Cell, Vol. 110, 315–325, August 9, 2002, Copyright ©2002 by Cell Press

Genomic Deletions Created upon LINE-1 Retrotransposition

Nicolas Gilbert,[1,2] Sheila Lutz-Prigge,[2] and John V. Moran[1]
Department of Human Genetics
Department of Internal Medicine
1241 East Catherine Street
University of Michigan Medical School
Ann Arbor, Michigan 48109

Summary

LINE-1 (L1) retrotransposition continues to impact the human genome, yet little is known about how L1 integrates into DNA. Here, we developed a plasmid-based rescue system and have used it to recover 37 new L1 retrotransposition events from cultured human cells. Sequencing of the insertions revealed the usual L1 structural hallmarks; however, in four instances, retrotransposition generated large target site deletions. Remarkably, three of those resulted in the formation of chimeric L1s, containing the 5′ end of an endogenous L1 fused precisely to our engineered L1. Thus, our data demonstrate multiple pathways for L1 integration in cultured cells, and show that L1 is not simply an insertional mutagen, but that its retrotransposition can result in significant deletions of genomic sequence.

Introduction

L1s are abundant retrotransposons that comprise ~17% of human DNA (Lander et al., 2001). Most L1s are retrotransposition-defective because they are 5′ truncated, internally rearranged, or mutated; however, ~60–100 L1s remain retrotransposition-competent (RC-L1s; Sassaman et al., 1997). RC-L1s are 6.0 kb and contain a 5′ untranslated region (UTR), two nonoverlapping open reading frames (ORF1 and ORF2), and a 3′ UTR ending in a poly (A) tail (Figure 1A; reviewed in Moran and Gilbert, 2002). ORF1 encodes a 40 kDa nucleic acid binding protein (Hohjoh and Singer, 1996, 1997), whereas ORF2 encodes a protein with endonuclease (L1 EN) and reverse transcriptase (L1 RT) activities (Feng et al., 1996; Mathias et al., 1991); both proteins are required for retrotransposition (Moran et al., 1996).

L1 retrotransposition requires the transcription of L1 RNA, its transport to the cytoplasm, and translation of its two open reading frames (reviewed in Moran and Gilbert, 2002). Recent genetic experiments indicate the L1-encoded proteins demonstrate a *cis*-preference (Esnault et al., 2000; Wei et al., 2001) by preferentially associating with their encoding transcript to form a ribonucleoprotein particle (RNP), which is a proposed retrotransposition intermediate (Hohjoh and Singer, 1996; Martin, 1991). The RNP presumably is transported to the nucleus, where L1 EN cleaves genomic DNA at a degenerate consensus target sequence (5′-TTTT/A-3′, and variants of that sequence; Cost and Boeke, 1998; Feng et al., 1996; Morrish et al., 2002). The liberated 3′ hydroxyl then serves as a primer for reverse transcription of L1 RNA by L1 RT in a process termed target-site primed reverse transcription (TPRT; Luan and Eickbush, 1995). The nascent L1 cDNA ultimately is joined to genomic DNA resulting in the typical L1 structural hallmarks (i.e., frequent 5′ truncations, the presence of a 3′ poly (A) tail, and either small duplications or small deletions of target site nucleotides). However, little is known about how L1 integration is completed.

L1s alter the genome by propagating themselves, as well as by mobilizing sequences derived from their 3′ flanks to new genomic locations (L1-mediated transduction; Moran et al., 1999). The L1-encoded proteins also may function *in trans* to mediate the mobility of Alu elements and processed pseudogenes, which comprise ~10% of genomic DNA (Boeke, 1997; Esnault et al., 2000; Wei et al., 2001). Thus, either directly or by the mobilization of cellular RNAs, L1 is responsible for ~30% of human DNA. Furthermore, nonallelic DNA recombination events between L1s also have been implicated in genomic rearrangement and disease (reviewed in Moran and Gilbert, 2002).

L1 integration is most clearly understood by comparing the sequences of the pre- and postintegration sites. However, because most L1s retrotransposed in the distant past and are now fixed in the human population, it often is difficult to obtain information about the preintegration site. Comparisons of DNA sequences between orthologous loci in humans and nonhuman primates and the identification of dimorphic L1s have yielded insight into the structures of preintegration sites (Ovchinnikov et al., 2001; Sheen et al., 2000). However, mutations and secondary DNA rearrangements occurring after the original retrotransposition event can complicate these analyses. Thus, the study of de novo insertions has relied primarily on the characterization of relatively few disease-producing mutations.

Here, we developed a system to rapidly clone and characterize new L1 retrotransposition events from human HeLa cells. Detailed analysis of 37 events demonstrated typical L1 structural hallmarks. Remarkably, our data also revealed that L1 retrotransposition can result in interstitial (>3 kb) deletions, and that the nascent L1-cDNA can undergo recombination with endogenous L1s, resulting in the formation of chimeric elements. Thus, we have uncovered mutational mechanisms by which L1 retrotransposition continues to alter the human genome.

Results

A System to Rapidly Recover New L1 Retrotransposition Events

We previously tagged L1s with a selectable marker (*mneoI*) that, upon retrotransposition, conferred resis-

[1]Correspondence: moranj@umich.edu (J.V.M.); gilbertn@umich.edu (N.G.)
[2]These authors contributed equally to this work.

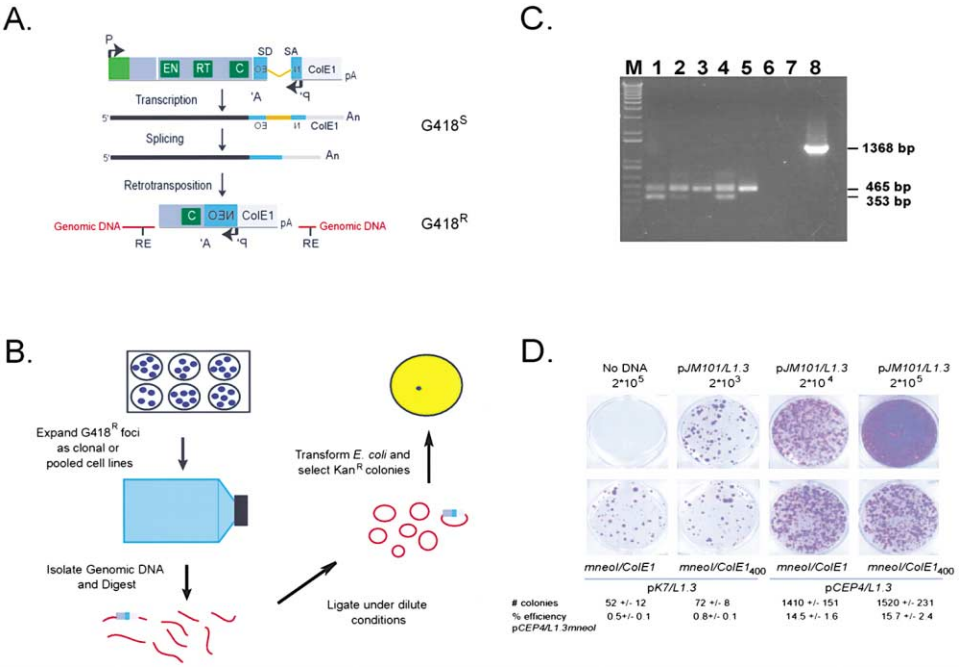

Figure 1. An Overview of the L1 Recovery System

(A) Rationale of the assay. The 3′UTR of a human RC-L1 was tagged with a reporter cassette designed to detect retrotransposition events. The gray rectangles indicate L1 ORF1 and L1 ORF2, respectively, and the relative positions of the endonuclease (EN), reverse transcriptase (RT), and cysteine-rich domain (C) are indicated. The positions of the L1 promoter (P) and the SV40 late polyadenylation signal (pA) also are indicated. The light blue rectangle indicates the *mneoI* gene and the lavender rectangle indicates the *ColE1* bacterial origin of replication. The relative positions of the prokaryotic/eukaryotic promoter (P′) and the thymidine kinase polyadenylation signal (A′) required for the expression of the reporter gene also are shown. The *mneoI* gene is interrupted by an intron (γ-globin intron 2, indicated by the orange line) in the opposite transcriptional orientation. SD and SA indicate the splice donor and splice acceptor sites. This arrangement ensures that a functional *Neo* transcript will only be translated following L1 retrotransposition. The putative structure of a resultant retrotransposition event that confers G418-resistance (G418^R) to HeLa cells is shown at the bottom of the figure. The red lines indicate flanking genomic DNA.

(B) Flowchart of the rescue procedure. Genomic DNA was extracted from cell lines derived from either a single G418^R focus or small pools of G418^R foci and the L1s were recovered using the protocol described (see Experimental Procedures).

(C) Results of RT-PCR analyses. RT-PCR was performed to monitor the presence of the spliced *mneoI* transcript. Molecular size standards (1 kb plus ladder; Gibco/BRL) are indicated in Lane M. The sizes of the unspliced *mneoI* product (Lane 8) and the respective cDNA products also are indicated. RT-PCR products from cells transfected with p*JM101/L1.3*, p*K7/L1.3mneoI/ColE1*, p*K7/L1.3mneoI*$_{400}$*/ColE1*, p*CEP4/L1.3mneoI/ColE1*, or p*CEP4/L1.3mneoI*$_{400}$*/ColE1* are indicated in Lanes 1–5, respectively. RT-PCR products from untransfected HeLa cells and no RNA control are indicated in Lanes 6 and 7, respectively.

(D) Results of the retrotransposition assay. L1 retrotransposition was assayed using the transient retrotransposition assay (Wei et al., 2000). The top panel displays the results from mock-transfected HeLa cells (No DNA) or HeLa cells (either 2×10^3, 2×10^4, or 2×10^5) transfected with p*JM101/L1.3* (Sassaman et al., 1997). The bottom panel displays the results from HeLa cells (2×10^5) transfected with p*K7/L1.3mneoI/ColE1*, p*K7/L1.3mneoI*$_{400}$*/ColE1*, p*CEP4/L1.3mneoI/ColE1*, and p*CEP4/L1.3mneoI*$_{400}$*/ColE1*, respectively. The relative retrotransposition efficiency of each of the constructs (relative to p*JM101/L1.3*, a previously characterized RC-L1 [Sassaman et al., 1997]). The standard deviation is indicated.

tance to G418 (G418^R) (Moran et al., 1996). To facilitate the cloning of new retrotransposition events, we modified the *mneoI* cassette to include a prokaryotic/eukaryotic promoter, a Shine Delgarno sequence, and a ColE1 origin of replication (Figure 1A). The modified cassette (*mneoI/ColE1*) was introduced either at the 3′ end or within the 3′ UTR of a human RC-L1 (L1.3) (Sassaman et al., 1997), and the resultant constructs were subcloned into a bacterial plasmid (p*K7*) or a mammalian expression vector (p*CEP4*), creating p*K7/L1.3mneoI/ColE1* and p*CEP4/L1.3mneoI/ColE1*. These changes allow expression of the retrotransposed *Neo* gene in both human and bacterial cells and ultimately enabled us to recover retrotransposition events as replicating plasmids in *E. coli* (Figure 1B).

To determine if *L1.3mneoI/ColE1* is expressed, HeLa cells were transfected with the above plasmids, and the presence of the spliced *mneoI* transcript was monitored by RT-PCR. Unexpectedly, we observed two products; one of the predicted size (465 bp), and a smaller product (353 bp) that DNA sequencing showed resulted from the use of a cryptic splice site present within the *mneoI* cassette (Figure 1C; lanes 1, 2, and 4). Because retrotransposition of incorrectly spliced products will not yield G418^R colonies, we created a silent mutation that destroyed the cryptic splice site (p*K7/L1.3mneoI*$_{400}$*/ColE1* and p*CEP4/L1.3mneoI*$_{400}$*/ColE1*; Figure 1C; lanes 3 and 5).

Both p*K7/L1.3mneoI/ColE1* and p*CEP4/L1.3mneoI/ColE1* retrotransposed in HeLa cells; though the retrotransposition efficiency was higher when L1.3 was expressed from the replication-proficient p*CEP4* vector (Figure 1D). Constructs harboring either *mneoI/ColE1*or *mneoI*$_{400}$*/ColE1* also retrotransposed at similar frequencies, indicating that the cryptic splice site does not affect the retrotransposition efficiency (Figure 1D). Moreover,

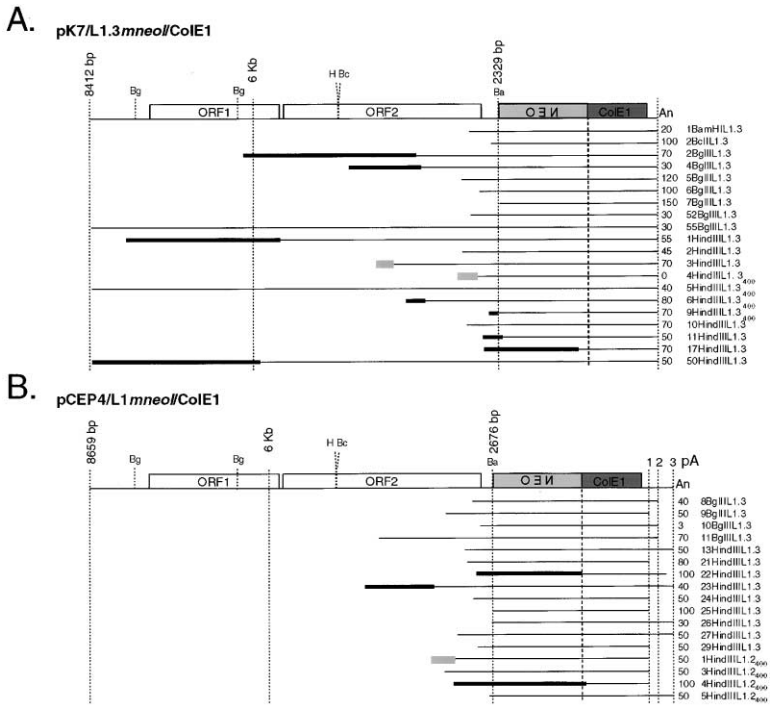

Figure 2. Recovered Events Demonstrate L1 Structural Hallmarks

(A) Size distributions of retrotransposition events from pK7/L1.3mneoI/ColE1. A schematic of a full-length L1 after retrotransposition from p*K7/L1.3mneoI/ColE1* is indicated. The relative positions of the L1 5′UTR, L1 ORF1, L1 ORF2, the retrotransposed *mneoI/ColE1* gene, and the SV40 pA signal are indicated. Lines below the figure indicate the lengths of the retrotransposed L1s. Bold black lines are used to indicate the relative sizes and positions of L1s that have undergone either an inversion/deletion or inversion/duplication. The names of the rescued clones are indicated at the right of the figure. Clones derived from the p*K7/L1.3mneoI*$_{400}$*/ColE1* vectors are denoted with a subscript 400. The estimated size of the L1 poly (A) tail, and the relative positions of the BglII (Bg), HindIII (H), BclI (Bc), and BamHI (Ba) restriction sites used in the rescue procedure also are indicated. Notably, 2329 bp must be retrotransposed to deliver the entire *mneoI/ColE1* cassette. The lengths of the chimeric L1s can only be estimated (see Figure 6A for a complete description). The gray shading indicates their respective 5′ ends.

(B) Size distributions of retrotransposition events from pCEP4/L1mneoI/ColE1. A schematic of a full-length L1 after retrotransposition p*CEP4/L1mneoI/ColE1* is indicated. Labels are the same as in Figure 3A; however, three possible polyadenylation signals exist (SV40pA$_1$ (pA1), L1pA (pA2), and SV40pA$_2$ (pA3), respectively). Thus, 2328, 2474, or 2676 bp must be retrotransposed to deliver the entire *mneoI/ColE1* cassette.

the retrotransposition efficiency of p*CEP4/L1.3mneoI/ColE1* was reduced by ~6-fold when compared to p*CEP4/JM101/L1.3*, which harbors the unmodified *mneoI* cassette (Sassaman et al., 1997; Wei et al., 2000). This decrease likely is due to the increased length of the retrotransposed product (i.e., the 880 bp ColE1 sequence) needed to confer G418-resistance to HeLa cells.

Recovered Retrotransposition Events Resemble Endogenous L1s

To clone the resultant retrotransposition events, genomic DNA was extracted from cell lines derived from either a single G418^R colony or small pools (10–150) of G418^R colonies and was digested with restriction enzymes that do not cleave within the retrotransposed *mneoI/ColE1*cassette. The digested DNA was ligated in dilute conditions and transformed into *E.coli*. Kan^R colonies contained plasmids harboring the retrotransposed *L1mneoI/ColE1* and flanking genomic DNA (Figure 1B).

We recovered 37 retrotransposition events that were derived from p*K7/L1.3mneoI/ColE1* (16 events), p*K7/L1.3mneoI*$_{400}$*/ColE1* (4 events), p*CEP4/L1.3mneoI/ColE1* (13 events), or p*CEP4*/L1.2*mneoI*$_{400}$*/ColE1* (4 events). L1.2 is a previously characterized RC-L1 (Dombroski et al., 1991; Moran et al., 1996). As expected, Southern blot analysis of eight cell lines derived from independent G418^R colonies confirmed that the sizes of the rescued plasmids were comparable to the sizes of the respective restriction fragments in genomic DNA that hybridized with a *Neo* gene probe (not shown).

Sequence analysis of the 37 clones revealed that they structurally resembled endogenous L1s (Figures 2A and 2B). Two clones (55BglII and 5HindIII) contained full-length (8.5 kb) L1 insertions. Thirty-five clones contained 5′ truncated L1s. Of these, eleven clones harbored internally rearranged L1s that contain either an inversion/deletion (9 events) or an inversion/duplication (2 events). Examination of the DNA sequences at the inversion junctions revealed that "twin priming" might be responsible for their formation (Ostertag and Kazazian, 2001).

Thirty-six of 37 L1s ended in a poly (A) tail that ranged in size from 3–150 nts, with an average length of ~60 residues. This is longer than the average poly (A) tail length present at the 3′ ends of Ta-subset L1 elements (~13 adenosines; Ovchinnikov et al., 2001). Thus, our data is consistent with the notion that dynamic mechanisms exist (e.g., replication slippage) to shorten the L1 poly (A) tail after insertion. Each insertion derived from p*K7/L1.3mneoI/ColE1* was polyadenylated at the sole SV40pA site present at the 3′ end of the construct (Figure 2A). By contrast, insertions derived from p*CEP4/L1.3mneoI/ColE1* were polyadenylated at one of three possible sites located at the 3′ end of the L1. Three events (8BglIIL1.3, 10BglIIL1.3, and 11BglIIL1.3) were polyadenylated at the native L1pA site, suggesting that it may be functional (pA2; Figure 2B). However, it is possible that those adenosine residues were added during TPRT, as proposed for the *Drosophila* I-factor non-LTR retrotransposon (Chaboissier et al., 2000).

New Retrotransposition Events Occur throughout the Genome

We used sequences flanking the L1s to identify the pre-integration site in the human genome working draft sequence (HGWD). In all but two instances (1HindIIIL1.2 and 4HindIIIL1.3; Table 1), both the 5′ and 3′ flanking

Table 1. Distribution of the Rescued Clones

Chromosome	Number of Events	Empty Site Chromosome Location/Accession Number	Name of Insertion
1	4	1p13.1 / AL3914	17HindIIIL1.3
		1p31.2 / AL391359	27HindIIIL1.3
		1q31.1 / AC073634	3HindIIIL1.3
		1q32.2 / AL355527	5HindIIIL1.3$_{400}$
2	3	2p13.1 / AC025072	2BclIL1.3
		2q24.1 / AC021837	9BglIIL1.3
		2q32.3 / 5′-AC073066, 3′-AC064834	5BglIIL1.3
3	2	3p25.1 / AC023230 in CSPG5 gene intron	8BglIIL1.3
		3q22.2 / AC026117	24HindIIIL1.3
4	1	4q26 / AC073559	21HindIIIL1.3
5	4	5p15.2 / AC026810	9HindIIIL1.3$_{400}$
		5q11.2 / 5′-AC092343, 3′-AC022128[a] in PDE4D gene intron	4HindIIIL1.3$_{400}$
		5q14.3 / AC044889	6BglIIL1.3
		5q15 / AC018806	1HindIIIL1.3
6	2	6q15 / AC016855	10BglIIL1.3
		6q23.2 / AL596188	10HindIIIL1.3
7	1	7q21.13 / AC006149	7BglIIL1.3
8	2	8p21.2 / AF263550	22HindIIIL1.3
		8q22.1 / AC027242	25HindIIIL1.3
9	0		
10	3	10p12.33 / AC069023	2HindIIIL1.3
		10p15.3 / AC022536	23HindIIIL1.3
		10q21.3 / AL133551	26HindIIIL1.3
11	3	11p15.1 / AC025620 in LDHC gene intron	5HindIIIL1.2$_{400}$
		11q12.1 / AP000781 in SSRP1 gene intron	1BamHIL1.3
		11q12.2 / AP000781 in PRG2 gene intron	4BglIIL1.3
12	3	12p12.3 / AC022334	11HindIIIL1.3
		12q22 / AC008128	50HindIIIL1.3
		12q22 / AC024190	6HindIIIL1.3$_{400}$
13	1	13q31.3 / AL356118 in GPC5 gene intron	11BglIIL1.3
14	2	14q11.2 / AL136419	55BglIIL1.3
		14q32.2 / AL163974, 3′ location unknown	1HindIIIL1.2$_{400}$
15	3	15q13.2 / AC021316 in KLF13 gene intron	3HindIIIL1.2$_{400}$
		15q21.3 / AC008131	2BglIIL1.3
		15q22.31 / AC027220	29HindIIIL1.3
16	1	16p12.2 / AC068150	52BglIIL1.3
17	0		
18	0		
19	0		
20	0		
21	1	21q22.3 / AL163301 in ADARB1 gene intron	13HindIIIL1.3
22	0		
X	1	Xq22.1 / Z70280	4HindIIIL1.2$_{400}$
Y	NA		

The features of the 37 rescued clones are listed. Column 1 indicates its chromosomal location. Column 2 indicates the number of events on a particular chromosome. Column 3 indicates the accession number of the preintegration site and whether retrotransposition occurred into a known gene. Column 4 indicates the name of the clone.

[a]There are multiple 3′ map locations, also including 3′-AC034225 or 3′-AC092343.

sequences matched unique sequences present in the HGWD with >99.5% identity. The recovered L1s were distributed among various chromosomes, and 8 of the 37 (~19%) rescued L1s integrated into known genes (Table 1). Thus, our data are consistent with previous analyses, which showed that L1s can insert into genes and that recent L1 insertions are distributed throughout the genome (Lander et al., 2001; Moran et al., 1999; Ovchinnikov et al., 2001).

Retrotransposition Events Predominantly Integrate at an L1 EN Consensus Site

Comparisons of the pre- and postintegration sequences revealed that 35 of 37 L1s integrated into a consensus L1 EN cleavage site (Figure 3). The integration site of 1HindIIIL1.2 could not be determined because the 3′ flanking sequence was absent from the HGWD. The other event integrated into an atypical site, lacked a poly (A) tail and was accompanied by a deletion of target site nucleotides (Figure 3; 4HindIIIL1.3; see below). Thus, it is a possible example of EN-independent retrotransposition (Morrish et al., 2002).

Target Site Alterations Generated upon Retrotransposition

Variable-length TSDs flanked 28 of 37 L1 retrotransposition events (Figure 3). Of these, canonical TSDs ranging from 10–16 bp flanked 18 events. By contrast, longer

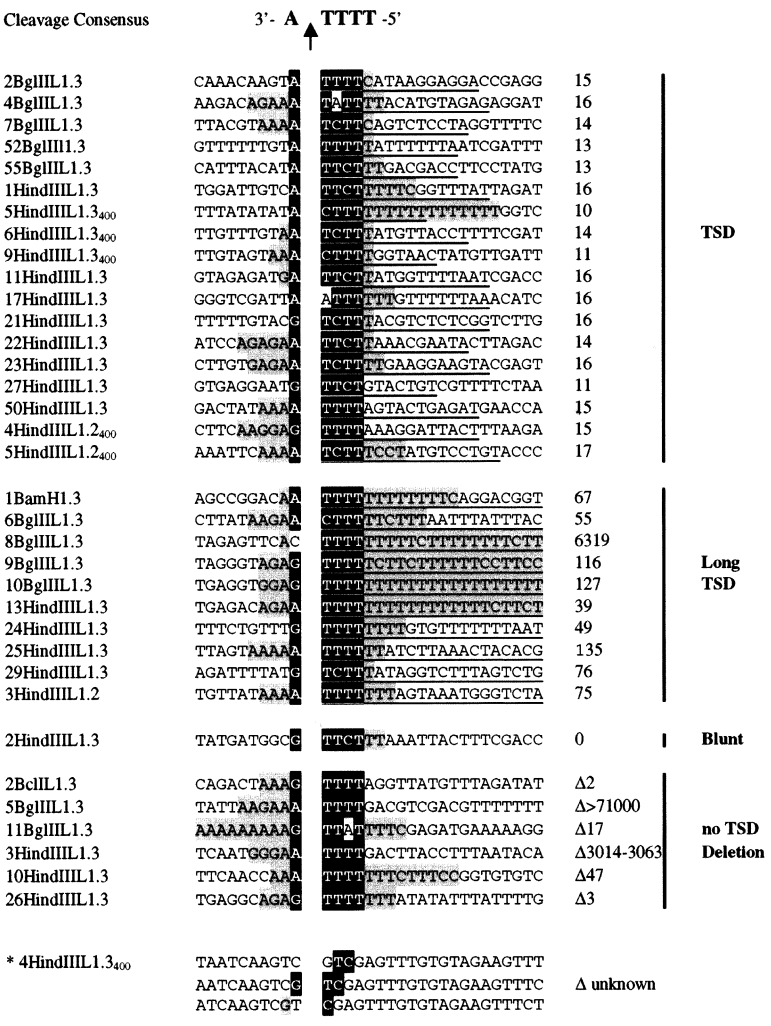

Figure 3. L1s Predominantly Integrate at an L1 EN Cleavage Site

Names of the rescued clones are indicated on the left. The "gap" in the center of the alignment indicates the bottom strand cleavage site. Black shading indicates nucleotides in agreement with the degenerate consensus L1 EN cleavage site sequence, which is indicated above the alignment (3′-A/TTTT-5′). Gray shading indicates stretches of pyrimidines or purines that are 5′ or 3′ of the bottom strand cleavage site, respectively. The sizes of the TSDs or deletions are indicated at the right of the figure. Underlining indicates nucleotides present in the TSD. The sequences of the long TSDs are available upon request. The asterisk indicates an L1 insertion (4HindIIIL1.3) that lacks a poly (A) tail. This L1 sequence also contains two ambiguous nucleotides (GT) that can be derived from either the L1 or the preintegration site. Thus, there are three possible cleavage sites (5′-GCTG/C, 5′-AGCT/G, or 5′-GAGC/T). The integration site of 1HindIIIL1.2 could not be determined (see text for details).

TSDs ranging from 39–135 bp flanked nine events. The longer TSDs are somewhat unorthodox, though TSDs of up to 60 bp have been identified in the HGWD (S.T. Szak and J.D. Boeke, personal communication). Finally, in one event (8BglIIL1.3), a TSD of 6319 bp flanked L1, and its generation resulted in the partial duplication of the first intron of the *CSPG5* gene. Thus, L1 retrotransposition can generate large TSDs. However, DNA recombination between the TSDs may render these events unstable during genome evolution, which is similar to the situation observed for the long terminal repeats of endogenous retroviruses (Liao et al., 1998).

Nine events lacked TSDs (Figure 3). In one event (2HindIIIL1.3), retrotransposition neither resulted in a gain nor loss of target site nucleotides; thus, it is an example of conservative retrotransposition. In eight events, retrotransposition resulted in the deletion of target site nucleotides. Four events (2BclIL1.3, 11BglIIL1.3, 10HindIIIL1.3, and 26HindIIIL1.3) were accompanied by small deletions that ranged in size from 2–47 bp, which is in accord with previous in vivo and in vitro studies (reviewed in Moran and Gilbert, 2002). The remaining events (5BglIIL1.3, 3HindIIIL1.3, 4HindIIIL1.3, and 1HindIIIL1.2) had novel structures, and each is characterized below.

Interstitial Deletions Are Created upon Retrotransposition

Sequence comparisons between the pre- and postintegration sites of 5BglIIL1.3 suggested that L1 retrotransposition resulted in a >71 kb deletion of target-site sequence (Figure 4A). To determine if the deletion existed prior to retrotransposition, we conducted a series of PCR amplifications on genomic DNA from either naïve HeLa cells or a polyclonal cell line containing the 5BglIIL1.3 insertion (SLP5BglII; Figure 4B). PCR reactions using the 5′ flank/3′ flank primers only amplified the predicted 3.2 kb product from SLP5BglII DNA; we never amplified a 221 bp product from naïve HeLa cells, which would be expected if a deletion occurred prior to retrotransposition.

To confirm that the HGWD in this interval was assembled correctly and that these sequences were present in HeLa cells, we performed PCR reactions with the following primers: (1) 5′ flank/5R; (2) 5′ flank/5R6K; (3) 5′ flank/5R12K; (4) 3F12K/3′ flank; (5) 3F6K/3′ flank; or (6) 3F/3′ flank. In every case, we amplified the expected sized products from both HeLa and SLP5BglII cells (see Figure 4B). Thus, our data indicate that retrotransposition resulted in a large target site deletion of at least 24 kb (derived by adding the sizes of the PCR products

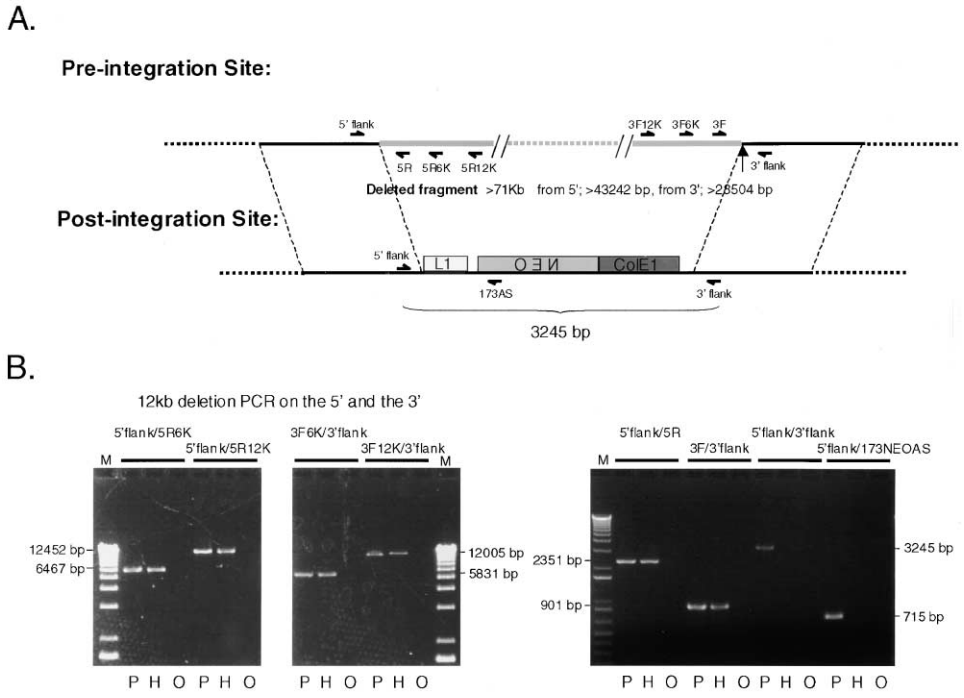

Figure 4. L1 Retrotransposition Can Result in Interstitial Deletions

(A) Schematic representation of the pre- and postintegration sequences. The structures of the preintegration site present in the HGWD and the postintegration site of 5BglIIL1.3 are indicated. Black lines indicate sequences present in both the pre- and postintegration sites. The thick gray lines indicate sequences that are present only in the preintegration site. The horizontal arrows indicate the relative positions of PCR primers. The vertical arrow indicates the L1 EN cleavage site, where integration began. The rectangles indicate the structure of the 2.9 kb retrotransposed *L1mneoI/ColE1*.

(B) PCR analyses. PCR reactions confirmed that at least 24 kb of DNA was deleted upon L1 retrotransposition. PCR primers are indicated above the gel. M indicates 1 kb plus molecular weight markers (Gibco/BRL). P indicates PCR products from SLP5BglII DNA, which contains the 5BglIIL1.3 insertion. H indicates PCR products from naïve HeLa cell DNA. O indicates PCR with no DNA. As an additional control, we demonstrated that PCR reactions using the 5′ flank/173NEOAS primer pair only amplified the expected sized product in SLP5BglII DNA. Sequencing of this product demonstrated that it was identical to the analogous segment in 5BglIIL1.3.

amplified by 5′ flank/5R12K (12 kb) and 3′ flank/3F12K (12 kb)). However, based upon the HGWD, the deletion may be >71 kb.

Chimeric L1s Are Created upon Retrotransposition

Three retrotransposition events (3HindIIIL1.3, 4HindIIIL1.3, and 1HindIIIL1.2) resulted in the generation of chimeric L1s that contain the 5′ end of an endogenous L1 precisely joined to the corresponding position of the *L1mneoI/ColE1* (Figure 5A). To confirm that the chimeric L1s were not generated by homeologous DNA recombination in *E. coli* during the rescue procedure, we conducted a series of PCR reactions on genomic DNA from clonal cell lines harboring the respective retrotransposition events. In each instance, PCR with a 5′ flanking DNA primer and 173NEOAS amplified products of the same size as those present in the rescued clone, and we confirmed that the sequences of the PCR products were identical to those in the rescued clones (Figure 5A and not shown).

Comparisons of the DNA sequences flanking the chimeric L1s with the respective preintegration sites in the HGWD revealed that both 4HindIIIL1.3 and 1HindIIIL1.2 integrated into repeat rich sequences that were incompletely assembled. PCR reactions using primers flanking each insertion amplified the resultant integration events (not shown); however, we could not amplify the preintegration site from naïve HeLa cell DNA. Thus, our data suggest that formation of these chimeric L1s was accompanied by a concomitant deletion of target site nucleotides, though we could not determine the sizes of the respective deletions.

By contrast, 3HindIIIL1.3 integrated into a fully assembled contig, and both PCR and Southern blot analyses confirmed that the formation of the chimeric L1 resulted in the concomitant deletion of 3.1 kb of intervening sequence (see Figures 5B–5E). Notably, we were unable to amplify a PCR product with 3HindIIIint/173NEOAS from genomic DNA derived from either the original polyclonal cell line containing the 3HindIIIL1.3 insertion or from two clonal cell lines derived from the original polyclonal cell line (see Figure 5B). Thus, our data indicate that the chimeric L1 present in 3HindIIIL1.3 did not result via a two step process in which a tandemly integrated L1 subsequently was resolved by homeologous DNA recombination, but that it most likely was generated during retrotransposition.

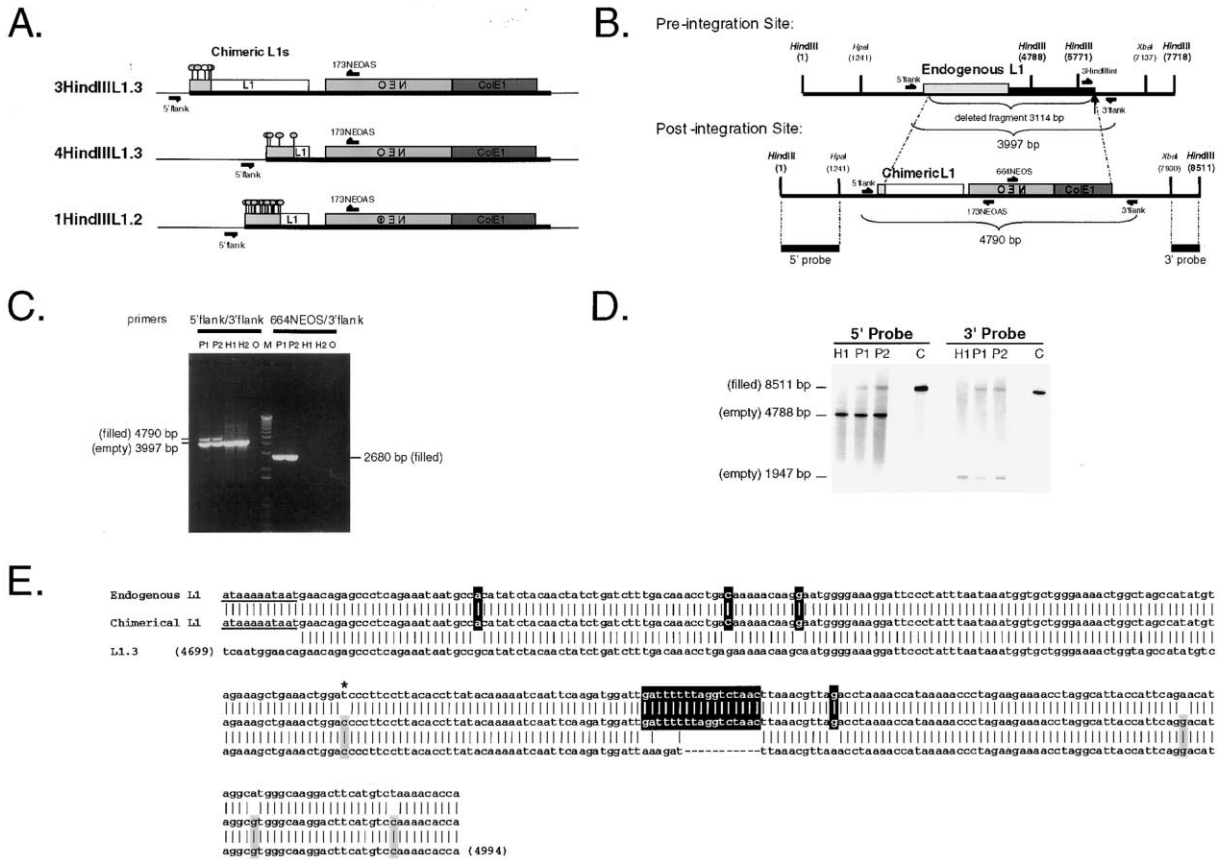

Figure 5. L1 Retrotransposition Can Result in Chimeric L1s

(A) Schematic representation of three chimeric L1s. The structures of 3HindIIIL1.3, 4HindIIIL1.3, and 1HindIIIL1.2 are indicated by the rectangles. Different shading indicates the junction between the endogenous L1 (gray rectangle) and the retrotransposed L1 (white rectangle). Lollipops indicate the relative positions of nucleotide sequence changes that differentiate the endogenous L1 from the retrotransposed *L1mneoI/ColE1*. The white lollipop indicates the position of a nucleotide specific to *L1.3mneoI/ColE1* (see description in 5E). There are 23 nucleotide differences in a 374 bp segment between 1HindIIIL1.2 and p*L1.2mneoI/ColE1*. Arrows indicate the relative position of PCR primers.

(B) Schematic representation of the pre- and postintegration sequences of 3HindIIIL1.3. The structures of the preintegration and the postintegration site of 3HindIIIL1.3 are shown. Labels are as in Figure 5A and the relative positions of the restriction sites are indicated. Also indicated are sequences deleted upon retrotransposition (3114 bp; the 5′ end of the deletion size was derived from the position of the first nucleotide polymorphism that exists between 3HindIIIL1.3 and p*L1.3mneoI/ColE1*). The sizes of the predicted PCR products generated from the 5′ flank and 3′ flank primer pairs for both the preintegration (3997 bp) and postintegration (4790 bp) site and the relative positions of the 5′ and 3′ probes used in Southern blot analysis also are indicated. Primer 3HindIIIint was used to determine whether tandemly integrated L1s were present.

(C) PCR analysis. Amplification of the pre-and postintegration sites confirmed that 3.1 kb of DNA was deleted upon the creation of the chimeric L1. PCR primers are indicated above the gel and their relative location is shown in Figure 5B. M indicates 1 kb plus molecular weight markers (Gibco/BRL). P1 and P2 indicate PCR products from the DNA of two different clonal cell lines containing 3HindIIIL1.3. H1 and H2 indicate PCR products generated either from HeLa cell DNA or from 50HindIIIL1.3 DNA, which contains a different (i.e., irrelevant) retrotransposed L1 sequence. O indicates PCR with no DNA.

(D) Southern analysis. Genomic DNA was isolated from naïve HeLa cells (H1) or from the two clonal cell lines containing the 3HindIIIL1.3 insertion (P1 and P2). 10 μg of DNA was restricted with HindIII and was subjected to Southern blot analysis using either the 1.2 kb 5′ probe or the 581 bp 3′ probe. Lane C contains DNA from the 3HindIIIL1.3 rescued clone, which was linearized with HindIII. It serves as both a size and hybridization control. The sizes of the respected fragments are indicated.

(E) Nucleotide sequence at the junction of 3HindIIIL1.3. Black shading indicates nucleotides diagnostic for the endogenous L1, whereas the gray shading indicates nucleotides diagnostic for the retrotransposed L1. The asterisk indicates an engineered sequence change in p*L1.3mneoI/ColE1*, which removes a BamHI site. SNP typing showed that the endogenous L1 contains the BamH1 site (data not shown). L1 nucleotides are numbered according to a reference L1 (L1.2; Dombroski et al., 1991). Underlining indicates the TSD flanking the endogenous L1.

Discussion

In summary, we developed a high-throughput system to characterize new L1 retrotransposition events from cultured cells, and as a result, have discovered mechanisms by which L1 retrotransposition continues to alter the human genome. Because our system allows the examination of retrotransposition events soon after their insertion, the possible effects of negative selection, DNA sequence mutation, and secondary DNA rearrangements

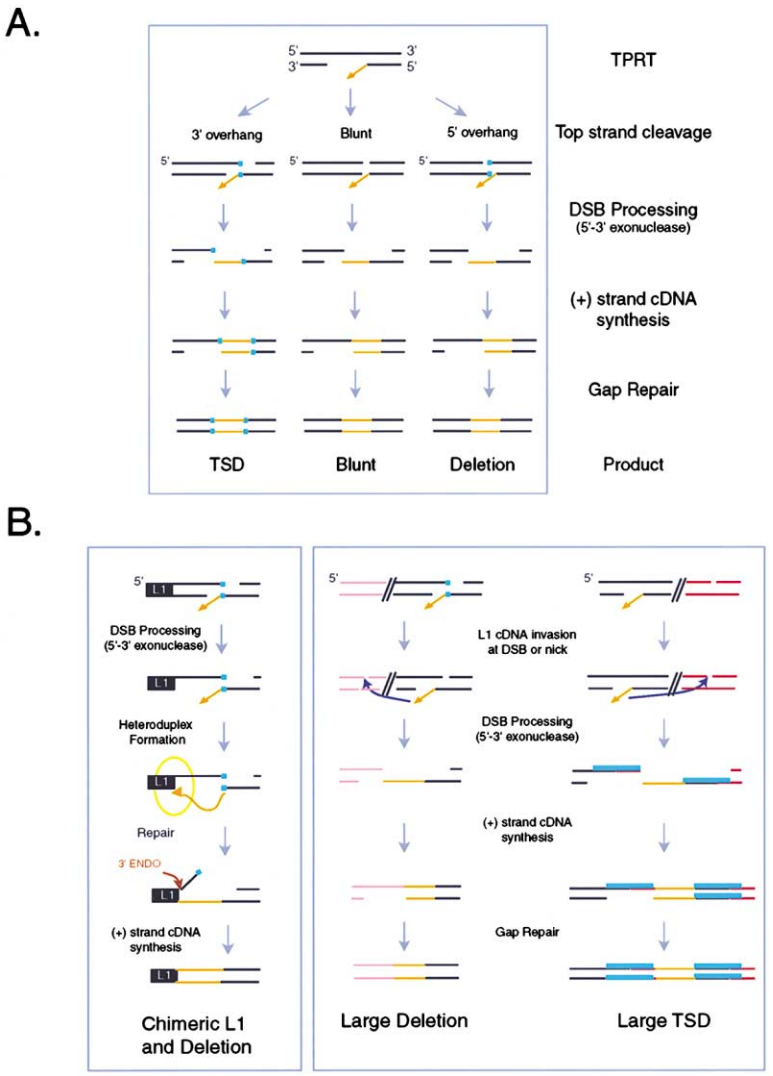

Figure 6. Models for L1 Integration

(A) Target site alterations created upon L1 retrotransposition. A model proposing how L1 retrotransposition results in various target site alterations (i.e., TSDs, blunt insertions, or small deletions) is shown. Each event is initiated by L1 EN cleavage of the bottom strand followed by L1 cDNA synthesis (represented by the orange arrow). The position of top strand cleavage can create a 3′ overhang (left), no overhang (i.e., a blunt end; middle), or a 5′ overhang (right). The proposed intermediate can be recognized as a substrate by the host repair machinery resulting in 5′-3′ exonucleolytic processing. Plus (+) strand L1 cDNA synthesis by either L1 RT or host enzymes followed by recombinational repair would lead to the final product. This model accounts for the generation of TSDs (indicated by the light blue rectangles on the left side), conservative insertions (middle) or small deletions (right side).

(B) Generation of other target site alterations. Modifications of the above model explain how retrotransposition can result in the formation of chimeric L1s, large deletions, or long TSDs. Details are provided in the text. L1 cDNA is indicated in orange. TSDs are indicated by the light blue rectangles. The black rectangle in the left panel indicates an endogenous L1. ENDO indicates a proposed substrate for a 3′ processing endonuclease. The pink lines in the center panel indicate endogenous DNA sequences located upstream of the L1 integration site. The red lines in the right panel indicate endogenous DNA sequences located downstream of the L1 integration site. A gap is used to indicate the position of a double-strand break or single-strand nick. The dark blue arrow indicates the initiation site of second (+) strand cDNA synthesis.

are minimized. Thus, we can accurately characterize both the pre- and postintegration sites.

All the events characterized in our study had L1 structural hallmarks. The majority of events (35 of 37) were 5′ truncated or contained internal rearrangements; however, two retrotransposition events were 8.5 kb in length. Thus, we conclude that engineered L1s can retrotranspose full-length copies of themselves. Moreover, because both events contained an intact 5′UTR, but lacked flanking vector sequences, our data provide experimental support that L1 transcription is initiated from the first nucleotide of the 5′UTR (Swergold, 1990). Finally, we note that three internally rearranged L1s were >6 kb in length; thus, 5 of 37 (~14%) insertions retrotransposed a "genomic equivalent" of L1 DNA. This is in general agreement with previous studies, which demonstrated that 2 of 13 (~15%) mutagenic L1 insertions are full-length (reviewed in Moran and Gilbert, 2002).

Our data provides strong experimental support for the hypothesis that L1 EN generates a sequence specific endonucleolytic nick in the bottom strand of the target sequence to initiate L1 retrotransposition. Thirty-five of 36 characterized retrotransposition events integrated at a L1 EN consensus site, whereas the other (4HindIIIL1.3) likely integrated into an existing DNA lesion via an EN-independent pathway (Morrish et al., 2002). Interestingly, though there is a clear L1 EN consensus "bottom strand" cleavage site, our data demonstrates that there is little or no target site preference for top strand cleavage. Thus, L1 EN either is an unusual enzyme that only displays sequence specific cleavage for a single DNA strand, or a yet undiscovered enzyme performs second strand cleavage (see Figure 6).

An unexpected outcome of our study is that L1 retrotransposition can result in a variety of target site alterations. How these modifications are generated requires further study. However, we propose that each event is initiated by an L1 EN cleavage of the bottom strand followed by TPRT. The duplication, conservation, or deletion of target site nucleotides then results from the variable position of top strand cleavage (Figure 6A), which by analogy to the R2 non-LTR retrotransposon of *Bombyx mori*, will occur after the initiation of (−) strand cDNA synthesis by TPRT (Luan and Eickbush, 1995; Luan et al., 1993). We propose that the resultant intermediate can be recognized as a substrate by the

host repair machinery and subsequently undergoes 5′-3′ exonucleolytic processing. Plus (+) strand L1 cDNA synthesis by either L1 RT or host enzymes followed by recombinational repair would then lead to the final integration product. According to our model, top strand cleavages downstream of the initial endonucleolytic nick ultimately will lead to the generation of TSDs (Figure 6A; left). By contrast, top strand cleavages in direct opposition or upstream of the initial endonucleolytic nick will result either in conservative insertions or the generation of small deletions (Figure 6A; center and right, respectively).

The other observed rearrangements can be explained by simple modifications of the above model (Figure 6B). For example, chimeric L1s can be generated by single-strand annealing if the (−) strand L1 cDNA undergoes heteroduplex formation and DNA recombination with a homeologous L1 located upstream of the integration site. Processing of the top strand 3′ overhang and completion of integration by recombinational repair (as above) will result in a chimeric L1 and the concomitant deletion of intervening genomic DNA (Figure 6B; left). This mechanism can account for the mosaic pattern of coconversion observed in the chimeric L1s (Figure 5A; Szostak et al., 1983), the formation of hybrid L1s (Hayward et al., 1997), and the generation of new L1 subfamilies (Saxton and Martin, 1998). By comparison, the large deletion observed in 5BglIIL1.3 can be generated if (−) strand L1 cDNA invades a double-strand break upstream of the initial integration site (Figure 6B; middle). Likewise, large TSDs can be generated if the (+) strand L1 cDNA synthesis is initiated from the 3′ hydroxyl residue of a single-strand nick downstream of the initial L1 integration site (see Figure 6B; right).

Future studies will determine whether the formation of long TSDs, interstitial deletions, and the generation of chimeric L1s occur at high frequencies naturally. However, the cell culture retrotransposition system has accurately recapitulated many aspects of L1 biology, including the phenomena of L1-mediated transduction and processed pseudogene formation (Esnault et al., 2000; Moran et al., 1999; Wei et al., 2001). Moreover, the accompanying paper by Symer et al. (2002 [this issue of *Cell*]) provides evidence for similar L1-induced rearrangements in a HCT116 colon cancer cell line. Thus, we hypothesize that the characterization of additional disease-producing de novo L1 insertions ultimately will uncover similar structural rearrangements. Indeed, recent studies have shown that mutagenic T_F subfamily L1 insertions into the *disabled 1* and *titin* genes of mouse resulted in deletions of 2.7 kb and 779 bp, respectively (Garvey et al., 2002; Kojima et al., 2000).

The generation of deletions and chimeric elements upon retrotransposition may not be unique to L1s. For example, the retrotransposition of a "young" Alu Y element into the CMP-N-acetylneuraminic acid hydroxylase gene resulted in the replacement of an "older" AluSq element and the deletion of an internal exon of that gene (Hayakawa et al., 2001). Similarly, retrotransposition of a young Alu Yb9 element into the APC gene resulted in the creation of a chimeric element, whose 5′ end consists of an older Alu Sq element (Su et al., 2000). There also are numerous examples of gene conversion in which young Alu elements replaced older Alu elements (Carroll et al., 2001; Kass et al., 1995), and we speculate that many of those events occurred by Alu cDNA-mediated recombination. Finally, de novo poly (A) tails have been found at the junctions of mutagenic deletions, which resulted in either breast cancer (a 6.2 kb deletion) or Alport syndrome (a 13.4 kb deletion; Segal et al., 1999; Wang et al., 2001). In both instances, the poly (A) tail integrated at an L1 EN consensus cleavage site. Thus, it is possible that L1 or Alu retrotransposition led to the formation of these deletions.

In closing, we have uncovered mechanisms by which L1 retrotransposition continues to alter the human genome, and have demonstrated that L1 is not simply an insertional mutagen. Indeed, our data highlight pathways by which L1 retrotransposition can act to reduce genome size and provide additional mutational mechanisms by which retrotransposition can affect tumor and perhaps germ cells.

Experimental Procedures

Oligonucleotides and Recombinant Plasmids

All oligonucleotide sequences are available at upon request or can be accessed at www.med.umich.edu/hg/RESEARCH/FACULTY/Moran/moranweb.htm.

pNGneoSD contains the *mneoI* cassette with a Shine Delgarno (SD) sequence. The SD sequence was introduced upstream of the *mneoI* initiator codon by PCR. A 1.8 kb product was amplified from p*mneoI* using the BamNeo and MluSDal primers. The product, which contained the SD sequence, was digested with MluI and BamHI and was used to replace the analogous fragment from p*mneoI*.

pNGneoEM7 is a modified version of p*NGneoSD* that contains the bacterial EM7 promoter. The 79 bp EM7 promoter from pHook3 (Invitrogen) was amplified using the EM75 and EM73 primers. The product was digested with MluI and was introduced into the MluI site of p*NGneoSD*.

pK7 is a plasmid containing the modified *mneoI* cassette from p*NGneoEM7*. The 1033 bp Amp gene from pBluescript (Stratagene) was amplified using the AscAmp5 and BamNotAmp3 primers. The 894 bp ColE1 origin from pBK-RSV (Stratagene) was amplified using the R1ColE5 and PstColE3 primers. The 249 bp SV40 polyadenylation signal from pBK-RSV (Stratagene) was amplified using the PstSV405 and AscSV403 primers. p*K7* was created by ligating the AscI-BamHI Amp fragment, the EcoRI and PstI ColE1 fragment, the PstI and AscI SV40pA fragment, and the EcoR1 and BamHI fragment from p*NGneoEM7*.

$pK7_{400}$ was generated as above, but contains a silent mutation that destroyed the crypic splice site present in the *mneoI* gene. Site-directed mutagenesis was performed using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene) using the SDM400S and SDM400AS primers.

pK7/L1.3mneoI/ColE1 contains the *mneoI/ColE1* cassette at the end of the L1.3 3′UTR. p*JM101/L1.3* no neo was digested with NotI and BamHI and the resultant fragment was cloned between the NotI and BamHI sites of p*K7*.

pK7/L1.3mneoI$_{400}$*/ColE1* contains the *mneoI*$_{400}$*/ColE1* cassette at the end of the L1.3 3′UTR. It was cloned as above.

pCEP4/L1.3mneoI/ColE1 contains the *mneoI/ColE1* cassette within the L1.3 3′UTR. p*K7* was digested with BamH1 and AscI. The ends were made blunt with 6 units of T4 DNA polymerase (NEB) and 100μM dNTP in 1× T4 DNA polymerase buffer (NEB). The resultant product was subcloned into the SmaI site present in the 3′UTR of p*JM101/L1.3* no neo.

pCEP4/L1.3 mneoI$_{400}$*/ColE1* contains the *mneoI*$_{400}$*/ColE1* cassette within the L1.3 3′UTR. It was cloned as above using p$K7_{400}$

pCEP4/L1.2 mneoI$_{400}$*/ColE1* contains the *mneoI*$_{400}$*/ColE1* cassette within the L1.2 3′UTR. Swapping the *NotI-BstZ17I* fragment from p*JM101/L1.2* with the corresponding fragment from p*CEP4/L1.3 mneoI*$_{400}$*/ColE1* created it.

DNA Preparation

Plasmid DNAs were purified on Qiagen midi prep columns (Qiagen). Rescued plasmids were purified using Wizard S/V miniprep kits (Promega). HeLa genomic DNA was isolated using either the Blood and Cell Midi prep kit (Qiagen) or the Cell and Tissue DNA isolation kit (Puregene, Gentra).

L1 Retrotransposition Assay

HeLa cells were grown at 37°C in an atmosphere containing 7% carbon dioxide and 100% humidity in Dulbecco's modified Eagle medium (DMEM) lacking pyruvate (Gibco BRL). DMEM was supplemented with 10% fetal bovine calf serum and 1× Penicillin-Streptomycin-L-glutamine (a 100 × stock is sold by Gibco BRL). Cell passage and cloning (by either limiting dilution or colony lifting) was performed using standard techniques. Retrotransposition was monitored using the transient retrotransposition assay (Wei et al., 2000).

RT-PCR

Total RNA was extracted using TRIZOL (Invitrogen) three days posttransfection. 50 μg of RNA in 25 μl of 1× buffer (20 mM Tris-HCl, 2 mM $MgCl_2$, and 50 mM KCl) were treated with 2 units of DNase 1 (Gibco/BRL). After incubation for 1 hr at 37°C, the enzyme was inactivated by a 20 min incubation at 65°C. 10 μg of treated RNA in 20 μl of 1× buffer (50 mM Tris-HCl, 8 mM $MgCl_2$, 37.5 mM KCl, 1 mM dithiothreitol, 1 mM of each dNTP, and 50 pmol oligo dT_{15}) was reverse transcribed using 25 U of AMV RT (Roche) for 1 hr at 42°C. 2 μl of the RT product was used as a template in PCR reactions with 437NEOS and 1785NEOAS using the following cycling conditions (one cycle 94°C for 3 min; 30 cycles: 94°C for 10 s, 60°C for 30 s, 72°C for 1 min; one cycle 72°C for 10 min). The products were resolved on 1.0% agarose-ethidium bromide gels.

Rescue Procedure

10 μg of DNA was digested overnight with HindIII, BglII, BclI or BamHI (NEB). The enzyme was either heat inactivated or removed using the Wizard DNA Clean-up Kit (Promega). The resultant fragments were ligated overnight at 14°C in a 500 μl volume with T4 DNA ligase (2,400 U; NEB). Ligations were concentrated by centrifugation through a Microcon-100 at 500 × g for 14 min, and the entire concentrated ligation was transformed into 1 ml of XL1-Blue MRF cells (Stratagene), which were made to competencies of $>1 \times 10^8$ transformants/μg (Inoue et al., 1990; Tang et al., 1994). One to 15 transformants were visible after overnight growth at 37°C on LB agar supplemented with 50 μg/ml kanamycin.

Sequence Analysis

Sequences flanking each L1 insertion were used as probes in BLAT searches to identify the preintegration site in the HGWD using either the 8/06/01 or the 12/22/01 freezes (http://genome.cse.ucsc.edu). Some rescued plasmids contained L1s that were truncated at a restriction site used in their recovery. To obtain the remaining L1 sequence, we used the HGWD to design an oligonucleotide primer from sequences presumed to flank the 5′ end of the retrotransposed L1. The 5′ primer then was used in conjunction with 173NEOAS to PCR amplify the 5′ flanking sequence from genomic DNA of cell lines harboring the relevant insertions. BigDye Terminator Cycle Sequencing (ABI PRISM) was performed on an Applied Biosystems DNA sequencer (Model ABI 377 or ABI3700) at the University of Michigan Core facilities.

PCR Amplification of the Integration Sites of 5BglII

Amplification of pre- and postintegration sequences was conducted following the protocol provided with the Expand Long Template PCR System (Roche). Each amplification was carried out in a 50 μl volume containing, 500 ng of genomic DNA, 300 nM of each primer, 350 μM of each dNTP, and 1 μl of enzyme mix. PCR was conducted using the following cycling conditions: (one cycle 94°C for 2 min; 30 cycles: 94°C for 10 s, 60°C for 30 s, and 68°C for 150 s; one cycle 68°C for 7 min). To amplify the 6 and 12 kb products, the extension time was extended to 10 min. The sequence of the L1 was verified by sequencing the PCR product. Amplification was conducted on genomic DNA isolated from a polyclonal cell line containing 5BglIIL1.3 and approximately 30 other clones.

PCR Amplification of the Integration Sites of the Chimeric L1s

The preintegration site of 3HindIIIL1.3 was amplified using 3HindIII5′ flank and 3HindIII3′ flank as primers. The postintegration sites of 3HindIIIL1.3, 4HindIIIL1.3 and 1HindIIIL1.2 were amplified using a respective 5′ flank primer and 173NEOAS as primers. Amplification conditions were the same as for 5BglIIL1.3 (see above) and were performed on genomic DNA from clonal cell lines containing the respective insertions. The sequence of the respective chimeric L1 was verified by sequencing the PCR product.

Southern Blot Analysis

Southern blotting was performed using the protocol provided in the Nytran membrane book (Schleicher and Schuell). 10 μg of genomic DNA was digested with HindIII and was resolved on a 0.8% agarose/TAE gel. After denaturing and neutralizing, the DNA was transferred to a Nytran SuPer Charge membrane (Schleicher and Schuell) using 20 × SSC buffer at pH 7.0 (SSC is 150 mM NaCl and 15 mM sodium citrate). The membranes were prehybridized for 3 hr at 42°C in Hyb buffer (5 × SSC, 2% Blocking Reagent [Roche], 50% formamide, 7% SDS, 0.1% N-laurosarcosine and 50 mM sodium phosphate buffer, [pH 7.0]). Probes (as indicated in Figure 5D) were obtained by digesting the rescued 3HindIII plasmid with HindIII and HpaI (5′ probe) or HindIII and XbaI (3′ probe). The digested DNA fragments were resolved on 0.7% agarose/ethidium bromide gels, were purified using a gel extraction kit (Qiagen), and were random prime labeled with [α-^{32}P]dCTP using the Redi-prime II labeling system (Amersham). Unincorporated counts were removed by centrifugation through a Microspin G50. Three million counts/ml of hybridization buffer were used in hybridizations, which were carried out overnight at 42°C in 1× Hyb buffer. Blots were washed twice in 2 × SSC, 0.5% SDS at 25°C for 30 min; once in 0.1 × SSC, 0.5% SDS at 37°C for 45 min; and once in 0.1 × SSC, 0.5% SDS at 68°C for 90 min. The washed filter was then exposed on X-OMAT film for 18 hr at −80°C.

Acknowledgments

Correspondence can be addressed to J.V.M. or N.G. We thank members of the University of Michigan DNA Sequencing Core for help with sequencing; Dr. Thomas Glover, Tammy Morrish, and other members of the Moran lab for critically evaluating the manuscript; and Aurélien Doucet and Tara Biagi for excellent technical assistance. This work was supported in part by grants to J.V.M. from the W.M. Keck Foundation, the National Institutes of Health (GM60518), and the March of Dimes. The University of Michigan Cancer Center helped defray some of the DNA sequencing costs.

Received: April 19, 2002
Revised: June 10, 2002

References

Boeke, J.D. (1997). LINEs and Alus-the polyA connection. Nat. Genet. *16*, 6–7.

Carroll, M.L., Roy-Engel, A.M., Nguyen, S.V., Salem, A.H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nguyen, L., Sammarco, M., et al. (2001). Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J. Mol. Biol. *311*, 17–40.

Chaboissier, M.C., Finnegan, D., and Bucheton, A. (2000). Retrotransposition of the I factor, a non-long terminal repeat retrotransposon of *Drosophila*, generates tandem repeats at the 3′ end. Nucleic Acids Res. *28*, 2467–2472.

Cost, G.J., and Boeke, J.D. (1998). Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry *37*, 18081–18093.

Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., and Kazazian, H.H., Jr. (1991). Isolation of an active human transposable element. Science *254*, 1805–1808.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retro-

transposons generate processed pseudogenes. Nat. Genet. *24*, 363–367.

Feng, Q., Moran, J.V., Kazazian, H.H., Jr., and Boeke, J.D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell *87*, 905–916.

Garvey, S.M., Rajan, C., Lerner, A.P., Frankel, W.N., and Cox, G.A. (2002). The muscular dystrophy with myositis (mdm) mouse mutation disrupts a skeletal muscle-specific domain of titin. Genomics *79*, 146–149.

Hayakawa, T., Satta, Y., Gagneux, P., Varki, A., and Takahata, N. (2001). Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. Proc. Natl. Acad. Sci. USA *98*, 11399–11404.

Hayward, B.E., Zavanelli, M., and Furano, A.V. (1997). Recombination creates novel L1 (LINE-1) elements in *Rattus norvegicus*. Genetics *146*, 641–654.

Hohjoh, H., and Singer, M.F. (1996). Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J. *15*, 630–639.

Hohjoh, H., and Singer, M.F. (1997). Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. EMBO J. *16*, 6034–6043.

Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene *96*, 23–28.

Kass, D.H., Batzer, M.A., and Deininger, P.L. (1995). Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. Mol. Cell. Biol. *15*, 19–25.

Kojima, T., Nakajima, K., and Mikoshiba, K. (2000). The disabled 1 gene is disrupted by a replacement with L1 fragment in yotari mice. Brain Res. Mol. Brain Res. *75*, 121–127.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Liao, D., Pavelitz, T., and Weiner, A.M. (1998). Characterization of a novel class of interspersed LTR elements in primate genomes: structure, genomic distribution, and evolution. J. Mol. Evol. *46*, 649–660.

Luan, D.D., and Eickbush, T.H. (1995). RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol. Cell. Biol. *15*, 3882–3891.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). 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.

Martin, S.L. (1991). Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol. Cell. Biol. *11*, 4804–4807.

Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. Science *254*, 1808–1810.

Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. Science *283*, 1530–1534.

Moran, J.V., and Gilbert, N. (2002). Mammalian LINE-1 retrotransposons and related elements. In Mobile DNA II, N. Craig, R. Craggie, M. Gellert and A. Lambowitz, eds. (Washington, DC: ASM Press), pp. 836–69.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell *87*, 917–927.

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T., Taccioli, G., Batzer, M.A., and Moran, J.V. (2002). Endonuclease-independent L1 retrotransposition: a mechanism of RT-mediated repair in mammalian cells. Nat. Genet. *31*, 159–165.

Ostertag, E.M., and Kazazian, H.H., Jr. (2001). Twin priming: a proposed mechanism for the creation of inversions in l1 retrotransposition. Genome Res. *11*, 2059–2065.

Ovchinnikov, I., Troxel, A.B., and Swergold, G.D. (2001). Genomic characterization of recent human LINE-1 insertions: evidence supporting random insertion. Genome Res. *11*, 2050–2058.

Sassaman, D.M., Dombroski, B.A., Moran, J.V., Kimberland, M.L., Naas, T.P., DeBerardinis, R.J., Gabriel, A., Swergold, G.D., and Kazazian, H.H., Jr. (1997). Many human L1 elements are capable of retrotransposition. Nat. Genet. *16*, 37–43.

Saxton, J.A., and Martin, S.L. (1998). Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. J. Mol. Biol. *280*, 611–622.

Segal, Y., Peissel, B., Renieri, A., de Marchi, M., Ballabio, A., Pei, Y., and Zhou, J. (1999). LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis. Am. J. Hum. Genet. *64*, 62–69.

Sheen, F.M., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M.A., and Swergold, G.D. (2000). Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. Genome Res. *10*, 1496–1508.

Su, L.K., Steinbach, G., Sawyer, J.C., Hindi, M., Ward, P.A., and Lynch, P.M. (2000). Genomic rearrangements of the APC tumor-suppressor gene in familial adenomatous polyposis. Hum. Genet. *106*, 101–107.

Swergold, G.D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol. Cell. Biol. *10*, 6718–6729.

Symer, D.E., Connelly, C., Szak, S.T., Caputo, E.M., Cost, G.J., Parmigiani, G., and Boeke, J.D. (2002). Human L1 retrotransposition is associated with genetic instability in vivo. Cell *110*, this issue, 327–337.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J., and Stahl, F.W. (1983). The double-strand-break repair model for recombination. Cell *33*, 25–35.

Tang, X., Nakata, Y., Li, H.O., Zhang, M., Gao, H., Fujita, A., Sakatsume, O., Ohta, T., and Yokoyama, K. (1994). The optimization of preparations of competent cells for transformation of *E. coli*. Nucleic Acids Res. *22*, 2857–2858.

Wang, T., Lerer, I., Gueta, Z., Sagi, M., Kadouri, L., Peretz, T., and Abeliovich, D. (2001). A deletion/insertion mutation in the BRCA2 gene in a breast cancer family: a possible role of the Alu-polyA tail in the evolution of the deletion. Genes Chromosomes Cancer *31*, 91–95.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Human L1 retrotransposition: *cis* preference versus *trans* complementation. Mol. Cell. Biol. *21*, 1429–1439.

Wei, W., Morrish, T.A., Alisch, R.S., and Moran, J.V. (2000). A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. Anal. Biochem. *284*, 435–438.