CMLS, Cell. Mol. Life Sci. 61 (2004) 2046–2059
1420-682X/04/162046-14
DOI 10.1007/s00018-004-4041-z

CMLS Cellular and Molecular Life Sciences

Review

Retroelements and formation of chimeric retrogenes

A. A. Buzdin

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya, Moscow 117997 (Russia), Fax +7 095 330 65 38, e-mail: anton@humgen.siobc.ras.ru

Received 28 January 2004; received after revision 10 March 2004; accepted 15 April 2004

Abstract. It is very likely that formation of new genes is the main pathway of molecular evolution in living organisms. Many such genes are products of preexisting reshuffling of genetic material. In these processes a very important role is played by mutations associated with the activity of transposable elements, mostly retroelements (REs) for higher eukaryotes. The life cycle of REs involves a stage of so-called reverse transcription of their RNA intermediates, i. e. synthesis of complementary DNA on an RNA template. Transcriptionally active sequences of RE origin are referred to as retrogenes. REs create chimeric genes by a variety of mechanisms: new RE insertions, recombinations between RE sequences, formation of functional gene active pseudogenes and template switches during reverse transcription of messenger RNA. The abovementioned events are also able to give rise to new RE families. These mechanisms are reviewed here along with the description of major RE groups.

Key words. Retroelements; DNA; RNA; chimeric; gene; retrotransposon; SINE; LINE.

Introduction

There are many examples of retroelement (RE) involvement in the formation of hybrid genes. RE activity itself is an inexhaustible source of many chimeric structures in the genome, such as transcriptionally silent DNA sequences, genes and new RE families. Our current understanding of these processes is certainly inseparably linked with the study the molecular genetic aspects of RE functioning. This review gathers together both fundamental data on RE life cycle and evolution, and a description of the most important mechanisms of their contribution to gene formation.

Transposable elements (TEs) are DNA fragments that are able to self-reproduce and to change their location into host genome, i. e. to transpose. TEs were discovered about 50 years ago in maize DNA by Barbara McClintock [1]. Since then TEs have been found in genomes of almost all organisms. Moreover, they are now known to make up a great portion of eukaryotic DNA. For example, TEs constitute more than 50% of maize (*Zea mays*) genome [2, 3], 22% of the *Drosophila* genome [4] and 42% of human DNA [5]. However, different TE groups have a strikingly different number of representatives, from a few copies to millions.

TEs differ from each other in structure and in transposition features. They can be subdivided into two principal classes [2, 3, 5–9]. Class (ii) representatives, called DNA transposons, use copies of their DNA to transpose via a 'copy and paste' mechanism. This review deals with class (i) TEs, which proliferate through their RNA intermediates. They use an RNA-dependant DNA polymerase, also called reverse transcriptase (RT), an enzyme that synthesizes a complementary DNA (cDNA) chain on an RNA template.

General characteristics of REs

The term 'retroelement' is applied to a vast class of nucleic acid sequences whose appearance and/or proliferation in a host genome are one way or another dependent on the direct transfer of genetic information from RNA to

DNA, called reverse transcription. This phenomenon was described for the first time in 1970 by Temin and Baltimore. They purified and characterized retroviral RT [10, 11]. Afterwards, RT sequences were found in very diverse genetic elements. Not only representatives of the viral realm did contain RT, such as retroviruses, hepadnaviruses and caulimoviruses, but also many eukaryotic TEs, mitochondrial group II introns, bacterial retrointrons and some plasmids. REs having their own RT genes are referred to as autonomous REs. They can be subdivided into two major groups: long terminal repeat (LTR)-containing elements and non-LTR retrotransposons. LTRs are usually 1000–1500-bp-long sequences flanking the retroelement 'body' in genomic DNA [12, 13]. Autonomous non-LTR REs are generally assigned to LINEs (long interspersed nuclear elements). LINEs are 3.5–8-kb-long sequences harbouring RT genes and frequently other genes encoding proteins necessary for their functioning [14, 15]. LINEs usually contain at their 3′ termini poly(A) stretches that play an important role in their integration into new genomic loci.

Non-autonomous REs lack RT genes and are classified as either SINEs (short interspersed nuclear elements) or processed pseudogenes. SINES are 50–700-bp-long sequences having, as a rule, an internal RNA polymerase III promoter. At their 3′ end they usually have a poly(A) tail as well. B1, B2, MIR famous Alu repeats, and many other REs belong to SINEs [13, 16].

Thus, REs are subdivided into three major systematic classes: LTR-containing elements, LINEs and SINEs. There is one more, rather unusual RE group, called retrointrons or mobile group II introns. The most important hypothesis explaining the RE origin was proposed by Temin [17]. It says that autonomous REs coevolved with the gene of the RT. The putative pathway of RE evolution started with the RT gene, first created non-LTR retrotransposons, and finally LTR-containing elements and retroviruses [13, 18]. Indeed, detailed sequence analysis clearly demonstrates the consecutive structural complexity, such as recruitment of new regulatory proteins, additional enzymatic activities – of RNase H, integrase, protease – from non-LTR to LTR-containing retroelements. In this way, retrointrons and LINEs are more ancient RE forms than retroviruses and LTR elements [19]. The mechanism of LINE and retrointron retroposition is also much simpler than that of LTR REs. According to a certain hypothesis [20], another offspring of an ancestral RT sequence is the gene for telomerase, a cellular telomere end-building enzyme that has RNA-dependant DNA polymerase activity. Interestingly, in certain *Drosophila* species the telomerase gene is inactivated, whereas telomere lengths are maintained by LINE integrations in chromosome termini [21, 22].

Retrointrons (group II introns)

Until recently, it was believed that retroelements are present exclusively in eukaryotic genomes. That is not the case, as we know now. Retrointrons, or group II introns, form one of the two classes of self-splicing introns, which exist in the genomes of prokaryotes and in eukaryotic organelles [23, 24]. RT sequence-based phylogenetic analysis revealed that group II introns are the oldest group among REs [25]. It is likely that group II mobile introns appeared for the first time in the genomes of bacteria. It doesn't seem surprising, as modern eukaryotic mitochondria and plastids where retrointrons were found are bacterial cell descendants, according to endosymbiotic theory [23]. Group II introns encode a single protein having both RT and endonuclease activities. Retrointrons are transcribed as parts of the genes they have inserted.

Retrointron RNA has ribozyme activity and self-splices from pre-messenger RNA (mRNA). Then spliced intronic RNA can be translated to give functional protein. During retroposition, retrointron RNA serves again as a ribozyme, making sequence-specific single-strand DNA breaks within the host genome. At the same time RNA covalently binds to the 5′ terminus of the DNA breaks. The second DNA strand is attacked by the retrointron protein, which reverse-transcribes the RNA (DNA synthesis is initiated at the newly formed 3′-hydroxyl of a target DNA). Subsequent RNA replacement by DNA followed by reparation of single-strand breaks completes mobile intron integration [24]. Retrointrons most probably gave rise to a very important group of eukaryotic genes, those of small nuclear and nucleolar RNAs, which take part in pre-mRNA splicing [25].

LINEs

As mentioned above, non-LTR REs are subdivided into two major classes: LINEs and SINEs [26]. Although they do differ in length (LINEs are 3.5–8 kb long; SINEs are 50–700 bp long), there is a substantial distinction between them in their principles of organization. LINEs are widely distributed in eukaryotes. These REs have been found in the genomes of fungi, plants, as well as in vertebrate and invertebrate animals. For example, nearly 17% of human DNA is occupied by LINEs [5].

LINE inserts can be identified by 10–20-bp-long genomic DNA duplications (called target site duplications), flanking each copy of these REs. LINEs contain 5–40 bp long oligo(A) sequences on their 3′ termini. However, sometimes there are other A-rich microsatellites instead of oligo(A) tails. Another distinguishing feature of LINEs is their frequent 5′ truncation, so that it is sometimes difficult to find the true RE 5′ end (e.g., primate L1 LINE-truncated copies were first described as the particular TE

[27]; short-cut F elements, LINEs from *Drosophila*, were known until recently as suffix retroposons [28]). Such truncations are probably due to abortive LINE RNA reverse transcription, when RT dissociates from its RNA template before having completed cDNA synthesis, and newly formed REs thus lack 5′ segments [29–32].

The number of copies per genome varies dramatically among different LINE families. It is believed that mammalian DNA contains REs of only one LINE family – L1 [27] – but this family is represented by a great number of copies, for instance about 5×10^5 human L1 elements occupy a total of 17% of human genomic DNA [5]. Most of these L1 copies are defective; only 80–100 human LINEs are transpositionally competent [33]. Approximately 4000 stable L1 inserts appeared during the last 6 million years of human genomic evolution, suggesting that the average rate of LINE accumulation in human DNA is about seven new copies per 10,000 years [34]. In the mouse, the number of active LINEs is significantly (nearly 30-fold) higher [35]. Notably, the ratio of retroposition frequency to the number of active REs is essentially the same for both mouse and human genomes, thus demonstrating similar mobility of active L1 members in both species [36].

In contrast, invertebrate genomes are usually examples of coexistence of several LINE families (e.g. ~10 different LINE families were found in *Drosophila*: F, Doc, G, R1, R2, HeT, jockey [37], BS [38], TART [39] and so on), but each of them is represented by a few thousand members [40]. The possible explanation for this phenomenon could be the hypothesis proposed by Petrov et al. [41], which explains the low RE copy number, in particular in *Drosophila*, by rampant deletion of DNA in unconstrained regions. In such cases all 'unnecessary' DNA, including REs, is quickly eliminated, and only essential sequences, whose loss leads to lethal mutations, survive in the genome. In mammals this mechanism functions with much slower DNA deletion rates, and numerous 'junk' sequences are thus perpetually being accumulated [4]. The transpositional activity of REs also varies greatly among mammalian and *Drosophila* genomes. For instance, TEs account for ~50% of spontaneous mutations in *Drosophila melanogaster* [42], in contrast to as low as 0.2% in the human genome [43].

LINEs are transcribed by cellular RNA polymerase II from an internal promoter located in their 5′-untranslated region (5′-UTR); as a matter of fact, the first internal RNA polymerase II promoter sequence was discovered for the jockey LINE element [44]). LINE 3′-terminal sequences generally have 3′-processing signal AATAAA [45], and that LINE oligo(A) sequence serves as a polyadenylation enhancer [46–48]. The full-sized LINE (+) RNA is, as in the case of retrointrons, both template for protein synthesis and transpositional RNA intermediate [49].

The presence of so many LINEs in eukaryotic genomes affects many cellular processes. LINE sequences are recombination hot spots, thus causing many genetic diseases [25, 50–52]. For instance, LINE recombination results in deletions within the β subunit of the phosphorylase kinase (PHKB) gene [51]; another example is homologous recombination between intronically located L1s, which causes deletions encompassing the 5′-terminal fragments of the paired type IV collagen genes *COL4A5* and *COL4A6*, thus giving rise to Alport syndrome [52]. Furthermore, unequal meiotic crossing over between L1 sequences on homologous chromosomes can create new gene families, as was the case for gamma-globin gene duplication [53]. LINE inserts often disrupt pre-existing gene exon-intron structures, causing various disorders: hemophilia A (L1 inserted into the factor VIII gene [54]), Duchenne muscular dystrophy (retrotransposition into the dystrophin gene [55, 56]) and many other diseases. Regardless of the fact that L1 insertions cause multiple recombination events, the frequency of such DNA rearrangements is much lower than one would expect taking into account their great copy number. To explain this phenomenon, it was hypothesized that mammalian cells have adopted a mechanism (as yet unknown) preventing host organisms from such recombinations [57].

Besides causing genomic instability, LINEs also participate in regulation of host gene expression [25, 29, 58–60]. For instance, the L1 element provides a polyadenylation signal for the human *NSBP1* gene [60]. LINE sequences can be used for cellular genes as enhancers (human apolipoprotein A) or as transcriptional repressors (rat insulin I [58], human gene *Ñ1D* [61]). The human gene *ZNF-177* 5′-UTR-located LINE sequence affects this gene expression at both transcriptional and translational levels [62]. Many other examples of LINE influence on gene expression are given in [25, 32, 36, 43, 58].

Another interesting peculiarity of LINEs is the ability to transfer their 3′-flanking DNA to new genomic loci, called L1 transduction [63, 64]. Its mechanism and consequences for genome evolution are better understood for the mammalian L1 group of LINEs. L1s have a rather weak polyadenylation signal; the RNA polymerase complex therefore sometimes gets through it and terminates an RNA synthesis on any polyadenylation site located downstream. Retropositions of LINE transcripts accompanied by 3′-flanking genomic sequences have been described in the literature [65–67]. Human genomic database analysis reveals that ~20% of all L1 inserts contain transduced DNA at the 3′ ends. The length of these sequences varies from 30 bp to 1 kb. Taken together, such transduced DNA makes up ~0.6–1% of the human genome (a fraction comparable to that occupied by exons [63, 64]).

In addition, as demonstrated recently by Speek and Nigumann for human LINEs, the L1 5′-UTR has an extra promoter function, driving transcription in an antisense

orientation relative to L1 [68, 69]. Each of these transcripts, therefore, mostly contains sequences complementary to genomic DNA, and only a small portion of LINE RNA. In particular, the authors have shown that some human genes are transcribed from such promoters [68]. Moreover, many normal L1 promoters are turned towards the host DNA, as many mammalian LINEs contain inversions [70].

LTR retrotransposons and endogenous retroviruses (LR/ERVs)

This group matches REs in complex organization. Their length varies from 4 to 12 kb. Integrated into genomic DNA, LR/ERVs are flanked by 77-bp–3.6-kb-long LTRs. They contain multiple regulatory sequences. LTRs can be found only in the DNA copies of these elements; they appear due to a rather complicated mechanism of LR/ERV reverse transcription.

LR/ERV genome is a single transcriptional unit; all its RNAs are RNA polymerase II transcribed using the promoter located in the element 5′-terminal LTR, and the polyadenylation signal from the 3′ LTR. LRs and ERVs resemble each other in their structure and mobility, except that LRs are unable to move from cell to cell, as they lack an important gene *env* for envelope protein. There are, however, some LRs that codine for Env-like proteins; the distinction between LRs and ERVs is therefore rather dodgy. The mechanisms of LR and ERV reverse transcription are similar and have been reviewed exhaustively in literature (see, e.g., [71, 72]), although some variations may be mentioned [73].

LR/ERVs actively initiate eukaryotic DNA structural rearrangements [74, 75]. For example, the LR family Ty1 accounts for most of the chromosomal rearrangements observed in yeasts [76]. LR/ERVs also have an influence on host genome transcription, mostly because LTRs serve as well-organized regulatory elements. An example is the rat MaLR element, inserted upstream of the promoter region of the gene *CYP2B1* (for cytochrome P450). It serves as a transcriptional repressor, competing with the gene promoter for binding of transcriptional factors (NF-κB and RBP Jκ/CBF1) [77]. Many LTRs have sequences that bind with high affinity to hormone receptors and transcriptional regulators, such as retinoic acid receptor (RAR), the hormone receptors GRE and PRE, transcriptional factors NF-κB, NFOC-1, YY1, SATB1, BMP, Oct-1, Myb, Sp1, Sp3 and others. Indeed, in many cases these proteins and hormones trigger LR/ERV expression [78–86]. For example, the transcriptional factor YY1 (Yin Yang-1), which usually serves as a repressor, specifically activates ERV HTDV/HERV-K transcription in human cells [79]. Another ERV group, HERV-H, is transcriptionally upregulated by Myb in teratocarcinoma lines [84].

Many LR/ERV sequences are involved in normal functioning of the host genome, e.g. ERV-9 LTR enhancer element determines beta-globin gene expression in human eritroid cells [87]. LTR promoters can be recruited by host genes as well [25]. LTRs of retroviral elements belonging to the HERV-E family contribute to the expression of human apolipoprotein C1 (*APOC1*) and endothelin B receptor (*EDNRB*) genes by providing alternative promoters [62, 88]. The LTR promoter of *EDNRB* is used even more often than the normal one, and the *APOC1* LTR-derived promoter initiates transcription of 15% of endothelin receptor mRNAs. Furthermore, LR/ERV sequences in gene introns can be alternatively spliced [12, 25, 58, 89, 90]. In many cases LR/ERV sequences are included in constitutive gene exons and translated [12, 25, 58, 91]. For instance, the fragment of the LR THE-1 element is a part of the human immunoglobulin heavy chain coding sequence [58]. As reported by Mager et al. [92], HERV-H solitary LTRs provide polyadenylation signals for at least two human genes, termed *HHLA2* and *HHLA3*. It is interesting that for baboon orthologous genes other RNA processing signals are used. Three other human genes recruit transcription termination signals from HERV-K-T47D elements [93]. LR/ERVs also take part in antisense regulation of gene expression [94]. Some endogenous retroviruses code for functional proteins. Such is the HERV-W Env protein, called syncytin, which participates in trophoblast fusion during placental development [95]. Besides creating new cellular genes, LR/ERVs also give rise to new chimeric retroelements.

For example, some retroviral genomes consist of different ERV representative fragments. Such a mosaic structure has been reported particularly for human HERV-W and HERV-E.PTN endogenous retroviruses [96–98]. They were most probably created due to recombinations between different ERV sequences: HERV-H and ERV-9 for HERV-W [97], HERV-E and HERV-I for HERV-E.PTN [98]. A recombinational origin was established for simian retrovirus BaEV as well [99], its *gag* and *pol* genes were taken from the PcEV (*Papio cynocephalus* endogenous retrovirus) genome, and *env* was adopted from simian endogenous retrovirus (SERV). Another chimeric LR/ERV representative, Circe element, was generated by a recombination event between LTR-retrotransposon Ulysses and LINE LOA [100]. Moreover, many LR/ERV-created chimeric REs should be assigned to SINEs.

SINEs and processed pseudogenes

SINEs are a very heterogeneous group of REs. Unlike autonomous REs, which have common ancestry at least of their RT gene, SINE representatives appeared many times in evolution irrespective of each other. They usually lack any protein-coding sequences and therefore use ‘exoge-

nous' RT for retroposition. It is generally accepted now that LINEs are used as the source of RT for SINE proliferation in genomes [49, 101]. These latter are thus 'parasites of parasites' (fig. 1). SINE sequences generally contain an oligo(A) tail or, less frequently, another A-rich stretch on their 3′ end [102].

SINEs are widely distributed in eukaryotes and exist, as LINEs do, in plants, in fungi and in vertebrate and invertebrate animals (reviewed by Jagadeeswaran in [103]). However, SINEs have not been found in *Drosophila* DNA or in yeasts [104] (not surprising, as the *Saccharomyces cerevisiae* genome lacks LINEs). SINEs are presented in different hosts by strikingly different copy numbers varying from hundreds to millions. About 12% of the human genome, for example, is occupied by SINEs, mostly Alu elements [5]. At the 5′-end SINEs typically harbour an RNA polymerase III promoter, whose presence implies their origin from aberrantly polyadenylated RNA polymerase III transcripts. A classic example is the mammalian SINE birth from a 7SL RNA copy with internal deletions (rodent B1 [105] and primate Alu elements [106]). These latter were very successful: there are more than 10^6 Alu copies in human DNA, which make up nearly 10% of the genome, roughly 1 Alu per 3 kb [5, 107]. Alu elements retain their transpositional activity, so that ~1 out of every 100–200 human births has a de novo Alu insertion [108].

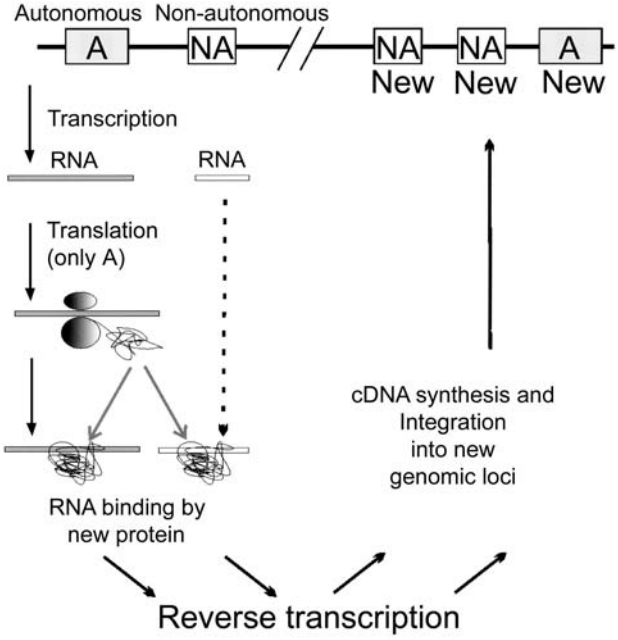

Figure 1. Schematic representation of autonomous and non-autonomous RE life cycle. Non-autonomous elements (SINEs) lack any functional protein-coding genes and, therefore, need foreign proteins for both reverse transcription and integration into the genome. Autonomous REs (LINEs, no cases reported to date for LTR-containing elements) are believed to provide their enzymatic machinery for SINE retroposition.

Alu repeats are ~300-bp-long dimers; their consensus sequence contains two tandem monomer fragments separated by an A-rich stretch [58, 109, 110]. Both monomers are homologous to 7SL RNA [58, 109]. The 5′ fragment of the left monomer includes a transfer RNA (tRNA)-like RNA polymerase III promoter. This promoter most probably appeared due to a 2-bp mutation in the 7SL copy [111]. Besides *Alu*, primate genomes also contain separate left- and right-monomer sequences. Rodent B1 SINEs are two times shorter than 7SL RNA-like elements homologous to left Alu monomer [58, 112, 113]. The Alu/B1 common ancestor most probably appeared before the divergence of rodent and primate lineages [111]. Alu/B1 SINEs were likely originated by a 7SL RNA copy inserted into genomic DNA. This copy mutated to Alu monomer, having an active promoter and internal deletion of 155 bp [58, 111]. The monomer started to settle the genome and one (or some) of its copies integrated just upstream of another similar element. The dimer formed in such a way then spread widely through the primate genome. This mechanism explains the presence in our genome of both Alu, and left and right monomers (termed, respectively, FLAM and FRAM, 'former Alu left/right monomer') [58, 109, 111].

Alu/B1 family members are involved in many cellular functions. Most frequently, they affect expression of neighbouring genes and cause chromosomal rearrangements. As reviewed by Makalowski and Brosius, extended chromosomal regions can be deleted or translocated due to Alu-Alu recombinations [58, 114]. An interesting example is the only currently reported gene distinguishing the human genome from genomes of other hominids, the gene encoding for CMP-N-acetylneuraminic acid hydroxylase, inactivated in *Homo sapiens* because of a similar Alu-mediated deletion [115]. Many inheritable disorders are known now to be due to Alu/B1 de novo insertions in certain genomic loci. For instance, Alu integration in the *APC* gene results in desmoid tumour development [116]; Alu jumping into the gene for factor IX causes hemophilia [43]. Numerous examples of Alu sequence influence on gene expression are also reported elsewhere [117–119].

Alu repeats serve as enhancers (e.g. for the human adenosindesaminase gene), as transcriptional modulators (e.g. for the *c-myc* gene), or as transcriptional silencers (e.g. for the *PCNA* gene); Alu elements are involved in alternative splicing, being in many cases included in protein coding sequences (e.g. for the integrin β1C-2 subunit). They work as insulators (for the keratin 18 gene), and probably have some other functions [25, 58, 109]. An Alu sequence constitutes a 120-bp-long 5′-terminal domain of neuron-specific RNA BC200 found in all higher primates [112]. This RNA of unknown function is expressed at a high level in neurons and is known to be

transferred to dendrites, probably participating in regulation of mRNA translation [58, 112].

An Alu sequence from the translated region of human casein kinase 2 liver isoform mRNA (gene *CK2alpha″*) provides a nuclear localization signal to this enzyme [120]. 5′-UTR-located Alu element influences *ZNF-177* expression at both transcriptional and translational levels [62]. It is important that ~4% of all human mRNAs contain Alu sequences in the 5′-UTRs [62]. One such gene, coding for sex hormone-binding globulin (SHBG), has a (TAAAA)n microsatellite in its 5′-UTR Alu. In different gene alleles the number of TAAAA monomers varies from 6 to 10. Depending on the number of monomers, the gene is expressed with different efficiency [121]. In addition, many human genes contain Alu repeats in protein-coding regions [122]. All these examples are given here just to illustrate the importance of 7SL-like SINEs for normal mammalian genome functioning.

A great number of SINEs (e.g. mammalian MIR [123] and rodent B2 elements [124], the TS element from tobacco [125] and many others) are homologous in their 5′ parts to tRNA sequences and are therefore called tRNA-like SINE. Their 3′-terminal domain is AT rich and resembles that of LINEs. In contrast to 7SL-derived elements, tRNA-like SINEs are widely distributed in eukaryotes [25, 58, 109, 126-128]. The human genome contains ~5 × 10^5 tRNA-like elements (2% of all human DNA [5]), mostly MIR elements [126, 129]. They consist of two regions: a conservative and a variable one (5′ and 3′ segments, respectively). The conservative region includes a tRNA promoter and core domain, whereas the variable part is similar to 3′-terminal sequences of different LINE families [126, 130, 131]. MIR-like elements have also been described in birds, reptiles, amphibians, pisces and in some invertebrates [126].

It is very likely that MIRs descended from retroviral strong-stop DNAs [131, 132]. The core domain of MIRs has conservative regions similar to fragments of lysine tRNA-primed retroviral LTRs. According to the model of MIR and LINE coevolution [131], retroviral strong-stop DNA integrated either into the LINE 3′-terminal part or slightly upstream. Alternatively, LINE 5′-truncated copy could insert near strong-stop DNA. The RE formed could be transcribed by RNA polymerase III and spread through the genome. In eukaryotic genome evolution, retropositionally active LINE lineages were continually replaced by newer LINE families [133]. By taking 3′ termini from different LINEs, MIRs were adapting themselves to such changes (LINE-coded proteins recognize specific sequences on their mRNA 3′ ends) [126, 131]. In papers [131, 134] the authors demonstrate high 3′-terminal sequence identity between SINEs and LINEs from the same genomes: turtle CR1-LINE and PolIII/SINE, piscine HpaI and Rsg-1, bovine Bov-tA and Bov-B and so on. This reflects a SINE molecular parasitism on LINEs. Another obvious case is the SINE from trypanosoma genome, RIME element, which consists of fused short 5′- and 3′-terminal parts of the LINE *ingi* from the same organism [135]: all 'unnecessary' DNA was removed from LINE, only the 5′ sequence with its internal promoter and the 3′ region with RT recognition sites were left.

Like many other REs, tRNA-like SINEs are able to serve as transcriptional regulators for host genes. Some of them provide polyadenylation signals for mammalian genes (e.g. the mouse gene for Glutathione-S-transferase uses the poly(A) site from the B2 element [58]). MIRs are involved in alternative splicing; they were found in several gene exons (for instance, the MIR sequence constitutes the second exon of the gene *ATM* [136]). In addition, B2 elements harbour the RNA polymerase II promoter, which is probably absolutely unnecessary for B2 retroposition. It was demonstrated that this promoter is activated by the transcriptional factor USF. B2 proliferation thus leads to transfer of functional RNA polymerase II promoters to new genomic loci [137].

Notable, the evolution of some SINEs can be precisely traced. For example, the rat ID element progenitor sequence (also termed master gene) survived in the genome and is still active [58]. One day one of the alanine tRNA copies mutated and became the BC1 sequence by acquiring an internal promoter sufficient for RNA polymerase III transcription initiation [58, 138]. BC1 codes for the 152-bp-long RNA which together with BC200 participates in translation regulation in neurons [138]. BC1 copies are now widely distributed in the rat genome and are called ID elements. ID sequences can stabilize mRNA structure (e.g. in the *pIL2* gene); they were also found in some enhancer regions [58].

There are also several examples of the recent formation of chimeric SINE families: MEN retroposon is a chimeric SINE whose 5′ region is homologous to B2 element, whereas the 3′ end was taken from the B1 sequence. There is also a rodent-specific chimeric family created by B1 and ID elements [113]. Some examples of small nuclear RNA-derived SINEs were also reported (the Bm1 element from *Bombyx mori* genome was born by the U1 snRNA [139]).

It should be noted, however, that the term 'short interspersed element' itself does not imply the presence of an RNA polymerase III promoter and origin similar to that described above. Every non-protein-coding short RE of another structure will be assigned to SINEs, which are a heterogeneous RE group. For example, the Cp1 element from chironomids is transcribed by RNA polymerase I [140, 141]; the human SINE-R element was originated by an LTR of HERV-K endogenous retrovirus [142]. Interestingly, SINE-R is included into a more complex retroposon called SVA. The latter consists of three parts: SINE-R, 15–23 tandemly repeated sequences (VNTR) and an Alu sequence (SVA = ***S***INE-R + ***V***NTR + ***A***lu) [143, 144].

Transcription starts with the Alu promoter. SVAs are present in human DNA in several thousand copies. The first SVA element probably appeared in the genome due to integration of several elements into the same genomic locus [144]. There is also another, separate group of SINEs, termed processed pseudogenes.

Processed pseudogenes

Not all sequences created by reverse transcription are transposable elements. The DNA of higher eukaryotes is full of pseudogenes, i.e. transcriptionally silent sequences homologous to known cellular genes; a considerable number of pseudogenes are of RE origin. These elements do not contain introns, in their functional homologs, terminate with an oligo(A) tail and are flanked by variable length short tandem repeats. Such pseudogenes are referred to as processed pseudogenes [26]. They most probably appeared due to reverse transcription of corresponding mRNAs, so that there are generally 1–10 (up to 100 in some cases) pseudogenes for each human gene [58], quite a number of 7SL RNA [145], of snRNAs (U1, U2, U3, U4, U5, U6, U7 [146–148]), of different tRNAs, of ribosomal 5S and 28S RNAs [149], as well as of mitochondrial mRNAs [150]. It is believed that LINE RT is used for processed pseudogene formation.

As long as RNA polymerase II-transcribed genes generally lack any promoter sequences in their mRNAs, corresponding retrogenes are mostly transcriptionally silent and quickly accumulate mutations, thus becoming 'genetic cargo', a raw material for the evolution. There are, however, several cases of functional retrogenes whose transcription is mediated by exogenous preexisting promoters [26]. Alternatively, mutations in pseudogene flanking sequences are able to create new promoters which codes too [58]. The characteristic example is the mouse retrogene *PMSE2b*, which codes the proteasome activator PA28 β subunit. The pseudogene is inserted into the L1 sequence under the control of the LINE promoter. This functional protein-producing retrogene is expressed in mouse tissues in parallel with the 'normal' gene [59]. Another example is the mouse *PHGP* gene for phospholipid hydroperoxide glutathione peroxidase. One of the *PHGP* pseudogenes is expressed tissue specifically from its 5′-adjacent sequence [151]. Two mouse zinc-finger protein genes, *Zfp352* and *Zfp353*, similarly originated from processed pseudogenes [152]. The 'silence' of RNA polymerase III transcript-derived retropseudogenes can be explained by their internal promoter insufficiency to drive transcription: in the case of the 7SL RNA gene, its 5′- adjacent sequence (absent from pseudogene) is absolutely necessary for transcription initiation [145]. In 7SL-like SINEs, the lace of this regulatory sequence is compensated by internal promoter structural alterations.

RNA-RNA recombination as a mechanism for chimeric retrogene formation

As it illustrated above, reverse transcription plays a great role in maintaining genome instability. Although the usual mechanism for RT action is the continuous cDNA synthesis on an RNA template, RT is also able to change templates during reverse transcription. The latter scenario is well known for the retroviruses. The RT jumps from one place on the template to another are necessary for the synthesis of retroviral LTRs. Moreover, usually retroviral particles contain two genomic RNA molecules [17], and the high template switch frequency significantly increases the retroviral genome variety and, therefore, their chances for successful adaptation to environmental changes [153]. According to the 'forced copy-choice model' [154], viral template RNAs have many damaged sites, and the RT has to change templates to complete the full cDNA synthesis. Such jumps often cause RT mistakes, thus increasing the mutation rate [155]. The probability of such RNA-RNA recombinations varies substantially for different regions within the same RNA molecule. The retroviral recombination hot spot is the so-called kissing loop area in the 5′-untranslated region [154]. Such recombination events most probably account for many retroviral genome mosaic structures.

The classic example is the Rous sarcoma virus *v-src* oncogene, derived from the cellular *c-src* gene [156]. The retroviral copy inserted into genomic DNA upstream of the *c-src* gene. The following DNA rearrangement resulted in *c-src* fusion with the retroviral 5′-terminal segment. The chimera was transcribed and packed in virion along with the normal viral genomic RNA molecule. The subsequent RNA recombination created the mosaic viral genome harbouring part of the captured cellular gene. Another example is the human FAM8A pseudogene, identified within the human HERV-K endogenous retroviral sequence [157]. The pseudogene replaces a 1.5-kb-long section of the viral *gag* gene. In both fragments flanking FAM8A insertion, significant sequence similarities between FAM8A and viral DNA were found. This chimera was likely formed due to at least two RNA-RNA recombinations. Finally, Giles et al. recently proposed that similar recombinations could give rise to pseudogene formation of some snRNAs [158].

Another mechanism of chimera creation is template switching during LINE-directed reverse transcription (see fig. 2). Recently, we identified in the human genome several types of chimeric retrogenes generated through a mechanism involving RNA recombination during reverse transcription of cellular RNAs [159]. The chimeras' components were DNA copies of various cellular transcripts: messenger RNAs, ribosomal RNAs, snRNAs, 7SL RNA and transposable elements. All the chimeras identified had common features: the 5′ and 3′ parts of the chimeras

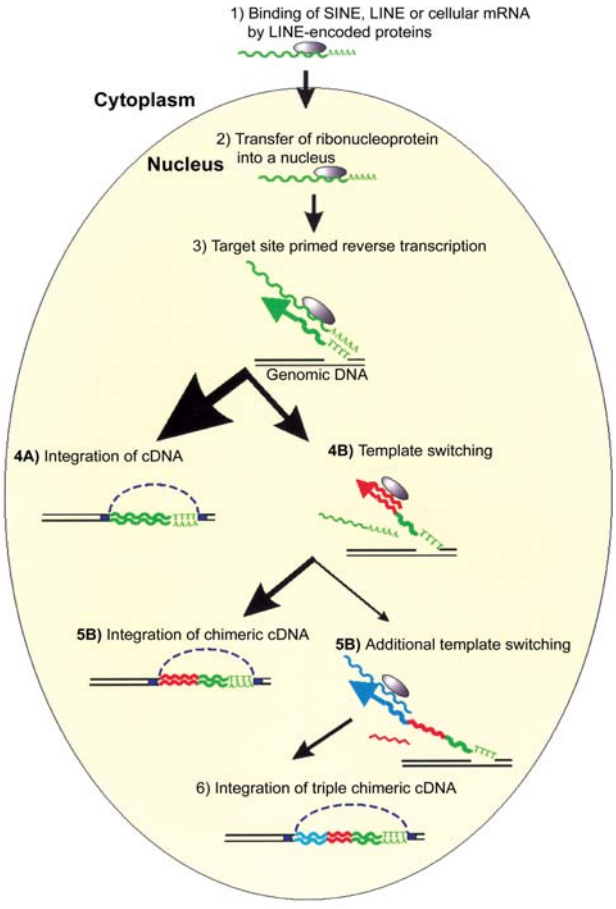

Figure 2. A probable mechanism for double and triple chimera formation by LINE enzymatic machinery. (Step 1) An L1 preintegration complex binds LINE, SINE or the host mRNA in the cytoplasm. (Step 2) The ribonucleoprotein formed is transferred to the nucleus. (Step 3) Reverse transcription of the bound mRNA primed by a genomic DNA single-stranded break within the TTTTAA sequence (target site primed reverse transcription). (Step 4A) Successful integration of the synthesized cDNA copy into genomic DNA. (Step 4B) Switch of templates to another RNA for reverse transcription. (Step 5A) Integration of the double chimera formed into genomic DNA. (Step 5B) The second template switch to another RNA with subsequent DNA reparation mediates formation of a triple chimeric retrogene insertion flanked by short direct repeats and carrying a poly(A) sequence at the 3′ terminus. The normal LINE integration pathway is steps (1), (2), (3), (4A).

were joined directly and had the same orientation, and the chimeras were flanked by direct repeats and carried at their 5′ ends T_2A_4 hexanucleotide or its derivatives with one or two nucleotide substitutions [159, 160]. The latter features (direct repeats and T_2A_4 henanucleotide) are most probably due to the property of LINE1 enzymatic machinery to preferably recognize T_2A_4 sites and to separate newly inserted DNA by short tandem repeats [161]. Newly formed retrogenes thus represent fused DNA copies of different cellular transcripts integrated into the host genome as a single entity.

This mechanism of transcript shuffling was demonstrated to be evolutionarily conserved in mammals [unpublished data]. No chimeras were found in invertebrate, fish and amphibian DNAs, whereas all mammalian genomes under study did contain fused retrogenes. The 5′ parts of chimeras were DNA copies of RNAs having nuclear or both nuclear and cytoplasmic localization, while 3′ parts were copies of cytoplasmic RNAs: mRNAs of cellular genes or of transposable elements. The evolutionary ages of the chimera parts correlated: the younger (i.e. the less divergent from the corresponding consensus sequence) were the 5′ parts; the younger were the 3′ sequences of the chimeras and vice versa. The data obtained show that the phenomenon of chimerization seems to have appeared at least 75 million years ago, before the divergence of primate and rodent ancestors [162], and it is still present today [160]. The further finding of similar mammalian chimeras, consisting of three components, accentuates the potential of the L1 machinery for template switching [unpublished data]. All the chimeras identified were similarly organized: the 5′ parts of triple chimeras were SINEs, middle parts were always U6 snRNA copies and 3′-terminal parts were LINEs. The detection of such retrotranscripts suggests that not only single but also double switches of templates occur in vivo during L1 reverse transcription. Many of the chimeras are expressed in different species, some of them in a tissue-specific manner [159].

Formation of certain L1 families might also involve RNA-RNA recombination due to a template switch after the major part of the L1 mRNA is reverse transcribed, resulting in fusion of the L1 3′ part with an entirely new nucleotide sequence. In particular, 5′-untranslated regions and the first third of the ORF1 of human, murine, rat and rabbit L1 families are known to be not homologous to each other [29].

Interestingly, similar recombinations were also described for L1-L1 [163] and R2-R2 non-LTR retrotransposon chimerization events [164]. In addition, LINE RT jumps probably created recently discovered SINE-like 5S ribosomal RNA (rRNA) retropseudogene from *Dictyostelium* genome [165]. The phenomenon described can thus be considered an evolutionarily conserved mechanism of new retrogene formation by shuffling parts of preexisting expressed sequences.

Concluding remarks

The meaning for evolution of the reverse flow of genetic information has been attracting scientists ever since reverse transcription was discovered. In many genomes REs occupy up to 30–40% of genetic information [26]. Just two RE families, L1 and Alu, constitute about 28% of all human DNA [5]. The fact that both the quantity of different RE families and their copy number were increasing during evolution suggests REs are not only useless ‘junk’ [166] or ‘selfish’ [167, 168] sequences.

Many REs are not simply selfish DNA, as they constitute protein-coding regions of functional genes. REs are also able to repair double-strand DNA breaks [169-171]. During LINE retroposition, the 3′ end of one broken DNA strand serves as a primer for reverse transcription. Such breaks are generally made by LINE endonuclease, but, in principle, any double-strand break can be used by the LINE retropositional machinery. In such case a new LINE copy appears at the site of 'cured' DNA break [170–172]. Notably, cell culture treatment with DNA-damaging reagents results in a dramatic increase of RE transcription [173]. It is possible that in such a way cells recruit LINEs to repair the DNA.

The major part of RE interactions with the cellular genome deal with the regulation of gene expression (fig. 3). Short retroposon sequences are saturated by transcriptional factor binding sites and serve as transcriptional promoters/enhancers for neighbouring genes [174]. The same has been reported for solitary LTRs of LR/ERVs [175, 176].

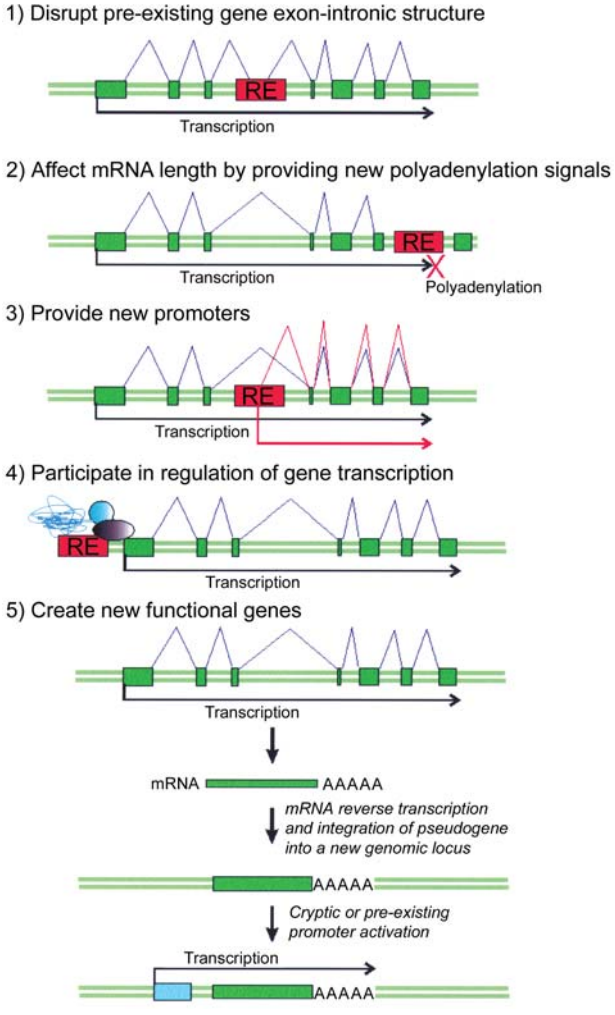

Figure 3. A model illustrating the most commonly used mechanisms of RE influence on host gene functioning.

In addition, REs may compete with gene promoters for the binding of transcriptional factors [177]. Many REs contain splicing enhancers as well as splice donor and/or acceptor sites, which can reshape preexisting gene exon-intronic structure [178]. SINEs, LINEs and LR/ERVs may provide polyadenylation signals, resulting in host gene 3′ truncations [178, 179]. Moreover, RE/cellular mRNA antisense interactions are assumed to play an important role in tight gene expression regulation at both translational [180] and transcriptional [181] levels.

Acknowledgements. The author expresses his deepest gratitude to professor Eugene Sverdlov for critical reading of the manuscript and valuable comments. I also apologize to all authors whose efforts contributed to the current understanding of what retroelements are, but were not cited here due to the rather wide scope of this review and limited space. This work was sponsored by grants from the Russian Foundation for Basic Research and by the Physico-Chemical Biological Program of the Russian Academy of Sciences.

1 McClintock B. (1956) Controlling elements and the gene. Cold Spring Harb. Symp. Quant. Biol. **21:** 197–216
2 Wessler S. R. (1998) Transposable elements and the evolution of gene expression. Symp. Soc. Exp. Biol. **51:** 115–122
3 Kidwell M. G. and Lisch D. (1997) Transposable elements as sources of variation in animals and plants. Proc. Natl. Acad. Sci. USA **94:** 7704–7711
4 Kapitonov V. V. and Jurka J. (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc. Natl. Acad. Sci. USA **100:** 6569–6574
5 Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J. et al. (2001) Initial sequencing and analysis of the human genome. Nature **409:** 860–921
6 Smit A. F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. **9:** 657–663
7 Jurka J. (1998) Repeats in genomic DNA: mining and meaning. Curr. Opin. Struct. Biol. **8:** 333–337
8 Labrador M. and Corces V. G. (1997) Transposable element-host interactions: regulation of insertion and excision. Annu. Rev. Genet. **31:** 381–404
9 Smit A. F. A. (1996) The origin of interspersed repeats in the human genome. Curr. Opin. Genet. Dev. **6:** 743–748
10 Baltimore D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature **226:** 1209–1211
11 Temin H. M. and Mizutani S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature **226:** 1211–1213
12 Urnovitz H. B. and Murphy W. H. (1996) Human endogenous retroviruses: nature, occurrence and clinical implications in human disease. Clin. Microbiol. Rev. **9:** 72–99
13 Leib-Mosch C. and Seifarth W. (1995) Evolution and biological significance of human retroelements. Virus. Genes **11:** 133–145
14 Volff J. N., Korting C., Froschauer A., Sweeney K. and Schartl M. (2001) Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. J. Mol. Evol. **52:** 351–360
15 Finnegan D. J. (1997) Transposable elements: how non-LTR retrotransposons do it. Curr. Biol. **7:** R245–248
16 Schmid C. W. (1998) Does SINE evolution preclude Alu function? Nucleic Acids Res. **26:** 4541–4550
17 Temin H. M. (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc. Natl. Acad. Sci. USA **90:** 6900–6903

18 Craig N. L., Craigie, R. C., Gellert, M. and Lambowitz, A. (2002) Mobile DNA II: ASM Press, Washington
19 Malik H. S., Eickbush T. H. (2001) Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res. **11:** 1187–1197
20 Eickbush T. H. (1997) Telomerase and retrotransposons: which came first? Science **277:** 911–912
21 Pardue M. L. and DeBaryshe P. G. (2003) Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. Annu. Rev. Genet. **37:** 485–511
22 Danilevskaya O. N., Traverse K. L., Hogan N. C., DeBaryshe P. G. and Pardue M. L. (1999) The two *Drosophila* telomeric transposable elements have very different patterns of transcription. Mol. Cell. Biol. **19:** 873–81
23 Zimmerly S., Hausner G. and Wu X. (2001) Phylogenetic relationships among group II intron ORFs. Nucleic Acids Res. **29:** 1238–1250
24 Martinez-Abarca F. and Toro N. (2000) Group II introns in the bacterial world. Mol. Microbiol. **38:** 917–926
25 Boeke J. D. and Stoye, J. P. (1997) Retrotransposons, endogenous retroviruses and the evolution of retroelements. In: Retroviruses, Coffin J. M., Hughes S. H., Varmus H. E. (eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
26 Weiner A. M., Deininger P. L. and Efstratiadis A. (1986) Nonviral retroposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55:** 631–661
27 Smit A. F., Toth G., Riggs A. D. and Jurka J. (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J. Mol. Biol. **246:** 401–417
28 Tchurikov N. A., Gerasimova T. I., Johnson T. K., Barbakar N. I., Kenzior A. L. and Georgiev G. P. (1989) Mobile elements and transposition events in the cut locus of *Drosophila melanogaster*. Mol. Gen. Genet. **219:** 241–248
29 Furano A. V. (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog. Nucleic Acid. Res. Mol. Biol. **64:** 255–294
30 Burke W. D., Malik H. S., Jones J. P. and Eickbush T. H. (1999) The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol. Biol. Evol. **16:** 502–511
31 Yang J. and Eickbush T. H. (1998) RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element. Mol. Cell. Biol. **18:** 3455–3465
32 Moran J. V. (1999) Human L1 retrotransposition: insights and peculiarities learned from a cultured cell retrotransposition assay. Genetica **107:** 39–51
33 Brouha B., Schustak J., Badge R. M., Lutz-Prigge S., Farley A. H., Moran J. V. et al. (2003) Hot L1s account for the bulk of retrotransposition in the human population. Proc. Natl. Acad. Sci. USA **100:** 5280–5285
34 Buzdin A., Ustyugova S., Gogvadze E., Lebedev Y., Hunsmann G. and Sverdlov E. (2003) Genome-wide targeted search for human specific and polymorphic L1 integrations. Hum. Genet. **112:** 527–533
35 Goodier J. L., Ostertag E. M., Du K. and Kazazian H. H. Jr (2001) A novel active L1 retrotransposon subfamily in the mouse. Genome Res. **11:** 1677–1685
36 Boissinot S., Chevret P. and Furano A. V. (2000) L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol. Biol. Evol. **17:** 915–928
37 Priimagi A. F., Mizrokhi L. J. and Ilyin Y. V. (1988) The *Drosophila* mobile element jockey belongs to LINEs and contains coding sequences homologous to some retroviral proteins. Gene **70:** 253–262
38 Udomkit A., Forbes S., Dalgleish G. and Finnegan D. J. (1995) BS a novel LINE-like element in *Drosophila melanogaster*. Nucleic Acids Res. **23:** 1354–1358
39 Levis R. W., Ganesan R., Houtchens K., Tolar L. A. and Sheen F. M. (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. Cell **75:** 1083–1093
40 Pimpinelli S., Berloco M., Fanti L., Dimitri P., Bonaccorsi S., Marchetti E. et al. (1995) Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. Proc. Natl. Acad. Sci. USA **92:** 3804–3808
41 Petrov D. A., Lozovskaya E. R. and Hartl D. L. (1996) High intrinsic rate of DNA loss in *Drosophila*. Nature **384:** 346–349
42 Finnegan D. J. (1992) Transposable elements. In: The Genome of *Drosophila melanogaster*. Lindsley D. L. and Zimm G. G. (eds), Academic Press
43 Kazazian H. H. (1998) Mobile elements and disease. Curr. Opin. Genet. Dev. **8:** 343–350
44 Mizrokhi L. J., Georgieva S. G. and Ilyin Y. V. (1988) jockey, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell **54:** 685–691
45 Birnstiel M. L., Busslinger M. and Strub K. (1985) Transcription termination and 3′ processing: the end is in site! Cell **41:** 349–359
46 Sewell E. and Kinsey J. A. (1996) Tad, a Neurospora LINE-like retrotransposon exhibits a complex pattern of transcription. Mol. Gen. Genet. **252:** 137–145
47 Agamalian N. S., Arkhipova I. R. and Surkov S. A. Il'in Iu. V. (1996) Regulating polyadenylation of jockey mobile genetic element transcripts belonging to the LINE class, in *Drosophila* cell culture. Mol. Biol. (Mosk.) **30:** 818–828
48 Perepelitsa-Belancio V. and Deininger P. (2003) RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat. Genet. **35:** 363–366
49 Eickbush T. H. (1992) Transposing without ends: the non-LTR retrotransposable elements. New Biol. **4:** 430–440
50 Kazazian H. H. Jr and Moran J. V. (1998) The impact of L1 retrotransposons on the human genome. Nat. Genet. **19:** 19–24
51 Burwinkel B. and Kilimann M. W. (1998) Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. J. Mol. Biol. **277:** 513–517
52 Segal Y., Peissel B., Renieri A., de Marchi M., Ballabio A., Pei Y. et al. (1999) LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis. Am. J. Hum. Genet. **64:** 62–69
53 Kazazian H. H. Jr (2000) Genetics. L1 retrotransposons shape the mammalian genome. Science **289:** 1152–1153
54 Van de Water N., Williams R., Ockelford P. and Browett P. (1998) A 20.7 kb deletion within the factor VIII gene associated with LINE-1 element insertion. Thromb. Haemost. **79:** 938–942
55 Narita N., Nishio H., Kitoh Y., Ishikawa Y., Minami R., Nakamura H. 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
56 McNaughton J. C., Hughes G., Jones W. A., Stockwell P. A., Klamut H. J. and Petersen G. B. (1997) The evolution of an intron: analysis of a long, deletion-prone intron in the human dystrophin gene. Genomics **40:** 294–304
57 Eickbush T. H. and Furano A. V. (2002) Fruit flies and humans respond differently to retrotransposons. Curr. Opin. Genet. Dev. **12:** 669–674
58 Brosius J. (1999) RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene **238:** 115–134
59 Zaiss D. M. and Kloetzel P. M. (1999) A second gene encoding the mouse proteasome activator PA28beta subunit is part of a LINE1 element and is driven by a LINE1 promoter. J. Mol. Biol. **287:** 829–835

60 King L. M. and Francomano C. A. (2001) Characterization of a human gene encoding nucleosomal binding protein NSBP1. Genomics **71:** 163–173
61 Rothbarth K., Hunziker A., Stammer H. and Werner D. (2001) Promoter of the gene encoding the 16 kDa DNA-binding and apoptosis-inducing C1D protein. Biochim. Biophys. Acta **1518:** 271–275
62 Landry J. R., Medstrand P. and Mager D. L. (2001) Repetitive elements in the 5′ untranslated region of a human zinc-finger gene modulate transcription and translation efficiency. Genomics **76:** 110–116
63 Goodier J. L., Ostertag E. M. and Kazazian H. H. Jr (2000) Transduction of 3′-flanking sequences is common in L1 retrotransposition. Hum. Mol. Genet. **9:** 653–657
64 Pickeral O. K., Makalowski W., Boguski M. S. and Boeke J. D. (2000) Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res. **10:** 411–415
65 Holmes S. E., Dombroski B. A., Krebs C. M., Boehm C. D. and Kazazian H. H. Jr (1994) A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nat. Genet. **7:** 143–148
66 Martin S. L. (1995) Characterization of a LINE-1 cDNA that originated from RNA present in ribonucleoprotein particles: implications for the structure of an active mouse LINE-1. Gene **153:** 261–266
67 Rozmahel R., Heng H. H., Duncan A. M., Shi X. M., Rommens J. M. and Tsui L. C. (1997) Amplification of CFTR exon 9 sequences to multiple locations in the human genome. Genomics **45:** 554–561
68 Nigumann P., Redik K., Matlik K. and Speek M. (2002) Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics **79:** 628–634
69 Speek M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol. Cell. Biol. **21:** 1973–1985
70 Ostertag E. M. and Kazazian H. H. Jr (2001) Biology of mammalian L1 retrotransposons. Annu. Rev. Genet. **35:** 501–538
71 Wilhelm M., Wilhelm F. X. (2001) Reverse transcription of retroviruses and LTR retrotransposons. Cell. Mol. Life Sci. **58:** 1246–1262
72 Sverdlov E. D. (2000) Retroviruses and primate evolution. Bioessays **22:** 161–171
73 Cheng Z. and Menees T. M. (2004) RNA branching and debranching in the yeast retrovirus-like element Ty1. Science **303:** 240–243
74 Hughes J. F. and Coffin J. M. (2001) Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. Nat. Genet. **29:** 487–489
75 Sverdlov E. D. (1998) Perpetually mobile footprints of ancient infections in human genome. FEBS Lett. **428:** 1–6
76 Umezu K., Hiraoka M., Mori M. and Maki H. (2002) Structural analysis of aberrant chromosomes that occur spontaneously in diploid *Saccharomyces cerevisiae*: retrotransposon Ty1 plays a crucial role in chromosomal rearrangements. Genetics **160:** 97–110
77 Lee S. H., Wang X. and DeJong J. (2000) Functional interactions between an atypical NF-kappaB site from the rat CYP2B1 promoter and the transcriptional repressor RBP-Jkappa/CBF1. Nucleic Acids Res. **28:** 2091–2098
78 Chene L., Nugeyre M. T., Barre-Sinoussi F. and Israel N. (1999) High-level replication of human immunodeficiency virus in thymocytes requires NF-kappaB activation through interaction with thymic epithelial cells. J. Virol. **73:** 2064–2073
79 Knossl M., Lower R. and Lower J. (1999) Expression of the human endogenous retrovirus HTDV/HERV-K is enhanced by cellular transcription factor YY1. J. Virol. **73:** 1254–1261
80 Schon U., Seifarth W., Baust C., Hohenadl C., Erfle V. and Leib-Mosch C. (2001) Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. Virology **279:** 280–291
81 Sjottem E., Anderssen S. and Johansen T. (1996) The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3′ to the TATA box. J. Virol. **70:** 188–198
82 Boronat S., Richard-Foy H. and Pina B. (1997) Specific deactivation of the mouse mammary tumor virus long terminal repeat promoter upon continuous hormone treatment. J. Biol. Chem. **272:** 21803–1810
83 Caricasole A., Ward-van Oostwaard D., Mummery C. and van den Eijnden-van Raaij A. (2000) Bone morphogenetic proteins and retinoic acid induce human endogenous retrovirus HERV-K expression in NT2D1 human embryonal carcinoma cells. Dev. Growth Differ. **42:** 407–411
84 de Parseval N., Alkabbani H. and Heidmann T. (1999) The long terminal repeats of the HERV-H human endogenous retrovirus contain binding sites for transcriptional regulation by the Myb protein. J. Gen. Virol. **80 (Pt 4):** 841–845
85 Inoue D., Santiago P., Horne W. C. and Baron R. (1997) Identification of an osteoclast transcription factor that binds to the human T cell leukemia virus type I-long terminal repeat enhancer element. J. Biol. Chem. **272:** 25386–25393
86 Akopov S. B., Nikolaev L. G., Khil P. P., Lebedev Y. B. and Sverdlov E. D. (1998) Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. FEBS Lett. **421:** 229–233
87 Lania L., Di Cristofano A., Strazzullo M., Pengue G., Majello B. and La Mantia G. (1992) Structural and functional organization of the human endogenous retroviral ERV9 sequences. Virology **191:** 464–468
88 Landry J. R. and Mager D. L. (2003) Functional analysis of the endogenous retroviral promoter of the human endothelin B receptor gene. J. Virol. **77:** 7459–7466
89 Kowalski P. E., Freeman J. D. and Mager D. L. (1999) Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes. Genomics **57:** 371–379
90 Feuchter-Murthy A. E., Freeman J. D. and Mager D. L. (1993) Splicing of a human endogenous retrovirus to a novel phospholipase A2 related gene. Nucleic Acids Res. **21:** 135–143
91 Kapitonov V. V. and Jurka J. (1999) The long terminal repeat of an endogenous retrovirus induces alternative splicing and encodes an additional carboxy-terminal sequence in the human leptin receptor. J. Mol. Evol. **48:** 248–251
92 Mager D. L., Hunter D. G., Schertzer M. and Freeman J. D. (1999) Endogenous retroviruses provide the primary polyadenylation signal for two new human genes. Genomics **59:** 255–263
93 Baust C., Seifarth W., Germaier H., Hehlmann R. and Leib Mosch C. (2000) HERV-K-T47D-Related long terminal repeats mediate polyadenylation of cellular transcripts. Genomics **66:** 98–103
94 Schneider P. M., Witzel-Schlomp K., Rittner C. and Zhang L. (2001) The endogenous retroviral insertion in the human complement C4 gene modulates the expression of homologous genes by antisense inhibition. Immunogenetics **53:** 1–9
95 Frendo J. L., Olivier D., Cheynet V., Blond J. L., Bouton O., Vidaud M. et al. (2003) Direct involvement of HERV-W Env glycoprotein in human trophoblast cell fusion and differentiation. Mol. Cell. Biol. **23:** 3566–3574
96 Blond J. l., Beseme F., Duret L., Bouton O., Bedin F., Perron H. et al. (1999) Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. J. Virol. **73:** 1175–1185
97 Kim H. S., Takenaka O. and Crow T. J. (1999) Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. J. Gen. Virol. **80 (Pt 10):** 2613–2619

98 Schulte A. M. and Wellstein A. (1998) Structure and phylogenetic analysis of an endogenous retrovirus inserted into the human growth factor gene pleiotrophin. J. Virol. **72:** 6065–6072

99 Mang R., Goudsmit J. and van der Kuyl A. C. (1999) Novel endogenous type C retrovirus in baboons: complete sequence, providing evidence for baboon endogenous virus gag-pol ancestry. J. Virol. **73:** 7021–7026

100 Losada A., Abad J. P., Agudo M. and Villasante A. (1999) The analysis of Circe, an LTR retrotransposon of *Drosophila melanogaster*, suggests that an insertion of non-LTR retrotransposons into LTR elements can create chimeric retroelements. Mol. Biol. Evol. **16:** 1341–1346

101 Dewannieux M., Esnault C., and Heidmann T. (2003) LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. **35:** 41–48

102 Nadir E., Margalit H., Gallily T. and Ben-Sasson S. A. (1996) Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proc. Natl. Acad. Sci. USA **93:** 6470–6475

103 Jagadeeswaran P., Forget B. G. and Weissman S. M. (1981) Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? Cell **26:** 141–142

104 Andersson A. C., Merza M., Venables P., Ponten F., Sundstrom J., Cohen M. et al. (1996) Elevated levels of the endogenous retrovirus ERV3 in human sebaceous glands. Journal Investigative Dermatology **106:** 125–128

105 Labuda D., Sinnett D., Richer C., Deragon J. M. and Striker G. (1991) Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved. J. Mol. Evol. **32:** 405–414

106 Ullu E. and Tschudi C. (1984) Alu sequences are processed 7SL RNA genes. Nature **312:** 171–172

107 Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G. et al. (2001) The sequence of the human genome. Science **291:** 1304–1351

108 Deininger P. L. and Batzer M. A. Mammalian retroelements. Genome Res. (2002) **12:** 1455–1465

109 Chu W. M., Ballard R., Carpick B. W., Williams B. R. and Schmid C. W. (1998) Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. Mol. Cell. Biol. **18:** 58–68

110 Hohenadl C., Leib-Mosch C., Hehlmann R. and Erfle V. (1996) Biological significance of human endogenous retroviral sequences. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. **13 Suppl. 1:** S268–273

111 Quentin Y. (1994) Emergence of master sequences in families of retroposons derived from 7sl RNA. Genetica **93:** 203–215

112 Skryabin B. V., Kremerskothen J., Vassilacopoulou D., Disotell T. R., Kapitonov V. V., Jurka J. et al. (1998) The BC200 RNA gene and its neural expression are conserved in Anthropoidea (Primates). J. Mol. Evol. **47:** 677–685

113 Kramerov D. A. and Vassetzky N. S. (2001) Structure and origin of a novel dimeric retroposon B1-diD. J. Mol. Evol. **52:** 137–143

114 Makalowski W., Mitchell G. A. and Labuda D. (1994) Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet. **10:** 188–193

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

116 Halling K. C., Lazzaro C. R., Honchel R., Bufill J. A., Powell S. M., Arndt C. A. et al. (1999) Hereditary desmoid disease in a family with a germline Alu I repeat mutation of the APC gene. Hum. Hered. **49:** 97–102

117 Le Goff W., Guerin M., Chapman M. J. and Thillet J. (2003) A CYP7A promoter binding factor site and Alu repeat in the distal promoter region are implicated in regulation of human CETP gene expression. J. Lipid. Res. **44:** 902–910

118 Ganguly A., Dunbar T., Chen P., Godmilow L. and Ganguly T. (2003) Exon skipping caused by an intronic insertion of a young Alu Yb9 element leads to severe hemophilia A. Hum. Genet. **113:** 348–352

119 Vila M. R., Gelpi C., Nicolas A., Morote J., Schwartz S. Jr, Schwartz S. et al. (2003) Higher processing rates of Alu-containing sequences in kidney tumors and cell lines with overexpressed Alu-mRNAs. Oncol. Rep. **10:** 1903–1909

120 Hilgard P., Huang T., Wolkoff A. W. and Stockert R. J. (2002) Translated Alu sequence determines nuclear localization of a novel catalytic subunit of casein kinase 2. Am. J. Physiol. Cell. Physiol. **283:** C472–483

121 Hogeveen K. N., Talikka M. and Hammond G. L. (2001) Human sex hormone-binding globulin promoter activity is influenced by a (TAAAA)n repeat element within an Alu sequence. J. Biol. Chem. **276:** 36383–36390

122 Nekrutenko A. and Li W. H. (2001) Transposable elements are found in a large number of human protein-coding genes. Trends Genet. **17:** 619–621

123 Broude N. E., Storm N., Malpel S., Graber J. H., Lukyanov S., Sverdlov E. et al. (1999) PCR based targeted genomic and cDNA differential display. Genet. Anal. **15:** 51–63

124 Daniels G. R. and Deininger P. L. (1985) Repeat sequence families derived from mammalian tRNA genes. Nature **317:** 819–822

125 Yoshioka Y., Matsumoto S., Kojima S., Ohshima K., Okada N. and Machida Y. (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc. Natl. Acad. Sci. USA **90:** 6562–6566

126 Gilbert N. and Labuda D. (1999) CORE-SINEs: eukaryotic short interspersed retroposing elements with common sequence motifs. Proc. Natl. Acad. Sci. USA **96:** 2869–2874

127 Gilbert N., Arnaud P., Lenoir A., Warwick S. I., Picard G. and Deragon J. M. (1997) Plant S1 SINEs as a model to study retroposition. Genetica **100:** 155–60

128 Mayorov V. I., Rogozin I. B., Elisaphenko E. A. and Adkison L. R. (2000) B2 elements present in the human genome. Mamm. Genome **11:** 177–179

129 Matassi G., Labuda D. and Bernardi G. (1998) Distribution of the mammalian-wide interspersed repeats (MIRs) in the isochores of the human genome. FEBS Lett. **439:** 63–65

130 Gilbert N. and Labuda D. (2000) Evolutionary inventions and continuity of CORE-SINEs in mammals. J. Mol. Biol. **298:** 365–377

131 Ohshima K., Hamada M., Terai Y. and Okada N. (1996) The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. Mol. Cell. Biol. **16:** 3756–3764

132 Gurskaya N. G., Diatchenko L., Chenchik A., Siebert P. D., Khaspekov G. L., Lukyanov K. A. et al. (1996) Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. Anal. Biochem. **240:** 90–97

133 Volff J. N., Korting C., Schartl M. (2000) Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. Mol. Biol. Evol. **17:** 1673–1684

134 Okada N., Hamada M. (1997) The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: a new example from the bovine genome. J. Mol. Evol. **44 Suppl. 1:** S52–56

135 Kimmel B. E., ole-MoiYoi O. K., Young J. R. (1987) Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. Mol. Cell. Biol. **7:** 1465–1475

136 Platzer M., Rotman G., Bauer D., Uziel T., Savitsky K., Bar-Shira A. et al. (1997) Ataxia-telangiectasia locus: sequence

analysis of 184 kb of human genomic DNA containing the entire ATM gene. Genome Res. **7:** 592–605
137 Ferrigno O., Virolle T., Djabari Z., Ortonne J. P., White R. J., Aberdam D. (2001) Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. Nat. Genet. **28:** 77–81
138 Shen M. R., Brosius J., Deininger P. L. (1997) BC1 RNA, the transcript from a master gene for ID element amplification, is able to prime its own reverse transcription. Nucleic Acids Res. **25:** 1641–1648
139 Herrera R. J., Wang J. (1991) Evidence for a relationship between the *Bombyx mori* middle repetitive Bm1 sequence family and U1 snRNA. Genetica **84:** 31–37
140 He H., Rovira C., Recco-Pimentel S., Liao C., Edstrom J. E. (1995) Polymorphic SINEs in chironomids with DNA derived from the R2 insertion site. J. Mol. Biol. **245:** 34–42
141 Bertioli D. J., Schlichter U. H., Adams M. J., Burrows P. R., Steinbiss H. H., Antoniw J. F. (1995) An analysis of differential display shows a strong bias towards high copy number mRNAs. Nucleic Acids Res. **23:** 4520–4523
142 Ono M., Kawakami M., Takezawa T. (1987) A novel human nonviral retroposon derived from an endogenous retrovirus. Nucleic Acids Res. **15:** 8725–8737
143 Broude N. E., Chandra A and Smith C. L. (1997) Differential display of genome subsets containing specific interspersed repeats. Proc. Natl. Acad. Sci. USA **94:** 4548–4553
144 Shen L., Wu L. C., Sanlioglu S., Chen R., Mendoza A. R., Dangel A. W. et al. (1994) Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon and breakpoint of gene duplication. J. Biol. Chem. **269:** 8466–8476
145 Ullu E., Weiner A. M. (1985) Upstream sequences modulate the internal promoter of the human 7SL RNA gene. Nature **318:** 371–374
146 Kristo P., Tsai M. J., O'Malley B. W. Characterization of three chicken pseudogenes for U1 RNA. DNA (1984) **3:** 281–286
147 Bark C., Pettersson U. (1989) Nucleotide sequence and organization of full length human U4 RNA pseudogenes. Gene **80:** 385–389
148 Soldati D., Schumperli D. (1990) Structures of four human pseudogenes for U7 small nuclear RNA. Gene **95:** 305–306
149 Ji H., Liu Y. E., Jia T., Wang M., Liu J., Xiao G. et al. (1997) Identification of a breast cancer-specific gene, BCSG1, by direct differential cDNA sequencing. Cancer Res. **57:** 759–764
150 Tourmen Y., Baris O., Dessen P., Jacques C., Malthiery Y., Reynier P. (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes. Genomics **80:** 71–77
151 Boschan C., Borchert A., Ufer C., Thiele B. J., Kuhn H. (2002) Discovery of a functional retrotransposon of the murine phospholipid hydroperoxide glutathione peroxidase: chromosomal localization and tissue-specific expression pattern. Genomics **79:** 387–394
152 Chen H. H., Liu T. Y., Huang C. J., Choo K. B. (2002) Generation of two homologous and intronless zinc-finger protein genes, zfp352 and zfp353, with different expression patterns by retrotransposition. Genomics **79:** 18–23
153 Kandel E. S., Nudler E. (2002) Template switching by RNA polymerase II in vivo. Evidence and implications from a retroviral system. Mol. Cell. **10:** 1495–1502
154 Hu W. S., Rhodes T., Dang Q., Pathak V. (2003) Retroviral recombination: review of genetic analyses. Front. Biosci. **8:** d143–155
155 Temin H. M. (1993) A proposal for a new approach to a preventive vaccine against human immunodeficiency virus. Proc. Natl. Acad. Sci. USA **90:** 4419–4420
156 Swanstrom R., Parker R. C., Varmus H. E., Bishop J. M. (1983) Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. Proc. Natl. Acad. Sci. USA **80:** 2519–2523
157 Jamain S., Girondot M., Leroy P., Clergue M., Quach H., Fellous M. et al. (2001) Transduction of the human gene FAM8A1 by endogenous retrovirus during primate evolution. Genomics **78:** 38–45
158 Giles K. E., Caputi M., Beemon K. L. (2004) Packaging and reverse transcription of snRNAs by retroviruses may generate pseudogenes. RNA **10:** 299–307
159 Buzdin A., Gogvadze E., Kovalskaya E., Volchkov P., Ustyugova S., Illarionova A. et al. (2003) The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination. Nucleic Acids Res. **31:** 4385–4390
160 Buzdin A., Ustyugova S., Gogvadze E., Vinogradova T., Lebedev Y., Sverdlov E. (2002) A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3′ terminus of l1. Genomics **80:** 402–406
161 Jurka J. (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. USA **94:** 1872–1877
162 Mouse genome sequencing consortium. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature **420:** 520–562
163 Hayward B. E., Zavanelli M., Furano A. V. (1997) Recombination creates novel L1 (LINE-1) elements in *Rattus norvegicus*. Genetics **146:** 641–654
164 Bibillo A., Eickbush T. H. (2002) The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. J. Mol. Biol. **316:** 459–473
165 Szafranski K., Dingermann T., Glockner G., Winckler T. (2003) Template jumping by a LINE reverse transcriptase has created a SINE-like 5S rRNA retropseudogene in *Dictyostelium*. Mol. Genet. Genomics
166 Ono S. (1972) So much 'junk' DNA in our genome. Brookhaven Symp. Biol. **23:** 366–370
167 Orgel L. E., Crick F. H. (1980) Selfish DNA: the ultimate parasite. Nature **284:** 604–607
168 Hickey D. A. (1982) Selfish DNA: a sexually-transmitted nuclear parasite. Genetics **101:** 519–531
169 Teng S. C., Kim B., Gabriel A. (1996) Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. Nature **383:** 641–644
170 Moore J. K., Haber J. E. (1996) Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature **383:** 644–646
171 Morrish T. A., Gilbert N., Myers J. S., Vincent B. J., Stamato T. D., Taccioli G. E. et al. (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat. Genet. **31:** 159–165
172 Biessmann H., Valgeirsdottir K., Lofsky A., Chin C., Ginther B., Levis R. W. et al. (1992) HeT-A, a transposable element specifically involved in 'healing' broken chromosome ends in *Drosophila melanogaster*. Mol. Cell. Biol. **12:** 3910–3918
173 Hagan C. R., Rudin C. M. (2002) Mobile genetic element activation and genotoxic cancer therapy: potential clinical implications. Am. J. Pharmacogenomics **2:** 25–35
174 Kazakov V. I., Tomilin N. V. (1996) Increased concentration of some transcription factor binding sites in human retroposons of the Alu family. Genetica **97:** 15–22
175 Banville D., Boie Y. (1989) Retroviral long terminal repeat is the promoter of the gene encoding the tumor-associated calcium-binding protein oncomodulin in the rat. J. Mol. Biol. **207:** 481–490
176 Friesen P. D., Rice W. C., Miller D. W., Miller L. K. (1986) Bidirectional transcription from a solo long terminal repeat of the retrotransposon TED: symmetrical RNA start sites. Mol. Cell. Biol. **6:** 1599–1607
177 Conte C., Dastugue B., Vaury C. (2002) Promoter competition as a mechanism of transcriptional interference mediated by retrotransposons. EMBO J. **21:** 3908–3916

178 Michel D., Chatelain G., Mauduit C., Benahmed M., Brun G. (1997) Recent evolutionary acquisition of alternative pre-mRNA splicing and 3′ processing regulations induced by intronic B2 SINE insertion. Nucleic Acids Res. **25:** 3228–3234
179 Krane D. E., Hardison R. C. (1990) Short interspersed repeats in rabbit DNA can provide functional polyadenylation signals. Mol. Biol. Evol. **7:** 1–8
180 Bladon T. S., McBurney M. W. (1991) The rodent B2 sequence can affect expression when present in the transcribed region of a reporter gene. Gene **98:** 259–263
181 Djikeng A., Shi H., Tschudi C., Ullu E. (2001) RNA interference in *Trypanosoma brucei:* cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. RNA **7:** 1522–1530