

RESEARCH

Open Access



Transcriptome-wide effects of inverted SINEs on gene expression and their impact on RNA polymerase II activity

Mansoureh Tajaddod^{1†}, Andrea Tanzer^{3†}, Konstantin Licht^{2†}, Michael T. Wolfinger^{2,3}, Stefan Badelt³, Florian Huber^{1,4}, Oliver Pusch², Sandy Schopoff¹, Michael Janisiw², Ivo Hofacker³ and Michael F. Jantsch^{2,5*}

Abstract

Background: Short interspersed elements (SINEs) represent the most abundant group of non-long-terminal repeat transposable elements in mammalian genomes. In primates, Alu elements are the most prominent and homogenous representatives of SINEs. Due to their frequent insertion within or close to coding regions, SINEs have been suggested to play a crucial role during genome evolution. Moreover, Alu elements within mRNAs have also been reported to control gene expression at different levels.

Results: Here, we undertake a genome-wide analysis of insertion patterns of human Alus within transcribed portions of the genome. Multiple, nearby insertions of SINEs within one transcript are more abundant in tandem orientation than in inverted orientation. Indeed, analysis of transcriptome-wide expression levels of 15 ENCODE cell lines suggests a *cis*-repressive effect of inverted Alu elements on gene expression. Using reporter assays, we show that the negative effect of inverted SINEs on gene expression is independent of known sensors of double-stranded RNAs. Instead, transcriptional elongation seems impaired, leading to reduced mRNA levels.

Conclusions: Our study suggests that there is a bias against multiple SINE insertions that can promote intramolecular base pairing within a transcript. Moreover, at a genome-wide level, mRNAs harboring inverted SINEs are less expressed than mRNAs harboring single or tandemly arranged SINEs. Finally, we demonstrate a novel mechanism by which inverted SINEs can impact on gene expression by interfering with RNA polymerase II.

Keywords: Alu elements, SINE, Double-stranded RNA, Gene regulation, RNA editing, Transcription, RNA Pol II, ADAR

Background

The non-long-terminal repeat (non-LTR) family of short interspersed elements (SINEs) comprises the numerically largest family of repetitive elements in the mammalian genome. SINEs are unusual in that they are found enriched in gene-rich regions and are often located in transcribed regions of genes [1]. Within genes, SINEs are mostly located in introns and untranslated regions

(UTRs). In rare cases, however, SINEs can reside within coding regions of genes, where they contribute to the formation of novel gene- or splicing variants [2].

SINEs depend on long interspersed elements (LINEs) for their transposition [3]. Transcription, and thus transposition, of SINEs is epigenetically repressed. Still, recent reports have shown that SINEs can transpose at a surprisingly high rate and thereby contribute significantly to genome variation between individuals as well as to somatic variation within individuals [4]. Moreover, exogenous factors such as heat-shock stress can boost transcription of SINEs [5].

SINEs can be of different origin and are rapidly evolving. Hence, SINEs of different species exhibit considerable heterogeneity [6]. In primates, however, a very abundant

* Correspondence: Michael.Jantsch@meduniwien.ac.at

†Equal contributors

²Department of Cell and Developmental Biology, Medical University of Vienna, Schwarzschanerstrasse 17, Vienna A-1090, Austria

⁵Department of Cell and Developmental Biology, Medical University of Vienna, Center of Anatomy and Cell Biology, Schwarzschanerstrasse 17, Vienna A-1090, Austria

Full list of author information is available at the end of the article



and surprisingly homogeneous population of SINEs has evolved, known as the Alu family of SINEs. Alu elements originate from a duplication of the 7SL RNA of the signal recognition particle and, consequently, are about 300 nucleotides in length. Rodent B1 elements, in contrast, are derived from a single 7SL RNA and are only about 150 nucleotides in length [7, 8]. Primate Alus are divided into several closely related subfamilies that apparently evolved in three timely distinct expansion waves while rodent B1 elements are more heterogeneous in sequence [9]. Besides changing the genomic landscape [1], SINEs can have a dramatic impact on the transcriptome by several means: first, SINE transcripts can impair polymerase II activity, thus repressing transcription at a global scale [5]. Second, primate Alu elements are very abundant and can be found in long noncoding RNAs and mRNAs. Thus, Alu elements containing long noncoding RNAs potentially base pair with Alus of inverted orientation located in mRNAs. Some, but not all, of these base-paired RNAs can be bound by the double-stranded RNA-binding protein STAUFEN. Binding of STAUFEN, in turn, may affect the stability of the bound RNAs [10]. Third, antisense SINEs have also been shown to be able to stimulate translation of mRNAs in a stress-dependent manner [11]. Lastly, insertion of SINEs can alter epigenetic marks and thereby influence the expression of nearby RNAs [12].

Multiple SINEs present in 3' UTRs base pair with each other if organized in inverted orientation. Frequently, such inverted SINEs (*i*SINEs) are substrates of RNA editing by adenosine deaminases that act on RNA (ADARs). ADARs bind double-stranded RNAs and, hence, the base-paired regions formed by *i*SINEs serve as substrates for these enzymes [13–16]. RNAs harboring double-stranded *i*SINEs were first reported to repress gene expression via nuclear retention [17, 18]. However, whether inosines, the product of adenosine deamination, trigger nuclear retention has been a matter of debate [19]. Moreover, inverted intramolecular base-paired Alu elements are bound by the protein STAUFEN, which has been proposed to regulate their translation [20, 21].

Studies in *Caenorhabditis elegans* and human cells have shown that mRNAs with double-stranded structures in their 3' UTRs are edited and repressed in their expression. However, these mRNAs are exported from the nucleus and are associated with ribosomes but are translationally repressed [19, 22, 23]. Thus, double-stranded structures formed by inverted SINEs may have different effects on individual RNAs based on cellular context or unknown factors.

Since it has been shown that *i*SINEs modulate gene expression through different pathways, we were interested to determine their impact on gene expression at a genome-wide level. Using available transcriptomic data we show that the presence of SINEs in 3' UTRs correlates

with reduced gene expression. This effect is strongest when SINEs are found in inverted orientation. In agreement with this finding, inverted SINEs are found at lower rates than tandemly arranged pairs of SINEs in the human genome, suggesting inverted insertions of SINEs are disfavored. To gain insight into the possible mechanism of *i*SINE-mediated mRNA repression, we used mouse embryonic fibroblasts (MEFs) of different genetic background and tested for their ability to repress reporter genes harboring *i*SINEs. These assays show that neither RNA editing nor binding by STAUFEN1 are underlying causes for the observed reduced gene expression of *i*SINE-containing reporters. Importantly, we show that inverted repeats impair gene expression at the RNA level, apparently by repressing transcriptional elongation. Thus, we present a novel mechanism for *i*SINE-triggered repression of gene expression.

Results

SINEs in inverted orientation are underrepresented in annotated genes

Several reports have indicated that multiple SINEs located in inverted orientation in individual mRNAs can negatively affect gene expression; however, several molecular mechanisms were proposed as the underlying cause [17, 21, 23]. Nonetheless, if *i*SINEs repress gene expression, we wondered whether they would be found at the same frequencies as tandemly arranged, duplicated SINEs (*d*SINEs) throughout the genome. We therefore analyzed the abundance and orientation of SINE insertions in the genome (Fig. 1a). For this analysis we used the ENCODE description of genic (transcribed) and intergenic partitions of the genome [24]. On average, Alus are found to be slightly more abundant in genic than intergenic regions. Within genic regions, Alus are less abundant in exons but accumulate in 3' UTRs and non-coding RNAs (Fig. 1a). The median distance between Alus is 748 nucleotides in the human genome and 68.5 % of all Alus fall within this range, indicating a tendency of Alus to cluster close to each other. Even more strikingly, 50 % of all Alus form clusters with their partner Alus within 300 nucleotides (Fig. 1b). Also, such clusters are slightly more abundant in genic than in intergenic regions. To clearly determine the relative position of Alus relative to each other and therefore to allow for a distinction between *i*SINEs and direct *d*SINEs, we determined the fraction of Alus with only a single second Alu within 300 nucleotides. About 21 % of Alus were arranged in such pairs and, again, the fraction of pairs was similar in genic and intergenic regions (Fig. 1b). When the relative arrangement of Alus in pairs was analyzed, however, 13 % were found arranged in *d*SINEs while only 7 % were found in an *i*SINEs configuration (Fig. 1b). Within *i*SINEs the head-to-head (*hi*SINE) and tail-to-tail (*ti*SINE) configuration was essentially the

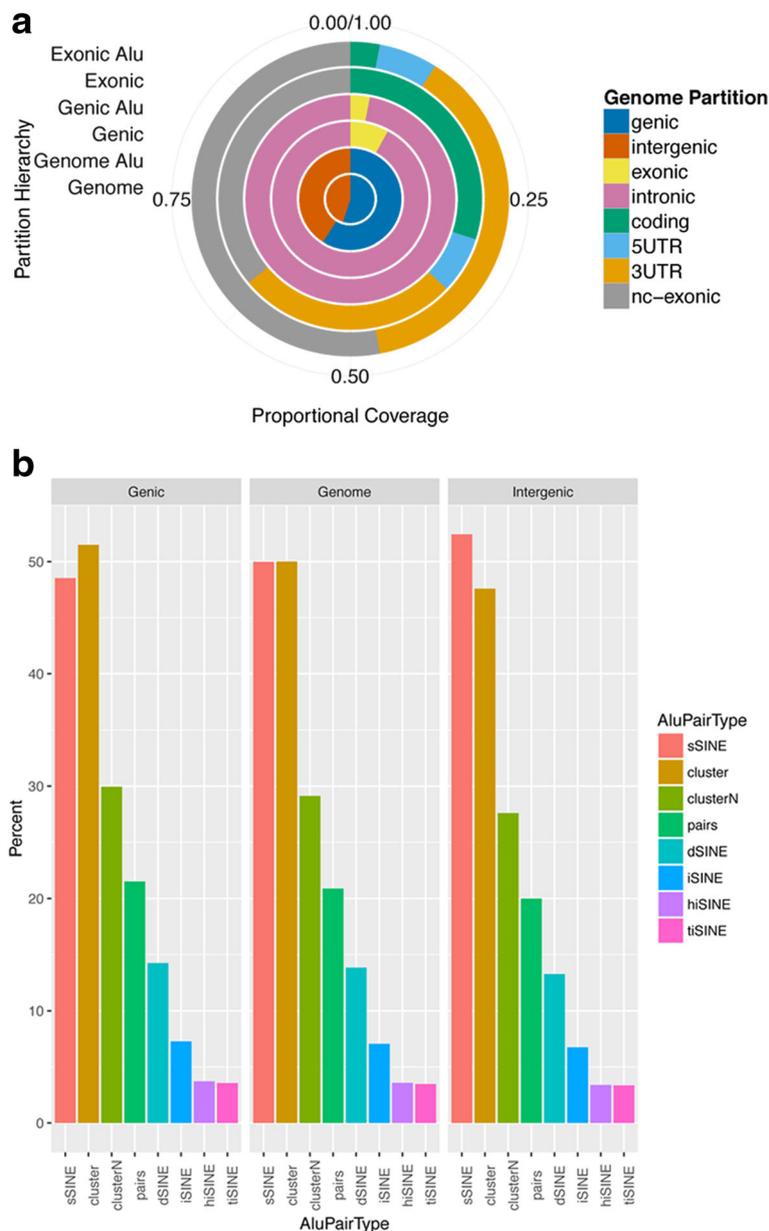


Fig. 1 *i*SINEs are less abundant than *d*SINEs in the human genome. **a** The proportional distribution of genome and “Aluome” to genic, intergenic, exonic and intronic partitions and different types of exons are shown. Alu elements are almost equally distributed to genic and intergenic regions. Within genic regions, however, a strong accumulation in non-coding regions, such as introns, 3’ UTRs, or non-coding RNAs, can be observed. **b** Comparison of the number of Alus in the genome in genic and intergenic regions. A large fraction of Alus accumulates in clusters, which are defined as a set of neighboring Alus that are at most 300 nucleotides apart from one another. Single SINEs (sSINE) thus have no other Alu in their vicinity. Pairs are clusters of size 2 and are further divided into *d*SINEs (direct tandem SINEs) and *i*SINEs (inverted SINEs). Based on their relative orientation to one another, *i*SINEs are grouped into *ti*SINEs (tail-to-tail) and *hi*SINEs (head-to-head). Clearly, *i*SINEs are less abundant than *d*SINEs

same (Fig. 1b). This suggests that either inverted *i*SINEs may lie under negative selective pressure or, alternatively, that tandemly arranged SINEs might be favored. In fact, both negative selection against inverted insertions of *i*SINEs and positive selection for tandemly arranged *d*SINEs have been discussed and were explained

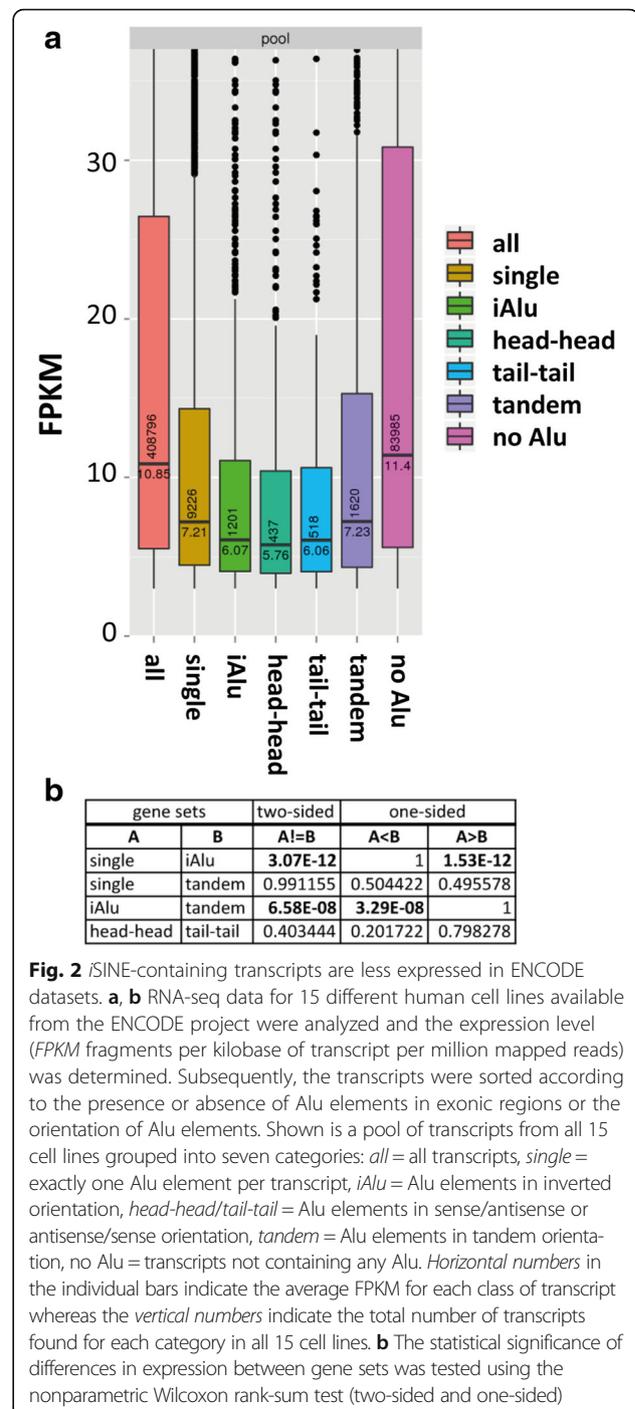
by recombination and deletion of *i*SINEs [25, 26] and favored tandem insertion mechanisms [27]. However, the fact that the ratio of *i*SINEs to *d*SINEs is roughly identical in genic and intergenic fractions of the genome indicates that *i*SINEs may not be selected against via a transcription-related process.

*i*SINE-containing transcripts are less expressed in ENCODE datasets

We next asked whether the presence of a single or multiple SINEs in transcripts can affect RNA levels. To comprehensively address this question, we analyzed ENCODE RNA-seq data from 15 different human cell lines and compared expression levels by counting fragments per kilobase of transcript per million mapped reads (FPKM) values. Of these, only those covered by at least 3 FPKM in each cell line were further considered. The resulting dataset was divided into transcripts not harboring any Alu elements (*noSINE*), harboring a single Alu (*1SINE*), exclusively tandemly duplicated Alus (*dSINE*), multiple Alus of which at least one adjacent pair was in the inverted orientation (*iSINE*), pairs of Alus in exclusively inverted head-to-head configuration (*head_head*) or tail-to-tail configuration (*tail_tail*) (Fig. 2a). The strongest difference in FPKM levels was observed when comparing transcripts with no Alu (*noSINE*) or any Alu, indicating that the presence of Alu elements, per se, negatively affects gene expression or is only tolerated in genes with low expression. Of interest to our study, however, was the comparison of tandem (*dSINES*) and inverted insertions of Alu elements (*iSINES*, *iAlu*). As expected, a clear and highly significant reduction in FPKM was observed when *iSINE*-containing transcripts were compared with those containing *dSINES* (Fig. 2b). Although it had been shown that *iSINES* can reduce gene expression, our data, for the first time, show a transcriptome-wide reduction of *iSINE*-containing transcripts.

Inverted SINEs in 3' UTRs modulate the RNA expression of reporter genes

To determine the impact of inverted SINEs on gene expression experimentally, we tested two different 3' UTRs each harboring two Alu elements in inverted orientation. Here, we picked the 3' UTR of the Nicotin (*Nicn1*) gene, which had already been proven to interfere with gene expression when fused to a green fluorescent protein (GFP)-expressing reporter [17]. As a second 3' UTR harboring two inverted SINEs we picked the 3' UTR of the InaD-like gene (*Inadl*). The 3' UTR of *Nicn1* contains an AluSp1 and an AluSp2 element in a tail-to-tail configuration spaced only 70 bp apart (the first Alu is oriented in sense "+" while the second Alu is oriented in antisense or complementary "c" orientation). The two Alu Sp elements in *Nicn1* are 81 % identical. The 3' UTR of *Inadl* contains an AluSx and an AluSg element in a head-to-head configuration; these are 79 % identical and spaced about 1000 nucleotides apart. As a control, one of the two SINEs was removed, leaving a single SINE in the construct (*1SINE*). As an additional control, one of the two SINEs was inverted, giving rise to a duplicated SINE (*dSINE*) (Fig. 3a, b). The Alu element-



containing UTRs were cloned downstream of the open reading frame (ORF) of firefly luciferase in pmirGLO. This vector simultaneously expresses renilla and firefly luciferase to allow easy quantification of changes in gene expression using a dual luciferase assay.

To check the impact of *iSINES* on reporter gene expression, the SINE-containing vectors were transfected into U2OS cells and RNA and protein levels were

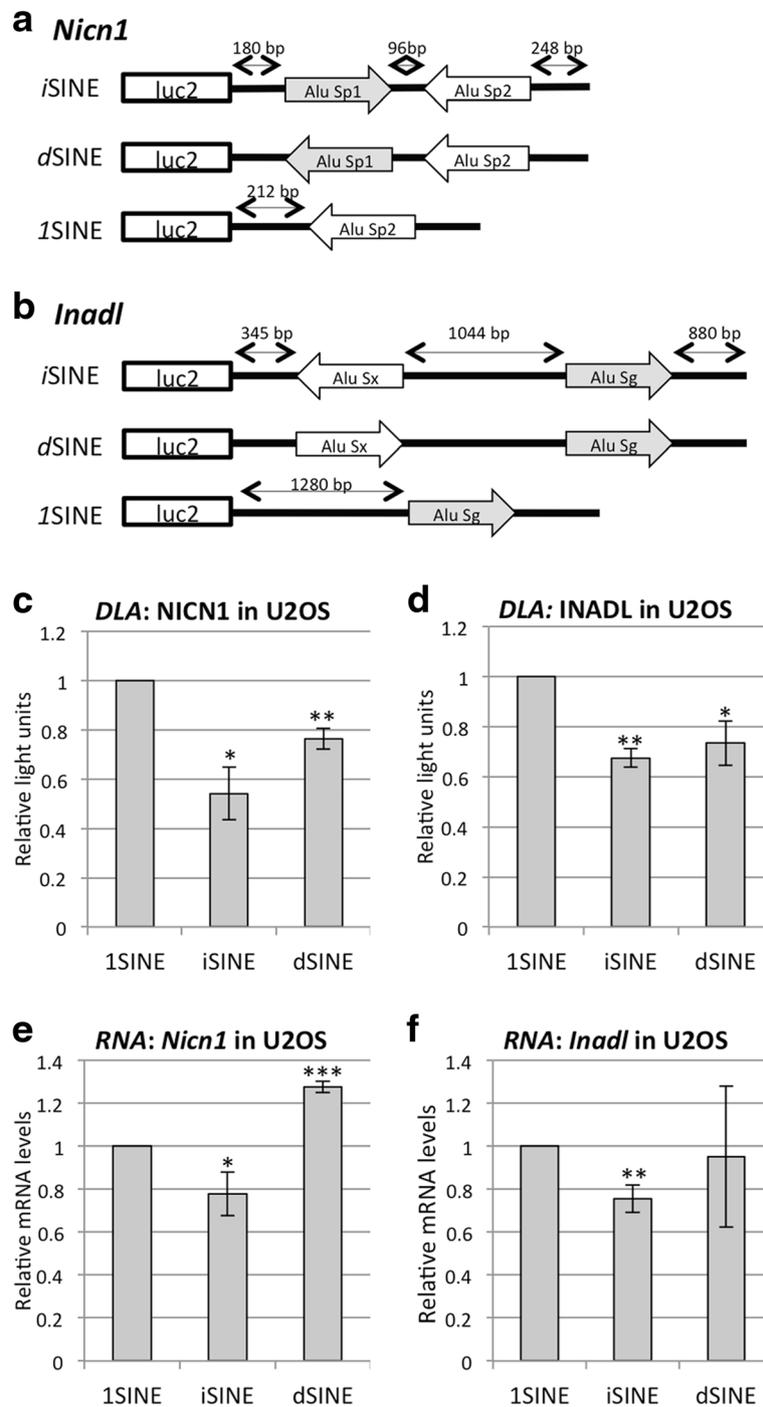


Fig. 3 *iSINEs* can repress reporter gene expression and RNA levels experimentally. The *iSINE*-containing 3' UTRs of **a** the *Nicn1* and **b** *Inadl* genes were inserted downstream of the firefly Luciferase ORF. As controls, one of the Alu elements was flipped to make a duplicated SINE (*dSINE*) and as a second control one of the ALUs was removed (*1SINE*). Arrows show SINE orientation and absolute distances are indicated. Reporter genes harboring different SINEs derived from **c** the *Nicn1* or **d** the *Inadl* 3' UTRs were transfected into U2OS cells and gene expression was determined using a dual luciferase assay (*DLA*). Total RNA was extracted from transfected cells with **e** the *Nicn1* or **f** the *Inadl* constructs and mRNA levels were measured using reverse transcription followed by quantitative PCR of total cDNA. Standard deviation is indicated by error bars. Asterisks indicate *p* values calculated with Student's *t*-test: **p* < 0.05, ***p* < 0.005, ****p* < 0.0005

measured using a dual luciferase assay. *Nicn1* and *Inadl* *i*SINEs showed a strong reduction in protein levels when compared to *IS*SINE controls (Fig. 3c, d). We also observed a reduction in protein levels of *d*SINE-containing reporter genes. However, the reduction was not as strong as for the respective *i*SINE constructs (Fig. 3c, d). Next, using quantitative PCR (qPCR), we quantified the mRNA levels of all SINE-harboring constructs. We observed that *Nicn1* and *Inadl* *i*SINEs led to a 35–45 % reduction in mRNA levels. Interestingly, *d*SINEs in 3' UTRs did not significantly reduce the mRNA levels compared to *IS*SINE controls (Fig. 3e, f). These data indicate that the presence of *i*SINEs in 3' UTRs led to a significant reduction in expressed RNA and protein levels.

The quality of double-stranded structures affects gene repression

The *i*SINEs in all 3' UTRs are able to undergo base-pairing interactions as they had been reported to be heavily edited by ADARs [13, 14, 28]. To determine whether the cloned *i*SINEs could form double-stranded structures in the context of the used reporter constructs and in cell lines, reporter constructs were transfected into mouse embryonic fibroblasts (MEFs) and the editing status was determined by sequencing of cDNAs. Indeed, *i*SINEs did become edited in MEFs, suggesting that the predicted double-stranded structures are also formed by the RNAs expressed from the reporter constructs (Additional file 1: Figure S1). Moreover, editing was more pronounced in the *Nicn1* sequencing traces, where 20 sites were found edited well above 50 %. Editing rates in the *Inadl* 3' UTR, in contrast, only reached a maximum of 30 % (Additional file 1: Figures S1a, b). This finding is consistent with the idea that the two more closely spaced SINEs in *Nicn1* are more likely to form a double-stranded structure than the more distantly spaced SINEs in *Inadl*, thereby affecting the extent of editing.

It has been noticed that not all *i*SINE-containing 3' UTRs lead to an equal reduction of gene expression [23]. Since *i*SINEs are able to form double-stranded structures, we reasoned that the reduction in gene repression might be related to the extent and stability of the formed double-stranded structures. To test this hypothesis we created new construct sets with identical, perfectly matching SINEs in inverted orientation. To do so, we picked part of the 3' UTR of the *Znf708* gene, which harbors an Alu Sc and an Alu Sg element spaced 180 bp apart in a head-to-head configuration. The two Alu elements are 77 % identical to each other. We replaced the Alu Sg element with an identical copy of the first Alu Sc element to create a 3' UTR with a perfect inverted SINE (*pi*SINE; Fig. 4a). Thus, these constructs should form a fully base-paired double-stranded

structure. Indeed, the *pi*SINEs in the 3' UTR of *Znf708* led to stronger repression in gene expression compared to the natural *Znf708* *i*SINE (Alu Sc–Alu Sg), supporting the idea that the quality of the formed double-stranded structure affects the strength of gene repression (Fig. 4b).

The observed *i*SINE-dependent repression of gene expression is sequence-dependent

Next, we asked whether the observed phenomenon depends on the Alu sequence or whether a stable secondary structure alone would be sufficient to repress gene expression. Therefore, we designed constructs that mimicked the secondary structure of an *i*SINE but did not share the same homologous sequence. All sequence designs were based on the *Znf708* UTR. Both Alu elements were replaced by artificial repeats while the remaining parts of the UTR were maintained. Using RNAfold, similar folding was predicted for all artificial UTRs and the *Znf708* UTR (Additional file 1: Figure S2). To confirm the folding state we also transfected all constructs into an editing-competent cell line. All constructs exhibited similar editing levels (Additional file 1: Figure S3), suggesting that the *Znf708* UTR as well as the artificial constructs form stable double-stranded structures.

While the original shortened *Znf708* UTR exhibited the expected repression, gene expression was not reduced for any of the artificial *i*SINE constructs compared to the respective *IS*SINE control (Fig. 4b–f). We conclude that, besides the structure, the sequence of Alu elements plays a role in the observed effect and constructs only mimicking the secondary structure cannot reproduce the gene repression. Therefore, the observed phenomenon appears to be specific for SINEs.

*i*SINE-mediated gene repression is species-independent

Next, we went on to test if we can detect *i*SINE-mediated gene repression in other species as well. Mice harbor B1 or B2 elements in their genomes, which belong to a SINE family similar to the Alu elements. B1 elements are half the length of Alu elements and are much less conserved than their primate counterparts. Consequently, the double-stranded regions formed between two antiparallel B1 elements are shorter and less extensive [29]. To test the ability of B1 elements to interfere with gene expression, the human Alu elements of the *Znf708* 3' UTR were replaced by two B1 elements of the mouse *Car5b* gene (Fig. 4g). The inverted B1 elements of *Car5b* affected luciferase expression but only to a minor extent, showing less than 20 % repression (Fig. 4h). Since our data show that the extent of the double-stranded structure can influence gene expression, we again stabilized the secondary structure by replacing the second naturally occurring B1 element by an inverted duplication of

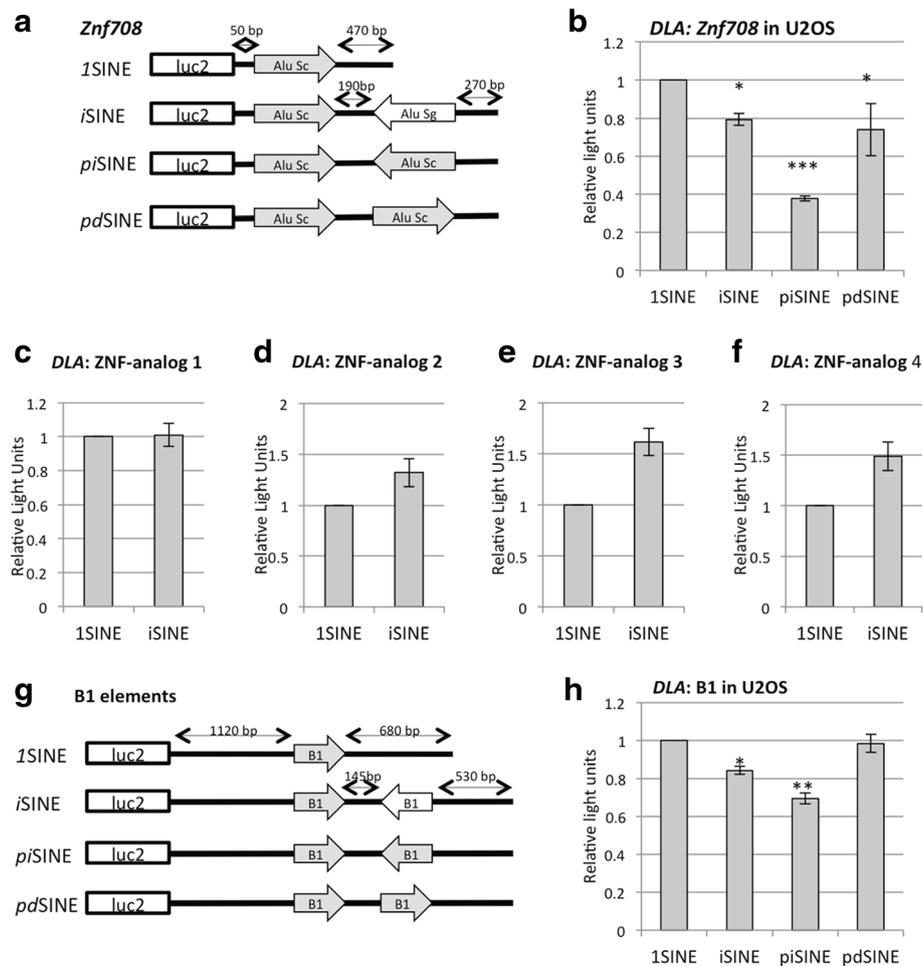


Fig. 4 Repression by *iSINEs* is dependent on secondary structures and sequence in a species-independent manner. **a** The *iSINE*-containing 3' UTR of the *Znf708* gene was inserted downstream of the firefly Luciferase ORF. To generate perfect complementarity, one Alu Sg was replaced by a duplication of the Alu Sc, giving rise to a perfect inverted SINE (*piSINE*). A perfect tandem SINE (*pdSINE*) and *1SINE* were used as controls and made by flipping or deleting the second SINE, respectively. **b** Dual luciferase assays of different SINE configurations derived from the *Znf708* 3' UTR demonstrate that the reduction of gene expression correlates with the extent of double-strandedness. **c-f** To evaluate whether the observed reduction in gene expression is specific for *iSINEs* or dependent on RNA structure alone, UTRs that resemble the secondary structure of the *Znf708* UTR but with different sequence context were designed. The *Znf708* analogues and respective controls were transfected into U2OS cells and gene expression was quantified using a dual luciferase assay. See Additional file 1: Figure S2 for minimum free energy structures of the *Znf708* UTR and the designed *Znf708* analogues. **g** To generate reporter constructs harboring mouse SINEs, B1 elements of the mouse *car5b* gene were used to replace the Alu elements in *Znf708*. **h** The B1-harboring reporter genes were transfected into U2OS cells and a dual luciferase assay was performed after 24 h. Standard deviations are indicated. Asterisks indicate *p* values calculated with Student's *t*-test: **p* < 0.05, ***p* < 0.005, ****p* < 0.0005

the first B1 element, thereby generating a construct with 100 % complementarity (*piSINE*) (Fig. 4g). The resulting formation of a short but perfect double-stranded structure led to a reduction in gene expression of more than 30 %, while the corresponding *pdSINE* had almost no influence on luciferase expression (Fig. 4h). This again demonstrates that the extent of complementarity and thus base pairing influences the strength of reporter gene repression. Taken together, our data indicate that regulation of gene expression by *iSINEs* is species-independent and is a common phenomenon, conserved at least between rodents and primates.

iSINEs repress gene expression independent of editing or other double-stranded RNA-binding proteins

The fact that RNAs expressed from *iSINE*-containing reporter plasmids are edited by ADARs strongly supports the idea that these RNAs form double-stranded structures. Recently, several studies have shown that different double-stranded RNA-binding proteins can repress gene expression by binding to RNA stem-loops [10, 17]. It was also suggested that the presence of inosines in RNAs would lead to nuclear retention, therefore repressing translation of such RNAs [17, 30]. We therefore tested whether proteins that bind double-stranded RNA,

such as ADARs and STAUFEN1, would be required for *i*SINE-mediated gene repression.

The most straightforward way to test for the involvement of genes or proteins is the use of cells derived from adequate genetic knock-out mice. Therefore, to test whether the presence of inosines would lead to nuclear retention of *i*SINEs, we performed experiments in mouse embryonic fibroblasts (MEFs) derived from mice

lacking ADAR1 and ADAR2. In MEFs lacking both active editing enzymes we still observed a significant reduction in protein and RNA levels, indicating that lack of editing does not interfere with *i*SINE-mediated gene repression (Fig. 5a–d). To test more directly for the suggested nuclear retention of *i*SINEs [17] we performed RNA-FISH using a firefly antisense probe to detect the RNA transcribed from *l*SINE, *d*SINE, and *i*SINE

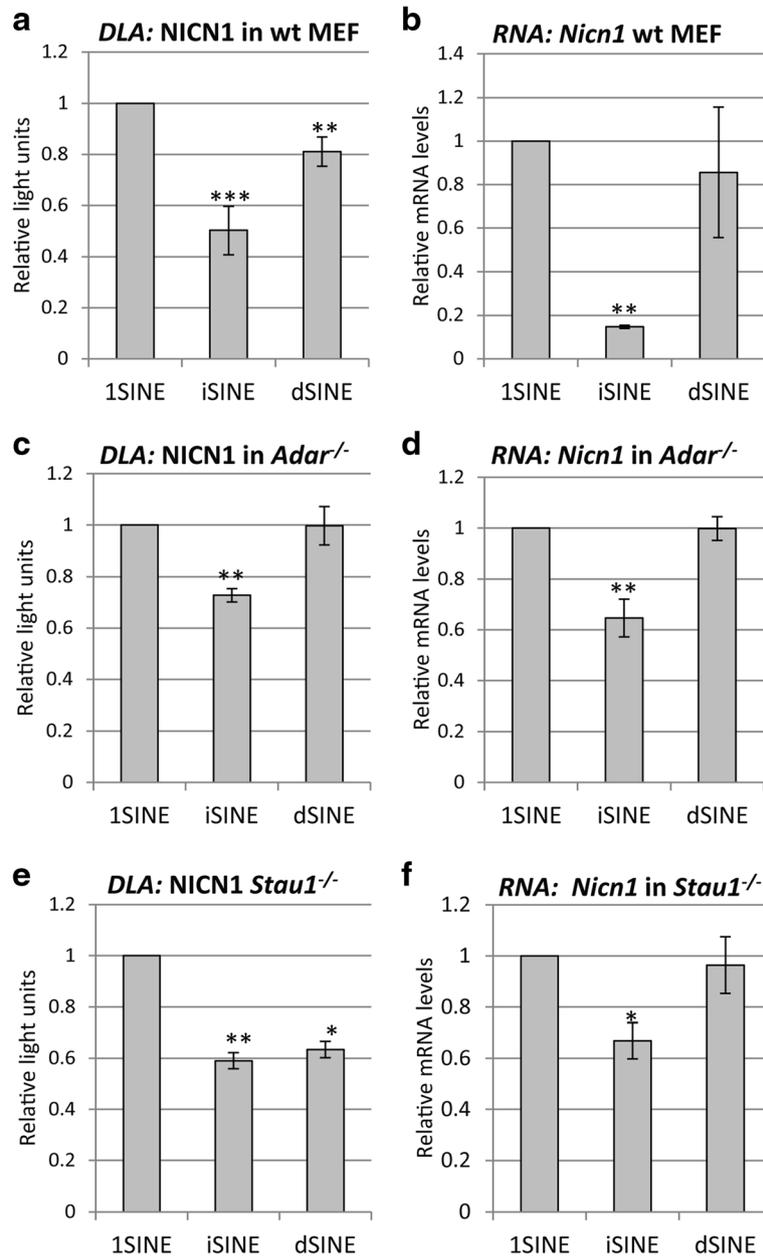


Fig. 5 *i*SINE-mediated gene repression is independent of ADARs and STAUFEN1. **a** To test whether human *i*SINEs lead to gene repression in mouse cells, *Nicn1* constructs were transfected in wild-type MEFs and luciferase activity was detected after 24 h. **b** mRNA levels of SINE-containing reporter genes were detected using RT-qPCR. **c, d** *Nicn1*-containing constructs were transfected in *Adar1*^{-/-}/*Adar2*^{-/-} mouse cells and after 24 h **c** protein and **d** RNA levels were measured. **e, f** *Stau1*^{-/-} MEF cells were also transfected and the **e** protein expression and **f** RNA levels were measured. Standard deviation is indicated. Asterisks indicate *p* values calculated with Student's *t*-test: **p* < 0.05, ***p* < 0.005, ****p* < 0.0005

constructs. At the same time, the encoded protein was detected using an antibody directed against the firefly luciferase protein (Additional file 1: Figure S4). This experiment showed clearly that firefly protein can be detected for all three constructs and that the RNA is readily exported to the cytoplasm. Thus, our data are in agreement with previous experiments that had reported efficient export of edited *iSINE*-containing RNAs from the nucleus to the cytoplasm [22].

In the next step we checked whether STAUFEN1 is involved in this phenomenon. As mentioned, Alu elements had been shown to interfere with gene expression by triggering STAUFEN-mediated RNA decay [10, 31]. We therefore tested the effect of *iSINEs* also in cells lacking STAUFEN1 protein. Again, even in the absence of STAUFEN1, a solid reduction in gene expression was triggered by the presence of *iSINEs* in 3' UTRs (Fig. 5e, f).

RNA transcribed from *iSINE*-containing constructs might get cleaved by DICER1 or DROSHA and lead to the observed reduction in mRNA levels. We therefore tested the effect of *iSINEs* in *Dicer1*^{-/-} MEFs. Interestingly, even in the absence of DICER, *iSINEs* led to reduced gene expression (Additional file 1: Figure S5b). Similarly, in cells with stable shRNA-mediated knockdown of DROSHA where *Drosha* mRNA levels are reduced by 50 %, *iSINE*-harboring constructs showed reduced expression (Additional file 1: Figure S5c). Lastly, double-stranded RNA can activate the double-stranded RNA-dependent kinase PKR, which phosphorylates eIF2-alpha leading to repression of translation [32]. However, the presence of *iSINEs* led to reduced gene expression also in MEFs lacking PKR activity [33] (Additional file 1: Figure S5d). Taken together, this indicates that the *iSINE*-mediated reduction in gene expression occurs independent of RNA editing and the double-stranded RNA binding or cleaving proteins STAUFEN, PKR, DICER, or DROSHA.

RNA polymerase II density decreases towards 3' UTR ends

Clearly, our data demonstrate that *iSINEs* in 3' UTRs lead to a reduction in mRNA levels. Moreover, the observed effect is independent of several known double-stranded RNA-binding proteins. Thus, we reasoned that reduced RNA levels may be the result of reduced RNA stability or a consequence of *iSINEs* interfering with transcription.

To test whether *iSINE*-containing RNAs have reduced stability, we compared the mRNA half-life of the *Nicn1* 3' UTR harboring an *iSINE* with control RNA (*ISINE*). For this purpose, mRNA transcription was blocked using Actinomycin D treatment, RNAs were collected at regular time intervals and, subsequently, mRNA levels were determined by real-time qPCR of cDNAs [34]. Our data show that *Nicn1 iSINE* mRNA was degraded as rapidly as *ISINE* mRNA (Fig. 6a). This observation suggests that

the reduction of mRNA levels is not caused by low mRNA stability.

All *iSINEs* investigated here are located in the 3' UTRs and might therefore affect polyadenylation of mRNAs and therefore maturation, translation, and turnover. To test this, we increased the distance between the stop codon and the first Alu by 950 nucleotides and the distance between the last Alu and the polyadenylation signal by 270 nucleotides in the *Znf708* constructs (Additional file 1: Figure S6a). Still we observed a similar decrease in protein levels as seen with shorter *Znf708* constructs (Additional file 1: Figure S6b). Lastly, we tested the length of the poly(A) tail in *ISINE*, *iSINE*, and *dSINE* constructs. All constructs showed the same poly(A) tail length (Additional file 1: Figure S6c). Taken together this suggests that there is no crosstalk between the polyadenylation signal and *iSINEs*.

Free Alu elements have been reported to interfere with the activity of RNA polymerase II (Pol II) [5]. We therefore wondered whether *iSINES* might interfere with Pol II function and thus performed RNA Pol II chromatin immunoprecipitation (ChIP) assays to determine the Pol II density along DNA in *iSINE* and control constructs [35]. *iSINE*-, *ISINE*-, and *dSINE*-harboring constructs were transfected and the density of Pol II was determined by ChIP. The DNA co-precipitated with a Pol II antibody was quantified by qPCR using amplicons for separate regions of the reporter genes (Fig. 6b). Our results show that Pol II density decreased downstream of the *Nicn1 iSINE* compared to the corresponding *ISINE*. However, no significant reduction in Pol II density was observed downstream of *dSINES* (Fig. 6c). We also checked Pol II density in the construct forming perfect double-stranded regions, the *Znf708 piSINE*. Here we also observed a significant reduction of Pol II density near the poly(A) site, corresponding well with the strong reduction of gene expression observed for the *Znf708 piSINE* (Fig. 6d). Again, no reduction in Pol II density was seen downstream of the *dSINE* construct.

To test whether this phenomenon could also be observed in a genomic context, we selected several endogenous genes containing either a single Alu in the 3' UTR or two Alus in inverted orientation (Fig. 7). As for the reporter constructs, we amplified different regions of the genes by qPCR after Pol II ChIP (Fig. 7a). Here a decrease in Pol II occupancy was also observed towards the 3' end for most of the genes containing inverted Alu elements, whereas the *ISINE*-containing control genes showed no particular trend (Fig. 7b–d). Taken together, our results demonstrate that the presence of *iSINES* interferes with transcription and that the strength of interference seemingly correlates with the stability of the double-stranded structure formed. Therefore, reduced RNA levels, and the subsequently reduced protein levels,

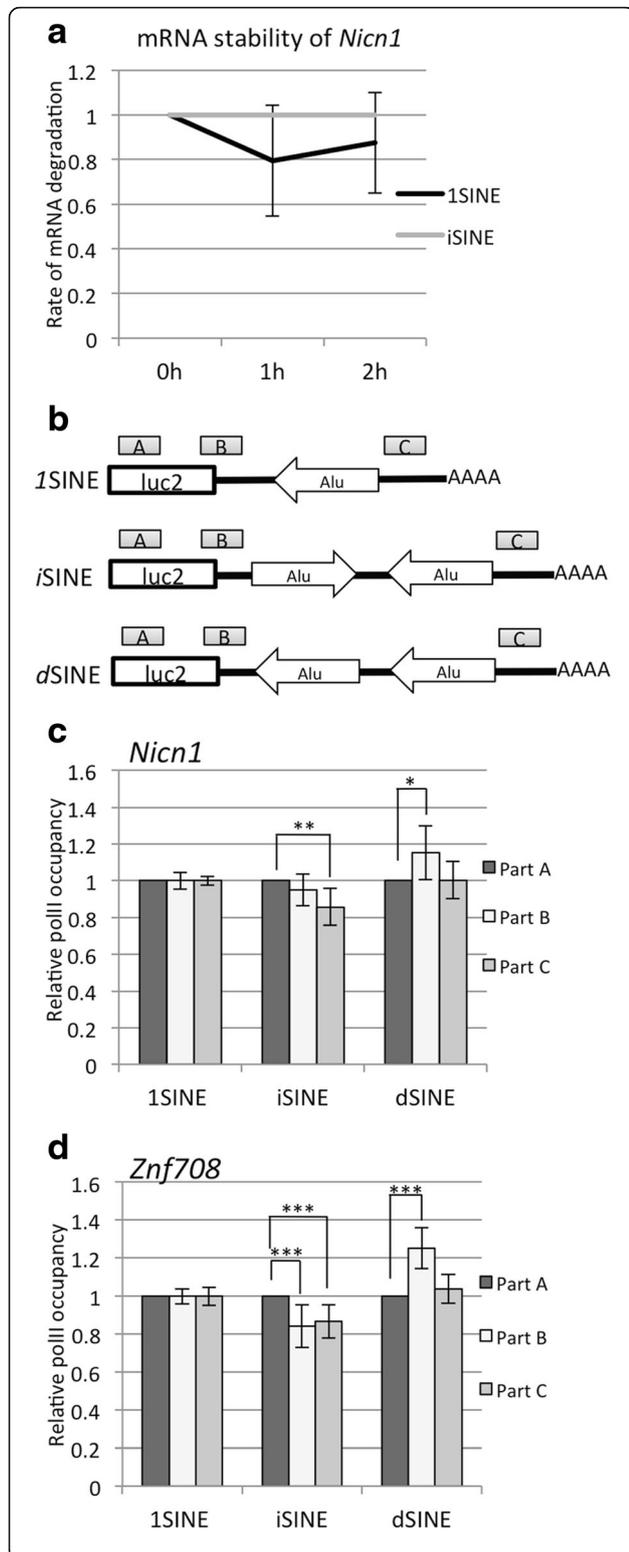


Fig. 6 *i*SINEs do not destabilize mRNA but interfere with RNA polymerase II (*polII*). **a** mRNA transcription was blocked using Actinomycin D. Subsequently, mRNAs were collected 0, 1, and 2 h after transcriptional inhibition and mRNA levels were determined by RT-qPCR. **b** The Pol II-immunoprecipitated DNA was analyzed by real time PCR. Three different regions of the reporter gene—the 5' coding region (A), stop codon (B), and near poly(A) signal (C)—were amplified. The amplicons are shown as grey boxes. Reporters harboring the **c** *Nicn1* 1SINE, *i*SINE, and *d*SINE or the **d** *Znf708* 1SINE, *i*SINE, and *pd*SINE were transfected in U2OS cells and the Pol II density along the genes was measured. Clearly, Pol II density decreases downstream of stable *i*SINEs. Error bars indicate standard deviation. Asterisks indicate *p* values calculated with Student's *t*-test: **p* < 0.05, ***p* < 0.005, ****p* < 0.0005

may be the result of cumulative reduced RNA production of *i*SINE-containing genes. These data are also in agreement with our transcriptome-wide analysis that demonstrated reduced RNA levels for *i*SINE-containing transcripts.

Discussion

SINEs are the numerically most abundant class of transposable elements in the genomes of higher metazoa and thus have a profound impact on the genomic landscape in these organisms. Most importantly, SINEs are frequently found within genes and are therefore transcribed as part of intronic sequences, UTRs, or even exons [36–38].

Recent research has demonstrated that the presence of SINEs in inverted orientation (*i*SINEs) in transcribed regions of genes can have a significant influence on gene expression through different proposed molecular mechanisms, including translational inhibition and nuclear retention [17, 21, 23].

We therefore studied whether *i*SINE-containing genes were found with equal frequency as *d*SINEs. Indeed, our data show that *i*SINE insertions into annotated transcripts are less abundant than *d*SINE insertions. Thus, either insertion of SINEs in inverted orientation is selected against as suggested by some [26] or, alternatively, duplicated *d*SINE insertions may be selected for, as suggested by others [27]. Clearly, although *i*SINEs have a negative impact on gene expression, they can be found in a considerable number of genes. This suggests that some *i*SINEs may have acquired regulatory functions that may have beneficial effects. Recently, it was shown that site-selective editing events are significantly increased in the vicinity of *i*SINEs [39]. Since ADAR enzymes bind to secondary structures, it has been suggested that *i*SINEs can act as a bait for ADAR enzymes and induce editing at sites located several hundred nucleotides from the Alu elements in the surrounding transcript [40]. Moreover, Ricci and colleagues [20] showed that *i*SINEs represent a major group of binding targets for STAUFEN1 and overexpression of STAUFEN1 mildly increases nucleocytoplasmic export

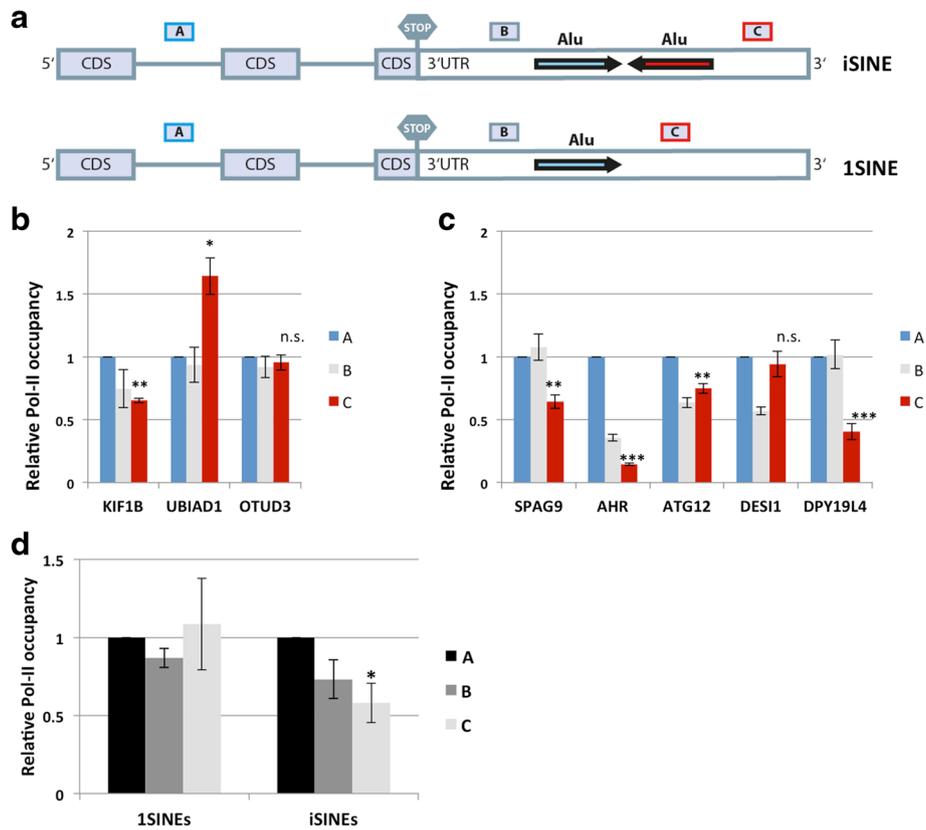


Fig. 7 Pol II occupancy decreases downstream of endogenous *i*SINEs. **a** To test for Pol II occupancy we performed ChIP assays for three endogenous genes carrying a single *1*SINE in their 3' UTR and five endogenous genes carrying a paired *i*SINE in their last UTR exon. Pol II ChIP was performed and the co-precipitated DNA was quantified by qPCR. We amplified three different regions: region A, about 10 kb upstream of the single or inverted SINE; region B, upstream of the *i*SINE or *1*SINE in close proximity; and region C, downstream of the *i*SINE or *1*SINE. Whereas for the *1*SINEs (**b**) we observed both a decrease and increase in Pol II occupancy, the Pol II occupancy clearly drops for four out of five genes carrying an *i*SINE (**c**). The average and the standard error of the mean (SEM) calculated from six biological replicates are plotted. **d** The average and SEM of the three *1*SINE- and five *i*SINE-containing genes. Asterisks indicate *p* values (region A to region C) calculated with Student's *t*-test: **p* < 0.05, ***p* < 0.005, ****p* < 0.0005; *n.s.* not significant

of the respective mRNA. In general, different RNA binding proteins, including ADAR1, p54nrb, STAU-FEN1, and PKR, have been shown to interact with *i*SINEs and subsequently affect mRNA modification, nuclear retention, mRNA transport, and translational repression, respectively [15, 21, 39, 41]. It seems that *i*SINEs may act as regulatory elements in the mRNA by providing a double-stranded RNA structure, which serves as a platform for double-stranded RNA-binding proteins. Based on the availability of double-stranded RNA-binding proteins in the cell, *i*SINEs might recruit them and trigger different cellular processes.

Capshew and colleagues [23] have suggested that the relative position of *i*SINEs in the 3' UTR can influence the impact of SINEs on gene expression. A minimal distance of 65 nucleotides from the *i*SINE to the stop codon would be required to repress gene expression [23]. However, the *Znf708* *i*SINE is located close to the stop codon (50 bp) and still leads to a significant

reduction in gene expression. Thus, at least in the context of the constructs used by us, we do not observe a position-dependent effect, suggesting that there may be other factors modulating the effects of *i*SINEs on gene expression.

Our data also indicate that the extent of double-stranded structures formed by *i*SINEs influences the strength of gene repression. This fact may explain why not all *i*SINEs in mRNA repress gene expression. Interestingly, artificial 3' UTRs that mimic the secondary structure of an *i*SINE but contain sequences that were not Alu-like did not reduce gene expression. This finding strongly suggests that, besides the formation of a double-stranded structure, the sequence would also be important.

While the impact of *i*SINEs on gene expression has been shown in other studies, our study shows a STAUFEN1-independent effect of *i*SINEs on RNA levels [17, 21, 23]. Similarly, our study also shows that other

RNA-binding proteins that recognize double-stranded RNAs are not responsible for the observed reduction in gene expression of *i*SINE-containing genes. It should be noted, however, that our reporter constructs did show different levels of expression in cell lines deleted for Staufen, ADAR, PKR, or Dicer. This may indicate that these factors do have some effect on gene expression but may also reflect the fact that these cell lines show different degrees of differentiation and also originate from different mouse strains. Importantly, however, irrespective of the genetic background, *i*SINE-containing constructs consistently showed a reduced level of expression, suggesting a more general mechanism of gene repression conserved in all cell lines tested. Indeed, our Pol II ChIP experiments showed that the presence of *i*SINEs leads to a reduction in Pol II density distal to *i*SINEs in the 3' UTR. We conclude that *i*SINEs interfere with Pol II activity and thus have a negative impact on mRNA transcription. Interestingly, in bacteria, double-stranded structures in RNAs have also been shown to interact with the bacterial RNA polymerase exit channel, prolong RNA polymerase pausing, and consequently reduce the transcriptional elongation rate [42]. Importantly, the crystal structures of bacterial RNA polymerase and eukaryotic Pol II are similar in the regions that interact with the transcriptional bubble [43]. Therefore, the long double-stranded structures formed by inverted SINEs used in our study might act in a comparable manner leading to an increase in Pol II pausing [44].

Conclusions

We analyzed the transcriptome-wide distribution and expression of SINEs. We found that SINEs in inverted orientation are underrepresented. Moreover, their expression is significantly reduced. Using reporter constructs we demonstrated that inverted SINEs lead to reduced RNA levels. *i*SINEs interfere with transcription most likely because they are subject to intramolecular base pairing. We therefore conclude that SINEs are not randomly inserted into the human transcriptome. In contrast, presumably due to their repressive effect on transcription, there is negative selection against multiple SINE insertions in inverted orientation.

Methods

Genomic distribution of Alu elements

Throughout the study we used GENCODE annotation v19 (human genome assembly GRCh37/hg19; <http://www.genecodegenes.org/releases/19.html>). Repeatmasker tracks were downloaded from the UCSC Genome Browser (<http://genome.ucsc.edu/>) for human genome assembly hg19 and entries of repeat *Family Alu* were extracted. Alus were clustered based on their distance to one another (maximum 300 nucleotides) and assigned to

the following groups: *ISINE* (single Alu), clusters (number of Alus in cluster >1); *dSINE* (pair of Alus; direct repeats, both Alus on same strand); *iSINE* (pair of Alus; inverted repeats, Alus are on different strands); *hiSINE* (*i*SINEs in a head-to-head orientation); *tiSINE* (*i*SINEs in a tail-to-tail orientation).

In order to remove multiple identities of genomic intervals in overlapping genes and transcript isoforms, genes and elements of genes were projected onto the genome in a hierarchical fashion on three levels: genomic (genic, intergenic), genic (exonic, intronic), and exonic (coding, 5p UTR, 3p UTR, non-coding). We used Pol II transcribed genes of the following gene types: protein coding, IG C gene, IG C pseudogene, IG D gene, IG J gene, IG J pseudogene, IG V gene, IG V pseudogene, TR C gene, TR D gene, TR J gene, TR J pseudogene, TR V gene, TR V pseudogene, polymorphic pseudogene, pseudogene, processed transcript, lincRNA, sense intronic, sense overlapping, 3' overlapping noncoding RNA, antisense.

For the enrichment analysis of Alus in genome partitions, we introduce the concept of the "Aluome". Like the genome, which is the set of nucleotides that make up the chromosomes, the Aluome is the set of nucleotides annotated as Alus. If Alus were evenly distributed in all genomic partitions, then the fraction of total nucleotides of all Alus in a given partition should be the same as the fraction of nucleotides of the genome in this partition. These proportional coverages are calculated for all genome partitions (genomic, genic, exonic) by intersecting Alu annotations with the respective genome partitions (minimum of one-nucleotide overlap).

Comparisons of positions of Alus, genes, and partitions were conducted using the intersect and closest tools of the BEDtools2 program package v2.25.0 [45] and custom Perl scripts. Statistical analysis was performed in R (<http://www.R-project.org/>) and data were plotted using ggplot2 [46].

Expression analysis of Alu-containing transcripts from ENCODE data

To test expression levels of transcripts containing no Alus, single Alu elements (*ISINEs*), tandemly repeated, i.e., duplicated (*dSINEs*), or inverted (*iSINEs* in either head-to-head [-,+] or tail-to-tail [+,-] orientation) Alu elements, GENCODE data from 15 available cell lines were analyzed (HepG2, HSMM, IMR90, MCF-7, NHEK, NHLF, K562, GM12878, Huvec, Sknsh, A549, AG04450, BJ, H1-hESC, HeLa-S3; see the "Availability of data and materials" section for details) [47]. To identify transcripts that contain Alus, we intersected Alu annotation and transcript annotation using *intersect*, a method of the BedTools package v2.16.2. Here, we considered only transcripts of genes that should be transcribed by Pol II, spliced and polyadenylated (e.g., like protein-coding mRNAs). Moreover,

we only considered Alu elements in mature transcripts fully contained in exonic regions.

Next, we formed pairs of neighboring Alus within a transcript of (i) the same Alu family and (ii) different families and annotated their relative orientation. Based on these annotations, we extracted subsets of human transcripts. Next, we plotted the distribution of expression values (FPKM cutoff of 3) for each data set, plotted them as box plots, and performed a Wilcoxon rank-sum nonparametric test for statistical significance of expression differences between groups of transcripts (e.g., *i*SINE versus *d*SINE).

Construction of renilla and firefly reporter constructs

The 3' UTRs of *Inadl*, *Nicn1*, and *Znf708* were cloned downstream of the open reading frame of firefly luciferase in pmirGLO, which also expresses renilla luciferase from the same plasmid as a reference (Promega, Madison, WI, USA). Alternatively, the 3' UTRs of interest were cloned downstream of renilla luciferase into phRL-TK (Promega, Madison, WI, USA). In those cases, the firefly luciferase-expressing plasmid pGL3 was used as a reference plasmid that was cotransfected (Promega, Madison, WI, USA).

Dual luciferase assay

To determine luciferase reporter expression, cells were transfected using Nanofectin (PAA, Pasching, Austria) or jetPEI (Polyplus transfection) following the manufacturer's instructions. For transfecting mouse embryonic fibroblasts (MEFs), Nanofectamin was used (PAA, Pasching, Austria). Six hours after transfection, cells were washed and incubated for 24 h prior to lysis and luciferase measurements. For luciferase (renilla and firefly) the dual luciferase assay (Promega, Madison, WI, USA) was used. Readings for experimental luciferase were normalized to readings for the reference construct. Experiments were done in at least three biological replicates.

MEF isolation and culture

To obtain MEFs of different genetic background, mice heterozygous for *Adar1*^{+/-}, *Adar2*^{+/-}, or both *Adar1*^{+/-}/*2*^{+/-} were intercrossed. Embryos were isolated from gravid mothers at embryonic day 11.5 (for *Adar1*^{-/-}/*2*^{-/-}). Embryos were genotyped by PCR using X and Y chromosome-specific primers and sex matched for further experiments. Homozygous and wild-type female embryos were homogenized with a syringe; cells were filtered through a cell strainer and cultured in DMEM supplemented with 20 % fetal calf serum, gentamycin, penicillin, and streptomycin. Cells were cultured for up to eight passages and used for transfection-based reporter assays.

Differentiation of embryonic stem cells

Mouse embryonic stem cells were differentiated using spontaneous differentiation of embryoid bodies (EBs) and then trypsinized and resuspended in differentiation media (DMEM supplemented with 20 % fetal bovine serum, 0.1 mM β -mercaptoethanol, and appropriate antibiotic) at 5×10^4 cells/ml. EBs were formed using the hanging drop method. For this, 300 cells were placed in a drop on the lid of a tissue culture dish. The dish was filled with phosphate-buffered saline (PBS) and cells were kept at 37 °C for 2 days. The newly formed EBs were transferred to gelatin-coated dishes and left for 7 days for spontaneous differentiation. EBs were trypsinized and used for further experiments.

RNA extraction

To determine RNA levels, cell lysates prepared for the dual luciferase assay were used immediately after lysis. Lysates were purified using the Qiagen RNeasy mini kit (Qiagen, Hilden, Germany). After purification, an extra round of DNase I and DpnI digestion was included to avoid plasmid DNA contamination. DNaseI and DpnI were heat inactivated and the RNA was precipitated with ethanol prior to RT-PCR or qPCR.

RT-PCR and qPCR

cDNA synthesis was done with random hexamers and RevertAid (RNaseH minus) mMuLV reverse transcriptase following the manufacturer's instructions (Thermo Fisher Scientific). As a control, MOCK reactions without RT enzyme were set up. For qPCR, a GoTaq qPCR master mix was used (Promega, Madison, WI, USA) on a BioRad iQ5 cycler (BioRad, Hercules, CA, USA). At least three biological and two technical replicates were done for each qPCR assay. Relative differences in RNA levels were determined by using the delta delta CT method [23, 48]. The sequences of the primers used to determine RNA levels of firefly and renilla luciferase by qPCR are listed in Additional file 2: Table S1.

Fluorescent in situ hybridization (FISH)

Coverslip-grown cells were washed three times in 1× PBS and fixed in 2 % paraformaldehyde/PBS for 15 minutes at room temperature. After washing again in 1× PBS, cells were permeabilized in 0.2–0.5 % Triton X-100, 2 mM vanadyl-ribonucleoside complex, 1× PBS. After washing slides in 2× SSC for 10 minutes at room temperature, 20 μ l of hybridization mix was added and the slides were sealed with a cover slip and incubated at 37 °C overnight. Coverslips were washed in 2× SSC, 50 % formamide for 5 minutes at 42 °C. Fluorescent signals were detected with an Alexa 488-labeled anti-FITC antibody. Cells were mounted in antifade containing DAPI.

Probe preparation

Specific DNA (200 ng) was labeled with FITC-dUTP via nick-translation (Roche). Labeled probe was precipitated and resuspended in 10 μ l of water. The probe (2 μ l) and yeast tRNA (20 μ g) were lyophilized and dissolved in 10 μ l deionized formamide (Ambion). The probe was denatured at 75 °C for 10 minutes and immediately chilled on ice. Hybridization buffer (10 μ l) was added to make a hybridization cocktail of 20 μ l per coverslip. Hybridization buffer consisted of 2 \times SSC, 10 % dextran sulfate, and four units of RNase inhibitor (Ribolock, Thermo Fisher Scientific, Waltham, MA, USA) per microliter.

Modeling of artificial Alu-like variants and editing detection

For the design of artificial Alu-like structures that were inserted in the Znf 3' UTR, we distinguished between spacer regions and Alu elements in the *Znf708* reference transcript. The spacer regions were kept constant during sequence design, while Alu elements were replaced by initially random sequences and then optimized to meet selected thermodynamic properties of the original transcript.

In particular, we optimized sequences to have the same minimum free energy (MFE) secondary structure, the same minimum free energy, and the same free energy of the secondary structure ensemble (EFE). Among all candidates fulfilling these criteria, we selected for a sequence with similar GC content and base-pairing probabilities. Choosing from sequences with the same MFE and EFE ensures that the probability of forming the MFE secondary structure is exactly the same. A comparison of the target and designed sequence and the corresponding secondary structure ensembles can be seen in Additional file 1: Figure S2.

To determine if the artificial constructs are edited, they were transfected into a cell line stably expressing rat ADAR2. RNA was extracted and cDNA was generated as described above. Subsequently, the first Alu (or the modeled sequence analogous to the first Alu) was amplified by PCR. Finally the PCR product was gel purified and submitted to Sanger sequencing.

RNA half-life determination

For detecting the rate of mRNA degradation, the mRNA transcription was blocked by Actinomycin D (10 μ g/ μ l) and the mRNA was collected in the regular time interval (0, 1, 2 h). The mRNA amount was quantified by qPCR. RT-PCR and qPCR were done as explained above.

Poly(A) tail determination

The length of the poly(A) tail was determined by PCR following a splint-linker ligation as described [49].

Pol II ChIP

Chromatin immunoprecipitation was performed as previously described by Hauser et al. [50] with the following changes: 3 \times 10-cm dishes (9×10^5 U2OS cells) were pooled for each immunoprecipitation. Cell lysates were incubated overnight with 2 μ g of Pol II antibody at 4 °C and Pol II antibody complexes were collected using 25 μ l (1×10^7) of Dynabeads Pan Mouse IgG beads (Invitrogen, Carlsbad, CA, USA). The amount of immunoprecipitated DNA was analyzed using qPCR on a BioRad iQ5 cycler (BioRad, Hercules, CA, USA) and the GoTaq qPCR master mix (Promega, Madison, WI, USA). Three different regions of the reporter gene (firefly), including early coding region (A), stop codon (B), and near poly(A) signal (C), were amplified with the primers listed in Additional file 2: Table S1. ChIP signals from regions B and C were normalized to region A. Signals in experimental constructs harboring *i*SINEs were compared to control constructs harboring *I*SINEs. The statistical significance of observed changes was determined using a Student's *t*-test. A non-specific rabbit IgG fraction was used as a control for the immunoprecipitation. For the endogenous targets region A is located approximately 10 kb upstream of the single or inverted SINE in an intronic region. Regions B and C are located either upstream or downstream of the single or inverted SINEs in close proximity. The primers are given in Additional file 2: Table S1.

Additional files

Additional file 1: Figure S1. Alu elements in *i*SINE containing constructs are edited, indicating that *i*SINEs can form double-stranded structures. **Figure S2.** Minimum free energy structures of the Znf708 UTR and the designed UTRs. **Figure S3.** ZNF-analogs mimicking the folding of the wildtype ZNF are edited, indicating that they form double-stranded structures. **Figure S4.** *i*SINEs do not lead to nuclear retention of RNAs. **Figure S5.** *i*SINE mediated gene repression is independent of dsRNA-activated kinase PKR and DICER or DROSHA activity. **Figure S6.** *i*SINEs do not interfere with mRNA polyadenylation. (PDF 2738 kb)

Additional file 2: Table S1. describing all DNA primers used in this study for regular and quantitative PCR analysis. (PDF 57 kb)

Acknowledgments

The authors would like to thank Rafael Schmid, Danijela Kurija, Johann Schmutzmeier, Peter Burg, and Celine Brunner for excellent technical assistance and Drs. Reis e Sousa and Oliver Schulz for the kind gift of PKR deficient cells. DICER1-deficient cells were a kind gift of Dr. Greg Hannon. ADAR1- and ADAR2-deficient mice were a kind gift of Prof. Peter Seeburg, MPI Heidelberg. Staufien-deficient mice were a kind gift of Dr. Michael Kiebler and Sabine Thomas, LMU Munich.

Funding

This work was supported by the Austrian Science Foundation grant number SFB F4313 P26845 and P26882 to M.J. MT was supported by the Doctoral College RNA-biology W1207 of the Austrian Science foundation. KL was supported by a stipend from the German Research Foundation DFG (LI 2431 2-1).

Availability of data and materials

ENCODE RNAseq samples used in this study are a subset of experiment GSE26284 and are available through Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26284>). Data used are listed here:

Cell line	Localization	RNA extract	Read type	Data version	GEO sample accession
A549	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM758564
AG04450	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM758561
BJ	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM758562
GM12878	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM758559
H1-hESC	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM758566
HeLa-S3	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM765402
HepG2	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM758575
HSMM	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM758578
HUVEC	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM758563
IMR90	Cell	longPolyA	2x101D	ENCODE Mar 2012 Freeze	GSM981249
K562	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM765405
MCF-7	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM765388
NHEK	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM765401
NHLF	Cell	longPolyA	2x76D	ENCODE Jan 2011 Freeze	GSM765394
SK-N-SH	Cell	longPolyA	2x101D	ENCODE Mar 2012 Freeze	GSM981253

Authors' contributions

MT, AT, KL, IH, and MFJ conceived the study. MT, KL, and MFJ wrote the manuscript. MT and KL performed wet-lab experiments. AT and MTW analyzed gene expression based on ENCODE RNAseq data and AT and FH analyzed genome annotations and genomic distribution of SINEs. SB designed sequences. OP did the DROSHA knockdown. SS initiated the study. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interest.

Author details

¹Department of Chromosome Biology, Max F. Perutz Laboratories, University of Vienna, Dr. Bohr Gasse 9/5, Vienna A-1030, Austria. ²Department of Cell and Developmental Biology, Medical University of Vienna, Schwarzspanierstrasse 17, Vienna A-1090, Austria. ³Institute for Theoretical Chemistry, University of Vienna, Währinger Strasse 17, Vienna A-1090, Austria. ⁴Present address: Center for molecular biology of the University Heidelberg, Im Neuenheimer Feld 282, Heidelberg D-69120, Germany. ⁵Department of Cell and Developmental Biology, Medical University of Vienna, Center of Anatomy and Cell Biology, Schwarzspanierstrasse 17, Vienna A-1090, Austria.

Received: 21 March 2016 Accepted: 10 October 2016

Published online: 25 October 2016

References

- Hasler J, Samuelsson T, Strub K. Useful 'junk': Alu RNAs in the human transcriptome. *Cell Mol Life Sci*. 2007;64:1793–800.
- Lev-Maor G, Sorek R, Shomron N, Ast G. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science*. 2003;300:1288–91.
- Hancks DC, Kazazian Jr HH. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev*. 2012;22:191–203.
- Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*. 2011;479:534–7.
- Mariner PD, Walters RD, Espinoza CA, Drullinger LF, Wagner SD, Kugel JF, Goodrich JA. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell*. 2008;29:499–509.
- Kramerov DA, Vassetzky NS. Origin and evolution of SINEs in eukaryotic genomes. *Heredity (Edinb)*. 2011;107:487–95.
- Kramerov DA, Vassetzky NS. Structure and origin of a novel dimeric retrotransposon B1-diD. *J Mol Evol*. 2001;52:137–43.
- Vassetzky NS, Ten OA, Kramerov DA. B1 and related SINEs in mammalian genomes. *Gene*. 2003;319:149–60.
- Rowold DJ, Herrera RJ. Alu elements and the human genome. *Genetica*. 2000;108:57–72.
- Gong CG, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*. 2011;470:284–8.
- Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*. 2012;491:454–457.
- Ishida K, Miyauchi K, Kimura Y, Mito M, Okada S, Suzuki T, Nakagawa S. Regulation of gene expression via retrotransposon insertions and the noncoding RNA 4.5S RNAH. *Genes Cells*. 2015;20:887–901.
- Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Szybel D, et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol*. 2004;22:1001–5.
- Athanasadiadis A, Rich A, Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol*. 2004;2:e391.
- Chen LL, Carmichael GG. Gene regulation by SINEs and inosines: biological consequences of A-to-I editing of Alu element inverted repeats. *Cell Cycle*. 2008;7:3294–301.
- Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs FJ, Rechavi G, Li JB, Eisenberg E, Levanon EY. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res*. 2014;24:365–76.
- Chen LL, DeCervo JN, Carmichael GG. Alu element-mediated gene silencing. *EMBO J*. 2008;27:1694–705.
- Hu SB, Xiang JF, Li X, Xu Y, Xue W, Huang M, Wong CC, Sagum CA, Bedford MT, Yang L, et al. Protein arginine methyltransferase CARM1 attenuates the

- paraspeckle-mediated nuclear retention of mRNAs containing IRALus. *Genes Dev.* 2015;29:630–45.
19. Hundley HA, Bass BL. ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends Biochem Sci.* 2010;35:377–383.
 20. Ricci EP, Kucukural A, Cenik C, Mercier BC, Singh G, Heyer EE, Ashar-Patel A, Peng LT, Moore MJ. Staufen1 senses overall transcript secondary structure to regulate translation. *Nat Struct Mol Biol.* 2014;21:26–35.
 21. Elbarbary RA, Li WC, Tian B, Maquat LE. STAU1 binding 3' UTR IRALus complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes Dev.* 2013;27:1495–510.
 22. Hundley HA, Krauchuk AA, Bass BL. *C. elegans* and *H. sapiens* mRNAs with edited 3' UTRs are present on polysomes. *RNA.* 2008;14:2050–60.
 23. Capshew CR, Dusenbury KL, Hundley HA. Inverted Alu dsRNA structures do not affect localization but can alter translation efficiency of human mRNAs independent of RNA editing. *Nucleic Acids Res.* 2012;40:8637–45.
 24. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res.* 2012;22:1760–74.
 25. Stenger JE, Lobachev KS, Gordenin D, Darden TA, Jurka J, Resnick MA. Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability. *Genome Res.* 2001;11:12–27.
 26. Lobachev KS, Stenger JE, Kozyreva OG, Jurka J, Gordenin DA, Resnick MA. Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.* 2000;19:3822–30.
 27. El-Sawy M, Deininger P. Tandem insertions of Alu elements. *Cytogenet Genome Res.* 2005;108:58–62.
 28. Riedmann EM, Schopoff S, Hartner JC, Jantsch MF. Specificity of ADAR-mediated RNA editing in newly identified targets. *RNA.* 2008;14:1110–8.
 29. Neeman Y, Levanon EY, Jantsch MF, Eisenberg E. RNA editing level in the mouse is determined by the genomic repeat repertoire. *RNA.* 2006;12:1802–9.
 30. Prasanth KV, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL. Regulating gene expression through RNA nuclear retention. *Cell.* 2005;123:249–63.
 31. Kim YK, Furic L, Parisien M, Major F, DesGroseillers L, Maquat LE. Staufen1 regulates diverse classes of mammalian transcripts. *EMBO J.* 2007;26:2670–81.
 32. Williams BRG. PKR; a sentinel kinase for cellular stress. *Oncogene.* 1999;18:6112–20.
 33. Abraham N, Stojdl DF, Duncan PI, Methot N, Ishii T, Dube M, Vanderhyden BC, Atkins HL, Gray DA, McBurney MW, et al. Characterization of transgenic mice with targeted disruption of the catalytic domain of the double-stranded RNA-dependent protein kinase, PKR. *J Biol Chem.* 1999;274:5953–62.
 34. Bensaude O. Inhibiting eukaryotic transcription: which compound to choose? How to evaluate its activity? *Transcription.* 2011;2:103–8.
 35. Sandoval J, Rodriguez JL, Tur G, Serviddio G, Pereda J, Boukaba A, Sastre J, Torres L, Franco L, Lopez-Rodas G. RNAPol-ChIP: a novel application of chromatin immunoprecipitation to the analysis of real-time gene transcription. *Nucleic Acids Res.* 2004;32:e88.
 36. Krehling J, Graveley BR. The origins and implications of Alu alternative splicing. *Trends Genet.* 2004;20:1–4.
 37. Lev-Maor G, Sorek R, Levanon EY, Paz N, Eisenberg E, Ast G. RNA-editing-mediated exon evolution. *Genome Biol.* 2007;8:R29.
 38. Lev-Maor G, Ram O, Kim E, Sela N, Goren A, Levanon EY, Ast G. Intronic Alu influence alternative splicing. *PLoS Genet.* 2008;4:e1000204.
 39. Daniel C, Silberberg G, Behm M, Ohman M. Alu elements shape the primate transcriptome by cis-regulation of RNA editing. *Genome Biol.* 2014;15:R28.
 40. Daniel C, Veno MT, Ekdahl Y, Kjems J, Ohman M. A distant cis acting intronic element induces site-selective RNA editing. *Nucleic Acids Res.* 2012;40:9876–86.
 41. Kim Y, Lee JH, Park JE, Cho J, Yi H, Kim VN. PKR is activated by cellular dsRNAs during mitosis and acts as a mitotic regulator. *Genes Dev.* 2014;28:1310–22.
 42. Toulkhonov I, Artsimovitch I, Landick R. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science.* 2001;292:730–3.
 43. Dangkulwanich M, Ishibashi T, Bintu L, Bustamante C. Molecular mechanisms of transcription through single-molecule experiments. *Chem Rev.* 2014;114:3203–23.
 44. Jonkers I, Kwak H, Lis JT. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife.* 2014;3:e02407.
 45. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
 46. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2009.
 47. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. Landscape of transcription in human cells. *Nature.* 2012;489:101–8.
 48. Hellems J, Mortier G, De Paep A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 2007;8:R19.
 49. Minasaki R, Rudel D, Eckmann CR. Increased sensitivity and accuracy of a single-stranded DNA splint-mediated ligation assay (sPAT) reveals poly(A) tail length dynamics of developmentally regulated mRNAs. *RNA Biol.* 2014;11:111–23.
 50. Hauser C, Schuettengruber B, Bartl S, Lagger G, Seiser C. Activation of the mouse histone deacetylase 1 gene by cooperative histone phosphorylation and acetylation. *Mol Cell Biol.* 2002;22:7820–30.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

