. Nat. Commun. *10*, 2313. https://doi.org/10.1038/s41467-019-10332-8.

201. Bizuayehu, T.T., Labun, K., Jakubec, M., Jefimov, K., Niazi, A.M., and Valen, E. (2022). Long-read single-molecule RNA structure sequencing using nanopore. Nucleic Acids Res. *50*, e120. https://doi.org/10.1093/nar/gkac775.

202. Aw, J.G.A., Lim, S.W., Wang, J.X., Lambert, F.R.P., Tan, W.T., Shen, Y., Zhang, Y., Kaewsapsak, P., Li, C., Ng, S.B., et al. (2021). Determination of isoform-specific RNA structure with nanopore long reads. Nat. Biotechnol. *39*, 336–346. https://doi.org/10.1038/s41587-020-0712-z.

203. Barsumian, E.L., Cunningham, C.M., Schlievert, P.M., and Watson, D.W. (1978). Heterogeneity of group A streptococcal pyrogenic exotoxin type B. Infect. Immun. *20*, 512–518. https://doi.org/10.1128/iai.20.2.512-518.1978.

204. Cer, R.Z., Donohue, D.E., Mudunuri, U.S., Temiz, N.A., Loss, M.A., Starner, N.J., Halusa, G.N., Volfovsky, N., Yi, M., Luke, B.T., et al. (2013). Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. Nucleic Acids Res. *41*, D94–D100. https://doi.org/10.1093/nar/gks955.

205. Huppert, J.L., and Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. Nucleic Acids Res. *33*, 2908–2916. https://doi.org/10.1093/nar/gki609.

206. Ye, C., Ji, G., Li, L., and Liang, C. (2014). detectIR: a novel program for detecting perfect and imperfect inverted repeats using complex numbers and vector calculation. PLoS One *9*, e113349. https://doi.org/10.1371/journal.pone.0113349.

207. Puig Lombardi, E., and Londoño-Vallejo, A. (2020). A guide to computational methods for G-quadruplex prediction. Nucleic Acids Res. *48*, 1–15. https://doi.org/10.1093/nar/gkz1097.

208. Rocher, V., Genais, M., Nassereddine, E., and Mourad, R. (2021). DeepG4: a deep learning approach to predict cell-type specific active G-quadruplex regions. PLoS Comput. Biol. *17*, e1009308. https://doi.org/10.1371/journal.pcbi.1009308.

209. Beknazarov, N., Jin, S., and Poptsova, M. (2020). Deep learning approach for predicting functional Z-DNA regions using omics data. Sci. Rep. *10*, 19134. https://doi.org/10.1038/s41598-020-76203-1.

210. Hosseini, M., Palmer, A., Manka, W., Grady, P.G.S., Patchigolla, V., Bi, J., O'Neill, R.J., Chi, Z., and Aguiar, D. (2023). Deep statistical modelling of nanopore sequencing translocation times reveals latent non-B DNA structures. Bioinformatics *39*, i242–i251. https://doi.org/10.1093/bioinformatics/btad220.

211. Shah, K.A., and Mirkin, S.M. (2015). The hidden side of unstable DNA repeats: mutagenesis at a distance. DNA Repair (Amst) *32*, 106–112. https://doi.org/10.1016/j.dnarep.2015.04.020.

212. Bacolla, A., Jaworski, A., Larson, J.E., Jakupciak, J.P., Chuzhanova, N., Abeysinghe, S.S., O'Connell, C.D., Cooper, D.N., and Wells, R.D. (2004). Breakpoints of gross deletions coincide with non-B DNA conformations. Proc. Natl. Acad. Sci. USA *101*, 14162–14167. https://doi.org/10.1073/pnas.0405974101.

213. Bacolla, A., Tainer, J.A., Vasquez, K.M., and Cooper, D.N. (2016). Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. Nucleic Acids Res. *44*, 5673–5688. https://doi.org/10.1093/nar/gkw261.

214. Du, X., Gertz, E.M., Wojtowicz, D., Zhabinskaya, D., Levens, D., Benham, C.J., Schäffer, A.A., and Przytycka, T.M. (2014). Potential non-B DNA regions in the human genome are associated with higher rates of nucleotide mutation and expression variation. Nucleic Acids Res. *42*, 12367–12379. https://doi.org/10.1093/nar/gku921.

215. Guiblet, W.M., Cremona, M.A., Harris, R.S., Chen, D., Eckert, K.A., Chiaromonte, F., Huang, Y.F., and Makova, K.D. (2021). Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. Nucleic Acids Res. *49*, 1497–1516. https://doi.org/10.1093/nar/gkaa1269.

216. McGinty, R.J., and Sunyaev, S.R. (2023). Revisiting mutagenesis at non-B DNA motifs in the human genome. Nat. Struct. Mol. Biol. *30*, 417–424. https://doi.org/10.1038/s41594-023-00936-6.

217. Weissensteiner, M.H., Cremona, M.A., Guiblet, W.M., Stoler, N., Harris, R.S., Cechova, M., Eckert, K.A., Chiaromonte, F., Huang, Y.F., and Makova, K.D. (2023). Accurate sequencing of DNA motifs able to form alternative (non-B) structures. Genome Res. *33*, 907–922. https://doi.org/10.1101/gr.277490.122.

218. Erwin, G.S., Gürsoy, G., Al-Abri, R., Suriyaprakash, A., Dolzhenko, E., Zhu, K., Hoerner, C.R., White, S.M., Ramirez, L., Vadlakonda, A., et al. (2023). Recurrent repeat expansions in human cancer genomes. Nature *613*, 96–102. https://doi.org/10.1038/s41586-022-05515-1.

219. Técher, H., Koundrioukoff, S., Nicolas, A., and Debatisse, M. (2017). The impact of replication stress on replication dynamics and DNA damage in vertebrate cells. Nat. Rev. Genet. *18*, 535–550. https://doi.org/10.1038/nrg.2017.46.

220. Zell, J., Rota Sperti, F., Britton, S., and Monchaud, D. (2021). DNA folds threaten genetic stability and can be leveraged for chemotherapy. RSC Chem. Biol. *2*, 47–76. https://doi.org/10.1039/d0cb00151a.

221. Krasilnikova, M.M., and Mirkin, S.M. (2004). Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Mol. Cell. Biol. *24*, 2286–2295. https://doi.org/10.1128/MCB.24.6.2286-2295.2004.

222. Sfeir, A., Kosiyatrakul, S.T., Hockemeyer, D., MacRae, S.L., Karlseder, J., Schildkraut, C.L., and de Lange, T. (2009). Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. Cell *138*, 90–103. https://doi.org/10.1016/j.cell.2009.06.021.

223. Dovrat, D., Dahan, D., Sherman, S., Tsirkas, I., Elia, N., and Aharoni, A. (2018). A live-cell imaging approach for measuring DNA replication rates. Cell Rep. *24*, 252–258. https://doi.org/10.1016/j.celrep.2018.06.018.

224. Mellor, C., Perez, C., and Sale, J.E. (2022). Creation and resolution of non-B-DNA structural impediments during replication. Crit. Rev. Biochem. Mol. Biol. *57*, 412–442. https://doi.org/10.1080/10409238.2022.2121803.

225. Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195–200. https://doi.org/10.1038/nature13124.

226. Viterbo, D., Michoud, G., Mosbach, V., Dujon, B., and Richard, G.F. (2016). Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. DNA Repair (Amst) *42*, 94–106. https://doi.org/10.1016/j.dnarep.2016.03.002.

227. Mirkin, S.M. (2007). Expandable DNA repeats and human disease. Nature *447*, 932–940. https://doi.org/10.1038/nature05977.

228. Owen, B.A., Yang, Z., Lai, M., Gajec, M., Badger, J.D., 2nd, Hayes, J.J., Edelmann, W., Kucherlapati, R., Wilson, T.M., and McMurray, C.T. (2005). (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. Nat. Struct. Mol. Biol. *12*, 663–670. https://doi.org/10.1038/nsmb965.

229. Kovtun, I.V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H., and McMurray, C.T. (2007). OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature *447*, 447–452. https://doi.org/10.1038/nature05778.

230. Bidichandani, S.I., Ashizawa, T., and Patel, P.I. (1998). The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am. J. Hum. Genet. *62*, 111–121. https://doi.org/10.1086/301680.

231. Potaman, V.N., Oussatcheva, E.A., Lyubchenko, Y.L., Shlyakhtenko, L.S., Bidichandani, S.I., Ashizawa, T., and Sinden, R.R. (2004). Length-dependent structure formation in Friedreich ataxia (GAA)n*(TTC)n repeats at neutral pH. Nucleic Acids Res. *32*, 1224–1231. https://doi.org/10.1093/nar/gkh274.

232. Gacy, A.M., Goellner, G.M., Spiro, C., Chen, X., Gupta, G., Bradbury, E.M., Dyer, R.B., Mikesell, M.J., Yao, J.Z., Johnson, A.J., et al. (1998). GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. Mol. Cell *1*, 583–593. https://doi.org/10.1016/s1097-2765(00)80058-1.

233. Brickner, J.R., Garzon, J.L., and Cimprich, K.A. (2022). Walking a tightrope: the complex balancing act of R-loops in genome stability. Mol. Cell *82*, 2267–2297. https://doi.org/10.1016/j.molcel.2022.04.014.

234. Groh, M., Lufino, M.M., Wade-Martins, R., and Gromak, N. (2014). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. *10*, e1004318. https://doi.org/10.1371/journal.pgen.1004318.

235. Loomis, E.W., Sanz, L.A., Chédin, F., and Hagerman, P.J. (2014). Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. PloS Genet. *10*, e1004294. https://doi.org/10.1371/journal.pgen.1004294.

236. Burnett, R., Melander, C., Puckett, J.W., Son, L.S., Wells, R.D., Dervan, P.B., and Gottesfeld, J.M. (2006). DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA.TTC repeats in Friedreich's ataxia. Proc. Natl. Acad. Sci. USA *103*, 11497–11502. https://doi.org/10.1073/pnas.0604939103.

237. Du, J., Campau, E., Soragni, E., Ku, S., Puckett, J.W., Dervan, P.B., and Gottesfeld, J.M. (2012). Role of mismatch repair enzymes in GAA·TTC triplet-repeat expansion in Friedreich ataxia induced pluripotent stem cells. J. Biol. Chem. *287*, 29861–29872. https://doi.org/10.1074/jbc.M112.391961.

238. Ohshima, K., Sakamoto, N., Labuda, M., Poirier, J., Moseley, M.L., Montermini, L., Ranum, L.P., Wells, R.D., and Pandolfo, M. (1999). A nonpathogenic GAAGGA repeat in the Friedreich gene: implications for pathogenesis. Neurology *53*, 1854–1857. https://doi.org/10.1212/wnl.53.8.1854.

239. Khristich, A.N., Armenia, J.F., Matera, R.M., Kolchinski, A.A., and Mirkin, S.M. (2020). Large-scale contractions of Friedreich's ataxia GAA repeats in yeast occur during DNA replication due to their triplex-forming ability. Proc. Natl. Acad. Sci. USA *117*, 1628–1637. https://doi.org/10.1073/pnas.1913416117.

240. Chintalaphani, S.R., Pineda, S.S., Deveson, I.W., and Kumar, K.R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. Acta Neuropathol. Commun. *9*, 98. https://doi.org/10.1186/s40478-021-01201-x.

241. Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W.Y., Humphrey, J., Jaunmuktane, Z., Sivakumar, P., Polke, J., et al. (2019). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. Nat. Genet. *51*, 649–658. https://doi.org/10.1038/s41588-019-0372-4.

242. Rafehi, H., Szmulewicz, D.J., Bennett, M.F., Sobreira, N.L.M., Pope, K., Smith, K.R., Gillies, G., Diakumis, P., Dolzhenko, E., Eberle, M.A., et al. (2019). Bioinformatics-based identification of expanded repeats: a non-reference intronic pentamer expansion in RFC1 causes CANVAS. Am. J. Hum. Genet. *105*, 151–165. https://doi.org/10.1016/j.ajhg.2019.05.016.

243. Cortese, A., Curro', R., Vegezzi, E., Yau, W.Y., Houlden, H., and Reilly, M.M. (2022). Cerebellar ataxia, neuropathy and vestibular areflexia syndrome (CANVAS): genetic and clinical aspects. Pract. Neurol. *22*, 14–18. https://doi.org/10.1136/practneurol-2020-002822.

244. Hisey, J.A., Radchenko, E.A., Ceschi, S., Rastokina, A., Mandel, N.H., McGinty, R.J., Matos-Rodrigues, G., Hernandez, A., Nussenzweig, A., and Mirkin, S.M. (2023). Preprint at bioRxiv. Pathogenic CANVAS (AAGGG)n repeats stall DNA replication due to the formation of alternative DNA structures. https://doi.org/10.1101/2023.07.25.550509.

245. Rafehi, H., Read, J., Szmulewicz, D.J., Davies, K.C., Snell, P., Fearnley, L.G., Scott, L., Thomsen, M., Gillies, G., Pope, K., et al. (2023). An intronic GAA repeat expansion in FGF14 causes the autosomal-dominant adult-onset ataxia SCA50/ATX-FGF14. Am. J. Hum. Genet. *110*, 105–119. https://doi.org/10.1016/j.ajhg.2022.11.015.

246. Pellerin, D., Danzi, M.C., Wilke, C., Renaud, M., Fazal, S., Dicaire, M.J., Scriba, C.K., Ashton, C., Yanick, C., Beijer, D., et al. (2023). Deep intronic FGF14 GAA repeat expansion in late-onset cerebellar ataxia. N. Engl. J. Med. *388*, 128–141. https://doi.org/10.1056/NEJMoa2207406.

247. Kosiol, N., Juranek, S., Brossart, P., Heine, A., and Paeschke, K. (2021). G-quadruplexes: a promising target for cancer therapy. Mol. Cancer *20*, 40. https://doi.org/10.1186/s12943-021-01328-4.

248. Del Mundo, I.M.A., Vasquez, K.M., and Wang, G. (2019). Modulation of DNA structure formation using small molecules. Biochim. Biophys. Acta Mol. Cell Res. *1866*, 118539. https://doi.org/10.1016/j.bbamcr.2019.118539.

249. Ivens, E., Cominetti, M.M.D., and Searcey, M. (2022). Junctions in DNA: underexplored targets for therapeutic intervention. Bioorg. Med. Chem. *69*, 116897. https://doi.org/10.1016/j.bmc.2022.116897.

250. McQuaid, K.T., Pipier, A., Cardin, C.J., and Monchaud, D. (2022). Interactions of small molecules with DNA junctions. Nucleic Acids Res. *50*, 12636–12656. https://doi.org/10.1093/nar/gkac1043.

251. Duskova, K., Lejault, P., Benchimol, É., Guillot, R., Britton, S., Granzhan, A., and Monchaud, D. (2020). DNA junction ligands trigger DNA damage and are synthetic lethal with DNA repair inhibitors in cancer cells. J. Am. Chem. Soc. *142*, 424–435. https://doi.org/10.1021/jacs.9b11150.

252. Zell, J., Duskova, K., Chouh, L., Bossaert, M., Chéron, N., Granzhan, A., Britton, S., and Monchaud, D. (2021). Dual targeting of higher-order DNA structures by azacryptands induces DNA junction-mediated DNA damage in cancer cells. Nucleic Acids Res. *49*, 10275–10288. https://doi.org/10.1093/nar/gkab796.

253. Rodriguez, R., Miller, K.M., Forment, J.V., Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., and Jackson, S.P. (2012). Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat. Chem. Biol. *8*, 301–310. https://doi.org/10.1038/nchembio.780.

254. Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N.J., Santos, N.D., Silvester, J., Wei, V., Garcia, J., et al. (2017). CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. Nat. Commun. *8*, 14432. https://doi.org/10.1038/ncomms14432.

255. McLuckie, K.I., Di Antonio, M., Zecchini, H., Xian, J., Caldas, C., Krippendorff, B.F., Tannahill, D., Lowe, C., and Balasubramanian, S. (2013). G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. J. Am. Chem. Soc. *135*, 9640–9643. https://doi.org/10.1021/ja404868t.

256. Zimmer, J., Tacconi, E.M.C., Folio, C., Badie, S., Porru, M., Klare, K., Tumiati, M., Markkanen, E., Halder, S., Ryan, A., et al. (2016). Targeting BRCA1 and BRCA2 deficiencies with G-quadruplex-interacting compounds. Mol. Cell *61*, 449–460. https://doi.org/10.1016/j.molcel.2015.12.004.

257. Hilton, J., Gelmon, K., Bedard, P.L., Tu, D., Xu, H., Tinker, A.V., Goodwin, R., Laurie, S.A., Jonker, D., Hansen, A.R., et al. (2022). Results of the phase I CCTG IND.231 trial of CX-5461 in patients with advanced solid tumors enriched for DNA-repair deficiencies. Nat. Commun. *13*, 3607. https://doi.org/10.1038/s41467-022-31199-2.

258. Xu, H., and Hurley, L.H. (2022). A first-in-class clinical G-quadruplex-targeting drug. The bench-to-bedside translation of the fluoroquinolone QQ58 to CX-5461 (Pidnarulex). Bioorg. Med. Chem. Lett. *77*, 129016. https://doi.org/10.1016/j.bmcl.2022.129016.

259. Nakamori, M., Panigrahi, G.B., Lanni, S., Gall-Duncan, T., Hayakawa, H., Tanaka, H., Luo, J., Otabe, T., Li, J., Sakata, A., et al. (2020). A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. Nat. Genet. *52*, 146–159. https://doi.org/10.1038/s41588-019-0575-8.

260. Hasuike, Y., Tanaka, H., Gall-Duncan, T., Mehkary, M., Nakatani, K., Pearson, C.E., Tsuji, S., Mochizuki, H., and Nakamori, M. (2022). CAG repeat-binding small molecule improves motor coordination impairment in a mouse model of dentatorubral-pallidoluysian atrophy. Neurobiol. Dis. *163*, 105604. https://doi.org/10.1016/j.nbd.2021.105604.

261. Lee, H.G., Imaichi, S., Kraeutler, E., Aguilar, R., Lee, Y.W., Sheridan, S.D., and Lee, J.T. (2023). Site-specific R-loops induce CGG repeat contraction and fragile X gene reactivation. Cell *186*, 2593–2609.e18. https://doi.org/10.1016/j.cell.2023.04.035.

262. Rastokina, A., Cebrián, J., Mozafari, N., Mandel, N.H., Smith, C.I.E., Lopes, M., Zain, R., and Mirkin, S.M. (2023). Large-scale expansions of Friedreich's ataxia GAA·TTC repeats in an experimental human system: role of DNA replication and prevention by LNA-DNA oligonucleotides and PNA oligomers. Nucleic Acids Res.gkad441. https://doi.org/10.1093/nar/gkad441.