Cell, Vol. 110, 277–280, August 9, 2002, Copyright ©2002 by Cell Press

LINE Drive: Retrotransposition and Genome Instability

Minireview

Haig H. Kazazian, Jr.,[1] and John L. Goodier
Department of Genetics
School of Medicine
475 Clinical Research Building
415 Curie Boulevard
University of Pennsylvania
Philadelphia, Pennsylvania 19105

The LINE-1 (L1) retrotransposon, the most important human mobile element, shapes the genome in many ways. Now two groups provide evidence that L1 retrotransposition is associated with large genomic deletions and inversions in transformed cells. If these events occur at a similar frequency in vivo, they have had a substantial effect on human genome evolution.

The human genome has become an exciting place to visit. Over the past 15 years, our view of it has changed considerably, offering a number of surprises and increasing interest in human genetics. In general, the genome has great stability; mutation occurs at a rate of 10^{-9} nucleotides per year. However, there are several mechanisms that create genome instability. For example, expanding trinucleotide and other short repeats account for 15–20 diseases (Cummings and Zoghbi, 2000). We have also learned about long-range rearrangements that produce deletions, duplications, and inversions of up to a few million base pairs. These rearrangements have caused not only common genomic polymorphisms, but also 10–15 contiguous gene syndromes in humans (Emanuel and Shaikh, 2001). In addition, we have discovered that the human genome is even more populated with transposable elements than other species whose genomes have been analyzed, and that these "jumping genes" can cause disease and alter the genome in a number of different ways (reviewed in Ostertag and Kazazian, 2001).

There are two main classes of transposable elements: DNA transposons and retrotransposons. Retrotransposons, the most important transposable elements in the human genome, are copied into RNA, the RNA is reverse-transcribed into DNA, and the DNA is inserted into the genome at a new location. Thus, these elements expand in number by a duplication (copy-and-paste) mechanism. The human genome is littered with remnants of the most prominent non-long-terminal repeat (non-LTR) retrotransposon, the L1 element, roughly a half million of which are 5′ truncated, inverted, or mutated to inactivity. It also contains roughly 5000 full-length 6 kb elements, 60–100 of which are still capable of retrotransposition. The full-length elements contain a 5′ untranslated region (5′ UTR) with an internal promoter, a 1 kb ORF1 that encodes a protein with RNA binding capability, a 4 kb ORF2 that encodes a protein with endonuclease and reverse transcriptase activities, a short 3′ UTR, and a poly(A) tail (Figure 1). Insertion into chromosomal DNA probably occurs by a process termed target-primed reverse transcription (TPRT); the endonuclease of ORF2 nicks a single strand of DNA, leaving a 3′-OH that serves as the primer for reverse transcription with the L1 RNA as template (Luan et al., 1993).

We already know that these elements can alter the genome in many ways beyond simple insertion of the element itself. (1) Often 3′ and occasionally 5′ flanking sequences are transduced along with the L1 during retrotransposition. The phenomenon of 3′ transduction is a potential mechanism for shuffling exons from one genomic site to another (Moran et al., 1999). Because 5′ truncation frequently limits the retrotransposed sequence to less than 1 kb, a transduced sequence may include only 3′ flanking DNA and completely lack L1 sequences. In addition, because L1 inversion occurs frequently during retrotransposition, 3′ transductions may contain inverted genomic sequence. Thus, transduction events have considerable potential for scrambling retrotransposed sequences and altering the genome (Figure 1). (2) Although L1 proteins have the curious property of preferentially acting to retrotranspose the element that encoded them (*cis* preference), they occasionally act in *trans* to retrotranspose non-L1 RNAs (Esnault et al., 2000; Wei et al., 2001). (3) Mispairing of homologous L1 sequences and unequal crossing-over can produce deletions and duplications of genomic sequences. A few cases of human disease have been caused by deletions of this type. (4) L1s contain within their 5′ UTR not only a sense strand promoter for their own transcription, but also an antisense promoter. This antisense Pol II promoter has been shown to provide an alternative transcription start site for a number of human Pol II-transcribed genes (Nigumann et al., 2002).

In sum, L1s account directly or indirectly for about one-third of the human genome. In fact, non-LTR retrotransposable elements have been extant since the Precambrian Era 500–600 million years ago. Since sequences greater than 200 million years old have mutated beyond recognition, it is likely that retrotransposons actually account for greater than half and perhaps the bulk of the human genome.

Now, two papers in this issue of *Cell* by Gilbert et al. (2002) and Symer et al. (2002) demonstrate that genome instability, principally in the form of substantial deletions, is associated with about 10% of L1 insertions in transformed human cells. New L1 retrotranspositions have been estimated to occur in vivo at a frequency of one new event in every 10 to 250 individuals born (reviewed in Ostertag and Kazazian, 2001). If sizable deletions accompany one in every 10 L1 retrotranspositions in vivo, the effect on genome evolution could be substantial. The finding is also remarkable because retrotransposition of both non-LTR elements of the L1 type and LTR elements of the yeast Ty variety have been rarely associated with significant deletions. In contrast, DNA transposons that generally move by a cut-and-paste process without using an RNA intermediate (e.g.,

[1]Correspondence: kazazian@mail.med.upenn.edu

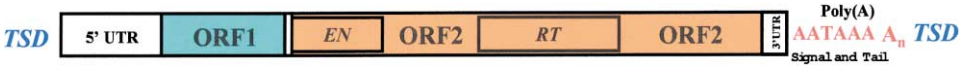

Figure 1. Structure of a Full-Length 6 kb Human L1 Element

Abbreviations: TSD, variable-length target site duplication; 5′ UTR, 5′ untranslated region; ORF1, first open reading frame; ORF2, second open reading frame; EN, endonuclease domain; RT, reverse transcriptase domain; 3′ UTR, 3′ untranslated region; AATAAA, hexanucleotide poly(A) signal; and A_n, the poly(A) tract abuts the hexanucleotide signal in humans.

P elements) are known to produce significant deletions upon leaving a site, and some bacterial transposons can produce genomic rearrangements upon insertion into a new location.

Our present knowledge of the characteristics of L1 retrotranspositions comes from three distinct sources. One source is insertions into the human genome characterized from database searches of the draft human genome sequence. The most valuable of these insertions are those of the youngest and most active subfamily of L1s (the Ta subfamily), which are on average only 2 million years old (Boissinot and Furano, 2001). About half of these elements are polymorphic as to presence or absence so that the genomic site that lacks the insertion can be characterized. A second source is the disease-causing insertions into the human and mouse genomes that are very recent (<100 years old). To date, 14 human and 8 mouse de novo insertions are known (see Ostertag and Kazazian, 2001). The third source is the new insertions obtained in transformed cells in culture. Both of the present papers examine cell culture insertions, while Symer et al. (2002) also compare their tissue culture data to data obtained from a database search.

The two papers in this volume employ a time-tested, plasmid-based strategy for studying individual retrotransposition events in cultured cells, an approach first used by Curcio and Garfinkel (1991) to examine Ty1 insertions in yeast. This assay has been particularly valuable in studying the biology of L1 retrotransposons (Moran et al., 1996). A retrotransposon is tagged with a reporter cassette consisting of an antisense copy of an auxotrophic or antibiotic resistance gene, interrupted by an intron placed in sense orientation. Transcription from the retrotransposon's promoter, splicing of the intron, reverse transcription, and insertion of the cDNA copy of the element into chromatin allows expression of the reporter gene. Transcipts generated from the promoter of the marker gene prior to retrotransposition fail to be expressed because the disrupting intron cannot be spliced due to its orientation. Such strategies capture retrotransposition events red-handed, before inserted sequences are blurred by time and the genomic forces of mutation, recombination, and gene conversion.

The authors of the present papers use the *mneoI* reporter cassette that confers G418 resistance upon cells acquiring an L1 insertion. Taking a cue from "self-cloning" or "plasposon" strategies that have facilitated recovery of bacterial transposon integrants (reviewed in Dennis and Zylstra, 1998), the authors cleverly reengineered *mneoI* to contain a bacterial origin of replication, Shine-Dalgarno sequence, and a bacterial promoter upstream of either the neoR gene (Gilbert et al., 2002) or a separate dihydrofolate reductase gene (Symer et al., 2002). DNA from G418-resistant cultured cells was isolated and cleaved with a restriction enzyme, and the fragments were self-ligated and transformed into bacterial cells. Thus, L1 integrants together with their flanking genomic DNA were speedily recovered from human cells as self-replicating bacterial clones selected on either kanamycin or trimethoprim plates. In this manner, the authors have significantly increased the number of retrotransposition events that can be characterized using this assay (Moran et al., 1996; Esnault et al., 2000; Wei et al., 2001).

While retrotransposition in cell culture mimics that in vivo in many respects, there are substantial differences between the two processes. Retrotransposition in the organism likely occurs in both male and female germ cells and may also occur very early in embryonic development. There is strong evidence that L1 retrotransposition is suppressed in somatic cells by mechanisms not clearly understood, although promoter repression, methylation, and RNA interference are among the usual suspects. Although one instance of L1 retrotransposition has been documented in a colon cancer (Miki et al., 1992), it is unclear whether that event occurred in a transformed cell or a primary epithelial cell. Also, retrotransposition in vivo occurs from one chromosome to another. In the tissue culture assay, retrotransposition events occur from an episome to a chromosomal site in cells that are transformed and therefore to some extent aneuploid and genetically unstable. That said, the new papers find that L1 insertions in tissue culture are quite similar to both recent in vivo insertions and database insertions in many respects, including occurrence of duplications and small deletions at the target site, sequence of the endonuclease cleavage site, occurrence of inversions and 3′ transductions, and the presence of a poly(A) tail.

As the present papers point out, however, there are some unusual features of the insertions in tissue culture. Although only about 1% of the L1s of all ages in the genome are full-length, some 30%–35% of the young Ta subfamily of elements is full-length (Boissinot and Furano, 2001). Among recent insertions, 2 of 14 in humans and 3 of 8 in mice are full-length, or 5/22 (23%) in all. Thus, one might expect that a substantial fraction of insertions would be full-length when a Ta element is used in the cultured cell assay. Combining the data from the two papers yields 8 insertions of 6 kb or longer among 75 insertions, or 10%. Likewise in the genome, nearly all target site duplications (TSDs) are 20 bp or shorter, whereas 10 of 28 TSDs in Gilbert et al. (2002) and 3 of 33 TSDs in Symer et al. (2002) are longer than 30 bp. In addition, Symer and colleagues found that 50% of their L1 insertions interrupted predicted genes, a bias not detected by other studies in HeLa cells (Moran

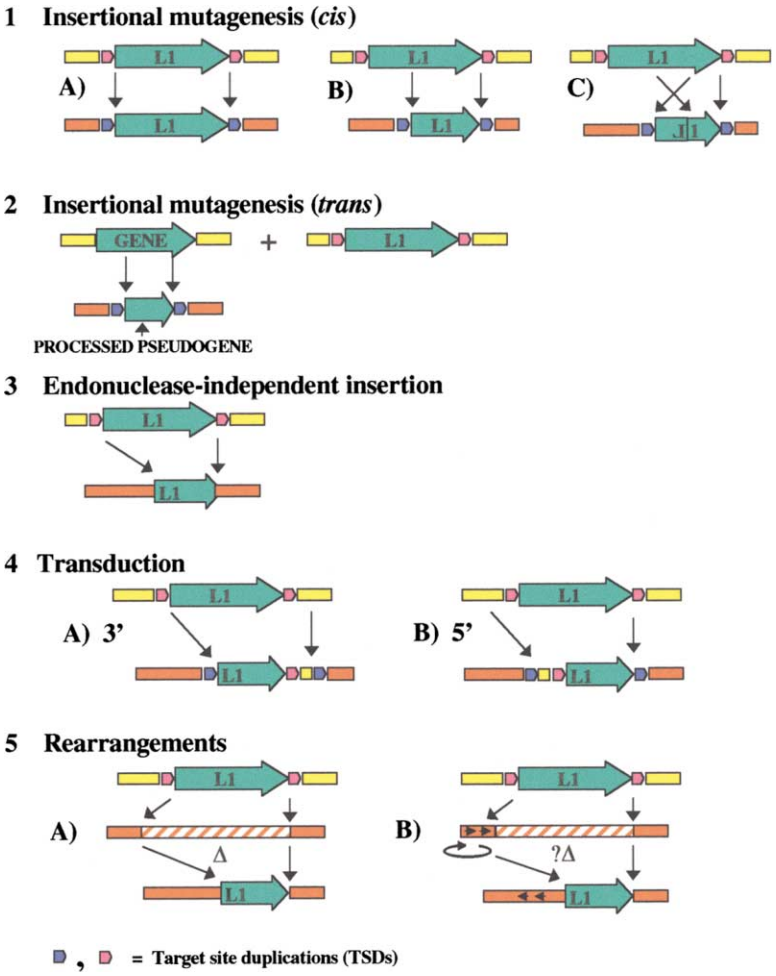

Figure 2. L1 Elements Alter Mammalian Genomes in Many Ways upon Retrotransposition

(1) L1 proteins tend to act in *cis* on their own encoding RNA to generate new retrotransposition events. Newly inserted copies may be full-length (A) and potentially active. However, more frequently they have 5′ truncations (B) or 5′ truncations and inversions (C) and are incapable of further retrotransposition.

(2) Occasionally L1 proteins act in *trans* to mobilize non-L1 RNAs. L1s have likely been responsible for the genome-wide dispersal of processed pseudogenes (roughly 1% of the genome) and of Alu nonautonomous retrotransposons (numbering 1.1 million copies or 10% of the human genome). A processed pseudogene forms when a spliced mRNA is reverse transcribed by L1 reverse transcriptase and the resulting DNA is inserted into the genome.

(3) Rarely, L1s are able to integrate into preformed double-strand breaks generated independently of the L1 endonuclease. The integrant structures are atypical, being truncated at their 3′ ends, and they lack TSDs.

(4) L1s are able to mobilize both 3′ and, less commonly, 5′ flanking DNA upon retrotransposition and carry these flanks to new genomic locations. (A) In the case of 3′ transduction, the weak L1 polyadenylation signal is occasionally bypassed by the RNA 3′ processing machinery in favor of a second downstream polyadenylation signal. (B) Sequence 5′ of an L1 may be transduced if (1) the L1 promoter initiates transcription at an abnormal upstream site, or (2) if the chance presence of a non-L1 upstream promoter generates an elongated L1 transcript. Thus, L1s may play a role in "exon shuffling." An L1 residing in the intron of a gene might transduce upstream or downstream exonic sequence to a new site, possibly into another gene.

(5) Two papers in this issue of *Cell* report large-scale deletions (A) and inversions (B) at the site of L1 insertions generated in a cell culture assay. The inserted L1s associated with these chromosomal rearrangements lack TSDs but may have a poly(A) tail. Note that the second junction of the inversion in (B) was not detected by Symer et al. (2002), so that whether the inversion involved a deletion or other rearrangement is unknown. Other genomic effects of L1s, such as recombination between mispaired L1s producing deletions or duplications and alternative transcription of some Pol II-dependent genes from a promoter within L1, are not shown here.

et al., 1999; Gilbert et al., 2002) or in the draft human genome sequence.

The most unusual feature of the insertions in tissue culture is the 7 large deletions and 1 potential chromosomal inversion observed among 79 insertions, an incidence of genomic rearrangements of 10%. One simple explanation for these rearrangements, which cannot be ruled out by the cell culture data, is that RNA has provided a template used to patch double-stranded and degraded DNA breaks within cells hosting the L1 insertions. Insertion of both L1 and Ty1 elements at breaks generated independently of the element-encoded endonucleases has been demonstrated in cell culture (Teng et al., 1996; Morrish et al., 2002). However, unlike the L1 endonuclease-independent events characterized to date (Morrish et al., 2002), four of the seven L1s reported in the present studies both possess a poly(A) tail and have inserted at a consensus L1 target site, suggesting that their associated rearrangements were mediated by retrotransposition.

To what extent can such rearrangements be observed in the organism? One test for deletions upon retrotransposition would be to find empty sites (sites into which an L1 insertion has occurred) that contain more genomic DNA than do those sites after L1 insertion. In other words, if one studies polymorphic young L1 elements in the human genome, are the empty sites larger than expected? Myers et al. (2002) were able to characterize the empty site for 115 polymorphic L1s from the Ta subfamily and found no evidence for deletions among them. However, the empty sites of five other polymorphic elements "produced nonspecific PCR results" and were excluded from the analysis. The failure to amplify these empty sites could be due to deletions accompanying retrotransposition, repetitive DNA at the empty sites, or assembly errors in the draft human genome sequence. Our laboratory has analyzed 25 polymorphic full-length Ta elements that do not overlap with those of Myers et al. (2002) and found empty sites of the expected size for all of them. Thus, these data would suggest that the frequency of DNA deletions created upon L1 retrotransposition is at most 5/145 or 3%.

However, some genomic deletions between 3 and 70 kb observed by Gilbert et al. (2002) and Symer et al. (2002) might be deleterious to the organism and eliminated from the genome. Even the youngest of the polymorphic Ta subfamily L1s in the draft human genome sequence must be thousands of years old. Thus, it is potentially more instructive to analyze insertions that have occurred into mammalian genomes in the past 100 years. No deletions have been found among the 14 known recent L1 insertions in humans. Among eight known recent L1 insertions in the mouse, six are straightforward endonuclease-dependent events, one appears to be an endonuclease-independent event associated with a deletion in the *disabled-1* gene (Kojima et al., 2000), and one may well be an endonuclease-dependent L1 retrotransposition associated with a deletion (Garvey et al., 2002). In this latter event, a 5′ truncated L1 derived from a subfamily of active mouse L1s is associated with a 779 bp deletion of the *titin* gene in the *mdm* mouse. Thus, the recent insertion data suggest a 5% incidence of endonuclease-mediated deletions associated with recent L1 insertions (1/22) and, together with the data from polymorphic Ta L1s, are in line with the cell culture data of the present papers (4/79 L1 insertions associated with endonuclease-mediated rearrangement events, or 5%).

In order to determine the frequency of in vivo genomic rearrangements occurring during retrotransposition, we need a mouse model of retrotransposition. We now know that such a mouse model is feasible, and that retrotransposition events produced by a human L1 transgene are stably inherited. After characterizing a large number of retrotransposition events that have occurred in the germline of transgenic mice, we should learn the degree to which L1s sculpt genomic DNA. In the meantime, the Moran and Boeke labs provide us with more tantalizing evidence for L1s as genome artisans.

Selected Reading

Boissinot, S., and Furano, A.V. (2001). Mol. Biol. Evol. *18*, 2186–2194.

Cummings, C.J., and Zoghbi, H.Y. (2000). Hum. Mol. Genet. *9*, 909–916.

Curcio, M.J., and Garfinkel, D.J. (1991). Proc. Natl. Acad. Sci. USA *88*, 936–940.

Dennis, J.J., and Zylstra, G.J. (1998). Appl. Environ. Microbiol. *64*, 2710–2715.

Emanuel, B.S., and Shaikh, T.H. (2001). Nat. Rev. Genet. *2*, 791–800.

Esnault, C., Maestre, J., and Heidmann, T. (2000). Nat. Genet. *24*, 363–367.

Garvey, S.M., Rajan, C., Lerner, A.P., Frankel, W.N., and Cox, G.A. (2002). Genomics *79*, 146–149.

Gilbert, N., Lutz-Prigge, S., and Moran, J.V. (2002). Cell *110*, this issue, 315–325.

Kojima, T., Nakajima, K., and Mikoshiba, K. (2000). Brain Res. Mol. Brain Res. *75*, 121–127.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Cell *72*, 595–605.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., and Nakamura, Y. (1992). Cancer Res. *52*, 643–645.

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

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

Morrish, T.A., Gilbert, N., Myers, J.S., Vincent, B.J., Stamato, T.D., Taccioli, G.E., Batzer, M.A., and Moran, J.V. (2002). Nat. Genet. *31*, 159–165.

Myers, J.S., Vincent, B.J., Udall, H., Watkins, W.S., Morrish, T.A., Kilroy, G.E., Swergold, G.D., Henke, J., Henke, L., Moran, J.V., et al. (2002). Am. J. Hum. Genet. *71*, 312–326.

Nigumann, P., Redik, K., Matlik, K., and Speek, M. (2002). Genomics *79*, 628–634.

Ostertag, E.M., and Kazazian, H.H., Jr. (2001). Annu. Rev. Genet. *35*, 501–538.

Symer, D.E., Connelly, C., Szak, S.T., Caputo, E.M., Cost, G.J., Parmigiani, G., and Boeke, J.D. (2002). Cell *110*, this issue, 327–338.

Teng, S.C., Kim, B., and Gabriel, A. (1996). Nature *383*, 641–644.

Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Mol. Cell. Biol. *21*, 1429–1439.