A YY1-binding site is required for accurate human LINE-1 transcription initiation

Jyoti N. Athanikar\(^1\), Richard M. Badge\(^3\) and John V. Moran\(^1,2,\)*

\(^1\)Department of Human Genetics and \(^2\)Department of Internal Medicine, The University of Michigan Medical School, Ann Arbor, MI 48109-0618, USA, and \(^3\)Department of Genetics, University of Leicester, University Road, Leicester, LE1 7RH, UK

Received May 11, 2004; Revised and Accepted June 22, 2004

ABSTRACT

The initial step in Long Interspersed Element-1 (LINE-1) retrotransposition requires transcription from an internal promoter located within its 5\(^{-}\)-untranslated region (5\(^{-}\)-UTR). Previous studies have identified a YY1 (Yin Yang 1)-binding site as an important sequence in LINE-1 transcription. Here, we demonstrate that mutations in the YY1-binding site have only minor effects on transcription activation of the full-length 5\(^{-}\)-UTR and LINE-1 mobility in a single round cultured cell retrotransposition assay. Instead, these mutations disrupt proper initiation of transcription from the +1 site of the 5\(^{-}\)-UTR. Thus, we propose that the YY1-binding site functions as a component of the LINE-1 core promoter to direct accurate transcription initiation. Indeed, this sequence may explain the evolutionary success of LINE-1 by enabling full-length retrotransposed copies to undergo autonomous retrotransposition in subsequent generations.

INTRODUCTION

Long Interspersed Element-1s (LINE-1s or L1s) are abundant non-long terminal repeat (non-LTR) retrotransposons in mammalian genomes that mobilize via a RNA intermediate by a process termed retrotransposition (1,2). Although LINE-1s comprise \(\sim 17\%\) of human genomic DNA, most elements are 5\(^{-}\) truncated, internally rearranged, or mutated and can no longer retrotranspose (3,4). However, the average human genome is estimated to contain \(\sim 80–100\) retrotransposition-competent LINE-1 (RC-L1s) elements (5,6), and some retrotransposition events have resulted in disease (7,8).

RC-L1s are \(\sim 6.0\) kb and contain a 5\(^{-}\)-untranslated region (5\(^{-}\)-UTR), two non-overlapping open reading frames (ORF1 and ORF2) and a 3\(^{-}\)-UTR that ends in a poly(A) tail (9,10). ORF1 encodes a 40 kDa RNA-binding protein (p40 or ORF1p) (11,12), whereas ORF2 has the potential to encode a 150 kDa protein with demonstrated endonuclease and reverse transcriptase activities (13,14). Both proteins are required for retrotransposition in cis (15), which probably occurs by target site-primed reverse transcription (TPRT) (13,16,17). LINE-1 elements also usually are flanked by variable sized target site duplications (TSDs), which are generated during retrotransposition (Figure 1). However, recent studies indicate that LINE-1 retrotransposition sometimes leads to other target site rearrangements (18–21). The initial step in human LINE-1 retrotransposition requires transcription from an internal promoter located within its 910-bp LINE-1 5\(^{-}\)-UTR, which lacks a TATA box (i.e. a TATA-less promoter). Swergold (22) demonstrated that the majority of promoter activity resides within the first 600-bp of the 5\(^{-}\)-UTR and that transcription initiates at or near the first nucleotide of the element at an unconventional start site (5\(^{-}\}-GGGGG-3\(^{-}\)). Subsequent studies revealed that the Yin Yang 1 (YY1) protein could bind the 5\(^{-}\)-UTR between position (+21 to +13) on the antisense strand, and that mutations in this sequence reduced the ability of a ‘minimal’ LINE-1 promoter (i.e. a promoter containing the first 155 bp of the LINE-1 5\(^{-}\)-UTR) to direct transcription of a heterologous reporter gene (23–25). Besides the YY1-binding site, two putative SRY-related transcription factor binding sites, a putative RUNX3 transcription factor binding site, as well as an antisense promoter have been identified within the 5\(^{-}\)-UTR (26–29). Thus, the 5\(^{-}\)-UTR most probably contains multiple cis-acting sequences that function in LINE-1 transcriptional regulation. However, how LINE-1 maintains the integrity of its 5\(^{-}\) end has long been a mystery.

We demonstrated previously that the 5\(^{-}\)-UTR is sufficient to drive LINE-1 transcription and retrotransposition in HeLa cells. In addition, we showed that a heterologous promoter [e.g. the cytomegalovirus (CMV) immediate early promoter] could replace the 5\(^{-}\)-UTR without drastically affecting retrotransposition (15). By comparison, deletion of both the 5\(^{-}\)-UTR and CMV promoter results in a three order of magnitude reduction in retrotransposition efficiency (15).

Here, we have used a single round cultured cell retrotransposition assay in conjunction with a heterologous luciferase-based reporter system to examine the function of the YY1-binding site in LINE-1 retrotransposition and transcription. Unexpectedly, we have found that mutations in the YY1-binding site have only minor effects on LINE-1 retrotransposition and the transcriptional activation potential of the full-length 5\(^{-}\)-UTR. Instead, these mutations disrupt the proper initiation of transcription from the +1 site of the 5\(^{-}\)-UTR. Thus, we propose that the YY1-binding site is a component of the LINE-1 core promoter that functions to direct accurate transcription initiation.

\*To whom correspondence should be addressed. Tel: +1 734 213 2921; Fax: +1 734 763 3784; Email: moranj@umich.edu
DNA preparation and DNA sequencing

Plasmid DNAs were purified on Qiagen Maxi or Midi Prep columns in accordance with the manufacturer’s instructions. Plasmid DNAs were extracted using phenol/chloroform/isoamyl alcohol (25:24:1) and were then precipitated using ethanol. Plasmids were sequenced on an Applied Biosystems DNA sequencer (ABI 3730) at the University of Michigan Core facilities.

Plasmid construction

pCEP4-based retrotransposition constructs. Wild-type (e.g. JM101/L1.3) and mutant (e.g. ΔΔJM101/L1.3) pCEP4-based LINE-1 constructs containing the mneoI retrotransposon cassettes have been described previously (5,15,30). Wild-type and mutant pCEP4-based LINE-1 constructs in this study were created using standard PCR mutagenesis techniques. Briefly, wild-type and mutant LINE-1 5'-UTR fragments were PCR amplified using PhiTurbo polymerase (Stratagene) with a 5'-primer harboring a NotI site (e.g. JA1; Supplementary Table 1) and a reverse primer complementary to sequences near the 3' end of the 5'-UTR, which contains an engineered EagI site (e.g. JA8; Supplementary Table 1). The primer pairs used to amplify each L1 5'-UTR fragment are provided in Supplementary Table 1. The resultant PCR fragments were digested with EagI and were subcloned into pΔΔJM101/L1.3 (30). Restriction mapping and DNA sequencing confirmed the orientation of the insert. Notably, a wild-type LINE-1 5'-UTR fragment containing an EagI site (1–910 EagI) was constructed and serves as an equivalent positive control.

Luciferase reporter constructs. Wild-type LINE-1 5'-UTR or mutant LINE-1 5'-UTR derivatives were PCR amplified from their respective pCEP4-based constructs with platinum Taq (BRL) using primers JA43 and JA44b (Supplementary Table 2). The ATG of ORF-1 in the reverse primer was destroyed so the first ATG is that of the luciferase gene. The resultant PCR fragments were digested with XhoI and HindIII, and were then subcloned into the pGL3 basic vector (Promega). Restriction mapping and DNA sequencing confirmed the orientation of the insert.

Cell culture and transfections

N-Tera 2D1 teratocarcinoma cells and PA-1 ovarian carcinoma cells were obtained from the American Type Culture Collection (Manassas, VA). HeLa cells and N-Tera 2D1 cells were grown in 7% CO2 at 37°C in MEM supplemented with 10% fetal bovine calf serum and 1x penicillin-streptomycin-l-glutamine (Invitrogen). PA-1 cells were grown in 7% CO2 at 37°C in MEM supplemented with 10% heat-inactivated fetal bovine serum and 1x non-essential amino acids (Invitrogen). PA-1 cells were plated in 6-well dishes at a density of 2 × 10^5 cells/well for luciferase experiments, or 2 × 10^5 cells/well for retrotransposition assays. Transfections were performed with the Fugene 6 transfection reagent (Roche Molecular Biochemicals) as described previously (31). Briefly, 1 μg of either a pCEP4-based LINE-1 construct or a luciferase reporter plasmid was co-transfected into cells with 0.5 μg of a β-galactosidase reporter construct.

**Figure 1.** An overview of the LINE-1 retrotransposition assay. (A) Rationale of the assay: depicted is a schematic diagram of a full-length retrotransposition-competent L1. ORF1 is indicated by the yellow rectangle. ORF2 is indicated by the blue rectangle. The 5'-UTR and 3'-UTR are indicated by the gray rectangles. A SV40 poly(A) signal present at the 3' end of the L1 is indicated by the (A)n. The relative position of the YY1-binding site on the antisense strand of the 5'-UTR is indicated, as are the sequences of the YY1-scramble (YY1-s) and YY1-forward (YY1-f) mutants. The mneoI retrotransposition indicator cassette was inserted into the 3' -UTR of wild-type and mutant L1 constructs (15,35). The cassette consists of a backward copy of the neomycin phosphotransferase gene (pink rectangle) and is interrupted by an intron (IVS-2 from the -globin gene), which is in the same transcriptional orientation as the L1. SD and SA indicate the splice donor and splice acceptor sites. The cassette also is equipped with its own promoter (P') and polyadenylation signal (A'). This arrangement ensures that a functional NEO transcript will only be translated following LINE-1 retrotransposition (15). The putative structure of a resultant retrotransposition event that confers G418-resistance (G418R) to HeLa cells is indicated at the bottom of the figure. The horizontal arrows indicate TSDs that are generated upon LINE-1 retrotransposition. (B) The results of the retrotransposition assay: retrotransposition was assayed in HeLa cells using the transient retrotransposition assay (31). Approximately 2 × 10^5 HeLa cells/well were transfected with JM101/L1.3 (WT), 1–910 EagI, YY1-s or YY1-f. No promoter indicates cells transfected with a LINE-1 construct that lacks a promoter (ΔΔJM101/L1.3) (30).

**MATERIALS AND METHODS**

Oligonucleotides

The names and sequences of oligonucleotides used in this study are provided in Supplementary Tables 1 and 2.

---

**Figure 1.** An overview of the LINE-1 retrotransposition assay. (A) Rationale of the assay: depicted is a schematic diagram of a full-length retrotransposition-competent L1. ORF1 is indicated by the yellow rectangle. ORF2 is indicated by the blue rectangle. The 5'-UTR and 3'-UTR are indicated by the gray rectangles. A SV40 poly(A) signal present at the 3' end of the L1 is indicated by the (A)n. The relative position of the YY1-binding site on the antisense strand of the 5'-UTR is indicated, as are the sequences of the YY1-scramble (YY1-s) and YY1-forward (YY1-f) mutants. The mneoI retrotransposition indicator cassette was inserted into the 3' -UTR of wild-type and mutant L1 constructs (15,35). The cassette consists of a backward copy of the neomycin phosphotransferase gene (pink rectangle) and is interrupted by an intron (IVS-2 from the -globin gene), which is in the same transcriptional orientation as the L1. SD and SA indicate the splice donor and splice acceptor sites. The cassette also is equipped with its own promoter (P') and polyadenylation signal (A'). This arrangement ensures that a functional NEO transcript will only be translated following LINE-1 retrotransposition (15). The putative structure of a resultant retrotransposition event that confers G418-resistance (G418R) to HeLa cells is indicated at the bottom of the figure. The horizontal arrows indicate TSDs that are generated upon LINE-1 retrotransposition. (B) The results of the retrotransposition assay: retrotransposition was assayed in HeLa cells using the transient retrotransposition assay (31). Approximately 2 × 10^5 HeLa cells/well were transfected with JM101/L1.3 (WT), 1–910 EagI, YY1-s or YY1-f. No promoter indicates cells transfected with a LINE-1 construct that lacks a promoter (ΔΔJM101/L1.3) (30).
(pCMV-βGal), which was used to normalize transfection efficiency (see below).

**Retrotropisation, luciferase and β-galactosidase assays**

Retrotropisation was monitored using a transient retrotropisation assay (31), except that a β-galactosidase reporter plasmid was used to normalize transfection efficiencies. Briefly, 48 h post-transfection, duplicate plates were rinsed in 1× PBS, lysed in 400 μl of 1× cell culture lysis buffer (Promega) and incubated for ~20 min. Whole cell lysates from each well were collected in a microcentrifuge tube and 50 μl of extract was used to measure β-galactosidase activity in a standard colorimetric assay. Cell extracts were added to 1 ml of Z buffer (0.1 M sodium phosphate, pH 7.5, 0.001 M MgSO4) and 200 μl of 2-nitrophenyl-β-D-galactopyranoside (4 mg/ml in 0.1 M sodium phosphate). The reaction was allowed to proceed at 37°C for ~1–2 h and the absorbance was measured at 420 nm. Retrotropisation efficiencies are reported as normalized values (i.e. #G418 resistant foci/β-galactosidase activity units).

Luciferase experiments were performed using a Luciferase Assay System (Promega). Briefly, cells were lysed in 400 μl of 1× cell lysis buffer (from 5× Cell Culture Lysis Reagent; Promega) for 20 min with gentle agitation. The resultant cell lysates were transferred to 1.5 ml microcentrifuge tubes and were then centrifuged at 14,000 r.p.m. for 5 min. Luciferase activities were measured using a luminometer (Monolight 3010; Pharmagen) using 20 μl of cell extracts and 100 μl of Luciferin reagent (Promega). β-Galactosidase assays were performed as noted above using 5 μl of the same cell lysate. Luciferase activities in relative light units (RLUs) are reported as normalized values (i.e. RLUs/β-galactosidase activity units).

**RNase protection assays**

Tissue culture cells were plated in T-175 mm² flasks at a density of 6 × 10⁶ cells/flask and were transfected with 30 μg of wild-type or mutant luciferase-based reporter constructs using 90 μl of the Fugene 6 transfection reagent (Roche Molecular Biochemicals) (31). Cells were harvested 72 h post-transfection and total RNA was prepared using the TRIzol reagent (Invitrogen). RNA samples were treated with 100 U of RNase-free DNase I (Roche Molecular Biochemicals) for 1 h at 37°C, extracted with acid-equilibrated phenol–chloroform, and precipitated with 2 volumes of 100% ethanol. PCR products containing a T7 promoter were generated from either the wild-type or the mutant luciferase-based reporter constructs using primers JA41 and pLuc-for (Supplementary Table 2). We used a Hybaid Thermocycler programmed as follows: 1 cycle of 95°C for 10 min, 55°C for 7 min, 72°C for 1 min, followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 1 min, followed by a final extension step of 72°C for 10 min. The PCR products were gel purified and cloned into pGEM T-Easy (Promega) and the resultant clones were sequenced on an Applied Biosystems DNA sequencer at the University of Michigan Core facilities.

**5′-Rapid amplification of cDNA ends (5′-RACE)**

Poly(A) RNA was isolated from ~300 μg of the total HeLa cell RNA samples described above using a MicroPoly(A) Pure Kit (Ambion). 5′-RACE was carried out using the FirstChoice RLM-RACE kit in accordance with the manufacturer’s instructions (Ambion). Briefly, 250 ng of poly(A) RNA was treated with calf intestinal phosphatase to remove the 5′-PO₄ from degraded mRNAs, rRNA, tRNA and DNA. Tobacco acid pyrophosphatase (TAP) treatment was then used to remove the 7-methyl-G cap from full-length mRNAs. The resultant poly(A) RNA was ligated with the 5′-RACE adapter using the procedures specified by the manufacturer. Reverse transcription using the SuperScriptII RT (Invitrogen) was conducted with random decamers; 1 μl of the RT reaction was then used in the first PCR with the outer 5′-RACE and pLuc-reverse primer pairs (Supplementary Figure 1 and Supplementary Table 2). An aliquot of 2 μl of the PCR was then used as the template in a nested PCR using the inner 5′-RACE and either the 300G (JA10) or the 600G (JA9) primer pairs (Supplementary Figure 1 and Supplementary Table 2). We used a Hybaid Thermocycler programmed as follows: 1 cycle of 95°C for 10 min, 55°C for 7 min, 72°C for 1 min, followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 1 min, followed by a final extension step of 72°C for 10 min. The PCR products were gel purified and cloned into pGEM T-Easy (Promega) and the resultant clones were sequenced on an Applied Biosystems DNA sequencer at the University of Michigan Core facilities.

**Analysis of LINE-1 5′-termini**

An annotated database of full-length, human-specific L1s showing high >98% identity to a RC-L1 sequence (L1.3, accession no. L19088) was screened for elements flanked by unambiguous TSDs (32). These elements were then screened for those containing an intact YY1 non-degenerate core (NDC) sequence (5′-CAAGATGGCCG-3′). A total of 229 LINE-1 sequences including 500 bp of the 5′-flanking genomic DNA were aligned using Clustal W (33) and the alignment was refined manually to annotate the position of the 5′-TSD. The distance from the 3′ end of the 5′-TSD to the 3′ most base of the YY1 NDC sequence (5′-CAAGATGGCCG-3′) for each sequence was then calculated to give an objective measure of the L1 5′-terminus length. Analysis of this distribution revealed a strong bimodality. Most sequences exhibited a terminal distance between 10 and 48 bases. However, 10 elements were excluded from the analysis because they had terminal distances >100 bp and likely are the result of transcription from upstream cellular promoters [also known as 5′-transduction (3)]. The frequency distribution of 5′-terminus distances was plotted in 1 nt bins and the nucleotide conservation calculated as the percentage proportion of the most frequent base at each position of the alignment, weighted by the number of sequences contributing to that position. The 219 sequences analyzed, their TSD sequences and the full alignment are available at http://www.leicester.ac.uk/ge/ajj/LINE1.
Constructs are depicted in column 1. The relative retrotransposition efficiencies in HeLa cells of each construct (percentage ± SD) are presented in column 2. The relative promoter activity of the luciferase-based expression constructs in HeLa cells (column 3), N-Tera 2D1 cells (column 4) and PA-1 cells (column 5) are indicated. In each instance, the reported percentages were normalized to 1–910 EagI. The number of times each assay was performed. ND, not performed. Absolute normalized luciferase values for 1–910 EagI in each cell type are indicated at the bottom of the table.

*The average normalized relative luciferase units for 1–910 EagI are as follows: HeLa, 1 038 935; N-Tera 2D1, 18 888 182; and PA-1, 5 259 602.

**RESULTS**

**The YY1-binding site is not required for LINE-1 retrotransposition in a single round retrotransposition assay**

To test whether the YY1-binding site is important in retrotransposition, we generated a series of wild-type and mutant LINE-1 expression constructs that contain an indicator cassette (mneo) in their respective 3’-UTRs (Figure 1A) (15,34,35). The retrotransposition indicator cassette consists of a selectable marker (N EO) in the opposite orientation of the LINE-1 transcript, a heterologous promoter (P) and a polyadenylation signal (A). In addition, the NEO gene is interrupted by an intron (IVS2 of the γ-globin gene) in the same orientation as the LINE-1 transcript. This arrangement ensures that G418-resistant foci will only arise when a transcript initiated from a promoter driving LINE-1 transcription is spliced, reverse transcribed and is integrated into chromosomal DNA, thereby allowing expression of the retrotransposed NEO gene from promoter P (15).

The first two constructs [JM101/L1.3ΔCMV (WT) and pJM101/L1.3ΔCMV 1–910EagI (1–910 EagI)] contain a wild-type L1 5’-UTR and differ only by the presence of an EagI site at position +904 (see Materials and methods). The EagI site was introduced to facilitate subsequent cloning of YY1-binding site mutants, and its presence had no significant effect on retrotransposition efficiency in the cultured cell assay (Figure 1B and Table 1). Mutant derivatives of the 5’-UTR contain either a scrambled YY1 sequence (YY1-s; Figure 1A) that can no longer bind the YY1 protein (25), or an intact, but inverted YY1 core binding sequence (YY1-f; Figure 1). Both mutants contain the engineered EagI site.

Unexpectedly, we found that the YY1-s mutant retrotransposed near the level of its corresponding control (1–910 EagI) when assayed in the context of a full-length 5’-UTR (Figure 1B and Table 1). By comparison, the YY1-s mutant displayed 4- to 5-fold less activity than its corresponding control (1–150-WT) when assayed in the context of a ‘minimal’ promoter, containing the first 150 bp of the 5’-UTR (Table 1).

A similar reduction in retrotransposition efficiency was observed when the YY1-f mutant was assayed in the context of a minimal promoter (Table 1).

**Mutations in the YY1-binding site have only minimal effects on the activation potential of the full-length 5’-UTR**

In light of previous data, we next tested whether the YY1-s and YY1-f mutations affect the ability of the 5’-UTR to activate the transcription of a luciferase reporter construct in HeLa cells, N-Tera 2D1 teratocarcinoma cells and PA-1 ovarian carcinoma cells. The latter two cell types are thought to be surrogates for endogenous cell types that naturally express LINE-1 elements (12,36,37). Consistently, we found that the wild-type 5’-UTR is ~18-fold more active in N-Tera 2D1 cells when compared to HeLa cells and is ~5-fold more active in PA-1 cells when compared to HeLa cells (Table 1).

The promoter activity of the YY1-s mutant construct was indistinguishable from its corresponding control when assayed in HeLa cells and only exhibited ~2-fold less activity than its corresponding control when assayed in either N-Tera 2D1 or PA-1 cells (1–910 YY1-s versus 1–910 EagI; Table 1). By comparison, in the context of a minimal promoter, the YY1-s mutant construct displayed 3-fold less activity than its corresponding control when assayed in HeLa cells, and displayed 4- to 10-fold less activity than its corresponding control when assayed in either N-Tera 2D1 or PA-1 cells (1–150-YY1-s versus 1–150-WT; Table 1). The reduction in promoter activity also correlated with a reduction in retrotransposition efficiency (Table 1).

Inversion of the YY1-binding site had no effect on the activation potential of the full-length promoter in any cell type examined (1–910 YY1-f versus 1–910 EagI; Table 1). However, in the context of a minimal promoter (1–150 YY1-f versus 1–150-WT; Table 1), the YY1-f mutation displays a dramatic reduction in promoter activity in the three cell lines.

Together, the above data indicate that the two YY1-binding site mutations have little effect on the activation potential of the full-length 5’-UTR. However, as putative downstream
regulatory sequences are deleted, both the sequence and orientation of the YY1-binding site are critical for allowing the YY1 protein to behave as a transcriptional activator. Differences in the effects of the YY1-binding site mutations on promoter activity in the three cell lines may be due to differences in the pool of available transcriptional activators. Consistently, we observed different activities of the CMV and SV40 promoters in the three cell lines (Table 1).

The YY1-binding site is needed for transcriptional initiation near the +1 site of the 5'-UTR

Since the YY1 protein can interact with components of the basal transcriptional machinery and can also function as an initiator in TATA-less promoters [reviewed in (38)], we wished to test whether the YY1-s mutant affects the fidelity of LINE-1 transcriptional initiation. To accomplish this objective, we transfected HeLa, N-Tera 2D1 and PA-1 cells with a

![Diagram of 5'-UTR and Luciferase](image)

Figure 2. Effect of the YY1-s mutation on LINE-1 transcriptional initiation. A schematic diagram of the 1–910 EagI wild-type expression construct is indicated at the top of the figure. The YY1-s construct is identical in structure, but harbors the YY1-scramble mutation in the LINE-1 5'-UTR. RPA probes were transcribed from PCR fragments amplified from either the wild-type or YY1-s mutant luciferase constructs; the size of the probe and the predicted size of the expected protected fragment is indicated below the schematic diagram. RPA was performed with total RNAs collected from HeLa, N-Tera 2D1 (N-Tera) and PA-1 cells transfected with either a wild-type (1–910 EagI) or YY1-s (1–910-YY1-s) mutant luciferase-based expression constructs. Experiments in the left-hand panel were conducted with a RPA probe derived from the 1–910 EagI wild-type construct. Experiments in the right-hand panel were conducted with a RPA probe derived from the YY1-s mutant construct. Neg. indicates a RNA sample derived from untransfected HeLa, N-Tera 2D1 or PA-1 cells, respectively; they serve as negative controls in the experiment. The RPA probe also was incubated with yeast RNA in the presence (+) or absence (−) of RNase to visualize the full-length probe and possible probe breakdown products. The center lane (M) indicates a 100 base RNA molecular weight marker (Ambion). The arrow at the left-hand side of the figure indicates the major protected product detected with the wild-type RPA probe. The vertical red bar in the right-hand side of the figure indicates the expected placement of products initiating at or near the first base of the 5'-UTR in the YY1-s experiment. We also detect a less intense band in the Neg. lanes, which probably represents protected RNAs derived from endogenously transcribed LINE-1 elements.
findings, our data indicate that LINE-1 transcription normally initiates near the first nucleotide of the 5′-UTR (22). Moreover, the data suggest that the aberrant migration of the RPA product may be due to the secondary structure adopted by the 5′-UTR.

We next performed RPA on whole cell RNAs derived from HeLa, N-Tera 2D1 or PA-1 cells transfected with a luciferase reporter construct that was under the control of the full-length 5′-UTR containing the YY1-s mutant. We did not observe a major protected product corresponding to transcripts that initiated at the +1 position of the 5′-UTR. Instead, we observed numerous protected products that varied in size, suggesting that transcription initiated at multiple sites both within and upstream of the 5′-UTR (Figure 2, right panel). A similar loss in initiation fidelity was observed in HeLa cells transfected with the YY1-f mutant (data not shown; however, see Figure 4). Notably, although the YY1-s mutation reduces transcription by ~2-fold in both N-Tera 2D1 and PA-1 cells when compared to its corresponding control (1–910-YY1-s versus 1–910 EagI; Table 1), the absolute expression levels are still sufficient to detect any correctly initiated full-length mRNAs that are present.

To independently confirm the RPA data, we conducted 5′-RACE to map the transcription initiation sites. Experiments generated from HeLa cells transfected with wild-type constructs (WT and 1–910 EagI) demonstrated that transcription generally initiated within the first five guanosine residues present in the 5′-UTR (Figure 4). Similarly, transcripts derived from HeLa cells transfected with the ∆5G construct initiated transcription at either the first base of its 5′-UTR (i.e. the adenosine at +6) or slightly upstream at the 3′ most guanosine residue present in the engineered NotI restriction site, which is consistent with data learned from the RPA (Figures 3, right panel and 4). In contrast, cells transfected with either the YY1-s or YY1-f mutants have variable transcription initiation sites located downstream of the YY1-binding site throughout the 5′-UTR (Figure 4). Thus, we conclude that both the sequence and orientation of the YY1 site is necessary for proper transcriptional initiation in the context of a full-length 5′-UTR. Interestingly, we detected only 5′-RACE products when RNA samples were treated with tobacco acid pyrophosphatase (see Materials and Methods), indicating that LINE-1 RNA contains a cap-like structure at its 5′-terminus (Supplementary Figure 1).

Endogenous LINE-1 elements have heterogeneous 5′ ends

The above data indicate that a functional YY1 site is important to direct transcriptional initiation near the terminus of the 5′-UTR. However, the variable initiation sites observed for the wild-type and ∆5G constructs (Figures 2–4) suggest that YY1 does not specify the precise nucleotide of transcription initiation. To test this model, we used a database of 219 human-specific L1s to objectively define the LINE-1 5′-terminus (and hence the likely transcriptional start site) without reference to the consensus sequence (see Materials and methods). Although the majority (i.e. 146/219 or 67%) of genomic elements begin within the first five guanosine residues that are deleted in the ∆5G construct, there is a broad distribution about this point (Figure 5, blue line and Supplementary Figure 2), indicating heterogeneity in initiation sites.
Sequence conservation among LINE-1 elements in this region also falls precipitously (Figure 5, red line), suggesting that transcription is occasionally initiated in the 5'-flanking genomic sequences. However, despite this observed sequence heterogeneity, ~96% of the LINE-1 sequences have a purine at their 5'-terminus. Together, these data are consistent with the hypothesis that YY1 directs transcription to initiate at heterogeneous purine residues located near the 5'-end of the LINE-1 element.

DISCUSSION

In summary, we have shown that mutations in the YY1-binding site have only minor effects on transcription activation of the full-length 5'-UTR and LINE-1 mobility in a single round cultured cell retrotransposition assay. Instead, these mutations disrupt proper initiation of transcription from near the +1 site of the 5'-UTR. Thus, we propose that the YY1-binding site functions as a component of the LINE-1 core promoter to direct accurate transcription initiation.

YY1 is an essential, widely expressed Kruppel-related zinc finger protein that can either activate or repress the transcription of many cellular and viral genes (39–41). It can also recruit components of the basal transcription machinery as well as histone modification enzymes to promoters and may play a role in directing transcriptional initiation (39,42–47). Thus, our data suggest that YY1 most probably interacts with other trans-acting factors that bind downstream sequences in the L1 5'-UTR to recruit RNA polymerase II to the 5' end of the element.

Analysis of young LINE-1 sequences in the human genome draft sequence corroborates our experimental data and reveals extensive heterogeneity at the 5' end of endogenous LINE-1 sequences (Figure 5). This 5'-end heterogeneity could be due to the following: (1) the use of multiple transcriptional initiation sites; (2) untemplated nucleotide additions that occur during (−) strand cDNA synthesis [(18,20) N. Gilbert, S. Lutz and J.V. Moran, unpublished data]; (3) 5'-truncation of the LINE-1 element during the process of retrotransposition; or (4) 5'-transduction (3). Although all of these processes most probably contribute to the observed LINE-1 5'-end heterogeneity, our data clearly demonstrate that there is no strict consensus sequence that is required for LINE-1 transcription initiation. Instead, they are consistent with the notion that YY1 directs transcription to initiate at heterogeneous purine residues located near the 5' end of the LINE-1 element.
Figure 5. Analysis of L1 5′-termini distribution and sequence conservation. An annotated non-redundant database of full-length human-specific L1s was used to identify 219 sequences with clearly defined TSDs (32). These elements and their TSDs were aligned and the distance from the base adjacent to the 5′-TSD to the 3′ end of the YY1 NDC sequence measured. This distribution is plotted against nucleotide position (blue line). For each position of the alignment, nucleotide conservation was calculated, and weighted by the number of sequences at that position (red line). Below the plot is aligned the 90% consensus sequence derived from the alignment, and the L1PA1 consensus sequence. The YY1 NDC sequence is underlined.

Figure 6. A model for how the YY1-binding site functions in LINE-1 retrotransposition. The model depicts retrotransposition events arising from a progenitor, full-length LINE-1 element that contains either a wild-type (left panel) or a mutant (right panel) YY1-binding site. In the wild-type scenario, a full-length LINE-1 element containing an intact 5′-UTR predominantly initiates transcription from the +1 site of the 5′-UTR allowing the internal promoter to be regenerated upon retrotransposition. The progenitor as well as the full-length retrotransposed product can then undergo autonomous retrotransposition in subsequent generations. By comparison, LINE-1 elements lacking a functional YY1-binding site cannot initiate transcription at the +1 position of the 5′-UTR and instead may initiate transcription from various positions within the 5′-UTR (represented by the numerous arrows). Thus, the resultant progeny will become progressively shorter, leading to the eventual formation of an element that cannot undergo autonomous retrotransposition. The red question marks (?) in the right-hand panel indicate uncertainties about whether the progeny LINE-1 element will be transcribed or whether it will remain retrotransposition-competent. It is likely that genomic sequences flanking the progenitor and progeny LINE-1 elements will influence their expression and ability to retrotranspose in vivo.
It is noteworthy that we have identified 10 LINE-1 sequences with intact YY1-binding sites that contain 5' extensions that range from 115 to 560 bp (available at http://www.leicester.ac.uk/ge/ajj/LINE1). These sequences may be derived from 5'-transduction, indicating that, even in the presence of a functional LINE-1 promoter, readthrough transcription from endogenous promoters occasionally may give rise to RC-L1 transcripts (3,22,32). Indeed, it is likely that the genomic context of a LINE-1 influences both its expression and retrotransposition capacity.

Given that mutations in the YY1-binding site have only minor effects on promoter activity and retrotransposition when present in the context of the full-length 5'-UTR, one might ask why has this sequence been conserved during evolution? Since the YY1-binding site plays an important role in directing transcriptional initiation near the +1 site of the LINE-1 5'-UTR, we propose that its presence ensures that full-length retrotransposed RNAs will retain an intact internal promoter, allowing them to undergo retrotransposition in subsequent generations (Figure 6). By comparison, although LINE-1 elements lacking a functional YY1-binding site may initiate transcription from internal sites within the 5'-UTR, the resultant progeny would probably lack an internal promoter, which rapidly would lead to their extinction as autonomously replicating elements. Ultimate proof of this hypothesis requires establishing a selection scheme that allows one to assay the retrotransposition potential of progeny LINE-1 elements in cultured cells (i.e. a two-round retrotransposition assay).

LINE-1 sequences and non-LTR retrotransposons in other organisms likely have evolved a variety of mechanisms to ensure that their respective progeny contain an internal promoter. For example, mouse and rat LINE-1 sequences, as well as avian CR1 elements may use a recombination-based mechanism to maintain their promoter structure (48–51). By comparison, the jockey element of Drosophila melanogaster contains a downstream promoter element that serves to ensure its transcription starts from a specific initiator sequence located at the +1 site of the element (52,53), whereas the D.melanogaster TART retrotransposon may use a promoter located in the 3' end of an upstream, tandemly localized HeT-A retrotransposon (54). Clearly, our data highlight yet another mechanism that non-LTR retrotransposons have evolved to maintain an internal promoter and show that there are multiple solutions to this perplexing problem.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS
We thank the members of the Moran lab for thoughtful feedback and discussions. We also thank Dr Timothy Osborne for the CMV-β-galactosidase construct and the University of Michigan Sequencing Core for all of their help. This work was supported by a grant from the National Institutes of Health (GM60518) awarded to J.V.M. J.N.A. was supported in part by a University of Michigan Cancer Biology postdoctoral training grant (CA009676) and then by a NRSA postdoctoral fellowship (F32GM20859). R.M.B. was supported by a Wellcome Trust Project Grant (069702/Z/02/Z).

REFERENCES


