RNA transcription and degradation of Alu retrotransposons depends on sequence features and evolutionary history

Till Baar, Sebastian Dümcke, Saskia Gressel, Björn Schwab, Alexander Dilthey, Patrick Cramer, Achim Tresch

1Institute of Medical Statistics and Computational Biology, Faculty of Medicine, University of Cologne, Cologne 50937, Germany
2Clem edi AG, Schlieren 8952, Switzerland,
3Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Göttingen 37077, Germany,
4Institute of Medical Microbiology and Hospital Hygiene, Medical Faculty, Heinrich-Heine-University Düsseldorf, Düsseldorf 40225, Germany,
5Center for Data and Simulation Science, University of Cologne, Cologne 50923, Germany

Abstract

Alu elements are one of the most successful groups of RNA retrotransposons and make up 11% of the human genome with over 1 million individual loci. They are linked to genetic defects, increases in sequence diversity, and influence transcriptional activity. Still, their RNA metabolism is poorly understood yet. It is even unclear whether Alu elements are mostly transcribed by RNA Polymerase II or III. We have conducted a transcription shutoff experiment by z-amanitin and metabolic RNA labeling by 4-thiouridine combined with RNA fragmentation (TT-seq) and RNA-seq to shed further light on the origin and life cycle of Alu transcripts. We find that Alu RNAs are more stable than previously thought and seem to originate in part from RNA Polymerase II activity, as previous reports suggest. Their expression however seems to be independent of the transcriptional activity of adjacent genes. Furthermore, we have developed a novel statistical test for detecting the expression of quantitative trait loci in Alu elements that relies on the de Bruijn graph representation of all Alu sequences. It controls for both statistical significance and biological relevance using a tuned k-mer representation, discovering influential sequence features missed by regular motif search. In addition, we discover several point mutations using a generalized linear model, and motifs of interest, which also match transcription factor-binding motifs.

Keywords: Alu elements; Alu transcription; RNA labeling; retrotransposons

Introduction

Alu elements are around 300-bp-long RNA retrotransposons (Lander et al. 2001, for review, see Deininger et al. 2011). They are classified as short interspersed nuclear elements (SINEs) and are highly abundant in the genome of higher primates (Gentles et al. 2005). Over 1 million Alu loci are currently annotated in the human genome. This means that 11% of the whole genome consists of just Alu sequences (Lander et al. 2001), making them one of the most successful groups of mobile elements.

Alu elements were discovered using a restriction endonucl


is itself a component of the SRP (Ullu and Tschudi 1984; Kriegs et al. 2007). They possess a dimeric structure with a left arm and a right arm, separated by a variable A-rich region, as is also shown later in Fig. 5 (Evgen’ev 2007). In addition, they possess 2 sequence features of note, the first being a bipartite RNA polymerase III (Pol-III) promoter located in the left arm, which is split into A box and B box (Paolella et al. 1983; Orioli et al. 2012). The second is the UGU(NR)-binding motif required for attachment to the ribosome exit tunnel (DagAn et al. 2004).

While only present in higher primates, Alu elements most likely evolved from B1 repeats in rodents, which became the free left and right Alu monomers (FLAM and FRAM) in primates, which lastly fused to form the Alu elements (Quentin 1992; Kriegs et al. 2007). However, Alu evolution through error-prone retrotransposition continues and has given rise to many Alu families (Richard Shen et al. 1991; Deininger et al. 1992). The evolutionary oldest is Aluj, followed by AluS, and finally, AluY, which shows the most transposition activity in humans (Batzer et al. 1996). An Alu family tree is shown in Fig. 1b (UCSC Genome Browser Annotation).

Despite the presence of a Pol-III promoter in their sequence, Alu elements are expected to be transcribed not only by Pol-III but also by RNA polymerase II (Pol-II) (Conti et al. 2015; Zhang et al. 2019). In this study, we investigate different hypotheses regarding Alu transcript origins further, using metabolic RNA labeling by 4-thiouridine [dynamic transcriptome analysis (DTA), Schwalb et al. 2012] coupled with the inhibition of Pol-II (see Methods, Alu RNA used interchangeably with Alu transcripts). This is of particular interest regarding previous experiments that utilized Pol-II inhibition, as it is common practice to rely on SINEs as a non-Pol-II-dependent negative control group in such cases.

We also explore the half-life of Alu RNA in the cell, which is expected to be low, as the transcripts are presumed to be unstable (An et al. 2004). Finally, we also analyze the totality of annotated Alu loci to detect sequence features that influence Alu transcription. While standard motif search has been applied to this problem with limited success in previous studies (Zhang et al. 2019), we use a base-level generalized linear model (Nelder and Wedderburn 1972), and alternatively a colored and compacted de Bruijn graph (Dury and Waterman 1995; Pevzner et al. 2001; Iqbal et al. 2012), which lead to the de novo discovery of highly significant motifs that can partly be assigned to known transcription factors.
Methods

Data sets
The sequencing data used in this investigation are described in detail in our previous publication Schwalb et al. (2016), exception for the sequencing data relating to the inhibition of RNA Pol-II by α-amanitin, which was newly generated. Both data sets are available in NCBI’s Gene Expression Omnibus, as described in Data Availability, including a table containing the positions, counts, and differential expression analysis results for the examined RNAs. The data set from Schwalb et al. (2016) is used throughout this investigation, except where α-amanitin inhibition is concerned.

KS62 cells
Human K562 erythroleukemia cells were obtained from DSMZ (Cat. # ACC-10; RRID: CVCL_0004). K562 cells were cultured in accordance with the DSMZ Cell Culture standards in RPMI 1640 medium (Thermo Fisher Scientific) containing 10% heat inactivated fetal bovine serum (Thermo Fisher Scientific), 1% penicillin–streptomycin, and 1% GlutaMAX supplement (Thermo Fisher Scientific) at 37°C in a humidified 5% CO2 incubator. K562 cells used in this study display the phenotypic properties, including morphology and proliferation rate, that have been described in the literature. Cells were verified to be free of mycoplasma contamination using the Plasma Test Mycoplasma Detection Kit (InvivoGen). Biological replicates were cultured independently.

α-Amanitin treatment
α-Amanitin is a toxic substance from the mushroom Amanita phalloides and a potent inhibitor of RNA Pol-II (Lindell et al. 1970; Kedinger et al. 1970; Stirpe and Fiume 1967; Jacob et al. 1970). Among the RNA polymerases, RNA Pol-II is inhibited at low α-amanitin concentrations, RNA Pol-III may be inhibited at high concentrations (>250 μM), while RNA polymerase I (Pol-I) and mitochondrial RNA polymerases remain unaffected by α-amanitin. Treatment conditions were optimized for selective RNA Pol-II inhibition (see RT-qPCR and Western blotting). For α-amanitin (5 or 15 μM) treatments of K562 cells, cells were treated for a time course of 0–9 h at 37°C in a humidified 5% CO2 incubator. For TT-seq and RNA-seq, K562 cells were treated for 8 h at 37°C in a humidified 5% CO2 incubator with 5 μM α-amanitin or solvent (water).

RT-qPCR
For each condition (5 or 15 μM α-amanitin or solvent), 5 × 105 cells were harvested at 3,000 × g for 2 min. Total RNA was isolated with QIAzol (QIAGEN) according to manufacturer’s instructions except for the addition of 10 ng RNA spike-in mix (6 spike-ins selected from ERCC RNA spike-in mix) (Schwalb et al. 2016) together with QIAzol. To remove possible genomic DNA contamination, isolated RNA (10 ng) was treated with TURBO DNase (Thermo Fisher Scientific) according to manufacturer’s instructions. For reverse transcription (RT), random hexamer priming [S’d(NNNNNN)-3’, N = G, A, T, or C] was used according to manufacturer’s instructions. Briefly, 1 μg of DNase-treated RNA, Random Hexamer primers (final concentration of 5 ng μl−1), and dNTP mix (final concentration of 0.5 mM) were mixed and incubated at 65°C for 5 min. Subsequently, Maxima H Minus Reverse Transcriptase (RT) (final concentration of 200 U) and 5× Maxima RT buffer (Thermo Fisher Scientific) were added (−RT reaction). For DNA contamination control, cDNA synthesis without RT (−RT reaction) was performed (RT was substituted with water). The (−/+) RT reactions were incubated in a PCR cycler at 25°C for 10 min, 50°C for 30 min, and 85°C for 5 min. Primers for quantitative PCR were designed by using the online primer design software Primer3 v.0.4.0 (Rozen and Skaltsky 2000). Briefly, the selection of Pol-II targets was based on half-lives measurements in K562 cells (Schwalb et al. 2016). Primers were targeted to Pol-II transcripts (PAIP1, CWC22, EGR1) as a positive control, RNA Pol-I transcripts (18S rRNA) as a negative control, and RNA Pol-III transcripts (U6 snRNA). Primer specificity (single product peak) was validated by melting profiles. Primer sequences, length, PCR efficiency values of primers (E) and targets are reported in Supplementary material S2. cDNAs (50 ng) were amplified with SYBR Select Master Mix (Thermo Fisher Scientific) according to the manufacturer’s instruction with a final primer concentration of 400 nM. PCRs were run in 96-well optical plates sealed with optical adhesive cover on a qTOWER 2.0/2.2 instrument (Analytik Jena AG). The following thermal cycling conditions were used (SYBR Select Master Mix reference, standard cycling mode): 50°C for 2 min, 95°C for 2 min, 40 cycles of 95°C for 15 s, and 60°C for 1 min. Two synthetic RNA spike-ins were used for normalization. The 2−ΔΔCt method was applied to calculate the normalized target gene expression fold change, with the amplification efficiency (E) for each target gene, slope of standard curve (S), and mean threshold cycle (Ct) (Livak and Schmittgen 2001).

Western blot
For α-amanitin (5 or 15 μg mL−1) treatments of K562 cells, cells were treated for 8 h at 37°C in a humidified 5% CO2 incubator. Water was used as a solvent. After 8 h of treatment, K562 cells were harvested at 3,000 × g for 2 min and washed twice in Dulbecco’s phosphate-buffered saline (DPBS). Cells were lysed in radioimmunoprecipitation assay (RIPA) lysis buffer for 45 min on ice and centrifuged at 14,000 × g for 15 min at 4°C. Samples were quantified with the Bradford method. Fifteen micrograms of cell lysate was denatured in 4× Loading Dye (including 100 mM DTT) at 70°C for 10 min. PrecisionPlus Protein All Blue Standard (Bio-Rad, # 161-0373, 10–250 kDa) was used as marker. NuPAGE 4–12% Bis–Tris Protein Gels and MOPS buffer were used according to manufacturer’s instructions. Transfer of a single NuPAGE onto PVDF membrane was performed in transfer buffer at 30 V for 1 h in a XCell II Blot Module (semiwet transfer unit) according to manufacturer’s instructions. Membrane blocking was performed in 5% milk PBS-T on a rocking surface for at least 1 h. Primary antibody was added overnight. The following primary antibodies were used in this study: N-terminal POLR2A/hRPB1 antibody, clone F-12 (Santa Cruz Biotechnology, sc-55492, Lot. # E2913; RRID: AB_1078991) as a loading control. HRP-coupled secondary antibody targeting mouse IgG (Aldrich, G8795, Lot. # 067M4785V; RRID: AB_630203) and GAPDH antibody, clone F-12 (Santa Cruz Biotechnology, sc-55492, Lot. # E2913; RRID: AB_630203) as marker. N-terminal POLR2A/hRPB1 antibody was used at a dilution of 1:3,000 and incubated on a rocking surface for 1 h. Chemiluminescence Substrate (ECL) working solution of SuperSignal West Pico PLUS Chemiluminescent Substrate was prepared according to manufacturer’s instructions. Proteins were visualized by chemiluminescence detection on INTAS. POLR2A/hRPB1 (RNA Pol-II) degradation, which is α-amanitin dose dependent (Nguyen et al. 1996), was monitored by Western blotting.

RNA spike-ins
Synthetic RNA spike-in controls are derived from selected RNAs of the ERCC RNA Spike-in Mix (Ambion) as described in Gressel et al. (2019) and Schwalb et al. (2016). Briefly, spike-ins (3 unlabeled and 3 4sU labeled) are in vitro transcribed using the
MEGAscript T7 kit (Ambion). In vitro transcription (IVT) of unlabeled spike-ins was performed following the manufacturer’s instructions. For IVT of 4sU-labeled spike-ins, 10% of UTP was substituted with 4-thio-UTP (Jena Bioscience). RNA spike-ins were purified with RNaClean XP beads (Beckman Coulter) following manufacturer’s instructions. The final RNA spike-in pool contained equal amounts of all RNA spike-ins.

**TT-seq and RNA-seq**

A detailed step-by-step protocol has been deposited in the protocols.io repository (Gressel et al. 2019a). TT-seq and RNA-seq were performed in 2 biological replicates including RNA spike-ins. Briefly, experiments were performed using 5 x 10^7 K562 cells per biological replicate. Cells were kept at optimal growth conditions and supplemented with 5 μg mL^-1 of α-amanitin or solvent (water) for 8h. After 7 h 55 min, a 4-thiouridine (4sU) labeling pulse (Sigma-Aldrich, T4509) was applied for 5 min using 500μM (see Supplementary material S3). Total RNA was isolated with the QiAzoI reagent (#79306) according to manufacturer’s instructions except for the addition of 150 ng of RNA spike-in pool with QiAzoI reagent as previously described (Schwalb et al. 2016; Gressel et al. 2019a). The Ovation Universal RNA-Seq System (NuGEN) was used for strand-specific library preparation as described (Gressel et al. 2019b). Purified cDNA libraries were analyzed by Fragment Analyzer prior to Illumina sequencing. Sequencing was performed on a HiSeq 2500 (Illumina) in paired-end mode with 50-bp read length.

**TT-seq/RNA-seq prepossessing and normalization parameters**

TT-seq and RNA-seq prepossessing and normalization were performed as detailed in Schwalb et al. (2016), with an alternative normalization applied as described under Differential Expression Analysis. Reads were mapped with STAR v2.7.3a (Dobin et al. 2016), allowing a maximum of 10 mismatches and minimal uniqueness filtering with an MAPQ value cutoff of 25S. The general error rate per library ranged between 0.6% and 0.9%. As Alu sequences are notoriously difficult to map (Sexton and Han 2019), we performed a mappability analysis (see Supplementary material S4) to make certain that our read alignment does not introduce biases. We simulated a homogeneous coverage of the genome and mapped the simulated reads using the same aligner settings. Our analysis shows that there is no bias caused by Alu mappability, which is also corroborated by the findings of Sexton and Han (2019), who show that the mappability of transposable elements can be improved through the use of paired-end read libraries to the point where a majority of elements are uniquely mappable. Association of the sequencing data with specific Alu elements or protein-coding genes used the GRCh37/hg19 repeat annotation or respectively the canonical gene annotation of the UCSC Table Browser (Karolchik et al. 2004), in addition to the RepeatMasker human Alu subfamily re-analysis (Price et al. 2004). We also tested using the transcription unit annotation from Gressel et al. (2019b) based on the GenoSTAN segmentation algorithm as an alternative to the UCSC canonical gene annotation, which leads to a similar mRNA read count distribution (Zacher et al. 2017, data not shown).

**RNA half-life estimation**

To estimate the half-life of expressed Alu elements and coding transcripts, we used MLE, described in detail in Supplementary material S5. Our estimation makes use of the labeled and total RNA sequencing fractions obtained through TT-seq and 4sU-seq, as detailed above. TT-seq and 4sU-seq data were used for the estimation, as the correlation between the 2 methods is suitably strong (r > 0.80 between replicates and r > 0.85 between sequencing methods). While 4sU-seq was designed to estimate half-lives, TT-seq was primarily designed to measure the polymerase processivity during transcription. Since Alu elements are short, TT-seq essentially measures the synthesis rate of Alu transcripts and can thus be applied for half-life estimation (see Supplementary material S6).

We assume that the total amount of transcripts in a cell remains constant and that transcript degradation follows exponential decay, meaning that the ratio of labeled to total transcripts LT increases exponentially to 1 over time. We further assume that the distribution of labeled and total counts follow a Poisson distribution. This allows us to formulate the likelihood function L of observing a specific number of labeled counts La and total counts Ta for any given Alu element or coding transcript a, as a parametrized combination of the underlying Poisson distribution:

\[
L(a; T_a; r_a; q, L_{qpk}, T_{qpk}) = \text{Pois}(T_a; \lambda_{clab} = C_{lab} \cdot r_a^{-1}) \\
\quad \times \text{Pois}(L_a; \lambda_{clab} = C_{lab} \cdot r_a^{-1})
\]

Here, \( r_a \) is the ratio between labeled and total molecules for any given Alu element or coding transcript a; q is the ratio between labeled and total spike-in molecules that were added; \( L_{qpk} \) and \( T_{qpk} \) are the number of labeled and total spike-in counts respectively; and \( C_{lab} \) and \( C_{tot} \) are given as:

\[
c_{tot} := \frac{L_a}{L_{qpk}} \cdot \frac{T_{qpk}}{q} \quad \text{and} \quad c_{lab} := \frac{L_{qpk}}{T_{qpk}} \cdot \frac{1}{q}
\]

This estimation does not depend on feature length, as we use the ratio between labeled and total counts. To find the optimal estimate of \( r_a \), we maximize the log likelihood function:

\[
\ell(r_a; L_a, T_a, q, L_{qpk}, T_{qpk}) = L_a \left( \ln(r_a) - \frac{L_{qpk}}{T_{qpk}} r_a \right) \\
+ T_a \left( \frac{L_{qpk}}{T_{qpk}} q - \ln(r_a) \right)
\]

From \( r_a \) when then obtain the degradation rate \( \delta_a \) by use of \( r_a = 1 - \exp(-\delta_a \Delta t) \), where \( \Delta t \) is the labeling pulse’s duration. This allows us to calculate the half-life as \( t_{1/2} = \ln(2)/\delta_a \).

As the used DTA methods rely on the labeling of U as 4sU, the percentage of U in an Alu sequence could potentially bias its half-life estimate. However, we can exclude this possibility, as no correlation exists between Alu half-life and U ratio (\( r < 0.01 \)).

**Differential expression analysis**

Differential expression analysis of the α-amanitin Pol-II inhibition experiment was performed using the DESeq2 package for R (Love et al. 2014), using un-normalized counts as required by the statistical model. The standard normalization strategy employed by DESeq2 internally relies on the assumption that there are no substantial, systematic global changes in (mRNA) expression between samples. Since this is most likely not given in the α-amanitin sample, we used a set of bona fide housekeeping transcripts. As it is known that the activity of mitochondrial polymerase is unaffected by α-amanitin (Menon 1971; Reid and Parsons 1971; Saccone et al. 1971), we chose mtRNAs for normalization. Alternatively, to check the robustness of our choice, we also
performed normalization based on a set of spike-in RNAs that had been added during sequencing library preparation (see RNA Spike-Ins). Similar results were obtained when spike-in RNAs were taken as reference as compared to mRNAs (Supplementary material S7c). The normalization constants obtained differ by a mean factor of 1.10, and consequently the results of our analysis hold under both regiments.

Rejection sampling was used to ensure that the obtained l2fc distributions are comparable (Flury 1990). To avoid biases arising from the substantially different expression level distribution of Alus and mRNAs, we compare the Alu l2fc distribution to the l2fc distribution of a large sample of mRNAs that were picked randomly such that their wild-type expression distribution is identical to that of the Alu elements. To generate this random sample, we use rejection sampling: we define expression breakpoints

\[ b_0 = \text{inf} < b_1 < b_2 < \cdots < b_{N-1} < b_N = \text{inf} \]

such that each bin \([b_n, b_{n+1})\] contains 1,500 Alu elements, excluding Alu elements with less than 5 reads. Thus, the empirical expression distribution

\[ f(n) = p(\text{Alu expression in}[b_{n-1}, b_n]) = 1/N, \quad n = 1, \ldots, N \]

is uniform. The same breakpoints are used to calculate the empirical expression distribution

\[ g(n) = p(\text{mRNA expression in}[b_{n-1}, b_n]), \quad n = 1, \ldots, N \]

of the mRNAs by calculating the relative abundance of mRNAs whose expression falls into the interval \([b_{n-1}, b_n]\). Let \( k = \min(g(n), n = 1, \ldots, N) \) and accept an mRNA sample that lies in the interval \( m \) with probability \( k/g(k) \). It can then be shown that the set of accepted samples approaches, for large numbers, the distribution \( f(n) \).

Generalized linear model

The generalized linear models (GLMs) to analyze the genomic sequences of all annotated Alu elements on base level were created with the glmnet package for R (Friedman et al. 2010). A Poisson family distribution response type was assumed with observed-expected standardization, an elastic mixing parameter \( \alpha \) of 1 (full Lasso penalty, no ridge regression penalty), no fitted intercept parameter, and 1,000× cross-validation. The input matrix for each model consisted of a binary encoding of the examined point mutation, base exchange, deletion, or insertion, for each position in the Alu consensus sequence. These matrices were obtained by aligning each individual Alu sequence against the Alu consensus sequence. These predictor matrices where then paired with the expression values as response variable for each individual model. The resulting effect sizes (\( \beta \) parameters) for the point mutations were then annotated along the Alu secondary structure according to Sinnett et al. (1991) (dot-bracket notation of the secondary structure in Supplementary material S8) to illustrate their context regarding Alu sequence features. The \( \beta \) parameters are the estimated regression coefficients of the GLM.

de Bruijn graph

The de Bruijn graph of all Alu sequences, regardless of expression, was generated using bifrost v1.0.5 (Holley and Melsted 2020), with a k-mer length of 9 and including 100 bp of flanking regions up- and downstream of each Alu sequence. \( k = 9 \) was chosen so that the distribution of Alu sequences per node in the graph was smooth, neither running into the lower nor the upper range of possible values (Supplementary material S9). Low-coverage k-mer connecting tips were kept by using the –keep-mercy (\( y \)) argument, and the compacted de Bruijn graph was colored using the –colors (-c) argument.

Our initial intent was to utilize the de Bruijn graph structure in the downstream analysis, but the compacted de Bruijn graph for \( k = 9 \) is almost complete: It contains 131,070 nodes, i.e. k-mers and 1,048, 544 in-edges, resulting in \( >7.99 \) edges per node (similarly \( >7.99 \) for out-edges). Note that the maximum in-degree and out-degree in the compacted graph is 8, not only 4, since each node can be traversed in forward and reverse complement direction (Holley and Melsted 2020). We, therefore, focused purely on the k-mers.

The resulting k-mers were then filtered according to 2 aspects. First, the expression of those Alu elements possessing the k-mer was compared with those not possessing it. Second, for any given k-mer, its suffixes and prefixes were generated. Each k-mer possesses 4 potential pre- and suffixes, as each sequence can be preceded or succeeded by one of the 4 bases.

When testing a k-mer, say XJY, with \( X, Y \in \{A, C, T, G\} \) and \( J \) a fixed 2-mer, for having an effect on Alu transcription, we have to guard against false positive due to the large number of k-mer tests. This is done by Bonferroni multiple testing correction for a family-wise error rate of at most 5%. We further need to filter out irrelevant findings resulting from the large number of observations that can render even small differences highly significant. To ensure that XJY is relevant for Alu transcription, we compare the group of Alu elements containing XJY to the group of Alu elements that contain the prefix X or the suffix JY, but not the full k-mer XJY. We compare the 2 groups with respect to their binarized expression (expressed with at least 1 read count/not expressed) using a Fisher test, and we require an OR of at least 2 or at most 0.5 for being considered a relevant difference.

Results

Alu expression increases with family age

Many Alu families that have emerged in the human genome over the last 65 million years vary not only in their sequence characteristics (Jurka and Smith 1988, but also in their activity (Bennett et al. 2008; Oler et al. 2012), active Alu elements being those that are not only transcribed but are also mobile have the capability to reinsert themselves into the genome through retrotransposition.

As was found by Bennett et al. (2008) in vitro, the activity of an Alu family appears to scale inversely with its age, the older an Alu element, the less likely it is to be still capable of successful reinsertion. We examined the transcriptional activity of Alu families, using the RNA-seq data from Schwab et al. (2016) (see Methods), as well as the stability of Alu transcripts, which is expected to be low in comparison to regular mRNAs (An et al. 2004 and Fig. 2, a–c). We used maximum likelihood estimation (MLE) based on metabolic RNA labeling to calculate the half-life \( t_{1/2} \) of Alu RNAs (see Methods), which relates to the degradation rate \( \delta \) by \( t_{1/2} = \ln(2)/\delta \), as further detailed in Supplementary material S5.

The global expression distribution of Alu transcripts (mean \( X = 85 \), median \( X = 27 \)) is about 2 orders of magnitude lower than that of mRNAs (mean \( X = 4088 \), median \( X = 740 \)), as is to be expected (Fig. 2a, Paulson and Schmid 1986). We found that members of the younger Alu families, mostly AluY, but also AluSQ10, exhibit overall less transcriptional activity than older Alus (Fig. 2b). This effect is most pronounced with the subfamilies of AluY. Since the
Alu expression and half-life—all subfigures based on data published by Schwalb et al. (2016), using the UCSC Genome Browser Annotation. a) Histogram of the read counts for both Alu elements and mRNA; Alu elements show less transcriptional activity overall (histogram y-axes: density). b) Read count distribution for all Alu families as shown in Fig. 1b. The colored bars represent the extent of the 50%, 75%, and 95% median percentile interval with decreasing opacity. The white dot indicates the arithmetic mean. Zero counts are disregarded in this subfigure. The upper lollipop plot shows the proportion of Alu elements per family with more than 0 read counts. c) Histogram of the half-life of both Alu elements and mRNA; Alu elements show stability comparable to regular transcripts. d) Half-life distribution for all Alu families (cf. Fig. 1b), binned by family. e) ECDF of the read counts of younger Alu families (AluY and subsequent, as well as AluSq10) in contrast to all older families, represented by the gray-shaded area. Younger Alu families show consistently lower read counts in comparison to the older ones.

Alu expression is independent of gene transcription

Alu elements are often presumed to be Pol-III transcripts, as they contain a bipartite Pol-III promoter (Orioli et al. 2012). Although there is experimental evidence for Alu elements being transcribed by Pol-III (Jagadeeswaran et al. 1981; Panning and Smiley 1993; Zhang et al. 2019), it has not been ruled out yet that (some) Alu elements might be transcribed by RNA Pol-II. We have therefore examined the possibility that Alu transcripts may also arise as side products of RNA Pol-II gene transcription, as hinted at by the results of Conti et al. (2015) and suggested by Zhang et al. (2019).

If Alu elements are also transcribed by Pol-II as side products of gene transcription, we would expect intragenic Alu elements to exhibit a higher transcript abundance than intergenic ones. Furthermore, the transcriptional activity of genes should be correlated with that of proximal Alu elements. The expression of those should be biased toward Alu elements lying in sense direction with regard to their associated gene.

To investigate this, we defined intragenic Alu elements as those that either directly overlap a gene or lie within 500-bp upstream or downstream of a gene according to the GRCh37/hg19 canonical gene annotation of the UCSC Table Browser (Karolchik et al. 2004). All other Alu elements are called intergenic. Over 96% of intragenic Alu elements overlap intronic regions according to the latest GENCODE annotation (Frankish et al. 2021). Effects of proximity to Pol-I and Pol-III transcripts have not been detected and cannot contribute a notable effect, since only a comparably tiny number of Alu elements would be affected. Of the expressed Alu elements, 37 loci lie in 500-bp proximity to rRNAs (excluding 5S rRNA), and 42 in proximity to tRNAs, using the respective UCSC Table Browser annotation.

We compared the read counts for intragenic and intergenic Alu transcripts, which show a significantly higher expression for intragenic Alus compared to intergenic ones (Fig. 3a, \( \bar{x} \approx 110, x = 49 \) intragenic, \( \bar{x} \approx 35, x = 6 \) intergenic, \( P < 10^{-5} \). Mann–Whitney
We also detected a less pronounced but significant difference in the expression of intragenic Alu elements lying in sense respectively in the antisense direction with regard to their associated gene (Fig. 3b, \(x = 44\) sense, \(x = 54\) antisense, \(P < 10^{-5}\), Mann–Whitney test).

We also correlated the transcript abundance of intragenic Alu elements and their associated genes for both Alu elements in the sense and antisense direction to their associated gene. The 2 groups show no difference in correlation strength (Spearman correlation \(r = 0.71\), \(P < 10^{-5}\) rand. test, Fig. 3c).

Lastly, we examined the number of sense and antisense Alu elements per family. Overall, 46% of all intragenic Alu elements lie in the sense direction with regard to their associated gene. No individual family exceeds a ratio of either 38% or 50% (Fig. 3d).

Furthermore, we combined Pol-II and Pol-III occupancy data drawn from the ENCODE portal (ENCSTR000EHL, ENCSR000EHQ, Dunham et al. 2012; Davis et al. 2018) with our RNA-seq data and RNA-seq data from ENCODE (ENCSTR000COM). While we found a modest correlation between our own sequencing’s mRNA transcript abundance and the data obtained from ENCODE (Spearman correlation \(r = 0.65\), our analysis did not detect global correlation between Pol-II or Pol-III occupancy and individual Alu expression according to our sequencing data (Spearman correlation \(r \leq 0.05\) for both, Supplementary material S7b).

**RNA Pol-II inhibition decreases Alu expression**

To further examine the origin of Alu transcripts, we performed a Pol-II inhibition experiment by incubating K562 cells with \(\alpha\)-amanitin at a concentration suited to efficiently inhibit the Pol-II activity while leaving the activity of Pol-III and mitochondrial RNA polymerases unaffected (see Methods). We quantified the degree of inhibition using mitochondrial RNAs as a negative control group, which are unaffected by \(\alpha\)-amanitin (see Methods). Similar results were obtained when spike-in RNAs were taken as

---

**Fig. 3.** Alu correlation and differential expression. a) Histogram of the read counts for transcribed intragenic (398,660) and intergenic (200,928) Alu elements. Intragenic Alu elements either directly overlap a gene or lie within 500-bp up- or downstream of a gene. Other Alu elements are intergenic. Intragenic Alu elements show a significantly higher expression than intergenic ones, which is likely explained by genome accessibility. b) Histogram of the read counts for transcribed intragenic Alu elements that lie in sense or antisense direction with respect to their associated gene, showing only minor differences. c) 2D kernel density estimation showing the correlation between the annotated Alu and mRNA read counts for intragenic sense (orange) and antisense (blue) Alu elements. The 2 groups show no difference in correlation strength. d) Number of sense and antisense Alu elements per family. No family exceeds a ratio of 38% or 50%. e) 2D density heatmap showing the DESeq2 differential expression of Alu elements, mRNAs, and tRNAs as control under \(\alpha\)-amanitin Pol-II inhibition. Semitransparent areas do not pass the significance threshold. Loci with a normalized mean expression below 0.1 are not shown, with affects 51% of all annotated Alu loci and practically no mRNAs or tRNAs. Both Alu elements and mRNAs show stronger significant downregulation than upregulation. f) Differential expression of Alu elements under \(\alpha\)-amanitin Pol-II inhibition, using family-wise aggregated read counts. All Alu families show downregulation.
reference (Supplementary material S7c). The analysis of tRNAs as an independent negative control group showed barely any differential expression, validating that our normalization procedure did not introduce a bias to putatively unaffacted transcripts. As a natural positive control group, we chose mRNAs that are bona fide Pol-II transcripts. Figure 3e shows the changes after α-amanitin inhibition for Alu elements, mRNAs, and tRNAs.

Due to the low coverage of individual Alu transcripts, a direct comparison based on individual expression folds would be misleading. We, therefore, aggregated all read counts mapping to 1 Alu family (or several families in the case of very small families, Fig. 3f) and calculated the expression fold of each Alu family. All Alu family (or several families in the case of very small families, leading. We, therefore, aggregated all read counts mapping to 1 Alu family (or several families in the case of very small families, Fig. 3f) and calculated the expression fold of each Alu family.

The abundance of Alu sequences in the genome offers the unique opportunity to use an analysis strategy resembling a genome-wide association study. We used a GLM to link Alu sequence features to their transcript abundance (see Methods). Due to the encoding of the sequence changes with respect to the Alu consensus sequence, this results in individual effect sizes (β parameters) for the 3 types of examined point mutations, base exchanges, deletions, and insertions, for each position in the Alu consensus sequence. Figure 5 shows the Alu secondary structure according to Sinnett et al. (1991) with the Euclidean norm of the effect sizes represented by base color (dot-bracket notation of the secondary structure in Supplementary material S8). However, the large number of available Alu sequences enabling this analysis also precludes any attempts to test for statistical significance.

Several noteworthy positions show a high correspondence to Alu transcription (Fig. 5): Position 48 is sensitive to all 3 types of point mutations, but primarily to insertions. It lies in the left arm at the juncture between inner loop 7 (I7) and inner loop 9 (I9), which contains the A box of the Pol-III promoter sequence. Position 79, also in the left arm in terminal loop 3 (T3), lies directly within the B box of the Pol-III promoter sequence. Like the previous position, it is sensitive to all 3 types of mutation, but primarily to insertions, which would bring the Alu consensus sequence closer to the consensus sequence of the tRNA promoter: GWTCRCANN (Paolella et al. 1983). Position 179 lies in the right arm near terminal loops 4 and 5 (T4, T5) and reacts only to insertions, and position 198, also in the right arm between inner loops 2 and 3 (I2, I3), reacts only to base exchanges.

Due to the large number of samples (i.e. Alu elements), we could also investigate interactions between paired bases in the secondary structure (see Methods), which resulted in 2 positions of interest, all in the vicinity of terminal loops 4 and 5 (T4, T5) of the right arm, which is also the region containing the SRP 9/14-binding motif: UGU(NR) (Weichenrieder et al. 2000). The pair of positions 135 and 180 reacts inversely to insertions in comparison to base exchanges and deletions. It forms the junction from which terminal loops 4 and 5 (T4, T5) branch off. The pair of positions 139 and 156 lies directly adjacent to the SRP-binding motif and reacts primarily to deletions, as well as to base exchanges.
Fig. 5. GLM and Alu secondary structure—secondary structure of the Alu element consensus sequence according to Sinnett et al. (1991), showing the left and right arm architecture with the variable region in between. Noteworthy sequence features include the bipartite Pol-III promoter split into A and B boxes (denoted by $s_A$ and $s_B$), and the SRP-binding motif used for attachment to the ribosome exit tunnel (denoted by $s_R$). Each base is colored according to the rescaled Euclidean norm of the 3 GLM effect sizes ($\beta$ parameters). Several noteworthy positions are annotated with bar graphs detailing the individual values. In the lower right corner, prominent results from the analysis of interactions between paired bases are shown.
As the GLM focused on individual bases or base pairs in the Alu consensus sequence, we also developed our own k-mer-based method to search for sequence motifs that influence Alu expression, better suited to the data at hand than regular motif search. First, to circumvent the problems that arise from the use of a reference sequence, we constructed a de Bruijn graph from the k-mers of all Alu sequences (k = 9, for the tuning of k and its justification, see Methods). Next, we filtered for k-mers with statistical significance and biological relevance. Each k-mer splits the group of all Alu sequences into those containing the k-mer and those that do not. At the same time, each Alu element can be either expressed (at least 1 read count) or not. This gives rise to a 2 x 2 contingency table, to which we apply a Fisher test and calculate the odds ratio (OR, see Methods). k-mers with both an OR of less than 0.5 or greater than 2 and a Bonferroni-corrected Fisher test P-value less than 0.05 were included in the downstream analysis, leaving us with 38 k-mers of interest. Finally, we performed another Fisher testing comparing expressed and nonexpressed Alu elements in the group containing a specific k-mer of interest vs the group of Alu elements containing the k-1 Prefix and suffix of that k-mer. This resulted in merely 12 remaining k-mers of interest after applying the same stringency criteria (Table 1). We also ensured that the reported k-mers originate from the Alu sequence itself and not from the 100-bp flanking region (see Methods and Supplementary material S11).

A comparison of those k-mers with transcription factor-bind- ing profiles from the JASPAR database reveals that 8 of the 12 k- mers show similarities to known motifs (Fornes et al. 2020). The most intriguing findings to us were the k-mers resembling the binding profiles of ARNT::HIF1A (MA0259.1, which is a master transcriptional regulator of hypoxia response), NFIX (MA0671.1, which is involved in the replication of adenovirus), and TFE3 (MA0831.2, which plays a role in the general immune response). Also, all listed transcription factors are Pol-II specific (see Supplementary material S10 for references and detailed statistics).

### Discussion

We presented a multifaceted analysis of the Alu elements found in the human genome in K562 cells, focusing on their RNA metabolism and sequence features that influence it.

The read count distributions of Alu elements in comparison to mRNA presented in Fig. 2, a, b, and c correspond well to previous in vitro findings, corroborating that Alu expression is generally lower in comparison to gene transcription and that the younger Alu families (AluY and subfamilies) are expressed at even lower rates than the older Alu families (AluJ and AluS) (Paulson and Schmid 1986; Bennett et al. 2008).

In contrast, the results of our RNA half-life estimation of Alu transcripts through the use of TT-seq (Schwalb et al. 2016) and 4uS-seq (Miller et al. 2011) are surprising. Previous studies suggest that Alu RNAs should be less stable than regular gene transcripts, as they contain adenine- and uracil-rich element motifs, a sequence feature linked to decreased transcript stability with an as yet unknown mechanism (An et al. 2004). In contrast, our MLE predicts that Alu transcripts and mRNAs should be of similar stability (Fig. 2, c and d). However, our MLE can only serve as an assessment to compare the relative half-life distributions; its predictions do not represent explicit half-life values. First, the MLE has to contend with the overall paucity of Alu read counts, which limits its explanatory power, an issue that could only be remedied by vastly increasing sequencing depths in future experiments. Second, to not exacerbate the already demanding estimation, our model contains some simplifications, such as the assumption of steady-state conditions, using a Poisson distribution to model read counts instead of a zero-inflated negative binomial distribution, and neglecting nonconstant labeling efficiencies for short labeling periods (see Supplementary material S5). As such, our half-life estimates point toward a need for further research regarding the stability of Alu transcripts, as the subject matter is less straightforward as it appears prima facie. Also, our data originate from K562 cells and are thus subject to the peculiarities of that cell line (Li et al. 2000). Nonetheless, taking into account the nonautonomous life cycle of Alu elements and their reliance on LINE-1 repeat translation (Boeke 1997; Dewannieux et al. 2003) in conjunction with their low expression, transcript persistence in the cell could potentially be instrumental for successful retrotransposition.

Regarding our noninterventional analysis concerning the origin of Alu transcripts, we observed a higher expression of intragenic Alu elements compared to intergenic ones (Fig 3a). This on its own cannot be taken as solid evidence either for or against a connection between Alu and gene transcription. Genome accessibility is a confounding factor, as intragenic regions are in general more accessible than intergenic ones (Guo et al. 2017). It should also be noted that a screening of Alu elements proximal to Pol-1 and Pol-III transcripts, while intriguing, falls outside the scope of our study, as our approach exploits the large group size of intragenic Alu elements to compensate for the low read counts. The difference in expression between intragenic Alu

---

**Table 1.** De Bruijn graph significant k-mers.

<table>
<thead>
<tr>
<th>k-mer</th>
<th>OR</th>
<th>JASPAR match</th>
<th>Matrix ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGCCGCCA</td>
<td>2.65</td>
<td>—</td>
<td>MA0671.1</td>
</tr>
<tr>
<td>ATGCCGCCT</td>
<td>2.72</td>
<td>NFIX</td>
<td>MA1536.1</td>
</tr>
<tr>
<td>CGGACTGCT</td>
<td>2.07</td>
<td>MEIS1</td>
<td>MA0498.2</td>
</tr>
<tr>
<td>CTCAACGCC</td>
<td>2.40</td>
<td>TFAD3</td>
<td>MA0803.1</td>
</tr>
<tr>
<td>CTCAGACGCT</td>
<td>2.12</td>
<td>ARNT::HIF1A</td>
<td>MA0259.1</td>
</tr>
<tr>
<td>GAAACCAGTC</td>
<td>2.27</td>
<td>TFE3</td>
<td>MA0831.2</td>
</tr>
<tr>
<td>GACACCGGC</td>
<td>2.56</td>
<td>GATA2</td>
<td>MA0036.1</td>
</tr>
<tr>
<td>GCCGCAGCT</td>
<td>2.35</td>
<td>MEIS1</td>
<td>MA0498.2</td>
</tr>
<tr>
<td>GGGCGGACT</td>
<td>2.04</td>
<td>GSX2</td>
<td>MA0893.2</td>
</tr>
<tr>
<td>TAGGCCGCC</td>
<td>2.22</td>
<td>TBX4</td>
<td>MA0806.1</td>
</tr>
<tr>
<td>TCAACCGCT</td>
<td>2.92</td>
<td>TBX5</td>
<td>MA0807.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR2C2 (var.2)</td>
<td>MA1536.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GSX2</td>
<td>MA0893.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MGA</td>
<td>MA0801.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA0831.2</td>
<td>MA0806.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA0801.1</td>
<td>MA0801.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TBX5</td>
<td>MA0807.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>USF1</td>
<td>MA0093.1</td>
</tr>
</tbody>
</table>

Shown are the k-mers (k = 9) from the de Bruijn graph created of all Alu sequences (incl. 100-bp flanking regions) that passed all significance and relevance thresholds: P ≤ 0.05 (Fisher test with Bonferroni multiple testing correction) and OR ≤ 0.5 or OR ≥ 2.0 for both test schemes. Listed are the k-mer sequence, associated OR (k-mer occurrence vs read counts), JASPAR match, and associated Matrix ID (see Supplementary material S10 for expanded statistics). The binding profile matches described in the Results are marked as bold.
elements that lie in sense or antisense direction with respect to their associated gene is so minor that we deem it irrelevant (Fig. 3b). Therefore, we analyzed the correlation between Alu and gene transcription (Fig. 3c). A certain level of correlation is to be expected regardless of Alu transcript origin due to genome accessibility, as mentioned above. However, the lack of a difference in correlation strength between sense and antisense points toward Alu transcripts not originating primarily as side products of Pol-II gene transcription. If that were the case, a stronger correlation between Alu elements in sense direction with respect to their associated gene and the expression of that gene should exist. This is further corroborated by the number of sense and antisense Alu elements present in each family (Fig. 3d). If Alu transcripts were primarily side products of Pol-II gene transcription, Alu insertion should show a bias toward aligning new Alu elements in sense direction with respect to their associated gene. Taken together, we interpret these findings as weakening the hypothesis that Alu transcripts are mainly created alongside Pol-II gene transcription. This does not rule out the possibility for a fraction of Alu elements being transcribed in this fashion, but it does not appear to be a major contributor to Alu expression in the cell.

Our analysis of Pol-II and Pol-III occupancy data did not detect any correlation between Alu expression and polymerase occupancy (Supplementary material 5b), which was a likely outcome given the nonconformity of the 2 measurements (Ehrensberger et al. 2013; Gressel et al. 2017). The issue of such a correlation was also addressed recently by Zhang et al. (2019), who utilized an extensive collection of data produced as part of the ENCODE project in conjunction with RAMPAGE (Batut and Gingeras 2013). While their study covered only 1.5% of all annotated Alu elements in the human genome due to strict read-mapping and analysis constraints, they identified a subset of Alu elements as Pol-III transcribed, also in accordance with chromatin immunoprecipitation-seq data. This suggests that the detection of correlation between Alu expression and polymerase occupancy, while possible, necessitates high-resolution data that make a global assessment demanding and beyond the explanatory power to be found in our data.

The differential expression analysis under α-amanitin inhibition of Pol-II suggests that a substantial part of Alu transcription is Pol-II dependent (Fig. 3, e and f). To the author’s knowledge, this is the first direct experimental evidence for the origin of Alu transcripts. If Alu transcripts would originate mainly from the Pol-III activity, no downregulation should be present. While this may appear to conflict with our previous findings that Alu expression is unlikely to be primarily a side product of Pol-II gene transcription, this leaves the possibility open that Alu transcripts arise from Pol-II activity independent from gene transcription. However, it seems more likely that Alu RNAs arise from different modes of transcription simultaneously, as also suggested by Conti et al. (2015) and Zhang et al. (2019). This is of particular interest regarding differential expression analyses of the past that utilized Pol-II inhibition, as it is common practice to rely on SINEs as a non-Pol-II-dependent negative control group in such experiments. Alus are the most common SINE in the human genome (Cordaux and Batzer 2009), which may call the results of these studies into question. Regarding our own results, the small but not negligible fraction of seemingly upregulated Alu and mRNA transcripts may be explained by inhibition of nuclear export (Bahar Halpern et al. 2015; Zander et al. 2016). To achieve noticeable differential expression of Alu elements under α-amanitin inhibition of Pol-II, an incubation time of 8h was required (see Methods). This exposes the cells to extreme levels of stress, as their protein metabolism is breaking down. This, in turn, may lead to the retention and accumulation of RNAs that are not related to shock response in the nucleus. As our DTA does not differentiate between nuclear and cytoplasmic RNA fractions, this could appear as apparent upregulation. Alternatively, protein translation inhibition, especially of proteins related to mRNA degradation (such as deadenylases or Xrn1), could be involved (Orban and Izaurralde 2005; Dobin et al. 2016). Changes in protein level, even after extended exposure, seem to be an uncommon effect of α-amanitin treatment, according to Flume and Stirpe (1966) and Stirpe and Flume (1967). A reduction in protein abundance or concentration would approximately lead to a proportional decrease in degradation rates. Hence, it would merely lead to a proportional shift in total expression levels. The reduction in the protein concentration might therefore lead to compensatory effects, which would make a quantitative, but not a qualitative change. Regarding Fig. 3e and f it should also be noted that an average l2fc of $-0.35 \pm 0.99$ for mRNAs may appear small when considered individually, especially in the 2D density heatmap, where strongly expressed genes do appear to be affected more, as random measurement noise is visually overrepresented and optically dominates the plot. However, calculating the overall differential expression shows a global reduction in mRNA levels by a factor of 3 (l2fc of $-1.64$), comparable to, for example, the effect of heat shock response (Mahat et al. 2016).

One influential position resulting from our GLM analysis of all Alu sequences is of particular note (see Fig. 5). Position 79 lies within the B box of the Pol-III promoter sequence and is most sensitive to insertions, which bring it closer to the consensus sequence of the tRNA promoter: GWTTCRANC (Paolella et al. 1983). This suggests that the promoter plays a role in Alu transcription, further substantiating the assumption that Alu RNAs arise from various methods of transcription.

It is remarkable that our new method employing de Bruijn graph k-mers does detect statistically significant and biologically relevant sequence motifs when previous attempts using regular de novo motif search did not (Zhang et al. 2019). We attribute this to some extent to the tuning of the parameter k (see Methods). We found that with a k either smaller or larger than 9, the high sequence similarity of Alu elements quickly leads to problems. With a k smaller than 9, many k-mers are possessed by all or almost all Alu elements, while with a k larger than 9, a high number of k-mers are unique to a few or even a single Alu sequence. In both unfavorable cases, a majority of k-mers are ruled out simply by the selection of k. This might appear as a limitation of our method, but in fact, it is an intrinsic limitation shared by all continuous-motif-based methods. Our method merely makes this problem noticeable and offers a solution through parameter tuning. We also see the potential for our approach to be applied to other data sets in the future.

Taken together, we present a compendium of results regarding the RNA metabolism of Alu elements. Our analyses affirm that older Alu families are more strongly expressed than younger Alu families, but the stability of Alu transcripts appears to be close to that of mRNA. Furthermore, we find evidence for Alu transcripts originating both from Pol-II and Pol-III activities, but no evidence for Alu transcripts being side products of Pol-II gene transcription. Finally, we have identified a list of sequence features that influence Alu transcription and are thus targets for further investigations and developed a novel method to test the statistical significance and biological relevance of de Bruijn graph k-mers.
Data availability

The sequencing data used in this publication have been deposited in NCBI’s Gene Expression Omnibus (Edgar et al. 2002) and are accessible through GEO Series accession number GSE75792 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75792) for the data described in Schwalb et al. (2016), and through GEO Series accession number GSE185485 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE185485) for the data relating to the inhibition of RNA Pol-II by α-amanitin. GSE75792 was used for all non-interventional analyses, while GSE185485 was used for the differential expression analysis.

Supplemental material is available at G3 online.

Acknowledgment

We thank the Regional Computing Center of the University of Cologne (RRZK) for providing computing time on the DFG-funded (funding number: INST 216/512/1FUGG) High Performance Computing (HPC) system CHEOPS as well as support.

Funding

TB is partly funded by the Sonderforschungsbereiche (SFB) 680 grant by the Deutsche Forschungsgemeinschaft (DFG).

Conflicts of interest

None declared.

Literature cited


Orban TL, Izaurralde E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA. 2005;11(4):459–469.


Communicating editor: L. M. Steinmetz