LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression?

Jeffrey S. Han and Jef D. Boeke*

Summary
LINE-1 (L1) retrotransposons are replicating repetitive elements that, by mass, are the most-abundant sequences in the human genome. Over one-third of mammalian genomes are the result, directly or indirectly, of L1 retrotransposition. L1 encodes two proteins: ORF1, an RNA-binding protein, and ORF2, an endonuclease/reverse transcriptase. Both proteins are required for L1 mobilization. Apart from the obvious function of self-replication, it is not clear what other roles, if any, L1 plays within its host. The sheer magnitude of L1 sequences in our genome has fueled speculation that over evolutionary time L1 insertions may structurally modify endogenous genes and regulate gene expression. Here we provide a review of L1 replication and its potential functional consequences. ***BioEssays*** **27:775–784, 2005.**

Introduction

The recent whole-genome sequencing binge has made it apparent that mammalian genomes are littered with enormous numbers of transposable elements interspersed within and between single-copy endogenous genes.[1–5] The most-abundant self-replicating class of human transposon is the LINE-1, or L1, element.[2,6] This sequence replicates through a round of transcription and reverse transcription, generating a new copy at a new genomic location (Fig. 1). This can lead to obvious deleterious effects—for example, the insertion of L1 into protein-coding sequences of a gene (see the website http://www.med.upenn.edu/genetics/labs/kazazian/kazazian_index.html, and the references therein), abolishing gene function. However, protein-coding sequences are a small target (only 1% of the genome[3]) and thus most new L1 integrations are in introns or intergenic regions.[7] The consequences of these L1 insertions on gene function have not been thoroughly studied.

Recent experiments have demonstrated that, in experimental systems, L1 sequence within a transcriptional unit can affect not only the structure of the target DNA,[8,9] but the structure and amount of mRNA produced.[10] This raises the question of whether L1 sequences integrated into endogenous genes have similar effects. In this review, we will provide an overview of L1 biology and discuss models of how L1 may alter transcriptional profiles and generate novel mRNA and protein isoforms. We also hypothesize on how these models could modify our view of mammalian genome evolution.

Historical overview

Before the age of genome sequencing, it was already clear that the DNA of many organisms contained more than just single-copy genes. This was discovered by analyzing the renaturation kinetics of sheared genomic DNA.[11] For example, in most bacteria and viruses, the renaturation of genomic DNA takes place over a narrow time range, suggesting that the DNA fragments are present at a similar concentration. This, in conjunction with measuring the mass of the respective genomes, led to the conclusion that the DNA sequences of these organisms, for the most part, are present in single copy. In contrast, when the renaturation kinetics of DNA from higher organisms such as humans is measured, large fractions renature at different times (Fig. 2A). These simple experiments showed that, in addition to single-copy sequences, there are families of reiterated sequence scattered throughout our genome.

The presence of highly repeated sequences is also evident by simply digesting human genomic DNA with a restriction enzyme and analyzing the result on an ethidium-stained gel. On top of the expected smear of variable length DNA, discrete bands can sometimes be detected. Cloning and characterization of *Kpn*I-generated bands led to the discovery of the *Kpn*I family of repetitive elements in humans,[12] and similar studies uncovered the homologous *Bam*HI family in mice.[13] Since full-length members of these families were over 6 kilobases long and interspersed between single-copy DNA, these were renamed as long interspersed nuclear elements (LINEs).[14]

Department of Molecular Biology and Genetics and High Throughput Biology Center, The Johns Hopkins University School of Medicine, Baltimore, MD
*Correspondence to: Jef D. Boeke, Department of Molecular Biology and Genetics and High Throughput Biology Center, The Johns Hopkins University School of Medicine, Baltimore, MD 21205.
E-mail: jboeke@jhmi.edu
DOI 10.1002/bies.20257
Published online in Wiley InterScience (www.interscience.wiley.com).

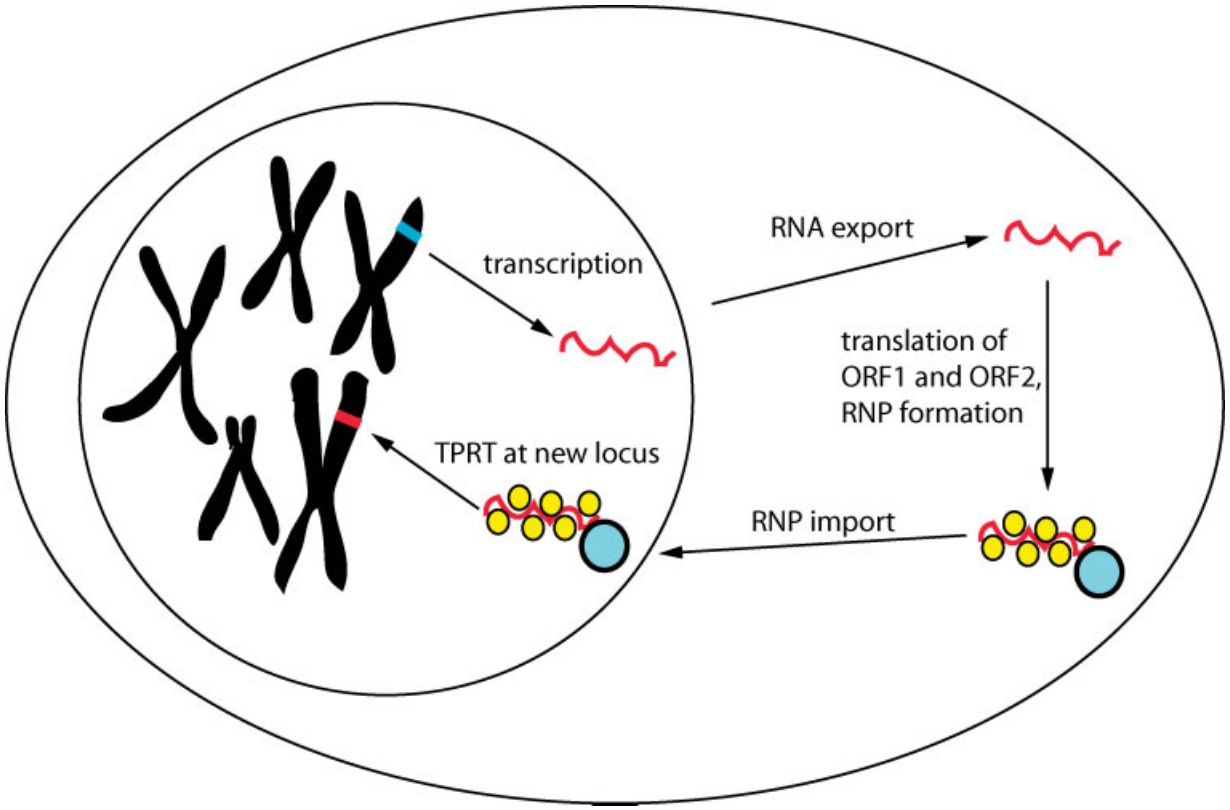

Figure 1. The life cycle of the L1 retrotransposon. A full-length, replication competent L1 (depicted as a blue band in the chromosomes) is transcribed. The resultant RNA is transported to the cytoplasm, where ORF1 and ORF2 are translated and form a ribonucleoprotein particle (RNP) with the L1 RNA that encoded them. When L1 retrotransposition is activated, the RNP is imported into the nucleus and the L1 RNA is reverse transcribed at the target site via target-primed reverse transcription (TPRT). This new L1 insertion is depicted as a red band in the chromosome. If the new insertion is full-length, it has the potential to also serve as a donor for subsequent generations of L1 retrotransposition events.

The other major component of mammalian repetitive DNA is short interspersed nuclear elements (SINEs), comprising predominantly Alu and Alu-like elements, and is less than 500 bp long (these were similarly discovered based on a pair of conserved *Alu*I sites).[14]

LINE-1, now usually referred to by the simpler name L1, is the most-recent lineage of elements of its type in mammals. It represents one of a multitude of quite diverse non-LTR retrotransposons found in host species ranging from fungi to insects to plants.[15] Inspection of the nucleotide sequence of a full-length L1 revealed two potential large open reading frames with substitution patterns consistent with protein-coding sequence including a reverse transcriptase-coding region.[16–18] This genomic structure was reminiscent of retrotransposons, and L1 was directly shown to be a mobile element by the identification of mutations resulting from new L1 insertions.[19] Further investigation has shown that L1 is indeed a retrotransposon, although it is structurally and mechanistically distinct from retroviruses and retroviral-like (LTR-containing) retrotransposons.[20] Thus L1 elements are also known as non-LTR retrotransposons, and retroviral-like retrotransposons are known as LTR-retrotransposons. A tissue-culture assay for retrotransposition activity has been useful in dissecting out the sequence requirements for optimal L1 activity.[21]

The completion of various eukaryotic sequencing projects[1–5] has confirmed these earlier studies demonstrating that essentially all eukaryotic genomes are strewn with interspersed transposable elements. In mammals, almost all of these fall into one of four classes: LINEs (autonomous non-LTR retrotransposons), SINEs (non-autonomous non-LTR retrotransposons), LTR-retrotransposons, and DNA transposons (Fig. 2B). These collectively account for at least 45% of human DNA,[2] and perhaps more given that older "relic" transposon copies may have been rendered unrecognizable by mutation. Of these element classes, L1 constitutes the largest fraction of the human genome, directly accounting for

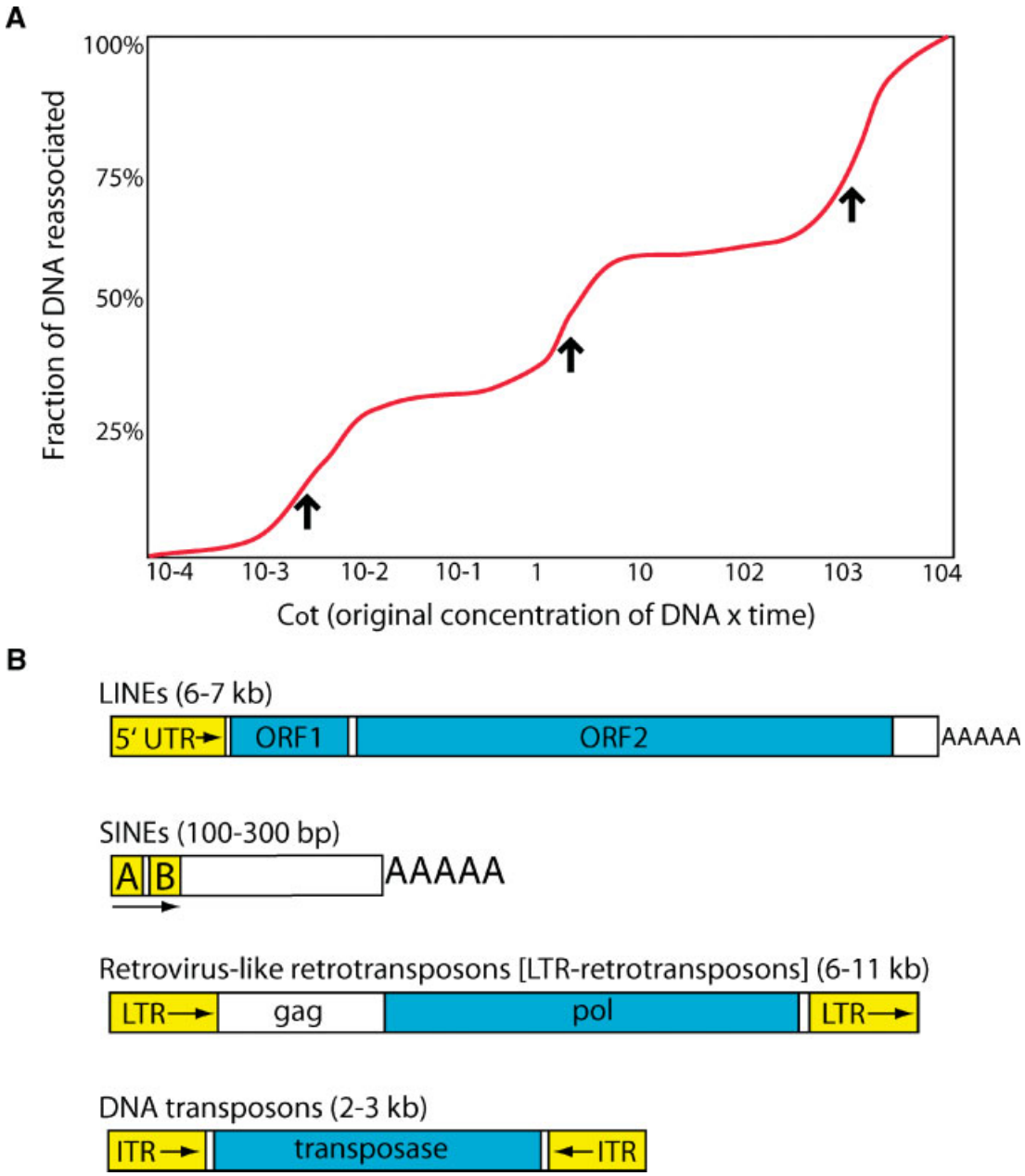

Figure 2. Repetitive elements are interspersed throughout mammalian genomes. **A:** A schematic of a C_ot curve of mammalian DNA. Fraction of DNA reassociated can be measured by UV absorbance. The C_ot curve shown represents three populations of sequences, each with a different copy number. The $C_ot_{1/2}$ of each of these populations is shown by the arrows and represents the time it takes for 1/2 of the population to reassociate. High-copy-number sequences are present at a higher number at a given concentration of total DNA and thus renature faster, and have a smaller $C_ot_{1/2}$. Single-copy genes have the largest $C_ot_{1/2}$. **B:** The major classes of transposable elements in the human genome. LINEs (L1) are retrotransposons that replicate by target-primed reverse transcription. SINEs are small Pol III transcripts that usurp the L1 machinery to also replicate via target-primed reverse transcription. Both LINEs and SINEs are also called non-LTR retrotransposons. Retrovirus-like retrotransposons (also called LTR-retrotransposons) are retrotransposons that replicate via a mechanism similar to retroviruses, but without an extracellular stage or a functional envelope protein. DNA transposons are cut and paste transposons.

at least 21% of the DNA.[2] Some of these L1 elements are within genes, and over 75% of human genes contain at least one L1 insertion,[10] usually as part of introns, 5′ untranslated sequence, or 3′ untranslated sequence.

The L1 life cycle

The structure of a full-length, genomic L1 is depicted in Fig. 2B, and a summary of the life cycle is depicted in Fig. 1. The 5′ untranslated region (5′UTR) contains an internal promoter that directs initiation of L1 element transcription at base 1.[22] The ''internal promoter'' structure makes sense for a retrotransposon, which must take its promoter with it to generate an active copy when it inserts at a new location. In this way, the L1 promoter resembles the promoter of eukaryotic tRNA genes, which are transcribed by RNA polymerase III (Pol III).[23] Indeed, the L1 5′ UTR sequence has been shown to be transcribed actively in vitro by Pol III,[24] but the preponderance of data, including the protein-coding capacity of the RNA, poly(A) tail, and in vivo α-amanitin inhibition, suggests that non-LTR retrotransposon transcripts are made by RNA polymerase II.[25,26] Within the 5′ UTR, several transcription factor-binding sites have been identified. SRY-family-binding sites[27] and a RUNX3-binding site[28] appear to be important for transcriptional activation, while a YY-1-binding site near the 5′ end directs accurate transcription initiation.[29] Since L1 transcription is presumed to occur predominantly in germ cells[30–33] and none of the currently identified transcription factors are germ-cell specific, we might expect that there are more factors waiting to be revealed.

After L1 RNA transcription and transport to the cytoplasm, the two L1-encoded proteins, ORF1 and ORF2, are translated. ORF1 is a non-sequence-specific RNA-binding protein with in vitro nucleic acid chaperone activity,[34,35] but an unknown mechanistic role in L1 replication. ORF2 contains endonuclease[36] and reverse transcriptase[37] activities, which have well-defined roles in the retrotransposition process. Both ORF1 and ORF2 must be encoded in *cis* for full activity, meaning that the L1 proteins preferentially act on the RNA that encoded them.[38,39] The L1 proteins and RNA assemble into a ribonucleoprotein particle[40] which is predominantly cytoplasmic, but presumably a small amount of it, or a subcomplex, must be transported into the nucleus. This transport may be mediated by a nuclear/nucleolar localization signal recently mapped to the N-terminus of ORF2.[41]

Once in the nucleus, L1 copies its RNA into DNA via a process called target-primed reverse transcription (TPRT). TPRT was first demonstrated for the R2 non-LTR retrotransposon in the silkworm *Bombyx mori*,[42,43] and subsequent evidence for it was obtained for L1.[36,44] During TPRT, the endonuclease domain of ORF2 makes a nick in genomic DNA, generating a 3′ hydroxyl. The reverse transcriptase of ORF2 uses this freed hydroxyl to prime reverse transcription of L1 RNA. At some point, the second strand of target DNA is nicked and used to prime second-strand synthesis. Depending on where the second-strand nick is with respect to the first strand, this may lead to a target site duplication or to a target site deletion.[45] The details of how L1 integration is resolved are not known. It is notable that most L1 integrants are severely 5′ truncated and non-functional for further replication,[7,46] possibly because L1 reverse transcription is relatively non-processive. Further discussion of the structural and functional characterization of ORF1/ORF2 is covered in greater detail elsewhere.[47]

Ways L1 can wreak havoc in the host genome

There is a multitude of ways that L1 elements can alter genome structure. By its very nature, L1 is an insertional mutagen. Of the hundreds of thousands of L1s in our genome, close to 10,000 are full-length and about 100 are active for retrotransposition in a tissue culture assay.[48] Thus our genome, to this day, is being continuously mutated by new L1 insertions. Given this, there are readily apparent ways that L1 can modify the genome in a detrimental manner. The most straightforward is the insertion of L1 into genic functional sequences—promoters, enhancers, exons, etc. Such insertions are expected to severely compromise gene function (Fig. 3A). In addition, homologous recombination between L1 elements at nonallelic sites has been shown to cause genome rearrangements (Fig. 3B),[49] although the frequency of this event is unknown and likely not high, at least in organisms that survive to gestation. De novo insertions created in a retrotransposition assay system using tissue culture cells as well as naturally occurring L1 insertions have shown that the process of L1 retrotransposition itself can also create large deletions and/or rearrangements (Fig. 3C).[8,9] Finally, the insertion of L1 into the intron of a gene can cause exon skipping or alternative splicing (Fig. 3D).[50–54]

With all the potential pitfalls of L1 retrotransposition, how could this mobile element not only survive in its host, but reproduce with such spectacular success? One might think that without some redeeming qualities, L1 elements would be selected against over evolutionary time and eventually go extinct. However, as originally pointed out by Hickey,[55] this is not necessarily the case. In sexually reproducing organisms purely parasitic transposons can in principle become fixed in the genome as long as their selection coefficient (the fraction by which host fitness is reduced by the transposon) is less than 0.5.[55,56] This is expected to select for aggressive transposons in outbred sexual populations (such as mammals), with the host evolving to negatively regulate transposon activity.[57] Recent evidence suggests that, in mammals, L1 (presumably an aggressive retrotransposon due to its ubiquitous distribution in our genome) is downregulated by cytosine methylation of its promoter,[58] supporting this ''genome defense'' hypothesis.[59]

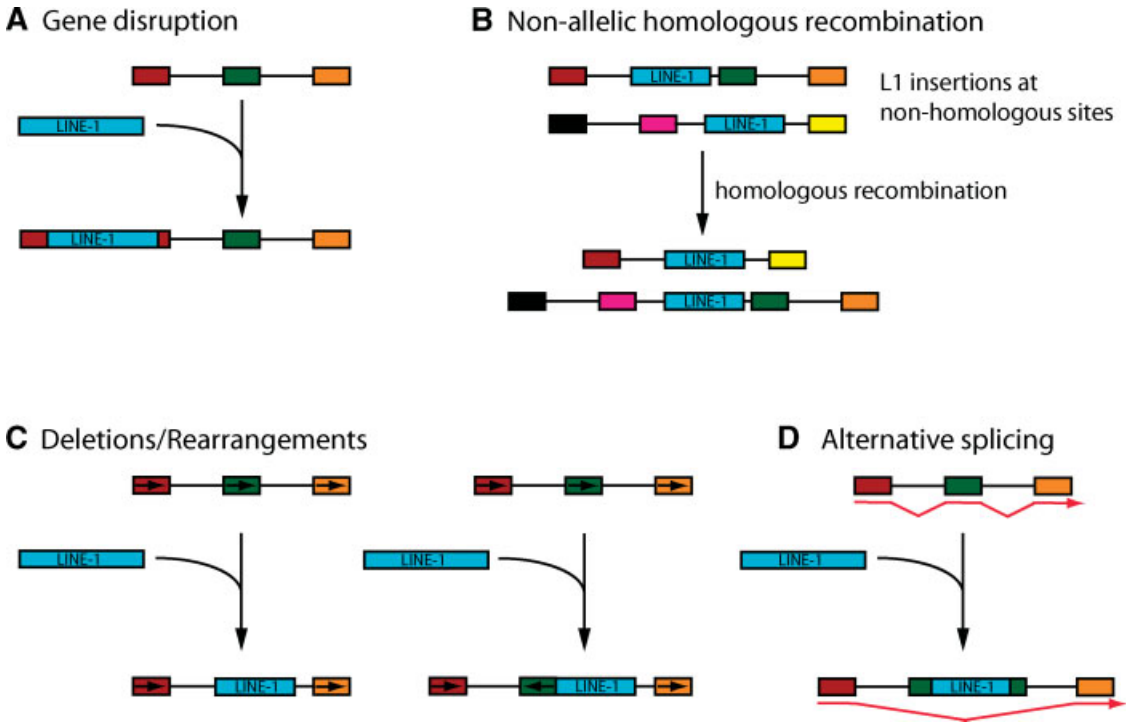

Figure 3. Mechanisms by which L1 retrotransposition can disrupt genes. Exons are depicted as blocks and introns are depicted as lines. **A:** Integrations into an essential component of the gene (in this example an exon). **B:** Homologous recombination between L1s at non-homologous sites. This could lead to gain or loss of genetic material on a chromosome, and could also lead to translocations. **C:** During the process of L1 retrotransposition, deletions and/or rearrangements at the target site have been reported. **D:** L1 insertions can cause exon skipping (the example shown) or activation of cryptic splice sites.

Since L1 can, in theory, thrive in mammalian genomes as a purely selfish element, there is no need for it to provide any useful function to the host. Even so, transposable elements have clearly been co-opted by the host to perform important functions. For example, *Drosophila* telomeres are maintained by the non-LTR retrotransposons TART and HeT-A,[60–62] and RAG1/RAG2 of the vertebrate immune system appears to have evolved from a DNA transposon.[63,64] This, combined with the awesome abundance of L1 sequence in our genome, especially within or adjacent to transcriptional units, has fueled speculation that L1 insertions may have functional consequences for the genes that they insert into. As discussed below, there is experimental evidence that L1 insertions can, in principle, do more than simply destroy gene function by insertional inactivation. To what extent this actually happens in the human genome is still open for debate.

A potpourri of L1-generated alleles

There are several ways that L1 has been hypothesized to affect gene function without necessarily destroying it. First, as mentioned above, the L1 promoter appears to be silenced by cytosine methylation.[58] If this leads to heterochromatin formation and this heterochromatin spreads into adjacent sequences, this could well attenuate the transcriptional activity of a nearby endogenous gene. This could create different alleles of the same gene, with each allele expressing the product at a different level (Fig. 4A). Although this is an active area of investigation, no specific examples of such a phenomenon have been found to date for L1 elements. In a similar mechanism, L1 promotes transcription from the 5′ UTR in either direction (there are both sense and anti-sense promoter activities in the L1 5′ UTR[22,65,66]), potentially leading to ectopic expression of genes. This ectopic expression, in turn, could be regulated epigenetically through DNA methylation (Fig. 4A). Again, we have yet to discover a naturally occurring example of this for L1, but it is noteworthy to mention that there are known mutants of the *agouti* gene in mice controlled by an IAP retrotransposon promoter (IAP is an LTR-retrotransposon), which is in turn controlled epigenetically.[67] A similar example exists in mouse for the *Cabp* gene, where an IAP insertion generates LTR-driven transcripts regulated by methylation.[68] It has recently been shown that there are many other chimeric

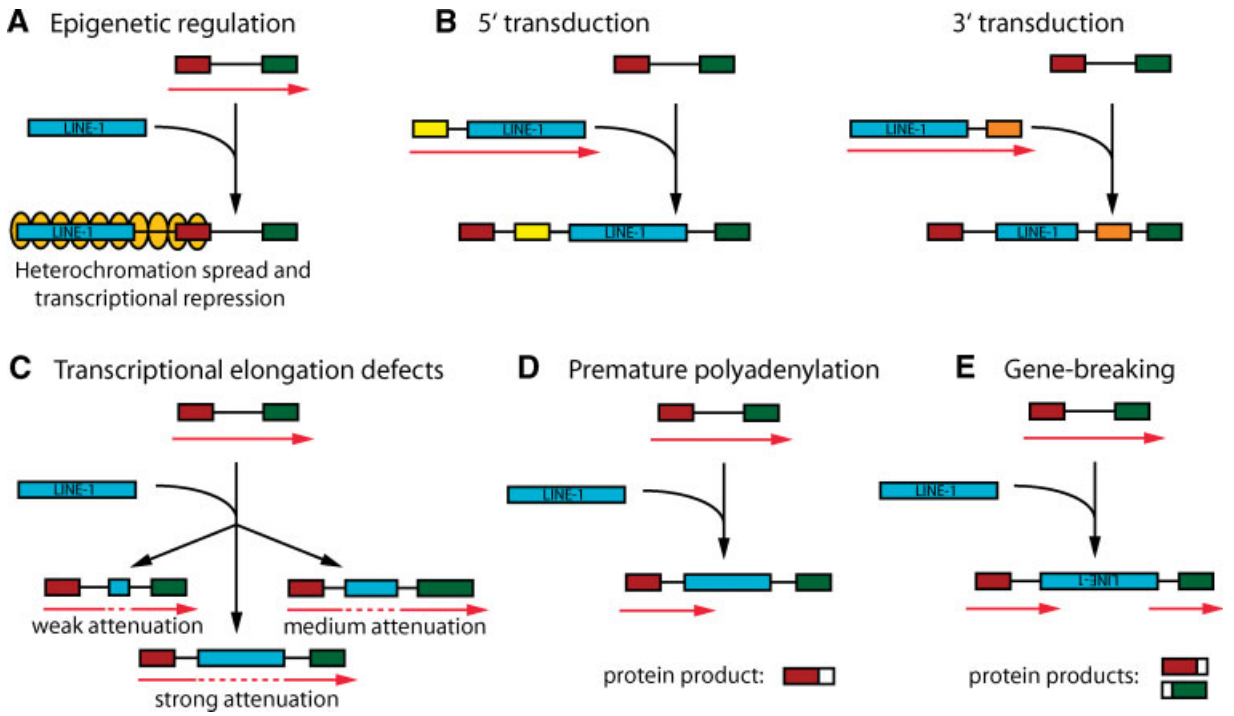

Figure 4. Ways that L1 can create novel functional alleles. **A:** Epigenetic modification. L1 sequences can be methylated, which is hypothesized to initiate heterochromatin formation and repress transcription. If this heterochromatin spreads into an adjacent gene, transcription of that adjacent gene may be decreased. **B:** 5′ and 3′ transduction. In 5′ transduction, an upstream endogenous promoter transcribes upstream sequence (shown as yellow) as well as L1 sequence. When this transcript is retrotransposed, a copy of the upstream sequence can be mobilized to a new site. In 3′ transduction, transcription proceeds past the L1 poly(A) signal into 3′ flanking sequence (shown as orange). After retrotransposition, a copy of the 3′ flanking sequence is mobilized to a new site. **C:** Transcriptional elongation. L1 insertions can attenuate transcription of endogenous genes by inserting into their introns. Longer insertions are predicted to have a greater inhibitory effect on transcriptional elongation. **D:** Premature polyadenylation. L1 insertions into endogenous genes can prematurely truncate transcription, producing a new mRNA isoforms. This can lead to translation into an intron (which has become the new 3′ untranslated region), generating a novel protein isoforms. **E:** Gene breaking. Full-length antisense L1 insertions into introns of endogenous genes can not only truncate transcription at cryptic L1 antisense poly(A) sites, but also generate a downstream transcript initiated at the L1 antisense promoter. This would produce two transcripts at the original gene, which could produce slightly altered protein products collectively similar to the original full-length protein.

transcripts generated from LTR-retrotransposons, typically originating from the LTR.[69] This shows the feasibility of such mechanisms with other types of retrotransposons.

Tissue-culture experiments have shown that when L1 retrotransposes, it can also co-transpose adjacent, non-L1 sequences to its new integration site. In this process, dubbed "3′ transduction", the presumed L1 polyadenylation signal is bypassed in favor of a downstream endogenous polyadenylation signal,[70] allowing downstream flanking host sequences to "come along for the ride" (Fig. 4B). Similarly, 5′ transduction occurs when a cellular promoter, by virtue of being upstream to the donor L1, transcribes both 5′ adjacent sequence and L1, which is then subjected to reverse transcription/integration (Fig. 4B).[9] These results have led to the proposal of a model whereby L1 can, on occasions, relocate useful non-L1 sequences such as promoters, enhancers, and even exons. This could, in turn, significantly alter the properties of the resulting allele. Many apparent natural examples of L1 3′ transduction have been found in the human genome,[71–73] establishing that this is not an artifact of the experimental system used. However, to date we are unaware of examples where an L1 5′ or 3′ transduced sequence has become an essential component of a cellular gene.

In the past year, we have seen that, as assayed by nuclear run-on, L1 sequences interfere with transcriptional elongation when fused to a reporter gene.[10] This effect appears to be orientation specific, occurring when the L1 sequences are transcribed in the sense orientation. Furthermore there is a length dependence to this phenomenon, A minimum of 1 kb of L1 sequence is required to observe a clear reduction of transcript levels, with longer sequences leading to ever-weaker RNA production.[10] Though these experiments were done on

plasmids, one can easily imagine that a similar effect might occur within L1 sequence copies naturally found within cellular genes. This is important because, as stated before, many L1 sequences are located within introns and RNA polymerase must elongate through these sequences even if they are to be spliced out later. If this transcriptional elongation effect does take place in the genome, L1 insertions in introns should attenuate levels of the target gene mRNA, without necessarily changing the mRNA sequence (Fig. 4C). The magnitude of this effect could range from subtle to large, depending on insert length (the ''molecular rheostat'' model, see reference 10)). Since most L1s in the genome are short (average around 1 kb) due to 5′ truncation, most of the effects are expected to be slight and difficult to perceive without careful examination. Even so, such slight effects can still be selected for or against over evolutionary time, and bioinformatic data suggest that this has happened over the course of human genome evolution, with weakly expressed genes containing, on average, much more L1 sequence than strongly expressed genes.[10] Furthermore, there are examples in the literature in which large ''sense strand'' L1 insertions in genes (the human retinitis pigmentosa 2 gene[74] and the mouse black-eyed white gene[75]) lead to diminished RNA levels as assayed by RT-PCR. No mechanistic studies were performed, so it is unclear how this occurs at the molecular level but, on the surface, these data strongly support the molecular rheostat model. Though intriguing, the bioinformatic data is merely correlative and therefore characterization of more allelic pairs differing only by the presence/absence of an insertion must be conducted to directly test the model. The effect may be context dependent and thus many alleles should be examined. If the molecular rheostat model is further validated, it would describe an important new way (in addition to promoter mutations, chromatin and epigentic effects) to produce alleles that lower expression levels.

L1 sequences also exist in the sequence of known proteins.[76] 64 known proteins were found to contain translated L1 sequences, and likewise 127 known proteins were shown to contain *Alu* sequences (*Alu* elements are a separate family of tiny retrotransposons that use the L1 machinery to retrotranspose[77–80]). Incorporation of these L1 sequences into cellular mRNA usually occurred via alternative splicing, which extended or truncated the coding region of the original protein.

In addition, insertion of L1 sequences can lead to translation of previously untranslated intron sequence, generating new protein isoforms (Fig. 4D). A well-characterized example of this exists for the human attractin gene, which codes for two differentially regulated isoforms and contains a short L1 insertion near the 3′ end.[81] The transmembrane mRNA isoform splices out the L1 sequence and thus contains all of the exons, including a membrane-spanning domain. In contrast, the mRNA encoding the soluble isoform of attractin reads into and terminates at the native poly(A) site of L1 sequence, and thus does not contain the membrane-spanning domain exons. Instead, the encoded protein extends into and terminates in intronic sequence, which has become the new 3′ untranslated region. This is a striking display of how the genetic diversity created by L1 can be exploited for cellular means. Related is another finding from plasmid-based assays that L1 contains various cryptic polyadenylation signals within its sequence, in both the sense and antisense directions.[10,82] If a new L1 integrant lands within a gene, this has the potential to produce a truncated RNA and thus a new mRNA isoform. Scouring EST databases has verified that truncated transcripts using these cryptic L1 poly(A) signals do in fact exist.[83] As outlined previously,[10] this could also lead to the production of new protein isoforms since, as with the soluble attractin gene, translation of the coding sequence of the last exon before truncation is expected to extend into the intron until it hits a fortuitous stop codon (Fig. 4D). If the insertion is close enough to the intron, L1 sequence may even be translated. It is of interest to note that different plasmid systems give different amounts of premature polyadenylation,[10,82] suggesting that the effects observed in chromosomes are likely context dependent.

Finally, as mentioned above, antisense promoter activity has previously been reported in the L1 5′UTR.[65,66] This raises the intriguing possibility that, although full-length or nearly full-length antisense L1 insertions in genes may generate a truncated transcript of the target gene via their cryptic polyadenylation signals, the remainder of the target gene transcript (downstream of the L1 insertion) may still be transcribed by the L1 antisense promoter (Fig. 4E). This potential ''gene-breaking'' represents a theoretical mechanism by which genes could be split into separable units, with new subtle mRNA and protein isoforms being generated. This may also be less deleterious than a simple truncation, since a simple truncation will result in untranscribed downstream sequences; in the case of gene-breaking, the downstream sequences will be transcribed (albeit on a separate transcript). Recent studies in our laboratory suggest that gene-breaking has occurred in the human and chimpanzee genomes.[83] Further work will focus on whether the corresponding ''split'' gene products are functional. A similar event has been shown to occur when an IAP retrotransposon integrated into the intron of the *Cabp* gene in mouse.[68] At this site, new transcripts upstream and downstream from the IAP insertion were generated. In this case, the upstream transcript truncates before entering the retroelement by an unknown mechanism.

L1 retrotransposition–an experiment in evolution?

Above we have detailed various proposed models in which the process of L1 retrotransposition may generate new alleles of

endogenous genes that are significantly different at the functional level, but not necessarily deleterious, as compared to their pre-insertion counterparts. Some of these L1-generated alleles have survived the test of time, and strongly suggest that L1, in addition to its own selfish agenda, may unknowingly be serving a useful function by creating a plethora of potential alleles that can be subjected to the forces of natural selection. These new creations can modulate both the quantity (expression variants) and quality (structurally distinct isoforms) of gene products. Although some experimental evidence for all of these models exists, not all have been explicitly demonstrated experimentally or by the discovery of natural examples. Part of this may be due to detection bias—while disease-causing insertions which, usually completely destroy gene function, are relatively easy to identify, subtle allelic variants with a slightly reduced expression level may not cause a clearly observable phenotype and will generally be missed unless a concerted effort is undertaken to find them. Efforts to find such alleles differing by expression level have been reported,[84,85] and efforts to coordinate these with the positions of retrotransposon insertions are now underway.

Due to the current lack of natural examples of functionally distinct, L1-generated alleles in mammalian genes, it is difficult to estimate how influential L1 insertions have been in determining what alleles we carry today. It is known that allelic variation in gene expression is common in the human genome.[84,85] The cause of this variation is a mystery, but it will be curious to see whether L1 insertions (via the molecular rheostat model) contribute significantly to this phenomenon. Answering this question will likely be a time-consuming task, requiring the structural characterization of many differentially expressed alleles. It has also been proposed that allele expression variability may explain human phenotypic variability.[84,85] If so, it may be the case that the L1 insertions scattered throughout our genome, most previously considered inconsequential to expression, actually help determine what makes us all so unique!

The examination of mammalian genomes has revealed remarkably similar gene complements.[1–5] However, different species of mammals are clearly distinct in terms of size, shape and behavior. How have these differences come about when the available sets of genes are essentially the same? The likely answer is that these genes are expressed at varying levels and as altered isoforms in different species. Certainly point mutations and genome rearrangements bear some of the responsibility for these changes. However, L1 is a mammalian retrotransposon family, and, as discussed at great length here, also has the potential to create these differences. Thus, we propose that L1 insertions may underlie at least some of the changes that have led to mammalian speciation. As more mammalian genome sequences and transcriptome profiles become available, it will be of great interest to see if some transcriptome discrepancies are correlated with L1 (or other transposon) insertions.

Conclusions

LINEs, of which L1 is the most recent lineage, are an ancient family of retroelements that have successfully reproduced in mammals for hundreds of millions of years. L1 has expanded to populate an enormous fraction of our genome and may be more than just "junk DNA". Research over the past few decades has provided insight into how L1 replicates and the dizzying array of mechanisms by which L1 can modify genome structure and function. Although L1 does not need to provide host benefit to survive, the variety of novel alleles that L1 insertions can generate make it worthwhile to contemplate whether a major component of our evolution is based on the fortuitous or semi-systematic selection of L1 mediated events.

References

1. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, et al. 2003. The dog genome: survey sequencing and comparative analysis. Science 301:1898–1903.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. Science 291:1304–1351.
4. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.
5. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493–521.
6. Smit AF. 1996. The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743–748.
7. Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, et al. 2002. Molecular archeology of L1 insertions in the human genome. Genome Biol 3:research0052.
8. Gilbert N, Lutz-Prigge S, Moran JV. 2002. Genomic deletions created upon LINE-1 retrotransposition. Cell 110:315–325.
9. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, et al. 2002. Human l1 retrotransposition is associated with genetic instability in vivo. Cell 110:327–338.
10. Han JS, Szak ST, Boeke JD. 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature 429:268–274.
11. Britten RJ, Kohne DE. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 161:529–540.
12. Shafit-Zagardo B, Maio JJ, Brown FL. 1982. KpnI families of long, interspersed repetitive DNAs in human and other primate genomes. Nucleic Acids Res 10:3175–193.
13. Meunier-Rotival M, Soriano P, Cuny G, Strauss F, Bernardi G. 1982. Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA. Proc Natl Acad Sci USA 79:355–359.
14. Singer MF. 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 28:433–434.
15. Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16:793–805.
16. Martin SL, Voliva CF, Burton FH, Edgell MH, Hutchison CA 3rd. 1984. A large interspersed repeat found in mouse DNA contains a long open reading frame that evolves as if it encodes a protein. Proc Natl Acad Sci USA 81:2308–2312.
17. Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB, et al. 1986. The sequence of a large L1Md element reveals a tandemly repeated 5′

end and several features found in retrotransposons. Mol Cell Biol 6:168–182.
18. Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, et al. 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics 1:113–125.
19. Kazazian HH Jr., Wong C, Youssoufian H, Scott AF, Phillips DG, et al. 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332:164–166.
20. Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr. 2003. Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 13: 651–658.
21. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, et al. 1996. High frequency retrotransposition in cultured mammalian cells. Cell 87: 917–927.
22. Swergold GD. 1990. Identification, characterization, and cell specificity of a human LINE- 1 promoter. Mol Cell Biol 10:6718–6729.
23. Paule MR, White RJ. 2000. Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res 28:1283–1298.
24. Kurose K, Hata K, Hattori M, Sakaki Y. 1995. 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.
25. Woodcock DM, Williamson MR, Doherty JP. 1996. A sensitive RNase protection assay to detect transcripts from potentially functional human endogenous L1 retrotransposons. Biochem Biophys Res Commun 222: 460–465.
26. Mizrokhi LJ, Georgieva SG, Ilyin YV. 1988. jockey, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell 54:685–691.
27. Tchenio T, Casella JF, Heidmann T. 2000. Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res 28:411–415.
28. Yang N, Zhang L, Zhang Y, Kazazian HH Jr. 2003. An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Res 31:4929–4940.
29. Athanikar JN, Badge RM, Moran JV. 2004. A YY1-binding site is required for accurate human LINE-1 transcription initiation. Nucleic Acids Res 32:3846–3855.
30. Branciforte D, Martin SL. 1994. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol Cell Biol 14:2584–2592.
31. Trelogan SA, Martin SL. 1995. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. Proc Natl Acad Sci USA 92:1520–1524.
32. Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, et al. 2002. A mouse model of human L1 retrotransposition. Nat Genet 32: 655–660.
33. Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, et al. 2004. Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. J Biol Chem 279:27753–27763.
34. Martin SL, Bushman FD. 2001. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE- 1 retrotransposon. Mol Cell Biol 21:467–475.
35. Kolosha VO, Martin SL. 2003. High-affinity, Non-sequence-specific RNA Binding by the Open Reading Frame 1 (ORF1) Protein from Long Interspersed Nuclear Element 1 (LINE-1). J Biol Chem 278:8112–8117.
36. Feng Q, Moran JV, Kazazian HH Jr., Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905–916.
37. Mathias SL, Scott AF, Kazazian HH Jr., Boeke JD, Gabriel A. 1991. Reverse transcriptase encoded by a human transposable element. Science 254:1808–1810.
38. Esnault C, Maestre J, Heidmann T. 2000. Human LINE retrotransposons generate processed pseudogenes. Nat Genet 24:363–367.
39. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, et al. 2001. Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol 21:1429–1439.
40. Martin SL. 1991. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol Cell Biol 11:4804–4807.
41. Goodier JL, Ostertag EM, Engleka KA, Seleme MC, Kazazian HH Jr. 2004. A potential role for the nucleolus in L1 retrotransposition. Hum Mol Genet 13:1041–1048.
42. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 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.
43. Yang J, Malik HS, Eickbush TH. 1999. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. Proc Natl Acad Sci USA 96:7847–7852.
44. Cost GJ, Feng Q, Jacquier A, Boeke JD. 2002. Human L1 element target-primed reverse transcription in vitro. Embo J 21:5899–5910.
45. Jensen S, Gassama MP, Heidmann T. 1994. Retrotransposition of the Drosophila LINE I element can induce deletion in the target DNA: a simple model also accounting for the variability of the normally observed target site duplications. Biochem Biophys Res Commun 202:111–119.
46. Boissinot S, Chevret P, Furano AV. 2000. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol Biol Evol 17: 915–928.
47. Moran JV, Gilbert N. 2002. Mammalian LINE-1 retrotransposons and related elements. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, D.C.: ASM Press. p 836–869.
48. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, et al. 2003. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci USA 100:5280–5285.
49. Burwinkel B, Kilimann MW. 1998. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. J Mol Biol 277:513–517.
50. Narita N, Nishio H, Kitoh Y, Ishikawa Y, Minami R, et al. 1993. 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.
51. Mulhardt C, Fischer M, Gass P, Simon-Chazottes D, Guenet JL, et al. 1994. The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intronic insertion of L1 element. Neuron 13: 1003–1015.
52. Takahara T, Ohsumi T, Kuromitsu J, Shibata K, Sasaki N, et al. 1996. 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.
53. Kohrman DC, Harris JB, Meisler MH. 1996. 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.
54. Meischl C, Boer M, Ahlin A, Roos D. 2000. A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. Eur J Hum Genet 8:697–703.
55. Hickey DA. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics 101:519–531.
56. Hickey DA. 1992. Evolutionary dynamics of transposable elements in prokaryotes and eukaryotes. Genetica 86:269–274.
57. Bestor TH. 1999. Sex brings transposons and genomes into conflict. Genetica 107:289–225.
58. Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431:96–99.
59. Yoder JA, Walsh CP, Bestor TH. 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–340.
60. Biessmann H, Mason JM, Ferry K, d'Hulst M, Valgeirsdottir K, et al. 1990. Addition of telomere-associated HeT DNA sequences ''heals'' broken chromosome ends in Drosophila. Cell 61:663–673.
61. Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen FM. 1993. Transposons in place of telomeric repeats at a Drosophila telomere. Cell 75: 1083–1093.
62. Sheen FM, Levis RW. 1994. Transposition of the LINE-like retrotransposon TART to Drosophila chromosome termini. Proc Natl Acad Sci USA 91:12510–12514.
63. Agrawal A, Eastman QM, Schatz DG. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394:744–751.
64. Zhou L, Mitra R, Atkinson PW, Burgess Hickmann A, Dyda F, et al. 2004. Transposition of *h*AT elements links transposable elements and V(D)J recombination. Nature 432:995–1001.

65. Nigumann P, Redik K, Matlik K, Speek M. 2002. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 79:628–634.
66. Speek M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol 21:1973–1985.
67. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 23:314–318.
68. Druker R, Bruxner TJ, Lehrbach NJ, Whitelaw E. 2004. Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. Nucleic Acids Res 32:5800–5808.
69. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, et al. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 7:597–606.
70. Moran JV, DeBerardinis RJ, Kazazian HH Jr. 1999. Exon shuffling by L1 retrotransposition. Science 283:1530–1534.
71. Goodier JL, Ostertag EM, Kazazian HH Jr. 2000. Transduction of 3′-flanking sequences is common in L1 retrotransposition. Hum Mol Genet 9:653–657.
72. Pickeral OK, Makalowski W, Boguski MS, Boeke JD. 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res 10:411–415.
73. Szak ST, Pickeral OK, Landsman D, Boeke JD. 2003. Identifying related L1 retrotransposons by analyzing 3′ transduced sequences. Genome Biol 4:R30.
74. Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, et al. 1998. Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nat Genet 19:327–332.
75. Yajima I, Sato S, Kimura T, Yasumoto K, Shibahara S, et al. 1999. An L1 element intronic insertion in the black-eyed white (Mitf[mi-bw]) gene: the loss of a single Mitf isoform responsible for the pigmentary defect and inner ear deafness. Hum Mol Genet 8:1431–1441.
76. Li WH, Gu Z, Wang H, Nekrutenko A. 2001. Evolutionary analyses of the human genome. Nature 409:847–849.
77. Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379.
78. Hagan CR, Sheffield RF, Rudin CM. 2003. Human Alu element retrotransposition induced by genotoxic stress. Nat Genet 35:219–220.
79. Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 35:41–48.
80. Boeke JD. 1997. LINEs and Alus–the polyA connection. Nat Genet 16: 6–7.
81. Tang W, Gunn TM, McLaughlin DF, Barsh GS, Schlossman SF, et al. 2000. Secreted and membrane attractin result from alternative splicing of the human ATRN gene. Proc Natl Acad Sci USA 97:6025–6030.
82. Perepelitsa-Belancio V, Deininger P. 2003. RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet 35:363–366.
83. Wheelan S, Aizawa Y, Han JS, Boeke JD. 2005. Submitted.
84. Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. 2003. Allelic variation in gene expression is common in the human genome. Genome Res 13:1855–1862.
85. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. 2002. Allelic variation in human gene expression. Science 297:1143.