MOBILE ELEMENTS AND THE HUMAN GENOME

Eline T. Luning Prak and Haig H. Kazazian Jr

Genomic DNA is often thought of as the stable template of heredity, largely dormant and unchanging, apart from perhaps the occasional point mutation. But it has become increasingly clear that DNA is dynamic rather than static, being subjected to rearrangements, insertions and deletions. Much of this plasticity can be attributed to transposable elements and their genomic relatives.

RETROTRANSPOSITION
The process of mobilizing a retrotransposon. It involves transcription, processing of the RNA, translation, reverse transcription of the transposon RNA, and integration of the reverse transcribed DNA into a new genomic location.

LONG TERMINAL REPEATS (LTRs)
Sequences of 300–1,000 base pairs that are directly repeated at the 5′ and 3′ ends of LTR retrotransposons and retroviruses. Because of the mechanism of reverse transcription, the 5′ and 3′ LTRs of an element or retrovirus are identical on insertion into new genomic sites.

V(D)J RECOMBINATION
Site-directed DNA recombination carried out by RAG1 and RAG2 to create an enormous diversity of immunoglobulins and T-cell receptors.

Department of Genetics, 475 Clinical Research Building, 415 Curie Boulevard, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. e-mail: kazazian@mail.med.upenn.edu

Transposable elements are repetitive mobile sequences that are dispersed throughout the genome. There are two broad classes of transposable element: DNA transposons and RETROTRANSPOSONS. Most DNA transposons move as pieces of DNA, cutting and pasting themselves into new genomic locations. In contrast, the genomic expansion of retrotransposons occurs by a replicative mechanism. Retrotransposons are duplicated through an RNA intermediate: the original transposon is maintained *in situ*, where it is transcribed; its RNA transcript is then reverse transcribed into DNA, which integrates into a new genomic location. There are two broad classes of retrotransposon: LONG TERMINAL REPEAT (LTR) and non-LTR retrotransposons. LTR retrotransposons are analogous to retroviruses, except that they cannot move from cell to cell. Non-LTR retrotransposons comprise a significant fraction (15–17%) of human genomic DNA[1].

Transposable elements can also be classified according to their degree of mechanistic self-sufficiency. Autonomous transposable elements, such as long interspersed nuclear elements (LINE-1 or L1 retrotransposons), are thought to encode essentially all of the machinery that they require to move. In contrast, non-autonomous transposable elements, such as short interspersed nuclear elements (SINEs), are entirely dependent on other transposable elements for their mobility. In some cases, the distinction between autonomous and non-autonomous elements becomes blurred. For example, it is unknown whether human endogenous retroviruses (HERVs), if mobilized, require the help of other elements or viruses *in trans*. Even 'autonomous' elements probably rely on the cellular machinery of their hosts.

Mobile elements, although sometimes harmful, have in all likelihood offered benefits to the human genome in the area of host defence (V(D)J RECOMBINATION), and may be involved in chromosomal structure/remodelling (TELOMERASE and, possibly, X-chromosome inactivation), as well as affording increased genome plasticity (transduction of 3′ flanking sequences) and facilitating gene expression (spliceosomal introns).

This review is divided into three parts. The first part describes the structure of transposable elements in the human genome and the mechanisms by which they move. The second part focuses on the consequences of transposable element dispersal and how genomes and transposable elements interact during the course of evolution. The review concludes with a brief section on harnessing transposons for biomedical applications.

Part I: Transposon structures and mechanisms

Autonomous transposable elements

L1 elements. L1 (non-LTR retrotransposon) elements are the most active autonomous transposable elements in the present-day human genome. Roughly 500,000 L1s reside in the human genome, most of which are 5′ TRUNCATED or rearranged[1]. Three to four thousand human L1s are full length (FIG. 1A), and of these an estimated 40–60 are active[2]. In contrast, the mouse genome has an estimated 3,000 active L1 elements[3]. Almost all the active human L1 elements that have been isolated so far belong to the TA SUBSET, a small family of full-length, transcribed L1 elements[4].

The mechanism of L1 retrotransposition is analo-

© 2000 Macmillan Magazines Ltd

TELOMERASE
A ribonucleoprotein enzyme with significant sequence similarity to reverse transcriptase of non-LTR retrotransposons. Telomerase maintains the ends of chromosomes and is expressed at increased levels in some transformed cells.

5′ TRUNCATION
Loss of the 5′ end of a retrotransposon (very common with L1 retrotransposition) upon insertion into a new genomic site.

TA SUBSET
A minor class of transcribed human L1 retrotransposons that has given rise to most cloned *de novo* L1 insertions.

REVERSE TRANSCRIPTASE
Enzyme used by retroviruses and retrotransposons to synthesize DNA from an RNA template.

GROUP II INTRON
A self-splicing intron of bacteria, mitochondria and chloroplasts that can excise from a specific site in a gene and insert into a specific site in another gene using activities encoded within its DNA.

INTRACISTERNAL A PARTICLE (IAP)
An LTR-containing retrotransposon of mice that resembles a retrovirus but has a defective *env* gene.

gous to the target-primed reverse transcription (TPRT) mechanism (FIG. 2) (BOX 1) established for the insect R2Bm element[5]. Unlike the endonuclease of the R2Bm element, the L1 endonuclease is not highly sequence-specific[6], but seems to preferentially recognize a DNA junction between pyrimidines and purines (dTn–dAn)[7–9]. This junction may have unusual secondary structure[10]. The L1 endonuclease consensus sequence seems to be slightly over-represented in non-coding regions of DNA[10]. Indeed, L1 insertions are frequently found in A+T-rich DNA or are recovered from within other genomic L1 insertions (which are themselves A+T rich)[1]. As coding sequences tend to be more G+C-rich, L1 elements, by preferring A+T-rich DNA, may be more successful genomic parasites, because they can reach higher copy numbers without causing too many deleterious insertions. However, L1 elements can and do insert into genes[11,12].

The L1 REVERSE TRANSCRIPTASE (RT) contains seven conserved sequences that are shared among the RTs of all retrotransposons and retroviruses[13]. These shared regions have been used to trace the genealogy of L1 elements and of other mobile elements. The L1 RT sequence and its mechanism (TPRT) are similar to those of GROUP II INTRONS found in bacteria, chloroplasts and mitochondria[14–17].

LTR retrotransposons. So far, no active LTR retrotransposon has been isolated in humans. However, a small fraction of several types of mouse LTR retrotransposon may be active. For example, an intact INTRACISTERNAL A PARTICLE (IAP) transgene that was tagged with a marker gene has been mobilized at low frequencies in cultured cells[18]. Ty1/3 elements in yeast and the *copia* retrotransposons of *Drosophila melanogaster* are examples of LTR retrotransposons in other species. A retrovirus that loses

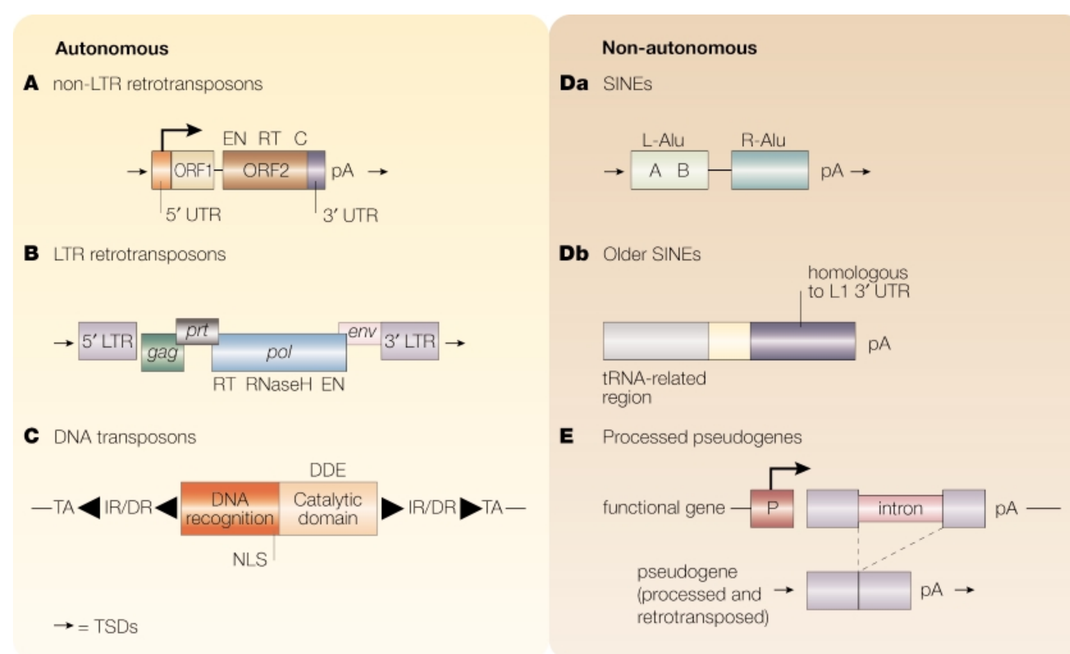


Figure 1 | **Structures of transposable element types. A** | A full-length (~6.1 kilobase) L1 element consists of a 5′ untranslated region (5′-UTR) that has promoter activity; two open reading frames (ORF1 and ORF2) that are separated by an intergenic spacer; followed by a 3′-UTR and a poly A tail. ORF1 encodes a 40 kDa protein that binds nucleic acids[122,123], whereas ORF2 contains reverse transcriptase (RT), an endonuclease (EN) domain and a cysteine-rich region (C). Genomic L1 insertions are often flanked by 7–20-nucleotide target-site duplications (TSDs), denoted here by arrows[8]. **B** | LTR retrotransposons, such as human endogenous retrovirus (HERV) proviral sequences, consist of partly overlapping coding regions for group-specific antigen (*gag*), protease (*prt*), polymerase (*pol*) and envelope (*env*) genes, flanked on both ends by long terminal repeats (LTRs) with promoter activity, and by TSDs (arrows). The *pol* gene contains domains for reverse transcriptase (RT), RNaseH and endonuclease (EN). **C** | The *Sleeping Beauty* DNA transposon is shown here. This transposon consists of a single 1020-base-pair ORF that encodes the transposase. The transposase has a DNA-binding domain, a nuclear localization signal (NLS) and a magnesium-dependent catalytic core with a conserved DDE motif that resembles a retroviral integrase[36,43]. Genomic insertions of Tc1/mariner elements are flanked by short terminal inverted repeats (IRs). These IRs can be less than 100 nucleotides and harbour a single transposon-binding site (for Tc1 or for mariner elements), or they can be more complex, with several binding sites and/or direct repeats (DR) that flank the IRs (IR/DR sequences), as shown here[36]. **Da** | SINES, such as Alu elements in humans and B1 elements in mice, are 300-nucleotide intronless sequences that consist of two G+C-rich fragments, the left (L-Alu) and the right (R-Alu) monomer, connected by an A-rich linker, and that end in a poly A tail (pA). The element is flanked by short direct repeats (arrows), which are duplications of the genomic insertion site. The element is transcribed from an internal RNA polymerase III promoter, but has no coding capacity. **Db** | Older SINEs consist of a transfer RNA-related region, a non-tRNA region, and a poly A tail. The 3′-tail region of some tRNA-derived SINEs resembles a portion of the 3′ UTR of LINEs[50]. **E** | Processed pseudogenes are non-functional coding sequences that lack introns and promoters, usually have poly A tails, and are flanked by target-site duplications[54]. They seem to be formed by retrotransposition. A functional or non-functional gene is transcribed from its promoter (P), processed, reverse transcribed and inserted into a new genomic location, where it may undergo further mutation.

© 2000 Macmillan Magazines Ltd

TRANSLATIONAL FRAMESHIFTING
Some viruses and transposable elements have sequences that can be read in more than one frame by the host cell ribosomes. Specific RNA sequences promote ribosomal slippage, facilitating the shift to a different reading frame.

PRIMER-BINDING SITE
Retroviral reverse transcriptase uses a transfer RNA primer to initiate DNA synthesis. The binding site for this tRNA is located near the 5′ end of the retroviral genome.

HOMINOIDS
A superfamily of the order Primates, comprising man and the anthropoid apes.

the function of its envelope (*env*) gene can become an LTR retrotransposon. For example, IAP elements have similar sequences to type D retroviruses and, similar to retroviruses, can express protease and polymerase proteins by TRANSLATIONAL FRAMESHIFTING[19]. However, unlike retroviruses, IAP elements no longer have functional *env* genes (because their *env* genes are heavily mutated). Conversely, an LTR retrotransposon that acquires an *env* gene could become a retrovirus. Some *gypsy* elements in *Drosophila* are probable examples of this latter pathway of a retrotransposon becoming a retrovirus.

In humans, the closest elements to LTR retrotransposons are endogenous retroviruses (HERVs). HERVs make up roughly 7% of the human genome (FIG. 1B)[1]. Most HERV sequences have accumulated several nonsense mutations and no longer encode functional retroviral proteins[20]. The HERV family that is most likely to be functional is a subfamily of elements named HERV-K (HML-2). Members of the HERV-K family are so named because they have a PRIMER-BINDING SITE for Lys-transfer RNA. There are 30–50 proviral copies of this family within the haploid human genome[21]. Retrovirus-like particles, typically found in germ-cell tumour-derived cell lines, are encoded by HERV-K, and patients with testicular cancer typically have high antibody titres against HERV-K, Gag and Env proteins[22]. Genes encoding intact enzymes from HERV-K subfamily members, such as protease, endonuclease, reverse transcriptase and RNase H, have been cloned from the human genome[23–25].

Although an almost intact HERV is present in humans, a mobile HERV element has not yet been identified[26]. Nevertheless, both the presence of human-specific HERV proviruses and their absence in corresponding genomic sites (empty sites) in other HOMINOIDS, indicate that a subset of these elements has been mobile during the evolution of hominoids[26–28]. Furthermore, there are thousands of 'solitary' LTRs, the result of homologous recombination events within proviruses, which are dispersed throughout the human genome (FIG. 3e). Some LTRs are unique to humans, again indicating that HERVs are or have been mobile recently[29].

How HERV elements are (or were) mobilized is unclear. Given their similarity to LTR retrotransposons from other species, the most probable mechanism is retrotransposition. In this model, full-length HERV transcripts are reverse transcribed in the cytoplasm in a retroviral fashion; the resulting double-stranded DNA returns to the nucleus and is integrated into chromosomal DNA. These events presumably parallel those of a retroviral replication cycle, but lack an extracellular phase.

Critical enzymes for some replication steps may be provided either by the particular retrotransposing HERV family itself, or *in trans* by other HERV elements. However, it is also possible that related exogenous retroviruses could provide such enzymatic activities. Indeed, a defective Moloney Murine Leukaemia Virus (MMLV) retrotransposes at low frequency in human cells when MMLV Gag and Pol proteins are provided *in trans* within the same cell[30]. In addition to mobilization within cells, HERV insertions could also acquire the capacity to move between cells by the packaging of HERV transcripts together into retroviral particles encoded by (temporarily present) exogenous retroviruses. Although no example has been found in humans, this type of cell-to-cell transfer followed by re-integration does occur with non-coding VL30 elements in the mouse, which package together into MMLV-derived retrovirus particles and complete the retrotransposition process in a newly infected cell[20,31].

Trans-mobilization may also apply to an interesting LTR transposon from mice called Etn (for early transposon element). Insertions of this element have caused several genetic diseases in laboratory strains of mice[32]. Etn sequences seem to be derived from type D retroviruses, called MusD[33]. Active Etn elements have very

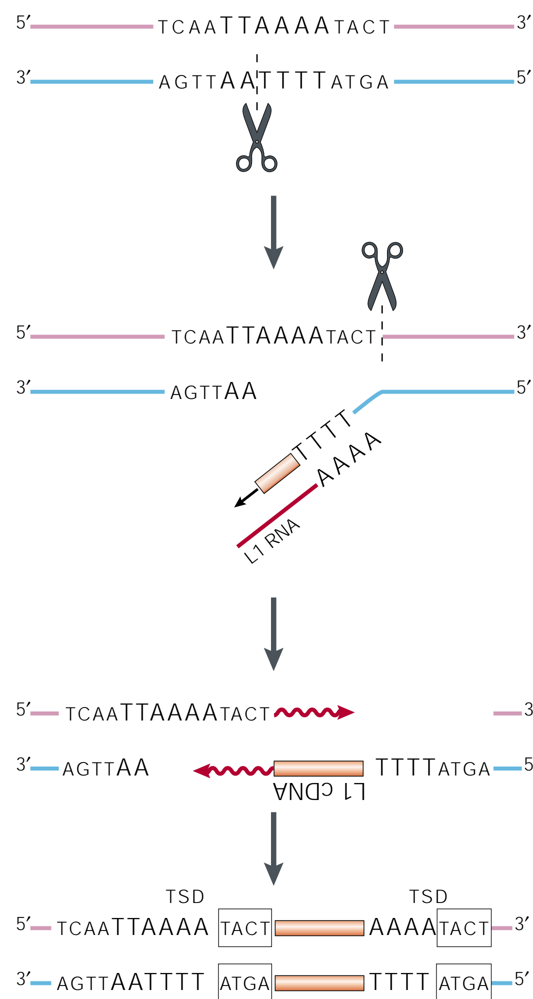
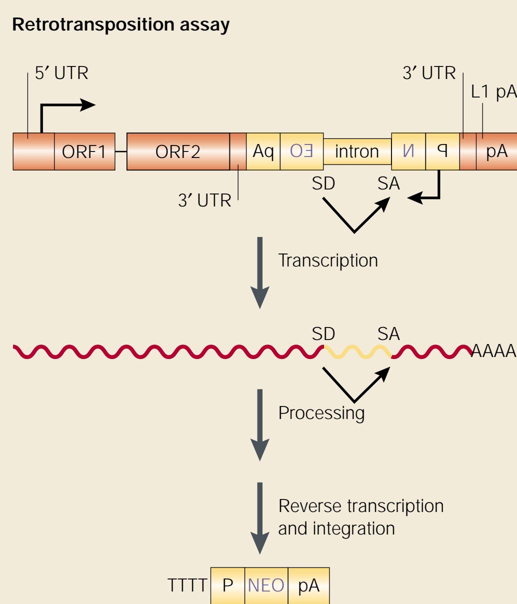


Figure 2 | **Target-primed reverse transcription (TPRT).** The L1 endonuclease cleaves between an A and a T on the bottom strand of DNA, exposing a 3′ hydroxy group. The target DNA functions as a primer from which complementary DNA is made (using the polyadenylated L1 RNA as a template). The second cleavage (on the sense strand of the chromosomal DNA) is staggered relative to the first, and repair or filling in of the gaps between the L1 cDNA and the chromosomal DNA results in a short target-site duplication (TACT, boxed).

© 2000 Macmillan Magazines Ltd

Box 1 | **Cell culture-based retrotransposition assay**



Retrotransposons can be tagged with intron-containing marker genes (yellow) that, when reverse transcribed, can move into various genomic locations[18,118]. The full-length transcript of the L1 element (red) includes the antisense marker gene (*NEO*), which is inserted into the L1 3′ UTR. When this L1–*NEO* hybrid transcript is processed, the intron, which interrupts the marker gene *NEO*, is spliced out. Reverse transcription of the processed transcript and integration of its DNA copy are coupled in the TPRT reaction. So the marker gene can only be activated if it has undergone retrotransposition. SD and SA are the splice donor and splice acceptor sites, respectively. Retrotransposed products were shown to resemble genomic L1 insertions, showing variable 5′ truncation, ending in poly A tracts and having target-site duplications[118].

This cell-culture-based retrotransposition assay has been used to compare the activities of different L1 retrotransposons and to map important functional domains within the L1 ORFs by mutagenesis. Mutations in some conserved regions (such as those encoding the Y(or F)ADD box in the fifth conserved segment) can abolish reverse transcriptase activity in yeast, as well as eliminate retrotransposition activity in cultured cells[118–121]. The cell culture assay also lends support to the hypothesis that proteins of an L1 element act preferentially on the RNA from that same element[118].

similar LTR repeats, as well as similar 5′- and 3′-internal sequences, to MusD, but lack the MusD *gag*, *pol* and *env* genes[33]. It is likely that Etn elements are mobilized by MusD proteins *in trans*.

Tc1/mariner transposon family. DNA transposon sequences constitute roughly 2.4–2.7% of the human genome[1,34]. Most DNA transposon copies in the human genome are remains of the hobo-Activator-Tam3-related DNA transposons (most of which are members of the human medium reiterated frequency repeat family, known as MER1)[35]. Another 40% of DNA transposon copies in human genomic DNA are members of the MER2 class, which includes the Tigger elements, so named because they resemble *pogo* elements (and in reference to one of the characters in Winnie the Pooh)[35]. Mariner elements, which represent a subset of the Tc1-mariner group of DNA transposons, comprise less than 3% of the DNA transposon copies in the human genome[1]. Tc1 from *Caenorhabditis elegans* and mariner from *Drosophila mauritania* are members of the Tc1/mariner superfamily of DNA transposons (FIG. 1C)[36].

So far, no active Tc1/mariner element (or any other DNA transposon for that matter) has been isolated from the human genome. However, by scanning the human genome database for inverted repeat (IR) sequences, mariner-like elements can be identified from the *irritans* subfamily (which is present in insects) and from the *cecropia* subfamily (discovered in the *Cecropia* moth), indicating that several horizontal transfer events of these elements have occurred in the past[37].

Attention has focused on Tc1/mariner superfamily members because of the surprising demonstrations that an engineered Tc1-like element (called *Sleeping Beauty*)[38], and more natural versions of Tc1/mariner transposons from other organisms, can be mobilized across species[39,40]. The broad host range of Tc1/mariner-like elements indicates a minimal need for *trans*-acting factors or a reliance on evolutionarily conserved cellular machinery.

The mechanism of Tc1/mariner transposition is one of excision and integration, which probably resembles that of many other DNA transposons, such as P-elements in *Drosophila melanogaster*, Tam1/Tam3 in plants and Tn*10* in bacteria[41]. This mechanism involves a series of hydrolysis and transesterification reactions (FIG. 4) and is strikingly similar to V(D)J recombination[42]. Interestingly, the RECOMBINASE-ACTIVATING GENE *RAG1*, one of the enzymes that carries out V(D)J recombination, may have one or more DDE motifs[36,43].

RECOMBINASE ACTIVATING GENE (*RAG*)
Encodes RAG proteins that facilitate the DNA rearrangements that result in the assembly of immunoglobulin and T-cell receptor genes from various gene segments.

Non-autonomous transposable elements

Alu elements. Some transposable elements are clearly mobilized *in trans*. Alu elements are SINEs that consist of a 280 nucleotide coding sequence that lacks introns (FIG. 1D)[44]. Alu sequences, of which there are about 1,000,000 in human genomic DNA[44], are transcribed but are not translated because they lack an ORF. As such, they are referred to as the 'parasite's parasite' because they rely entirely on other mobile DNA elements for their mobility[45].

L1 elements have been proposed to be the main gen-

© 2000 Macmillan Magazines Ltd

erators of Alu and of processed pseudogene retrotranspositions. L1s are thought to mobilize Alus because of the similarity of their target-site duplications, the similarity of their insertion sites (the DNA nick for Alu insertions is probably made by L1 endonuclease), and the similarity of their RNA expression patterns in the male germ line[44,46].

Alu elements may have always relied on L1 elements for their mobility. Phylogenetic analysis indicates that non-LTR retrotransposons, such as L1, originated at the inception of eukaryotic life[13,47]. In contrast, Alu elements are relative newcomers, originating perhaps at the outset of mammalian radiation[48]. Nearly all SINEs (except for Alu and the rodent B1 elements) are related to tRNA[44]. Interestingly, it has been found that the 3′ ends of several species of tRNA-derived SINEs are identical to some of the LINEs found in the same species[49,50]. So the parasitism of LINEs by SINEs could extend to the evolutionary borrowing of sequences. However, there is no significant sequence homology between the most abundant SINEs in present-day humans, primates and rodents, and their corresponding LINEs.

The parasitism of LINEs by SINEs remains difficult to reconcile with the observation that LINEs seem to insert preferentially into A+T-rich regions, whereas SINEs, such as Alus, accumulate in G+C-rich regions[1]. One theory is that Alu elements integrate randomly but those that are actively transcribed (and are therefore more likely to reside in G+C-rich regions of the genome) are more likely to become fixed in the population[1]. This explanation in turn predicts that Alu RNA might have some advantageous function. It has been proposed that Alu-mediated inhibition of PRKR (interferon-inducible double-stranded (ds)RNA-dependent protein kinases) transiently derepresses translation, facilitating the rapid production of stress proteins[51].

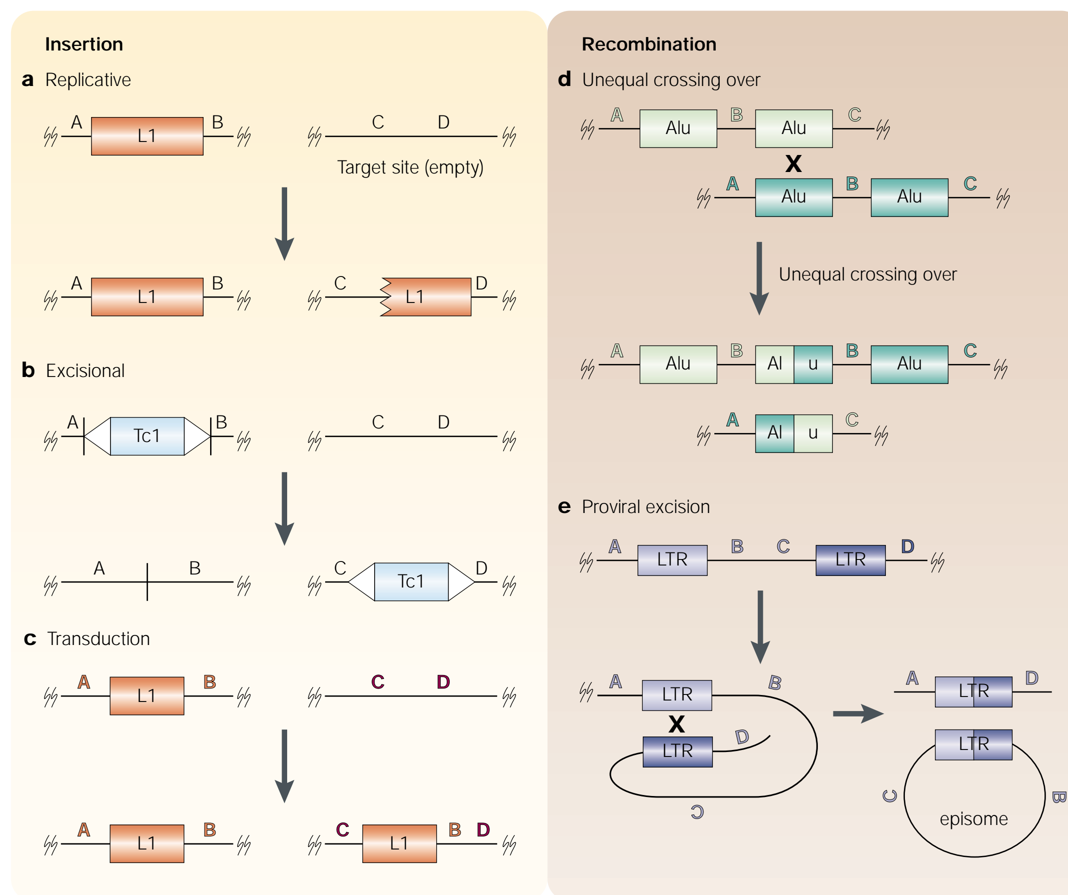


Figure 3 | **Transposon genome-shaping pathways. a** | Replicative insertion. The replicative insertion pathway, used by retrotransposons such as L1, results in a net increase in the copy number of the element (from one to two). A copy of the element is inserted into a second genomic site (the target site is between C and D). The inserted copy is usually 5′ truncated (so the L1 box between C and D is smaller than the original L1 element between A and B). **b** | Excisional insertion. The excisional pathway, used by DNA transposons such as *Sleeping Beauty*, results in a transfer of the transposon from one genomic site (between A and B) to another genomic site (between C and D). **c** | Transduction. L1 elements can sometimes carry downstream flanking sequences (region B) with them when they retrotranspose. **d** | Unequal crossing-over. Alu elements may promote homologous-recombination events because they provide short regions of homology at fairly frequent intervals (Alu elements are about 3 kilobases apart, on average). Unequal crossing over between Alu elements on homologous chromosomes can result in heritable duplications and deletions of the intervening region. **e** | Proviral excision. Long terminal repeats (LTRs) of HERVs can also participate in recombination reactions that generate solitary LTRs. Proviral excision results in extrusion of the provirus (region B and C and one of the two LTRs on the episome) and the retention of a solitary LTR flanked by sequences A and D.

PROCESSED PSEUDOGENES
Copies of the coding sequences of genes that lack promoters and introns. They contain poly A tails and are flanked by target-site duplications.

© 2000 Macmillan Magazines Ltd

A second problem for L1s in serving as the mobilizers of Alus is that the proteins of an L1 element seem to act most efficiently on the RNA from that same element (*cis* preference)[11]. But there are more Alus than L1s in the human genome, so how were Alus able to out-compete L1s for their own *cis*-acting machinery? Alus consist of two monomers with homology to signal-recognition particle (SRP) RNA[44]. Perhaps Alu RNA concentrates on large ribosomal subunits because it is bound by SRP 9/14 proteins, two signal recognition particle proteins that interact with ribosomes[52,53]. A high concentration of Alu RNA near the large ribosomal subunit may allow Alu to hijack the L1 retrotransposition machinery[52].

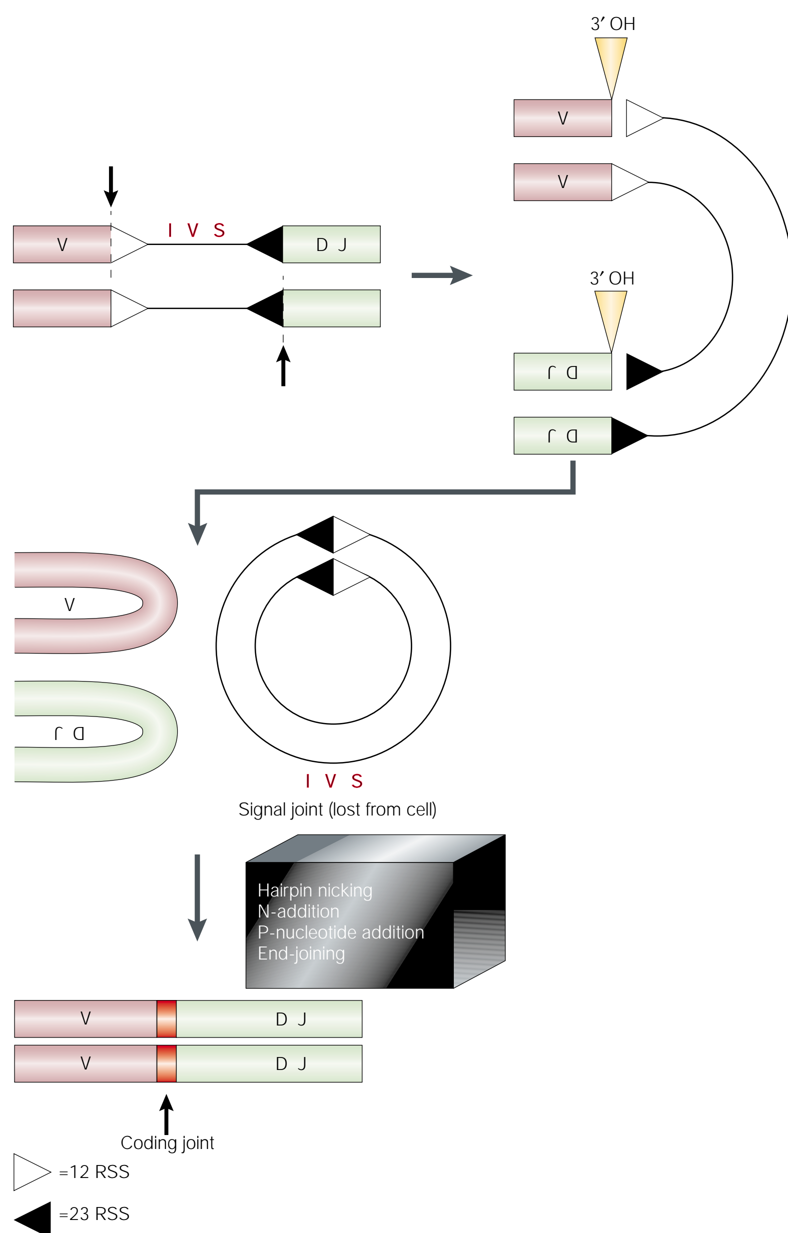


Figure 4 | **V(D)J recombination.** Immunoglobulin gene rearrangement or V(D)J recombination, similar to Tn*10* transposition, occurs by a series of hydrolysis and transesterification reactions[41]. The deletional rearrangement that recombines an upstream variable region gene (V) with a downstream fused diversity (D) and joining (J) segment (V to DJ rearrangement) is shown. First, the recombinase (Rag1/2) nicks the DNA at the recombination signal sequences (RSSs), generating 3′ hydroxy groups (3′ OH). These 3′ OH groups act as nucleophiles and attack the opposite strand, forming hairpin loops and double-stranded breaks. Then the hairpins are nicked and the resulting 3′ OH nucleophiles attack the target site. Following processing (N-addition, P-nucleotide addition and end joining), the coding joint is formed. The intervening sequence (IVS), in the DNA fragment that contains the precisely fused signal sequence, is liberated as an episome and eventually lost. The filled triangles correspond to 23-base-pair RSSs and the open triangles correspond to 12-base-pair RSSs, which mark the 3′ and 5′ ends of the IVS, respectively. Both strands of DNA are shown.

Processed pseudogenes. Similar to Alus, PROCESSED PSEUDOGENES are non-autonomous, transposed sequences. Processed pseudogenes are non-functional genes that have been moved from their native location through an RNA intermediate (FIG. 1E)[54]. On the basis of sequence data for human chromosome 22, processed pseudogenes are estimated to constitute about 0.5% of the human genome[55].

As with Alus, L1 elements are believed to mobilize processed pseudogenes. L1 elements can mobilize pseudogenes, albeit inefficiently in cultured cells[56]. However, the relevance of these findings to physiological L1 function remains unclear. L1s may possess the machinery to move mRNAs in a sensitive episomal cell-culture assay, but this does not establish their dominant role in this process *in vivo*.

Part II: Consequences of movement

Mobile elements influence genomic structure and function through several mechanisms, including insertion, transduction of 3′ flanking sequences, and recombination.

Insertions. The first active human L1 element was cloned by analysing the *de novo* insertion of an L1 element into the *factor VIII* gene of a haemophilia A patient[57,58]. L1 insertions have turned up in several other genetic disorders, including Duchenne muscular dystrophy, type 2 retinitis pigmentosa, β-thalassaemia and chronic granulomatous disease[59–64]. L1 elements can also insert somatically, as shown by the presence of an L1 insertion in the adenomatous polyposis coli (APC) gene of adenocarcinoma cells from a colon cancer patient, but not in the surrounding normal tissue[65].

It is perhaps not surprising that several of the genetic disorders in which L1 elements have been involved are X-chromosome linked. This may reflect the greater ease of detecting allelic loss in males. Alternatively, or in addition, L1 elements may inherently prefer the X chromosome. The density of L1 elements on the X chromosome is roughly twice that on autosomes (26% of total available sequence versus 13%)[1,66]. L1 elements may be attracted to the X chromosome because of its high A+T content. On the other hand, L1s themselves have a high A+T content, so this ‘preference’ may be a consequence rather than a cause of L1 insertions. An interesting model proposed by Lyon is that L1 elements serve as booster stations, helping to propagate the signal that inactivates one of the two female X chromosomes[67]. L1s are, in fact, concentrated in the Xq13 region, which contains

© 2000 Macmillan Magazines Ltd

the X-chromosome-inactivation centre[66]. L1s are also concentrated around inactivated genes and are in short supply in regions of the X chromosome that escape inactivation[66].

Chromosomal predilections aside, L1 insertions seem to be a relatively infrequent cause of genetic disease in man[68]. This is in contrast to the situation in mice, in which retrotransposable elements, particularly defective LTR retrotransposons (such as IAPs, Etns and mammalian apparent LTR retrotransposons), contribute to roughly 10% of all characterized spontaneous mutations[11].

Alu insertions have probably altered the expression of a larger number of human genes than have L1 elements[45]. In addition to disrupting exons, Alus and L1s can insert into non-coding regions and modify gene regulation. For example, some Alu insertions contain transcriptional regulatory sequences, such as a retinoic-acid-response element[69]. There is an interesting example of an Alu element residing in the intron of the human Wilms tumour suppressor gene. Here the element acts as a transcriptional silencer[70], whereas other Alu elements, along with other transposable elements, have been implicated as enhancers or insulators[71,72]. In a similar vein, solitary LTRs derived from HERV-Ks may function as bidirectional promoters[73].

L1-mediated transduction. In addition to duplicating themselves, L1s can carry with them genomic flanking sequences that are downstream of their 3′ UTRs (FIG. 3c). The idea of transposable element insertions bringing new functions to genes was taken one step further when Holmes *et al.* realized the importance of a 3′ transduction event seen in an L1 insertion into the human *dystrophin* gene[59]. Moran *et al.* confirmed 3′ transduction experimentally by placing the intron-containing antisense neomycin resistance gene downstream of the L1 polyadenylation signal[12]. L1s were able to retrotranspose sequences from their 3′ flanks to new genomic locations in a cultured cell assay. Furthermore, L1 could retrotranspose a promoterless neomycin cassette into a transcribed gene, forming new fusion gene products. This indicates that exons downstream of active L1s could be shuffled into new sites, to create new genes[12]. The ability of L1 elements to transduce their 3′ flanks is probably facilitated by their weak polyadenylation signal, which can be bypassed in favour of a stronger downstream poly A signal. Perhaps a weak polyadenylation signal allows L1s residing in introns to be minimally disruptive to their hosts[12].

3′ transduction by L1s has been corroborated by human genome database searches[74,75]; 15–23% of human L1 insertions contain 3′ transduced sequences that account for 0.5–1.0% of the human genome[74,75]. Furthermore, 3′ transduction events have been found in the mouse[74]. Because L1 insertions are often 5′ truncated, some of these shuffling events may lack L1 sequence altogether. Genome database searches for 3′ transduction therefore probably underestimate the extent of L1-mediated sequence shuffling. However, so far no known functional sequences have been found in L1-transduced sequences.

Recombination. Apart from influencing gene expression and function through insertions, the large number of Alu elements, and to a lesser extent L1 elements, may promote homologous recombination events (FIG. 3d). There are many reported homologous recombination events between Alu elements, compared with only three known L1 examples (R. Gatti, personal communication)[45,76,77]. The higher frequency of Alu recombinations may be because of the larger number of Alu elements (recombination between more closely spaced regions of homology tend to occur more efficiently), the tendency of Alus to reside in G+C-rich regions (increasing the chance of an adverse or detectable recombination event), or because Alu elements themselves may contain recombinogenic sequences[76,78].

An intriguing example of how repetitive elements may participate in recombination events was reported for the Charcot-Marie-Tooth disease locus (*CMT*)[79]. CMT disease and hereditary neuropathy with liability to pressure palsies (HNPP) (see also PMP22) are owing to a duplication and deletion, respectively, arising from unequal crossing over between misaligned CMT repeat sequences that flank the *CMT* gene. The crossover point was localized to a 1.7-kilobase hot-spot region in roughly 75% of CMT and HNPP patients[79]. The CMT repeat region which contained the hot spot also contained a defective mariner-like element. Reiter and colleagues speculate that this element could participate in strand-cleavage events at the hot spot[79].

In recent years, it has become increasingly clear that non-homologous DNA recombination events may directly or indirectly involve transposable elements. The most spectacular example of this is V(D)J recombination, a site-directed DNA rearrangement of immunoglobulin gene segments that is carried out by recombinase-activating gene products (RAG1 and RAG2)[42,80] (FIG. 4). Human recombinase-activating genes (*RAGs*) are structurally and functionally similar to DNA transposons, because: first, they are encoded in a short, intronless stretch of DNA[42]; second, they carry out the same cleavage and strand-transfer reactions as the Mu, Tn*5* and Tn*10* transposons, using hydrolysis and transesterification chemistry and generating hairpin intermediates[81–83]; third, they are sequence-specific, recognizing heptamer–nonamer recombination signal sequences[80]; and last, RAG proteins can carry out transposition reactions *in vitro*[82–84]. It would seem that the ancestor of *RAG1* and *RAG2* originally may have been a DNA transposon that was co-opted by the genomes of higher organisms to serve the adaptive immune system[85].

As V(D)J recombination relies on double-strand-break repair enzymes (such as XRCC4, DNA ligase IV, DNA-PK (the DNA-dependent protein kinase) and Ku proteins) the question is raised as to whether these enzymes might aid transposons as well. Perhaps it is not merely because of their high copy number that transposable elements are often found at translocation breakpoints. Ty1 retrotransposons are frequently recov-

© 2000 Macmillan Magazines Ltd

ered at double-strand breaks that are repaired by non-homologous end-joining in yeast[86–88]. Another possibility is that transposons themselves may participate directly in translocations. The RAGs may engage in one-sided transposition reactions that join an immunoglobulin or T-cell receptor (TCR) recombination signal sequence to a site on another chromosome[89]. This could help to explain why so many oncogenic translocations involve the TCR or immunoglobulin loci. Ordinarily, the immunoglobulin recombination signal sequences form signal joints and coding joints, limiting their ability to instigate DNA damage.

Transposons and us: symbiosis or parasitism? Mobile DNA has been coined selfish, existing merely to propagate itself. To some extent, this reputation is deserved because in propagation and genome colonization, mobile elements have been successful, collectively inhabiting over 42% of euchromatic DNA in humans[1]. To expand successfully, transposons and retrotransposons, as in any parasite, must minimize the damage they inflict on their hosts. So they may move in a highly restricted subset of cells or a narrow window of development, or take advantage of a situation in which DNA damage is occurring for another reason[86,88]. The germ line is a popular place for transposon activity, perhaps because lethal mutations can be counterselected during development or during fertilization without too many ill effects[46]. Indeed, members of the SRY family of transcription factors, which are expressed in germ cells, can bind to the human L1 promoter and upregulate its transcription about tenfold[90]. Another strategy for minimizing host cell damage is to target insertions to non-essential regions of the genome. Alternatively, one mobile element could insert into and inactivate another mobile element, thereby reducing the danger of insertional mutagenesis from either element.

Viewed from the perspective of the host, what can be done to reduce the frequency of transposable element disruptions? A number of genomic defence strategies have been proposed, the most popular of which is DNA methylation. However, the data on this model are not straightforward. Cytosine methylation is known to abrogate RNA polymerase II-dependent transcription and to reduce homologous recombination[91]. So the mobility of repetitive DNA may be reduced by methylation[92]. Consistent with this theory, day 9.5 homozygous DNA methyltransferase-1 (*Dmnt1*) knockout mouse embryos have much higher levels of IAP transcripts than their wild-type littermates[93]. Furthermore, an atypical CpG island in the 5′-UTR of active human L1 elements is methylated and, following 5-azadeoxycytidine treatment (which causes demethylation), sense-strand L1 transcription increases[94]. However, in the invertebrate *Ciona intestinalis*, repetitive DNA, including several copies of an LTR retrotransposon, is largely unmethylated, whereas genes are largely methylated[95]. If methylation shuts down mobile DNA, one might expect the level of methylation to correlate with genome size, or at least with the abundance of repetitive sequences[96]. But neither prediction has compelling evidence to support it[97].

Other organisms use various post-transcriptional mechanisms to constrain the movements of mobile elements. These may or may not apply to humans. One fascinating defensive tactic, termed RNA interference (RNAi), consists of the silencing of a gene's expression by dsRNA from that same gene[98]. RNAi has been shown in several model organisms, and can silence transposon activity. Mutated strains of *C. elegans* that are resistant to RNAi have been isolated and, interestingly, some of these mutants show the mobilization of transposable elements such as Tc1 (REFS 99,100). In *Drosophila*, another post-transcriptional mechanism, known as co-suppression, has been linked to silencing of a class of retrotransposons known as I elements[101,102]. Crossing *Drosophila* males possessing active I elements (inducers) with *Drosophila* females lacking these elements (reactive) results in high frequency retrotransposition, chromosomal abnormalities and female sterility. This syndrome, termed hybrid dysgenesis, can be suppressed by passaging part of an I element sequence as a transgene in the reactive female germ line (and then crossing the transgenic female with an inducer male). This homology-dependent silencing mechanism is dependent on transgene copy number and requires transcription and/or chromatin structure modification[101,102].

The relationship between transposons and their hosts is probably not entirely antagonistic, as several host genes have a high degree of homology to one or more transposable elements. For example, the enzyme telomerase, responsible for the maintenance of chromosome ends (telomeres), is similar to the reverse transcriptase (RT) enzyme of non-LTR retrotransposons[103,104]. There is much debate as to whether telomerase originated from non-LTR retrotransposon RT or vice versa[105,106]. Another example of evolutionary borrowing between transposons and their hosts arises in bacterial and yeast group II introns. These self-splicing group II introns are believed to be the ancestors of nuclear spliceosomal introns[107].

Part III: transposons as tools

Potential for research and biotechnology. Retro-transposons, including L1s, Alus and processed pseudogenes, are useful for phylogenetic studies because they enter the genome but cannot exit from it. Many recent retrotransposon insertions in human beings are polymorphic and are useful biallelic markers in gene mapping and in studies of human and primate evolution. In addition, mutations within sets of retrotransposons can be used to discover their lineage and to group the elements into subfamilies. For example, the Ta1 subfamily of L1 elements in humans is a subdivision of the Ta subset[4], and for several reasons seems to be the youngest and the most active of human L1 elements. First, roughly two thirds of genomic Ta1 insertions are polymorphic in the genome[108]. Second, most of the 14 new L1 insertions involve members of the Ta1 subset[108]. Third, full-length Ta1 element sequences are over 99% identical to each other[108]. Last, Ta1 subfamily members are not present in chimpanzees, suggesting

© 2000 Macmillan Magazines Ltd

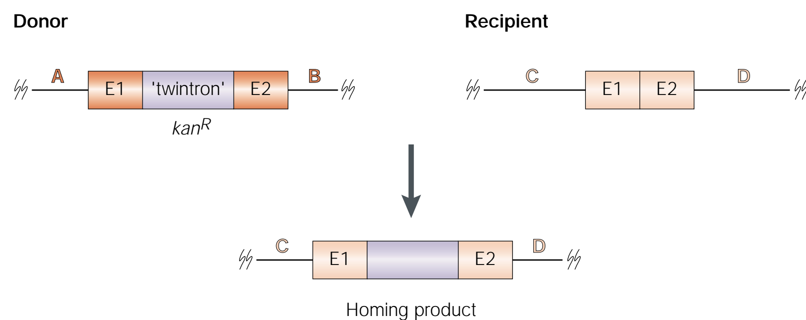


Figure 5 | **Retrohoming.** Retrohoming is the process whereby group II introns self splice, are reverse transcribed by their own reverse transcriptase and insert into an intronless allele (fused exons 1 and 2, E1E2) of the same gene (E1–intron–E2). Using a two-plasmid system, homing from the donor plasmid (in which the kanamycin-resistance gene *kan*^{R} is carried by the intron) to a recipient plasmid (which is chloramphenicol resistant, *cam*^{R}) could be followed by the acquisition of kanamycin resistance (retrohoming products are *cam*^{R}, *kan*^{R})[16]. The term 'twintron' refers to the fact that a group I intron is embedded in the group II intron.

(along with other data) that the Ta1 subfamily arose roughly 2.5 million years ago[108].

A similar type of analysis has been carried out with Alu elements, which can be subdivided into AluY (the most recently dispersed), AluS (intermediate) and AluJ (the oldest elements)[48,109]. Mobile elements have also been used to probe the phylogenetic relationships between more distantly related species. For example, the analysis of two families of SINEs convincingly showed that the terrestrial ancestors of whales were more closely related to hippos and ruminants than they were to pigs and camels[110]. Horizontal transmission of DNA transposons, which results in the transfer of a transposon from the genome of one species to the genome of another species, can confound phylogenetic analysis. This is particularly problematic with DNA transposons that reside in a wide range of species.

Transposons have found increasing use in molecular biology. For example, the hyperactive form of the Tn5 DNA transposon can be cloned next to a gene of interest and be used to generate intramolecular transposition products[111]. These intramolecular transpositions result in hundreds of deletions or inversions in a given target gene, providing a rapid way of making reagents for protein structure–function analysis[111]. Transposons can be made to jump randomly into large cloned pieces of DNA, facilitating the generation of large plasmid libraries with different transposable-element insertions, from which individual subclones can be sequenced directly from the transposon tag[112].

Because transposons can tag the sequences that they disrupt, they may prove to be powerful insertional mutagens *in vitro* and possibly *in vivo*. With their broad host range, Tc1/mariner transposons are prime candidates for transposon-tagged mutagenesis in vertebrate organisms, such as the mouse. To that end, Luo and colleagues were able to show chromosomal transposition of the *Sleeping Beauty* transposon in mouse embryonic stem cells at frequencies of 0.5–5 $\times 10^{-6}$ transfected cells[113]. Like DNA transposons, retrotransposons may prove to be useful as insertional mutagens because they can jump into genes and tag the genes that they disrupt. Furthermore, because retrotransposition is replicative, retrotransposons may be useful in models of cancer in which several sequential mutations may be required in a subset of cells. However, several potential obstacles remain for germline or somatic transposon-mediated mutagenesis. At present, it is unknown how frequently elements such as L1 retrotranspose in transgenic mice. Furthermore, the competence of different organ systems and cell types to support chromosomal retrotransposition is unknown.

Transposable elements may also be useful as gene-delivery vehicles. Unlike viruses, transposons can be assembled from endogenous sequences (such as L1) which are not inherently immunogenic. Transposon and retrotransposon insertions are typically single copy, unlike most DNA-based transfection methods, and therefore potentially less liable to be inactivated by methylation[114]. Recently, Kay's group showed chromosomal integration and long-term expression of factor IX delivered to haemophilic mice by the *Sleeping Beauty* transposase[115].

Bacterial group II introns are another potentially useful tool for targeted gene delivery or disruption in mammalian systems. Group II introns are catalytic RNAs that can self-splice and be reverse-transcribed by their own encoded RT. In a process termed retrohoming (FIG. 5), the group II intron is replicated from an intron-containing allele to a non-intron-containing allele of the same gene[16,116]. In addition to retrohoming, group II introns can invade ectopic chromosomal sites by retrotransposition, providing a potential mechanism for the dispersal of spliceosomal introns in the eukaryotic genome[15]. Lambowitz's group has recently elucidated the nature of the target-site recognition rules of a group II intron. So they could design an intron on one episome that inserts into a specific gene target, such as *CCR5*, on another episome in transformed mammalian cells[117].

In summary, mobile elements have had an important role in the evolution of the human genome. Although parasitic in nature, their sequences have contributed in many ways to cellular function. In the future, their contributions to our well-being may extend to their use as tools for gene discovery and gene delivery.

Links

DATABASE LINKS **copia | PRKR | factor VII | haemophilia A | Duchenne muscular dystrophy | type 2 retinitis pigmentosa | β-thalassaemia | chronic granulomatous disease | APC | Wilms tumour suppressor gene | dystrophin | CMT | HNPP | PMP22 | Rag2 | RAG1 | RAG2 | SRY family | Factor IX | CCR5**

FURTHER INFORMATION **Table with human retrotransposon insertions | RepBase | Transposon DNA database at the Pasteur Institute | Infobiogen | RepeatMaker | Clustal W | GCG | BLAST**

ENCYCLOPEDIA OF LIFE SCIENCES **DNA transposition: classes and mechanisms | Genome organization/ human | Developmentally programmed DNA rearrangements | Repetitive DNA: evolution**

© 2000 Macmillan Magazines Ltd

1. Smit, A. F. A. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
A learned and up-to-date review of transposable elements and their remnants in mammalian genomes.
2. Sassaman, D. M., Dombroski, B. A., Moran, J. V. *et al.* Many human L1 elements are capable of retrotransposition. *Nature Genet.* **16**, 37–43 (1997).
3. DeBerardinis, R., Goodier, J., Ostertag, E. & Kazazian, H. H.. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nature Genet.* **20**, 288–290 (1998).
4. Skowronski, J., Fanning, T. G. & Singer, M. F. Unit length LINE-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8**, 1385–1397 (1988).
5. 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).
6. Yang, J., Malik, H. S. & Eickbush, T. H. 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 (1999).
7. Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).
This study provides the first conclusive evidence that non-LTR retrotransposons contain an endonuclease activity that is critical for their retrotransposition.
8. Jurka, J. & Klonowski, P. Integration of retroposable elements in mammals: selection of target sites. *J. Mol. Evol.* **43**, 685–689 (1996).
9. Martin, F. C., Maranon, C., Olivares, M., Alonso, C. & Lopez, M. C. Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the Ape family of DNA repair enzymes. *J. Mol. Biol.* **247**, 49–59 (1995).
10. Cost, G. J. & Boeke, J. D. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* **37**, 18081–18093 (1998).
11. Kazazian, H. H. & Moran, J. V. The impact of L1 retrotranpsosons on the human genome. *Nature Genet.* **19**, 19–20 (1998).
12. Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H. Exon shuffling by L1 retrotransposition. *Science* **283**, 1530–1534 (1999).
Use of a cultured-cell assay to show that L1s insert into transcribed genes and can retrotranspose sequences derived from their 3′ flanks to new genomic locations.
13. Malik, H. S., Burke, W. D. & Eickbush, T. H. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16**, 793–805 (1999).
A comprehensive phylogenetic analysis of non-LTR retrotransposable elements using an extended reverse transcriptase domain to resolve non-LTR elements into 11 clades.
14. Eickbush, T. H. Introns gain ground. *Nature* **404**, 940–943 (2000).
15. Cousineau, B. *et al.* Retrotransposition of a bacterial group II intron. *Nature* **404**, 1018–1021 (2000).
Self-splicing group II introns from bacteria can retrotranspose into ectopic sites, possibly accounting for the genomic dispersal of spliceosomal introns.
16. Cousineau, B. *et al.* Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94**, 451–462 (1998).
17. Zimmerly, S., Guo, H., Perlman, P. S. & Lambowitz, A. M. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82**, 545–554 (1995).
18. Heidmann, O. & Heidmann, T. Retrotransposition of a mouse IAP sequence tagged with an indicator gene. *Cell* **64**, 159–170 (1991).
19. Fehrmann, F., Welker, R. & Krausslich, H.-G. Intracisternal A-type particles express their proteinase in a separate reading frame by translational frame-shifting, similar to D-type retroviruses. *Virology* **235**, 352–359 (1997).
20. Boeke, J. D. & Stoye, J. P. in *Retroviruses* (eds Coffin, J. M., Hughes, S. H. & H. E. Varmus, H. E.) 343–436 (Cold Spring Harbor, New York, 1997).
21. 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).
22. Goedert, J. J. *et al.* High prevalence of antibodies against HERV-K10 in patients with testicular cancer but not with AIDS. *Ca. Epid. Biomarkers Prev.* **8**, 293–296 (1999).
23. Mueller-Lantzsch, N., Sauter, M., Weiskircher, A., Kramer, K., Best B. & Grasser, F. Human endogenous retroviral element K10 (HERV-K10) encodes a full length *gag* homologous 73 kDa protein and a functional protease. *AIDS Res. Hum. Retroviruses* **9**, 343–350 (1993).
24. Kitamura, Y., Ayukawa, T., Ishikawa, T., Kanda, T. & Yoshiike, K. Human endogenous retrovirus K10 encodes a functional integrase. *J. Virol.* **70**, 3302–3306 (1996).
25. Berkhout, B., Jebbink, M. & Zsiros, J. Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus. *J. Virol.* **73**, 2365–2375 (1999).
26. Mayer, J. *et al.* An almost-intact human endogenous retrovirus K on human chromosome 7. *Nature Genet.* **21**, 257–258 (1999).
A cosmid clone was isolated from human chromosome 7, which contained a full-length HERV–K provirus with intact open reading frames for all of the retroviral genes. On the basis of sequence differences in the LTRs, the provirus may have integrated as recently as 1.2 million years ago.
27. Mayer, J., Meese, E. & Mueller-Laztsch, N. M. Human endogenous retrovirus K homologous sequences and their coding capacity in old world primates. *J. Virol.* **72**, 1870–1875 (1998).
28. Barbulescu, M. *et al.* Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. *Curr. Biol.* **9**, 861–868 (1999).
29. Medstrand, P. & Mager, D. Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* **72**, 9782–9787 (1998).
30. Tchenio, T. & Heidmann, T. Defective retroviruses can disperse in the human genome by intracellular transposition. *J. Virol.* **55**, 2113–2118 (1991).
31. Itin, A. & Keshet, E. Apparent recombinants between virus-like (VL30) and murine leukemia virus-related sequences in mouse DNA. *J. Virol.* **47**, 178–184 (1983).
32. Moon, B. C. & Friedman, J. M. The molecular basis of the obese mutation in ob2J mice. *Genomics* **42**, 152–156 (1997).
33. Mager, D. L. & Freeman, J. D. Novel mouse type D endogenous proviruses and Etn elements share long terminal repeat and internal sequences. *J. Virol.* **74**, 7221–7229 (2000).
34. Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
35. Smit, A. F. & Riggs, A. D. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl Acad. Sci. USA* **20**, 1443–1448 (1996).
36. Plasterk, R. H. A., Izsvak, Z. & Ivics, Z. Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet.* **15**, 326–332 (1999).
37. Oosumi, T., Balknap, W. R. & Garlick, B. Mariner transposons in humans. *Nature* **378**, 672 (1995).
38. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvak, Z. Molecular reconstruction of sleeping beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510 (1997).
Construction of a synthetic Tc1-like DNA transposon (*Sleeping Beauty*) using parts of transposons from eight different species of fish. *Sleeping Beauty* was able to mediate cut-and-paste transpositions in fish, mouse and human cells.
39. Gueiros-Filho, F. J. & Beverley, S. M. *Trans*-kingdom transposition of the *Drosophila* element mariner within the protozoan *Leishmania*. *Science* **276**, 1716–1719 (1997).
40. Schouten, G. J., van Luenen, H. G. A. M., Nerra, N. C. V., Valerio, D. & Plasterk, R. H. A. Transposon Tc1 of the nematode *Caenorhabditis elegans* jumps in human cells. *Nucleic Acids Res.* **12**, 3013–3017 (1998).
41. Fugmann, S. D., Lee, A. I., Shockett, P. E., Villey, I. J. & Schatz, D. G. The RAG proteins and V(D)J recombination: complexes, ends and transposition. *Annu. Rev. Immunol.* **18**, 495–527 (2000).
42. Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517–1523 (1990).
43. Dreyfus, D. H. *et al.* Epstein Barr virus infection of T cells: implications for altered T-lymphocyte activation, repertoire development and autoimmunity. *Immunol. Rev.* **152**, 89–110 (1996).
44. Schmid, C. W. Does SINE evolution preclude Alu function? *Nucleic Acids Res.* **26**, 4541–4550 (1998).
45. Deininger, P. L. & Batzer, M. A. Minireview: Alu repeats and human disease. *Mol. Gen. Metab.* **67**, 183–193 (1999).
46. 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).
47. 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).
48. Kapitonov, V. & Jurka, J. The age of Alu subfamilies. *J. Mol. Evol.* **42**, 59–65 (1996).
49. Ogiwara, I., Miya, M., Ohshima, K. & Okada, N. Retropositional parasitism of SINEs and LINEs in Elasmobranchs. *Mol. Biol. Evol.* **16**, 1238–1250 (1999).
50. Okada, N., Hamada, M., Ogiwara, I. & Ohshima, K. SINEs and LINEs share common 3′ sequences: a review. *Gene* **205**, 229–243 (1997).
51. Chu, W. M., Ballard, R., Carpick, B. W., Williams, B. R. & Schmid, C. W. Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR *Mol. Cell. Biol.* **18**, 58–68 (1998).
52. Boeke, J. D. LINEs and Alus — the polyA connection. *Nature Genet.* **16**, 6–7 (1997).
53. Chang, D. Y. *et al.* A human Alu RNA-binding protein whose expression is associated with the accumulation of small cytoplasmic Alu RNA. *Mol. Cell. Biol.* **14**, 3949–3959 (1994).
54. Maestre, J., Tchenio, T., Dhellin, O. & Heidmann, T. mRNA retrotransposition in human cells: processed pseudogene formation. *EMBO J.* **14**, 6333–6338 (1995).
55. Dunham, I. *et al.* The DNA sequence of chromosome 22. *Nature* **402**, 489–495 (2000).
56. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24**, 363–367 (2000).
57. Kazazian, H. H. *et al.* Haemophilia A resulting from the *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332**, 164–166 (1988).
58. Dombrowski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F. & Kazazian, H. H. Isolation of an active human retrotransposable element. *Science* **254**, 1805–1808 (1991).
59. Holmes, S. E., Dombroski, B. A., Krebs, C. M., Boehm, C. D. & Kazazian, H. H. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nature Genet.* **17**, 143–148 (1994).
60. 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).
61. Schwahn, U. *et al.* Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nature Genet.* **19**, 327–332 (1998).
62. Divoky, V. *et al.* A novel mechanism of β–thalassemia. The insertion of L1 retrotransposable element into β globin IVSII. *Blood* **88**, 148a (1996) .
63. Kimberland, M. L. *et al.* Full-length human L1 insertions retain the capacity for high-frequency retrotransposition in cultured cells. *Hum. Mol. Genet.* **8**, 1557–1560 (1999).
64. Meischl, C., de Boer, M., Ahlin, A. & Roos, D. Intronic insertion of a LINE–1 fragment as the cause of chronic granulomatous disease. *Eur. J. Hum. Genet.* **8**,697–703 (2000).
65. 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).
66. Bailey, J. A., Carrel, L., Chakravarti, A. & Eichler, E. E. Molecular evidence for a relationship between LINE-1 elements and X-chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl Acad. Sci.* **97**, 6634–6639 (2000).
67. Lyon, M. F. X-chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* **80**, 133–137 (1998).
68. Kazazian, H. H. An estimated frequency of endogenous insertional mutations in humans. *Nature Genet.* **22**, 130 (1999).
69. Vansant, G. & Reynolds, W. F. The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. *Proc. Natl Acad. Sci. USA* **92**, 8229–8233 (1995).
70. Britten, R. J. Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**, 177–182 (1997).
71. Hewitt, S. M., Fraizer, G. C. & Saunders, G. F. Transcriptional silencer of the Wilms tumor gene WT1 contains an Alu repeat. *J. Biol. Chem.* **270**, 17908–17912 (1995).
72. Yang, Z., Boffelli, D., Boonmark, N., Schwartz, K. & Lawn, R. Apolipoprotein (a) gene enhancer resides within a LINE element. *J. Biol. Chem.* **273**, 891–897 (1998).
73. Domansky, A. N. *et al.* Solitary HERV-K LTRs possess bi-

© 2000 Macmillan Magazines Ltd

directional promoter activity and contain a negative regulatory element in the U5 region. *FEBS Lett.* **472**, 191–195 (2000).
74. Goodier, J. L., Ostertag, E. M. & Kazazian, H. H. Transduction of 3′ flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* **9**, 653–657 (2000).
75. Pickeral, O. K., Makalowski, W., Boguski, M. S. & Boeke, J. D. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genet. Res.* **10**, 411–415 (2000).
76. Burwinkel, B. & Kilimann, M. W. Unequal homologous recombination between LINE1 elements as a mutational mechanism in human genetic disease. *J. Mol. Biol.* **277**, 513–517 (1998).
77. Segal, Y. *et al.* LINE-1 elements at the sites of molecular rearrangements in Alport Syndrome-diffuse leiomyomatosis. *Am. J. Hum. Genet.* **64**, 62–69 (1999).
78. Rudiger, N. S., Gregersen, N. & Kielland-Brandt, M. C. One short well conserved region of Alu sequences is involved in human gene rearrangements and has homology with prokaryotic *chi. Nucleic Acids Res.* **23**, 256–260 (1995).
79. Reiter, L. T. *et al.* A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-line element. *Nature Genet.* **12**, 288–297 (1996).
80. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. Sequences at the somatic recombination sites of immunoglobulin light chain genes. *Nature* **280**, 288–294 (1979).
81. Kennedy, A. K., Guhathakurta, A., Kleckner, N. & Harriford, D. B. Tn10 transposition via a DNA hairpin intermediate. *Cell* **95**, 125–134 (1998).
82. Agrawal, A., Eastman, Q. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744–751 (1998).
83. Melek, M., Gellert, M. & van Gent, D. C. Rejoining of DNA by the RAG1 and RAG2 proteins. *Science* **280**, 301–303 (1998).
Reference 82 and 83 show that the RAG1 and RAG2 proteins, together, form a DNA transposase which can excise a piece of DNA flanked by recombination signal sequences and move it to another DNA molecule.
84. van Gent, D. C., Mizuuchi, K. & Gellert, M. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271**, 1592–1594 (1996).
85. Thompson, C. B. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* **3**, 531–539 (1995).
86. Moore, J. K. & Haber, J. E. Capture of retrotransposon DNA at the sites of chromosomal double strand breaks. *Nature* **383**, 644–646 (1996).
87. Yu, X. & Gabriel, A. Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell* 4, 873–881 (1999).
88. Teng, S.-C., Kim, B. & Gabriel, A. Retrotransposon reverse-transcriptase mediated repair of chromosomal breaks. *Nature* **383**, 641–644 (1996).
89. Hiom, K., Melek, M. & Gellert, M. DNA Transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463–470 (1998).
90. Tchenio, T., Casella, J.-F. & Heidmann, T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **28**, 411–415 (2000).
91. Bender, J. Cytosine methylation of repeated sequences in eukaryotes: the role of DNA pairing. *Trends Biochem. Sci.* **23**, 252–256 (1998).
92. Bestor,T. H. & Tycko, B. Creation of genomic methylation patterns. *Nature Genet.* **12**, 363–367 (1996).
93. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20**, 116–117 (1998).
94. Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P. & Warren, W. D. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J. Biol. Chem.* **272**, 7810–7816 (1997).
95. Simmen, M. W. *et al.* Nonmethylated transposable elements and methylated genes in a chordate genome. *Science* **283**, 1164–1167 (1999).
96. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
97. Regev, A., Lamb, M. J. & Jablonka, E. The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol. Biol. Evol.* **15**, 880–891 (1998).
98. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans. Nature* **391**, 806–811 (1998).
Double-stranded RNA injected into *C. elegans* can specifically interfere with expression of the corresponding endogenous gene in injected animals and their progeny.
99. Ketting, R. F., Haverkamp, T. H. A., van Luenen, H. G. A. M. & Plasterk, R. *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141 (1999).
This elegant study provides genetic evidence that links activation of the Tc1 transposon to resistance to RNA interference in some strains of *C. elegans.* One of the mutated genes in the interference-resistant strains of worms, *mut-7*, was found to encode a protein with homology to RNaseD, which, in turn, may be involved in the negative regulation of transposable elements.
100. Tabara, H. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans. Cell* **99**, 123–132 (1999).
101. Chaboissier, M. C., Bucheton, A. & Finnegan, D. J. Copy number control of a transposable element, the I-factor, a LINE-like element in *Drosophila. Proc. Natl Acad. Sci. USA* **95**,11781–11785 (1998).
102. Jensen, S., Gassama, M. P. & Heidmann, T. Taming of transposable elements by homology-dependent gene silencing. *Nature Genet.* **21**, 209–212 (1999).
103. Nakamura, T. M. *et al.* Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955–959 (1997).
104. Eickbush, T. H. *The Evolutionary Biology of Viruses* (ed. Morse, S. S.) 121–157 (Raven, New York, 1994).
105. Nakamura, T. M. & Cech, T. R. Reversing time: origin of telomerase. *Cell* **92**, 587–590 (1998).
106. Eickbush, T. H. Telomerase and retrotransposons: which came first? *Science* **277**, 911–912 (1997).
This paper describes the evolutionary relationships among various transposable elements containing reverse transcriptase and telomerase, which has sequence and functional similarity to reverse transcriptase. One possible phylogenetic tree gives non-LTR retrotransposons the most ancient RTs and predicts that the RT of one of these older parasites may have been recruited (subsequently evolving into telomerase) to maintain chromosomes ends).
107. Sharp, P. Five easy pieces. *Science* **254**, 663 (1991).
108. Boissinot, S. Chevret, P. & Furano, A. V. L1 (LINE-1) retrotransposition evolution and amplification in recent human history. *Mol. Biol. Evol.* **17**, 915–928 (2000).
109. Batzer, M. A. *et al.* Standardized nomenclature for Alu repeats. *J. Mol. Evol.* **42**, 3–6 (1996).
110. Shimamura, M. *et al.* Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* **388**, 666–670 (1997).
111. York, D., Welch, K., Goryshin, I. Y. & Reznikoff, W. S. Simple and efficient generation of *in vitro* nested deletions and inversions: Tn5 intramolecular transposition. *Nucleic Acids Res.* **26**, 1927–1933 (1998).
112. Hoffman, L. M. & Loomis, K. B. Protein modification using the new EZ:TN™ in-frame linker insertion kit. *Epicentre Forum* **7**, 4–6 (2000).
113. Luo, G., Ivics, Z., Izsvak, Z. & Bradley, A. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **95**, 10769–10773 (1998).
114. Garrick, D., Fiering, S., Martin, D. I. K. & Whitelaw, E. Repeat-induced gene silencing in mammals. *Nature Genet.* **18**, 56–59 (1998).
115. Yant, S. R. *et al.* Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. *Nature Genet.* **25**, 35–41 (2000).
The first use of a transposable element as a gene delivery vehicle in mice. The authors show that the *Sleeping Beauty* DNA transposon can insert the *factor IX* gene into chromosomal DNA by transposition. Some of the haemophilic mice that received factor IX–transposon DNA expressed therapeutic levels of *factor IX* for over five months.
116. Moran, J. V., Zimmerly, S., Eskes, R., Kennell, J. C. & Lambowitz, A. M. Mobile group II introns of yeast mitochondrial DNA are novel, site-specific retroelements. *Mol. Cell. Biol.* **15**, 2829–2838 (1995).
117. Guo, H. *et al.* Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289**, 452–457 (2000).
118. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
The first demonstration that cloned human L1 elements can actively retrotranspose at high frequencies in cultured mammalian cells.
119. Mathias, S. L., Scott, A. F., Kazazian H. H., Boeke, J. D. & Gabriel, A. Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808–1810 (1991).
120. Clements, A. P. & Singer, M. F. The human LINE-1 reverse transcriptase: effects of deletions outside the common reverse transcriptase domain. *Nucleic Acids Res.* **26**, 3528–3535 (1998).
121. Dombroski, B. A. *et al.* An *in vivo* assay for the reverse transcriptase of human retrotransposon L1 in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **14**, 4485–4492 (1994).
122. Hohjoh, H. & Singer, M. F. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* **15**, 630–639 (1996).
123. 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).

Acknowledgements

We are very grateful to the reviewers and to J. Mayer, J. Moran, M. Carmen Seleme and E. Ostertag for their thoughtful comments on the manuscript. E.L.P. is supported by a grant from the National Cancer Institute and H.H.K. is supported by grants from the NIH.

© 2000 Macmillan Magazines Ltd