

Review



**Cite this article:** Herbert A. 2020 ALU non-B-DNA conformations, flipons, binary codes and evolution. *R. Soc. Open Sci.* **7**: 200222. <http://dx.doi.org/10.1098/rsos.200222>

Received: 17 February 2020

Accepted: 18 May 2020

**Subject Category:**

Genetics and genomics

**Subject Areas:**

evolution/biophysics/genetics

**Keywords:**

evolution, Z-DNA, triplex, quadruplex, flipon, ALU

**Author for correspondence:**

Alan Herbert

e-mail: [alan.herbert@insideoutbio.com](mailto:alan.herbert@insideoutbio.com)

# ALU non-B-DNA conformations, flipons, binary codes and evolution

Alan Herbert

Discovery, InsideOutBio, Inc., 42 8th Street, Unit 3412, Charlestown, MA 02129, USA

 AH, 0000-0002-0093-1572

ALUs contribute to genetic diversity by altering DNA's linear sequence through retrotransposition, recombination and repair. ALUs also have the potential to form alternative non-B-DNA conformations such as Z-DNA, triplexes and quadruplexes that alter the read-out of information from the genome. I suggest here these structures enable the rapid reprogramming of cellular pathways to offset DNA damage and regulate inflammation. The experimental data supporting this form of genetic encoding is presented. ALU sequence motifs that form non-B-DNA conformations under physiological conditions are called flipons. Flipons are binary switches. They are dissipative structures that trade energy for information. By efficiently targeting cellular machines to active genes, flipons expand the repertoire of RNAs compiled from a gene. Their action greatly increases the informational capacity of linearly encoded genomes. Flipons are programmable by epigenetic modification, synchronizing cellular events by altering both chromatin state and nucleosome phasing. Different classes of flipon exist. Z-flipons are based on Z-DNA and modify the transcripts compiled from a gene. T-flipons are based on triplexes and localize non-coding RNAs that direct the assembly of cellular machines. G-flipons are based on G-quadruplexes and sense DNA damage, then trigger the appropriate protective responses. Flipon conformation is dynamic, changing with context. When frozen in one state, flipons often cause disease. The propagation of flipons throughout the genome by ALU elements represents a novel evolutionary innovation that allows for rapid change. Each ALU insertion creates variability by extracting a different set of information from the neighbourhood in which it lands. By elaborating on already successful adaptations, the newly compiled transcripts work with the old to enhance survival. Systems that optimize flipon settings through learning can adapt faster than with other forms of evolution. They avoid the risk of relying on random and irreversible codon rewrites.

# 1. Introduction

The small 140–300 base-pair (bp) ALU size belittles the extent of evolutionary improvisations enabled in Hominidae by this class of short interspersed nuclear element (SINE) repeat family. They are named for the restriction enzyme Alu I from *Arthrobacter luteus* that cuts within them, producing short fragments when two ALU elements are nearby. ALUs are derived from the non-coding 7SL RNA of the signal recognition protein. They spread invasively by retrotransposition, with each attack initiated by a different family, generating over one million copies to occupy almost 11% of the human genome [1,2]. The most efficient transposition was by ALUs that are dimeric, composed of a left and right arm derived from 7SLRNA joined by a deoxyadenosine-rich spacer and with a 3' polydeoxyadenosine tail of variable length (figure 1a). Like other DNA, ALUs are packaged into nucleosomes, wrapped around histone octamers. DNase hypersensitive sites are present every 10.2 bp, where the energetic cost of DNA bending is lowest. The flexing of d(CG) dinucleotides, which repeat every 32 nucleotides, contributes to this pattern [9]. The core nucleosome footprints from –80 to +69 of ALU with the spacer that separates the Alu monomers positioned in the linker region between each nucleosome pair [10,11]. The stiffness (i.e. resistance to bending) of the adenosine-rich ALU spacer sequence contributes to their exclusion from nucleosomes [12].

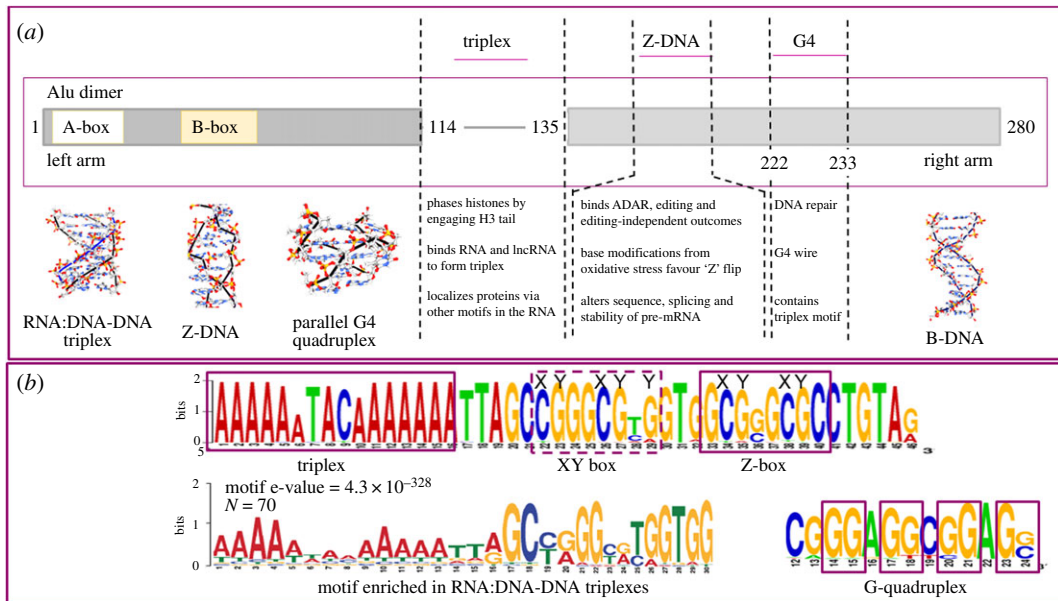
## 2. Why ALUs are interesting?

There are many reasons to consider ALUs of interest, reflecting the many different factors driving their evolution. First, ALUs have played a very large role in shaping the human genome [13]. They are widely dispersed and mostly found in RNA polymerase II (RP2) transcribed genes where they can act as signals for splicing and transcriptional termination [2,14]. Only a minority (0.2%) of ALUs undergo exonization, often with deleterious effects. Around 86% of ALU exons are inserted in the reverse orientation and contain an in-frame stop codon or a frameshift resulting in loss of protein function [15]. Second, the spread of ALUs through the genome during evolution offers a rapid way to alter the read-out of genetic information, independently of codon mutation. They are bound by a wide range of transcription factors [16] that enable the co-regulation of their host genes, even when those genes are present on different chromosomes [17]. Third, the extent of ALU transcription allows cells to monitor their health. High levels of ALU RNAs are sensed and initiate inflammatory responses leading to disease [18]. Mechanisms such as RNA editing [19–21], originally tasked with stopping the existential threat posed by ALU spread throughout the genome [4], now regulate innate immunity [22]. Many cancers exploit these pathways to remain immunologically silent [23–26]. Fourth, as will be reviewed here, the existing data are consistent with a model where the non-B structures (NoBs) ALUs form enable dynamic responses to environmental changes. In this synthesis, ALU sequences vary genomic read-out by changing conformation, acting as programmable switches to alter the transcripts produced. Once the invader, ALU elements now enhance the evolution of their host.

## 3. What does ALU RNA do?

ALU RNAs have roles that vary by how they are transcribed. A minority of ALUs (1.5%) undergo transcription by RNA polymerase III (RP3), using the same A and B boxes as 7SL RNA (figure 1) [14]. Many RP3 transcripts are expressed in a tissue-specific manner. An analysis of 27 cell lines, 16 primary cell types, and 45 tissues using a recently developed high-stringency method for identifying RP3 transcripts revealed that their cell-specific expression is predicted by chromatin modification, not by differences in ALU sequence [14]. Transcription is highest for ALU elements that lie within 10 000 bases of genes transcribed by RP2. The RP2 and RP3 transcription units cooperatively induce local chromosomal loop formation [27,28]. Some RP3 units produce an enhancer RNA (eRNA) that guides the assembly of transcription factories [29,30]. Other RP3 transcripts act in *trans* to inhibit RP2, with the ALU right arm binding to the RP2 cleft [31]. ALU self-cleavage relieves this inhibition during heat shock. The ALU-Y ribozyme cut site is at position 51 of the left arm. Cleavage is enhanced by the enhancer of *zeste 2* polycomb repressive complex 2 subunit (EZH2), even with catalytically inactive versions of the protein [32,33]. ALUs can inhibit translation by inhibiting ribosome assembly [34,35]. In addition, they induce phosphorylation of PKR (double-stranded RNA (dsRNA) activated protein kinase encoded by EIF2AK2) to inhibit translation and initiate stress granule formation [36].

RP2 ALU transcripts have different roles, acting in *cis* to alter RNA processing and turnover [37]. The ALU elements stimulate alternative splicing. They enable the production of circular RNAs (cRNAs),



**Figure 1.** ALU sequences, non-B-DNA conformations and modifications. (a) An Alu dimer consisting of left and right arms with a triplex forming adenosine-rich spacer. The A- and B-boxes promote transcription by RNA polymerase III. Triplex [3] and Z-DNA formation [4] dynamically change the information read-out from the genome, making different outcomes possible. ALUs can also potentially stack two-layered intramolecular quadruplexes to form G4Q wires [5]. DNA models are displayed from the following sources: triplex model (PDB:TRX3) with the third strand in blue, Z-DNA (PDB:4LB5), a parallel strand G4Q (PDB:1NYD), B-DNA (PDB:1BNA). (b) Sequence Weblogos (generated at <https://weblogo.berkeley.edu/logo.cgi>). The upper logo is from the consensus ALU sequences with canonical triplex and Z-forming sequences boxed by solid lines (table 1 in [1]). The dotted box contains sequences where base modifications decrease the energetic cost of flipping their conformation from right- to the left-handed DNA. Cytidine residues with the potential to undergo methylation and further oxidation are marked with an 'X' [6] while those favouring oxidation of guanosine to form 8-oxo-dG are marked with 'Y' [7], resulting in an 'XY' DNA sensor for oxidative stress. The lower left logo summarizes experimentally determined, triplex-forming ALU sequences. The difference between this sequence and the consensus sequence reflects the evolutionary adaptations that improve targeting of RNA strands to specific ALU duplexes [8]. The lower right logo captures the sequence reported to form an intramolecular quadruplex with the four G quartet forming residues boxed [5].

which act as sponges for microRNA to alter phenotype [38]. RP2 also transcribes short (100–300 bp) intronic ALUs that enhance phase transitions leading to nucleolus formation [39,40] and sets of AluACA transcripts with unknown function [41]. ALUs form double-stranded RNAs (dsRNA) when inverted repeat elements (AIRE) are close enough to base-pair with each other. AIRE localize messenger RNAs to nuclear paraspeckles [42] and cytoplasmic stress granules [43]. They are substrates for dsRNA editing [19–21]. The different ALU roles reflect an evolutionary history with layers upon layers of adaptations. A similar evolutionary complexity is attached to ALU NoBs.

## 4. What are non-B-DNA structures?

Here I discuss Z-DNA, triplexes and quadruplexes (figure 1), referencing the work of many talented scientists. Z-DNA duplexes are twisted to the left rather than the right as in B-DNA. Z-DNA forms directly from B-DNA by flipping base-pairs over, causing the strands connecting the bases to zig-zag to the left. Triplexes are right-handed with a third strand wrapped around the two strands of B-DNA. Quadruplexes are usually right-handed with four strands of DNA wrapped around each other. The most studied quadruplexes are built from a quartet of guanosine hydrogen-bonded to each other (G4Q). Here I focus on their role in the biology of ALUs, trying as we proceed to separate evidence from *in vitro*, *in vivo* and *in silico* studies from other sources. The reader is reminded that this field is rapidly evolving with many recently developed technologies paving the way to new discoveries. The discussion is intended as a guide to future experimentation.

## 5. Why are non-B-DNA structures interesting?

Alternative DNA structures epitomize a non-traditional way of encoding genetic information [44,45]. I refer to the sequences that dynamically switch conformation under physiological conditions as flipons [44]. Flipons trade energy for information [44]. They are dissipative structures [46], requiring work to change their conformation. The flip enables targeting of the cellular machinery to active genes instead of to other sequences. By localizing RNA processing complexes, flipons alter the transcripts compiled from a gene. The RNAs produced depend on context and change with flipon conformation.

Much of this discussion is an interpretive review of the literature, intended to highlight what we don't know and where we need to explore. The presentation starts with the following question. If NoBs are so important *in vivo*, why is this entire field still controversial? Many readers will be aware of the view that NoBs lack biological relevance. Indeed NoBs were long ago penned into history as a biological dead end, at least for Z-DNA, first by the professional journal editors [47] and then by the philosophers [48]. The opinion is largely based on the belief that absence of proof is equivalent to proof of absence [49], the famous black swan fallacy. Each to their own measure of success, but the case for the biological relevance of NoBs is now very strong.

## 6. Why is the science of non-B-DNA structures difficult?

The study of NoBs has always been hard because of the many different sources of potential experimental bias. The question with NoBs experiments is always the same: are the findings valid or just an artefact? Experiments on NoBs are easily confounded. NoBs can form *in vitro* during procedures designed to show their presence *in vivo*. Procedures may inadvertently release bound histones and generate enough negative supercoiling (NSC) to either flip right-handed DNA to the left-handed conformation or to create single-stranded regions that fold as G4Q [50,51]. An example of this effect is the binding of Z-DNA-specific antibodies to *Drosophila* polytene chromosomes. Results depend on fixation conditions [52]. Similar problems are reported for antibody staining of G4Q [51], which are remarkably stable under a wide range of conditions [53]. With *in vivo* approaches, introduction of a Z-DNA binding protein or a G4Q binding drug into a cell may stabilize the conformation in an unphysiological manner. In some NoBs studies, artefacts only become known with time. One example is the co-immunoprecipitation of RNA complexes that form by rapid exchange of RNAs between proteins only after cell lysis [54]. Another is the preferential enrichment by CHIP-Seq, regardless of the antibody used, of certain sequences. The problem was first identified through the efforts of the Encode Project and summarized by a sequence blacklist that include examples like centromeric repeats [55]. Earlier results based on this technique that relate to NoBs must therefore be interpreted cautiously [56]. There is also a reagent bias. For example, permanganate only modifies thymine bases [57] and kethoxal only reacts with guanosine bases [58]. Together they provide a complementary approach for footprinting single-stranded regions associated with NoBs *in vivo* but individually may miss certain sequences. A different type of bias arises from data missing due to experimental design: it is routine to exclude from genome-wide analyses sequences like ALU. They are removed from analysis of d(CpG) modifications by protocol [59] and also from some *in vivo* footprinting assay results [45]. Also heavily edited ALUs are often discarded without analysis as they do not map to the genome. The proper pipeline is needed to study such sequences [60]. To increase our understanding of flipon biology, we need to use different data collection protocols.

The biochemical identification of NoBs binding proteins is also challenging and requires rigorous controls with structural studies to confirm findings. For Z-DNA, there have been many solo publications in which conformation-specific proteins are claimed, but where verification by biochemical and structural studies is absent [61–63]. Subsequent studies often show these proteins to have a different functions altogether, unrelated to NoBs. One example is Zuotin, a ribosomal co-chaperone for nascent peptides originally proposed as a Z-DNA binding protein [64]. The same is true for the claim that most Z-DNA binding proteins are actually lipid binding proteins [65]. Even when good reagents are available against proven Z-DNA binding proteins, like the p150 isoform of the adenosine deaminase RNA-specific enzyme ADAR, incorrect statements echo through the literature. The finding that protein p150 location in a limited number of tumour cells is cytoplasmic after interferon induction has been highly cited [66], despite the clear nuclear localization of p150 demonstrated in the original publication [66]. The presence of p150 in both the cytosol and the nucleus is consistent with the nucleocytoplasmic shuttling of p150 [67–69]. Recent surveys of normal



tissues show that p150 is mostly nuclear in murine thymus, spleen and prefrontal cortex [70–72]. Conversely, the cytoplasmic localization during stress of ADAR p110 isoform that lacks the Z-DNA binding domain, is not often referenced [73].

Additional challenges to identifying conformation-specific non-B binding proteins also exist. Proteins like the high-mobility group family are disordered random coils in solution and only fold on binding DNA. They engage a variety of NoBs, including triplexes and quadruplexes, none specifically [74,75]. Another issue arises with sequences that adopt more than one NoB, raising the question as to which interaction is biologically relevant. For example, the telomeric sequences that form G4Q *in vitro* stabilize a triplex *in vivo* in which the G-strand overhang folds back onto a telomere proximal sequence to form a T-loop [76,77]. There is also debate as to whether the relevant NoBs are formed from DNA, RNA or a hybrid. The direct answer is that all three classes of NoB duplexes can serve an informational role. The better response is to focus on the transient nature of the NoBs formed *in vivo*. They allow a cell to switch from one phenotypic state to another. Their persistence is often associated with disease.

The work on Z-DNA satisfies the necessary criteria for demonstrating the *in vivo* relevance of NoBs: *in vitro* biochemical validation to show function; angstrom resolution structures to eliminate ersatz explanations; *in vivo* studies to confirm mechanism; Mendelian genetics to confirm disease causality. The development of well-controlled assays [78] enabled the characterization of the Z $\alpha$  domain in the double-stranded RNA editing enzyme ADAR p150 isoform, then the identification of a host of homologues. ADAR Z $\alpha$  binds specifically to Z-DNA without sequence preference [79–81]. The initial biochemical characterization demonstrating structure-specific binding [81] was confirmed by NMR and crystallographic studies [79,82,83]. Mendelian genetics and biochemical characterization of the disease-causing variants then established a biological role for the Z-duplex in the negative regulation of type I interferon responses [84]. Recent publications show that many cancers exploit this property to silence immune responses [26].

Further validation for a role for the Z-duplex in biology is provided by studies of the related protein, Z-DNA binding protein 1 (ZBP1, also known as DAI or DLM-1) [85–89]. ZBP1 has two Z $\alpha$  domains, the first of which is absent in many splice isoforms [90,91]. Deletion or mutation of Z $\alpha$  residues in ZBP1 confirm a central role for the Z-conformation in RIPK3-dependent necroptosis that occurs in inflammatory bowel disease (IBD), dermatitis and influenza virus infection. Mutations to Z-binding residues in the second Z $\alpha$  domain ameliorate these phenotypes [85–87,89,92]. E3L, a vaccinia virus Z $\alpha$  orthologue, was recently shown to inhibit ZBP1-induced necroptosis during infection [93]. ZBP1 plays other roles in other cell death pathways involving apoptosis and pyroptosis [92]. The interactions between ADAR p150 and ZBP1 and their specific roles requires further study. One possible generalization is that ZBP1 induces early inflammatory responses that are later negatively regulated by ADAR.

## 7. What is the connection between ALUs and flipons?

Z $\alpha$ -dependent RNA editing of ALU RNAs by ADAR provides evidence that Z-flipons regulate biological responses such as the type I interferon responses in innate immunity [4]. The finding raises many questions. Has natural selection exploited the spread through the genome of Z-flipons by ALU to regulate additional pathways? If so, is the same true for other classes of flipon? What is the evidence? We have only partial answers to these questions at this time. Not all the necessary experiments have been done and not all the relevant data collected. What follows is an interpretation of the data we have with many of the gaps highlighted as hypotheses that require experimental evaluation.

## 8. Z-flipons

ALU include sequences where the base-pairs flip over to form Z-DNA (Z-Box, figure 1b), a structure in which the double-helical backbone zags left instead of twisting right as it does in Watson–Crick B-DNA [4]. The energetic cost of the transition varies with the length and nucleotide composition of a DNA segment. Z-DNA formation is favoured with alternating pyrimidine, purine sequences such as d(CG) repeats [4]. Not all sequences can flip under physiological conditions, giving informational value to those that can. Predicted Z-DNA forming sites in the human genome are present at transcription start sites (TSS) of the annotated genes on chromosome 22 [94]. Z-DNA forms in the MYC promoter as measured by Z-DNA-specific antibodies [95] and stabilizes the open chromatin structure at the CSF1

promoter, a process dependent upon the BAF complex (BRG1- or HBRM-associated factors) [96]. A genome-wide *in silico* survey reveals that the majority of predicted Z-DNA forming elements lie in repeat elements, either as simple sequence repeats such as  $d(AC)_n$  and  $d(GT)_n$  or in the right arm of ALU dimers, with the propensity to form Z-DNA varying by ALU family (figure 1) [4].

The existence of proteins that bind Z-DNA with high affinity provide evidence for the biological relevance of this conformation. The best-known example involves the p150 isoform of ADAR (adenosine deaminase RNA-specific enzyme) with its  $Z\alpha$  domain that binds Z-DNA in a structure-specific but sequence-independent manner [79,83,97], even to sequences like  $d(AT)_n$  that do not form Z-DNA easily *in vitro* [81]. Mutations to the amino acid residues contacting Z-DNA, like N173S and P193A, are causal for the type I interferonopathies characteristic of Mendelian diseases, such as Aicardi–Goutières syndrome [84], revealing an important regulatory role for the Z-duplex in innate immunity. The P193A mutation is associated with reduced double-stranded RNA (dsRNA) editing [98] and increased induction of the interferon response factor 3 (IRF3) gene [99].

Studies have revealed that ADAR  $Z\alpha$  binds a number of genomic sites. ChIP-Seq based on an ADAR  $Z\alpha$  dimer identified a total of 391 Z-forming sequences, mostly (46%) in promoters and associated with active histone marks such as H3K4me3 and H3K9ac [100]. The mapped genes are enriched for replication-dependent histones prone to citrullination [101], cell cycle factors, transcription factors and topoisomerase 3A, which is essential for resolution of D-loop recombination intermediates [102]. The centromeric repeats found in an earlier study [56] are most likely Encode blacklist sequences [55], as they lack a strong Z-forming motif. A recent study of the ADAR  $Z\alpha$  regulation of fear extinction in mice found that 80% (97/122) of RNA editing sites overlap with Adar1 DNA binding, with an overrepresentation of SINE/LINE elements nearby. Both DNA binding and RNA editing were absent when a loss-of-function  $Z\alpha$  (N175A/Y179A) mutant was tested in the *in vivo* assay [70]. This response also depends upon prion formation by the cytoplasmic polyadenylation element binding protein CPEB3 [103]. Whether Z-formation affects memory by altering CPEB3 splicing or the processing of miRNA loci that regulate CPEB3 expression is unknown at present.

ZBP1  $Z\alpha$ -dependent functions have been investigated in a number of studies. In mouse models, ZBP1 activates necroptosis mediated by RIPK3–MLKL [86,89,92]. In these studies, tissue-specific gene knockouts were used to induce disease by preventing RIPK1/FADD complex-induced inactivation of ZBP1/RIPK3 kinase signalling through caspase 8 activation. Mouse models used epidermis-specific RIPK1 deficiency to produce skin inflammation and intestinal epithelial-specific FADD deficiency to induce colitis. The necroptosis produced was due to Z-binding by ZBP1 as either deletion or mutation of the  $Z\alpha$  domains prevented disease. Activation of ZBP1 was by endogenous retroelement expression and dependent on the second  $Z\alpha$  domain. In skin lesions, SINEs were upregulated and long terminal repeat elements downregulated relative to normal tissue [86]. Skin lesions were partially ameliorated by a combination of reverse-transcriptase inhibitors, suggesting that DNA/RNA hybrids can trigger ZBP1 [86]. Activation of ZBP1 by endogenous retroviruses (ERV) is also reported in human inflammatory bowel disease (IBD) patients, possibly due to low expression levels of the repressive methyltransferase SETDB1 [87]. In mouse models with selective *Setdb1* knockout in intestinal stem cells, increased levels of ERV in diseased tissue relative to normal was also present. In models of influenza viral infection, where the virus replicates in the nucleus, ZBP1 binds defective viral genomes in immunoprecipitation assays, although binding to ALU elements or other retroelements was not evaluated. Forced nuclear localization of ZBP1 in this model increased necroptosis [88]. Collectively these results reveal an important role for ZBP1 and endogenous flipons in regulating cell death.

What initiates a change in Z-flipons conformation? The unwinding of B-DNA powers the flip from B-DNA to Z-DNA. Enzymes generate the required energy as they work to transcribe RNA. Polymerases generate negative supercoils (NSC) faster than topoisomerases relax them [45,104–106]. The negative supercoiling they produce accumulates upstream of the polymerase and stabilizes the left-handed conformation [107]. NSC is highest at the TSS where Z-DNA forming sequences are found [94–96]. The topoisomerase I activity here is lower (as judged by use of camptothecin to trap catalytically active enzyme on DNA), even though levels of the enzyme are much higher than in gene bodies (as judged by ChIP-Seq) [108]. These effects are enhanced when upstream DNA is anchored to limit their dissipation along the helix. Anchoring can be either by scaffold attachment or by DNA:RNA hybrid formation [109]. The energy stored in NSC and Z-DNA is then available for assembly of transcriptional complexes and to load RNA polymerase at the promoter [94–96].

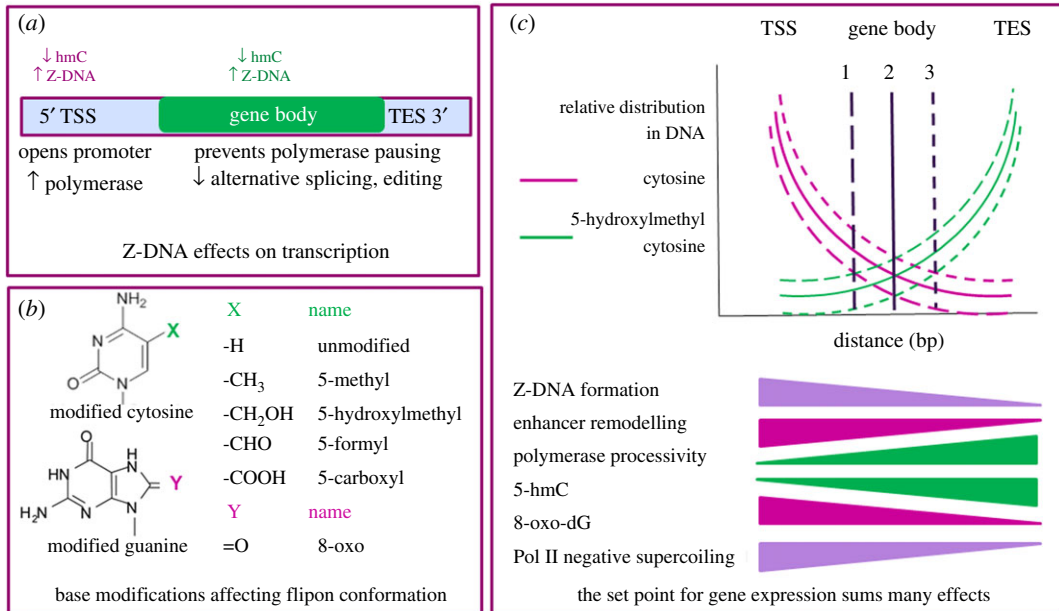
The energetic cost of Z-formation *in vitro* is lowered by polyamines such as spermine and spermidine, but not by acetylspermidine or putrescine [110]. The levels of spermine and spermidine vary during the cell cycle and increase during oxidative stress [111] with catabolism due to acetylation by spermidine/spermine

N1-acetyltransferase 1 (SAT1). Polyamines exist in cells as nuclear aggregates (NAPs) [112]. *In vitro*, NAPs can promote Z-DNA formation and form higher order structures that wrap about DNA to form fibrils. The biological role of fibrils is not clear. One untested possibility involves a role in the production of neutrophil extracellular traps (NETs). NETs are made of chromatin structures from neutrophils extruded into the extracellular space as a result of pyroptosis, which is induced through ZBP1 interactions with RIPK1 and dependent on NFκB activation of the NLRP3 inflammasome pathway [113–115]. The NETs are stabilized by chlorinated polyamines and are pro-inflammatory [116]. *In principle*, NAPs could activate ZBP1 by promoting Z-DNA formation and lead to NET formation via gasdermin [117]. A similar outcome could account for a rare monogenic form of systemic lupus erythematosus (SLE), a disease in which anti-Z-DNA antibodies are frequent [118,119]. In two multiplex families, an X-linked recessive form of SLE was due to SAT1 loss of function variants that would be anticipated to increase spermine levels. In general, NET degradation is impaired in SLE, [120] and increases the risk of lupus nephritis [121].

The energetic cost of Z-DNA formation *in vitro* is lowered by a number of base modifications. These include the 8-oxo-7,8-dihydro-2'-deoxyguanosine (commonly called 8-oxo-dG) adduct [7], which is produced by a non-enzymatic, chemical reaction during oxidative stress [122]. ALUs are particularly susceptible to formation of 8-oxo-dG due to the presence of many cytidine, guanosine dinucleotide steps that undergo rapid oxidation [123]. These steps also favour Z-DNA formation. The combination of sequence and 8-oxo-dG in the Z-box then enhances a flip to the Z-conformation (figure 1*b*). Interestingly, ALU elements in cells are highly susceptible to X-ray-induced oxidative damage, but in a survey of tumours, there is no increase in the incidence of single nucleotide variations compared to the flanking sequences (fig. 5*i* in [122]), suggesting that the ALU sequence containing the Z-DNA forming motif is constrained by function.

Modification of cytidine in the 5-position also alters the *in vitro* energetics of Z-DNA formation in an unexpected manner [6]. While 5-methylcytosine (5mC) is a well-known modification, base oxidation produces other derivatives of cytidine (C). These lesser known products arise during oxidative stress [124] and through the action of the ten-eleven-translocation (TET) dioxygenase protein family, which promotes the sequential formation from 5mC of 5-hydroxycytosine (5hC), 5-formylcytosine (5fC) and then 5-carboxylcytosine (5cC). The reactions occur in the order C → 5mC → 5hC → 5fC → 5cC → C (figure 1*b*) [125]. While 5mC promotes the Z-DNA flip *in vitro*, 5hC inhibits the transition, stabilizing the B-DNA flipon conformation instead [6,126]. The effect of 5fC and 5cC on Z-DNA formation is interesting. If both cytosines in a CpG step are modified, there is a steric clash between the adducts that inhibits Z-DNA formation [126] (figures 2*b* and 3*a*). If the modified cytosines are spaced to prevent the clash, the flip to Z-DNA is actually enhanced [6]. The Z-forming potential of a CpG island then varies with the density of methylation. At high levels of 5mC, flipons will probably be locked into the B-conformation by cytosine oxidation, due to steric clashes preventing Z-DNA formation. In hemi-methylated regions, clashes will no longer occur, with 5fC and 5cC now favouring the flip to Z-DNA. Hemi-methylated d(CpG) dinucleotides are found *in vivo*. They flank CCCTC-binding factor (CTCF) sites that anchor chromatin loops [127,128] and are maintained during replication [127]. The hemi-CpGs cluster together and are present only on one strand, with rotational symmetry about the CTCF motif. They are enriched in gene bodies, but not promoter regions, which are unmethylated [127]. A different way to relieve d(5fCpG) and d(5cCpG) steric classes is by transcription. The modifications then exist only on the DNA strand with the RNA unmodified in nascent transcripts. The cytosine adducts can then enhance Z-DNA formation by the hybrid, and promote any Z-dependent RNA processing. The regulation of Z-flipon conformation *in vivo* by 5hC and its oxidative products needs experimental validation. There is indirect evidence for this possibility. The ratio of 5fC and 5cC to 5hC varies between different repeat elements, between young and old members of a repeat families and between different cell types, suggesting active control of this reaction pathway [129]. The variation depends in part on levels of the repair enzyme thymine DNA glycosylase (TDG), which excises 5fC and 5cC from DNA. The reaction is stimulated by the base excision repair enzyme NEIL1 (nei like DNA glycosylase 1), even when NEIL1 is catalytically dead. Currently it is believed that 5hC removal in animals requires its oxidation to 5fC [129,130].

Cytosine adducts potentially impact gene expression in a number of ways. They may directly influence placement of nucleosomes as 5fC forms a Schiff's base with H3 histone tails during the assembly of octamers [131]. Other effects of these modifications depend on their location in a gene. In the case of 5hC, experimentally measured transcription decreases when 5hC accumulates in the promoter region, but increases when 5hC accrues in the gene body (figure 2*c*) [132]. One testable explanation for the apparently contradictory effects on transcription involves the inhibition of Z-DNA formation by 5hC. At the promoter, low levels of 5hC probably favours Z-DNA formation, potentiating a role in enhancer and promoter assembly. In contrast, high levels of 5hC in the gene body potentially diminish inhibition of RP2 processivity by Z-DNA, thereby increasing transcription rate [4].



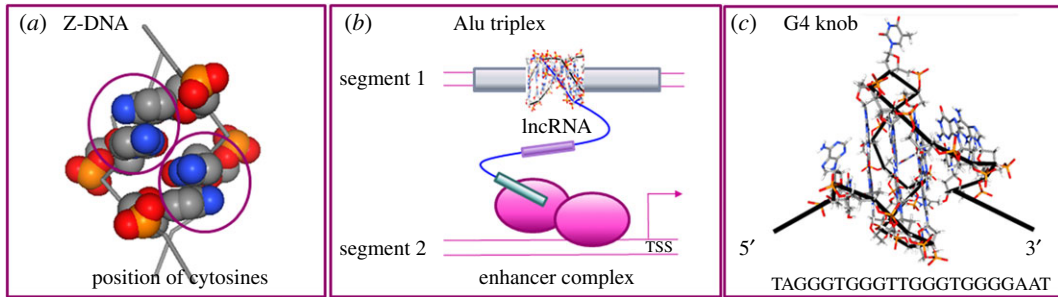
**Figure 2.** Z-DNA formation. (a) 5hC modifications lower Z-DNA formation by a flipon sequence. The consequences depend on whether 5hC forms in the promoter region where it inhibits promoter opening associated with Z-formation or in the gene body where it promotes RNA polymerase processivity by preventing Z-formation. (TSS, transcription start site; TES transcription end site) [132]. (b) The modifications to cytosine (labelled X as in figure 1) and guanine (labelled Y as in figure 1) described in the text. (c) Differences in negative supercoiling and base modifications, as indicated by the wedges, determine the optimal set points for expression of different messages from each gene. The conformation of Z-flipons within a gene body allows a dynamic change to the transcripts compiled from a gene. The vertical lines represent the best set points for three different genes (gene 1: long dash, gene 2: no dash, gene 3: short dash). The vertical lines (labelled by gene number) intersect the other two transcript-specific lines at the point where read-out is optimal for each gene, balancing enhancer and transcriptional activity with the extent of base modification. (8-oxo-dG: 8-oxo-7,8-dihydro-2'-deoxyguanosine).

Low rates of polymerase pausing are associated with a decrease in alternative RNA splicing [133,134]. Z-flipon-induced alternative splicing depends then, *in principle*, on removal of 5hC to allow the sequence to flip. Formation of Z-DNA pauses RP2 and provides time to fold the RNA and assemble a RNA splicing complex. Alternative splicing in T-Cells is associated with CTCF, which as described above, has binding sites flanked by tracts of hemi-methylated d(CpG) [127]. CTCF has greatest affinity for sites containing 5cC and binds poorly to 5mC, 5hc and 5fc [133]. Whether the effects of 5cC on CTCF-dependent alternative splicing are causal, coincidental or an effect is currently unknown. One scenario is that 5cC promotes Z-formation adjacent to the CTCF binding site. Subsequently, RP2 pauses, CTCF docks via 5cC and scans for its cognate motif. Once CTCF engages, alternative RNA splicing initiates.

Modification of 5hC to 5cC then links flipon conformation to altered RNA processing. As an analogy, Z-flipons within genes act similarly to traffic signals distributed along a highway. When lights are all set to green, traffic speed is high as there is nothing to interrupt the flow. Turning lights to red stops the traffic, enabling pedestrians to cross or cars to enter from other directions. Controlling each light with a programmable switch, like a flipon, enables optimization for particular outcomes. The most adaptable signal implementation is one in which switches are set independently. That is how an engineer would design the system. Nature probably began with flipons that were initially in sync with ALUs from the same family as they all interact with the same factors. With time, more adaptive permutations arose through natural selection.

Collectively, the different types of epigenetic modifications alter the B-Z equilibrium, changing the transcripts produced from a gene. Chemical modifications, like 8-oxo-dG and 5hC modulate the flip to 'Z' and probably affect RNA processing *in vivo*. A number of proteins also affect outcomes. Polymerase pausing due to Z-formation stalls the trailing RP2 [135], providing time for splicing and editing substrates to fold and protein complexes to assemble (figure 2c) [132]. Topoisomerase activity in the gene body controls transcription rate [108], potentially by limiting Z-DNA formation. TET, DNMT3A (DNA methyltransferases 3A) and repair enzymes like TDG, NEIL1 and oxoguanine glycosylase (OGG1) determine the extent and persistence of adducts, potentially regulating their





**Figure 3.** Flipon conformations. (a) A Z-DNA helix with the cytosines highlighted by a spacefill representation. The approximation of the C5 residues from adjacent steps is circled, and illustrates the potential for steric clashes with 5fC and 5cC (blue: hydrogen, red: oxygen, grey: carbon, yellow: phosphorus) (PDB:4LB5) [126]. (b) A triplex bound by a lncRNA directs assembly of complexes according to the motifs present in the RNA (indicated by coloured boxes), either by local looping or through interactions with a non-contiguous DNA segment. (c) A G4Q formed from a single DNA strand that emphasizes its distinctive non-B-DNA structure and loops with exposed unpaired bases susceptible to modification (PDB: 2LXQ).

steady state levels. It is therefore reasonable to propose that the optimal setpoint for any particular gene in a specific tissue with respect to transcription and RNA processing is altered by a wide range of factors that impact NoBs formation. The best permutation of flipon conformations and the transcripts compiled then vary by context (figure 2c), changing the information extracted from their genomic neighbourhood.

How do these events relate to ALU conformation? In addition to the Z-Box [4] (figure 1b), ALUs contain a second potential Z-forming sequence, the XY-box (figure 1b, XY-box within the dotted line where 'X' over the base represents a cytosine variant and 'Y' a guanine adduct). The energetic cost of flipping the 'XY' box is Z-DNA is much higher than for the Z-Box. For the flip to occur under physiological conditions it is likely that oxidative cytosine and guanine modifications in the XY-box are required. The XY-box lies immediately 5' to the canonical Z-Box. These two segments potentially form an extended ALU Z-DNA 'XYZ' block, with the inducible 'XY' flipon sequence acting as a sensor to signal oxidative stresses that cause base modifications (figure 1b). The increased length of Z-DNA might preferentially trigger DNA repair or chromatin modifications that silence the damaged region, as reported for the ADAM12 gene [136]. The 'XYZ' flip might also trigger cell-death via ZBP1 when cell-survival pathways are overwhelmed. The interplay between DNA modification and Z-DNA formation highlights an important potential regulatory role in DNA damage responses for low-complexity ALU sequences. The XYZ boxes extend the role ALUs play in innate and cancer immunity [26,84]. The widespread distribution of ALUs enables genome-wide network for surveillance of DNA damage.

What is the role of epigenetic protein modifications in flipon biology? In addition to base modifications, Z-DNA formation is associated with histone epigenetic marks that regulate gene read-out. An increase in Z-DNA formation *in vivo* occurs following trichostatin A inhibition of histone deacetylases, reflecting the boost in transcriptional activity [137]. Potential Z-DNA forming sequences overlap DNA motifs, such as the tetranucleotide d(CGCG), that are present in regions of unmethylated DNA and associated with the active histone marks H3K4me3 (histone 3 lysine 4 trimethylation) and H3K27ac (histone 3 lysine 27 acetylation) [138]. Whether flipons help set chromatin states is unknown [139]. The best possible example so far is the CSF1 promoter, whose open chromatin state depends on the BAF complex and Z-forming sequences [96]. There is a paucity of data on whether Z-flipons affect ALU chromatin in a similar manner. However, ALU DNA sequences are strongly associated with active histone marks [140], they bind over 75 transcription factors [16] and they form Z-DNA [4]. Use of ALU Z-flipons to regulate histone state and transcription factor engagement is a reasonable proposal, one that requires further exploration.

What is the interplay between B- and Z-conformations of a flipon? The B-conformation of ALU is bound in a sequence-specific manner by many proteins. For example, the TET enzymes involved in cytosine oxidations bind strong Z-DNA forming sequences as B-DNA (e.g. the TET1 motif is (A)CGCGCT [141]), with a dissociation constant in the low micromolar range [142]. Similarly, E-box binding transcription factors like MAX (MYC associated factor) also bind an alternating purine/pyrimidine motif ((C)CACGTG) capable of forming Z-DNA at NSC levels found *in vivo*. The E-box proteins bind these sequences as B-DNA with nanomolar affinity. The ADAR  $\alpha$  domain binds the same sequences MAX binds with similar low nanomolar affinity, but to the Z-conformation [79]. Both

B- and Z-DNA protein binding to these sequences is affected by epigenetic modifications in the same manner. The affinity of E-boxes is higher for 5mC modification ( $K_d = 11$  nM) than for 5hmC by a factor of 10 ( $K_d > 110$  nM) [143], while Z-formation is inhibited by 5hmC and promoted by 5mC, as described above [6,126]. What then is the molecular choreography that explains how B-DNA, sequence-specific E-Box proteins and Z-DNA, structure-specific  $Z\alpha$  domains target the same sequence? A plausible explanation is that flipons help localize transcription factors to the sequence. One possible model, but not so far experimentally addressed, involves Z-Z junctions formed by out of alternation base-pairs in the left-handed duplex. The junctions cause partial melting of the helix, producing a kink [81,144]. This distortion accentuates any base modification present. The adducts then could act as flags to initiate docking of the E-Box protein that then scans the surrounding sequence [145]. If its motif is present, the protein engages and flips the sequence back to the B-conformation. In ALU, the two Z-Z junctions are formed by an out of alternation d(GTG) sequence that separates the XY- and Z-boxes. Modification of both cytosines on the complementary d(CAC) strand will promote the flip and flag the region for the E-Box protein docking. A number of transcription factors have specificity for sequence variants of this ALU 'Z-Z' seed region [146]. If a sequence match exists, the energy stored as Z-DNA is then available to power enhancersome assembly. A dock and search strategy based on Z-flipons provides *in principle* an efficient mechanism for shrinking the search space for transcription factors to explore. It targets the cellular machinery to active chromatin.

If Z-formation is needed to assemble enhancersome in normal cells, then what leads to Z-formation prior to the initiation of transcription? Base modification due to oxidative processes or mutagens that promote flipping is one way *in principle* to lower the energetic barrier to Z-formation. Modifications by PAD peptidyl arginine deiminases is another way. These enzymes catalyse arginine deamination and citrullination of histone tails. They generate NSC by unwrapping DNA from histones cores [101]. Chromatin remodelling is a further option. In this highly dynamic process, ATP-dependent motors slide histone octamers along the DNA duplex to alter nucleosome phasing. In the process, they create nucleosome-free regions [147] and NSC that is probably sufficient for Z-formation [96]. Such chromatin reorganization is essential during development [148]. Its initiation depends upon a set of specific epigenetic marks present on the pioneer nucleosomes bound with high affinity to G-rich sequences transmitted from sperm [148–150]. In the case of the CSF1 promoter, this type of remodelling is sufficient to initiate Z-formation in cells [96]. A similar process probably accounts for the epigenetic memory associated with interferon responses [151].

In addition to changes in chromatin structure, Z-flipons alter the information content of RNA transcripts by enhancing base-modification. One example involves the p150 isoform of the ADAR. ADAR isoforms deaminate adenosine in regions of dsRNA to give inosine, which is read out as guanosine. A major subset of p150 substrates originate from chromosomal segments with AIRE that lie close to each other (less than 5000 bp separation) [19,60]. Transcripts from these regions are mostly products of RP2 [2]. They form double-stranded RNA editing substrates by folding back on themselves. *In silico* analysis reveals that editing efficiency varies with the propensity of ALU sequences to form Z-DNA [4]. Z-formation enables recognition of ALU substrates by the p150  $Z\alpha$  domain that binds the Z-duplex in a structure-specific, but sequence-independent manner [81,152]. This process has recently been documented in mouse prefrontal cortex [70]. Whether it is ALU Z-DNA, Z-RNA, a Z-DNA:RNA hybrid that the domain binds is still an open question as  $Z\alpha$  binds all things 'Z'. Binding to the DNA/RNA hybrid formed during transcription potentially facilitates the handoff of ADAR from Z-DNA to Z-RNA by pausing RNA polymerases long enough for the editing substrate to finish folding and for the dsRNA binding domains to engage. By latching ADAR to its dsRNA editing substrates, the  $Z\alpha$  domain improves the efficiency of editing by the ADAR catalytic domain [152].

Potential Z-specific substrates that do not rely on AIRE involve those with nonsynonymous edits conserved between human and mice [153]. Examples include DACT3 (R259G), a regulator of  $\beta$ -catenin and MEX3B (Q189R), a regulator of translation and the cyclin dependent kinase CDK13 (Q103R). Both DACT3 and MEX3B show Z-specific permanganate hypersensitivity in the Burkitt's lymphoma Raji cells, but CDK13 does not [45]. DACT3 editing along with that for CDK13 (Q103R) is reduced in mice in which ADAR is catalytically dead. It is enhanced in the spleens of normal mice where the ADAR p150 is the major isoform expressed [72]. The functional consequences of these edits require further study. In a few cases, editing by the ADAR family of enzymes can compensate for otherwise deleterious G-to-A codon mutations. [154,155]. The mutations necessitate the retention of ADAR enzymes in the genome. Here, flipons potentially compensate for maladaptive codon variants [155].

Besides acting as substrates for editing, AIRE dsRNAs also act as enhancers. An analysis of 10 650 edited sites in non-repeat sequences revealed a higher than expected presence of edited AIRE in 2 kb windows either side of the modified adenosine, more often downstream (56%) than upstream (44%) [156]. *In vitro* assays with model substrates revealed that AIRE dsRNAs increase editing efficiency of short dsRNA that would not otherwise undergo modification. One substrate analysed was NEIL1, whose editing *in vivo* is interferon-induced and almost certainly due to the ADAR p150 isoform [157]. The K242R recoding shifts enzyme substrate specificity from thymine glycol to guanidinohydantoin, a highly mutagenic oxidation product of 8-oxo-dG [158]. The ALU inverted repeat elements that enhance editing are separated by a strong Z-forming d(AC)<sub>n</sub> dinucleotide repeat sequence. Any p150-dependent editing of the transcript is probably due to either recognition of Z-DNA or a Z-DNA/Z-RNA hybrid, as the dinucleotide repeat remains single-stranded in the folded RNA editing substrate. Further study of this particular editing site will help define the important parameters for Z-dependent editing.

How does Z-dependent RNA editing affect phenotype? RNA editing impacts the proteome of a cell in many ways beyond the recoding of codons. Base modification changes secondary structure, stability, splicing and the sequence-specific pairing with other RNAs. It also impacts processing of microRNA precursors and the function of long non-coding RNAs (lncRNAs) [159]. Failure to edit dsRNAs is causal for Mendelian type I interferonopathies and arises from 'loss of function' mutations to the ADAR catalytic domain [160] and to its Z $\alpha$  domain [84]. The Z $\alpha$  P193A variant is associated with reduced dsRNA editing [98] and increased induction of the IRF3 gene [99]. The Z $\alpha$  mutation may also affect outcomes that are independent of editing. ADAR p150, but not p110, interacts with the RNA helicase DHX9 to suppress crRNA formation by AIRE without an increase in detectable RNA edits [161]. ADAR constructs lacking the catalytic domain target enzyme complexes involved in microRNA production that involve interactions with drosha ribonuclease III (DROSHA) and dicer 1 ribonuclease (DICER), either directly or by binding components like the DGCR8 microprocessor complex subunit [4,161–165]. ADAR p150 and p110 isoforms target different steps in miRNA processing [166,167], suggesting that some Z-flipons pattern phenotype by regulating microRNA levels. Z-flipons may explain the small size of mice in which the lethality of a catalytically dead ADAR enzyme is overcome by genetic inactivation of the type I interferon pathway [168].

## 9. T-flipons

Triplex formation is favoured in purine-rich dsDNA segments and involves docking of an additional nucleic acid strand almost always into the major groove of B-DNA. The third strand can be either purine (anti-parallel to the dsDNA purine strand, RYR (R: purine, Y: pyrimidine)) or pyrimidine (parallel to the dsDNA purine strand, YRY) [169]. Triplexes can form by intramolecular rearrangement when NSC of plasmids produce single-stranded regions that fold back onto a base-matched duplex region to form H-DNA [169]. Formation is favoured under acidic conditions (pH 4) with cytosines protonated [170]. The discussion here refers to intermolecular triplexes. These structures are only slightly underwound (11–12 bp per turn versus 10.4 bp in B-DNA) and not directly dependent on NSC for formation. Indirectly, supercoiling helps correctly align the triplex strands by increasing the flexibility of chromatin loops [171,172].

Most commonly the third triplex strand is RNA, while the duplex can be dsDNA, dsRNA or a DNA-RNA hybrid. The use of DNA as the third strand is energetically disfavoured when duplexes are made with a RNA purine strand [173]. *In vitro* triplex thermal stability depends on its base-composition and its length [174]. Formation of dC-dG:dC (': indicates non-Watson-Crick base-pairing to the third strand) triplets is pH dependent, as it depends on protonation of cytosines on the third strand. Triplets can also incorporate non-canonical triplets such as G-C:U [175] and G-C:A [169]. In general, YRY structures are more stable than RYR as the triplets are isomorphous, stacking without distorting the helical backbone [176].

Recent studies indicate a role for triplexes in gene expression with RNA as the third strand (TSR) (figure 3b). Triplex formation by the MEG3 lncRNA represses TGF $\beta$  (transforming growth factor beta) pathway gene expression by targeting the polycomb repressive complex 2 to these genes. The lncRNA uses a d(GA)<sub>n</sub>-rich sequence motif to bind the target genes [177]. Another triplex-forming lncRNA, PARTICLE, induces locus-specific methylation in response to low-dose irradiation [178]. In some cases, regulation is quite complex. With the sphingosine kinase 1 (SPHK1) gene, an antisense RNA KHPS1 forms a triplex at the SPHK1 enhancer, then recruits transcription factors E2F1 and p300 to initiate synthesis of eRNA-Sphk1. The eRNA displaces CTCF from its binding site, restructuring the domain for expression of SPHK1 mRNA [179]. Both YRY and RYR triplexes can form with lncRNAs,

which are non-coding RNAs that direct tissue-specific developmental programmes [180]. Whether gene silencing or RNA expression ensues depends on the non-triplex forming sequence motifs present in the lncRNA. These motifs affect the assembly and phase separation of transcription factories [30,181–183].

The ALU adenosine-rich spacer between the left and right arms of dimers also form triplexes. The variable nature of this sequence allows specific docking of a third strand (figure 1a). Genome-wide screens for triplex forming sequences reveals an enrichment for the ALU spacer motif in DNA-DNA:RNA hybrids (figure 1) [8]. The triplexes are stabilized by the interaction of the histone H3 N-terminus tail with the 2'-hydroxyl group of the RNA strand [184]. ALU RNA:DNA:DNA triplexes also involve lncRNAs. An *in silico* analysis of the NIH epigenome roadmap data reveals that 34% of triplexes formed by lncRNAs in super enhancer regions overlap SINEs [3]. Tissue-specific enhancers in these regions guide developmental programmes [185,186].

ALU T-flipons allow RNAs with low complexity sequences to regulate the read-out of genes, a model similar to that envisioned by Britten & Davidson [17]. Novel evolutionary outcomes potentially arise when an ALU T-flipon spreads to a new chromosomal location, enabling regulation of a different set of genes. Changes in the tissue-specific expression of TSRs has the potential to produce additional phenotypic variability. The gain or loss of protein-binding motifs in TSRs are another possible way to vary outcomes. The many motif combinations possible allow the assembly of new cellular machines with novel functionalities [30,181–183]. Novel TSRs probably arise rapidly as the constraints in non-coding regions are fewer than in exons. They create a medley of protein ensembles, some of which enhance survival and others with no apparent benefit.

## 10. G-flipons

Four-stranded quadruplex structures (G4) held together by a quartet of guanines paired with both Watson–Crick and Hoogsteen hydrogen bonds are another class of flipons. G4 variants exist in which the phosphate backbones are parallel, anti-parallel or a mixture of both [51,187]. The quadruplex structures may be intermolecular (figure 1a) or intramolecular (figure 3c) with single-stranded regions connecting the quartets in various ways [188]. The *in vitro* stability depends on the presence of a metal ion at their centre. Certain metals favour parallel strand structures and while others show preference for anti-parallel folds [189]. Their functional relevance can be examined with controls unable to form G4Q. Some controls use modified residues like 7-deaza-guanosine that disrupt Hoogsteen bonding [190] and others introduce double or triple point mutations that prevent G4Q folding [187].

NSC alone is unlikely to promote quadruplex formation *in vivo*, given that other alternative NoBs form more easily [45,191]. However, transcription of linear templates *in vitro* can drive quadruplex formation in regions 5', but not 3' to a polymerase [192] and is due to duplex melting [193]. Enrichment for G4 forming sequences in humans is highest in the promoters and in the 5' regions of genes and occurs at approximately equal frequency on coding and template strands [194,195]. Formation of quadruplexes in these regions is probably very rapid as it only requires refolding of a single DNA strand (figure 3c).

G4Q formation is modified *in vivo* by guanine adducts, such as 8-oxo-dG produced by mitochondria-generated reactive oxygen species and by X-rays [122]. The modifications lower B-DNA melting temperature and facilitate G4Q formation in G-rich regions by opening up the duplex [196]. G4Q formation is favoured when there are more than four G-tracks in the neighbourhood, as the extra tracks can substitute for those damaged by oxidation [197]. About 46% of predicted G4Q forming sequences have a fifth track, including those in VEGF, c-MYC, KRAS, HIF-1 $\alpha$  RET, HSP90, AR, PDGFR- $\beta$ BCL-2 and HIF-1 $\alpha$  promoters [197]. Damaged residues are readily accommodated in G4Q loops (figure 3c) and seed the assembly of break-excision repair (BER) complexes composed of OGG1, NEIL1, NEIL2, NEIL3 and NTH1 [197–200].

The effects of 8-oxo-dG on G4Q has been extensively studied using plasmid substrates engineered *in vitro* and tested post-transfection *in vivo*. Lesions on the template strand resolve most rapidly and downregulate gene expression by a BER-independent process [198]. Transcription stops completely when excision of 8-oxo-dG on the template strand produces an apurinic site [201]. Lesions on the coding strand are repaired more slowly and upregulate the hypoxia-induced transcription of vascular epithelial growth factor (VEGF) RNA [201,202]. Lesions to the G4Q that are enriched on the coding strand of DNA repair gene promoters also enhance gene transcription [195]. They initiate protective responses to resolve DNA damage and to reset gene expression [203]. Both processes involve chromatin remodelling. As a class, G4Q forming regions *in vivo* have epigenetic



signatures characteristic of nucleosome-depletion and elevated transcription. They shape regulatory chromatin [204].

Are there G4Q binding proteins? Crystal structures of proteins bound to G4Q indicate that many proteins recognize this conformation indirectly. The proteins either bind bases in single-stranded loops or interact with the stem formed as B-DNA transitions to G4Q [205–207]. The proteins either lock the quadruplex strands in place or dismantle it. Many of proteins that dismantle G4Q are ATP-dependent helicases. They tug on single-stranded regions to tease the structure apart. The complementary sequence then captures the single strand to reform B-DNA, probably preventing the helicase from slipping backwards [205]. The ratchet mechanism allows helicases to slide nucleosomes along DNA, clearing the way for replication, repair, recombination and transcription complexes to assemble [205,207].

Modern genomes place G-rich sequences at promoters and use them to bootstrap development (as described above). Helicases ensure that formation of G4Q is dynamic and reversible. These enzymes enable G-flipons to switch from one conformation to another. When flipons freeze in the G4Q conformation, outcomes are often bad. ‘Loss of function’ helicase variants cause a wide range of monogenic disorders. The various phenotypes reflect the sequence and structural preferences of each enzyme, and the tissues that express them [208]. Their loss is compensated for by lower fidelity mechanisms that remove the G4Q blocking DNA transcription and replication [207]. These rescue pathways can also fail, leading to an increase in genome fragility. Over a longer time period, G-flipons prone to freezing are deleted from the genome [209].

Do ALUs have G-flipons? ALUs contain a single G4Q forming sequence capable of forming an intramolecular parallel-stranded quadruplex *in vitro*. The structure is made from two G-quartets stacked on each other [210]. The arrangement lacks the stability of canonical G4Q sequences made with three or more layers (figures 1a and 3a). *In vitro*, ALU wires can form by stacking on top of each other the two intramolecular G4Q made from different ALU elements [5]. Whether ALU wires form *in vivo* is uncertain, but such higher order structures *in principle* could promote nuclear organization or zip together homologous chromosomes [5]. They would be sensitive to DNA damage and destabilized by 8-oxo-dG. Alternatively, intermolecular ALU G4Q can form with quartets composed of guanines from different ALU strands [210]. There is currently no published data confirming a biological role for ALU encoded G-flipons.

It is more likely that G-flipons are spread through the genome by elements like the primate-specific SINE-VNTR-AU (SVA) rather than by ALU alone. SVA are a subset of retrotransposons where ALU is associated with a variable number of tandem repeats (VNTR). VNTR have a repeat motif length  $\geq 7$  bp (the short tandem repeat (STR) length is 1–6 bp) [211]. VNTRs/STRs exist that are either *ab initio* and unique to humans or have a repeat length expanded when compared to apes, a total of 1584 nonredundant events. Around 92% of these human-specific repeats are associated with transposable elements. SVAs contain about half of the *ab initio* insertions and 18% of the expansions. Intronic SVAs are enriched for motifs with long homopolymer guanine and cytosine stretches with a GC content greater than 45%. They affect gene expression. For example, SVA formed *ab initio* are frequent in genes whose expression is downregulated in inhibitory neuronal cells. SVA provide evidence that the evolutionary history and biology of G-flipons differs from that of other classes.

## 11. *In vivo* evidence for flipons

The evidence for formation of non-B-DNA structures *in vivo* is strong for Z-DNA where Mendelian genetics reveal that this flipon class regulates Type I interferon responses [84]. A separate role in the initiation of inflammatory responses and necrosis is supported by mouse genetics [86–88]. Other Mendelian diseases provide evidence of biological roles for triplex and G-flipon classes, especially that subset of disorders where helicase mutations affect triplex and quadruplex resolution. Those mutations freeze flipons in the NoBs conformation [208]. Instances of Mendelian disease may also be expected from mutations to lncRNAs, but detection is challenging when multiple weak and redundant triplex interactions underlie formation of high avidity complexes [212,213]. When mutation in one triplex forming motif is partially covered by another, effects may be quantitative and carry a low risk of disease [214]. The lncRNA HOTAIR, which has five separate triplex forming sites [215], illustrates this difficulty. Obtaining a phenotype for HOTAIR by reverse mouse genetics has been challenging and controversial with findings varying by strain background [216]. Similar difficulties exist for other lncRNA, even those highly conserved between species [217]. Another approach is to

correlate expression of lncRNAs with various diseases. In cancer, elevated levels of lncRNAs, like MALAT1, THOR, SAMMSON and SchLAP1 predict lower survival [218,219]. Whether those lncRNAs are causal, coincidental or an effect of cancer is currently unresolved. Other studies show that knockout of MALAT1 reduces metastases in some tumour backgrounds [220]. In such cases, MALAT1 might compensate for 'loss of function' cancer cell mutations that prevent adaptation to new niches or derepress 'gain of function' mutations that promote spread by reducing cell attachment. Collectively the findings imply that genetic interactions are important in T-flipon biology. Mapping of phenotypes require carefully designed genetic screens.

Other evidence for non-B-DNA structures comes from genome-wide sequencing methods that use permanganate and other chemicals to fingerprint the single strand regions characteristic of each conformation [45]. For example, permanganate hypersensitive regions exist at junctions between B- and Z-DNA, one at each end of the NoBs. A limitation of these protocols is often the exclusion of repeat elements, like ALU from analysis [45], so cataloguing of all NoBs by this method is incomplete. Psoralen cross-linking of DNA reveals that many actively transcribed chromatin domains *in vivo* are both underwound and enriched for ALUs [105]. G4Q-specific reagents such as BRACO-19, pyridostatin and Phen-DC3 allow capture of quadruplexes inside cells [195]. Cross-linkable, chemically conjugated oligonucleotides demonstrate that triplexes form *in vivo* and are addressable to particular genomic targets [221]. Many other high throughput approaches, underlying much of the biology described above, have helped define flipon effects on RNA editing, epigenetic modification and genetic susceptibility. Each represents only a snapshot of a very complex picture that has multiple themes and a confounding cast of characters.

New tools under development will increase our understanding of these processes by providing time-resolved identification and imaging of non-B-DNA conformations inside intact cells. These include reagents for click chemistry at sites of transcriptionally active DNA [58], supra-resolution microscopy and fast field-resolved infrared spectroscopy [222]. These approaches enable study of proteins that modulate flipon conformation in real time and the effects of small molecule perturbagens on the processes these proteins regulate. Drugs for topoisomerases, polymerases, helicases and epigenetic modulators are already available. The low cost of whole genome sequencing will enable further Mendelian studies in disease families to map variants in hypothesized NoBs to phenotypes and help elucidate their biological functions. It is already possible to identify genotypes of interest in variant databases, but there is no phenotypic annotation available or mechanism to trace the individuals or families whose DNA was sequenced. One example concerns the  $Z\alpha$  P193A homozygotes in humans, who are clearly viable but for whom no phenotypes are published. Reverse genetics in animal models is now easier with the DNA editing tools now available. In the future they will provide useful insights into the biologically relevant mechanisms and the genetic interactions underlying flipon phenotypes.

## 12. ALU flipons and evolution

ALU NoBs are an example of an instructive genetic code that specifies how messages are compiled. By comparison, codons represent a semantic genetic code that maps nucleotide triplets to amino acids. In the simplest form, an instructive code is binary, with the composition of a message contingent on context. Sequences that dynamically change conformation to alter genomic read-out are one way to genetically implement instructive codes. The name flipon captures their switch-like nature.

Flipons act locally. They alter the information extracted from their neighbourhood. The number of transcripts possible expands combinatorically as the flipon count in a gene increases, enhancing variability in transcript level, alternative splicing, base modification and three-dimensional RNA structure. Stochastic differences in flipons conformation within a population of cells is expected to produce a gradient of phenotypes within a tissue. Selection will optimize for flipon sets that work well enough most of the time, under most conditions, for sufficient cells, tissues or organisms to survive. Exploration and engagement of some flipon sets may not occur until something happens that compels them to compile something different. Until then, many of their transcripts will be triaged along with others that fail quality controls. Not all flipons will be useful, reflecting the random nature of their insertion. However, their exaptation may lead to new elaborations and novel protein assemblies. The medley of messages they compile will coexist with the old. There is no need to completely abandon previously successful adaptations, nor to risk irreversibly rewriting the genome by randomly mutating codons. There is a metabolic cost to the flipon strategy. It takes work to switch flipon conformations. Despite all this expenditure of energy, most of the RNA transcribed will never

be used and will require more work to dispose of it, the only gain being entropy [223]. The trade-off is a larger transcript space to explore and exploit. Quite complex systems that solve for survival can then evolve. Eventually these systems will discover ways to learn. Optimization of flipon settings is then easier than with other forms of evolution.

ALU transposition enables the spread flipon adaptations throughout the genome. Other sequences can just hitch a ride. The mash of flipons, codons, VNTRs, STRs and other motifs rapidly generates variability for natural selection to act on. This mechanism remains active in humans. Genome Wide Association Studies revealing a disproportionately high rate of recent ALU and SVA insertions at many quantitative trait loci [211,224]. ALU transposition also influences speciation. There is a 15-fold difference in ALU activity between humans and great apes [225], with evidence of positive selection in *Homo sapiens* [226,227]. The insertion of new ALUs after the human–chimpanzees split is higher in genes coding for nervous system development and function. ALU RNA editing also occurs more frequently in the human brain [228] with evidence for the positive selection of the ADAR p150 N-terminus that contains the Z $\alpha$  domain [229]. Along with other proposed mechanisms [230], ALUs prime the genome for rapid evolution.

### 13. Summary

ALUs have a direct impact on evolution, through the DNA damage they cause and the chromosomal rearrangements they provoke [1,231–236]. ALUs also impact evolution by the non-B-DNA structures they encode. The sequences capable of adopting these alternative conformations are called flipons.

Flipons are programmable binary switches whose conformation can be set independently of each other. They act in combination to greatly increase the coding capacity of linearly encoded genomes. Flipons act in *cis* through locally produced Z-DNA and in *trans* through triplexes formed from RNAs transcribed elsewhere. Along with G4Q that sense DNA damage, they alter the transcripts compiled. They act by changing chromatin state and phasing nucleosomes. The complex machines they assemble alter cell-specific outcomes. The cellular memories they create refashion future responses [151,237]. The Mendelian diseases they cause are instances when these processes either freeze or fail.

By their very nature, invasive ALU elements cannot help what they do. They spread by inserting themselves into active genes throughout the genome. The non-B-DNA structures they form provide an instructive code for compiling novel transcripts from wherever they land. By carelessly generating phenotypic diversity, ALUs unintentionally accelerate the evolution of many things new and unexpected.

**Ethics.** No human participants were involved in this study.

**Data accessibility.** All the data reviewed are published or available from public sources. No proprietary algorithms or material were used for this review.

**Competing interests.** There are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome. The author is the founder the company InsideOutBio that is committed to open science and working across disciplines. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest and conducted using publicly available data sources.

**Funding.** I received no funding for this study.

**Acknowledgements.** The author acknowledges the work of many scientists who prepared the data referenced in this review. I would like to thank Dr Levens and his laboratory for making available bed files with locations determined by permanganate footprinting of NoBs in the genome ([https://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/software/nonbdna/nonB\\_DNA\\_predicted.tar](https://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/software/nonbdna/nonB_DNA_predicted.tar)) and Drs He, Wu and Lyu for providing bed files of their kethoxal data.

### References

1. Batzer MA, Deininger PL. 2002 Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3**, 370–379. (doi:10.1038/nrg798)
2. Deininger P. 2011 Alu elements: know the SINES. *Genome Biol.* **12**, 236. (doi:10.1186/gb-2011-12-12-236)
3. Soibam B. 2017 Super-lncRNAs: identification of lncRNAs that target super-enhancers via RNA:DNA:DNA triplex formation. *RNA* **23**, 1729–1742. (doi:10.1261/rna.061317.117)
4. Herbert A. 2019 Z-DNA and Z-RNA in human disease. *Commun. Biol.* **2**, 7. (doi:10.1038/s42003-018-0237-x)
5. Varizhuk AM *et al.* 2018 Polymorphism of G4 associates: from stacks to wires via interlocks. *Nucleic Acids Res.* **46**, 8978–8992. (doi:10.1093/nar/gky729)
6. Wang S, Long Y, Wang J, Ge Y, Guo P, Liu Y, Tian T, Zhou X. 2014 Systematic investigations of different cytosine modifications on CpG dinucleotide sequences: the effects on the B-Z transition. *J. Am. Chem. Soc.* **136**, 56–59. (doi:10.1021/ja4107012)

7. Wang J, Wang S, Zhong C, Tian T, Zhou X. 2015 Novel insights into a major DNA oxidative lesion: its effects on Z-DNA stabilization. *Org. Biomol. Chem.* **13**, 8996–8999. (doi:10.1039/c5ob01340b)
8. Sentürk Cetin N, Kuo C-C, Ribarska T, Li R, Costa IG, Grummt I. 2019 Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. *Nucleic Acids Res.* **47**, 2306–2321. (doi:10.1093/nar/gky1305)
9. Salih F, Salih B, Kogan S, Trifonov EN. 2008 Epigenetic nucleosomes: Alu sequences and CG as nucleosome positioning element. *J. Biomol. Struct. Dyn.* **26**, 9–16. (doi:10.1080/07391102.2008.10507219)
10. Englander EW, Howard BH. 1995 Nucleosome positioning by human Alu elements in chromatin. *J. Biol. Chem.* **270**, 10 091–10 096. (doi:10.1074/jbc.270.17.10091)
11. Tanaka Y, Yamashita R, Suzuki Y, Nakai K. 2010 Effects of Alu elements on global nucleosome positioning in the human genome. *BMC Genomics* **11**, 309. (doi:10.1186/1471-2164-11-309)
12. Struhl K, Segal E. 2013 Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **20**, 267–273. (doi:10.1038/nsmb.2506)
13. Chuong EB, Elde NC, Feschotte C. 2017 Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86. (doi:10.1038/nrg.2016.139)
14. Zhang XO, Gingeras TR, Weng Z. 2019 Genome-wide analysis of polymerase III-transcribed Alu elements suggests cell-type-specific enhancer function. *Genome Res.* **29**, 1402–1414. (doi:10.1101/gr.249789.119)
15. Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. 2007 Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol.* **8**, R127. (doi:10.1186/gb-2007-8-6-r127)
16. Shrestha S, Sewell JA, Santos CS, Forchielli E, Pro SC, Martinez M, Bass JIF. 2019 Discovering human transcription factor physical interactions with genetic variants, novel DNA motifs, and repetitive elements using enhanced yeast one-hybrid assays. *Genome Res.* **29**, 1533–1544. (doi:10.1101/gr.248823.119)
17. Britten RJ, Davidson EH. 1969 Gene regulation for higher cells: a theory. *Science* **165**, 349–357. (doi:10.1126/science.165.3891.349)
18. Ahmad S, Mu X, Yang F, Greenwald E, Park JW, Jacob E, Zhang CZ, Hur S. 2018 Breaching self-tolerance to Alu duplex RNA underlies MDAS-mediated inflammation. *Cell* **172**, 797–810 e713. (doi:10.1016/j.cell.2017.12.016)
19. Athanasiadis A, Rich A, Maas S. 2004 Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391. (doi:10.1371/journal.pbio.0020391)
20. Kim DD, Kim TT, Walsh T, Kobayashi Y, Matisse TC, Buyske S, Gabriel A. 2004 Widespread RNA editing of embedded Alu elements in the human transcriptome. *Genome Res.* **14**, 1719–1725. (doi:10.1101/gr.2855504)
21. Levanon EY *et al.* 2004 Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.* **22**, 1001–1005. (doi:10.1038/nb996)
22. Pfaller CK, Donohue RC, Nersisyan S, Brodsky L, Cattaneo R. 2018 Extensive editing of cellular and viral double-stranded RNA structures accounts for innate immunity suppression and the proviral activity of ADAR1p150. *PLoS Biol.* **16**, e2006577. (doi:10.1371/journal.pbio.2006577)
23. Gannon HS *et al.* 2018 Identification of ADAR1 adenosine deaminase dependency in a subset of cancer cells. *Nat. Commun.* **9**, 5450. (doi:10.1038/s41467-018-07824-4)
24. Ishizuka JJ *et al.* 2019 Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature* **565**, 43–48. (doi:10.1038/s41586-018-0768-9)
25. Liu H *et al.* 2019 Tumor-derived IFN triggers chronic pathway agonism and sensitivity to ADAR loss. *Nat. Med.* **25**, 95–102. (doi:10.1038/s41591-018-0302-5)
26. Herbert A. 2019 ADAR and immune silencing in cancer. *Trends Cancer* **5**, 272–282. (doi:10.1016/j.trecan.2019.03.004)
27. Crepaldi L, Policarpi C, Coatti A, Sherlock WT, Jongbloets BC, Down TA, Riccio A. 2013 Binding of TFIIC to SINE elements controls the relocation of activity-dependent neuronal genes to transcription factories. *PLoS Genet.* **9**, e1003699. (doi:10.1371/journal.pgen.1003699)
28. Ferrari R *et al.* 2020 TFIIC binding to Alu elements controls gene expression via chromatin looping and histone acetylation. *Mol. Cell* **77**, 475–487 e411. (doi:10.1016/j.molcel.2019.10.020)
29. Policarpi C, Crepaldi L, Brookes E, Nitarska J, French SM, Coatti A, Riccio A. 2017 Enhancer SINES link Pol III to Pol II transcription in neurons. *Cell Rep.* **21**, 2879–2894. (doi:10.1016/j.celrep.2017.11.019)
30. Palikyras S, Papanonis A. 2019 Modes of phase separation affecting chromatin regulation. *Open Biol.* **9**, 190167. (doi:10.1098/rsob.190167)
31. Kassube SA, Fang J, Grob P, Yakovchuk P, Goodrich JA, Nogales E. 2013 Structural insights into transcriptional repression by noncoding RNAs that bind to human Pol II. *J. Mol. Biol.* **425**, 3639–3648. (doi:10.1016/j.jmb.2012.08.024)
32. Zovoilis A, Cifuentes-Rojas C, Chu HP, Hernandez AJ, Lee JT. 2016 Destabilization of B2 RNA by EZH2 activates the stress response. *Cell* **167**, 1788–1802 e1713. (doi:10.1016/j.cell.2016.11.041)
33. Hernandez AJ, Zovoilis A, Cifuentes-Rojas C, Han L, Bujisic B, Lee JT. 2020 B2 and ALU retrotransposons are self-cleaving ribozymes whose activity is enhanced by EZH2. *Proc. Natl Acad. Sci. USA* **117**, 415–425. (doi:10.1073/pnas.1917190117)
34. Berger A, Ivanova E, Gareau C, Scherrer A, Mazroui R, Strub K. 2014 Direct binding of the Alu binding protein dimer SRP9/14 to 40S ribosomal subunits promotes stress granule formation and is regulated by Alu RNA. *Nucleic Acids Res.* **42**, 11 203–11 217. (doi:10.1093/nar/gku822)
35. Ivanova E, Berger A, Scherrer A, Alkalaeva E, Strub K. 2015 Alu RNA regulates the cellular pool of active ribosomes by targeted delivery of SRP9/14 to 40S subunits. *Nucleic Acids Res.* **43**, 2874–2887. (doi:10.1093/nar/gkv048)
36. Samuel CE. 2019 Adenosine deaminase acting on RNA (ADAR1), a suppressor of double-stranded RNA-triggered innate immune responses. *J. Biol. Chem.* **294**, 1710–1720. (doi:10.1074/jbc.TM118.004166)
37. Chen LL, Yang L. 2017 ALU alternative regulation for gene expression. *Trends Cell Biol.* **27**, 480–490. (doi:10.1016/j.tcb.2017.01.002)
38. Panda AC. 2018 Circular RNAs act as miRNA sponges. *Adv. Exp. Med. Biol.* **1087**, 67–79. (doi:10.1007/978-981-13-1426-1\_6)
39. Caudron-Herger M, Pankert T, Seiler J, Nemeth A, Voit R, Grummt I, Rippe K. 2015 Alu element-containing RNAs maintain nucleolar structure and function. *EMBO J.* **34**, 2758–2774. (doi:10.15252/embj.201591458)
40. Caudron-Herger M, Pankert T, Rippe K. 2016 Regulation of nucleolus assembly by non-coding RNA polymerase II transcripts. *Nucleus* **7**, 308–318. (doi:10.1080/19491034.2016.1190890)
41. Ketele A, Kiss T, Jady BE. 2016 Human intron-encoded AluACA RNAs and telomerase RNA share a common element promoting RNA accumulation. *RNA Biol.* **13**, 1274–1285. (doi:10.1080/15476286.2016.1239689)
42. Elbarbary RA, Maquat LE. 2017 Distinct mechanisms obviate the potentially toxic effects of inverted-repeat Alu elements on cellular RNA metabolism. *Nat. Struct. Mol. Biol.* **24**, 496–498. (doi:10.1038/nsmb.3416)
43. Fitzpatrick T, Huang S. 2012 3'-UTR-located inverted Alu repeats facilitate mRNA translational repression and stress granule accumulation. *Nucleus* **3**, 359–369. (doi:10.4161/nud.20827)
44. Herbert A. 2019 A genetic instruction code based on DNA conformation. *Trends Genet.* **35**, 887–890. (doi:10.1016/j.tig.2019.09.007)
45. Kouzine F *et al.* 2017 Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **4**, 344–356 e347. (doi:10.1016/j.cels.2017.01.013)
46. Prigogine I. 1978 Time, structure, and fluctuations. *Science* **201**, 777–785. (doi:10.1126/science.201.4358.777)
47. Marx JL. 1985 Z-DNA: still searching for a function. *Science* **230**, 794–796. (doi:10.1126/science.2997919)
48. Morange M. 2007 What history tells us IX. Z-DNA: when nature is not opportunistic. *J. Biosci.* **32**, 657–661. (doi:10.1007/s12038-007-0065-5)
49. Herbert A, Rich A. 1996 The biology of left-handed Z-DNA. *J. Biol. Chem.* **271**, 11 595–11 598. (doi:10.1074/jbc.271.20.11595)
50. Safina A *et al.* 2017 FACT is a sensor of DNA torsional stress in eukaryotic cells. *Nucleic Acids Res.* **45**, 1925–1945. (doi:10.1093/nar/gkw1366)
51. Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. In press. The regulation and functions of DNA and RNA G-quadruplexes. *Nat.*



- Rev. Mol. Cell Biol.* (doi:10.1038/s41580-020-0236-x)
52. Hill RJ, Stollar BD. 1983 Dependence of Z-DNA antibody binding to polytene chromosomes on acid fixation and DNA torsional strain. *Nature* **305**, 338–340. (doi:10.1038/305338a0)
  53. Lane AN, Chaires JB, Gray RD, Trent JO. 2008 Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36**, 5482–5515. (doi:10.1093/nar/gkn517)
  54. Mili S, Steitz JA. 2004 Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694. (doi:10.1261/ma.7151404)
  55. Amemiya HM, Kundaje A, Boyle AP. 2019 The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354. (doi:10.1038/s41598-019-45839-z)
  56. Li H, Xiao J, Li J, Lu L, Feng S, Droge P. 2009 Human genomic Z-DNA segments probed by the Z alpha domain of ADAR1. *Nucleic Acids Res.* **37**, 2737–2746. (doi:10.1093/nar/gkp124)
  57. Carey MF, Peterson CL, Smale ST. 2010 Potassium permanganate probing of Pol II open complexes. *Cold Spring Harb. Protoc.* **2010**, pdb prot5479. (doi:10.1101/pdb.prot5479)
  58. Wu T, Lyu R, You Q, He C. 2020 Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. *Nat. Methods* **17**, 515–523. (doi:10.1038/s41592-020-0797-9)
  59. Takai D, Jones PA. 2002 Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA* **99**, 3740–3745. (doi:10.1073/pnas.052410099)
  60. Paz-Yaacov N, Bazak L, Buchumenski I, Porath HT, Danan-Gotthold M, Knisbacher BA, Eisenberg E, Levanon EY. 2015 Elevated RNA editing activity is a major contributor to transcriptomic diversity in tumors. *Cell Rep.* **13**, 267–276. (doi:10.1016/j.celrep.2015.08.080)
  61. Azorin F, Rich A. 1985 Isolation of Z-DNA binding proteins from SV40 minichromosomes: evidence for binding to the viral control region. *Cell* **41**, 365–374. (doi:10.1016/s0092-8674(85)80009-x)
  62. Lafer EM, Sousa RJ, Rich A. 1988 Z-DNA-binding proteins in *Escherichia coli* purification, generation of monoclonal antibodies and gene isolation. *J. Mol. Biol.* **203**, 511–516. (doi:10.1016/0022-2836(88)90017-4)
  63. Zhang S, Lockshin C, Herbert A, Winter E, Rich A. 1992 ZuoTn, a putative Z-DNA binding protein in *Saccharomyces cerevisiae*. *EMBO J.* **11**, 3787–3796.
  64. Zhang Y *et al.* 2014 Structural basis for interaction of a cotranslational chaperone with the eukaryotic ribosome. *Nat. Struct. Mol. Biol.* **21**, 1042–1046. (doi:10.1038/nsmb.2908)
  65. Krishna P, Kennedy BP, Waisman DM, van de Sande JH, McGhee JD. 1990 Are many Z-DNA binding proteins actually phospholipid-binding proteins? *Proc. Natl Acad. Sci. USA* **87**, 1292–1295. (doi:10.1073/pnas.87.4.1292)
  66. Patterson JB, Thomas DC, Hans SL, Samuel CE. 1995 Mechanism of interferon action: double-stranded RNA-specific adenosine deaminase from human cells is inducible by alpha and gamma interferons. *Virology* **210**, 508–511. (doi:10.1006/viro.1995.1370)
  67. Fritz J, Strehblow A, Taschner A, Schoppoff S, Pasierbek P, Jantsch MF. 2009 RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1. *Mol. Cell. Biol.* **29**, 1487–1497. (doi:10.1128/MCB.01519-08)
  68. Poulsen H, Nilsson J, Damgaard CK, Egebjerg J, Kjems J. 2001 CRM1 mediates the export of ADAR1 through a nuclear export signal within the Z-DNA binding domain. *Mol. Cell. Biol.* **21**, 7862–7871. (doi:10.1128/MCB.21.22.7862-7871.2001)
  69. Strehblow A, Hallegger M, Jantsch MF. 2002 Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain. *Mol. Cell. Biol.* **13**, 3822–3835. (doi:10.1091/mbc.e02-03-0161)
  70. Marshall PR *et al.* In press. Dynamic regulation of Z-DNA in the mouse prefrontal cortex by the RNA-editing enzyme Adar1 is required for fear extinction. *Nat. Neurosci.* (doi:10.1038/s41593-020-0627-5)
  71. Nakahama T, Kato Y, Kim JI, Vongpipatana T, Suzuki Y, Walkley CR, Kawahara Y. 2018 ADAR1-mediated RNA editing is required for thymic self-tolerance and inhibition of autoimmunity. *EMBO Rep.* **19**, e46303. (doi:10.15252/embr.201846303)
  72. Costa Cruz PH, Kato Y, Nakahama T, Shibuya T, Kawahara Y. 2020 A comparative analysis of ADAR mutant mice reveals site-specific regulation of RNA editing. *RNA* **26**, 454–469. (doi:10.1261/ma.072728.119)
  73. Sakurai M *et al.* 2017 ADAR1 controls apoptosis of stressed cells by inhibiting Staufen1-mediated mRNA decay. *Nat. Struct. Mol. Biol.* **24**, 534–543. (doi:10.1038/nsmb.3403)
  74. Lange SS, Vasquez KM. 2009 HMGB1: the jack-of-all-trades protein is a master DNA repair mechanic. *Mol. Cell. Biol.* **29**, 571–580. (doi:10.1002/mc.20544)
  75. Reeves R. 2015 High mobility group (HMG) proteins: modulators of chromatin structure and DNA repair in mammalian cells. *DNA Repair (Amst.)* **36**, 122–136. (doi:10.1016/j.dnarep.2015.09.015)
  76. Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H, de Lange T. 1999 Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514. (doi:10.1016/s0092-8674(00)80760-6)
  77. Doksani Y, Wu JY, de Lange T, Zhuang X. 2013 Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* **155**, 345–356. (doi:10.1016/j.cell.2013.09.048)
  78. Herbert AG, Rich A. 1993 A method to identify and characterize Z-DNA binding proteins using a linear oligodeoxynucleotide. *Nucleic Acids Res.* **21**, 2669–2672. (doi:10.1093/nar/21.11.2669)
  79. Herbert A, Alfken J, Kim YG, Mian IS, Nishikura K, Rich A. 1997 A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc. Natl Acad. Sci. USA* **94**, 8421–8426.
  80. Herbert A, Lowenhaupt K, Spitzner J, Rich A. 1995 Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc. Natl Acad. Sci. USA* **92**, 7550–7554. (doi:10.1073/pnas.92.16.7550)
  81. Herbert A, Schade M, Lowenhaupt K, Alfken J, Schwartz T, Shlyakhtenko LS, Lyubchenko YL, Rich A. 1998 The Z $\alpha$  domain from human ADAR1 binds to the Z-DNA conformer of many different sequences. *Nucleic Acids Res.* **26**, 3486–3493. (doi:10.1093/nar/26.15.3486)
  82. Schade M, Turner CJ, Kuhne R, Schmieder P, Lowenhaupt K, Herbert A, Rich A, Oschkinat H. 1999 The solution structure of the Z $\alpha$  domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA. *Proc. Natl Acad. Sci. USA* **96**, 12 465–12 470.
  83. Schwartz T, Rould MA, Lowenhaupt K, Herbert A, Rich A. 1999 Crystal structure of the Z $\alpha$  domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284**, 1841–1845.
  84. Herbert A. 2019 Mendelian disease caused by variants affecting recognition of Z-DNA and Z-RNA by the Z $\alpha$  domain of the double-stranded RNA editing enzyme ADAR. *Eur. J. Hum. Genet.* **28**, 114–117. (doi:10.1038/s41431-019-0458-6)
  85. Maelfait J, Liverpool L, Bridgeman A, Ragan KB, Upton JW, Rehwinkel J. 2017 Sensing of viral and endogenous RNA by ZBP1/DAI induces necroptosis. *EMBO J.* **36**, 2529–2543. (doi:10.15252/emj.201796476)
  86. Jiao H *et al.* 2020 Z-nucleic-acid sensing triggers ZBP1-dependent necroptosis and inflammation. *Nature* **580**, 391–395. (doi:10.1038/s41586-020-2129-8)
  87. Wang R *et al.* 2020 Gut stem cell necroptosis by genome instability triggers bowel inflammation. *Nature* **580**, 386–390. (doi:10.1038/s41586-020-2127-x)
  88. Zhang T *et al.* 2020 Influenza virus Z-RNAs induce ZBP1-mediated necroptosis. *Cell* **180**, 1115–1129 e1113. (doi:10.1016/j.cell.2020.02.050)
  89. Devos M *et al.* 2020 Sensing of endogenous nucleic acids by ZBP1 induces keratinocyte necroptosis and skin inflammation. *J. Exp. Med.* **217**, e20191913. (doi:10.1084/jem.20191913)
  90. Schwartz T, Behlke J, Lowenhaupt K, Heinemann U, Rich A. 2001 Structure of the DL1M-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.* **8**, 761–765. (doi:10.1038/nsb0901-761)
  91. Ha SC, Kim D, Hwang HY, Rich A, Kim YG, Kim KK. 2008 The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA. *Proc. Natl Acad. Sci. USA* **105**, 20 671–20 676. (doi:10.1073/pnas.0810463106)
  92. Kesavardhana S, Malireddi RKS, Burton AR, Porter SN, Vogel P, Pruetz-Miller SM, Kannekanti TD. 2020 The Z $\alpha$ 2 domain of ZBP1 is a molecular switch regulating influenza-induced PANoptosis and perinatal lethality

- during development. *J. Biol. Chem.* (doi:10.1074/jbc.RA120.013752)
93. Koehler H, Cotsmire S, Langland J, Kibler KV, Kalman D, Upton JW, Mocarski ES, Jacobs BL. 2017 Inhibition of DAI-dependent necroptosis by the Z-DNA binding domain of the vaccinia virus innate immune evasion protein, E3. *Proc. Natl Acad. Sci. USA* **114**, 11 506–11 511. (doi:10.1073/pnas.1700999114)
  94. Champ PC, Maurice S, Vargason JM, Camp T, Ho PS. 2004 Distributions of Z-DNA and nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation. *Nucleic Acids Res.* **32**, 6501–6510. (doi:10.1093/nar/gkh988)
  95. Wittig B, Wolff S, Dorbic T, Vahrson W, 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.
  96. Liu R, Liu H, Chen X, Kirby M, Brown PO, Zhao K. 2001 Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* **106**, 309–318.
  97. Schade M, Turner CJ, Lowenhaupt K, Rich A, Herbert A. 1999 Structure–function analysis of the Z-DNA-binding domain Z $\alpha$  of dsRNA adenosine deaminase type I reveals similarity to the ( $\alpha$ + $\beta$ ) family of helix–turn–helix proteins. *EMBO J.* **18**, 470–479. (doi:10.1093/emboj/18.2.470)
  98. Mannion NM *et al.* 2014 The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep.* **9**, 1482–1494. (doi:10.1016/j.celrep.2014.10.041)
  99. Beyer U *et al.* 2017 Rare ADAR and RNASEH2B variants and a type I interferon signature in glioma and prostate carcinoma risk and tumorigenesis. *Acta Neuropathol.* **134**, 905–922. (doi:10.1007/s00401-017-1774-y)
  100. Shin SJ, Ham S, Park J, Seo SH, Lim CH, Jeon H, Huh J, Roh TY. 2016 Z-DNA-forming sites identified by ChIP-Seq are associated with actively transcribed regions in the human genome. *DNA Res.* **23**, 477–486. (doi:10.1093/dnares/dsw031)
  101. Zhai Q, Wang L, Zhao P, Li T. 2017 Role of citrullination modification catalyzed by peptidylarginine deiminase 4 in gene transcriptional regulation. *Acta Biochim. Biophys. Sin. (Shanghai)* **49**, 567–572. (doi:10.1093/abbs/gmx042)
  102. Zhang T, Wallis M, Petrovic V, Challis J, Kalitsis P, Hudson DF. 2019 Loss of TOP3B leads to increased R-loop formation and genome instability. *Open Biol.* **9**, 190222. (doi:10.1098/rsob.190222)
  103. Fioriti L *et al.* 2015 The persistence of hippocampal-based memory requires protein synthesis mediated by the prion-like protein CPEB3. *Neuron* **86**, 1433–1448. (doi:10.1016/j.neuron.2015.05.021)
  104. Fernández X, Diaz-Ingelmo O, Martínez-García B, Roca J. 2014 Chromatin regulates DNA torsional energy via topoisomerase II-mediated relaxation of positive supercoils. *EMBO J.* **33**, 1492–1501. (doi:10.15252/emboj.201488091)
  105. Naughton C *et al.* 2013 Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* **20**, 387–395. (doi:10.1038/nsmb.2509)
  106. Teves SS, Henikoff S. 2013 Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.* **21**, 88–94. (doi:10.1038/nsmb.2723)
  107. Peck LJ, Wang JC. 1983 Energetics of B-to-Z transition in DNA. *Proc. Natl Acad. Sci. USA* **80**, 6206–6210.
  108. Baranello L *et al.* 2016 RNA polymerase II regulates topoisomerase 1 activity to favor efficient transcription. *Cell* **165**, 357–371. (doi:10.1016/j.cell.2016.02.036)
  109. Belotserkovskii BP, Mirkin SM, Hanawalt PC. 2013 DNA sequences that interfere with transcription: implications for genome function and stability. *Chem. Rev.* **113**, 8620–8637. (doi:10.1021/cr400078y)
  110. Thomas TJ, Gunnia UB, Thomas T. 1991 Polyamine-induced B-DNA to Z-DNA conformational transition of a plasmid DNA with (dG-dC)n insert. *J. Biol. Chem.* **266**, 6137–6141.
  111. Miller-Fleming L, Olin-Sandoval V, Campbell K, Ralsler M. 2015 Remaining mysteries of molecular biology: the role of polyamines in the cell. *J. Mol. Biol.* **427**, 3389–3406. (doi:10.1016/j.jmb.2015.06.020)
  112. Iacomino G, Picariello G, D'Agostino L. 2012 DNA and nuclear aggregates of polyamines. *Biochim. Biophys. Acta Mol. Cell Res.* **1823**, 1745–1755. (doi:10.1016/j.bbamcr.2012.05.033)
  113. Kuriakose T, Man SM, Malireddi RK, Karki R, Kesavardhana S, Place DE, Neale G, Vogel P, Kanneganti TD. 2016 ZBP1/DAI is an innate sensor of influenza virus triggering the NLRP3 inflammasome and programmed cell death pathways. *Sci. Immunol.* **1**, aag2045. (doi:10.1126/sciimmunol.aag2045)
  114. Lin J, Kumari S, Kim C, Van TM, Wachsmuth L, Polykratis A, Pasparakis M. 2016 RIPK1 counteracts ZBP1-mediated necroptosis to inhibit inflammation. *Nature* **540**, 124–128. (doi:10.1038/nature20558)
  115. Malireddi RKS, Kesavardhana S, Kanneganti TD. 2019 ZBP1 and TAK1: master regulators of NLRP3 inflammasome/pyroptosis, apoptosis, and necroptosis (PAN-optosis). *Front. Cell. Infect. Microbiol.* **9**, 406. (doi:10.3389/fcimb.2019.00406)
  116. Papayannopoulos V. 2018 Neutrophil extracellular traps in immunity and disease. *Nat. Rev. Immunol.* **18**, 134–147. (doi:10.1038/nri.2017.105)
  117. Sollberger G *et al.* 2018 Gasdermin D plays a vital role in the generation of neutrophil extracellular traps. *Sci. Immunol.* **3**, eaar6689. (doi:10.1126/sciimmunol.aar6689)
  118. Lafer EM, Valle RP, Moller A, Nordheim A, Schur PH, Rich A, Stollar BD. 1983 Z-DNA-specific antibodies in human systemic lupus erythematosus. *J. Clin. Invest.* **71**, 314–321. (doi:10.1172/jci110771)
  119. Thomas TJ, Meryhew NL, Messner RP. 1990 Enhanced binding of lupus sera to the polyamine-induced left-handed Z-DNA form of polynucleotides. *Arthritis Rheum.* **33**, 356–365. (doi:10.1002/art.1780330308)
  120. Yu Y, Su K. 2013 Neutrophil extracellular traps and systemic lupus erythematosus. *J. Clin. Cell. Immunol.* **4**, 139. (doi:10.4172/2155-9899.1000139)
  121. Hakkim A, Furnrohr BG, Amann K, Laube B, Abed UA, Brinkmann V, Herrmann M, Voll RE, Zychlinsky A. 2010 Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proc. Natl Acad. Sci. USA* **107**, 9813–9818. (doi:10.1073/pnas.0909927107)
  122. Poetsch AR, Boulton SJ, Luscombe NM. 2018 Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome Biol.* **19**, 215. (doi:10.1186/s13059-018-1582-2)
  123. Ming X, Matter B, Song M, Veliath E, Shanley R, Jones R, Tretyakova N. 2014 Mapping structurally defined guanine oxidation products along DNA duplexes: influence of local sequence context and endogenous cytosine methylation. *J. Am. Chem. Soc.* **136**, 4223–4235. (doi:10.1021/ja411636j)
  124. Cadet J, Wagner JR. 2013 DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harb. Perspect. Biol.* **5**, a012559. (doi:10.1101/cshperspect.a012559)
  125. Hofer A, Liu ZJ, Balasubramanian S. 2019 Detection, structure and function of modified DNA bases. *J. Am. Chem. Soc.* **141**, 6420–6429. (doi:10.1021/jacs.9b01915)
  126. Vongsutilers V, Shinohara Y, Kawai G. 2020 Epigenetic TET-catalyzed oxidative products of 5-methylcytosine impede Z-DNA formation of CG decamers. *ACS Omega* **5**, 8056–8064. (doi:10.1021/acsomega.0c00120)
  127. Xu C, Corces VG. 2018 Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science* **359**, 1166–1170. (doi:10.1126/science.aan5480)
  128. Li Y *et al.* 2020 The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472–476. (doi:10.1038/s41586-019-1910-z)
  129. Papin C *et al.* 2017 Combinatorial DNA methylation codes at repetitive elements. *Genome Res.* **27**, 934–946. (doi:10.1101/gr.213983.116)
  130. Slyvka A, Mierzejewska K, Bochtler M. 2017 Neilike 1 (NEIL1) excises 5-carboxylcytosine directly and stimulates TDG-mediated 5-formyl and 5-carboxylcytosine excision. *Sci. Rep.* **7**, 9001. (doi:10.1038/s41598-017-07458-4)
  131. Raiber E-A *et al.* 2018 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat. Chem.* **10**, 1258–1266. (doi:10.1038/s41557-018-0149-x)
  132. Zhang X *et al.* 2016 DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nat. Genet.* **48**, 1014–1023. (doi:10.1038/ng.3610)
  133. Marina RJ *et al.* 2016 TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. *EMBO J.* **35**, 335–355. (doi:10.15252/emboj.201593235)
  134. Mayer A, Landry HM, Churchman LS. 2017 Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr. Opin. Cell Biol.* **46**, 72–80. (doi:10.1016/j.cob.2017.03.002)
  135. Ditlevsen JV, Tomaletti S, Belotserkovskii BP, Teijeiro V, Wang G, Vasquez KM, Hanawalt PC. 2008 Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA

- polymerase. *Nucleic Acids Res.* **36**, 3163–3170. (doi:10.1093/nar/gkn136)
136. Ray BK, Dhar S, Shakya A, Ray A. 2011 Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression. *Proc. Natl Acad. Sci. USA* **108**, 103–108. (doi:10.1073/pnas.1008831108)
137. Zhang F, Huang Q, Yan J, Chen Z. 2016 Histone acetylation induced transformation of B-DNA to Z-DNA in cells probed through FT-IR spectroscopy. *Anal. Chem.* **88**, 4179–4182. (doi:10.1021/acs.analchem.6b00400)
138. Wang M *et al.* 2019 Identification of DNA motifs that regulate DNA methylation. *Nucleic Acids Res.* **47**, 6753–6768. (doi:10.1093/nar/gkz483)
139. Allis CD. 2018 Pursuing the secrets of histone proteins: an amazing journey with a remarkable supporting cast. *Cell* **175**, 18–21. (doi:10.1016/j.cell.2018.08.022)
140. Su M, Han D, Boyd-Kirkup J, Yu X, Han JDJ. 2014 Evolution of Alu elements toward enhancers. *Cell Rep.* **7**, 376–385. (doi:10.1016/j.celrep.2014.03.011)
141. Xu C, Liu K, Lei M, Yang A, Li Y, Hughes TR, Min J. 2018 DNA sequence recognition of human CXXC domains and their structural determinants. *Structure* **26**, 85–95 e83. (doi:10.1016/j.str.2017.11.022)
142. Hu L *et al.* 2015 Structural insight into substrate preference for TE1-mediated oxidation. *Nature* **527**, 118–122. (doi:10.1038/nature15713)
143. Wang D, Hashimoto H, Zhang X, Barwick BG, Lonial S, Boise LH, Vertino PM, Cheng X. 2017 MAX is an epigenetic sensor of 5-carboxylcytosine and is altered in multiple myeloma. *Nucleic Acids Res.* **45**, 2396–2407. (doi:10.1093/nar/gkw1184)
144. de Rosa M, de Sanctis D, Rosario AL, Archer M, Rich A, Athanasiadis A, Carrondo MA. 2010 Crystal structure of a junction between two Z-DNA helices. *Proc. Natl Acad. Sci. USA* **107**, 9088–9092. (doi:10.1073/pnas.1003182107)
145. Gordan R, Shen N, Dror I, Zhou T, Horton J, Rohs R, Bulyk ML. 2013 Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* **3**, 1093–1104. (doi:10.1016/j.celrep.2013.03.014)
146. Zemojtel T, Kielbasa SM, Arndt PF, Behrens S, Bourque G, Vingron M. 2011 CpG deamination creates transcription factor-binding sites with high efficiency. *Genome Biol. Evol.* **3**, 1304–1311. (doi:10.1093/gbe/evr107)
147. Brahma S, Henikoff S. 2020 Epigenome regulation by dynamic nucleosome unwrapping. *Trends Biochem. Sci.* **45**, 13–26. (doi:10.1016/j.tibs.2019.09.003)
148. Hota SK, Bruneau BG. 2016 ATP-dependent chromatin remodeling during mammalian development. *Development* **143**, 2882–2897. (doi:10.1242/dev.128892)
149. Vavouri T, Lehner B. 2011 Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS Genet.* **7**, e1002036. (doi:10.1371/journal.pgen.1002036)
150. Blossley R, Schiessel H. 2018 The latest twists in chromatin remodeling. *Biophys. J.* **114**, 2255–2261. (doi:10.1016/j.bpj.2017.12.008)
151. Kamada R *et al.* 2018 Interferon stimulation creates chromatin marks and establishes transcriptional memory. *Proc. Natl Acad. Sci. USA* **115**, E9162–E9171. (doi:10.1073/pnas.1720930115)
152. Herbert A, Rich A. 2001 The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. *Proc. Natl Acad. Sci. USA* **98**, 12 132–12 137. (doi:10.1073/pnas.211419898)
153. Li JB, Levanon EY, Yoon JK, Aach J, Xie B, Leproust E, Zhang K, Gao Y, Church GM. 2009 Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213. (doi:10.1126/science.1170995)
154. Higuchi M, Maas S, Single FN, Hartner J, Rozov A, Burnashev N, Feldmeyer D, Sprengel R, Seeburg PH. 2000 Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406**, 78–81. (doi:10.1038/35017558)
155. Jiang D, Zhang J. 2019 The preponderance of nonsynonymous A-to-I RNA editing in coleoids is nonadaptive. *Nat. Commun.* **10**, 5411. (doi:10.1038/s41467-019-13275-2)
156. Daniel C, Silberberg G, Behm M, Ohman M. 2014 Alu elements shape the primate transcriptome by cis-regulation of RNA editing. *Genome Biol.* **15**, R28. (doi:10.1186/gb-2014-15-2-r28)
157. George CX, Ramaswami G, Li JB, Samuel CE. 2016 Editing of cellular self-RNAs by adenosine deaminase ADAR1 suppresses innate immune stress responses. *J. Biol. Chem.* **291**, 6158–6168. (doi:10.1074/jbc.M115.709014)
158. Yeo J, Goodman RA, Schirle NT, David SS, Beal PA. 2010 RNA editing changes the lesion specificity for the DNA repair enzyme NEIL1. *Proc. Natl Acad. Sci. USA* **107**, 20 715–20 719. (doi:10.1073/pnas.1009231107)
159. Nishikura K. 2016 A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **17**, 83–96. (doi:10.1038/nrm.2015.4)
160. Rice GI *et al.* 2012 Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. *Nat. Genet.* **44**, 1243–1248. (doi:10.1038/ng.2414)
161. Aktas T, Avsar Ilik I, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, Manke T, Backofen R, Akhtar A. 2017 DHX9 suppresses RNA processing defects originating from the *Alu* invasion of the human genome. *Nature* **544**, 115–119. (doi:10.1038/nature21715)
162. Ota H, Sakurai M, Gupta R, Valente L, Wulff BE, Ariyoshi K, Iizasa H, Davuluri RV, Nishikura K. 2013 ADAR1 forms a complex with Dicer to promote microRNA processing and RNA-induced gene silencing. *Cell* **153**, 575–589. (doi:10.1016/j.cell.2013.03.024)
163. Qi L *et al.* 2017 An RNA editing/dsRNA binding-independent gene regulatory mechanism of ADARs and its clinical implication in cancer. *Nucleic Acids Res.* **45**, 10 436–10 451. (doi:10.1093/nar/gkx667)
164. Bahn JH, Ahn J, Lin X, Zhang Q, Lee JH, Civelek M, Xiao X. 2015 Genomic analysis of ADAR1 binding and its involvement in multiple RNA processing pathways. *Nat. Commun.* **6**, 6355. (doi:10.1038/ncomms7355)
165. Quinones-Valdez G *et al.* 2019 Regulation of RNA editing by RNA-binding proteins in human cells. *Commun. Biol.* **2**, 19. (doi:10.1038/s42003-018-0271-8)
166. Nemlich Y *et al.* 2013 MicroRNA-mediated loss of ADAR1 in metastatic melanoma promotes tumor growth. *J. Clin. Invest.* **123**, 2703–2718. (doi:10.1172/JCI62980)
167. Chen T *et al.* 2015 ADAR1 is required for differentiation and neural induction by regulating microRNA processing in a catalytically independent manner. *Cell Res.* **25**, 459–476. (doi:10.1038/cr.2015.24)
168. Liddicoat BJ, Piskol R, Chalk AM, Ramaswami G, Higuchi M, Hartner JC, Li JB, Seeburg PH, Walkley CR. 2015 RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **349**, 1115–1120. (doi:10.1126/science.aac7049)
169. Frank-Kamenetski MD, Mirkin SM. 1995 Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95. (doi:10.1146/annurev.bi.64.070195.000433)
170. Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA, Frank-Kamenetski MD. 1987 DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature* **330**, 495–497. (doi:10.1038/330495a0)
171. Szabo Q, Bantignies F, Cavalli G. 2019 Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668. (doi:10.1126/sciadv.aaw1668)
172. Zhou Z, Giles KE, Felsenfeld G. 2019 DNA-RNA triple helix formation can function as a cis-acting regulatory mechanism at the human  $\beta$ -globin locus. *Proc. Natl Acad. Sci. USA* **116**, 6130–6139. (doi:10.1073/pnas.1900107116)
173. Roberts RW, Crothers DM. 1992 Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **258**, 1463–1466. (doi:10.1126/science.1279808)
174. Idili A, Vallee-Belisle A, Ricci F. 2014 Programmable pH-triggered DNA nanoswitches. *J. Am. Chem. Soc.* **136**, 5836–5839. (doi:10.1021/ja500619w)
175. Shefer K, Brown Y, Gorkovoy V, Nussbaum T, Ulyanov NB, Tzfati Y. 2007 A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol. Cell. Biol.* **27**, 2130–2143. (doi:10.1128/MCB.01826-06)
176. Chandrasekaran AR, Rusling DA. 2018 Triplex-forming oligonucleotides: a third strand for DNA nanotechnology. *Nucleic Acids Res.* **46**, 1021–1037. (doi:10.1093/nar/gkx1230)
177. Mondal T *et al.* 2015 MEG3 long noncoding RNA regulates the TGF- $\beta$  pathway genes through formation of RNA–DNA triplex structures. *Nat. Commun.* **6**, 7743. (doi:10.1038/ncomms8743)
178. O’Leary VB *et al.* 2015 PARTICLE, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. *Cell Rep.* **11**, 474–485. (doi:10.1016/j.celrep.2015.03.043)
179. Blank-Giwonja A, Postepska-Igielska A, Grumt I. 2019 lncRNA KHPS1 activates a poised enhancer by triplex-dependent recruitment of

- epigenomic regulators. *Cell Rep.* **26**, 2904–2915 e2904. (doi:10.1016/j.celrep.2019.02.059)
180. Li Y, Syed J, Sugiyama H. 2016 RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem. Biol.* **23**, 1325–1333. (doi:10.1016/j.chembiol.2016.09.011)
181. Shin Y, Chang YC, Lee DSW, Berry J, Sanders DW, Ronceray P, Wingreen NS, Haataja M, Brangwynne CP. 2018 Liquid nuclear condensates mechanically sense and restructure the genome. *Cell* **175**, 1481–1491 e1413. (doi:10.1016/j.cell.2018.10.057)
182. Fay MM, Anderson PJ. 2018 The role of RNA in biological phase separations. *J. Mol. Biol.* **430**, 4685–4701. (doi:10.1016/j.jmb.2018.05.003)
183. Frank L, Rippe K. 2020 Repetitive RNAs as regulators of chromatin-associated subcompartment formation by phase separation. *J. Mol. Biol.* (doi:10.1016/j.jmb.2020.04.015)
184. Maldonado R, Schwartz U, Silberhorn E, Längst G. 2019 Nucleosomes stabilize ssRNA-dsDNA triple helices in human cells. *Mol. Cell* **73**, 1243–1254.e1246. (doi:10.1016/j.molcel.2019.01.007)
185. Adam RC *et al.* 2015 Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521**, 366–370. (doi:10.1038/nature14289)
186. Ryu J, Kim H, Yang D, Lee AJ, Jung I. 2019 A new class of constitutively active super-enhancers is associated with fast recovery of 3D chromatin loops. *BMC Bioinformatics* **20**, 127. (doi:10.1186/s12859-019-2646-3)
187. Lightfoot HL, Hagen T, Tatum NJ, Hall J. 2019 The diverse structural landscape of quadruplexes. *FEBS Lett.* **593**, 2083–2102. (doi:10.1002/1873-3468.13547)
188. Oganessian L, Karlseder J. 2009 Telomeric armor: the layers of end protection. *J. Cell Sci.* **122**, 4013–4025. (doi:10.1242/jcs.050567)
189. Bhattacharyya D, Mirihana Arachchilage G, Basu S. 2016 Metal cations in G-quadruplex folding and stability. *Front. Chem.* **4**, 38. (doi:10.3389/fchem.2016.00038)
190. Wanrooij PH, Uhler JP, Shi Y, Westerlund F, Falkenberg M, Gustafsson CM. 2012 A hybrid G-quadruplex structure formed between RNA and DNA explains the extraordinary stability of the mitochondrial R-loop. *Nucleic Acids Res.* **40**, 10 334–10 344. (doi:10.1093/nar/gks802)
191. Sekibo DAT, Fox KR. 2017 The effects of DNA supercoiling on G-quadruplex formation. *Nucleic Acids Res.* **45**, 12 069–12 079. (doi:10.1093/nar/gkx856)
192. Zhang C, Liu HH, Zheng KW, Hao YH, Tan Z. 2013 DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix. *Nucleic Acids Res.* **41**, 7144–7152. (doi:10.1093/nar/gkt443)
193. Selvam S, Koirala D, Yu Z, Mao H. 2014 Quantification of topological coupling between DNA superhelicity and G-quadruplex formation. *J. Am. Chem. Soc.* **136**, 13 967–13 970. (doi:10.1021/ja5064394)
194. Huppert JL, Bugaut A, Kumari S, Balasubramanian S. 2008 G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res.* **36**, 6260–6268. (doi:10.1093/nar/gkn511)
195. Fleming AM, Zhu J, Ding Y, Visser JA, Zhu J, Burrows CJ. 2018 Human DNA repair genes possess potential G-quadruplex sequences in their promoters and 5'-untranslated regions. *Biochemistry* **57**, 991–1002. (doi:10.1021/acs.biochem.7b01172)
196. Yennie CJ, Delaney S. 2012 Thermodynamic consequences of the hyperoxidized guanine lesion guanidinohydantoin in duplex DNA. *Chem. Res. Toxicol.* **25**, 1732–1739. (doi:10.1021/tx300190a)
197. Fleming AM, Zhou J, Wallace SS, Burrows CJ. 2015 A role for the fifth G-track in G-quadruplex forming oncogene promoter sequences during oxidative stress: do these 'spare tires' have an evolved function? *ACS Cent. Sci.* **1**, 226–233. (doi:10.1021/acscentsci.5b00202)
198. Fleming AM, Zhu J, Ding Y, Burrows CJ. 2017 8-Oxo-7,8-dihydroguanine in the context of a gene promoter G-quadruplex is an on-off switch for transcription. *ACS Chem. Biol.* **12**, 2417–2426. (doi:10.1021/acscmbio.7b00636)
199. Fleming AM, Burrows CJ. 2017 8-Oxo-7,8-dihydroguanine, friend and foe: epigenetic-like regulator versus initiator of mutagenesis. *DNA Repair (Amst.)* **56**, 75–83. (doi:10.1016/j.dnarep.2017.06.009)
200. Takahashi S, Kim KT, Podbevsek P, Plavec J, Kim BH, Sugimoto N. 2018 Recovery of the formation and function of oxidized G-quadruplexes by a pyrene-modified guanine tract. *J. Am. Chem. Soc.* **140**, 5774–5783. (doi:10.1021/jacs.8b01577)
201. Fleming AM, Ding Y, Burrows CJ. 2017 Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl Acad. Sci. USA* **114**, 2604–2609. (doi:10.1073/pnas.1619809114)
202. Pastukh V, Roberts JT, Clark DW, Bardwell GC, Patel M, Al-Mehdi AB, Borchert GM, Gillespie MN. 2015 An oxidative DNA 'damage' and repair mechanism localized in the VEGF promoter is important for hypoxia-induced VEGF mRNA expression. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **309**, L1367–L1375. (doi:10.1152/ajplung.00236.2015)
203. Hauer MH, Gasser SM. 2017 Chromatin and nucleosome dynamics in DNA damage and repair. *Genes Dev.* **31**, 2204–2221. (doi:10.1101/gad.307702.117)
204. Hansel-Hertsch R *et al.* 2016 G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272. (doi:10.1038/ng.3662)
205. Tippana R, Chen MC, Demeshkina NA, Ferré-D'Amaré AR, Myong S. 2019 RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36. *Nat. Commun.* **10**, 1855. (doi:10.1038/s41467-019-09802-w)
206. Vasilyev N, Polonskaia A, Darnell JC, Darnell RB, Patel DJ, Serganov A. 2015 Crystal structure reveals specific recognition of a G-quadruplex RNA by a  $\beta$ -turn in the RGG motif of FMRP. *Proc. Natl Acad. Sci. USA* **112**, E5391–E5400. (doi:10.1073/pnas.1515737112)
207. Voter AF, Qiu Y, Tippana R, Myong S, Keck JL. 2018 A guanine-flipping and sequestration mechanism for G-quadruplex unwinding by RecQ helicases. *Nat. Commun.* **9**, 4201. (doi:10.1038/s41467-018-06751-8)
208. Lerner LK, Sale JE. 2019 Replication of G quadruplex DNA. *Genes (Basel)* **10**, 96. (doi:10.3390/genes10020095)
209. Puig Lombardi E, Holmes A, Verga D, Teulade-Fichou MP, Nicolas A, Londono-Vallejo A. 2019 Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species. *Nucleic Acids Res.* **47**, 6098–6113. (doi:10.1093/nar/gkz463)
210. Weldon C, Eperon IC, Dominguez C. 2016 Do we know whether potential G-quadruplexes actually form in long functional RNA molecules? *Biochem. Soc. Trans.* **44**, 1761–1768. (doi:10.1042/bst20160109)
211. Sulovari A *et al.* 2019 Human-specific tandem repeat expansion and differential gene expression during primate evolution. *Proc. Natl Acad. Sci. USA* **116**, 23 243–23 253. (doi:10.1073/pnas.1912175116)
212. Finotti A, Fabbri E, Lampronti I, Gasparello J, Borgatti M, Gambari R. 2019 MicroRNAs and long non-coding RNAs in genetic diseases. *Mol. Diagn. Ther.* **23**, 155–171. (doi:10.1007/s40291-018-0380-6)
213. Sanders DW *et al.* 2020 Competing protein-RNA interaction networks control multiphase intercellular organization. *Cell* **181**, 306–324 e328. (doi:10.1016/j.cell.2020.03.050)
214. Hon CC *et al.* 2017 An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **543**, 199–204. (doi:10.1038/nature21374)
215. Kalwa M *et al.* 2016 The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. *Nucleic Acids Res.* **44**, 10 631–10 643. (doi:10.1093/nar/gkw802)
216. Sellaeri L, Bartolomei MS, Bickmore WA, He L, Stubbs L, Reik W, Barsh GS. 2016 A Hox-embedded long noncoding RNA: is it all hot air? *PLoS Genet.* **12**, e1006485. (doi:10.1371/journal.pgen.1006485)
217. Gao F, Cai Y, Kapranov P, Xu D. 2020 Reverse-genetics studies of lncRNAs—what we have learnt and paths forward. *Genome Biol.* **21**, 93. (doi:10.1186/s13059-020-01994-5)
218. Eissmann M *et al.* 2012 Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol.* **9**, 1076–1087. (doi:10.4161/ma.21089)
219. Kopp F. 2019 Molecular functions and biological roles of long non-coding RNAs in human physiology and disease. *J. Gene Med.* **21**, e3104. (doi:10.1002/jgm.3104)
220. Amodio N, Raimondi L, Juli G, Stamato MA, Caracciolo D, Tagliaferri P, Tassone P. 2018 MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. *J. Hematol. Oncol.* **11**, 63. (doi:10.1186/s13045-018-0606-4)
221. Duca M, Vekhoff P, Oussedik K, Halby L, Arimondo PB. 2008 The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* **36**, 5123–5138. (doi:10.1093/nar/gkn493)
222. Pupeza I *et al.* 2020 Field-resolved infrared spectroscopy of biological systems. *Nature* **577**, 52–59. (doi:10.1038/s41586-019-1850-7)
223. Kurosaki T, Popp MW, Maquat LE. 2019 Quality and quantity control of gene expression by



- nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* **20**, 406–420. (doi:10.1038/s41580-019-0126-2)
224. Payer LM *et al.* 2017 Structural variants caused by Alu insertions are associated with risks for many human diseases. *Proc. Natl Acad. Sci. USA* **114**, E3984–E3992. (doi:10.1073/pnas.1704117114)
225. Hormozdiari F *et al.* 2013 Rates and patterns of great ape retrotransposition. *Proc. Natl Acad. Sci. USA* **110**, 13 457–13 462. (doi:10.1073/pnas.1310914110)
226. Tsigros A, Rigoutsos I. 2009 Alu and B1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. *PLoS Comput. Biol.* **5**, e1000610. (doi:10.1371/journal.pcbi.1000610)
227. Rishishwar L, Wang L, Wang J, Yi SV, Lachance J, Jordan IK. 2018 Evidence for positive selection on recent human transposable element insertions. *Gene* **675**, 69–79. (doi:10.1016/j.gene.2018.06.077)
228. Paz-Yaacov N, Levanon EY, Nevo E, Kinar Y, Harmelin A, Jacob-Hirsch J, Amariglio N, Eisenberg E, Rechavi G. 2010 Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proc. Natl Acad. Sci. USA* **107**, 12 174–12 179. (doi:10.1073/pnas.1006183107)
229. Forni D *et al.* 2015 Diverse selective regimes shape genetic diversity at ADAR genes and at their coding targets. *RNA Biol.* **12**, 149–161. (doi:10.1080/15476286.2015.1017215)
230. Suzuki IK. 2020 Molecular drivers of human cerebral cortical evolution. *Neurosci. Res.* **151**, 1–14. (doi:10.1016/j.neures.2019.05.007)
231. Vasquez KM, Wang G. 2013 The yin and yang of repair mechanisms in DNA structure-induced genetic instability. *Mutat. Res.* **743–744**, 118–131. (doi:10.1016/j.mrfmmm.2012.11.005)
232. Lin L, Jiang P, Park JW, Wang J, Lu Z-x, Lam MPY, Ping P, Xing Y. 2016 The contribution of Alu exons to the human proteome. *Genome Biol.* **17**, 15. (doi:10.1186/s13059-016-0876-5)
233. Jangi M, Sharp PA. 2014 Building robust transcriptomes with master splicing factors. *Cell* **159**, 487–498. (doi:10.1016/j.cell.2014.09.054)
234. Song X *et al.* 2018 Predicting human genes susceptible to genomic instability associated with Alu/Alu-mediated rearrangements. *Genome Res.* **28**, 1228–1242. (doi:10.1101/gr.229401.117)
235. Del Mundo IMA, Zewail-Foote M, Kerwin SM, Vasquez KM. 2017 Alternative DNA structure formation in the mutagenic human c-MYC promoter. *Nucleic Acids Res.* **45**, 4929–4943. (doi:10.1093/nar/gkx100)
236. Bourque G *et al.* 2018 Ten things you should know about transposable elements. *Genome Biol.* **19**, 199. (doi:10.1186/s13059-018-1577-z)
237. Lio C-WJ, Rao A. 2019 TET enzymes and 5hmC in adaptive and innate immune systems. *Front. Immunol.* **10**, 210. (doi:10.3389/fimmu.2019.00210)