© 1998 Nature Publishing Group http://www.nature.com/naturegenetics

The impact of L1 retrotransposons on the human genome

Haig H. Kazazian, Jr & John V. Moran

The 'master' human mobile element, the L1 retrotransposon, has come of age as a biological entity. Knowledge of how it retrotransposes *in vivo*, how its proteins act to retrotranspose other poly A elements and the extent of its role in shaping the human genome should emerge rapidly over the next few years. We review the impact of retrotransposons and how new insight is likely to lead to important practical applications for these intriguing mobile elements.

Evolutionary biologists hypothesize that the earliest life forms contained self-replicating RNA genomes. The advent of polymerases that make DNA copies of RNA templates allowed the conversion of information from unstable ribose-based polymers to more stable deoxyribose-based polymers through the process of reverse transcription. In this way, reverse transcription appears to have played a pivotal role in the formation of the first DNA genomes.

Although reverse transcription has been ongoing during genome evolution, its impact is only just being realized. It is now apparent that reiterative rounds of reverse transcription served to expand both the size and complexity of the human genome. The chief perpetrators in this process seem to have been a small number of autonomously mobile DNA sequences known as long interspersed nuclear elements (LINEs or L1s). At least one-quarter of the human genome consists of sequences which either are derived directly from retrotransposition-competent L1s or likely resulted from the promiscuous action of L1-encoded reverse transcriptase on other transcripts, including Alu elements and cellular mRNAs (Table 1; refs 1,2). Although other autonomous mobile sequences may exist in the human genome, their contribution to its total mass is far less dramatic than that of L1 elements.

Weeds in the garden—human mobile elements

As a result of the genome project, we now know that only approximately 3% of the human genome is comprised of exonic sequences. The remainder, so-called 'junk DNA', is composed largely of introns, simple repeated sequences and mobile elements or their remnants. Mobile elements fall into three major classes: DNA-based transposable elements, autonomous retrotransposons and non-autonomous retrotransposons (Fig. 1; ref. 3).

DNA-based transposons, which resemble bacterial transposons, have the potential to encode a transposase that shares homology to the $DD_{35}E$ family of integrases[4]. These elements transpose through a DNA intermediate *via* a 'cut and paste' or 'copy and paste' mechanism. This class comprises about 1.6% of the genome and its most prevalent members are mariner elements[2,5]. Although no transposition-competent human mariner element has been identified, an active mariner element, called 'sleeping beauty', has been reconstructed from the consensus transposase sequences of salmon. This element demonstrates 'cut and paste' transposition in HeLa cells, raising the possibility that active human elements exist[6].

Autonomous retrotransposons are mobilized *via* an RNA intermediate and fall into two subclasses: those that contain and those that lack long terminal repeats (LTRs). LTR-containing retrotransposons resemble retroviruses in structure, but lack a functional envelope gene. The most abundant members of this class are the human endogenous retroviruses (HERVs) which comprise about 1–2% of the genome[7]. Like mariner elements, no retrotransposition-competent HERVs are known, but some elements are expressed[7], suggesting HERVs capable of retrotransposition may exist. Retrotransposons lacking LTRs are mainly L1 elements, which comprise roughly 15% of the genome[2] and are concentrated in AT-rich regions[8]. Although over 100,000 L1s exist, the vast majority are incapable of retrotransposition because they contain deleterious mutations[9]. Greater than 95% of L1s are variably 5′ truncated and about 10% of those are internally rearranged. Thus, it seems that a relatively small number of L1s gave rise to a substantial portion of human genomic DNA.

The final class of mobile elements, the non-autonomous retrotransposons, is composed mainly of Alu elements and processed pseudogenes, both of which end in a poly A tail and lack protein-coding capacity. Their mobilization requires a cellular source of reverse transcriptase, which is most likely encoded by retrotransposition-competent L1s.

Snakes in the grass—retrotransposons cause disease

Mobile elements were first recognized as potential causal agents of human disease in 1988 when two separate insertions of truncated L1s were found to disrupt the factor VIII gene, resulting in haemophilia A (ref. 10). Six additional recently retrotransposed L1 insertions were subsequently found, one in the factor VIII gene[11], three in the *DMD* gene[12,13] (E. Bakker and G. van Ommen, pers. comm.), one in *APC* (ref. 14) and another in the β-globin gene (V. Divoky and J. Prchal, pers. comm.). While five of these occurred either in the germ line or during early development, the L1 insertion in *APC* was found in a colon cancer but not in the constitutional tissue of the patient[14], indicating that L1 retrotransposition can occur in somatic tissues.

Recent L1 insertions have a number of interesting structural features. First, each insertion differs in sequence from every other, suggesting that each arose from a distinct progenitor element. Second, although seven of the eight insertions described above are 5′ truncated to lengths varying from 538 bp to 3.8 kb, all of the protein-coding regions (to the extent they are present) have

Department of Genetics, University of Pennsylvania, School of Medicine, 415 Curie Blvd, Philadelphia, Pennsylvania 19104, USA. Correspondence should be addressed to H.H.K. e-mail: kazazian@mail.med.upenn.edu

© 1998 Nature Publishing Group http://www.nature.com/naturegenetics

remained intact. The eighth, an L1 insertion into the β-globin gene, is a full-length (6.0-kb) insertion (V. Divoky and J. Prchal, pers. comm.). These data raise the possibility that not all mutagenic L1 insertions are 'dead on arrival' and that some may retain retrotransposing ability. Indeed, full-length mouse L1 insertions appear capable of undergoing subsequent rounds of retrotransposition (see below). Third, seven of the eight insertions belong to a class of transcriptionally-active L1s, the Ta subset[15], suggesting that the majority of retrotransposition-competent L1s are derived from this subset of expressed L1 elements.

Table 1 • Interspersed repeats in the human genome

GC level (database size)	Alu	LINE 1	HERV	DNA transposons	Total (includes MIR, LINE2, MaLR, and other sequences)
< 43% GC (4102 kB)	5.7%	20.5%	1.7%	1.8%	38.5%
43–52% GC (1724 kB)	17.9%	6.1%	0.7%	2.0%	34.1%
> 52% GC (1225 kB)	20.2%	4.6%	0.5%	0.5%	30.3%
Fraction of total genome (3×109 bp)	10.0%	14.6%	1.3%	1.6%	35%

Interspersed repeats in the human genome. Data are derived from an analysis of all Human GenBank entries of greater than 40 kb (a total of seven mB from 40 loci) by Smit[2]. Sequences were pooled by GC content to demonstrate skewed distribution in AT- and GC-rich DNA.

Alu retrotransposition has also been associated with human disease. Eleven insertions of full-length Alu elements have been identified as the molecular basis of disease in the past seven years, including insertions into the factor IX gene in haemophilia[16], the *NF1* gene in neurofibromatosis[17], the *FGFR2* gene in Apert syndrome (M. Oldridge, D.M. McDonald-McGinn, E.H. Zackai and A.O.M. Wilkie, pers. comm.), the *APC* gene in desmoid tumours (K.C. Halling, pers. comm.), the *XLA* and *XSCID* genes in X-linked immunodeficiency syndromes (T. Lester, pers. comm.) and *BRCA2* in breast cancer[18]. In contrast with the L1 insertion into *APC* which occurred in somatic tissue, Alu insertions into the *APC* and *BRCA2* genes occurred either in the germ line or during early development.

It is difficult to estimate the fraction of human disease-associated mutations due to L1 and Alu insertions. The number observed so far (19 in total) is clearly an underestimate because most mutation-detection strategies currently utilize PCR and therefore fail to detect large insertions. As a putative estimate of the number of independent human mutations associated with disease, the Human Gene Mutation database, established by D.N. Cooper, E.V. Ball, P.D. Stenson and M. Krawczak (http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html), currently contains approximately 12,770 different mutations in various genes. This number is, however, a modest underestimate as it is based on single entries for each given mutation and it does not take into account multiple independent-origin (*de novo*) mutations at the same site. With this caveat in mind, the frequency with which

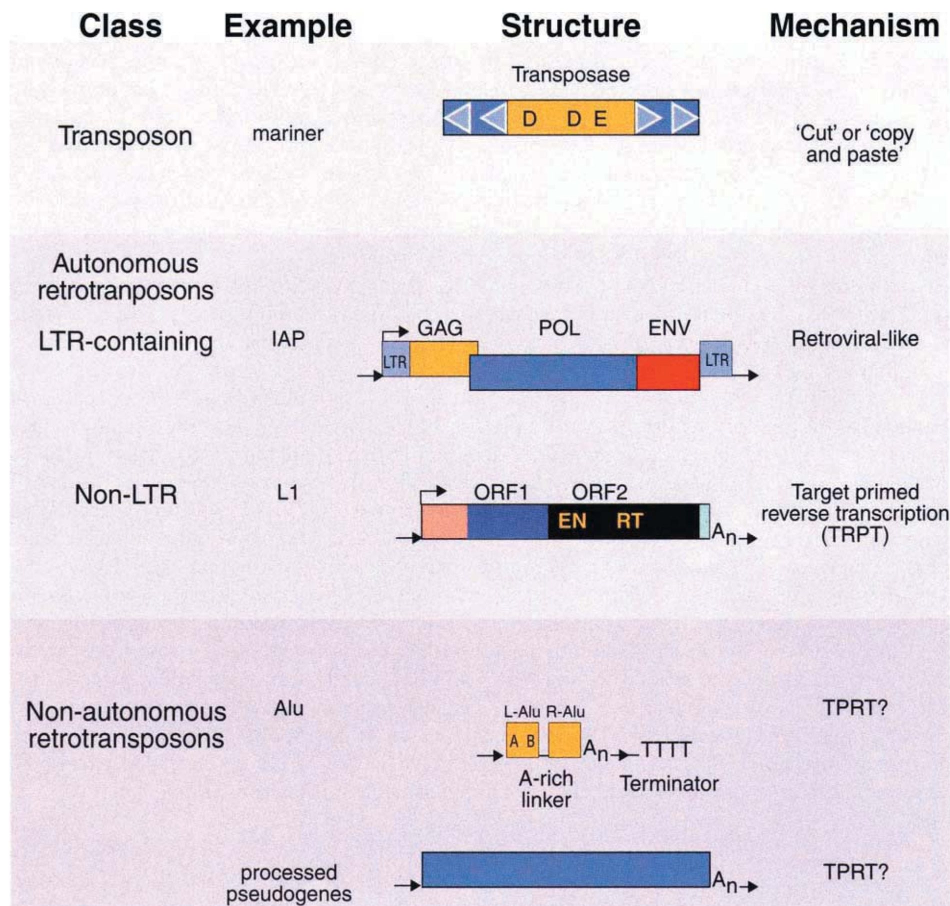

Fig. 1 Three classes of mobile elements in mammalian genomes. DNA transposons, as illustrated by mariner, transpose by either a 'cut-and-paste' or 'copy-and-paste' mechanism, using a transposase. The large arrowheads represent inverted repeats. The amino-acid residues, two Asp (D) and one Glu (E), which are critical to transposase activity are indicated. Autonomous retrotransposons, as exemplified by IAP and L1 elements, retrotranspose through an RNA intermediate. To date, all characterized IAPs are defective due to deletions or nonsense mutations and presumably are mobilized in *trans* by other cellular reverse transcriptases. LTR-retrotransposons use a retroviral-like mechanism, while non-LTR retrotransposons most likely use target-primed reverse transcription (TPRT). Arrows at each end of these retrotransposons represent target site duplications that are usually present and the arrows above the 5′ ends of the elements represent their internal promoters. Non-autonomous retrotransposons illustrated by Alu and processed pseudogenes are thought to use the TPRT mechanism of non-LTR retrotransposons for mobility. Again, the arrows at each end represent target-site duplications. In the Alu element, 'A' and 'B' represent regions which contain sequence similarity to RNA polymerase III promoter sequences. The structures are not drawn to scale. Mariner elements are about 1.4 kb, while IAP and L1 elements are 6–7 kb in length; Alu elements are approximately 300 bp and composed of left (L) and right (R) Alu monomers.

© 1998 Nature Publishing Group http://www.nature.com/naturegenetics

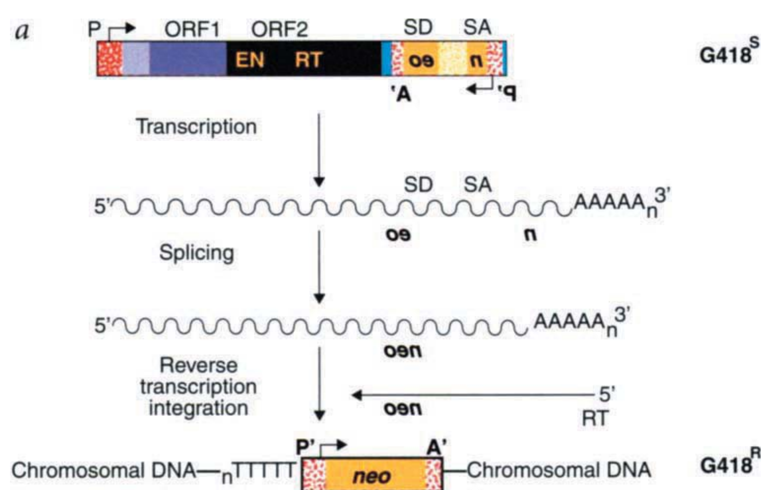

Fig. 2 Retrotransposition in cultured cells. ***a***, An overview of the L1 retrotransposition assay in HeLa cells. An active L1 was tagged with an indicator gene containing an antisense copy of the *neo* gene disrupted by IVS-2 of the γ-globin gene in the sense orientation. The splice donor (SD) and splice acceptor (SA) sites of the intron are indicated. The *neo* gene is flanked by a heterologous promoter (P') and a polyadenylation signal (A') denoted by hatched rectangles. After transfection of HeLa cells, transcripts originating from the promoter (P) which drives L1 expression are capable of splicing the intron, but contain an antisense copy of the *neo* gene. G418-resistant ($G418^R$) colonies (authentic retrotransposition events) arise only when the transcript is expressed from the P promoter and then reverse transcribed and integrated into chromosomal DNA. (Figure adapted from ref. 38.) ***b***, Results of a typical retrotransposition assay. $G418^R$ foci were fixed to flasks and stained with Giemsa for visualization. In the upper flask, an active L1 has readily retrotransposed, while, in the lower flasks, L1s with missense mutations in conserved amino acids of either the endonuclease (EN) or reverse transcriptase (RT) domains of ORF2 showed a reduction in retrotransposition frequency of 2–3 orders of magnitude.

mutations can be ascribed to retrotransposition events is estimated to be approximately 19 of 12,770, or about one in 670. Although six of eight L1 insertions associated with disease have been observed in genes located on the X chromosome, this over-representation is probably due to the hemizygosity of X-linked genes in males rather than preference for retrotransposition on the X chromosome.

The rarity of retrotransposition events causing disease in humans contrasts with the frequency in mice. Seventeen disease-associated retrotransposition events have been observed among approximately 160 spontaneous alleles at 86 characterized loci in the mouse genome (B.A. Taylor, pers. comm.). Of these, four are L1 insertions[19–23]. The $L1_{spa}$ insertion in the spastic mouse[19,20] and $L1_{Orl}$ insertion in the Orleans reeler mouse[21] are full-length and maintain intact open reading frames. Unlike the situation in humans, however, a number of spontaneous mouse mutants are due to insertion of defective LTR-retrotransposons. These elements include intracisternal A particles (IAP; Fig. 1), early transposons (ETn) and mammalian apparent LTR-retrotransposons (MaLR). Seven insertions of IAPs, which are endogenous retroviral-like elements present in approximately 1,000–2,000 copies in the mouse genome[24] are responsible for agouti and other phenotypes[25–30]. Six insertions of ETn and MaLR have been found in other spontaneous mouse mutants[31–36]. Thus, the estimated frequency of retrotransposition events among spontaneous mouse mutants is 17 in 160, or about 10%, a frequency that is more than 60-fold greater than the estimated frequency in humans. This difference is due primarily to the mobility of defective LTR-based elements in mice.

Be fruitful and multiply—the structure of active L1s

Inspection of the mutagenic insertions described so far suggests that full-length progenitors of L1 insertions are probably the best source of active mammalian retrotransposons. Approximately 3,000–4,000 human L1s are full length, but most of these are rendered inactive by nonsense and frameshift mutations. Candidates for active, retrotransposition-competent L1s were first identified through the isolation and characterization of L1.2 and LRE2, the likely progenitors of factor VIII gene and *DMD* gene insertions, respectively[37,12]. Each element has a 5′ untranslated region (UTR) that contains an internal promoter, two non-overlapping ORFs (ORF1 and ORF2) separated by a 66-bp intergenic spacer and a 205-bp 3′–UTR with an unorthodox polyadenylation signal followed by a poly A tail (Fig. 1).

The establishment of a cultured-cell assay for L1 retrotransposition demonstrated conclusively that these L1s are autonomous retrotransposons that do not rely on endogenous L1s in *trans* for their retrotransposition[38] (Fig. 2). Sassaman *et al.* developed a selective screening strategy to isolate other active L1s directly from the Ta subset[39]. Currently, seven retrotransposition-competent L1s have been identified from an estimated 30–60 present within the human genome[39].

Identification of retrotransposition-competent mouse L1s was a far less arduous task. Since $L1_{spa}$ and $L1_{Orl}$ are full-length insertions with intact ORFs, they are themselves candidates for active elements. Both elements readily retrotranspose in cultured mouse cells, indicating that the mutagenic insertions may retain the capacity to retrotranspose *in vivo*[40]. $L1_{spa}$ and $L1_{Orl}$ belong to a newly-discovered L1 subfamily, T_F, with approximately 2000 full-length members[40], suggesting that this subfamily (like the Ta subset in present-day humans) may contain the bulk of retrotransposition-competent L1s in present-day mice. Moreover, the number of active T_F elements may be further expanding the mouse genome.

Copy and paste—a proposed mechanism

L1 is the proposed 'master' mobile element in the human genome. It is thought that the proteins derived from retrotransposition-competent L1s also act on Alus and cellular mRNAs to instigate their retrotransposition[41]. Thus, elucidating the mechanism of L1 retrotransposition is likely to be crucial to understanding the mechanism of insertion of other non-autonomous elements. Recently, a potential mechanism has emerged for L1 retrotransposition based on the work of Eickbush and colleagues on the insect non-LTR element R2BM (ref. 42). Although data supporting this mechanism have been generated, unanswered questions remain at every step in the proposed pathway of retrotransposition.

Active elements are first transcribed from their internal promoter (ref. 43; Fig. 3). L1 transcription is thought to be limited to undifferentiated cells[15], early germ cells[44] and undifferenti-

© 1998 Nature Publishing Group http://www.nature.com/naturegenetics

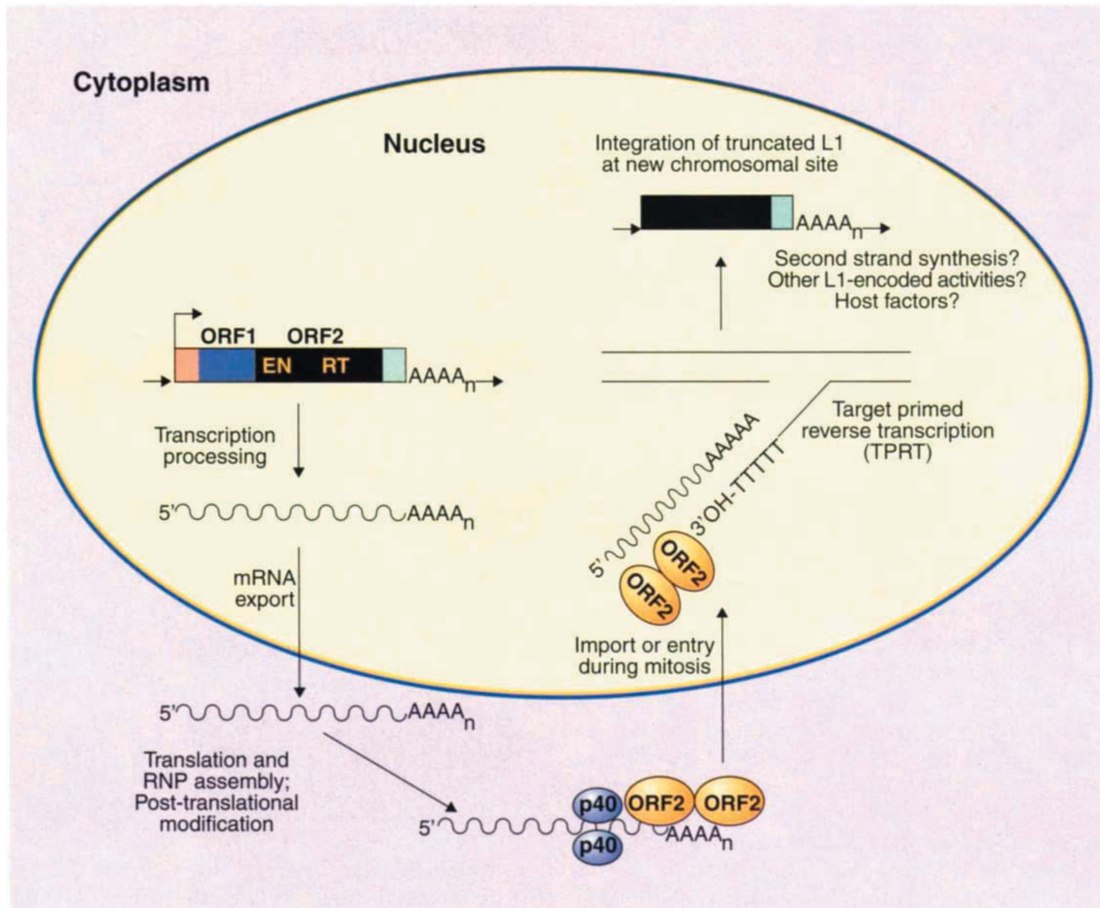

Fig. 3 A proposed mechanism of L1 retrotransposition. An active L1 is transcribed in the nucleus and is subsequently transported to and translated in the cytoplasm. The two L1 protein products, p40 and ORF2 protein, complex with their encoding L1 transcript in ribonucleoprotein particles (RNPs). The complex is then transported to recipient DNA sequences where target-primed reverse transcription (TPRT) occurs. The new, integrated L1 copy is usually truncated at its 5′ end. Although many questions remain at all steps in the pathway, there is good evidence for RNA-protein complex assembly and TPRT, suggesting the general validity of the proposed mechanism.

ated tumour cells[45]. This may result from undermethylation at CpG residues in undifferentiated cells. In fact, it has been suggested that a major purpose of CpG methylation is to reduce the expression of mobile elements in differentiated cells[46]. The length of the L1 transcript and its poly A tail suggests that transcription is carried out by RNA polymerase II. The presence of an internal promoter, however, in addition to other data[47], makes it plausible that a combination of polymerases act in L1 transcription. The ubiquitous transcription factor YY1 binds to a functionally important site in the internal promoter[48], but the role of this protein and those of other transcription factors remain unclear.

After transcription, the bicistronic L1 transcript is translated in the cytoplasm. The product of ORF1, p40, is critical for retrotransposition in HeLa cells[38] and it is easily detected by polyclonal antibodies in the cytoplasm of various undifferentiated cultured cells and tumour lines[49]. Human p40 is a sequence-specific RNA-binding protein which specifically binds to the L1 transcript near the 5′ end of ORF2 (ref. 50). Indeed, ribonucleoprotein particles (RNPs) containing p40 and L1 transcripts have been isolated from human and mouse cells[51,52]. If the proposed retrotransposition pathway is correct, these RNPs should be enriched for the RNA of active L1s. Interestingly, the only sequenced L1 cDNA obtained from mouse RNPs has two intact ORFs[53], suggesting that it could be derived from a functional L1.

Translation of ORF2 appears very much reduced compared with that of ORF1 (ref. 54). Although the ORF2 protein is also thought to bind the L1 transcript and to accompany it into the nucleus, this protein has been difficult to detect and its nuclear presence has not been confirmed. Whether p40 and ORF2 proteins are both critical to the transport of L1 RNA into the nucleus or whether L1 RNA bound to ORF2 protein reaches chromatin passively after nuclear breakdown in mitosis is unknown.

Once in the nucleus, it is likely that L1 RNA undergoes target-primed reverse transcription in order to carry out retrotransposition. The protein encoded by the insect R2BM element was shown *in vitro* to display endonuclease activity which makes a single-strand nick leaving a 5′ phosphate group and a 3′ hydroxyl group. The 3′ hydroxyl group then serves as a primer for the reverse transcriptase encoded by the R2BM protein[42]. Feng *et al.* showed that L1 ORF2 protein contains an N-terminal endonuclease (EN) domain[55], which has striking similarities to an *Escherichia coli* endonuclease, *Exo*III. Key residues in the *Exo*III catalytic site are completely conserved in L1 and all other non-LTR retrotransposons that lack site specificity[56]. Although L1 endonuclease also has similarities with apurinic/apyrimidinic (AP) endonucleases, it lacks conserved amino acids that form two helical loops essential to the structure of AP endonucleases[57]

© 1998 Nature Publishing Group http://www.nature.com/naturegenetics

and shows no preference for apurinic DNA[55]. Furthermore, in contrast to *Exo*III, the L1 endonuclease has a very low turnover number, prefers target sites composed of $(Py)_n/(Pu)_n$ (where '/' indicates the nick site) and is critical for retrotransposition in HeLa cells[55] (Fig. 2*b*).

The ORF2 protein also contains a reverse transcriptase (RT) domain (Fig. 1), composed of seven regions of sequence similarity with RTs of other retrotransposons and retroviruses[58]. This RT domain is critical for retrotransposition in HeLa cells[38] (Fig. 2). While it is unclear precisely how the enzyme functions, it is thought that the 5′ truncation is due to its relatively poor processivity. Interestingly, the L1 RT domain has considerable sequence similarity to the catalytic subunit of telomerase[59,60], raising the possibility that a retrotransposon protein was co-opted for an essential cellular function. Alternatively, it is feasible that telomerase predated the RT of non-LTR retrotransposons.

Although there is some insight into the mechanism of retrotransposition, the subsequent steps of second-strand cleavage of the target DNA at a downstream site (which is necessary for target-site duplication), the second-strand DNA synthesis and ligation of the retrotransposing DNA to genomic DNA still remain to be explained. Furthermore, any proposed retrotransposition mechanism must also take into account inversions of the 5′ region of truncated L1 insertions that have been occasionally observed[10,12,14].

To thine own self be true—*cis* preference

The protein products of a particular L1 are thought to bind usually to the RNA of that same L1. This is described as *cis* preference because binding to the encoding RNA is preferred, but not exclusive. *Cis* preference was suggested by the lack of nonsense and frameshift mutations in mutagenic L1 insertions[10] and the finding that two human insertions were exact, although truncated, copies of their retrotransposition-competent progenitors[12,37,38]. Recently, we used cotransfection of active and mutant L1s into HeLa cells to demonstrate directly a *cis* preference of L1 proteins for their encoding RNA *in vivo*. Although the *cis* preference was evident in these experiments, low-level *trans* complementation could not be ruled out. Heidman and colleagues have recently shown low-level reverse transcription of cellular messages in HeLa cells transfected with active L1.2 (ref. 61). *Cis* preference is important to the organism because it limits mutagenesis due to retrotransposition of RNAs with 3′ poly A tails, especially defective L1 transcripts and cellular mRNAs.

But then how does one explain the retrotransposition of Alu elements if these events are under the concerted direction of EN and RT enzymatic activity of L1s? As Alu elements number between 500,000–1,000,000 copies in the human genome, they have been even more successful 'genomic colonizers' than L1s themselves. Boeke has suggested that the EN and RT activities of the L1 ORF2 protein are available to Alu by dint of cellular proximity[62]. Two signal-recognition particle proteins, SRP9 and SRP14, form a heterodimer and bind with high affinity to the 7SL domain of Alu[63–65]. This complex most likely binds to the large ribosomal subunit, leaving Alu poly A in close proximity to L1 ORF2 protein. Thus, as Boeke puts it, the Alu poly A RNAs are "hanging around the ribosomal neighborhood, presumably greatly increasing their chances of hijacking a retrotransposon RT" (ref. 62). Support for the role of L1s in Alu retrotransposition comes from the similar sequence and length of target-site duplications (TSDs) at the 5′ and 3′ ends of L1 and Alu elements, suggesting that retrotransposition of both L1 and Alu elements is mediated by the same endonuclease, L1 EN (ref. 66). It is unclear, however, why only a minor subset of transcribed Alus appears to be capable of retrotransposition[67] and also why the genomic distributions of Alus (GC-rich DNA) and L1s (AT-rich DNA) differ markedly[2] (Table 1). It is possible that the structure of the RNA complexed with L1 protein(s) plays a role in determining target-site specificity.

Although Luan *et al.* showed that 250 nucleotides at the 3′ end of the R2BM transcript are crucial to the RT activity of R2BM (ref. 68), the poly A tail itself, and not the 3′ UTR sequences, appears critical to the RT activity of L1. Holmes *et al.* found that a transcript with 489 nucleotides of non-L1 sequence just preceding the poly A tail could be readily retrotransposed *in vivo*[12]. In HeLa cells, Moran *et al.* demonstrated that the 3′ UTR is dispensable for retrotransposition of an active L1 and that 138 bp of non-L1 sequence prior to the poly A could be reverse transcribed[38]. Thus, unlike R2BM RT, L1 RT does not seem to require a specific RNA sequence upstream of its poly A. This has led to the hypothesis that L1 RT binds to the poly A stretch itself, and that although it greatly prefers its own RNA, it will act rarely on other cellular mRNAs and with greater frequency on Alu RNA because of the proximity of Alu RNAs to L1 RNA on the ribosome.

Next in LINE—potential uses

Harnessing the power of retrotransposons is envisaged to lead to practical applications. Our knowledge of L1 biology, however, is currently too limited to put human retrotransposons to use. Nevertheless, L1s have enticing potential applications. They could be used as insertional mutagens in mice through transfection of ES cells and subsequent blastocyst injection. Alternatively, retrotransposition could be carried out in the sperm of transgenic mice, and sperm containing retrotransposition events could be sorted by the presence of a fluorescent tag and used to produce mutant progeny. Genes which, when disrupted by L1 insertion, result in a phenotype of interest can then be cloned by utility of the tagged L1 element. As human L1s are capable of high-frequency retrotransposition in mouse LTK- cells[38], these strategies offer potential for effective insertional mutagenesis in mice.

L1s also have potential as a gene delivery vehicle. One vehicle presently being assessed in cultured mammalian cells is a chimaeric vector containing an adenoviral backbone carrying an active human L1. Attractive features of using human L1s in gene delivery are the potential for ongoing delivery of multiple gene copies and the endogenous nature of L1 DNA. One drawback, however, is our current inability to control both truncation of the inserted sequence and the sites of L1 insertion—most retrotransposed genes would be inactive due to truncation and a fraction of them would disrupt genes. Further research into the intricate biological nature of these mobile elements will likely lead to possible ways of overcoming these limitations and extending the scope of practical applications.

Acknowledgements

We thank B. Taylor for unpublished information on characterized mouse mutations and R. DeBerardinis, J. Goodier, E. Ostertag and E. Luning Prak for critique of the manuscript. H.H.K is supported by a grant from the NIH and J.V.M was supported in part by a Damon Runyon-Walter Winchell Cancer Research Fund Fellowship.

npg © 1998 Nature Publishing Group http://www.nature.com/naturegenetics

1. Smit, A.F., Toth, G., Riggs, A.D. & Jurka, J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246**, 401–417 (1995).
2. Smit, A.F. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6**, 743–748 (1996).
3. Temin, H.M. Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. *Mol. Biol. Evol.* **2**, 455–468 (1985).
4. Rice, P., Craigie, R. & Davies, D.R. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* **6**, 76–83 (1996).
5. Plasterk, R.H. The Tc1/mariner transposon family. *Curr. Top. Microbiol. Immunol.* **204**, 125–143 (1996).
6. Ivics, Z., Hackett, P.B., Plasterk, R.H. & Izsvák, Z. Molecular Reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510 (1997).
7. Lower, R., Lower, J. & Kurth, R. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci. USA* **93**, 5177–5184 (1996).
8. Korenberg, J.R. & Rykowski, M.C. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* **53**, 391–400 (1988).
9. Fanning, T.G. & Singer, M.F. LINE-1: a mammalian transposable element. *Biochim. Biophys. Acta* **910**, 203–212 (1987).
10. Kazazian, H.H., Jr. *et al.* Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164–166 (1988).
11. Woods-Samuels, P. *et al.* Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. *Genomics* **4**, 290–296 (1989).
12. Holmes, S.E., Dombrowski, B.A., Krebs, C.M., Boehm, C.D. & Kazazian, H.H., Jr. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nature Genet.* **7**, 143–148 (1994).
13. Narita, N. *et al.* Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.* **91**, 1862–1867 (1993).
14. Miki, Y. *et al.* Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52**, 643–645 (1992).
15. Skowronski, J., Fanning, T.G. & Singer, M.F. Unit-length Line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8**, 1385–1397 (1988).
16. Vidaud, D. *et al.* Haemophilia B due to a de novo insertion of a human-specific Alu subfamily member within the coding region of the factor IX gene. *Eur. J. Hum. Genet.* **1**, 30–36 (1993).
17. Wallace, M.R. *et al.* A de novo Alu insertion results in neurofibromatosis type 1. *Nature* **353**, 864–866 (1991).
18. Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T. & Nakamura, Y. Mutation analysis in the BRCA2 gene in primary breast cancers. *Nature Genet.* **13**, 245–247 (1996).
19. Kingsmore, S.F. *et al.* Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nature Genet.* **7**, 136–141 (1994).
20. Mulhardt, C. *et al.* The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intronic insertion of L1 element. *Neuron* **13**, 1003–1015 (1994).
21. Takahara, T. *et al.* Dysfunction of the Orleans reeler gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. *Hum. Mol. Genet.* **5**, 989–993 (1996).
22. Kohrman, D.C., Harris, J.B. & Meisler, M.H. Mutation detection in the med and medJ alleles of the sodium channel Scn8a. Unusual splicing due to a minor class AT-AC intron. *J. Biol. Chem.* **271**, 17576–17581 (1996).
23. Perou, C.M., Pryor, R.J., Naas, T.P. & Kaplan, J. The bg allele mutation is due to a LINE1 element retrotransposition. *Genomics* **42**, 366–368 (1997).
24. Kuff, E.L. & Lueders, K.K. The intracisternal A-particle gene family: structure and functional aspects. *Adv. Cancer Res.* **51**, 183–276 (1988).
25. Duhl, D.M., Vrieling, H., Miller, K.A., Wolff, G.L. & Barsh, G.S. Neomorphic agouti mutations in obese yellow mice. *Nature Genet.* **8**, 59–65 (1994).
26. Michaud, E.J. *et al.* Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* **8**, 1463–1472 (1994).
27. Argeson, A.C., Nelson, K.K. & Siracusa, L.D. Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus. *Genetics* **142**, 557–567 (1996).
28. Gardner, J.M. *et al.* The mouse pale ear (ep) mutation is the homologue of human Hermansky-Pudlak syndrome. *Proc. Natl. Acad. Sci. USA* **94**, 9238–9243 (1997).
29. Hamilton, B.A. *et al.* The vibrator mutation causes neurodegeneration via reduced expression of PITP alpha: positional complementation cloning and extragenic suppression. *Neuron* **18**, 711–722 (1997).
30. Kuster, J.E., Guarnieri, M.H., Ault, J.G., Flaherty, L. & Swiatek, P.J. IAP insertion in the murine LamB3 gene results in junctional epidermolysis bullosa. *Mamm. Genome* **8**, 673–681 (1997).
31. Herrmann, B.G., Labeit, S., Poustka, A., King, T.R. & Lehrach, H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617–622 (1990).
32. Steinmeyer, K. *et al.* Inactivation of muscle chloride channel by transposon insertion in myotonic mice. *Nature* **354**, 304–308 (1991).
33. Adachi, M., Watanabe-Fukunaga, R. & Nagata, S. Aberrant transcription caused by the insertion of an early transposable element in an intron of the Fas antigen gene of lpr mice. *Proc. Natl. Acad. Sci. USA* **90**, 1756–1760 (1993).
34. Moon, B.C. & Friedman, J.M. The molecular basis of the obese mutation in ob2J mice. *Genomics* **42**, 152–156 (1997).
35. Shiels, A. & Bassnett, S. Mutations in the founder of the MIP gene family underlie cataract development in the mouse. *Nature Genet.* **12**, 212–215 (1996).
36. Loftus, S.K. *et al.* Murine model of Niemann-Pick C disease: mutation in a cholesterol homeostasis gene. *Science* **277**, 232–235 (1997).
37. Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F. & Kazazian, H., Jr. Isolation of an active human transposable element. *Science* **254**, 1805–1808 (1991).
38. Moran, J.V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
39. Sassaman, D.M. *et al.* Many human L1 elements are capable of retrotransposition. *Nature Genet.* **16**, 37–43 (1997).
40. Naas, T.P. *et al.* An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.* **17**, 590–597 (1998).
41. Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D. & Gabriel, A. Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808–1810 (1991).
42. Luan, D.D., Korman, M.H., Jakubczak, J.L. & Eickbush, T.H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).
43. Swergold, G.D. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* **10**, 6718–6729 (1990).
44. Branciforte, D. & Martin, S.L. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol. Cell. Biol.* **14**, 2584–2592 (1994).
45. Bratthauer, G.L. & Fanning, T.G. Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* **7**, 507–510 (1992).
46. Yoder, J.A., Walsh, C.P. & Bestor, T.H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
47. Kurose, K., Hata, K., Hattori, M. & Sakaki, Y. RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. *Nucleic Acids Res.* **23**, 3704–3709 (1995).
48. Becker, K.G., Swergold, G.D., Ozato, K. & Thayer, R.E. Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. *Hum. Mol. Genet.* **2**, 1697–1702 (1993).
49. Leibold, D.M. *et al.* Translation of LINE-1 DNA elements in vitro and in human cells. *Proc. Natl. Acad. Sci. USA* **87**, 6990–6994 (1990).
50. Hohjoh, H. & Singer, M.F. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* **15**, 630–639 (1996).
51. Hohjoh, H. & Singer, M.F. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.* **16**, 6034–6043 (1997).
52. Martin, S.L. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* **11**, 4804–4807 (1991).
53. Martin, S.L. Characterization of a LINE-1 cDNA that originated from RNA present in ribonucleoprotein particles: implications for the structure of an active mouse LINE-1. *Gene* **153**, 261–266 (1995).
54. McMillan, J.P. & Singer, M.F. Translation of the human LINE-1 element, L1Hs. *Proc. Natl. Acad. Sci. USA* **90**, 11533–11537 (1993).
55. Feng, Q., Moran, J.V., Kazazian, H.H. & Boeke, J.D. Human L1 retrotransposon encodes a conserved endonuclease required fo retrotransposition. *Cell* **87**, 905–916 (1996).
56. Mol, C.D., Kuo, C.F., Thayer, M.M., Cunningham, R.P. & Tainer, J.A. Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature* **374**, 381–386 (1995).
57. Gorman, M.A. *et al.* The crystal structure of the human DNA repair endonuclease HAP1 suggests the recognition of extra-helical deoxyribose at DNA abasic sites. *EMBO J.* **16**, 6548–6558 (1997).
58. Xiong, Y. & Eickbush, T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**, 3353–3362 (1990).
59. Nakamura, T.M. *et al.* Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955–959 (1997).
60. Eickbush, T.H. Telomerase and retrotransposons: which came first? *Science* **277**, 911–912 (1997).
61. Dhellin, O., Maestre, J. & Heidmann, T. Functional difference between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcriptase. *EMBO J.* **16**, 6590–6602 (1997).
62. Boeke, J.D. LINEs and Alus—the polyA connection. *Nature Genet.* **16**, 6–7 (1997).
63. Chang, D.Y. & Maraia, R.J. A cellular protein binds B1 and Alu small cytoplasmic RNAs in vitro. *J. Biol. Chem.* **268**, 6423–6428 (1993).
64. Chang, D.Y. *et al.* A human Alu RNA-binding protein whose expression is associated with accumulation of small cytoplasmic Alu RNA. *Mol. Cell. Biol.* **14**, 3949–3959 (1994).
65. Chang, D.Y., Sasaki-Tozawa, N., Green, L.K. & Maraia, R.J. A trinucleotide repeat-associated increase in the level of Alu RNA-binding protein occurred during the same period as the major Alu amplification that accompanied anthropoid evolution. *Mol. Cell. Biol.* **15**, 2109–2116 (1995).
66. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1872–1877 (1997).
67. Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A. & Deininger, P.L. cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J. Mol. Biol.* **271**, 222–234 (1997).
68. Luan, D.D. & Eickbush, T.H. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol. Cell. Biol.* **15**, 3882–3891 (1995).