Alu RNA fold links splicing with signal recognition particle proteins

Ivana Borovská [1], Igor Vořechovský [2] and Jana Královičová [1,3,*]

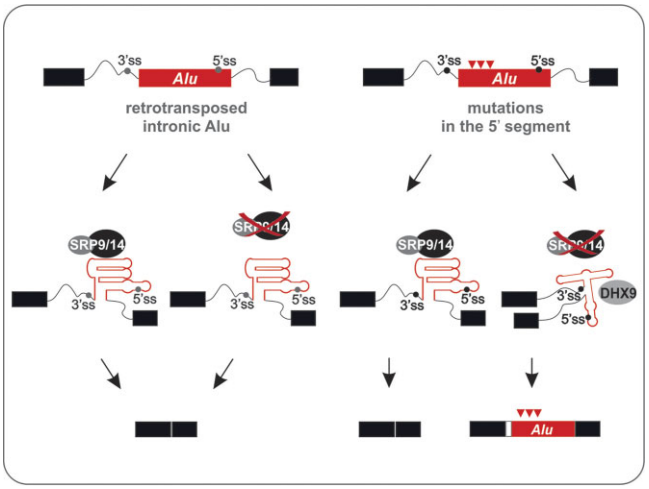
[1]Institute of Molecular Physiology and Genetics, Centre of Biosciences, Slovak Academy of Sciences, Bratislava840 05, Slovak Republic, [2]Faculty of Medicine, University of Southampton, HDH, MP808, Southampton SO16 6YD, United Kingdom and [3]Institute of Zoology, Slovak Academy of Sciences, Bratislava 845 06, Slovak Republic

Received March 16, 2023; Revised May 23, 2023; Editorial Decision May 24, 2023; Accepted May 31, 2023

ABSTRACT

Transcriptomic diversity in primates was considerably expanded by exonizations of intronic *Alu* elements. To better understand their cellular mechanisms we have used structure-based mutagenesis coupled with functional and proteomic assays to study the impact of successive primate mutations and their combinations on inclusion of a sense-oriented *AluJ* exon in the human *F8* gene. We show that the splicing outcome was better predicted by consecutive RNA conformation changes than by computationally derived splicing regulatory motifs. We also demonstrate an involvement of SRP9/14 (signal recognition particle) heterodimer in splicing regulation of *Alu*-derived exons. Nucleotide substitutions that accumulated during primate evolution relaxed the conserved left-arm *AluJ* structure including helix H1 and reduced the capacity of SRP9/14 to stabilize the closed *Alu* conformation. RNA secondary structure-constrained mutations that promoted open Y-shaped conformations of the *Alu* made the *Alu* exon inclusion reliant on DHX9. Finally, we identified additional SRP9/14 sensitive *Alu* exons and predicted their functional roles in the cell. Together, these results provide unique insights into architectural elements required for sense *Alu* exonization, identify conserved pre-mRNA structures involved in exon selection and point to a possible chaperone activity of SRP9/14 outside the mammalian signal recognition particle.

GRAPHICAL ABSTRACT



INTRODUCTION

*Alu*s are non-autonomous retrotransposons that occupy ∼11% of the human genome in more than one million copies (1,2). *Alu* elements evolved from the 7SL RNA, which encodes the RNA moiety of the signal recognition particle (SRP) (3), a cytoplasmic ribonucleoprotein (RNP) that interacts with the ribosome to control co-translational translocation of proteins into the endoplasmic reticulum (4). The elongation arrest activity of mammalian SRP depends both on the S domain, which binds SRP19, SRP54, and SRP68/72 proteins, and the *Alu* domain, which interacts with the SRP9/14 heterodimer (5,6). SRP9/14 binds with high specificity both to 7SL and *Alu* RNAs transcribed from various loci (7). This interaction supports formation of conserved structural elements in *Alu* RNA and the closed conformation of the SRP *Alu* domain (8–10), which enters the ribosomal translation elongation factor binding sites (11). The *Alu* consensus is about 300 nucleotides (nts) long and consists of left and right monomers separated by adenine stretches and terminating with a long poly(A) tail of variable length (12). The left arms harbour the A and

*To whom correspondence should be addressed. Tel: +421 2 3229 5538; Email: jana.kralovicova@savba.sk

B boxes of RNA polymerase III (Pol III) promoters (13), which permit transcription of new repeat units that can be inserted at new genomic locations by retrotransposition (14). The maximum amplification rate of *Alu* retrotranspositions was detected around 30 million years ago (15). During primate evolution *Alu*s accumulated numerous base mutations that serve as a basis for their classification into three main families, termed J, S and Y. The most ancient family is *Alu*J while the youngest family is *Alu*Y (16,17).

Although previously regarded as junk DNA, *Alu* repeats play a fundamental role in the regulation of gene expression (18). Under physiological conditions, *Alu* elements are epigenetically silenced (19–21) but their expression can be dramatically increased following various types of stress (22), virus infection (23), heat shock (24,25), and malignant transformation (26,27). *Alu* RNAs transcribed by Pol III can block transcription by binding RNA polymerase II (24). Dimeric *Alu* transcripts can be processed into more stable cytoplasmic *Alu*s (sc*Alu*) (28) that interfere with translation initiation and stress granules formation (29–31). The sc*Alu* consists of the left arm with a half-life of about 3 h (28,32) and maintains interactions with SRP9/14 (33,34). *Alu* RNAs transcribed by RNA polymerase II can influence mRNA nuclear export (35), affect translation (29), and induce ADAR-dependent RNA editing (36). *Alu*s embedded in the 3' untranslated regions (UTRs) can alter translation efficiency by base pairing sense and antisense sequences (37), activate Staufen 1-mediated mRNA decay (37), act as micro RNA targets (38), stimulate circular RNA biogenesis by backsplicing (39), and contribute to formation of long noncoding RNAs in vertebrates (40). Upon transcription intronic *Alu*s become important parts of pre-mRNAs and a source of new coding sequences in a process known as exonization (41,42). Over 90% of the youngest cassette exons in primates originated from repetitive sequences, with *Alu*s comprising about 62% of these events (43,44). Most of the newly born exons are generated via mutations that create splice sites (41,42,45) or splicing enhancer elements (46,47). Exonizations greatly enhance transcriptomic and proteomic diversity but can also lead to genetic disease (48). Activation of such cryptic exons may introduce premature stop codons and create frameshifts (43), yet low-inclusion exons can be tolerated. They are frequently associated with alternative splicing (AS) that allows them to be evolutionarily tested without compromising original proteomic repertoires (49,50). *Alu* exons are most often derived from antisense right arms, employing poly(U) signals in the pre-mRNA as polypyrimidine tracts (PPT), critical exon recognition motifs in vertebrates (42,43,47,48,51), and require only one or a few mutations for activation (41,42,47). The vast exonization potential of *Alu*s thus provides a useful model for studying how small exons are recognized by the cell in the sea of long introns, an unresolved problem in biology, and which *trans*-acting factors are required for *Alu* exon adaptation.

Recognition of exon-intron boundaries requires multiple pre-mRNAs motifs, including branch points, PPTs, 3' and 5' splice sites (3'ss, 5'ss) and auxiliary splicing signals known as exonic or intronic splicing enhancers (ESE, ISE) and silencers (ESS, ISS) (52,53). These *cis*-acting regulatory elements provide interaction platforms for small nuclear ribonucleoprotein particles (snRNPs) (54) and *trans*-acting factors, typically various families of RNA-binding proteins (55). Their accessibility is influenced by sequence context, regional variations in GC content, and by pre-mRNA structure and folding dynamics (56–58). As with splicing, pre-mRNA folding is largely co-transcriptional where the direction of transcription dictates the order of structure formation (59). RNA molecules can be trapped in low energy-inactive conformations and their conversion to functional structures may require specific RNA chaperones (60). Despite the development of transcriptome-wide RNA structural probing (61–64), it remains elusive how exactly pre-mRNA structures influence dynamic rearrangements of spliceosome assemblies and which *Alu* structural motifs are required for exon selection.

In this study, we have investigated molecular mechanisms that control a previously described disease-causing exonization of the sense-oriented left arm *Alu*J in intron 18 of the human *F8* gene (45,65). We have identified mutations in the conserved core of its 5' segment that accumulated during evolution and were essential for exon inclusion in mature transcripts. We found that the evolutionary pressure to conserve the RNA secondary structure was accompanied by a loss of SRP9/14-mediated closed conformation in exonization of the sense *Alu*. Our work also uncovered additional SRP9/14-sensitive *Alu* exons and a role for DEAD-box helicase DHX9 in this process.

MATERIALS AND METHODS

Plasmid preparations

The human wild-type *F8* reporter (F8wt) was prepared by cloning *F8* intron 18 with adjacent exons between HindIII/ApaI sites of pcDNA3.1/myc-His A (ThermoFisher) (Figure 1A and Supplementary Figure S1). The mutated reporter (F8*Alu*) was created by overlap extension PCR, introducing exon-activating mutation *F8* c.5998 + 530 C > T into F8wt. *F8* constructs representing main *Alu* families were prepared by replacing a 113-nt segment of F8*Alu* by *Alu*J, *Alu*S, and *Alu*Y sequences (*uppercase* in Figure 1A). The *Alu5′*con minigene was created by replacing 56 nts of the *Alu* segment of the F8*Alu* exon with a conserved sequence common to the consensus of three *Alu* families (*green box* in Figure 1A). The hybrid *PKP2* reporter was prepared by subcloning PCR amplicons containing *PKP2* exon 6 and portions of its native flanking introns into XhoI/XbaI sites of the *U2AF1* reporter (Figure 2E) (66). SRP14 cDNA was subcloned between BamHI/XbaI sites of pcDNA3.1/myc-His A (ThermoFisher) with the *myc* tag at the C terminus, employing the p14-9VN construct (Addgene cat. # 50930) as a template. Cloning PCR primers are in Supplementary Table S1. All reporters were propagated in *E. coli* DH5α. Plasmid DNA was isolated using the GeneJET Plasmid Miniprep Kit (ThermoFisher). All constructs were sequenced prior to transfections to exclude undesired mutations.

Cell cultures and transfections

Human embryonal kidney (HEK) 293 cells (DSMZ, cat.# ACC305) were grown under standard conditions in DMEM

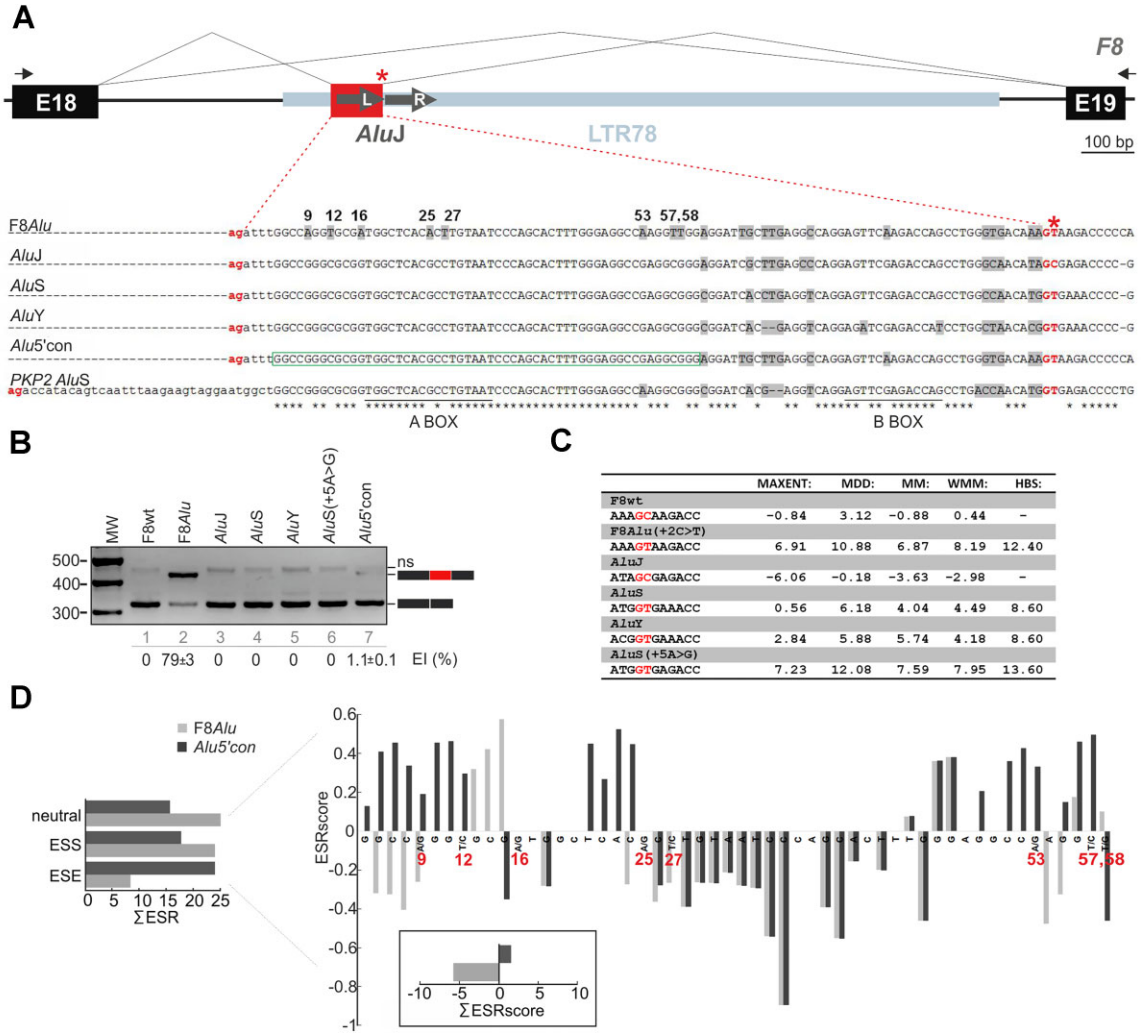


Figure 1. Evolutionary history of sense *Alu*J exonization in *F8*. (**A**) Schematics and sequences of minigene reporters. Introns are shown as horizontal lines, canonical exons as black boxes and the *Alu*J exon as a red box. Spliced products are denoted by hairlines above the transcript. Dark grey arrows represent left (L) and right (R) *Alu*J arms. Black arrows denote PCR primers. The lower panel shows alignment of the splicing-proficient F8*Alu* construct, which has mutation C > T (asterisk) that optimized the 5'ss (Supplementary Figure S1) and led to cryptic *Alu* exon activation and haemophilia A (45,65), with consensus sequences of main *Alu* families and with *PKP2* exon 6 derived from the left arm of sense *Alu*S (Figure 2E). *Alu*5'con, the splicing-deficient construct derived from F8*Alu* where the 5' segment of *Alu*J exon was replaced with the *Alu* consensus (*green box*). *Alu* sequences are in uppercase; the remaining exonic sequences are in lower case. The 3'ss and 5'ss are in red. The numbering starts from the first nucleotide of the F8*Alu* exon and is used consistently throughout the text and all figures. *F8Alu*-specific substitutions and mismatches between *Alu* families are highlighted in grey; two RNA Pol III promoters are underlined. (**B**) Splicing of *Alu* exons in wt and mutated minigenes; F8wt, construct with the GC 5'ss. *Alu*S(+5A > G), the *Alu*S construct that carries the 5'ss of *PKP2* exon 6 shown in panel C. Spliced RNA products are to the right; ns, nonspecific PCR product; MW, 100-bp size marker. (**C**) The intrinsic 5'ss strength of *Alu* exons from panels A and B. (**D**) Sums of ESR counts (left panel) and ESR scores (right panel) across the 5' segment of splicing-proficient F8*Alu* and splicing-deficient *Alu*5'con constructs. Higher values predict higher exon inclusion levels in mature transcripts.

supplemented with 10% (v/v) bovine calf serum (Biosera). Transfections were carried out in 12-well plates using 150 ng of reporter plasmids and jetPRIME (Polyplus) according to manufacturer's recommendations. Cells were harvested for RNA isolation 24 h after transfection. For depletion experiments, the cells were treated with small interfering RNAs (siRNA) to a final concentration of 80 nM (Supplementary Table S1). After 24 h, the cells were split into 12-well plates and transfected with the indicated plasmid reporters. For SRP9/14 rescue experiments the cells received the first hit with siRNA SRP14(1) at a final concentration of 80 nM and the second hit at a final concentration of 50 nM together with 50 ng of reporters and 100 ng of SRP14 plasmids. Cells were harvested for RNA and protein lysate preparations 24 h later.

RNA isolation and RT-PCR

Total RNA was isolated using TRI Reagent (Molecular Research Center) according to the manufacturer's protocol and treated with DNase I (Promega). Complementary DNA was synthetised with oligo d(T) primers using the Moloney Murine Leukaemia Virus Reverse Transcriptase (RT, Promega). RT-PCRs were carried out using a combination of gene- (F8F) and vector- (Pl4) specific primers, except for *PKP2* constructs where we used primers U2AF35e

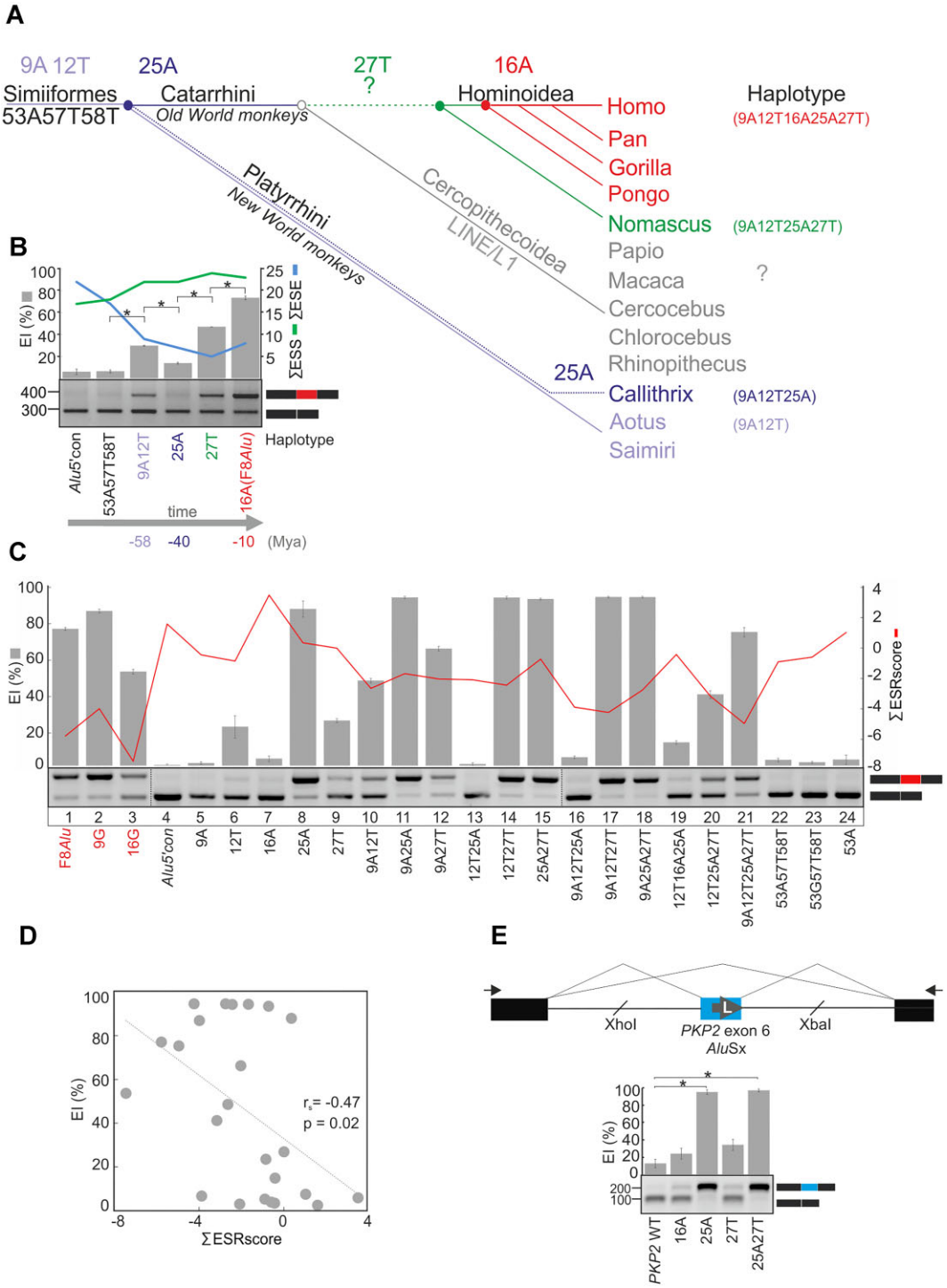


Figure 2. Phylogenetic changes in sense *Alu*J exon that promote or repress splicing. (**A**) *F8 Alu*J phylogeny. Genomic alignments of primate species are in Supplementary Figure S2. Lineage-specific haplotypes shown to the right were reconstructed in panel B plasmids.The exact assignment of substitution 27C > T was precluded by a LINE1 element retroposition in *Cercopithecoidea*. (**B**) *Alu* exon inclusion (EI) induced by *Alu*5'con substitutions created to mimick succession of phylogenetic changes. Apart from the indicated mutation(s), each minigene contains additional substitutions that accumulated during earlier primate evolution, reflecting lineage-specific haplotypes (coloured in panel A). Their approximate evolutionary span is shown at the bottom. Mya, million years ago. (**C**) RT-PCR of HEK293 cells transfected with reporters carrying various combinations of *F8 Alu* substitutions that accumulated during evolution. Mutations were introduced into F8*Alu* (red) and *Alu*5'con (black) minigenes. Red line denotes ∑ESRscores for exon positions 5 to 58, columns show corresponding EI levels (%). (**D**) Negative correlation between the ∑ESRscores and EI for constructs in panel C. (**E**) The impact of evolutionary changes in *F8Alu*J exon on exonization of *PKP2 Alu*S. Heterologous splicing *PKP2* reporter with an *Alu*S-derived exon is in the upper panel. Grey arrow represents the left arm of sense *Alu*S. The lower panel shows EI of mutated *PKP2* reporters with *F8Alu*-specific changes. Mutations are numbered as in panel A and Figure 1A.

and Pl4 (Supplementary Table S1). PCR products were separated on 1.5% agarose gels with addition of ethidium bromide (Promega) for visualisation. Signal intensities of the spliced products were measured using the Amersham Imager 600 (GE Healthcare). Amplifications of endogenous transcripts were carried out using primers shown in Supplementary Table S1.

Immunoblotting

Cells were washed with PBS and lysed in the RIPA buffer (ThermoFisher). Protein concentrations were determined by the Pierce BCA protein assay kit (ThermoFisher). Lysates were fractionated on 10% SDS-PAGE, transferred onto nitrocellulose membranes and incubated with antibodies against SRP9 (ProteinTech, 11195-1-AP), SRP14 (ProteinTech, 11528-1-AP), DHX9 (ProteinTech, 17721–1-AP), DHX36 (Abcam, ab70269), GAPDH (Novus, NB300-322) and secondary antibodies (Abcam, ab205718). Membranes were visualised using the Pierce ECL Western Blotting Substrate (ThermoFisher) according to the manufacturer's instructions. Chemiluminescent signals were measured with the Amersham Imager 600 (GE Healthcare).

RNA probe synthesis

RNA probes were produced with HiScribe™ T7 High Yield RNA Synthesis Kit (New England BioLabs) using PCR products as templates. Probe design was guided by the minimal *Alu* domain structure identified previously (8,9). Probe 5' sequences contained the first 66 nts of mutated *Alu* segments, followed by an invariable 3' stem terminating with a GUAA tetraloop. The amplicons were prepared with PCR primers in Supplementary Table S1 and the indicated minigene DNAs as templates. Forward primers included T7 promoter sequences. Synthetized RNAs were purified using TRI Reagent and resuspended in DNase/RNase free water. RNAs for structure mapping were 3′end-labeled using pCp-Cy5 (Jena Bioscience), and T4 RNA ligase (ThermoFisher). The labelling reaction was carried out at 4°C overnight.

Pulldown assays

30 μg of each *in vitro* transcribed RNA probe was oxidized with freshly prepared 5 mM Na-m-periodate solution dissolved in 0.1 M NaOAc, pH 5.0. Oxidation was carried out in the dark at room temperature for 1 h. Oxidized RNA was precipitated with ethanol, resuspended in 0.1 M NaOAc, pH 5.0, and denatured at 75°C for 3 min before coupling with beads. Adipic acid dihydrazide agarose beads (Sigma-Aldrich) were washed with the same solution and incubated with RNA overnight in the dark at 4°C. The next day beads were washed twice with 2 M NaCl and twice with buffer D (20 mM HEPES, pH 7.5, 0.2 mM EDTA, 100 mM KCl, 0.5 mM DTT, 6% v/v glycerol). RNA-beads complexes were incubated in a freshly prepared solution containing Hela nuclear extract (Ipracell), heparin (1 mg per reaction) and buffer D for 30 min at room temperature. Finally, the beads were washed five times with buffer D. Bound proteins were resolved on SDS-PAGE, transferred onto nitrocellulose membrane and incubated with the indicated antibodies.

RNA folding

Two microgram of probe RNAs were denatured at 95°C for 90 s, cooled on ice and folded in a buffer containing 20 mM TrisHCl, pH 8.0, 10 mM $MgCl_2$, 10 mM KCl, 200 mM NaCl and 8% v/v glycerol at 37°C for 45 min. RNA samples were loaded onto 6% native polyacrylamide gels (1× TBE) and separated (150 V) at 4°C. For denaturing gels, samples were run with a 95% formamide loading buffer on 6% gels containing 8 M urea. The gels were stained with ethidium bromide and visualized with the Amersham Imager 600 (GE Healthcare).

Preparation of recombinant SRP9/14

SRP9/14 expression plasmid was created by amplification of a coding segment of p14-9VN (Addgene plasmid # 50930) using primers SRP14F NcoI and SRP9R XhoI (Supplementary Table S1). The amplicon was inserted into NcoI/XhoI sites of pET-28a (Novagen). The hybrid protein was expressed in BL21 (DE3) pLysS Competent Cells (Promega). The cells were grown to an OD of 0.8 and protein expression was induced by 1 mM IPTG at 37°C for 3 h. Bacterial pellets were dissolved in buffer A (50 mM Tris–HCl, pH 8.0, 300 mM NaCl, 10% glycerol (v/v) and 3.6 mM β-mercaptoethanol) containing the cOmplete™, an EDTA-free Protease Inhibitor Cocktail (Roche) and were homogenized using SONOPULS GM Mini 20 (Bandelin Electronic). Proteins were purified using the Ni Sepharose 6 Fast Flow beads (GE Healthcare), washed five times with buffer A containing 20 mM imidazole and eluted with buffer A with 300 mM imidazole. The purified heterodimer was dialyzed against a storage buffer (10 mM Tris–HCl, pH 8.0, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 10% glycerol (v/v) and 1 mM β-mercaptoethanol) using the Slide-A-Lyzer™ G2 Dialysis Cassettes (ThermoFisher) at 4°C overnight and stored at –80°C.

RNA structure mapping

Structural probing was performed using 1.5 μg of Cy5-labeled probes and endonucleases RNase A or RNase T1, which cleave at single-stranded pyrimidines or guanines, respectively. RNAs were denatured at 95°C for 90 s, cooled on ice and incubated in the folding buffer at a final concentration of 100 mM KCl, 40 mM HEPES, pH 7.5, and 5 mM $MgCl_2$ at 37°C for 45 min. RNAs were mixed with 5 μg of yeast RNA and digested using conditions that were optimized to allow roughly a single cleavage per RNA molecule. Reactions containing 0.001 ng RNase A or 0.5 U RNase T1 were incubated at room temperature for 3 or 15 min, respectively. Cleaved RNAs were purified with TRI Reagent, resuspended in denaturing loading buffer containing Orange G, and fractionated on 8% gels with 8 M urea at 55 W for 3 h. Samples were run in parallel with a T1 marker prepared by a limited digestion of the denatured probe. Gels were imaged using Typhoon 9210 (GE Healthcare). For structural probing in the presence of SRP9/14, the RNAs were first incubated in the folding buffer to allow structure formation as described above. Subsequently, 16.5 μg of protein was added to ensure a molar excess of SRP9/14 over RNA and reaction mixtures were incubated at 37°C for 10 min.

The endonuclease was added after protein–RNA complex formation and products were analyzed as described above. Intensities of RNase T1-cleavages were measured using ImageQuant TL (GE Healthcare) and normalized to the intensity of nucleotide 71G located in the GUAA loop.

Bioinformatic and statistical analyses

Sequences of *Alu* families were obtained from Dfam (https://www.dfam.org; release 3.6). Sequence alignment of primate orthologs across of *F8* introns was created with Ensembl reference sequences (http://www.ensembl.org; accessed on 10 December 2021) using Clustal Omega (v. 1.2.4). RNA secondary structures were predicted by RNAfold 2.4.18 (http://rna.tbi.univie.ac.at) using default and alternative folding options. The intrinsic strength of cryptic splice sites was estimated by maximum entropy scores (67) and by H-bond scores designed to quantify the 5' ss complementarity to the U1 small nuclear RNA (68).

To determine ESE/ESS profiles of the indicated *Alu*s we employed ESEseq and ESSseq scores defined previously (69). ESEseq and ESSseq scores indicate the strength of hexamer motifs, with positive values for ESE and negative values for ESS (69). To predict exon inclusion of mutated constructs we calculated a sum of ESEseq and ESSseq scores ($\sum$ESRscores) for all hexamers that covered the analyzed segments (Dataset S1).

Endogenous transcripts for testing of *Alu* exons were selected from previously characterized *Alu* exons (45,70) that had a full-length sense-oriented left *Alu* arm expressed in HEK293 cells (71).

To compare exon inclusion levels of wt and mutated minigenes, we used one-way ANOVA with post hoc Tukey–Kramer tests. To compare inclusion levels of *Alu* exons in depletion experiments we used unpaired Student's *t*-test. Bar graphs reporting exon inclusion data show mean ± standard deviation (SD) from at least three independent replicates. Statistically significant changes relative to controls are shown as $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. The Spearman correlation coefficient (r_s) was computed with SigmaPlot, v.11.

RESULTS

Exonization potential of sense left-arm *Alu* in *F8*

We chose to study a left-arm sense *Alu* exon that was activated by a C > T mutation optimizing its 5'ss (Figure 1A). The exon employed a 3'ss/PPT derived from a more ancient long terminal repeat (45,65). *In vitro* structural probing suggested that the optimized GT 5'ss is more accessible than its wt GC counterpart, possibly improving the interaction with U1 snRNP components (45). However, a lack of correlation between the intrinsic 5'ss strength of exonized *Alu*s and their inclusion in mature transcripts pointed to the importance of 3'ss and/or cross-exon motifs (45). To identify them and to test whether the composite transposon can form a 'pre-exon' we created a minigene consisting of *F8* intron 18 and flanking exons (Figure 1A, *upper panel,* and Supplementary Figure S1). Transfection of HEK293 cells with the mutated *F8* construct (F8*Alu*) confirmed *Alu* exon activation (Figure 1B, lanes 1, 2). Surprisingly, replacements of the *Alu* left arm with ancestral sequences of main *Alu* families in the F8*Alu* construct (72) (Figure 1A, *lower panel*) revealed only transcripts lacking the *Alu* exon (Figure 1B, lanes 3–5), despite the presence of the C > T substitution in younger *Alu*S and *Alu*Y (Figure 1A). Because their decoy 5'ss differed from the optimal consensus (A/C)AG|GT(A/G)AGT (73), we examined transition A > G at intron position + 5, which strengthens the 5'ss of the *Alu*S construct (Figure 1C). Although the same 5'ss was used by an *Alu*S exon in *PKP2* (74), it still failed to activate the F8*Alu* exon (Figure 1B, lane 6), pointing to the importance of other substitutions in *Alu* exonization. Alignment of F8*Alu* and other *Alu* families revealed that their 5' half is almost invariant while the human F8*Alu* exon gained 8 point mutations during primate evolution (highlighted in Figure 1A, *lower panel*). To examine how these substitutions alter ESEs/ESSs and exon inclusion in mRNA, we replaced this F8*Alu* segment with the consensus sequence for main *Alu* families to create construct *Alu*5'con (*green box in* Figure 1A). The *Alu*5'con failed to activate the exon despite having a higher ESE density and higher $\sum$ESRscore than splicing proficient F8*Alu* construct (Figure 1B, lane 7, Figure 1D). Together, these results showed that the decoy 5'ss present in *Alu*J was *per se* insufficient for exon selection and that the *Alu*J exonization in *F8* required additional alterations in its 5' segment.

Evolution-driven exonization of *F8Alu* and ESEs/ESSs

Alignment of genomic sequences of *F8* in 15 primates revealed lineage-specificity of the 8 mutations (Figure 2A and Supplementary Figure S2). We therefore set out to compare their impact on *Alu* exonization in the context of primate haplotypes as opposed to individual substitutions. Transient transfections into HEK293 cells showed that the earliest *Alu* exon-activating mutations were 9A12T in *Simiiformes* (Figure 2B). The *Alu* exon inclusion was then suppressed by a more recent mutation 25A that arose before the split into Old and New World Monkeys and was restored again by a younger mutation 27T in *Cercopithecoidea*. The most recent mutation 16A, which completed the haplotype of the 5' segment of F8*Alu*, further enhanced *Alu* exon inclusion (Figure 2B).

We then extended this analysis to individual mutations and their double to quadruple combinations that were created on both F8*Alu* (Figure 2C, lanes 1–3) and *Alu*5'con (Figure 2C, lanes 4–24) backgrounds. Here, the most exon promoting effect among single substitutions was observed for 25A (Figure 2C, lane 8 versus 4), which contrasted with the exon repression when the same mutation was introduced into the *Simiiformes* haplotype 9A12T53A57T58T (Figure 2*cf. panels* B and C). Mutations 53A, 57T58T, 9A, and the *Hominidae*-specific mutation 16A had no effect (lanes 22–24, 5, 7). Surprisingly, exon inclusion levels of these constructs negatively correlated with $\sum$ESRscores calculated for each substitution in the *Alu*5'con segment or their combinations (Figure 2D).

To test the importance of the 5' segment for activation of exons derived from other sense *Alu*s, we selected an *AluS* exonization model in *PKP2* (Figure 1A, *lower panel*) (74). Examination of a heterologous *PKP2* reporter (Figure 2E) upon transfection revealed only marginal exon ac-

tivation despite a strong 5'ss (Figure 2E and Figure 1A, C). As splicing of this exon could be tissue-/developmental stage-specific, we amplified a panel of cDNAs from 18 human tissues, but we detected only transcripts lacking the *Alu*S exon (data not shown). In contrast, introducing the 5' segment-specific changes into the *PKP2* reporter, such as 16A and 27T, promoted *Alu*S exon. The *Alu*S exon was almost completely activated by mutations 25A and 25A27T (Figure 2E).

Together, we identified exonic *Alu* variants that arose during evolution and activated splicing of the sense *Alu* exons. Their succession in primate evolution did not gradually raise *Alu*J exon inclusion to the level observed in human *F8* but showed a context-dependent interplay of closely linked variants, involving neutral, additive and opposite haplotype effects (*cf.* Figure 2B and C). Finally, exon inclusion did not correlate with their predicted ESR profiles, implicating other regulatory factors that are distinct from simple ESE/ESS creation or abrogation.

Secondary structure of *Alu* RNA is the main determinant of *Alu* exon inclusion

*Alu*s are folded into a conserved three-way junction structure, preserving structural features of 7SL RNA (75). Positioning newly identified F8*Alu*-specific substitutions to the *Alu* secondary structure revealed that mutations at positions 9–27 were restricted to helix H1 while mutations at positions 53–58 affected helix H31 (Figure 3A). Helix mutations could explain the opposite splicing outcomes when present as solitary changes *vs.* their combinations that accumulated during primate evolution. Most notably, mutation 25G > A activated the *Alu*J exon in the presence of 12C (Figure 2C, lane 8) but repressed the same exon in the presence of 12T (Figure 2B, lane 4). As a solitary change, mutation 12C > T would impair Watson-Crick (WC) 12–25 base pairing which would be restored by mutation 25G > A in the course of primate evolution.

To test in more detail the impact of helix H1 stability on *Alu* exon inclusion, we first created constructs on the *Alu*5'con background that contained all nucleotide combinations at positions 12 and 25. Transient transfections revealed a gradual increase of exon inclusion with decreasing base pairing (Figure 3B). This limited initial sample size did not reveal significant correlation between exon inclusion and ∑ESRscores ($r_s = 0.24$, $P = 0.37$) or predicted thermodynamic stability of the 5' segment ($r_s = 0.43$, $P = 0.1$). Nevertheless, RNAs with the splice-supporting substitutions showed more variable motifs at loop L1 or at the central three-way junction (J12) (Figure 3B, *right panel*, and Supplementary Figure S3A). Next, we extended the analysis to more constructs with *F8*-specific mutations in helix H1 (Figures 2C and 3C, Supplementary Table S2 and Supplementary Figure S3B) and observed a weak correlation between exon inclusion and predicted thermodynamic stabilities of the 5' segment and a significant correlation with the number of hydrogen bonds in helix H1, but not with ∑ESRscores (Figure 3D). Finally, deletion of helix H1 from *Alu*5'con while leaving loop L1 in primary transcripts markedly increased *Alu* exon inclusion (Figure 3E).

These results strongly support a concept that specific RNA structural features play a major role in *Alu* exonization. The position and identity of complementary bases coupled with altered stability of helix H1 modulate splicing outcome more predictably than merely sequence-guided ESE/ESS profiles.

Splicing-proficient substitutions impair compact folding of *Alu* RNA

To examine how *Alu* exon-modifying substitutions in primates affect the *Alu*J RNA structure, we first prepared RNA probes that represented splicing-deficient (*Alu*5'con) and splicing-proficient (F8*Alu*) constructs (Figure 4A and Supplementary Figure S4). Both probes had identical migration on denaturing gels, however, in native conditions *Alu*5'con displayed a faster-running compact band while F8*Alu* was slower and showed a more diffused pattern (Figure 4B), consistent with a major change in RNA conformation. To explore which structural alterations account for different mobilities, we labeled 3' ends of each probe with Cy5 and treated them with single strand-specific endonucleases RNase T1 and RNase A. Nuclease probing with both RNAs showed positive signals at positions 71G and 72U, which map to the helix 32-closing GUAA tetraloop (Figure 4A, C). F8*Alu* also revealed RNase-accessible guanosines 11G, 13G and 15G of helix H1, 18G, 19G and 38G, which are involved in base pairing between loops L1 and L2 (8,10), and 29G at junction J12 where *Alu*5'con lacked corresponding cleavage products (Figure 4C, D).

To determine whether splicing-proficient substitutions at complementary positions 12 and 25 of helix H1 (Figure 3B) induce structural alterations similar to F8*Alu*, we subjected these transcripts to RNase T1 treatment. Mutations predicted to abrogate base pairing revealed cleavage products that signify a variable extent of helix H1 relaxation. In contrast, the cleavage products were absent in 7SL RNA that contained additional helix H1-stabilizing changes, except for 71G in the helix H32 closing loop (Supplementary Figures S4 and S5).

Taken together, these results demonstrate a substantial impact of lineage-specific and splice-supporting mutations on structural assemblies of the *Alu* exon. They also point to an essential role of base pair relaxation at complementary position 12–25 in the disruption of conserved *Alu* domain-like structure and in *Alu* exon recognition.

Signal recognition particle 9/14 heterodimer and helicase DHX9 regulate *Alu* exon splicing

High-resolution structural analysis of *Alu* RNA revealed that its closed conformation (Figure 4D) is stabilized by SRP9/14 heterodimer, which acts like a clamp (8,10). As conformational differences between *Alu* RNA variants are critical for SRP9/14 interaction (10,76) we tested whether the F8 *Alu* RNA can bind SRP9/14 using pull-down assays. Western blotting of protein fractions bound to *Alu*5'con and F8*Alu* revealed the presence of a strong SRP9/14 signal from the former probe but the absence from the latter (Figure 5A). Interestingly, the opposite was observed for the DEAD-box family helicase DHX9, which binds to

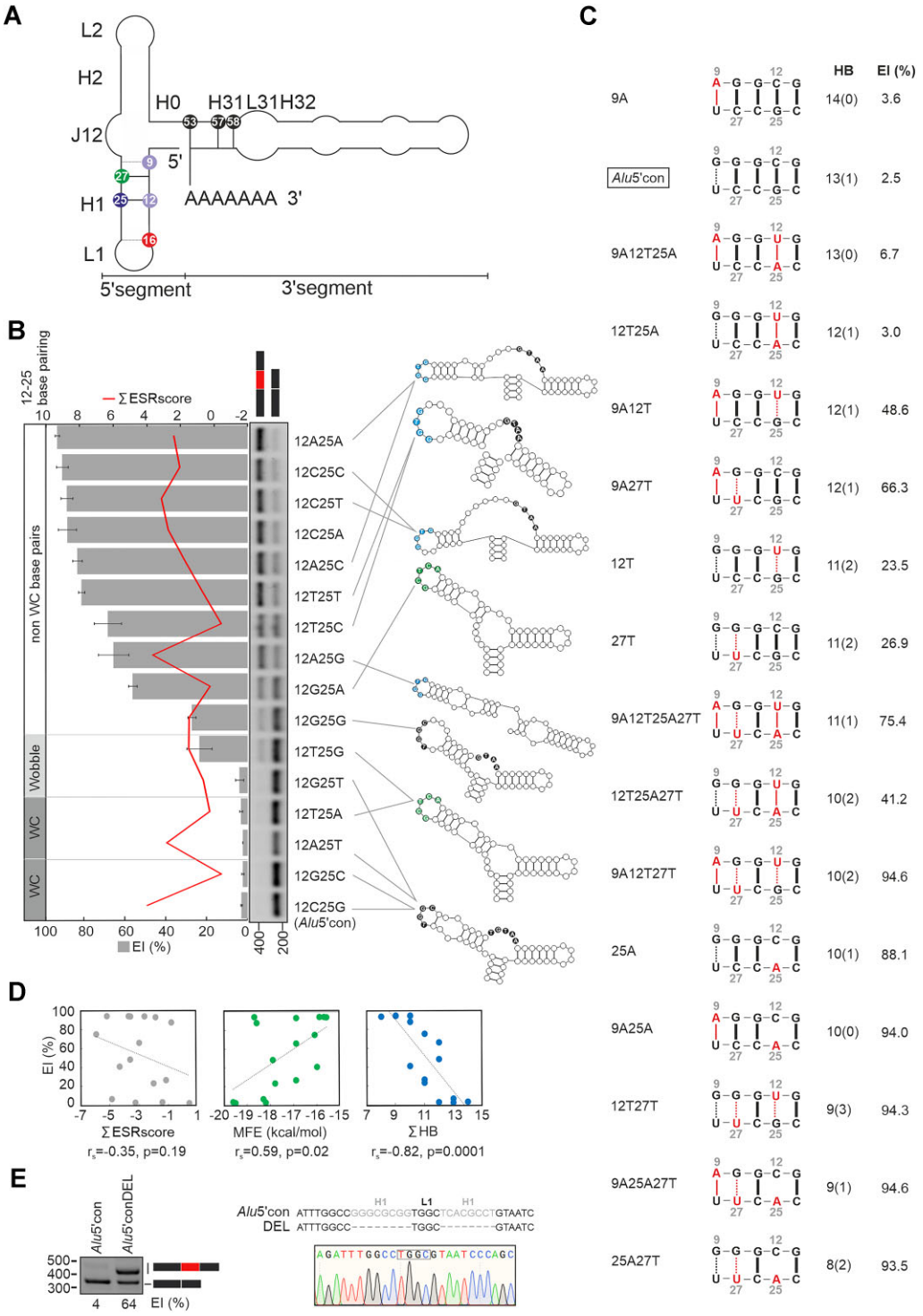


Figure 3. *Alu* exon activation is controlled by helix H1 stability. (**A**) RNA secondary structure of the left arm *Alu.* Nomenclature of helices, loops and junction is according to Ahl *et al.* (10). Human-specific substitutions in F8*Alu* are highlighted. (**B**) Destabilization of helix H1 coupled with a release of junction J12 from the structure tend to increase exon inclusion (EI). EI levels are ordered from low to high. Mutations at complementary positions 12–25 were introduced on the *Alu*5'con background. Predicted folding of the first 48 nts of mutated *Alu* segments is shown to the right (see also Supplementary Figure S3A). Variable motifs at loop L1 and junction J12 exposed in predicted RNA conformations are coloured. (**C**) Base pairing, the number of hydrogen bonds (HB) in helix H1 and EI. *F8*-specific mutations are in red. Number of wobble base pairs is in parenthesis. (**D**) Correlation between EI and ∑ESRscores, predicted thermodynamic stability, and HB for constructs in panel C (for more details, see Supplementary Figure S3B). MFE, minimum free energy. (**E**) Deletion of helix H1 activates *Alu*J exon inclusion in the splicing-deficient construct *Alu*5'con. Deleted nucleotides are shown in grey (*right panel*).

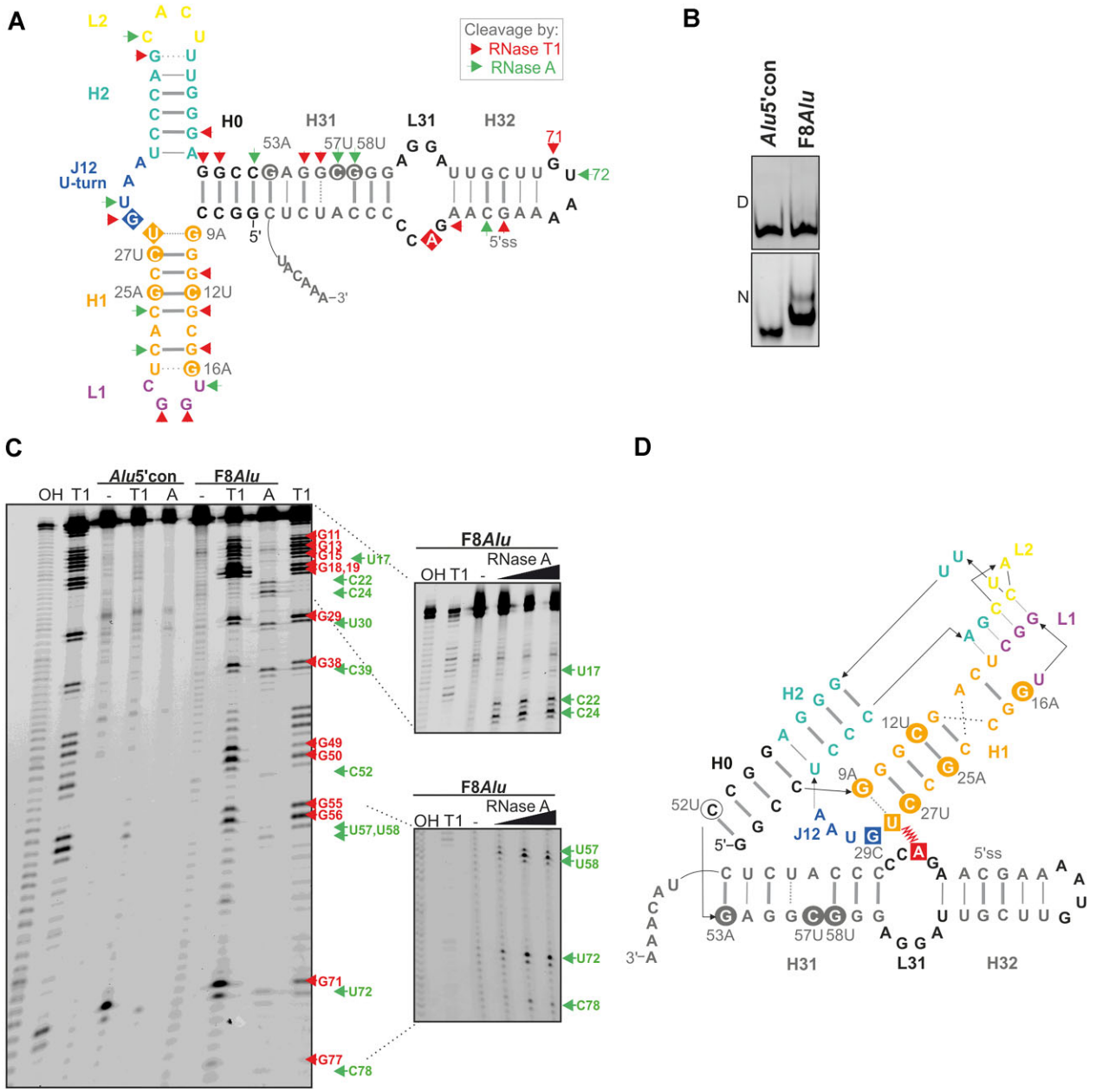


Figure 4. *F8*-specific substitutions induce major structural changes of the *Alu*5'con RNA. (**A**) RNA secondary structure of the *Alu*5'con probe. The model is based on folding of the minimal *Alu* domain proposed previously (8,9). Nomenclature of helices, loops and junction is as in Figure 3A. F8*Alu* mutations are circled; structurally important positions (10) are on a squared background. Solid lines represent canonical WC base pairs, dotted lines indicate wobble pairs. RNase T1 and RNase A cleavage sites identified in panel C are represented by red and green arrowheads, respectively. (**B**) Electrophoretic mobility of the indicated RNA probes on native (N) or denaturing (D) 6% gels. (**C**) Denaturing PAGE with Cy5-labeled *Alu*5'con and F8*Alu*. Probes were (mock)-digested with limiting amounts of RNase T1 and RNase A. Examples of dose-dependent cleavages (10^{-4}, 10^{-3} and 10^{-2} ng of RNase A per reaction) are shown in *right panels*. The cleavage products are numbered to the right. T1, OH, ladders generated by RNase T1 cleavage and NaOH treatment, respectively. (**D**) The *Alu*5'con probe (panel A) in closed conformation. The model is based on a complete *Alu* domain structure in a complex with SRP9/14 (8,10). Tertiary base pairs between loops L1 and L2 are shown as hairlines. Interaction of J12 with adenine in loop L31 is highlighted by a red zigzag line.

inverted-repeat *Alu* elements (77), and for DHX36 (Figure 5A). We detected both helicases by mass spectrometry with previously derived RNA probes (45) among proteins most enriched in the pull-down assay (Figure 5A and data not shown).

The association of *Alu* exon activation with a lack of SRP9/14 binding to the F8*Alu* probe prompted us to examine the effect of SRP9/14 and both helicases on F8*Alu* splicing. We transfected *Alu*5'con and F8*Alu* into HEK293 cells individually depleted of each protein (Figure 5B and Supplementary Figure S6). We also transfected mutated reporters F8-16G and 12C25C, which showed most pronounced changes in exon inclusion and RNA folding (Figures 2C, 3C, and Supplementary Figure S5). The reduced expression of SRP9/14 (SRP9/14-) increased exon usage, with the strongest increase observed for the F8-16G

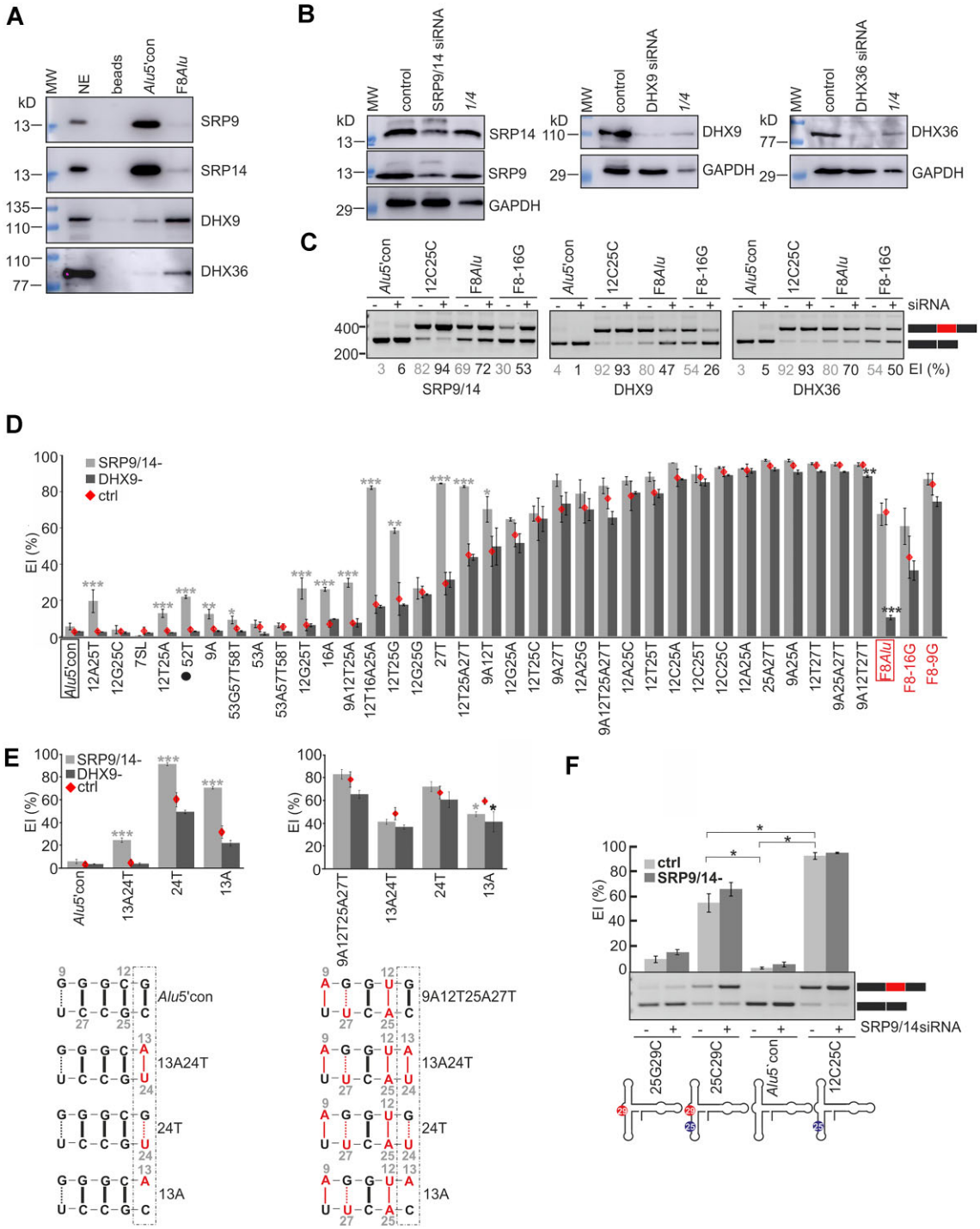


Figure 5. *Alu*J exonization is influenced by SRP9/14 and DHX9. (**A**) RNA pull-down assay with *Alu*5'con and F8A*lu* probes (Figure 4A and Supplementary Figure S4). Immunoblotting was carried out with antibodies to the right. NE, Hela nuclear extract; beads, RNA-free control. (**B**) Immunoblots of cell lysates from RNA interference-mediated depletion of SRP9/14, DHX9, and DHX36 in HEK293 cells. Antibodies are to the right. MW, size marker; control, scrambled siRNAs; 1/4, a quarter of the control lysate. (**C**) Exon inclusion (EI) for the indicated constructs in depleted (+) and control (–) HEK293 cells. Depleted proteins are at the bottom. (**D**) EI levels for a panel of 38 constructs in SRP9/14- and DHX9- cells. Constructs are ordered according to EI in mock-depleted cultures (red diamonds). Grey and black asterisks show statistically significant deviations in SRP9/14- and DHX9- cultures from controls, respectively. A black dot marks a variant in the *SDCCAG8 Alu*Y exon that was activated in SRP9/14- cells (*see also* Figure 8). Representative gels from these experiments are in Supplementary Figure S7. (**E**) Splicing analysis of the indicated constructs in SRP9/14- and DHX9- cells (*upper panels*). Tested base pairs at position 13–24 are marked by a rectangle. Nucleotides different from the *Alu*5'con sequence are in red *(lower panels*). (**F**) Substitution 29G > C in J12, previously shown to reduce SRP9/14 affinity (10,76), did not affect exon inclusion in SRP9/14- cells. +/–, SRP9/14-/mock-depleted HEK293 cells.

mutant (Figure 5C). Unlike SRP9/14-, depletion of helicase DHX9, induced exon skipping of F8*Alu*-derived minigenes but did not alter inclusion of *Alu*5'con or 12C25C exons. Finally, no inclusion changes were associated with diminished expression of DHX36 irrespective of the splicing reporter (Figure 5C).

To further assess how SRP9/14 and DHX9 influence splicing of exons with alterations in helices H1 and H31, we transfected a large set of *Alu*5'con- and F8*Alu*-derived mutants (Figures 2C and 3B) into SRP9/14- and DHX9- cells. Splicing analysis of 38 reporters revealed that SRP9/14 depletion affected exon inclusion of *Alu*5'con-based constructs with A-U/U-A or wobble G-U/U-G base pairs at position 12–25 and the G-U pair at position 10–27 (Figure 5D, Supplementary Figures S7 and S8). Contrary to SRP9/14- cells, depletion of DHX9 induced significant exon skipping only in F8*Alu*. This trend was also noticeable for F8-based reporters F8-16G and F8-9G with wobble pairs at the closing loop (16G-21U) and at the base of helix H1 (9G-28U), respectively. However, sensitivity to SRP9/14 was not determined solely by *F8*-specific changes. It was also supported by gradual base pair relaxation at position 13–24 created on the *Alu*5'con background (Figure 5E).

SRP9/14 makes contact with junction J12 of *Alu* RNA, which contains a core motif 28U29G30U (8,10) (Figure 4A). Because cytosine 29C decreased SRP9/14 affinity to *Alu* RNA (10,76,78), we created constructs 25G29C and 25C29C by introducing 29C in *Alu*5'con and splicing proficient 12C25C reporters. This mutation affected both splicing and RNA structure, but it did not modify responses to SRP9/14 depletion (Figure 5F and Supplementary Figure S5).

Together, these data show that SRP9/14 heterodimer and DHX9 helicase are involved in splicing regulation of *Alu* exon and that their effects can be modulated by changes in *Alu* structure, particularly by helix H1-stabilizing substitutions.

Tertiary contacts within *Alu* RNA fine-tune *Alu* exon recognition

The terminal loops L1 and L2 of the *Alu*'s 5' segment interact via tertiary base pairs. The structural motif in the 5' segment with AC dinucleotide in loop L31 stabilized the *Alu* RNA 3' segment in the closed conformation (Figure 4D). These contacts are guided by guanosine 6G, which fixes the backbone of 53G and 54A of helix H31, and are supported by guanosine 5G that pairs with 52C in helix H0 (8,10,79). To determine *Alu* exon regulation by the nucleotides involved in tertiary contacts, we created ancestral mutations in helix H31 (positions 53 and 57, 58) of F8*Alu* and F8-16G reporters. Whereas F8-16G reduced splicing (Figure 2C), presumably by supporting the loop-loop interaction (Figure 4D), the *F8*-specific 16A, predicted to pair with complementary 21U in loop L1 and to restrict loops interaction, promoted exon inclusion regardless of the helix H31 haplotype (Figure 6A, Structures 1–4 versus 5–8 Structures). The *F8*-specific 53A reduced exon inclusion (Structures 1 versus 3, 2 versus 4, 5 versus 7, and 6 versus 8), and combination of mutation 53A and the ancestral haplotype 16G57C58G made the *Alu* exon sensitive to SRP9/14 depletion (Structure 6). The role of compact structure in *Alu*-exon selection was also supported by enzymatic probing. RNAs derived from reporter 53G had a higher exon inclusion than F8*Alu* and showed increased accessibility of nucleotides G45-G47 in helix H2. Probes with ancestral nucleotides that induced exon skipping (16G57C58G and F8-16G) revealed high accessibility of helix H31 and loop L31 (Figure 6B and Supplementary Figure S5).

Collectively, these results showed that substitutions at positions involved in tertiary contacts within *Alu* RNA regulate *Alu* exon inclusion, influence its response to SRP9/14 and promote changes in RNA structure that predict splicing outcomes.

Modulation of *Alu* structure by SRP9/14 can influence splicing

To determine if SRP9/14 could regulate splicing by restoration of the closed RNA conformation disrupted by mutations, we used recombinant SRP heterodimers in enzymatic footprinting of RNAs derived from SRP9/14-sensitive (27T) and -resistant (F8*Alu*, 12C25C and 25C29C) reporters. As a control, we employed probe 3L1 29C based on *Alu* RNAs previously shown to reduce SRP9/14 affinity and impair *Alu* folding (80). The probe contained three changes in loop L1 and mutation 29G > C (Supplementary Figure S4). The footprinting in the presence of SRP9/14 lacked cleavages for probes 27T and 12C25C. The three remaining RNAs showed reduced accessibility of guanosines spanning positions 11G to 29G, presumably resulting from protection by protein binding (Figure 7A) (81). Probes 25C29C and 3L1 29C, predicted to alter SRP9/14 interaction, revealed enhanced sensitivity of guanosine 38G, which is involved in contacts between helices H1 and H2 (Figure 7A,B, and 4D) (8,10,79). Additionally, RNAs derived from the splicing proficient 25C29C and F8*Alu* showed enhanced accessibility of sequences for helices H0 and H1, respectively (Figure 7A, B).

Next, we employed the pull-down assay with RNA probes used in the footprint. Immunoblotting with SRP9/SRP14 antibodies showed significant differences in signal intensity between RNAs, which corresponded to the SRP9/14 sensitivity (Figures 5D and 6). The only exception was 12C25C which showed a strong SRP9/14 interaction (Figure 7C) despite high exon inclusion (Figures 3C and 5D). Finally, incubation with DHX9 antibody revealed binding of DHX9 not only to DHX9 sensitive F8*Alu* but also to 12C25C and 25C29C RNAs, which did not respond to DHX9 depletion (Figure 5D and Supplementary Figure S7).

Taken together, nuclease protection and RNA pull-down demonstrate that SRP9/14 binding and its capacity to induce RNA conformational changes are reflected in splicing of *Alu* exons in SRP9/14- cells. *Alu* exons of SRP9/14-sensitive constructs can be probably locked in a splice-disfavoring rearrangement through SRP9/14 binding (e.g. 27T) while exons with splice-supporting variants form an alternative and/or SRP9/14-resistant structures (e.g. 12C25C, 25C29C) or require the assistance of DHX9 (e.g. F8*Alu*) (Figure 7D).

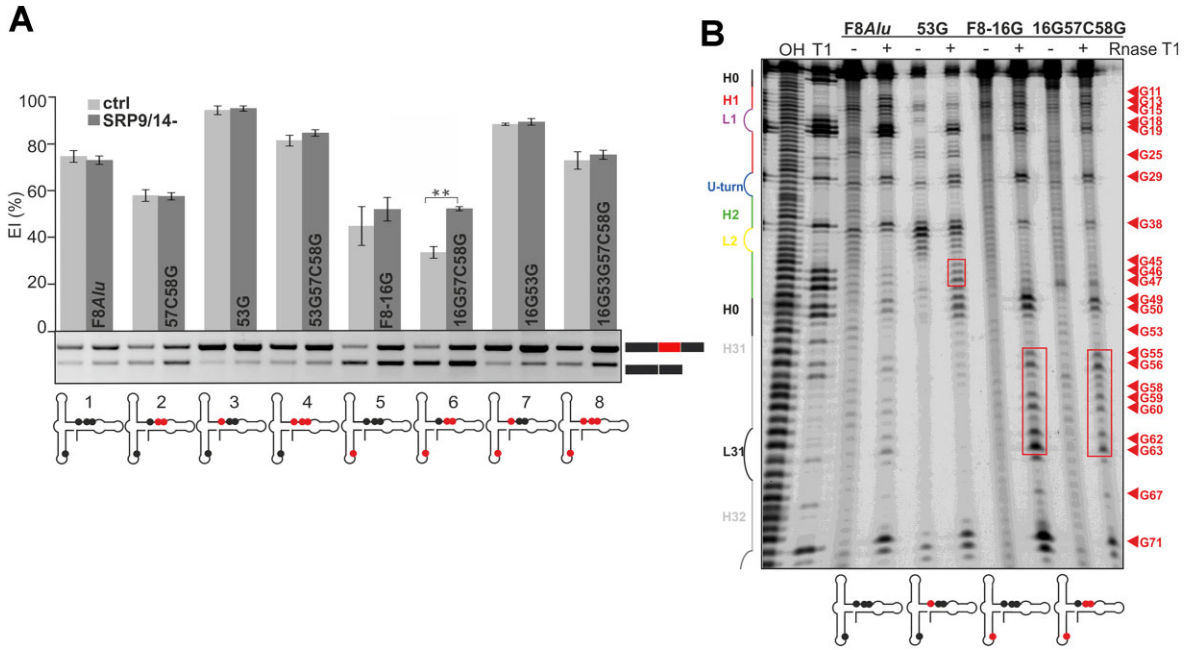


Figure 6. *Alu* exon inclusion is fine-tuned by substitutions at positions involved in tertiary interactions within the *Alu* RNA. (**A**) Exon inclusion (EI) for the indicated constructs in SRP9/14- cells. Mutated positions in the F8*Alu* structure are denoted by red dots. (**B**) RNase T1 cleavage products of Cy5-labeled RNA probes derived from minigenes shown in panel A. Red rectangles indicate guanosines that show differential nuclease sensitivity as compared to F8*Alu*. Red triangles represent cleavage sites.

Towards a universal mode of endogenous *Alu* exon recognition

To test the generality of the observed conformation-dependent *Alu*J exon usage, we examined splicing of other sense *Alu*-containing transcripts in SRP9/14- and helicases-depleted cells. First, we examined the wt *PKP2* construct (Figure 2E) and its mutated counterparts. As anticipated, SRP9/14 depletion activated exon inclusion of the wt and mutated *PKP2* reporters while depletion of DHX9 marginally induced exon skipping (Figure 8A, B). We observed no effect on *Alu*-exon splicing of *PKP2* in DHX36-cells (data not shown). This result suggested that the antagonistic effect of SRP9/14 and DHX9 on exon inclusion may not be limited to *F8 Alu* but could affect a wider range of endogenous *Alu* exons. We therefore selected a set of 15 human transcripts containing sense *Alu* left arms in exonic sequences (Supplementary Table S3) and examined their splicing upon SRP9/14, DHX9 and DHX36 depletion. Although six transcripts did not produce alternative splicing outcomes in HEK293 cells and were thus not informative, another 6 out of the remaining 9 transcripts (*SDCCAG8, ERCC1, CC2D2A, CAPN2, BIRC5* and *PKP2*) had SRP9/14 and/or DHX9 sensitive *Alu* exons (Figure 8C, D and Supplementary Figure S9). Both in SRP9/14- and DHX9- cells the most pronounced effect was observed for *Alu*Y in *SDCCAG8* intron 7 (Figure 8C, D and Supplementary Figure S9). *SDCCAG8* encodes a centrosome-associated protein and its deficiency was linked to skeleton, limbs, retina and kidney abnormalities (82,83). In SRP9/14- cells, activation of the *Alu*Y exon induced truncated transcripts with *Alu* in its terminal exon, and/or transcripts with an *Alu*-cassette exon introducing a stop codon. Moreover, inclusion of the *Alu*Y exon was associated with activation of cryptic exon 7b, previously reported in the Bardet-Biedl syndrome (Figure 8D) (84). The *SDCCAG8Alu* sequence differs from the invariable 5' segment of *Alu* families at positions 52T and 55A (Figure 8C) and ancestral 52C is involved in stabilization of the *Alu* closed conformation (Figure 4D). Substitution 52C > T at the *Alu*5'con plasmid induced *Alu* exon inclusion in SRP9/14- cultures (Figure 5D), consistent with activation of endogenous *SDCCAG8 Alu* exon upon SRP9/14 depletion (Figure 8D). Finally, the expression of isoform 202 of the excision repair cross-complementation group 1 protein (ERCC1) was suppressed in SRP9/14- cells at the expense of isoform 201, which has an alternative, *Alu*s containing 3'UTR (Figure 8D).

Taken together, we have identified additional *Alu*-derived exons regulated by SRP9/14 and/or DHX9 and modulated by sequence variants that have a universal impact on *Alu* exon selection.

DISCUSSION

In this study, we have characterized the exonization potential of sense-orientated *Alu*s from a structural perspective. Substitution C > T proposed to release decoy GC 5'ss from helix H32 (45) was sufficient for the *F8 Alu*J exonization, however, the corresponding substitution in evolutionary younger *Alu*S and *Alu*Y families (72) was not (Figure 1). This permitted identification of mutations that facilitated selection of *Alu* exons not only in *F8* but also in other transcripts (Figures 2E, 8D). These data imply that the left arm *Alu*s can function as an independent unit through ligand interactions and/or RNA structural rearrangements, similar

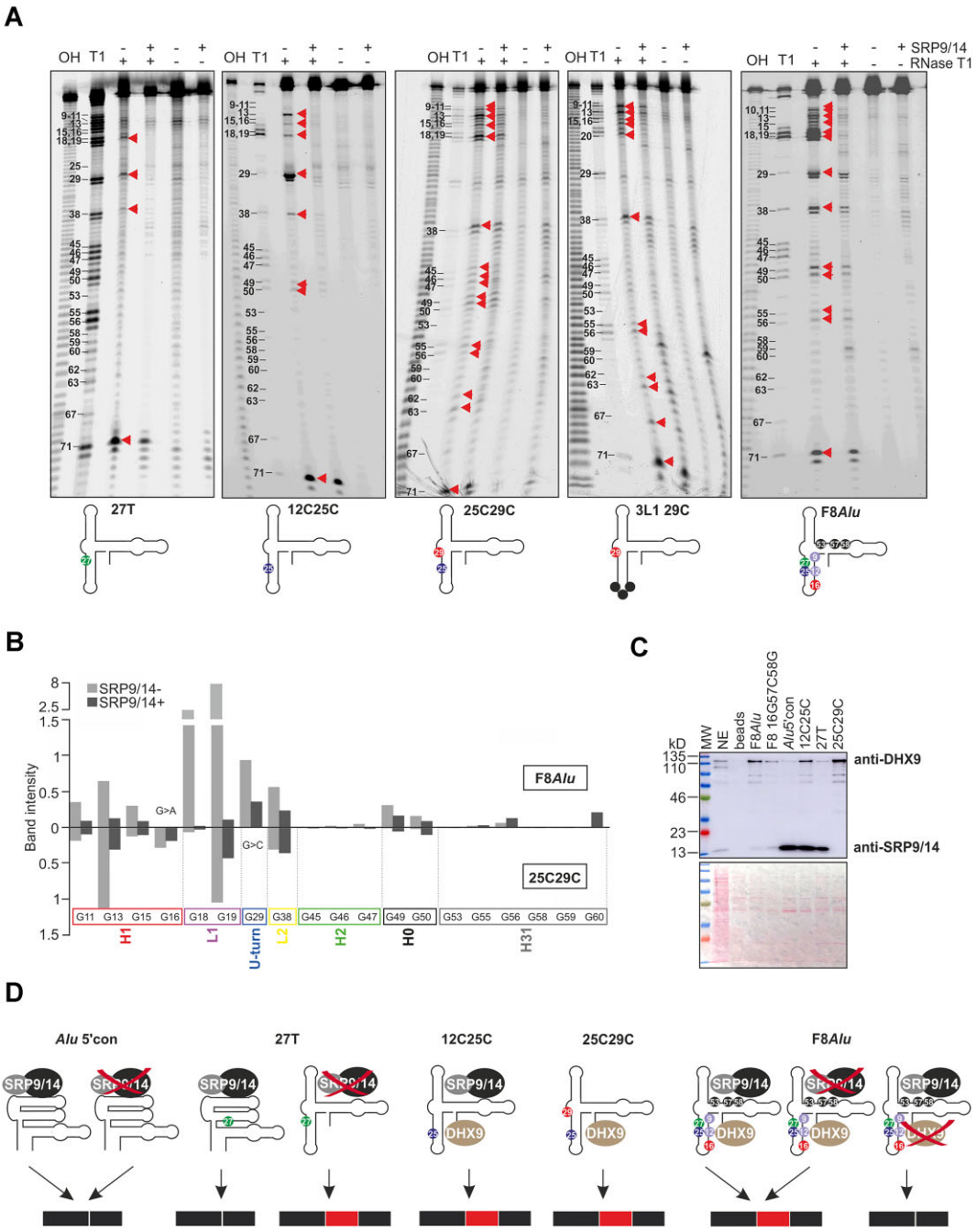


Figure 7. Restricted capacity of SRP9/14 to induce folding of splicing-proficient *Alu* exons. (**A**) RNase T1 cleavage of the indicated RNA probes. Their sequences are in Supplementary Figure S4. The reactions contained 1.5 μg of Cy5-labeled RNAs in the absence (–) and presence (+) of recombinant SRP9/14 (16.5 μg). The colour scheme for probe-specific substitutions at the bottom corresponds to Figure 3A. The marker was prepared using F8*Alu* RNA (panels 1, 3 and 4) or *Alu*5'con (panels 2 and 5) as a template. Some guanosines in helices H1, H31, and H32 of the *Alu*5'con marker have not been completely cleaved under denaturing conditions. (**B**) Cleavage-product intensities of nuclease-treated probes in the presence (+) and absence (–) of SRP9/14. Means are shown for F8*Alu* (*above*) and mutant 25C29C (*below horizontal axis*). 16G > A and 29G > C are probe-specific variants that do not permit cleavage by RNase T1. (**C**) RNA pull-down assay followed by immunoblotting with the indicated antibodies. NE, Hela nuclear extracts; beads, RNA-free control. The RNA probes (*top*) were used for structural analysis in Figures 7A, 4C and Supplementary Figure S5. (**D**) Models of *Alu* structure-guided activation or repression of the indicated *Alu* exons and proposed interactions between *Alu*, SRP9/14 and DHX9.

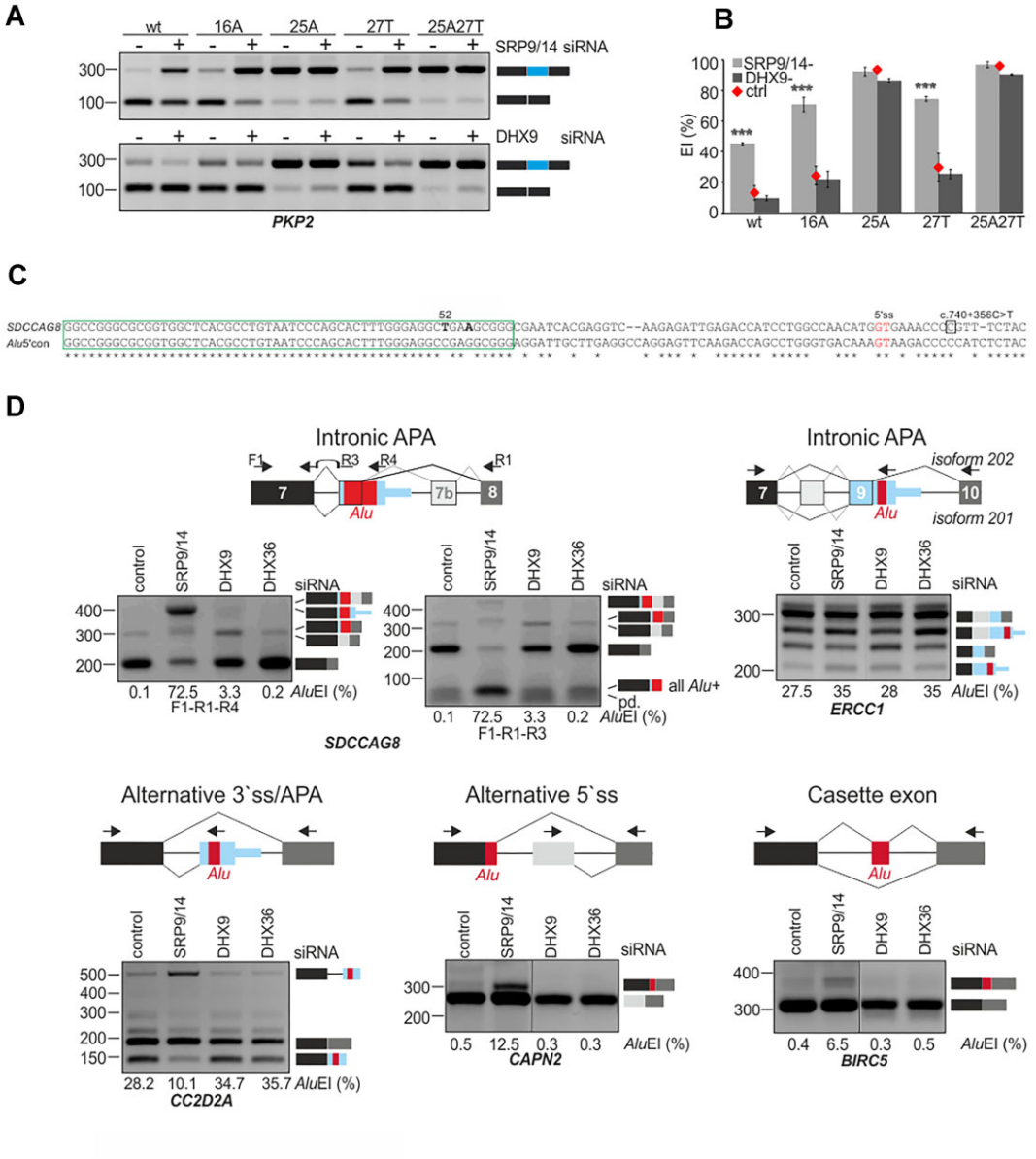


Figure 8. Identification of endogenous *Alu*-derived exons sensitive to SRP9/14 depletion**.** (**A**) *Alu* exon responses of hybrid *PKP2* minigenes to SRP9/14 depletion are determined by the same substitutions as in *Alu*5'con reporters. (**B**) Exon inclusion levels for transcripts in panel A. (**C**) Alignment of *Alu*Y in intron 7 of the *SDCCAG8* gene and *Alu*5'con. c.740 + 356C > T, a mutation associated with *Alu*-exon activation in *SDCCAG8* (84). (**D**) Splicing patterns of endogenous transcripts in cells depleted of proteins indicated at the top. Control, scrambled siRNAs. Sense-oriented *Alu*s embedded in tested exons are in red. RNA products are schematically shown to the right. Amplification primers (*arrows*) are in Supplementary Table S1. *Alu*EI (%), the mean relative abundance of transcripts containing *Alu* exon calculated from two experiments. APA, alternative polyadenylation. pd, primer dimers.

to a stem-loop structure derived from an exonic mammalian interspersed repeat in the *FGB* gene (85).

Up to ∼20% of *Alu* sequences are made of CpG residues, common methylation sites involved in silencing of *Alu* retrotransposition activities (86). Retrotransposition can also be reduced by weakening internal Pol III promoter (87), decreasing SRP9/14 interaction and by *Alu* structural changes that disrupt their ribosome-binding conformation (10,76). Five out of the 8 *Alu* substitutions in *F8* altered CpGs, with two of them in box A of Pol III (Figure 1A) (88), which may reflect increased mutation rates and stronger purifying selection within CpGs as compared to control sequences (89,90).

Evolution-driven splicing activation of the *F8Alu*J exon was not associated with a linear increase in exon inclusion in the primate lineage, and was coupled with a loss of ESEs and gain of ESSs (Figure 2B), arguing against a gradual loss of inhibitory motifs (91,92) and supporting the importance of structural determinants of exon inclusion. Restoration of secondary structures and WC base-pairing may occur via compensatory mutations and may involve an intermediate GU wobble base pairing (93–95), as observed for

positions 12–25 (Figures 3 and 4A). For example, the ancestral 12C25G WC pair in helix H1 could be substituted with F8*Alu*-specific 12T25A to maintain exon skipping (Figures 2C and 3C). By contrast, the replacement of 12C25G with the wobble pair showed *Alu* exon activation as well as repression (Figure 3B). We speculate that this difference can be due to their non-isostericity (96,97) and/or their position within the helix. Similarly, the noncanonical base geometry of the purine pair GG (98) limited exon inclusion, in contrast to exon activation by GA or AG pairs (Figure 3B), which can be positioned in multiple ways through their hydrogen-bonding capabilities (99).

The positional and context-dependent impact of primate *Alu* exon substitutions were also seen for closely linked mutations 9A and 12T (Figure 2C). Substitution 9A replaces the wobble pair with a WC pair at the base of helix H1 while mutation 12T replaces a WC pair with a wobble pair (Figures 3B, C and 4). The former, but not the latter wobble pair, is involved in fixation of the closed conformation (Figure 4D) (10). This may explain additive effects on exon inclusion for double mutation 9A12T (Figure 2C).

Proper formation of helix H1 and interaction between loops L1 and L2 are critical for the *Alu* closed conformation, which is further stabilized by SRP9/14 binding (8,10,79,80). We have shown that the effect of siRNA-mediated depletion of SRP9/14 on splicing of *Alu* exons was mutation-dependent (Figure 5). The SRP9/14-sensitive substitutions may alter loop-loop interactions (mutation 16A) and helix H1 stability (mutations 27T, 12T25A, 9A) (Figure 5D). The sensitivity of *Alu*-exon to SRP9/14 was also associated with variants in helix H31 and with the *SDCCAG8*-specific mutation 52C > T (Figures 5D, 6A and 8C, D), which is at a position important for stabilization of the central three-way junction and the closed conformation (Figure 4D) (10). The endonuclease resistance of *Alu*5`con RNA sharply contrasted with F8*Alu* and RNAs derived from other splicing-proficient minigenes, which revealed cleavages at positions involved in tertiary contacts (Figure 4C and 6B and Supplementary Figure S5). In the presence of SRP9/14, RNAs derived from the SRP9/14-sensitive reporter 27T showed complete nuclease protection whereas probes F8*Alu* and 25C29C of SRP9/14-resistant exons did not (Figure 7A, B). Therefore, they may not adopt a conserved *Alu* structure and resemble an open conformation, such as described for the *P. falciparum* SRP *Alu* domain, which remains unaltered upon the SRP9/14 binding (100). Thus, SRP9/14 directs folding of not only 7SL RNA progeny (80) but may act as an RNA chaperone of an *Alu* embedded in a pre-mRNA and regulate *Alu* exon splicing. Speculatively, the conserved structure of *Alu* exon could interact with snRNPs in a mode analogous to the SRP RNA and the sarcin-ricin loop of large rRNA (11). Interaction of *Alu* RNA with snRNPs may compete with binding of snRNP-specific proteins, similar to competition of the SRP *Alu* domain with elongation factors for binding sites at ribosomes (10,11).

DHX9 helicase was identified in human prespliceosomes (101) and was implicated in mRNA/pre-mRNA binding and coordination of RNA editing and splicing (102,103). DHX9 resolves double-stranded RNA formed by *Alu* repeats during transcription (77). Binding of DHX9 only to a subset of probes in our RNA pull-downs (Figure 7C) point to the *Alu* RNA structure-shaped recognition, reminiscent of DHX9 interaction with the primer binding site segment of HIV-1 RNA (104). It is possible that the folding landscape of the DHX9-sensitive pre-mRNAs may need its chaperone activity to ensure splicing-proficient RNA remodelling, which is supported by a distinct migration pattern of F8*Alu* RNA in native conditions (Figure 4B and Supplementary Figure S5B), and diminished splicing of F8*Alu* in DHX9- cultures (Figure 5C, D). Under physiological conditions, DHX9 could also compete with SRP9/14 binding and RNA-protein assembly dynamics (Figure 7D). However, this scenario does not exclude a possibility that DHX9 interaction with *Alu* exons may also stimulate structural changes in adjacent RNA segments that help recruit other proteins to RNA-protein complexes.

Our search for other *Alu* exons regulated by SRP9/14 revealed six additional transcripts (Figure 8). Most of their *Alu*s were embedded in UTRs, in agreement with a more abundant *Alu* exonization within UTRs than in coding regions (43). This suggests that SRP9/14 may play a role additional to the co-translation translocation, such as *Alu* transcript metabolism and/or splicing. While much of the *Alu* expression is autonomous (105) and under tight epigenetic control (106), high levels of *Alu* RNAs are a common feature of cellular responses to different types of stress, such as viral infection (107–109). Moreover, many families of transposed elements can be upregulated in cancer, with more than half possibly resulting from a loss of DNA methylation (110). The transient increase of *Alu* RNA could therefore sequestrate SRP9/14, which is present in 20-fold excess over SRP in primate cells (33), and modulate the expression of *Alu*-containing transcripts. The concept of the fine-tune regulation of *Alu* exons by SRP9/14-supported conformational changes can be extened to a variety of cellular processes that involve *Alu* RNA, such as translation inhibition (30), stress response (31) and modulation of immune responses by *Alu* Z-flipons (111).

Altogether, our results show that RNA conformation changes rather than an ESE/ESS evolution determine splicing outcomes of the sense *Alu* exons. Secondary structure-constrained nucleotide substitutions that accumulated in *Alu* helix H1 during primate evolution and promoted exon usage altered the conserved *Alu* conformation, reduced its binding by SRP9/14 heterodimer and increased its sensitivity to DHX9. We have also demonstrated the involvement of SRP9/14 heterodimer in the splicing regulation of a number of endogenous *Alu*-containing transcripts. Finally, these results highlight novel aspects of the promiscuous function of SRP proteins outside the mammalian signal recognition particles, which is reminiscent of chaperone activities of ribosomal proteins (112,113).

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We wish to thank Katarína Vondráškova for technical assistance and to Peter Barath (Institute of Chemistry, SAS) and members of his group for mass spectrometry analysis.

FUNDING

VEGA [2/0016/22 to J.K.]; Slovak Research and Development Agency [APVV-18–0096 to J.K.]. Funding for open access charge: Slovak Research and Development Agency [APVV-18-0096]; Vedecká Grantová Agentúra MŠVVaŠ SR a SAV [2/0016/22].

Conflict of interest statement. None declared.

REFERENCES

1. Grover,D., Kannan,K., Brahmachari,S.K. and Mukerji,M. (2005) ALU-ring elements in the primate genomes. *Genetica*, **124**, 273–289.
2. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
3. Ullu,E. and Tschudi,C. (1984) Alu sequences are processed 7SL RNA genes. *Nature*, **312**, 171–172.
4. Walter,P. and Blobel,G. (1982) Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature*, **299**, 691–698.
5. Lakkaraju,A.K., Mary,C., Scherrer,A., Johnson,A.E. and Strub,K. (2008) SRP keeps polypeptides translocation-competent by slowing translation to match limiting ER-targeting sites. *Cell*, **133**, 440–451.
6. Akopian,D., Shen,K., Zhang,X. and Shan,S.O. (2013) Signal recognition particle: an essential protein-targeting machine. *Annu. Rev. Biochem.*, **82**, 693–721.
7. Bovia,F., Wolff,N., Ryser,S. and Strub,K. (1997) The SRP9/14 subunit of the human signal recognition particle binds to a variety of Alu-like RNAs and with higher affinity than its mouse homolog. *Nucleic Acids Res.*, **25**, 318–326.
8. 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.
9. Weichenrieder,O., Kapp,U., Cusack,S. and Strub,K. (1997) Identification of a minimal Alu RNA folding domain that specifically binds SRP9/14. *RNA*, **3**, 1262–1274.
10. 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.
11. Halic,M., Becker,T., Pool,M.R., Spahn,C.M., Grassucci,R.A., Frank,J. and Beckmann,R. (2004) Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature*, **427**, 808–814.
12. Jelinek,W.R. and Schmid,C.W. (1982) Repetitive sequences in eukaryotic DNA and their expression. *Annu. Rev. Biochem.*, **51**, 813–844.
13. Perez-Stable,C. and Shen,C.K. (1986) Competitive and cooperative functioning of the anterior and posterior promoter elements of an Alu family repeat. *Mol. Cell. Biol.*, **6**, 2041–2052.
14. Rogers,J. (1983) Retroposons defined. *Nature*, **301**, 460.
15. Britten,R.J. (1994) Evidence that most human Alu sequences were inserted in a process that ceased about 30 million years ago. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 6148–6150.
16. Jurka,J. and Milosavljevic,A. (1991) Reconstruction and analysis of human Alu genes. *J. Mol. Evol.*, **32**, 105–121.
17. Batzer,M.A., Deininger,P.L., Hellmann-Blumberg,U., Jurka,J., Labuda,D., Rubin,C.M., Schmid,C.W., Zietkiewicz,E. and Zuckerkandl,E. (1996) Standardized nomenclature for Alu repeats. *J. Mol. Evol.*, **42**, 3–6.
18. Gussakovsky,D. and McKenna,S.A. (2021) Alu RNA and their roles in human disease states. *RNA Biol*, **18**, 574–585.
19. Schmid,C.W. (1991) Human Alu subfamilies and their methylation revealed by blot hybridization. *Nucleic Acids Res.*, **19**, 5613–5617.
20. Varshney,D., Vavrova-Anderson,J., Oler,A.J., Cowling,V.H., Cairns,B.R. and White,R.J. (2015) SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. *Nat. Commun.*, **6**, 6569.
21. Jordà,M., Díez-Villanueva,A., Mallona,I., Martín,B., Lois,S., Barrera,V., Esteller,M., Vavouri,T. and Peinado,M.A. (2017) The epigenetic landscape of Alu repeats delineates the structural and functional genomic architecture of colon cancer cells. *Genome Res*, **27**, 118–132.
22. Hwang,Y.E., Baek,Y.M., Baek,A. and Kim,D.E. (2019) Oxidative stress causes Alu RNA accumulation via PIWIL4 sequestration into stress granules. *BMB Rep*, **52**, 196–201.
23. Panning,B. and Smiley,J.R. (1993) Activation of RNA polymerase III transcription of human Alu repetitive elements by adenovirus type 5: requirement for the E1b 58-kilodalton protein and the products of E4 open reading frames 3 and 6. *Mol. Cell. Biol.*, **13**, 3231–3244.
24. Mariner,P.D., Walters,R.D., Espinoza,C.A., Drullinger,L.F., Wagner,S.D., Kugel,J.F. and Goodrich,J.A. (2008) Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell*, **29**, 499–509.
25. Liu,W.M., Chu,W.M., Choudary,P.V. and Schmid,C.W. (1995) Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.*, **23**, 1758–1765.
26. Tang,R.B., Wang,H.Y., Lu,H.Y., Xiong,J., Li,H.H., Qiu,X.H. and Liu,H.Q. (2005) Increased level of polymerase III transcribed Alu RNA in hepatocellular carcinoma tissue. *Mol. Carcinog.*, **42**, 93–96.
27. Zhou,S. and Van Bortle,K. (2023) The Pol III transcriptome: basic features, recurrent patterns, and emerging roles in cancer. *Wiley Interdiscip. Rev. RNA*, e1782.
28. Maraia,R.J., Driscoll,C.T., Bilyeu,T., Hsu,K. and Darlington,G.J. (1993) Multiple dispersed loci produce small cytoplasmic Alu RNA. *Mol. Cell. Biol.*, **13**, 4233–4241.
29. Häsler,J. and Strub,K. (2006) Alu RNP and Alu RNA regulate translation initiation in vitro. *Nucleic Acids Res.*, **34**, 2374–2385.
30. Ivanova,E., Berger,A., Scherrer,A., Alkalaeva,E. and 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.
31. Berger,A., Ivanova,E., Gareau,C., Scherrer,A., Mazroui,R. and 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**, 11203–11217.
32. Chu,W.M., Liu,W.M. and Schmid,C.W. (1995) RNA polymerase III promoter and terminator elements affect Alu RNA expression. *Nucleic Acids Res.*, **23**, 1750–1757.
33. Bovia,F., Fornallaz,M., Leffers,H. and Strub,K. (1995) The SRP9/14 subunit of the signal recognition particle (SRP) is present in more than 20-fold excess over SRP in primate cells and exists primarily free but also in complex with small cytoplasmic Alu RNAs. *Mol. Biol. Cell*, **6**, 471–484.
34. Hsu,K., Chang,D.Y. and Maraia,R.J. (1995) Human signal recognition particle (SRP) Alu-associated protein also binds Alu interspersed repeat sequence RNAs. Characterization of human SRP9. *J. Biol. Chem.*, **270**, 10179–10186.
35. Elbarbary,R.A. and Maquat,L.E. (2015) CARMing down the SINEs of anarchy: two paths to freedom from paraspeckle detention. *Genes Dev*, **29**, 687–689.
36. Ramaswami,G., Lin,W., Piskol,R., Tan,M.H., Davis,C. and Li,J.B. (2012) Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods*, **9**, 579–581.
37. Gong,C. and Maquat,L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*, **470**, 284–288.
38. Spengler,R.M., Oakley,C.K. and Davidson,B.L. (2014) Functional microRNAs and target sites are created by lineage-specific transposition. *Hum. Mol. Genet.*, **23**, 1783–1793.
39. Zhang,X.O., Wang,H.B., Zhang,Y., Lu,X., Chen,L.L. and Yang,L. (2014) Complementary sequence-mediated exon circularization. *Cell*, **159**, 134–147.
40. 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.
41. Lev-Maor,G., Sorek,R., Shomron,N. and Ast,G. (2003) The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science*, **300**, 1288–1291.
42. Sorek,R., Lev-Maor,G., Reznik,M., Dagan,T., Belinky,F., Graur,D. and Ast,G. (2004) Minimal conditions for exonization of intronic sequences: 5' splice site formation in alu exons. *Mol. Cell*, **14**, 221–231.
43. Sela,N., Mersch,B., Gal-Mark,N., Lev-Maor,G., Hotz-Wagenblatt,A. and 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.
44. Zhang,X.H. and Chasin,L.A. (2006) Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 13427–13432.
45. Alvarez,M.E.V., Chivers,M., Borovska,I., Monger,S., Giannoulatou,E., Kralovicova,J. and Vorechovsky,I. (2021) Transposon clusters as substrates for aberrant splice-site activation. *RNA Biol*, **18**, 354–367.
46. Lei,H., Day,I.N. and Vorechovský,I. (2005) Exonization of AluYa5 in the human ACE gene requires mutations in both 3' and 5' splice sites and is facilitated by a conserved splicing enhancer. *Nucleic Acids Res.*, **33**, 3897–3906.
47. Lei,H. and Vorechovsky,I. (2005) Identification of splicing silencers and enhancers in sense Alus: a role for pseudoacceptors in splice site repression. *Mol. Cell. Biol.*, **25**, 6912–6920.
48. Vorechovsky,I. (2010) Transposable elements in disease-associated cryptic exons. *Hum. Genet.*, **127**, 135–154.
49. Sorek,R., Ast,G. and Graur,D. (2002) Alu-containing exons are alternatively spliced. *Genome Res*, **12**, 1060–1067.
50. Alekseyenko,A.V., Kim,N. and Lee,C.J. (2007) Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA*, **13**, 661–670.
51. Ram,O., Schwartz,S. and Ast,G. (2008) Multifactorial interplay controls the splicing profile of Alu-derived exons. *Mol. Cell. Biol.*, **28**, 3513–3525.
52. Wang,Y., Liu,J., Huang,B.O., Xu,Y.M., Li,J., Huang,L.F., Lin,J., Zhang,J., Min,Q.H., Yang,W.M. *et al.* (2015) Mechanism of alternative splicing and its regulation. *Biomed Rep.*, **3**, 152–158.
53. Chen,S.Y., Li,C., Jia,X. and Lai,S.J. (2019) Sequence and evolutionary features for the alternatively spliced exons of eukaryotic genes. *Int. J. Mol. Sci.*, **20**, 3834.
54. Wahl,M.C., Will,C.L. and Lührmann,R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
55. Dominguez,D., Freese,P., Alexis,M.S., Su,A., Hochman,M., Palden,T., Bazile,C., Lambert,N.J., Van Nostrand,E.L., Pratt,G.A. *et al.* (2018) Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell*, **70**, 854–867.
56. Hiller,M., Zhang,Z., Backofen,R. and Stamm,S. (2007) Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.*, **3**, e204.
57. Zhang,J., Kuo,C.C. and Chen,L. (2011) GC content around splice sites affects splicing through pre-mRNA secondary structures. *BMC Genomics*, **12**, 90.
58. Goren,A., Ram,O., Amit,M., Keren,H., Lev-Maor,G., Vig,I., Pupko,T. and Ast,G. (2006) Comparative analysis identifies exonic splicing regulatory sequences–The complex definition of enhancers and silencers. *Mol. Cell*, **22**, 769–781.
59. Saldi,T., Riemondy,K., Erickson,B. and Bentley,D.L. (2021) Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. *Mol. Cell*, **81**, 1789–1801.
60. Doetsch,M., Schroeder,R. and Fürtig,B. (2011) Transient RNA-protein interactions in RNA folding. *FEBS J*, **278**, 1634–1642.
61. Spitale,R.C., Flynn,R.A., Zhang,Q.C., Crisalli,P., Lee,B., Jung,J.W., Kuchelmeister,H.Y., Batista,P.J., Torre,E.A., Kool,E.T. *et al.* (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, **519**, 486–490.
62. Bevilacqua,P.C., Ritchey,L.E., Su,Z. and Assmann,S.M. (2016) Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.*, **50**, 235–266.
63. Sun,L., Fazal,F.M., Li,P., Broughton,J.P., Lee,B., Tang,L., Huang,W., Kool,E.T., Chang,H.Y. and Zhang,Q.C. (2019) RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.*, **26**, 322–330.
64. Christy,T.W., Giannetti,C.A., Houlihan,G., Smola,M.J., Rice,G.M., Wang,J., Dokholyan,N.V., Laederach,A., Holliger,P. and Weeks,K.M. (2021) Direct mapping of higher-order RNA interactions by SHAPE-JuMP. *Biochemistry*, **60**, 1971–1982.
65. Pezeshkpoor,B., Zimmer,N., Marquardt,N., Nanda,I., Haaf,T., Budde,U., Oldenburg,J. and El-Maarri,O. (2013) Deep intronic 'mutations' cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA. *J. Thromb. Haemost.*, **11**, 1679–1687.
66. Kralovicova,J., Knut,M., Cross,N.C. and Vorechovsky,I. (2015) Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3' splice-site organization and activity of U2AF-related proteins. *Nucleic Acids Res.*, **43**, 3747–3763.
67. Yeo,G. and Burge,C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
68. Freund,M., Asang,C., Kammler,S., Konermann,C., Krummheuer,J., Hipp,M., Meyer,I., Gierling,W., Theiss,S., Preuss,T. *et al.* (2003) A novel approach to describe a U1 snRNA binding site. *Nucleic Acids Res.*, **31**, 6963–6975.
69. Ke,S., Shang,S., Kalachikov,S.M., Morozova,I., Yu,L., Russo,J.J., Ju,J. and Chasin,L.A. (2011) Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res*, **21**, 1360–1374.
70. Mandal,A.K., Pandey,R., Jha,V. and Mukerji,M. (2013) Transcriptome-wide expansion of non-coding regulatory switches: evidence from co-occurrence of Alu exonization, antisense and editing. *Nucleic Acids Res.*, **41**, 2121–2137.
71. Královicová,J., Ševcíková,I., Stejskalová,E., Obuca,M., Hiller,M., Stanek,D. and Vorechovský,I. (2018) PUF60-activated exons uncover altered 3' splice-site selection by germline missense mutations in a single RRM. *Nucleic Acids Res.*, **46**, 6166–6187.
72. Teixeira-Silva,A., Silva,R.M., Carneiro,J., Amorim,A. and Azevedo,L. (2013) The role of recombination in the origin and evolution of Alu subfamilies. *PLoS One*, **8**, e64884.
73. Horowitz,D.S. and Krainer,A.R. (1994) Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends Genet*, **10**, 100–106.
74. Krull,M., Brosius,J. and Schmitz,J. (2005) Alu-SINE exonization: en route to protein-coding function. *Mol. Biol. Evol.*, **22**, 1702–1711.
75. Larsen,N. and Zwieb,C. (1991) SRP-RNA sequence alignment and secondary structure. *Nucleic Acids Res.*, **19**, 209–215.
76. Bennett,E.A., Keller,H., Mills,R.E., Schmidt,S., Moran,J.V., Weichenrieder,O. and Devine,S.E. (2008) Active Alu retrotransposons in the human genome. *Genome Res*, **18**, 1875–1883.
77. Aktaş,T., Avşar Ilık,İ., Maticzka,D., Bhardwaj,V., Pessoa Rodrigues,C., Mittler,G., Manke,T., Backofen,R. and Akhtar,A. (2017) DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature*, **544**, 115–119.
78. Chang,D.Y., Newitt,J.A., Hsu,K., Bernstein,H.D. and Maraia,R.J. (1997) A highly conserved nucleotide in the Alu domain of SRP RNA mediates translation arrest through high affinity binding to SRP9/14. *Nucleic Acids Res.*, **25**, 1117–1122.
79. Kempf,G., Wild,K. and Sinning,I. (2014) Structure of the complete bacterial SRP Alu domain. *Nucleic Acids Res.*, **42**, 12284–12294.
80. 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.
81. Strub,K., Moss,J. and Walter,P. (1991) Binding sites of the 9- and 14-kilodalton heterodimeric protein subunit of the signal recognition particle (SRP) are contained exclusively in the Alu domain of SRP RNA and contain a sequence motif that is conserved in evolution. *Mol. Cell. Biol.*, **11**, 3949–3959.
82. Huangfu,D., Liu,A., Rakeman,A.S., Murcia,N.S., Niswander,L. and Anderson,K.V. (2003) Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature*, **426**, 83–87.
83. Airik,R., Schueler,M., Airik,M., Cho,J., Ulanowicz,K.A., Porath,J.D., Hurd,T.W., Bekker-Jensen,S., Schrøder,J.M., Andersen,J.S. *et al.* (2016) SDCCAG8 Interacts with RAB Effector

Proteins RABEP2 and ERC1 and Is Required for Hedgehog Signaling. *PLoS One*, **11**, e0156081.
84. Otto,E.A., Hurd,T.W., Airik,R., Chaki,M., Zhou,W., Stoetzel,C., Patil,S.B., Levy,S., Ghosh,A.K., Murga-Zamalloa,C.A. *et al.* (2010) Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat. Genet.*, **42**, 840–850.
85. Kralovicova,J., Patel,A., Searle,M. and Vorechovsky,I. (2015) The role of short RNA loops in recognition of a single-hairpin exon derived from a mammalian-wide interspersed repeat. *RNA Biol.*, **12**, 54–69.
86. Xing,J., Hedges,D.J., Han,K., Wang,H., Cordaux,R. and Batzer,M.A. (2004) Alu element mutation spectra: molecular clocks and the effect of DNA methylation. *J. Mol. Biol.*, **344**, 675–682.
87. Ullu,E. and Weiner,A.M. (1985) Upstream sequences modulate the internal promoter of the human 7SL RNA gene. *Nature*, **318**, 371–374.
88. Orioli,A., Pascali,C., Pagano,A., Teichmann,M. and Dieci,G. (2012) RNA polymerase III transcription control elements: themes and variations. *Gene*, **493**, 185–194.
89. Arndt,P.F., Hwa,T. and Petrov,D.A. (2005) Substantial regional variation in substitution rates in the human genome: importance of GC content, gene density, and telomere-specific effects. *J. Mol. Evol.*, **60**, 748–763.
90. Schmidt,S., Gerasimova,A., Kondrashov,F.A., Adzhubei,I.A., Adzuhbei,I.A., Kondrashov,A.S. and Sunyaev,S. (2008) Hypermutable non-synonymous sites are under stronger negative selection. *PLoS Genet*, **4**, e1000281.
91. Attig,J. and Ule,J. (2019) Genomic accumulation of retrotransposons was facilitated by repressive RNA-binding proteins: a hypothesis. *Bioessays*, **41**, e1800132.
92. Schwartz,S., Gal-Mark,N., Kfir,N., Oren,R., Kim,E. and Ast,G. (2009) Alu exonization events reveal features required for precise recognition of exons by the splicing machinery. *PLoS Comput. Biol.*, **5**, e1000300.
93. Rousset,F., Pélandakis,M. and Solignac,M. (1991) Evolution of compensatory substitutions through G.U intermediate state in Drosophila rRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 10032–10036.
94. Kirby,D.A., Muse,S.V. and Stephan,W. (1995) Maintenance of pre-mRNA secondary structure by epistatic selection. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 9047–9051.
95. Ananth,P., Goldsmith,G. and Yathindra,N. (2013) An innate twist between Crick's wobble and Watson-Crick base pairs. *RNA*, **19**, 1038–1053.
96. Westhof,E. (2014) Isostericity and tautomerism of base pairs in nucleic acids. *FEBS Lett.*, **588**, 2464–2469.
97. Westhof,E., Yusupov,M. and Yusupova,G. (2019) The multiple flavors of GoU pairs in RNA. *J. Mol. Recogn.*, **32**, e2782.
98. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
99. Olson,W.K., Li,S., Kaukonen,T., Colasanti,A.V., Xin,Y. and Lu,X.J. (2019) Effects of noncanonical base pairing on RNA folding: structural context and spatial arrangements of G·A pairs. *Biochemistry*, **58**, 2474–2487.
100. Soni,K., Kempf,G., Manalastas-Cantos,K., Hendricks,A., Flemming,D., Guizetti,J., Simon,B., Frischknecht,F., Svergun,D.I., Wild,K. *et al.* (2021) Structural analysis of the SRP Alu domain from Plasmodium falciparum reveals a non-canonical open conformation. *Commun. Biol.*, **4**, 600.
101. Hartmuth,K., Urlaub,H., Vornlocher,H.P., Will,C.L., Gentzel,M., Wilm,M. and Lührmann,R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 16719–16724.
102. Zhang,S., Herrmann,C. and Grosse,F. (1999) Pre-mRNA and mRNA binding of human nuclear DNA helicase II (RNA helicase A). *J. Cell Sci.*, **112**, 1055–1064.
103. Bratt,E. and Ohman,M. (2003) Coordination of editing and splicing of glutamate receptor pre-mRNA. *RNA*, **9**, 309–318.
104. Song,Z., Gremminger,T., Singh,G., Cheng,Y., Li,J., Qiu,L., Ji,J., Lange,M.J., Zuo,X., Chen,S.J. *et al.* (2021) The three-way junction structure of the HIV-1 PBS-segment binds host enzyme important for viral infectivity. *Nucleic Acids Res.*, **49**, 5925–5942.
105. Matera,A.G., Hellmann,U. and Schmid,C.W. (1990) A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.*, **10**, 5424–5432.
106. Kong,Y., Rose,C.M., Cass,A.A., Williams,A.G., Darwish,M., Lianoglou,S., Haverty,P.M., Tong,A.J., Blanchette,C., Albert,M.L. *et al.* (2019) Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat. Commun.*, **10**, 5228.
107. Chu,W.M., Ballard,R., Carpick,B.W., Williams,B.R. and Schmid,C.W. (1998) Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol. Cell. Biol.*, **18**, 58–68.
108. Jang,K.L. and Latchman,D.S. (1989) HSV infection induces increased transcription of Alu repeated sequences by RNA polymerase III. *FEBS Lett*, **258**, 255–258.
109. Hsu,P.S., Yu,S.H., Tsai,Y.T., Chang,J.Y., Tsai,L.K., Ye,C.H., Song,N.Y., Yau,L.C. and Lin,S.P. (2021) More than causing (epi)genomic instability: emerging physiological implications of transposable element modulation. *J. Biomed. Sci.*, **28**, 58.
110. Park,S.Y., Seo,A.N., Jung,H.Y., Gwak,J.M., Jung,N., Cho,N.Y. and Kang,G.H. (2014) Alu and LINE-1 hypomethylation is associated with HER2 enriched subtype of breast cancer. *PLoS One*, **9**, e100429.
111. Herbert,A. (2021) To "Z" or not to "Z": Z-RNA, self-recognition, and the MDA5 helicase. *PLoS Genet.*, **17**, e1009513.
112. Ameres,S.L., Shcherbakov,D., Nikonova,E., Piendl,W., Schroeder,R. and Semrad,K. (2007) RNA chaperone activity of L1 ribosomal proteins: phylogenetic conservation and splicing inhibition. *Nucleic Acids Res.*, **35**, 3752–3763.
113. Kovacs,D., Rakacs,M., Agoston,B., Lenkey,K., Semrad,K., Schroeder,R. and Tompa,P. (2009) Janus chaperones: assistance of both RNA- and protein-folding by ribosomal proteins. *FEBS Lett*, **583**, 88–92.