9462–9471 *Nucleic Acids Research, 2016, Vol. 44, No. 19* *Published online 4 July 2016*
doi: 10.1093/nar/gkw599

Inverted repeat *Alu* elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function

Isabel Chillón[1,2] and Anna M. Pyle[1,2,3,*]

[1]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA, [2]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA and [3]Department of Chemistry, Yale University, New Haven, CT 06511, USA

Received April 13, 2016; Revised June 1, 2016; Accepted June 22, 2016

ABSTRACT

LincRNA-p21 is a long intergenic non-coding RNA (lincRNA) involved in the p53-mediated stress response. We sequenced the human lincRNA-p21 (hLincRNA-p21) and found that it has a single exon that includes inverted repeat *Alu* elements (IR*Alus*). Sense and antisense *Alu* elements fold independently of one another into a secondary structure that is conserved in lincRNA-p21 among primates. Moreover, the structures formed by IR*Alus* are involved in the localization of hLincRNA-p21 in the nucleus, where hLincRNA-p21 colocalizes with paraspeckles. Our results underscore the importance of IR*Alus* structures for the function of hLincRNA-p21 during the stress response.

INTRODUCTION

When cells are exposed to stress signals like DNA damage, telomere erosion or hypoxia, a cellular stress response is triggered, leading to cell-cycle arrest, apoptosis and senescence (1). A key regulator of the cellular stress response is the protein p53, which is a transcription factor that can activate and repress many target genes (2,3). Recently, it was discovered that among p53 targets are genes for long non-coding RNAs (lncRNAs) (4). One of these p53-targeted lncRNAs is the long intergenic non-coding RNA p21 (lincRNA-p21, formally known as Tp53cor1) (5).

LincRNA-p21 was originally discovered in mice, where it is a transcript of ∼3 kb located proximal to the gene encoding the cell-cycle regulator p21/Cdkn1a (Figure 1A) (5). Under stress conditions, such as DNA damage, p53 activates transcription of mouse lincRNA-p21 (mLincRNA-p21), which accumulates in the nucleus and associates with the heterogeneous nuclear ribonucleoprotein K (hnRNP-K) to regulate transcription of specific target genes (5–7). LincRNA-p21 is also present in humans (hLincRNA-p21, formally known as TP53COR1). The hLincRNA-p21 has been proposed to promote apoptosis through a feedback mechanism that enhances p53 transcriptional activity in the nucleus (8). However, unlike mLincRNA-p21, hLincRNA-p21 is also exported to the cytoplasm, where it was found to repress the translation of specific target genes through activation of the RISC complex and to induce glycolysis under hypoxic conditions (9–11).

One limitation to further investigation into the role of hLincRNA-p21 is that its sequence has not yet been completely annotated. Furthermore, some reports on hLincRNA-p21 were carried out using truncated constructs (Figure 1A) (9,11–13). In addition, homology between mouse and human lincRNA-p21 is limited to a short region of 130 nt at the 5′ end of each transcript. The increasing evidence of an important role for hLincRNA-p21 in severe diseases, including colorectal cancer (14,15), skin cancer (16), liver cancer (17), rheumatoid arthritis (13) and atherosclerosis (8) makes it crucial that we obtain a deeper understanding of the molecular properties of hLincRNA-p21.

To this end, we have characterized the complete sequence of hLincRNA-p21 and we have discovered that the hLincRNA-p21 gene is transcribed into two isoforms that contain inverted repeat *Alu* elements (IR*Alus*) that influence function of the RNA. Both sense and antisense *Alu* elements of hLincRNA-p21 are highly conserved among primates and they fold as independent domains. We show that the IR*Alus* are involved in hLincRNA-p21 nuclear localization in the cell and that hLincRNA-p21 colocalizes with the lncRNA NEAT1 in paraspeckles during the course of the stress response. Our data establish the active isoforms of hLincRNA-p21 and their subcellular localization, and we demonstrate the critical role that tertiary structural elements play in hLincRNA-p21 function.

*To whom correspondence should be addressed. Tel: +1 203 432 5633; Fax: +1 203 432 5316; Email: anna.pyle@yale.edu

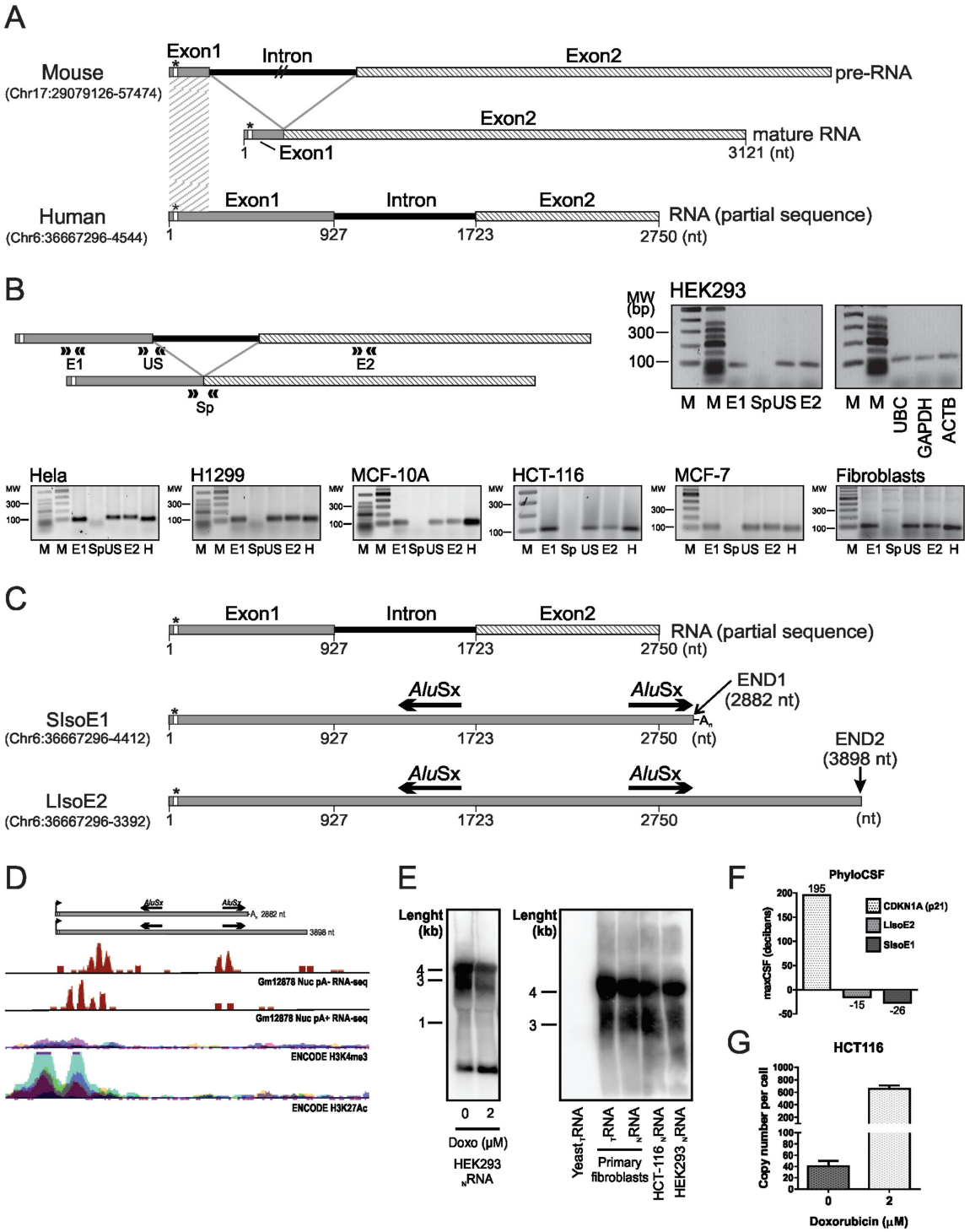

Figure 1. The human lincRNA-21 is a single exon lncRNA that contains IR*Alu* elements. (**A**) Diagram of precursor (Pre-RNA) and mature mouse lincRNA-p21 (mLincRNA-p21) and of the previously reported partial sequence of human lincRNA-p21 (hLincRNA-p21) [adapted from Yoon *et al.* (9)]. The asterisk represents the p53 protein binding site. The shaded box represents a region of sequence homology between mLincRNA-p21 and hLincRNA-p21. (**B**) RT-PCR performed on total RNA from HEK293, Hela, H1299, MCF-10A, HCT-116 and MCF-7 cell lines and on primary dermal fibroblasts. Four different primer pairs were used to amplify portions of exon 1 (E1), the putative exon1-intron junction (US), the putative exon1–exon2 junction (Sp) and exon 2 (E2). H stands for housekeeping gene (tRNA-Lys for HeLa and ACTB for all other cell lines). MW is a DNA size ladder expressed in base pairs. (**C**) Comparative diagram of the full-length sequence isoforms of hLincRNA-p21 as obtained by 3′ RACE on HEK293 cells and the previously reported partial sequence of hLincRNA-p21. (**D**) Schematic representation of hLincRNA-p21 with associated UCSC Genome Browser tracks depicting ENCODE H3K4me3 and H3K27Ac ChIP-seq coverage as well as CSHL Long RNA-seq coverage (26). (**E**) Northern blot analysis of hLincRNA-p21 in nuclear RNA from HEK293 and HCT-116 cell lines and of total and nuclear RNA from primary dermal fibroblasts. A probe hybridizing with the 5′ end of hLincRNA-p21 was used. (**F**) Maximum CSF scores of hLincRNA-p21 isoforms and CDKN1A (p21) RNAs determined by analysis with PhyloCSF (27). (**G**) Quantification of hLincRNA-p21 copy number per cell determined in the HCT-116 cell line by qRT-PCR.

MATERIALS AND METHODS

Rapid amplification of cDNA ends (RACE)

Total RNA was extracted from HEK293 cells using Trizol (Thermo). The 3′ RACE protocol for detecting polyadenylated transcripts was performed using the FirstChoice RLM-RACE Kit (Thermo), following manufacturer's recommendations. The 3′ RACE protocol for detecting non-polyadenylated transcripts was performed by ligating total RNA to a 3′ RNA adapter using T4 RNA ligase (NEB) at 37°C for 2 h followed by overnight incubation at 16°C. The cDNA was generated using Thermoscript reverse transcriptase (Thermo) with a primer complementary to the RNA adapter. The sequence of the 3′ end of the molecule was determined by nested polymerase chain reaction (PCR) with primers that anneal to the RNA adapter and specific primers that anneal to known regions of the hLincRNA-p21 sequence (9). The list of primers and oligonucleotides used for 3′ RACE are provided in Supplementary Table S1.

Cell lines, DNA damage induction and transfections

HEK293 (ATCC® CRL-1573), HeLa (ATCC® CCL-2) and MCF-7 cell lines were grown in Dulbecco's modified Eagle's medium medium containing high glucose (Thermo) supplemented with fetal bovine serum to a final concentration of 10%. NCI-H1299 (ATCC® CRL5803) cell line was grown in RPMI-1640 medium (Thermo) and HCT-116 (ATCC® CCL247) and U-2 OS (ATCC® HTB96) cell lines in McCoy's 5a medium modified (Thermo) supplemented with fetal bovine serum to a final concentration of 10%. Normal human primary dermal fibroblasts from neonatal foreskin (ATCC® PCS-201-010) were grown in fibroblast basal medium supplemented with Fibroblast Growth Kit–Low Serum (ATCC PCS201041). MCF-7 and the non-tumorigenic breast MCF-10A cell line were a generous gift of Dr Brian Adams (Harvard University). To induce DNA damage, cells were treated with 0.5, 1 or 2 μM doxorubicin hydrochloride (Sigma) for 12 h. HEK293 cells (2×10^5 cells/well) were transfected with 250 ng of DNA using Lipofectamine 2000 (Thermo), according to manufacturer's instructions. Transfected cells were incubated in complete medium for 36 h.

Northern blot analysis

Northern blot analysis was performed using the NorthernMax® Kit (Thermo), following manufacturer's instructions. Nuclear RNA was extracted using the Nuclei Pure Prep Nuclei Isolation Kit (Sigma) and further purified with the RNeasy Maxi Kit (Qiagen). A ^{32}P body-labeled RNA probe complementary to the first 920 nt of hLincRNA-p21 was used for hybridization. Hybridization was performed overnight at 60°C in a hybridization oven with orbital shaking.

Total RNA extraction and Nuclear/Cytoplasmic RNA fractionation

Total RNA was extracted using Trizol (Thermo) according to manufacturer's instructions. Nuclear and cytoplasmic RNA fractionation was performed as in Hu *et al*. (18). After extraction, total and fractionated RNAs were treated with 10 U of TURBO DNase (Thermo) for 1 h at 37°C.

Reverse transcription PCR (RT-PCR) and quantitative reverse-transcription PCR (qRT-PCR)

The cDNA was generated by reverse transcription (RT) using random hexamers (Thermo) and SuperScript II reverse transcriptase (Thermo) from total or fractionated RNA. The PCR reaction was performed using Platinum Taq DNA Polymerase (Thermo). The qRT-PCR was performed on cDNA using the LightCycler® 480 SYBR Green I Master (Roche). Supplementary Table S1 reports a list of all primers used. See Supplementary Data for detailed descriptions.

RNA fluorescence *in situ* hybridization (RNA-FISH)

RNA fluorescence in situ hybridization (RNA-FISH) was performed as in Byron *et al*. (19) and Trcek *et al*. (20) with some modifications. Briefly, fixed cells were hybridized with 10 ng of RNA probe, followed by overnight incubation. RNA probes were obtained by *in vitro* transcription in the presence of DIG-11-UTP (Roche) for hLincRNA-p21 and ChromaTide® Alexa Fluor® 488-5-UTP (Thermo) for NEAT1 RNA and MS2 tag. After washing, indirect fluorescent labeling of hLincRNA-p21 was performed using Cy3-conjugated anti-digoxigenin antibody (Jackson Immuno Research). See Supplementary Data for detailed descriptions.

Confocal imaging and colocalization analysis

RNA-FISH images were acquired using a Zeiss LSM 510 confocal microscope (Carl Zeiss Inc.) equipped with a 63X Plan-Apo/1.4 NA Oil with DIC capability objective. Cells were imaged using 0.75 μm z-dimension axis steps across a range of approximately 3–4 μm and a pixel dwell time of 6.39–25.6 μs. To calculate the fraction of NEAT1 lncRNA that colocalizes with hLincRNA-p21, the plugin Coloc2, written for Image J (NIH) (21), was used. Using nuclei as a region of interest, Mander's coefficients with thresholds (22) and the associated Costes P-values (23) were obtained for every cell.

Sedimentation-velocity analytical ultracentrifugation (SV-AUC)

Sedimentation-velocity analytical ultracentrifugation (SV-AUC) was performed according to Chillon *et al*. (24). See Supplementary Data for detailed descriptions.

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE)

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) analysis was performed according to Chillon *et al*. (24). See Supplementary Data for detailed descriptions.

Analysis of RNA modifications

Total RNA was extracted from HEK293 cells using Trizol (Thermo) and reverse transcribed with SuperScript II reverse transcriptase (Thermo) using either random hexamers (Thermo) or a hLincRNA-p21 specific primer (Supplementary Table S1). Total DNA was extracted from HEK293 cells using DNAZOL (Thermo). Two regions of hLincRNA-p21 covering the sense and the antisense *Alu* sequences were amplified by PCR, cloned in the pGEM-T Easy Vector (Promega) and sequenced. Sequences from cDNA and genomic DNA were aligned using SeqMan (DNASTAR, Inc). Supplementary Table S1 reports the list of primers used.

Covariation analysis

Covariation analysis was performed as described previously (25). See Supplementary Data for detailed descriptions.

RESULTS AND DISCUSSION

hLincRNA-p21 contains IR*Alu* elements in its sequence

A partial 2750-nt long sequence of hLincRNA-p21 had been obtained previously by Yoon *et al*. (9), who proposed that hLincRNA-p21 may contain an intron comprising nucleotides 928 to 1722, in analogy with mLincRNA-p21 (Figure 1A). To determine the splicing levels of hLincRNA-p21 *in vivo*, we performed RT-PCR on total RNA extracted from six different human cell lines, including one non-tumorigenic and five cancerous cell lines, and from one primary culture, and we amplified different regions of hLincRNA-p21 (Figure 1B). No product was observed from primers that span the putative exon–exon junction in any cell culture analyzed, leading us to conclude that hLincRNA-p21 does not contain an intron and is in fact a single-exon lncRNA (Figure 1C).

We next determined the 5′ and 3′ boundaries of hLincRNA-p21. At the 5′ end, hLincRNA-p21 was reported to start at position chr6:36,667,296 (GRCh38/hg38) (9) and this transcription start site can also be identified in RNA-seq data obtained from the ENCODE project (Figure 1D). However, the 3′ boundary of hLincRNA-p21 had remained completely uncharacterized. We detected two different 3′ end sites by 3′-RACE (Figure 1C). Using conventional oligo(dT) priming, we detected a 3′ end site (end site 1) located 2882 nt downstream of the transcription start site and adjacent to a genomically encoded polyadenosine tract and a polyadenylation signal. We refer to this isoform as SIsoE1 (Short Isoform Ending at site 1). However, when priming was achieved by linking an RNA oligonucleotide adapter to the 3′ end of total cellular RNA, we detected a second 3′ end site (end site 2), located 3898 nt downstream of the transcription start site. At this end site, hLincRNA-p21 is non-polyadenylated and we refer to this isoform as LIsoE2 (Long Isoform Ending at site 2). Both end sites are supported by RNA-seq data obtained from the ENCODE project (Figure 1D) (26). Moreover, we confirmed the existence of these two hLincRNA-p21 isoforms by northern blotting (Figure 1E), which reveals the presence of two major bands migrating at approximately 3000 and 4000 nt in length, respectively.

To analyze if the sequence of these two isoforms of hLincRNA-p21 possess protein-coding capacity, we examined their coding potential using PhyloCSF (27) and compared it with that of the gene encoding the cell-cycle regulator p21(CDKN1A) (Figure 1F). The analysis revealed that, while the CDKN1A gene has an expectedly high coding potential (positive value), neither of the two hLincRNA-p21 isoforms has any coding potential (negative values) (Figure 1F). Based on this analysis, we confirm that hLincRNA-p21 is a *bona fide* non-coding RNA. Moreover, to analyze if hLincRNA-p21 is in fact an intergenic RNA, we inspected the region surrounding hLincRNA-p21 in the genome using publicly-available RNA-seq datasets and checked the latest RefSeq records (Annotation Release 107). We did not detect any transcripts overlapping with hLincRNA-p21 in the antisense orientation.

Surprisingly, both hLincRNA-p21 isoforms contain two inverted repeat *Alu* elements (IR*Alus*) (Figure 1C). The sense element of these IR*Alus* is located at positions 2589–2895 and the antisense *Alu* element is located at positions 1351–1651. The distance between the sense and the antisense *Alu* elements in hLincRNA-p21 is 937 nt, which is similar to the distance between functional inverted *Alu* elements identified in other RNAs (28–31). Both *Alu* elements belong to the Sx1 subfamily of short interspersed nuclear elements (SINEs). As expected for *Alu* elements, they are present only in lincRNA-p21 sequences from primates, where they show a high degree of sequence conservation (Supplementary Figure S1).

Since *Alu* elements in certain pre-mRNAs induce the formation of circular RNA isoforms (32), we checked for such a possibility by conducting RT-PCR using divergent primers. We did not observe any circular forms of hLincRNA-p21.

We also determined the abundance of hLincRNA-p21 in cells in the presence and absence of stress. Compared to mLincRNA-p21, which is present at 2 copies per cell in the absence of stress and at 8 copies per cell in the presence of stress (5,6), hLincRNA-p21 is more abundant, with 40 copies per cell in the absence of stress and more than 600 copies per cell in the presence of the chemotherapeutic drug doxorubicin (Figure 1G). Similar levels of hLincRNA-p21 in different cell lines were reported elsewhere (11).

In summary, hLincRNA-p21 is a linear, single exon lncRNA that contains IR*Alu* elements. Such properties clearly distinguish hLincRNA-p21 from mLincRNA-p21, despite the fact that these RNAs are located in syntenic regions in human chromosome 17 and mouse chromosome 6, respectively. Homology analyses of these genes show that hLincRNA-p21 and mLincRNA-p21 only share 130 nt at their 5′ end, which include a p53-response element and the binding site for protein hnRNP K (Figure 1A) (5,6,8), while the rest of their gene sequences are different. This difference suggests that mlincRNA-p21 and hLincRNA-p21 evolved independently and diverged after the split of mouse and human lineages.

The hLincRNA-p21 inverted repeat *Alu* elements form independent structural domains *in vitro*

Next, we determined the secondary structure of the IR*Alu* elements within hLincRNA-p21 using an approach previously applied to other lncRNAs (25). To this end, we transcribed hLincRNA-p21 *in vitro* under non-denaturing conditions (24) and obtained high yields of pure and homogenous RNA (Supplementary Figure S2A). Sedimentation-velocity analytical ultracentrifugation showed that hLincRNA-p21 folds at near physiological concentrations of $MgCl_2$ ($K_{1/2Mg} = 5$ mM) and reaches maximum compaction at 15 mM $MgCl_2$ (Supplementary Figure S2B). We determined the secondary structure of the sense and antisense *Alu* elements in the context of the long isoform using SHAPE, at both 5 and 15 mM $MgCl_2$, using an established protocol (33). Reactivity values obtained at 5 and 15 mM $MgCl_2$ correlated well for all nucleotides (Supplementary Figure S2C and D). We observed that both sense and antisense *Alu* elements are composed of a left and a right arm, defined such that the left arm of the sense *Alu* is homologous to the right arm of the antisense *Alu*, and *vice versa* (Figure 2A and B). The two arms of each *Alu* element are connected by a single-stranded region with high SHAPE reactivity. Each arm consists of a 5′ domain characterized by a central three-way junction and a 3′ domain characterized by a long stem-loop. Residues located in the terminal loops of the 5′ domain in each arm display low SHAPE reactivity, which is consistent with putative interactions via two or three base pairs, as observed for other *Alu* elements (dotted lines, Figure 2A and B) (34). Overall, the secondary structures we have determined for the IR*Alus* in hLincRNA-p21 are almost identical to those reported through enzymatic and comparative analyses for other *Alu* elements and *Alu*-related RNAs (35,36). Our structures also agree with the reported crystal structures of a mammalian *Alu* element and of the *Alu* domain of the signal recognition particle (37,38). Considering that RNA may fold co-transcriptionally (39,40), the separate organization of the two IR*Alus* into two independently folded units is likely to be energetically more favorable than a previously proposed foldback structure in which base pairs are formed between the sense and the antisense *Alu* elements (28,41).

Functional confirmation of our structures comes from a covariation analysis of the hLincRNA-p21 *Alu* elements (Figure 2A and B, Supplementary Figure S3A and B). We compared sequences of lincRNA-p21 from 10 primate species. Species beyond primates were excluded because *Alu* elements are restricted to this taxonomic order (42). Overall, we observed a high level of sequence conservation across the *Alu* elements of lincRNA-p21 (Figure 2A and B, insets, red boxes; Supplementary Figure S3A and B), including consistent half-flips and covariant base-pairs that preserve their structural architecture (Figure 2A and B, insets, blue and green boxes, respectively; Supplementary Figure S3A and B). Thus, our analysis reveals selective evolutionary pressure to preserve the *Alu* elements embedded in lincRNA-p21.

Given that *Alu* sequences frequently contain post-transcriptional modifications, such as adenosine-to-inosine editing (28), we also examined whether nucleotides in hLincRNA-p21 are edited. To this end, we sequenced hLincRNA-p21 cDNA from HEK293 cells using random hexamers and primers that anneal specifically to hLincRNA-p21 and we compared the sequence of the cDNA with that of the genomic DNA extracted from the same cells (Figure 2A and B, Supplementary Figure S2E). Surprisingly, all cDNA sequences obtained from random and from sequence-specific primers, but none of the genomic DNA sequences, contained unusual uridine-to-cytosine transitions in both the sense and the antisense *Alu* sequences. No modifications were observed in cDNA sequences outside of the *Alu* regions. In addition, we calculated the frequency of occurrence and the positions of these observed substitutions in the secondary structure. We observed that 80–90% of the events are localized within single-stranded regions or in the vicinity of loops, and would not affect the conserved structure of hLincRNA-p21 *Alu* elements. Interestingly, we noted that some uridines that undergo U-to-C interchanges correspond to cytidines in the genome of primate ancestors. This observation is notable in light of a recently described process named 'RNA memory of evolution', by which RNA editing is used to revert DNA mutations to their ancestral sequence (43). Nevertheless, we are unsure whether these observations are the result of bona-fide U-to-C editing process (44,45) or the signature of some type of RNA modification, as no enzyme is known to perform U-to-C RNA editing.

In summary, we found that the sense and antisense *Alu* elements form independent structural domains in the context of hLincRNA-p21 and that these domains are highly conserved among primates.

The structure of the IR*Alu* elements contributes to hLincRNA-p21 nuclear localization

In many cells, mRNAs containing inverted repeat elements in their 3′ untranslated regions are retained in the nucleus in paraspeckles through interaction with the dsRNA-binding protein p54nrb (31,46). Therefore, we investigated if the secondary structure of the IR*Alu* elements of hLincRNA-p21 affects the localization of this RNA to the nucleus. To this end, we transfected HEK293 cells with hLincRNA-p21 constructs containing mutations in the IR*Alu* elements and compared their cellular distribution with that of wild-type (WT) constructs of both hLincRNA-p21 isoforms SIsoE1 and LIsoE2 (Figure 3). Specifically, we disrupted the secondary structure of the IR*Alu*s by replacing the nucleotides within the 5′ portion of each helical stem (Figure 3A). We produced three types of mutants: those that contain mutations only in the sense *Alu* element (dS-Alu), only in the antisense *Alu* element (dA-Alu) and in both *Alu* elements (dIR-Alu). In addition, we used a truncated construct (5E), which comprises the first 920 nt of hLincRNA-p21 and thus preserves all known regulatory elements of this RNA, including the p53 response element and the hRNP K binding region (5), but lacks both *Alu* elements. If the IR*Alus* are necessary for nuclear retention, then the resulting mutants, with disrupted IR*Alu* secondary structure or missing IR*Alu*s, should escape from the nucleus. To quantify expression levels of WT and mutant hLincRNA-p21 in the nuclear and cytoplasmic compartments, we conducted cellu-

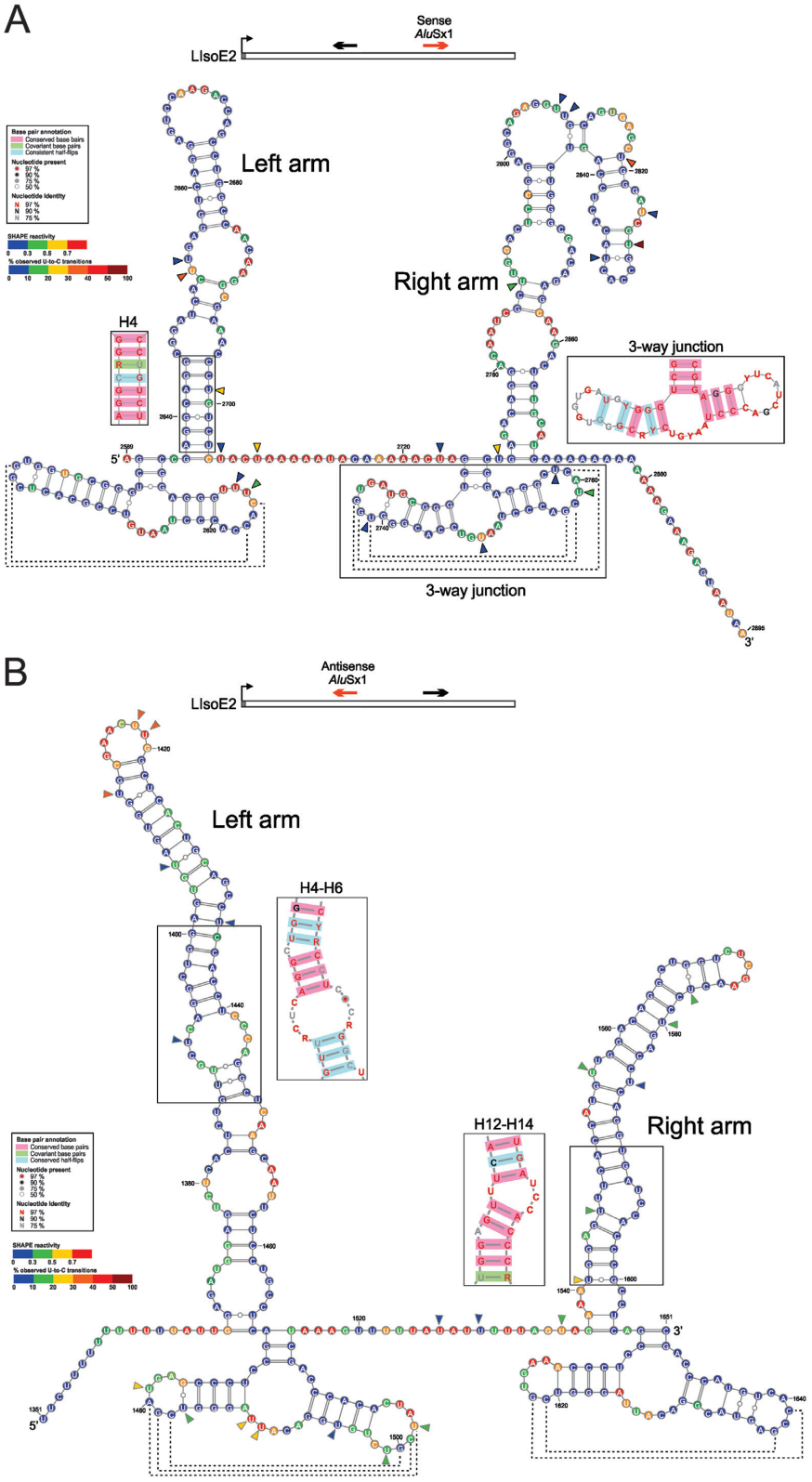

Figure 2. Inverted repeat *Alu* elements (IR*Alus*) in hLincRNA-21 fold as independent structural domains *in vitro*. (**A** and **B**) Experimentally-derived secondary structures of sense (**C**) and antisense (**D**) *Alu* sequences of hLincRNA-p21. A schematic diagram of hLincRNA-p21 LIsoE2 isoform highlights the relative position of each structure in the RNA sequence (top). Colors of the circles around each nucleotide indicate SHAPE reactivity. The dotted lines represent putative tertiary contacts between the terminal loops of the three-way junctions. Data values are the mean of two biological replicates. Observed U to C transitions are indicated by colored arrowheads. The frequency values represent the fraction of observed transitions of $n = 93$ sequences from eight biological replicates for the sense *Alu* and $n = 35$ sequences from three biological replicates for the antisense *Alu*. Covariation of lincRNA-p21 in human and nine other primates is shown in boxes for selected regions (See also Supplementary Figure S3).

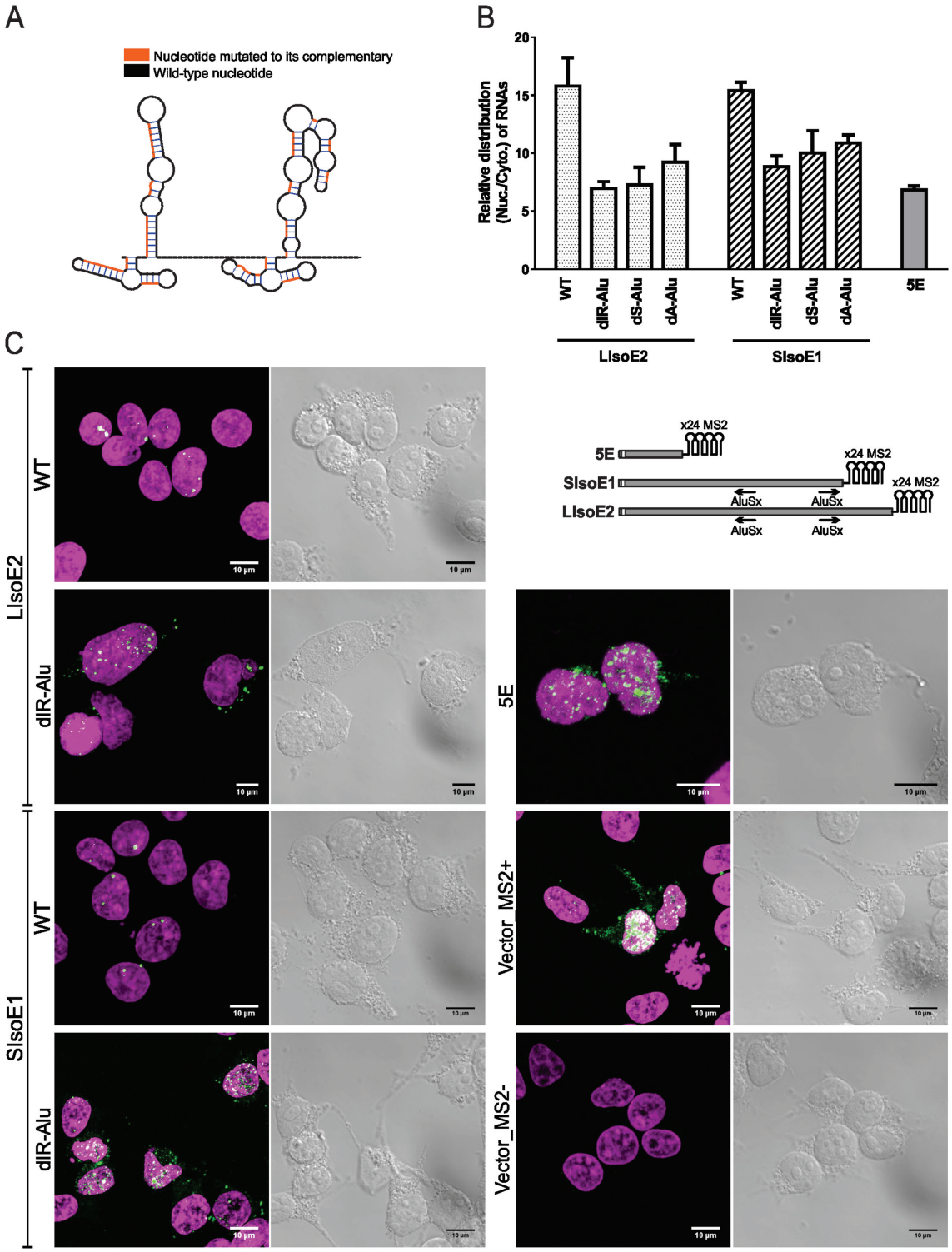

Figure 3. The secondary structure of the inverted repeat *Alu* (IR*Alu*) elements is involved in hLincRNA-p21 nuclear localization. (**A**) Schematic representation of an *Alu* element indicating the strategy followed to disrupt the secondary structure of the RNA helices for each *Alu* element. (**B**) Relative subcellular distribution of various hLincRNA-p21 constructs quantified by qRT-PCR using nuclear and cytoplasmic RNA fractions. Wild-type (WT), disrupted IR*Alu* elements (dIR-Alu), disrupted sense *Alu* element (dS-Alu) and disrupted antisense *Alu* element (dA-Alu) constructs derived from the short isoform ending at site 1 (SIsoE1) and the long isoform ending at site 2 (LIsoE2) were used. A truncated 5E construct, retaining the first 920 nt of hLincRNA-p21 was also included. β-actin was used for normalization. Data represent the mean ± SEM of three biological replicates. The *P*-values obtained from unpaired *t*-test are <0.05 for all samples, assuming a level of significance $\alpha = 0.05$. (**C**) Representative RNA-FISH maximum projection images depicting cellular distribution of WT and selected mutant hLincRNA-p21 constructs transfected in HEK293 cells. Cells were harvested 36 h after transfection and RNA-FISH was performed using a probe complementary to the MS2 hairpins (green). Nuclei were identified with NucRed staining (magenta). Schematic representations of the MS2-tailed constructs used for the heterologous expression of hLincRNA-p21 are also shown (upper right).

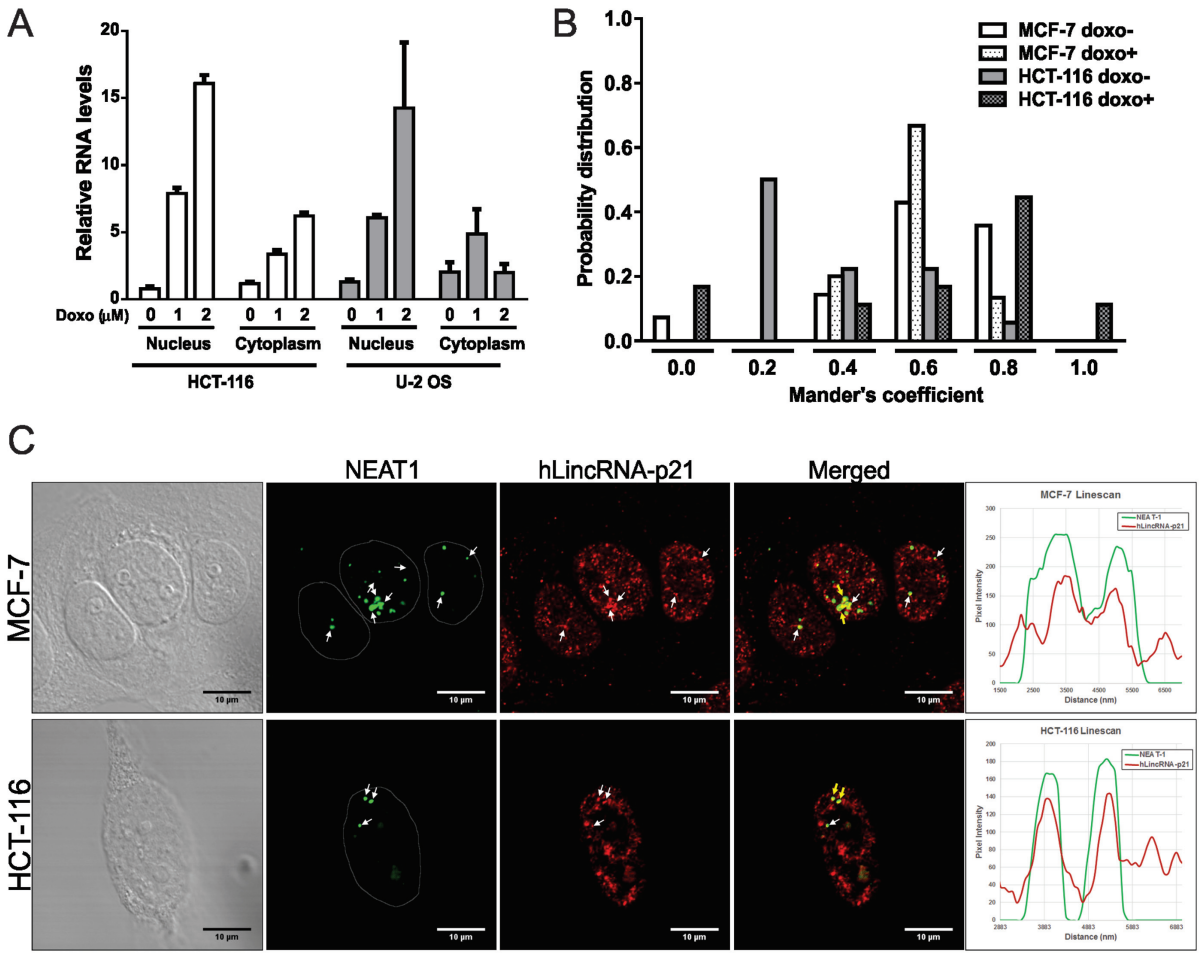

Figure 4. The hLincRNA-p21 colocalizes with paraspeckles in the nucleus. (**A**) Quantification of endogenous hLincRNA-p21 by qRT-PCR in the nuclear and cytoplasmic fractions after the treatment with increasing concentrations of doxorubicin in different cell lines for 12 h. Data represent the mean ± SEM of two biological replicates. (**B**) The distribution of the fraction of NEAT1 spots that colocalize with hLincRNA-21 in doxorubicin-treated and untreated cells ($n = 14$ at 0 μM and $n = 15$ at 1 μM doxorubicin for MCF-7 and $n = 18$ at 0 μM and $n = 18$ at 1 μM doxorubicin for HCT-116). Mander's coefficients, indicating the degree of overlap, were obtained using Coloc2 (ImageJ) from RNA-FISH experiments performed in different cell lines. (**C**) Representative RNA-FISH single confocal plane images showing cellular distribution of endogenous hLincRNA-p21 and colocalization with NEAT1 in different cell lines at 1 μM doxorubicin. Overlapping foci between NEAT1 and hLincRNA-p21 are highlighted by arrows. Linescan analyses (right panel) were performed through the spots indicated by yellow arrows in the merge panel.

lar fractionation followed by qRT-PCR (Figure 3B, Supplementary Figure S4A and B). We observed ∼2-fold enrichment in the nucleus of the WT constructs compared to the mutant constructs (Figure 3B). Importantly, restoring the secondary structure of the sense *Alu* element by compensatory mutations to a disrupted construct partially recovered the WT phenotype (78% capacity of nuclear retention of the restored construct, dS-res, compared to 57% of the disrupted construct, dS-Alu, in respect to WT, Supplementary Figure S4C). We confirmed these results by RNA-FISH using the above constructs modified with an MS2 tag extension and a probe complementary to the MS2 hairpins, allowing us to discriminate between endogenous and transfected molecules (Figure 3C). Although nuclear delocalization of the mutant constructs is statistically significant (Figure 3B), there is still a considerable amount of nuclear retention for all of the mutant RNAs. This observation suggests that other factors, such as the interaction of the 5′ end of hLincRNA-p21 with protein hnRNP K (5,8), may also regulate the cellular localization of hLincRNA-p21 in addition to IR*Alus*.

Taken together, these data show that the structures of the IR*Alu* elements formed by hLincRNA-p21 are an important determinant for nuclear localization of this lncRNA.

hLincRNA-p21 colocalizes with paraspeckles *in vivo* during the course of the stress response

To gain insight into the functional relevance of hLincRNA-p21 nuclear localization, we next sought to examine the subcellular distribution of the endogenous hLincRNA-p21 in different cell lines. We quantified the expression of hLincRNA-p21 by qRT-PCR in the nuclear and cytoplasmic compartments of two cell lines before and after inducing DNA damage by the chemotherapeutic drug doxorubicin (Figure 4A and Supplementary Figure S5A). We observed that the levels of hLincRNA-p21 are equally abundant in the nucleus and cytoplasm of cells in the absence

of doxorubicin and that they increase exponentially after DNA damage. Specifically, in the presence of doxorubicin, hLincRNA-p21 was enriched up to 17-fold for U-2 OS cells and 3-fold for HCT-116 cells, indicating that this lncRNA does indeed predominantly accumulate in the nucleus. In line with previous reports (9,11), we also observed a fraction of hLincRNA-p21 in the cytoplasm after hLincRNA-p21 induction (30% in HCT-116 and <10% in U-2 OS cells).

To ascertain whether hLincRNA-p21 is incorporated into paraspeckles, we examined the colocalization of hLincRNA-p21 and the lncRNA NEAT1 (also known as MEN epsilon), an essential paraspeckle component (47–49), by RNA-FISH (Figure 4B and C, Supplementary Figure S5B and C). We investigated the colocalization of hLincRNA-p21 and NEAT1 in untreated cells as well as in the presence of doxorubicin. Doxorubicin concentration was restricted to 1 μM as paraspeckles disorganize and the expression of lncRNA NEAT1 decreases exponentially in response to higher concentrations of this chemotherapeutic drug (Supplementary Figure S5D) or to other types of stresses (18,50). We scored the correlation between hLincRNA-p21 and NEAT1 and found that 40–60% of the total signal of NEAT1 colocalized with hLincRNA-p21 (Figure 4B and Supplementary Figure S5B), both in the absence and presence of DNA damage for MCF-7 and HCT-116 cells, which is considered a case of medium to high colocalization (51).

Taken together, these data show that hLincRNA-p21 is predominantly localized in the nucleus, where it accumulates in the presence and in the absence of stress conditions. Thus, hLincRNA-p21 differs significantly from some mRNAs that contain inverted repeat (IR) elements, e.g. from mouse CTN-RNA, which is released from paraspeckles upon cellular stress, and undergoes cleavage of its IR elements and consequent translocation of an IR-depleted mRNA to the cytoplasm (46).

In conclusion, our work reveals that hLincRNA-p21 possesses inverted repeat *Alu* elements (IR*Alus*) in its mature, single exon functional isoforms. These IR*Alu* elements form structural domains that are conserved in primates and which regulate the cellular localization of hLincRNA-p21 over the course of the stress response. These observations help explain properties of hLincRNA-p21 that were previously considered contradictory, such as the ambiguous localization of this RNA and its morphological differences with respect to mLincRNA-p21. Future work should determine whether hLincRNA-p21, like mLincRNA-p21, has an active role in transcriptional regulation. Our results suggest a regulatory role that is synergistic with NEAT1, which has recently been found to localize at genomic sites over active genes (52). However, we cannot rule out that hLincRNA-p21 is a target of NEAT1 transcriptional regulation or that it is retained in paraspeckles as an inactive form. Independent of its exact physiological role, our work shows that the molecular mechanism of hLincRNA-p21 can be fully understood only in the context of its RNA structure. Structural domains, like IR*Alu*s, may be crucial in determining how this lncRNA acts as a regulatory checkpoint during the stress response and it may serve as potential new target for chemotherapies.

ACCESSION NUMBERS

The human lincRNA-p21 sequences reported in this paper have been deposited in GenBank with accession numbers KU881768 (bankIt 1900853) for LIsoE2 (Long Isoform Ending at site 2) and KU881769 (bankIt 1900853) for SIsoE1 (Short Isoform Ending at site 1).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Brian Adams (Harvard University) for the generous gift of MCF-7 and MCF-10A cells. We thank Dr Joseph Wolenski and Dr Sandra Wolin (Yale University) for help setting up and interpreting the microscopy assays, Dr Valerie Horsley (Yale University) for providing cell culture facilities and Dr Erik Jagdmann for the synthesis of 1M7. We also thank all members of the Pyle lab, and in particular Dr Olga Fedorova for synthesizing the RNA oligonucleotide for the RACE experiments, Dr Nathan Pirakitikulr for help on SHAPE analysis and Dr Joseph Liberman for constructive criticism and review of the manuscript. We are also thankful to Dr Nadya Dimitrova (Yale University) and Dr Marco Marcia (EMBL Grenoble) for insightful discussions and critical review of the manuscript.

FUNDING

Howard Hughes Medical Institute. Funding for open access charge: Howard Hughes Medical Institute.
Conflict of interest statement. None declared.

REFERENCES

1. Horn,H.F. and Vousden,K.H. (2007) Coping with stress: multiple ways to activate p53. *Oncogene*, **26**, 1306–1316.
2. Murray-Zmijewski,F., Slee,E.A. and Lu,X. (2008) A complex barcode underlies the heterogeneous response of p53 to stress. *Nat. Rev. Mol. Cell Biol.*, **9**, 702–712.
3. Riley,T., Sontag,E., Chen,P. and Levine,A. (2008) Transcriptional control of human p53-regulated genes. *Nat. Rev. Mol. Cell Biol.*, **9**, 402–412.
4. Guttman,M., Amit,I., Garber,M., French,C., Lin,M.F., Feldser,D., Huarte,M., Zuk,O., Carey,B.W., Cassady,J.P. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
5. Huarte,M., Guttman,M., Feldser,D., Garber,M., Koziol,M.J., Kenzelmann-Broz,D., Khalil,A.M., Zuk,O., Amit,I., Rabani,M. *et al.* (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, **142**, 409–419.
6. Dimitrova,N., Zamudio,J.R., Jong,R.M., Soukup,D., Resnick,R., Sarma,K., Ward,A.J., Raj,A., Lee,J.T., Sharp,P.A. *et al.* (2014) LincRNA-p21 activates p21 in cis to promote polycomb target gene expression and to enforce the G1/S checkpoint. *Mol. Cell*, **54**, 777–790.
7. Bao,X., Wu,H., Zhu,X., Guo,X., Hutchins,A.P., Luo,Z., Song,H., Chen,Y., Lai,K., Yin,M. *et al.* (2015) The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res.*, **25**, 80–92.
8. Wu,G., Cai,J., Han,Y., Chen,J., Huang,Z.P., Chen,C., Cai,Y., Huang,H., Yang,Y., Liu,Y. *et al.* (2014) LincRNA-p21 regulates neointima formation, vascular smooth muscle cell proliferation, apoptosis, and atherosclerosis by enhancing p53 activity. *Circulation*, **130**, 1452–1465.

9. Yoon,J.H., Abdelmohsen,K., Srikantan,S., Yang,X., Martindale,J.L., De,S., Huarte,M., Zhan,M., Becker,K.G. and Gorospe,M. (2012) LincRNA-p21 suppresses target mRNA translation. *Mol. Cell*, **47**, 648–655.
10. Chou,S.D., Murshid,A., Eguchi,T., Gong,J. and Calderwood,S.K. (2015) HSF1 regulation of beta-catenin in mammary cancer cells through control of HuR/elavL1 expression. *Oncogene*, **34**, 2178–2188.
11. Yang,F., Zhang,H., Mei,Y. and Wu,M. (2014) Reciprocal regulation of HIF-1alpha and lincRNA-p21 modulates the Warburg effect. *Mol. Cell*, **53**, 88–100.
12. Tran,U.M., Rajarajacholan,U., Soh,J., Kim,T.S., Thalappilly,S., Sensen,C.W. and Riabowol,K. (2015) LincRNA-p21 acts as a mediator of ING1b-induced apoptosis. *Cell Death Dis.*, **6**, e1668.
13. Spurlock,C.F. 3rd, Tossberg,J.T., Matlock,B.K., Olsen,N.J. and Aune,T.M. (2014) Methotrexate inhibits NF-kappaB activity via long intergenic (noncoding) RNA-p21 induction. *Arthritis Rheumatol.*, **66**, 2947–2957.
14. Zhai,H., Fesler,A., Schee,K., Fodstad,O., Flatmark,K. and Ju,J. (2013) Clinical significance of long intergenic noncoding RNA-p21 in colorectal cancer. *Clin. Colorectal Cancer*, **12**, 261–266.
15. Wang,G., Li,Z., Zhao,Q., Zhu,Y., Zhao,C., Li,X., Ma,Z., Li,X. and Zhang,Y. (2014) LincRNA-p21 enhances the sensitivity of radiotherapy for human colorectal cancer by targeting the Wnt/beta-catenin signaling pathway. *Oncol. Rep.*, **31**, 1839–1845.
16. Hall,J.R., Messenger,Z.J., Tam,H.W., Phillips,S.L., Recio,L. and Smart,R.C. (2015) Long noncoding RNA lincRNA-p21 is the major mediator of UVB-induced and p53-dependent apoptosis in keratinocytes. *Cell Death Dis.*, **6**, e1700.
17. Yang,N., Fu,Y., Zhang,H., Sima,H., Zhu,N. and Yang,G. (2015) LincRNA-p21 activates endoplasmic reticulum stress and inhibits hepatocellular carcinoma. *Oncotarget*, **6**, 28151–28163.
18. Hu,S.B., Xiang,J.F., Li,X., Xu,Y., Xue,W., Huang,M., Wong,C.C., Sagum,C.A., Bedford,M.T., Yang,L. *et al.* (2015) Protein arginine methyltransferase CARM1 attenuates the paraspeckle-mediated nuclear retention of mRNAs containing IRAlus. *Genes Dev.*, **29**, 630–645.
19. Byron,M., Hall,L.L. and Lawrence,J.B. (2013) A multifaceted FISH approach to study endogenous RNAs and DNAs in native nuclear and cell structures. *Curr. Protoc. Hum. Genet.*, Chapter 4, Unit 4.
20. Trcek,T., Chao,J.A., Larson,D.R., Park,H.Y., Zenklusen,D., Shenoy,S.M. and Singer,R.H. (2012) Single-mRNA counting using fluorescent in situ hybridization in budding yeast. *Nat. Protoc.*, **7**, 408–419.
21. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
22. Manders,E.M.M., Verbeek,F.J. and Aten,J.A. (1993) Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.*, **169**, 375–382.
23. Costes,S.V., Daelemans,D., Cho,E.H., Dobbin,Z., Pavlakis,G. and Lockett,S. (2004) Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys. J.*, **86**, 3993–4003.
24. Chillon,I., Marcia,M., Legiewicz,M., Liu,F., Somarowthu,S. and Pyle,A.M. (2015) Native purification and analysis of long RNAs. *Methods Enzymol.*, **558**, 3–37.
25. Somarowthu,S., Legiewicz,M., Chillon,I., Marcia,M., Liu,F. and Pyle,A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Mol. Cell*, **58**, 353–361.
26. Rosenbloom,K.R., Sloan,C.A., Malladi,V.S., Dreszer,T.R., Learned,K., Kirkup,V.M., Wong,M.C., Maddren,M., Fang,R., Heitner,S.G. *et al.* (2013) ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.*, **41**, D56–D63.
27. Lin,M.F., Jungreis,I. and Kellis,M. (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*, **27**, i275–i282.
28. Athanasiadis,A., Rich,A. and Maas,S. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.*, **2**, e391.
29. Bass,B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.*, **71**, 817–846.
30. Chen,L.L. and Carmichael,G.G. (2008) Gene regulation by SINES and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. *Cell Cycle*, **7**, 3294–3301.
31. Chen,L.L., DeCerbo,J.N. and Carmichael,G.G. (2008) Alu element-mediated gene silencing. *EMBO J.*, **27**, 1694–1705.
32. Liang,D. and Wilusz,J.E. (2014) Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.*, **28**, 2233–2247.
33. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
34. Huck,L., Scherrer,A., Terzi,L., Johnson,A.E., Bernstein,H.D., Cusack,S., Weichenrieder,O. and Strub,K. (2004) Conserved tertiary base pairing ensures proper RNA folding and efficient assembly of the signal recognition particle Alu domain. *Nucleic Acids Res.*, **32**, 4915–4924.
35. Labuda,D. and Zitkiewicz,E. (1994) Evolution of secondary structure in the family of 7SL-like RNAs. *J. Mol. Evol.*, **39**, 506–518.
36. Sinnett,D., Richer,C., Deragon,J.M. and Labuda,D. (1991) Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units. *J. Biol. Chem.*, **266**, 8675–8678.
37. Ahl,V., Keller,H., Schmidt,S. and Weichenrieder,O. (2015) Retrotransposition and crystal structure of an Alu RNP in the ribosome-stalling conformation. *Mol. Cell*, **60**, 715–727.
38. Weichenrieder,O., Wild,K., Strub,K. and Cusack,S. (2000) Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature*, **408**, 167–173.
39. Lai,D., Proctor,J.R. and Meyer,I.M. (2013) On the importance of cotranscriptional RNA structure formation. *RNA*, **19**, 1461–1473.
40. Heilman-Miller,S.L. and Woodson,S.A. (2003) Effect of transcription on folding of the Tetrahymena ribozyme. *RNA*, **9**, 722–733.
41. Kapusta,A., Kronenberg,Z., Lynch,V.J., Zhuo,X., Ramsay,L., Bourque,G., Yandell,M. and Feschotte,C. (2013) Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.*, **9**, e1003470.
42. Deininger,P. (2011) Alu elements: know the SINEs. *Genome Biol.*, **12**, 236–236.
43. Chen,L. (2013) Characterization and comparison of human nuclear and cytosolic editomes. *Proc. Natl. Acad. Sci.*, **110**, E2741–E2747.
44. Sharma,P.M., Bowman,M., Madden,S.L., Rauscher,F.J. III and Sukumar,S. (1994) RNA editing in the Wilms' tumor susceptibility gene, WT1. *Genes Dev.*, **8**, 720–731.
45. Liu,Z., Song,W. and Dong,K. (2004) Persistent tetrodotoxin-sensitive sodium current resulting from U-to-C RNA editing of an insect sodium channel. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 11862–11867.
46. Prasanth,K.V., Prasanth,S.G., Xuan,Z., Hearn,S., Freier,S.M., Bennett,C.F., Zhang,M.Q. and Spector,D.L. (2005) Regulating gene expression through RNA nuclear retention. *Cell*, **123**, 249–263.
47. Clemson,C.M., Hutchinson,J.N., Sara,S.A., Ensminger,A.W., Fox,A.H., Chess,A. and Lawrence,J.B. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell*, **33**, 717–726.
48. Sasaki,Y.T., Ideue,T., Sano,M., Mituyama,T. and Hirose,T. (2009) MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 2525–2530.
49. Sunwoo,H., Dinger,M.E., Wilusz,J.E., Amaral,P.P., Mattick,J.S. and Spector,D.L. (2009) MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.*, **19**, 347–359.
50. Fox,A.H., Bond,C.S. and Lamond,A.I. (2005) P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol. Biol. Cell*, **16**, 5304–5315.
51. Cabili,M.N., Dunagin,M.C., McClanahan,P.D., Biaesch,A., Padovan-Merhar,O., Regev,A., Rinn,J.L. and Raj,A. (2015) Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.*, **16**, 20.
52. West,J.A., Davis,C.P., Sunwoo,H., Simon,M.D., Sadreyev,R.I., Wang,P.I., Tolstorukov,M.Y. and Kingston,R.E. (2014) The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell*, **55**, 791–802.