nature
structural &
molecular biology

L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells

Nuo Yang & Haig H Kazazian, Jr.

LINE-1s, or L1s, are highly abundant retrotransposons comprising 17% of the human genome. Most L1s are retrotransposition defective; nonetheless, there are ~100 full-length L1s potentially capable of retrotransposition in the diploid genome. L1 retrotransposition may be detrimental to the host and thus needs to be controlled. Previous studies have identified sense and antisense promoters in the 5′ UTR of full-length human L1. Here we show that the resulting bidirectional transcripts can be processed to small interfering RNAs (siRNAs) that suppress retrotransposition by an RNA interference (RNAi) mechanism. We thus provide evidence that RNAi triggered by antisense transcripts may modulate human L1 retrotransposition efficiently and economically. L1-specific siRNAs are among the first natural siRNAs reported in mammalian systems. This work may contribute to understanding the regulatory role of abundant antisense transcripts in eukaryotic genomes.

Long interspersed element-1 (LINE-1 or L1) is a class of highly abundant retrotransposons comprising ~17% of the human genome[1]. A full-length human L1 is ~6.0 kilobases and contains a 5′ untranslated region (UTR), two open reading frames (ORF1 and ORF2) and a 3′ UTR terminating in a polyadenylated tail[2] (see below). L1 is an important autonomous non–long terminal repeat retrotransposon, which moves by a 'copy and paste' mechanism[3]. The original retrotransposon is transcribed into RNA and exported to the cytoplasm, where ORF1 and ORF2 proteins are translated and form a ribonucleoprotein complex preferentially with the precursor transcript (*cis*-preference)[4,5]. ORF1 encodes a 40-kDa protein (p40) with RNA-binding[6] and nucleic acid–chaperone activity[7]; ORF2 encodes a protein with endonuclease[8,9] and reverse transcriptase[10] activities. Both ORF1 and ORF2 are required for retrotransposition[2,6–10]. L1s integrate into the genome by a mechanism called target-primed reverse transcription[11], using the free 3′-OH at the endonuclease cut site on the genomic DNA as a primer and the L1 RNA as a template.

The impact of L1 on genome evolution is substantial and can be either constructive or destructive[3]. On one hand, L1 can give rise to new genes via exon shuffling by 3′ transduction[12] and occasionally repair double-strand breaks by endonuclease-independent integration into DNA lesions[13]. On the other hand, L1 retrotransposition can affect normal gene expression[3] and at least 14 cases of genetic disease caused by human L1 insertions have been reported in human patients[3]. Bioinformatic analysis of L1 distribution on human autosomes and sex chromosomes has also demonstrated that full-length L1 is deleterious to the host[14]. Although the large majority of L1 retrotransposons in mammals are defective owing to 5′ truncation, inversion or mutation[2], the average diploid human genome has 80–100 full-length L1s potentially capable of retrotransposition[3]. Therefore, certain mechanisms must exist to alleviate the threat of human L1 retrotransposition by controlling its activity.

RNA silencing refers to a particular collection of phenomena in which small regulatory noncoding RNAs of ~19–28 nucleotides (nt), derived from double-stranded RNAs (dsRNAs) or stem-loop precursors, trigger repression of homologous sequences[15,16]. RNA silencing is highly conserved across eukaryotes. Small RNAs induce gene silencing through various mechanisms, including cleavage of cognate RNAs by RNAi, blocking productive translation of messenger RNAs by the microRNA pathway, and transcriptional gene silencing by induction of heterochromatization of specific target DNAs.

Since its first elucidation in *Caenorhabditis elegans*[17], the mechanism of RNAi has been revealed through a combination of genetic, molecular and computational approaches. We now know that siRNAs 21–23 nt in length[18] are processed from long dsRNAs by the enzyme Dicer, a member of the RNase III family of ribonucleases[19]. The siRNA is then incorporated into an RNA-induced silencing complex (RISC)[19], which cleaves perfectly complementary mRNA species[20].

RNA silencing may have initially evolved to defend against invading nucleic acids, including viruses and genomic parasites (for example, transposable elements) and subsequently been co-opted to serve in gene regulation[21]. For example, increased mobilization of endogenous transposons has been observed in some RNAi-deficient *C. elegans* strains[22,23], and it has since been proven that fortuitous read-through transcription of dispersed Tc1 transposon copies can form dsRNA and that RNAi mediated by this dsRNA silences transposition[24]. Furthermore, recent studies in mammalian systems have shown that knockdown of *Dicer1* in one-cell mouse embryos by synthetic siRNA increases transcript abundance of murine MuERV-L and IAP retrotransposons[25], and *Dicer1* knockout mouse embryonic stem cells

Department of Genetics, University of Pennsylvania School of Medicine, Rm. 475 Clinical Research Building, 415 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA. Correspondence should be addressed to H.H.K. (kazazian@mail.med.upenn.edu).

Received 7 July; accepted 4 August; published online 27 August 2006; doi:10.1038/nsmb1141

express a higher level of L1 and IAP transcripts[26]. However, the control of transposable elements by RNAi in mammalian systems is yet to be confirmed.

The 5′ UTR of human L1 is approximately 900 nt and contains an internal promoter that drives transcription at or near position +1 (ref. 27). Notably, a ubiquitously active antisense promoter (ASP) has been identified in the 5′ UTR, giving rise to many chimeric transcripts of adjacent cellular genes[28,29]. Thus, it is possible that dsRNA with the L1 sequence is produced by bidirectional transcription of the 5′ UTR. Here we show that dsRNA derived from the human L1 5′ UTR can be processed into siRNAs and reduce retrotransposition in cell culture–based assays by an RNAi-related mechanism.

RESULTS

Characterization of the ASP in the human L1 5′ UTR

To confirm that the ASP in the human L1 5′ UTR is indeed transcriptionally active, we transfected HeLa cells with luciferase reporter constructs whose expression is driven by various promoters and performed northern blots with a luciferase-specific riboprobe. In this way, we detected an ASP-derived transcript along with a sense 5′ UTR transcript (**Fig. 1a**). As measured by density quantification software, transcript abundance of the sense 5′ UTR promoter (SP) was comparable to that of the *SV40* promoter but greater than that of the ASP by about eight-fold (**Fig. 1a**). This result was consistent with their promoter activities in a luciferase reporter assay (**Fig. 1b**). It is noteworthy that, according to the northern blot, ASP seems to start transcription near the middle of the 5′ UTR (**Fig. 1a**), unlike the SP, which initiates transcription around +1 (refs. 27,30,31).

To better understand the structure of ASP-derived transcripts, we identified their transcription start sites by 5′ rapid amplification of complementary DNA ends (RACE). We found that ASP-driven transcription initiates at heterogeneous start sites in two major regions, +378–431 and +480–497. Furthermore, within each region several nucleotides could serve as initiation sites (**Fig. 1c**). We also showed by mapping reverse-transcription (RT)-PCR that no other transcription start site exists upstream of position +500 (**Fig. 1d**). Finally, we characterized the promoter activity of ASP using a luciferase reporter assay (**Fig. 1e**). We found that ASP activity is located in the 3′ region of the 5′ UTR, after about position +400, where transcription by ASP is initiated. In addition, sequences further upstream of the initiation sites—for example, upstream of +510 and +663—also contribute significantly to ASP activity ($P < 0.001$ and $P = 0.001$, respectively). In summary, the above results indicate that the ASP in the human L1 5′ UTR is transcriptionally functional and that bidirectional transcription of the 5′ UTR has the potential to generate dsRNA from +1 to +500.

Bidirectional transcripts generate L1-specific siRNAs

In diverse human and monkey cell lines, abundant RNA that anneals with an L1 probe has been identified. These transcripts are predominantly nuclear, heterogeneous in size and not polyadenylated, and, most importantly, they originate from both strands of the human L1 sequence (discussed in ref. 27). These observations suggest that dsRNAs with the L1 sequence may exist in wild-type primate cells and may serve as an endogenous mechanism for retrotransposon silencing. In light of the above characterization of ASP, we conducted northern blots with human cultured cells in search of endogenous siRNAs generated from the human L1 5′ UTR dsRNA. We used riboprobes containing the sense or antisense strand sequence of the putative dsRNA generated by ASP and SP transcription or containing the rest of the 5′ UTR (**Fig. 2a**). We detected abundant ~21-nt siRNAs and degraded L1 transcripts of both sense and antisense strands with probes complementary to the putative dsRNA region (+1–500) (**Fig. 2b**). siRNAs were more readily detected in wild-type 293 cells

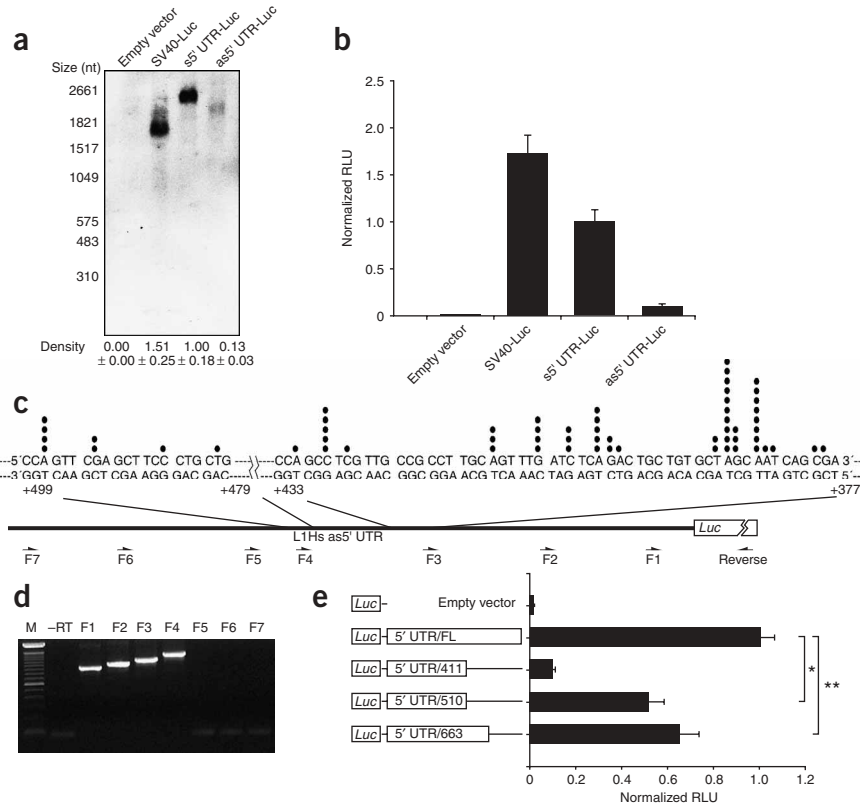

Figure 1 Characterization of the ASP in the human L1 5′ UTR. (**a**) Detection of the ASP-derived transcript by northern blotting. Total RNA was extracted from HeLa cells transfected with pGL3-Basic vector containing either no promoter (empty vector), *SV40* control promoter (SV40-Luc) or human L1 5′ UTR in the sense (s5′ UTR-Luc) or antisense direction (as5′ UTR-Luc). Density quantitation of the signal ($n = 3$) is below gel. (**b**) Luciferase reporter assay of expression from *SV40* control promoter (SV40-Luc), sense human L1 5′ UTR (s5′ UTR-Luc) and ASP (as5′ UTR-Luc). pGL3-Basic vector containing no promoter (empty vector) was used as a negative control. RLU, relative luciferase units. Error bars show s.d. ($n = 9$). (**c**) Transcription start sites of ASP identified by 5′ RACE. Dot, one occurrence of transcription initiation at the specific nucleotide; arrows, primers for mapping RT-PCR. (**d**) Electrophoresis of the mapping RT-PCR products amplified by the primers in **c**. M, molecular weight markers. (**e**) Characterization of the ASP activity in full-length (FL) and truncated constructs (last nucleotide listed after slash) by luciferase reporter assay (*, $P < 0.001$; **, $P = 0.001$; $n = 9$). Error bars show s.d.

than in HeLa cells, and the discrepancy between the abundances of sense and antisense siRNAs is probably due to cleavage of the passenger siRNA by Argonaute-2 (AGO2) in RISC[32]. Notably, no siRNA could be detected with a probe from position +601 to +906 or with an irrelevant probe against the gene encoding green fluorescent protein (GFP) (**Fig. 2b**), suggesting that the siRNAs are specifically produced from the dsRNA region of the 5′ UTR.

As the expression of L1 siRNAs was barely detectable in wild-type HeLa cells, these cells provided an excellent host for transfection experiments. To further demonstrate that the identified siRNAs were L1 specific, we transfected HeLa cells with various luciferase reporter constructs. The ~21-nt siRNAs could be detected only in HeLa cells transfected with the construct whose expression is driven by the full-length human L1 5′ UTR, and not in cells transfected with the empty vector (**Fig. 2c**, left), indicating that the siRNAs are L1 specific. In addition, when the amount of RNA loaded was increased, siRNAs were better resolved from the background and observed even in untransfected HeLa cells (**Fig. 2c**, right). Finally, we transfected HeLa cells with luciferase reporter constructs whose expression is driven by various 5′ UTR fragments containing either complete or impaired ASP activity and performed a northern blot as above. As expected, impaired ASP activity resulting from deletions after position +412 substantially decreased siRNA abundance as compared to that derived from the full-length 5′ UTR, and the decrease was largely correlated with the lost ASP activity (**Fig. 2d**). As reported in the previous section, ASP activity is about 10% of SP activity (**Fig. 1a**,**b**). Therefore, consistent with what we observed here (**Fig. 2d**), the intactness of the ASP but not that of the sense promoter may be a deciding factor in siRNA production.

ASP gives rise to inhibition of L1 retrotransposition

Having detected human L1-specific siRNAs derived from bidirectional transcripts of the 5′ UTR, we set out to confirm their inhibitory effect on retrotransposition. To do this, we replaced the full-length 5′ UTR with a panel of 5′ UTR fragments truncated at the 3′ end in the cell culture–based retrotransposition assay and the luciferase reporter assay (**Fig. 3a**). For the retrotransposition assay, an L1 element under investigation is tagged in its 3′ UTR with an antisense enhanced GFP (*EGFP*) cassette, which is driven by a heterologous cytomegalovirus (CMV) promoter, disrupted by a γ-globin intron in the same orientation as L1 and followed by a thymidine kinase (*TK*) polyadenylation signal[9,33]. Therefore, EGFP expression occurs only after L1 transcription, splicing of the intron, reverse transcription and insertion of the L1 cDNA copy back into the genome—that is, after a retrotransposition event (**Fig. 3b**). In this

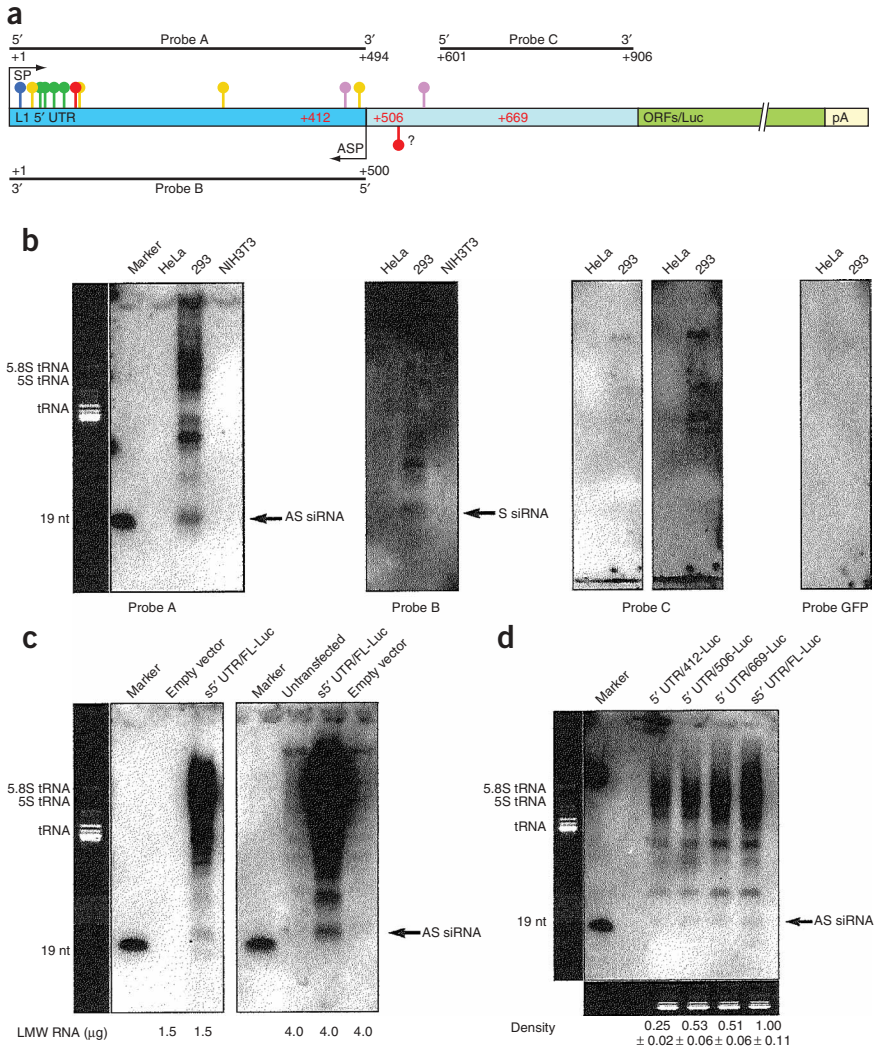

Figure 2 Bidirectional transcripts of 5′ UTR generate L1-specific siRNAs. (**a**) Schematic presentation of L1 5′ UTR. SP, sense promoter; ASP, antisense promoter. Known transcription factor–binding and methylation-sensitive sites involved in the transcriptional regulation of human L1 5′ UTR are denoted by colored 'lollipop' symbols at left. Blue, YY1 (+21–13)[30]; yellow, methyl-CP2-responsive (+36, +101, +304 and +481)[49]; green, CpGs (+52, +58, +61 and +70)[48]; pink, SRY (+472–477 and +572-577)[34]; red, RUNX3 (+83–101 and +526–508)[35]. Probes used in northern blot are shown. Deletions of the 5′ UTR after positions +412, 506 and 669 were used in this study. Sequence is numbered according to the sense strand. (**b**) Detection of human L1 5′ UTR–specific siRNAs in wild-type cells. Each lane contains 1.5 μg low–molecular weight (LMW) RNA. Arrows, sense (S) and antisense (AS) siRNAs; marker, molecular weight marker of 19 nt. Left, the ethidium bromide–stained gel before electrotransfer. Hybridization probe is indicated under each gel. For probe C, both short and long exposure is shown. (**c**) Left, northern blot detection of 5′ UTR–specific siRNAs in HeLa cells transfected with pGL3-Basic vector containing no promoter (empty vector) or the full-length sense human L1 5′ UTR (s5′ UTR-Luc). Each lane contains 1.5 μg LMW RNA. Right, lanes loaded with 4 μg of LMW RNAs extracted from wild-type HeLa cells (untransfected) or HeLa cells transfected with pGL3-Basic containing the full-length sense human L1 5′ UTR (s5′ UTR-Luc) or no promoter (empty vector). (**d**) Comparison of siRNA abundance generated by the transfected luciferase reporter constructs containing sense nucleotides +1–412 (s5′ UTR/412-Luc), +1–506 (s5′ UTR/506-Luc), +1–669 (s5′ UTR/669-Luc) or full-length 5′ UTR (s5′ UTR/FL-Luc) in HeLa cells. The ethidium bromide-stained gel before electrotransfer is aligned to the left and below (as RNA loading control). Density quantitation of the signal ($n = 3$) is below gel.

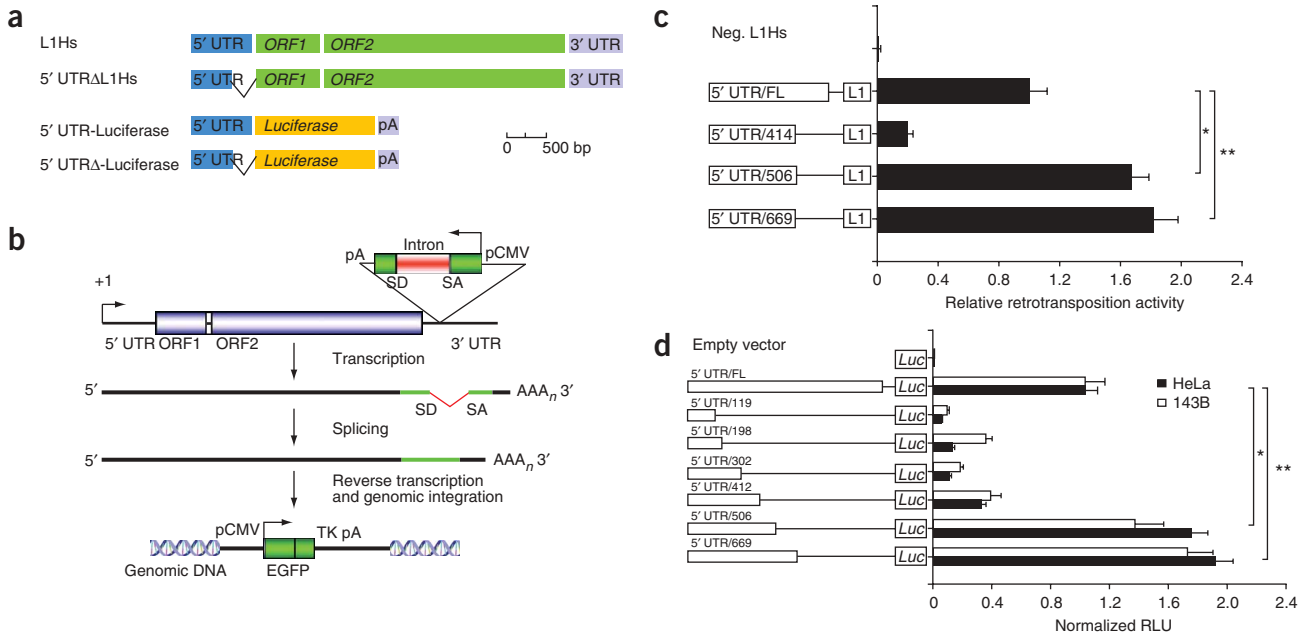

Figure 3 ASP encodes an inhibitory effect on L1 retrotransposition. (**a**) Schematic of the constructs used for retrotransposition assay and luciferase reporter assay, in which the human L1 5′ UTR is truncated variably at the 3′ end. L1Hs, human L1; UTR, untranslated region; ORF, open reading frame; pA, polyadenylation signal. (**b**) Cell culture–based assay for L1 retrotransposition[9]. The L1 under investigation is tagged with an antisense *EGFP* cassette at its 3′ UTR. pCMV, CMV immediate early promoter; intron, γ-globin intron; pA, polyadenylation signal; SD, splice donor site; SA, splice acceptor site. (**c**) ASP encodes an inhibitory effect on L1 retrotransposition. Constructs in **c** and **d** are denoted as in **Figure 1e**. *, $P = 0.001$; **, $P < 0.001$; $n = 6$; error bars show s.d. (**d**) ASP reduces the transcript abundance derived from L1 5′ UTR. Luciferase reporter assay of the 5′ UTR fragments truncated at the 3′ end, in HeLa (black) and 143B *TK*$^-$ cells (white). All activities are normalized to that of 5′ UTR/FL. RLU, relative luciferase units. Error bars show s.d. HeLa: *, $P = 0.001$; **, $P < 0.001$. 143B: *, $P = 0.026$; **, $P = 0.002$. $n = 9$.

experiment, we found that the 5′ UTR deletion after position +414 almost eliminated L1 retrotransposition, which was not unexpected, as putative transcription factor–binding sites such as SRY[34] (**Fig. 2a**) are present in the deleted region. More notably, deletions after +506 and +669, which resulted in partial loss of ASP activity (**Fig. 1e**), increased L1 retrotransposition by up to 83% (**Fig. 3c**, $P < 0.001$). We have previously observed a correlation between L1 transcript abundance and retrotransposition[35]. To determine whether the detected retrotransposition difference originated at the transcript level, we performed luciferase reporter assays with the corresponding 5′ UTR fragments. In both cell lines we tested, HeLa and 143B *TK*$^-$, the results were consistent with those observed in the retrotransposition assay. Luciferase activity, which closely correlates with transcript abundance, was markedly reduced by the 5′ UTR deletion after +412. Thus, 5′ UTR deletion after about position +412 greatly affects both sense and antisense promoter activity. Moreover, the transcript abundance was almost doubled by deletions after +506 and +669 (**Fig. 3d**; $P = 0.001$ and $P < 0.001$, respectively, for HeLa; $P = 0.026$ and 0.002, respectively, for 143B). In summary, our data suggest that the ASP in the full-length human L1 5′ UTR gives rise to an inhibitory effect on retrotransposition, most probably by reducing transcript abundance.

ASP induces mRNA degradation of the 5′ UTR–derived transcript

To develop a more natural system for verification and further study, we used the Flp-In system (Invitrogen) to establish 293-derived cell lines in which a single copy of a *d2EGFP* transgene (encoding a destabilized variant of the wild-type GFP), with expression driven by various human L1 5′ UTR fragments (**Fig. 4a**), was introduced into an active chromatin structure. This system has two unique advantages. First, it allows us to study L1 expression from a natural genomic environment rather than from a transfected plasmid. Second, as every transgene is inserted into the same genomic locus, we could eliminate positional effects from flanking sequences in different cell lines.

We first compared the *d2EGFP* transcript abundance between cell lines using quantitative real-time RT-PCR. Consistent with the results of the transfection experiment (**Fig. 3d**), impaired ASP activity significantly increased the transcript abundance derived from L1 5′ UTR fragments (**Fig. 4b**, $P < 0.001$). At the protein level, d2EGFP expression could not be detected in the no-promoter control cell line by FACS or fluorescence microscopy but was readily seen in the cell line where expression was driven by the full-length L1 5′ UTR (**Fig. 4c**). Also consistent with the transfection experiment, deletion after position +412 almost abolished d2EGFP expression, whereas deletion after +672, which resulted in the partial loss of ASP activity, greatly increased it.

To rule out the possibility that deletion after +672 increased transcript abundance by eliminating negative regulatory sequences affecting transcription, we performed chromatin immunoprecipitation (ChIP) assays to quantify RNA polymerase II (Pol II) loading on the transgene promoter of each cell line. As expected, no Pol II loading was detected in the no-promoter control line, and very little loading was detected in the cell line containing positions +1–412 of the 5′ UTR, compared to the line with the full-length 5′ UTR. This agrees with our observation that the sequence after +412 is also important for 5′ UTR–derived transcription (**Fig. 3d**). Most notably, no appreciable difference was observed between the lines driven by the full-length 5′ UTR and positions +1–672 of the 5′ UTR (**Fig. 4d**), indicating that the deletion effect was at a post-transcriptional level.

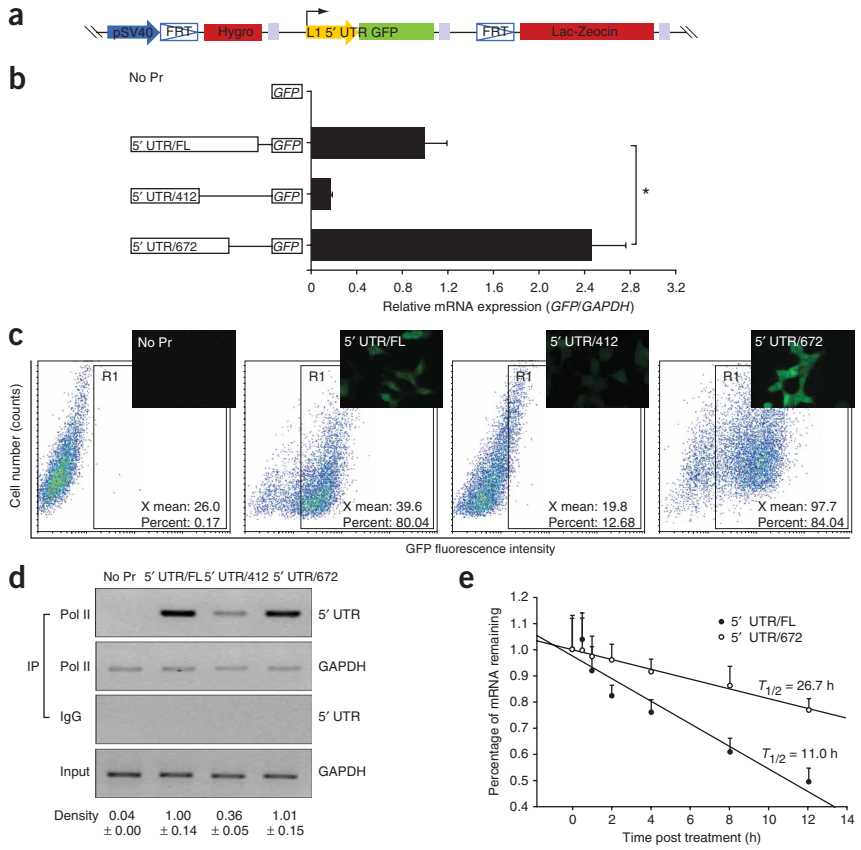

Figure 4 ASP induces the mRNA degradation of 5′ UTR–derived transcripts. (**a**) Schematic figure of transgene structure in Flp-In cell lines. (**b**) Quantitative real-time RT-PCR analysis of the *d2EGFP* transcript abundance in the Flp-In cell lines driven by no promoter (No Pr), full-length 5′ UTR (5′ UTR/FL) or positions +1–412 (5′ UTR/412) or +1–672 (5′ UTR/672), respectively. *GAPDH* was used as internal control. Data are normalized to that of 5′ UTR/FL (*, $P < 0.001$; $n = 4$). Error bars show s.d. (**c**) Typical FACS density plots of d2EGFP expression in the established Flp-In cell lines. X mean, geometric mean of GFP fluorescence intensity; %, percentage of the gated d2EGFP-positive cells among the recorded events. (**d**) ChIP assay of Pol II loading on the transgene promoters of the Flp-In cell lines used in **b**. The promoter of a housekeeping gene, *GAPDH*, was used as internal control. Immunoprecipitation (IP) was performed with antibody to Pol II or control IgG. DNA before IP was used as the input control. Standardized density quantification of Pol II loading on the transgene promoter, relative to the *GAPDH* control promoter, is shown below gel ($n = 3$). (**e**) Degradation curve of the *d2EGFP* transcript driven by the full-length 5′ UTR or positions +1–672. $T_{1/2}$ denotes half-life. *GAPDH* and β-*actin* (data not shown) were used as internal controls for real-time RT-PCR. Error bars show s.d. ($n = 6$).

Finally, to confirm that the ASP decreases transcript abundance by post-transcriptional gene silencing—that is, through mRNA degradation—we compared *d2EGFP* transcript stability in the Flp-In cell lines where expression was driven by either full-length 5′ UTR or positions +1–672. Cells were treated with a strong Pol II transcription inhibitor, actinomycin D, and *d2EGFP* transcript abundance was quantified at various time points by real-time RT-PCR. We observed a large increase (over two-fold greater half-life) in the stability of *d2EGFP* mRNA, concomitant with the loss of ASP activity (**Fig. 4e**). Together, our data indicate that the inhibitory effect of the ASP in the full-length L1 5′ UTR controls retrotransposition at a post-transcriptional level by inducing mRNA degradation.

RNAi suppresses human L1 retrotransposons *in vivo*

To provide further evidence that an RNAi mechanism elicited by the antisense transcript of the 5′ UTR can suppress L1 expression and retrotransposition, we targeted the expression of *Dicer1*, which encodes a crucial component in the RNAi pathway, with synthetic siRNAs. We managed to decrease *Dicer1* expression by up to ~70% at the mRNA level (**Fig. 5a**) and almost abolished its expression at the protein level in both HeLa (**Fig. 5b**) and 293 cells (data not shown). Knockdown of *Dicer1* in our Flp-In cell lines increased *d2EGFP* expression driven by the full-length 5′ UTR but had no effect on the no-promoter control line or the line driven by positions +1–672 of the 5′ UTR (**Fig. 5c**). Furthermore, in wild-type, untransfected HeLa cells, knockdown of *Dicer1* resulted in a significant increase in the abundance of the endogenous full-length L1 transcript, as revealed by quantitative real-time RT-PCR (**Fig. 5d**, $P = 0.001$). As the change in endogenous L1 may be masked by noise from readthrough transcripts, we further studied the effect of *Dicer1*-targeting siRNAs on the retrotransposition of a transfected L1 (**Fig. 5e**). Knockdown of *Dicer1* significantly increased L1 retrotransposition driven by the full-length 5′ UTR (by roughly two-fold; $P = 0.003$) but not that driven by positions +1–669 of the 5′ UTR, which have impaired ASP activity ($P = 0.617$).

DISCUSSION

Previous studies have indicated that RNA silencing can protect cells against transposable elements in plants and invertebrates[21]. Our data provide the first evidence that transposable elements in mammals may also be controlled by this evolutionarily conserved mechanism. The human L1-specific siRNAs observed here are among the first natural siRNAs reported in mammals. Notably, our dsRNA input is the bidirectional transcripts derived from the internal SP and the ASP located in the 5′ UTR of L1. Abundant antisense transcripts with potential regulatory functions are encoded by eukaryotic genomes[36], and our work may contribute to understanding the role of those antisense transcripts in the regulation of transcriptional output.

The ASP in the human L1 5′ UTR has been shown to generate numerous chimeric transcripts[29], and its function as an active promoter is further demonstrated by our northern blots (**Fig. 1a**). Like the sense 5′ UTR[30,31], the ASP initiates transcription at variable sites (**Fig. 1c**). This characteristic of L1 promoters is reminiscent of initiator elements that act as TATA-less promoters driving several cellular genes[37,38]. The ASP transcription start sites revealed by 5′ RACE are consistent with the previous data suggesting that there are several ASP transcription start sites in the middle of the 5′ UTR[28] and can be roughly grouped into two regions about 80 nt apart (**Fig. 1c**). Among the start sites, there are definite 'hot spots'—for example, positions +388 and +389—from which transcripts arise more

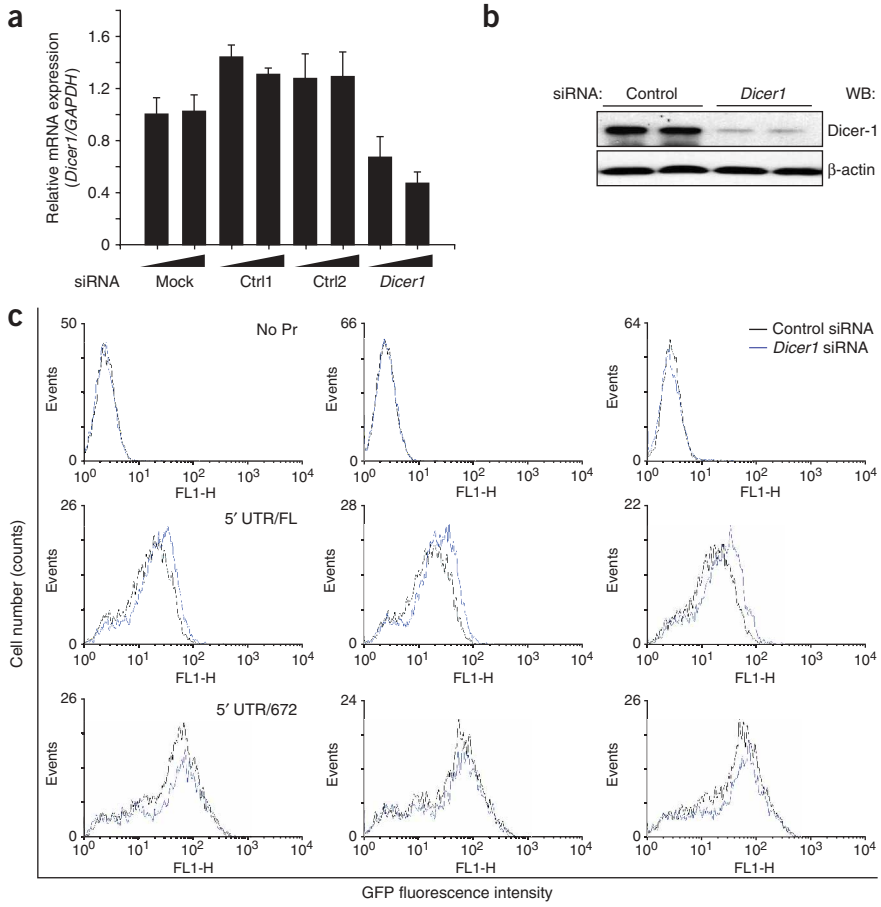

Figure 5 RNAi is an intrinsic mechanism for human L1 retrotransposon silencing. (**a**) Real-time RT-PCR of endogenous *Dicer1* knockdown at the mRNA level by synthetic siRNAs in HeLa cells. Cultured cells were treated with transfection reagent alone (Mock) or with siCONTROL nontargeting RNA no. 1 (Ctrl1), siCONTROL nontargeting pool (Ctrl2) and SMARTpool siRNAs targeting *Dicer1* at increasing concentrations. *GAPDH* was used as internal control. Data was normalized to that of the 40 nM mock group. (**b**) Western blot of endogenous Dicer1 protein knockdown by synthetic siRNAs in HeLa cells. β-actin was used as internal control. Cultured cells were treated with Ctrl2 or *Dicer1*-targeting siRNAs. (**c**) FACS analysis of d2EGFP expression in Flp-In cell lines driven by no promoter (No Pr), full-length 5′ UTR (5′ UTR/FL) or positions +1–672 (5′ UTR/672), after transfection with control (black) or *Dicer1*-targeting (blue) siRNAs. Experiments were done in triplicate, and a representative experiment is shown. (**d**) Real-time RT-PCR analysis of the transcript abundance of full-length human L1 in HeLa cells transfected with control or *Dicer1*-targeting siRNAs (*, $P = 0.001$; $n = 6$). Error bars show s.d. (**e**) Retrotransposition activity of transfected human L1 whose expression is driven by the full-length 5′ UTR (L1Hs/FL) or positions +1–669 (L1Hs/669) in the presence of control (black) or *Dicer1*-targeting siRNAs (white) in HeLa cells. A negative control L1 lacking retrotransposition activity was performed in parallel. Error bars show s.d. (*, $P = 0.003$; **, $P = 0.617$; $n = 6$).

frequently. Also, the +480–497 region seems to be less frequently used than the +378–435 region; however, this difference might be due to the bias toward shorter fragments during PCR amplification and TOPO TA cloning of 5′ RACE products.

Previously, it has been shown that synthetic L1 dsRNA can be converted into small RNAs *in vitro* and that these diced small RNAs can reduce the abundance of a 5′ UTR–driven reporter transcript and inhibit L1 retrotransposition in a transient transfection assay[39]. Here we demonstrate a specific RNAi effect derived from the intrinsic bidirectional transcription of the 5′ UTR that reduces the stability of L1 RNA *in vivo*. Recapitulating such an effect in an appropriate RNAi-deficient animal model would provide important support. However, the embryonic lethality of *Dicer1* and *Ago2* knockout mice[40,41] has prevented us from carrying out such an experiment at present.

Using deletion analysis and a luciferase reporter assay, we located the ASP activity in the region downstream of about position +400 (**Fig. 1e**). Further characterization of the ASP—for example, identification of its transcription regulatory factors—should contribute to a better understanding of related mechanisms and effects. Indeed, we have previously found that mutations in the RUNX factor–binding site (+526–508) within the ASP region (**Fig. 2a**) decreases ASP activity but increases human L1 retrotransposition[35]. As we already know that RUNX3 activates human L1 transcription and retrotransposition by binding another RUNX site (+83–101; **Fig. 2a**) in the 5′ UTR[35], it is possible that both the sense and antisense promoters in the 5′ UTR are regulated by the same transcription factor via interactions at different sites. Both of these sites are well conserved in human L1PA8 to L1PA1 families over 40 million years of evolution[42], further suggesting that they function in L1 regulation.

There are several noteworthy aspects of our siRNA data. First, L1 siRNAs are more abundant in 293 than in HeLa cells. Perhaps the transcription efficiency for both strands of a particular sequence differs between these cell lines of different tissue origins. Indeed, coexpression of sense-antisense pairs with complex tissue-specific regulation has been reported[36], and diverse spatial and temporal small RNA expression patterns have been observed[43].

Second, as abundant non-polyadenylated L1 transcripts of both strands exist exclusively in the nucleus[27] and specific, potent RNAi induced by dsRNAs lacking a 5′ cap and a polyadenylated tail has been reported to occur solely in the nucleus[44,45], we speculate that the RNAi of L1 may also be confined to the nucleus. Indeed, this could explain why native dsRNAs of L1 do not elicit an interferon response, even though dsRNAs longer than 30 nt usually do in mammalian cells[46].

Lastly, we observed an increase in d2EGFP expression with the deletion of 5′ UTR positions +673–906 in Flp-In 293 cells even in the presence of endogenous L1 siRNAs (**Fig. 4**). One explanation could be that the concentration of small RNAs specific for young L1s and identical in sequence to the 5′ UTR of $L1_{RP}$ (an active full-length human L1 retrotransposed into the *RP2* gene in an X-linked retinitis pigmentosa patient, reviewed in ref. 2) is low. Another possibility is that L1 siRNA has a *cis*-preference—that is, siRNAs generated by one L1 may act preferentially on that specific element. This might occur if the silencing complex is packaged in the proximity of the element. Indeed, *cis*-preference occurs in L1 retrotransposition, as proteins encoded by a particular L1 act preferentially on that element[4,5]. From an evolutionary standpoint, *cis*-preference, whether in the RNAi effect or the retrotransposition mechanism, minimizes the effect of one transposable element on other transposable elements and cellular genes and is thus less likely to be eliminated by negative selection.

Other mechanisms have been reported or implicated to suppress L1 retrotransposition, such as methylation[47–49], premature transcript termination[50] and transcriptional elongation defects[51]. According to our data, the siRNAs that trigger an RNAi effect are encoded by the human L1 itself. The large majority of L1s in the mammalian genome are truncated at the 5′ end. They lack the 5′ UTR and are incapable of retrotransposition. Only those full-length elements containing an intact 5′ UTR are potentially retrotransposition competent and need to be controlled. Therefore, RNAi suppression triggered exclusively by an intact L1 5′ UTR is a highly economical control mechanism that has probably been selected for during evolution.

Our work is consistent with previous knowledge that a defective RNAi pathway increases the transcript abundance of murine retrotranspons, such as L1 and IAP[25,26]. It is further corroborated by a recent report that one class of ~20- to 24-nt siRNAs from mouse oocytes is derived from retroelements including LINEs, which can mediate mRNA degradation and presumably can suppress retrotransposons through the RNAi pathway[52]. Our experimental data, in particular the ChIP and mRNA degradation assay results (**Fig. 4d,e**), strongly suggest that L1 siRNAs function at a post-transcriptional level via mRNA degradation. However, the possibility of another mechanism for control through RNA silencing, such as the existence of L1-specific miRNAs, cannot be ruled out.

In summary, we show that an RNAi effect can suppress L1 retrotransposition by ~50%. This effect may be highly important over an evolutionary timescale. With other factors held constant, we calculate that a 50% decrease in retrotransposition rate in the human lineage after a divergence point, say between humans and rhesus 25 million years ago, would lead to half as many active L1s in the human lineage compared to the rhesus lineage in ~450 years of evolution (see **Supplementary Notes** online). In addition, RNAi triggered by L1 ASP may also affect cellular gene expression—for example, through a decrease in chimeric L1-cellular gene transcripts or the insertion of full-length L1s into genes.

METHODS

Cell culture. Mouse NIH3T3, HeLa, human osteosarcoma cell line 143B (*TK*$^-$) (all from American Type Culture Collection) and Flp-In 293 cells (Invitrogen) were cultured as described[35]. Zeocin (100 μg ml^{-1}) or hygromycin B (200 μg ml^{-1}) was added into the complete medium for culturing the original Flp-In 293 and variant Flp-In cell lines, respectively.

Transfection of synthetic siRNA. Synthetic SMARTpool siRNAs targeting human *Dicer1* or siCONTROL nontargeting siRNA no. 1 and an siCONTROL nontargeting pool (which silences firefly luciferase; all from Dharmacon) were transfected into cultured cells using Lipofectamine 2,000 (Invitrogen) following the manufacturer's instructions. Flp-In cell lines were cultured in a 24-well plate in antibiotic-free DMEM containing 10% (v/v) FBS. Upon 70%–80% confluency, we transfected siRNAs at 40 nM or 100 nM. Each experimental group was transfected in triplicate and the experiment was repeated twice. Forty-eight hours after transfection, total RNA and protein were extracted to confirm the decreased *Dicer1* expression or to quantify the endogenous L1 transcript abundance by real-time RT-PCR and western blotting. For the retrotransposition assay with siRNAs added, wild-type HeLa cells were cultured in six-well plates in antibiotic-free DMEM containing 10% (v/v) FBS and transfection was performed upon 90% confluency. Each transfection received 1 μg plasmid DNA containing *EGFP* cassette–tagged L1 and 100 nM siRNAs. Duplicate transfection was performed for each experiment group and the experiment was repeated twice. Forty-eight hours after transfection, the transfection mixture was replaced with medium containing 10% (v/v) FBS and 2.5 μg ml^{-1} puromycin. Puromycin selection was continued until the FACS analysis on day 6 after transfection.

Establishment of Flp-In 293 5′ UTR–d2EGFP cell lines. $L1_{RP}$ 5′ UTR variants driving expression of the *d2EGFP* gene (Clontech) were inserted into a pcDNA5/FRT vector (Invitrogen) with the CMV promoter deleted to create transgene constructs. Following the manufacturer's protocol (Invitrogen), a Flp-In 293 host cell line with a single Flp recombinase target (FRT) site located at a transcriptionally active genomic locus was cotransfected with the Flp-In transgene constructs and the Flp recombinase expression vector. The transiently expressed recombinase efficiently integrated the expression vector at the genomic FRT site to create the stable Flp-In cell lines. Each cell line was created by duplicate transfection using different plasmid preps. The established cell lines were examined for d2EGFP fluorescence expression by FACS on a FACScalibur system (BD Biosciences).

5′ rapid amplification of cDNA ends. HeLa cells were transfected with pGL-5′UTRas, which contains the full-length human L1 5′ UTR upstream of the luciferase gene[35]. Total RNA was extracted with the RNeasy Mini RNA kit (Qiagen) and the 5′ RACE system (Invitrogen) was used. The 5′ RACE PCR product was inserted into the TA cloning system (Invitrogen) and 100 clones were sequenced.

Northern blot. Total RNA was extracted using the RNeasy Mini RNA kit (Qiagen) from wild-type HeLa cells or HeLa cells transfected with pGL-5′UTRas or control plasmids in which the luciferase gene was driven by the $L1_{RP}$ 5′ UTR (pGL-5′UTR) or the *SV40* promoter (pGL3-Control; Promega). One microgram of each RNA sample was electrophoresed in a 1% (w/v) formaldehyde agarose gel and then transferred to a positively charged nylon membrane (Roche) in a Trans-Blot SD semidry electrophoretic transfer cell (Bio-Rad). Prehybridization was performed at 68 °C for 2 h, followed by overnight hybridization at 68 °C with a luciferase-specific riboprobe labeled with digoxigenin-11-UTP. On the next day, the membrane was washed, immunodetected with anti–digoxigenin-AP Fab fragments (Roche), visualized with the chemiluminescence substrate CDP-Star, ready-to-use (Roche) and exposed. Northern blotting for siRNA detection was as described[53].

Western blot. Total protein was extracted from cultured cells using Mammalian Protein Extraction Reagent (Pierce) at 48 h after transfection with the SMARTpool siRNA controls or *Dicer1*-targeting siRNAs. Total protein (7.5 μg) was electrophoresed in a 10% SDS-PAGE gel, transferred onto a PVDF membrane (Millipore), blocked for 2 h at room temperature with 5% (w/v) skim milk in TTBS and incubated overnight with primary monoclonal antibody to human Dicer1 (Abcam) diluted 1:500 or monoclonal antibody to β-actin (Sigma-Aldrich) diluted 1: 20,000. The membrane was then probed with secondary peroxidase-conjugated anti-mouse (diluted 1:5,000 in TTBS; Amersham Biosciences) for 1 h and visualized with an ECL detection kit (Amersham Biosciences).

Chromatin immunoprecipitation. The ChIP assay to detect Pol II loading on the promoter of Flp-In cells was performed using the Upstate ChIP assay kit following the manufacturer's instruction. Sheared DNA was incubated at 4 °C overnight with 2 μg immunoprecipitating rabbit polyclonal IgG antibody to

Pol II (N-20, Santa Cruz) or 2 μg normal rabbit control IgG (Upstate). Input and chromatin-immunoprecipitated DNA was used as PCR template for detection of Pol II loading on the transgene promoters (with primers 5′-AACCCACTGCTTACTGGCTTATC-3′ and 5′-GTACCTCAGATGGAAATGCA GAA-3′) or on the *GAPDH* promoter (with primers 5′-GACACCATGGG GAAGGTGAA-3′ and 5′-GAGTAGGGACCTCCTGTTTC-3′)[54] using the PCR Core kit (Roche).

mRNA degradation assay. Flp-In cell lines containing full-length 5′ UTR or positions +1–672 were seeded in six-well plates. Upon ~80% confluency, cells were treated with actinomycin D (Sigma) added in complete medium to a final concentration of 1 μg ml^{-1}. Total RNA was extracted at 0, 0.5, 1, 2, 4, 8 and 12 h after treatment, and triplicate samples were prepared for each time point and cell line. Total RNA (2 μg) was treated with DNase I and reverse-transcribed with random hexamer using Superscript First-Strand Synthesis System for RT-PCR (Invitrogen). Real-time PCR was then performed to quantify the transcript abundance of *d2EGFP* from each sample as described[35]. The experiment was repeated twice.

Fluorescence microscopy. Fluorescence microscopy was performed as described[53].

Luciferase reporter assay. The luciferase reporter assay was performed as described[35].

Retrotransposition assay. Retrotransposition activities of L1 constructs were measured as described[33].

Statistical analysis. Statistical analysis of data was performed using the SPSS statistics software package. All results are expressed as mean ± s.d., and $P < 0.05$ is used for significance.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank Z. Mourelatos, P. Sniegowski, S. Liebhaber and J. Mayer for helpful discussion and technical advice and D. Babushok and J. Goodier for critical reading of the manuscript. This work was supported by US National Institutes of Health grant GM 045398.

AUTHOR CONTRIBUTIONS

N.Y. contributed to the concept, performed the experiments, analyzed the data and wrote the manuscript. H.H.K. contributed to the concept, analyzed the data and revised the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare they have no competing financial interests.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Ostertag, E.M. & Kazazian, H.H., Jr. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).
3. Kazazian, H.H., Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
4. Wei, W. *et al.* Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell. Biol.* **21**, 1429–1439 (2001).
5. Kulpa, D.A. & Moran, J.V. *Cis*-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* **13**, 655–660 (2006).
6. Hohjoh, H. & Singer, M.F. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.* **16**, 6034–6043 (1997).
7. Martin, S.L. & Bushman, F.D. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* **21**, 467–475 (2001).
8. Feng, Q., Moran, J.V., Kazazian, H.H., Jr. & Boeke, J.D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).
9. Moran, J.V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
10. Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D. & Gabriel, A. Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808–1810 (1991).
11. Luan, D.D., Korman, M.H., Jakubczak, J.L. & Eickbush, T.H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).
12. Moran, J.V., DeBerardinis, R.J. & Kazazian, H.H., Jr. Exon shuffling by L1 retrotransposition. *Science* **283**, 1530–1534 (1999).
13. Morrish, T.A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* **31**, 159–165 (2002).
14. Boissinot, S., Entezam, A. & Furano, A.V. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **18**, 926–935 (2001).
15. Bayne, E.H. & Allshire, R.C. RNA-directed transcriptional gene silencing in mammals. *Trends Genet.* **21**, 370–373 (2005).
16. Kim, V.N. Small RNAs: classification, biogenesis, and function. *Mol. Cells* **19**, 1–15 (2005).
17. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
18. Zamore, P.D., Tuschl, T., Sharp, P.A. & Bartel, D.P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).
19. Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
20. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
21. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476 (2004).
22. Tabara, H. *et al.* The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
23. Ketting, R.F., Haverkamp, T.H., van Luenen, H.G. & Plasterk, R.H. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141 (1999).
24. Sijen, T. & Plasterk, R.H. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**, 310–314 (2003).
25. Svoboda, P. *et al.* RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269**, 276–285 (2004).
26. Kanellopoulou, C. *et al.* Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* **19**, 489–501 (2005).
27. Swergold, G.D. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* **10**, 6718–6729 (1990).
28. Speek, M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* **21**, 1973–1985 (2001).
29. Nigumann, P., Redik, K., Matlik, K. & Speek, M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* **79**, 628–634 (2002).
30. Athanikar, J.N., Badge, R.M. & Moran, J.V.A. YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* **32**, 3846–3855 (2004).
31. Lavie, L., Maldener, E., Brouha, B., Meese, E.U. & Mayer, J. The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res.* **14**, 2253–2260 (2004).
32. Matranga, C., Tomari, Y., Shin, C., Bartel, D.P. & Zamore, P.D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).
33. Ostertag, E.M., Prak, E.T., DeBerardinis, R.J., Moran, J.V. & Kazazian, H.H., Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28**, 1418–1423 (2000).
34. Tchenio, T., Casella, J.F. & Heidmann, T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **28**, 411–415 (2000).
35. Yang, N., Zhang, L., Zhang, Y. & Kazazian, H.H., Jr. An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.* **31**, 4929–4940 (2003).
36. Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566 (2005).
37. Weis, L. & Reinberg, D. Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol. Cell. Biol.* **17**, 2973–2984 (1997).
38. Parks, C.L. & Shenk, T. The serotonin 1a receptor gene contains a TATA-less promoter that responds to MAZ and Sp1. *J. Biol. Chem.* **271**, 4417–4430 (1996).
39. Soifer, H.S., Zaragoza, A., Peyvan, M., Behlke, M.A. & Rossi, J.J. A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Res.* **33**, 846–856 (2005).
40. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nat. Genet.* **35**, 215–217 (2003).
41. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
42. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).
43. Lu, C. *et al.* Elucidation of the small RNA component of the transcriptome. *Science* **309**, 1567–1569 (2005).
44. Shinagawa, T. & Ishii, S. Generation of Ski-knockdown mice by expressing a long double-strand RNA from an RNA polymerase II promoter. *Genes Dev.* **17**, 1340–1345 (2003).
45. Robb, G.B., Brown, K.M., Khurana, J. & Rana, T.M. Specific and potent RNAi in the nucleus of human cells. *Nat. Struct. Mol. Biol.* **12**, 133–137 (2005).

46. Elbashir, S.M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
47. Bourc'his, D. & Bestor, T.H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).
48. Hata, K. & Sakaki, Y. Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* **189**, 227–234 (1997).
49. Yu, F., Zingler, N., Schumann, G. & Stratling, W.H. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* **29**, 4493–4501 (2001).
50. Perepelitsa-Belancio, V. & Deininger, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* **35**, 363–366 (2003).
51. Han, J.S., Szak, S.T. & Boeke, J.D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268–274 (2004).
52. Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* **20**, 1732–1743 (2006).
53. Yang, N., Zhang, L. & Kazazian, H.H., Jr. L1 retrotransposon-mediated stable gene silencing. *Nucleic Acids Res.* **33**, e57 (2005).
54. Yu, J., Li, Y., Ishizuka, T., Guenther, M.G. & Lazar, M.A.A. SANT motif in the SMRT corepressor interprets the histone code and promotes histone deacetylation. *EMBO J.* **22**, 3403–3410 (2003).