Chapter 23
Recent Inserts of Transposable Elements Affect Structure and Functions of Human Genome

Anton Buzdin, Maria Suntsova, Olga Bantysh, Elena Aleksandrova, Anastasia Zabolotneva, Elena Gogvadze, and Nurshat Gaifullin

Abstract Transposable elements (TEs) are selfish fragments of DNA able to reproduce themselves into the host genomes. TEs typically occupy ~40–50% of the mammalian genomes. In our studies, we focus on evolutionary recent TE inserts that appeared in the DNA of human ancestor lineage after divergence with the chimpanzee ancestry, *i.e.* less than ~6 million years ago. These human specific elements (hsTEs) represent only a minor fraction of the whole TE cargo of the human genome. hsTEs are represented by the four families called HERV-K(HML-2), L1, Alu and SVA. The number of human specific copies for HERV-K(HML-2), L1, Alu and SVA families is approx. 150, 1,200, 5,500 and 860 copies per genome, respectively. Taken together, hsTEs shape ~6.4 megabases of human DNA, which is about 6-times lower than what is occupied by the human specific simple nucleotide polymorphisms, and 23-times smaller than the overall length of human specific deletions and duplications. However, although modest in terms of genomic proportion, hsTEs should be regarded as the perspective candidates for being molecular genetic agents of human speciation. Unlike most of random mutations and duplications, each novel insert of hsTE has provided to the recipient genomic locus a set of functional transcriptional factor binding sites positively selected during the TE evolution. For example, clusters of novel inserts of Alu elements may serve as CpG islets, SVA elements provide functional splice sites and polyadenylation signals, whereas L1 and HERV-K(HML-2) elements donate enhancers, promoters, splice sites and polyadenylation signals. Significant proportion of the human-specific genomic deletions,

A. Buzdin (✉) • M. Suntsova • O. Bantysh • E. Aleksandrova • A. Zabolotneva • E. Gogvadze
Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry,
Miklukho-Maklaya 16/10, Moscow 117997, Russia
e-mail: bu3din@mail.ru

N. Gaifullin
Lomonosov Moscow State University, Lomonosovsky pr. 31-5,
Moscow 119192, Russia

C.E. Mothersill et al. (eds.), *Radiobiology and Environmental Security*,
NATO Science for Peace and Security Series C: Environmental Security,
DOI 10.1007/978-94-007-1939-2_23, © Springer Science+Business Media B.V. 2012

duplications and translocations has been also generated due to ectopic recombinations between the different individual TE inserts. Among the other, we report for the first time a detailed functional characteristics of the HERV-K(HML-2) hsTEs done at the genome-wide level. We have identified 65 active *in vivo* human specific promoters contributed by these elements. We also identified three cases of the hsTE-mediated human specific transcriptional regulation of functional protein-coding genes taking part in brain development during embryogenesis. We found ~180 human specific polyadenylation signals transferred by the SVA elements into the introns of known functional genes. Scaling of these data to the total number of the hsTEs predicts that hundreds of human genes are regulated by these elements. Finally, we discovered the first exclusively human specific TE family, represented by ~80 members formed by a combination of a part of a CpG islet of human gene *MAST2* ansd of the 3′-terminal part of an SVA retrotransposon. According to our estimates, this family, termed CpG-SVA, was far more active than the ancestral SVA family. Our data indicate that *MAST2* regulatory sequence was recruited during the evolution to provide effective CpG-SVA transcription in human testicular germ-line cells.

Keywords Human evolution • Genetic instability • Transposable elements • Human specific promoters • Antisense transcripts • Regulation of gene expression • Brain development • Hybrid family of retrotransposons

23.1 Introduction

23.1.1 Recent Evolution of the Human Genome

23.1.1.1 Major Genetic Differences Between Humans and Chimpanzees

Understanding of the genetic basis that accounts for the obvious differences in phenotypes of humans and their closest relatives, chimpanzees, is one of the most interesting tasks of modern life sciences. This task is also challenging, mostly due to strikingly high similarities in their genome structure and organization [25, 26, 106, 117]. Indeed, an average divergence between human and chimpanzee DNAs is about 1.24% [36], being as low as only 0.5% in protein coding regions [54]. Human and chimpanzee ancestor lineages diverged relatively recently in evolution, approximately 6 million years ago [57]. At present, we still don't know exactly what genetic traits make us humans, but a number of functionally important differences between human and great ape genomes have been identified. In general, they can be classified into the four major groups:

1. diverse chromosomal organization, including deletions, inversions, duplications and translocations
2. variations in copy number, genomic localization and functional status among the pre-existing common sequences

3. differences in protein coding regions
4. lineage-specific genomic insertions of transposable elements (TEs).

23.1.1.2 Non-TE Differences

Millions of mutated loci, mostly single nucleotide polymorphisms (SNPs), are known to be polymorphic in humans. However, they should not be considered here due to the lack of a functional role in human speciation for these sequences, which are presented only in a fraction of human population. Human and chimpanzee genomes, both approximately $3 \cdot 10^9$ base pairs in size, share ~98.8% sequence identity [36], thus making identification of functional human-specific sequences finding the needle in a haystack. Theoretically, recent success in human and chimpanzee DNA sequencing projects [25, 116] has provided an instrument for the direct comparison of genomes with the subsequent association of genomic changes with interspecies differences at the level of protein expression. However, in practice the chimpanzee genome draft, currently available in public databases, is not sufficiently accurate for such a comprehensive study [20], although many successful large-scale bioinformatical screenings have been performed [56, 81, 110, 111]. A plenty of lineage-specific substitutions, deletions, insertions, duplications, expressed pseudogenes, anonymous RNAs, transposable elements have been identified nowadays, but the omnibus study still remains to be done.

Cytogenetic differences. The comparison of human and African great ape karyotypes using fluorescent in situ hybridization has revealed the most important lineage-specific distinction, which is the fusion in human lineage of two ancestral chromosomes (human chromosome 2), corresponding to chimpanzee chromosomes 12 and 13 [141]. Other major points are numerous changes in centromeric and telomeric regions [62, 97, 102, 114] and lineage-specific rearrangements and amplifications of several gene families in non-recombining parts of Y chromosome [47]. Apart from translocations, insertions and deletions have together given rise to at least 150 Mb of genomic DNA sequence that is either present or absent in humans as compared to chimpanzees, according to the recent estimation by Kehrer-Sawatzki and Cooper [72]. Interestingly, mostly chromosome ends were the "hot spots" of recent genome evolution [68].

Emerging or inactivation of functional genes. Few functional genes are known to distinguish human and ape DNA. First of all, this is the functional deletion of an exon within the protein coding sequence of human gene *CMP* for syalic acid hydroxylase. Mutation caused by the human specific insertion of an Alu retroelement into 92 bp-long *CMP* exon, disrupted normal open reading frame for this enzyme and resulted in the lack of N-glycolyl neuraminic acid (Neu5Gc) on a surface of human cell membranes [22, 65]. Neu5Gc, thus, is replaced in humans by its precursor, N-acetyl neuraminic acid (Neu5Ac). This absence of Neu5Gc is the major biochemical distinction between human and chimpanzee, which, theoretically, may influence intercellular interactions and embryo development. Some other ancestor genes, mostly encoding olfactory receptors [45], have been lost or

pseudogenised in the human lineage due to premature stop codon accumulations [56, 134]. On the contrary, transcribed human specific sequence termed *c1orf37-dup* gene, encoding for short transmembrane protein of unknown function, is selectively expressed in several human tissues including brain [140]. Also, a number of new copies of genes involved in immune response such as leukocyte receptors or antigens, have been acquired, lost or mutated during the recent human lineage evolution [43]. Finally, Pollard et al. [110] recently reported a novel human specific gene *HAR1F* for a putative regulatory RNA that is expressed specifically in the developing human neocortex from 7 to 19 gestational weeks, a crucial period for cortical neuron specification and migration.

Gene duplications. Gene duplications may influence cell physiology by providing additional copies of transcribed genes, thus escaping the original qualitative control of gene expression. For example, 7–11 copies of the olfactory receptor gene *OR-A* reside in human DNA, whereas the chimpanzee genome posesses only one copy of that gene. Different human copies are transcribed with different specificities, depending on their new genomic context [78, 127]. Similarly, eight genes for keratinocyte growth factor KGF were mapped in human DNA, in contrast to only five copies in the chimpanzee [143].

Lineage specific nucleotide substitutions. Millions of human specific single nucleotide substitutions, short deletions, duplications or microsatellite amplifications have been documented to the date [72, 117]. Many of them have been mapped in the regulatory genomic regions or in protein coding sequences. For example, chimpanzee dopamin receptor gene *D4* has 12 bp long deletion, as compared with its human ortholog [86]. However, the biological significance of these numerous changes accounting for a total of ~36 megabases in our DNA (mostly single nucleotide substitutions) is still unclear.

Differences in gene expression. Identifying differentially transcribed sequences may be a better solution for the direct finding of functional genes that might be involved in human speciation [37, 40]. For example, Nadezhdin et al. managed to identify differential transcription of a gene for transthyretin, the carrier of thyroid hormones, in the cerebella of humans and chimpanzees [101]. However, one has to compare samples from the same sex/physiological state groups of tissue donors. Due to an extremely limited number of the available chimpanzee tissue specimens, no reliable comparizon has been made so far, and the observed interspecies differences in gene expression remain frequently less in amplitude than the intraspecies ones [37].

23.1.1.3 Insertions of Transposable Elements

TEs are DNA fragments, capable of self-reproducing and changing their location into the host genome, *i.e.* to transpose. These selfish repetitive elements proliferate either directly via their DNA copies (DNA transposons), or through RNA intermediates (retroelements) utilizing the mechanism termed ‘reverse transcription’ and the RNA-dependant DNA polymerase enzyme, called reverse transcriptase (RT). The newly formed DNA copy of the element then integrates into the genome, using

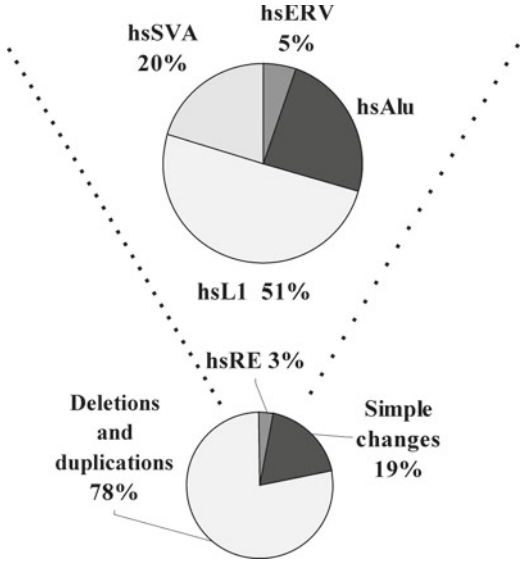

Fig. 23.1 Endogenous retroviruses occupy ~5% of the DNA shaped by human specific transposable elements, which, in turn, form only 3% of the total lineage specific DNA

a combination of host and self-encoded proteins, depending on the TE origin [11]. Retroelements, which constitute >42% of human DNA, are the only class of TEs, able to transpose in mammals [10, 124]. Four retroelement families (L1, Alu, SVA and HERV-K(HML-2)) were transpositionally active after the divergence of human and chimpanzee ancestries, thus forming relatively modest fraction of human-specific inserts (~7,800 copies [98], compared to a total of ~3 millions of human retroelements [77, 132]).

Together, human specific retroelements constitute approximately 6.4 megabases of the human DNA (Fig. 23.1), which is sixfold lower than that formed by short nucleotide substitutions, and 23-fold lower than human specific deletions/duplications. However, such a modest proportion is somewhat compensated by the active role of functional Genome Reshapers that is being played by human retrotransposons [10, 31, 123, 124, 136]. TEs are known to be recombination hot spots (e.g., human specific Alu-Alu recombinations resulted in deletion of at least 400 kb of human DNA [120]). It is known that retroelements can modify the activity of pre-existing human genes [9, 10, 31, 131]. At least one third of all human specific retroelements has been mapped within or close to genes [98]. Therefore, REs may well be one of the causative agents responsible for the phenotypic differences between *Homo sapiens* and its closest relatives, *Pan paniscus* and *Pan troglodytes* chimpanzees. These differences can be envisioned to arise not from the appearance of any new and/or disappearance of old genes but due to variations in the regulation of some genes common for the related species.

The first group, (~1,200 human specific members), is the L1 family of authonomous retrotransposons. The full-length primate L1s are about 6 kb long elements encoding two open reading frames, for RT/integrase and RNA binding protein.

However, L1 inserts are mostly 5′-truncated deficient copies originated, most probably, due to abortive reverse transcription [5]. The next two groups, Alu (~300 bp long) and SVA (~1.5 kb in size) retroposons, are non-authonomous TEs that recruit "heterologous" RT of the L1 origin for their own proliferation [133]. These two groups, presented in human DNA by ~5500 and ~860 lineage-specific copies, respectively, lack any protein coding genes and can be regarded as the parasites of L1 retrotranspositional machinery [10]. Finally, authonomous HERV-K (HML-2) endogenous retroviruses are the most complex group of human TEs. They harbor three typical retroviral functional genes and one additional gene encoding for a small regulatory protein.

23.1.2 Transposable Elements as Genome Reshapers

Repetitive sequences occupy a huge fraction of essentially every eukaryotic genome. Repetitive sequences cover more than 50% of mammalian genomic DNAs, whereas gene exons and protein coding sequences occupy only ~3% and 1%, respectively. Numerous genomic repeats include genes themselves. Those generally encode "selfish" proteins necessary for the proliferation of TEs in the host genome. The major part of evolutionary "older" TEs accumulated mutations over time and fails to encode functional proteins. However, repeats have important functions also on the RNA level [49]. Repetitive transcripts may serve as multifunctional RNAs by participating in the antisense regulation of gene activity and by competing with the host-encoded transcripts for cellular factors. Moreover, polymorphic intron-located L1 and Alu elements have been shown recently to decrease transcription of the corresponding alleles when compared to the expression of retroelement-free alleles [80, 128, 129]. In addition, genomic repeats include regulatory sequences like promoters, enhancers, splice sites, polyadenylation signals and insulators, which actively reshape cellular genome and transcriptome.

23.1.2.1 TEs as Transcriptional Promoters

Whole-genome analysis revealed that about 25% of all human promoters contain REs in their sequence [130]. Moreover, 7–10% of experimentally characterized transcription factor binding sites (TFBS) were shown to be derived from repetitive sequences including simple sequence repeats and transposable elements [109]. TFBS that originated from repeats evolve more rapidly than non-repetitive TFBS but still show signs of sequence conservation on functionally critical bases. Such rapidly evolving TFBS are likely to direct species-specific regulation of gene expression, thus participating in evolutionary process (Fig. 23.2).

In the majority of examples reported to date, REs act as alternative promoters, but may also represent the only known promoter for some human genes. For example, L1 and Alu sequences act as the unique promoter for *HYAL-4* gene, necessary for hyaluronan catabolism [130]. The application of novel high-throughput techniques

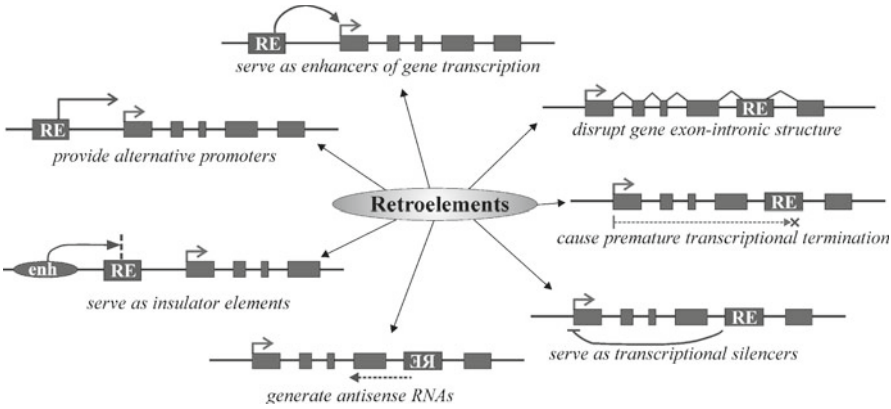

Fig. 23.2 Different mechanisms of RE influence on gene transcription

such as cap analysis of gene expression (CAGE) and paired-end ditag (PET) sequencing recently revealed 51,197 endogenous retrovirus (ERV)-derived promoter sequences. In 1,743 cases, ERVs were located in gene proximal or 5′ untranslated regions. 114 ERV-derived transcription start sites drive transcription of 97 human genes, producing chimeric transcripts initiated within LTR and read-through into known gene sequences [24].

23.1.2.2 TEs as Transcriptional Enhancers

There are many examples of TE enhancer activity in human tissues *in vivo*. For example, the ERV9 LTR element upstream of the DNase I hypersensitive site 5 (HS5) of the locus control region in the human β-globin cluster is responsible for controlling expression of this cluster in erythroid cells [87]. The enhancer of human apolipprotein A was shown to reside within LINE element [139]. Alu sequence is a part of enhancer element located in the last intron of the human CD8 alpha gene [58].

23.1.2.3 TEs as the Alternative Splice Sites

Apart from the modulation of transcription, TEs can also regulate splicing of pre-mRNA. In a genome wide comparison of the genomes of human and mouse, a total of 3,932,058 and 3,122,416 transposable elements have been identified in human and mouse, respectively. Interestingly, 60% of transposons in human are located in intronic sequences, whereas introns occupy only 24% of the genome [119]. All TE families in human can "exonize", *i.e.* be included in the exons of mature mRNA. TEs that are shared between human and mouse exhibit the same percentage of exonization in the two species, but the exonization level of a primate-specific retroelement Alu is far greater than that of other human transposons. This results in a higher overall level of transposon exonization in human than in mouse (1,824 exons compared with 506

exons, respectively) [119]. Alus are the most abundant repetitive elements in the human genome. The major burst of *Alu* retroposition took place 50–60 million years ago and has since dropped to a frequency of one new retroposition for every 20–125 new births [3, 28]. Alus are presented by more than 1.1 million of copies [21], and over 0.5 million of them reside in introns of human protein coding genes [84]. Almost all Alu-derived exons are alternatively spliced. Alu-derived exons typically have significantly weaker splicing signals compared to non-repetitive constitutively spliced exons and other alternatively spliced exons. However, at least six Alu-containing exons (in genes *FAM55C*, *NLRP1*, *ZNF611*, *ADAL*, *RPP38* and *RSPH10B*) are constitutively spliced in human tissues [85, 91, 122].

In some genes, Alu elements strikingly increased the average amount of sequence divergence between human and chimpanzee up to more than 2% in the 3′-UTRs. Moreover, 20 out of the 87 transcripts carrying Alu insert either in the 5′- or in the 3′-UTR contained more than 10% structural divergence in length. In particular, two-thirds of this structural divergence was found in the 3′-UTRs, and variable transcription start sites were conspicuous in the 5′-UTRs [118]. In both 5′- and 3′-UTR sequences, presence of an Alu element may be important for post transcriptional regulation of gene expression, for example by affecting protein translation, alternative splicing and mRNA stability [60]. Alu exonization might have played a certain role in human speciation. For example, there is a muscle-specific inclusion of an Alu-derived exon in mRNA of gene *SEPN1* (gene implicated in a form of congenital muscular dystrophy), which appeared due to a human-specific splicing change after the divergence of humans and chimpanzees [83]. The second example is the above mentioned functional deletion of an exon of human gene *CMP* for syalic acid hydroxylase.

Overall, the proportion of proteins with TE-encoded exons (approximately 0.1%), although probably underestimated, is much less than what the data at transcript level suggest (approximately 4%) [55].

23.1.2.4 TEs as Providers of Polyadenylation Signals

mRNA polyadenylation is an essential step for the maturation of almost all eukaryotic mRNAs, and is tightly coupled with termination of transcription in defining the 3′-end of genes. A polyadenylation signal (AAUAAA) nearby the 3′ end of pre-mRNA is required for poly(A) synthesis. The protein complex involved in the pre-mRNA polyadenylation is coupled with RNA polymerase II during the transcription of a gene, and only RNA polymerase II – products are terminally polyadenylated with the remarkable exception of two polyadenylated polymerase III – transcribed RNAs [7]. Autonomous retrotransposons encode proteins and utilize functional poly(A) signals at the 3′-termini of their genes. Therefore, insertions of these elements in genes in the sense orientation can influence the expression of neighboring genes by providing new poly(A) signals. For example, 5′ LTR of the retrovirus HERV-F may function as the alternative polyadenylation site for gene *ZNF195* [73]. Human genes *HHLA2* and *HHLA3* utilize HERV-H LTRs as the major polyadenylation signals [89]. Recently it was estimated that ~8% of all mammalian poly(A) sites are associated with TEs [82].

In general, there is a clearly seen strong negative selective pressure on the intron-located autonomous TE inserts oriented in the same transcriptional direction as the enclosing gene [12, 29, 131, 137, 142]. Indeed, all protein-coding intronic retroelements (including LINEs and LTR retrotransposons) oriented sense to gene transcription are underrepresented in all investigated genomes compared to statistically expected ratio of sense/antisense inserts. In contrast, non-autonomous retroelements like Alu don't employ polyadenylation of their transcripts and, thus, may have only casual AAUAAA sequences. However, such poly(A) signals are very weak and are highly affected by the surrounding sequence [113].

23.1.2.5 TEs as the Antisense Transcriptional Regulators

It has been demonstrated that TE inserts in gene introns are preferentially fixed in the antisense orientation relatively to enclosing gene transcriptional direction [96, 131]. Therefore, promoters of the intronic TEs may drive transcription of the RNAs that are complementary to gene introns and/or exons. Moreover, some retrotransposons are also known to possess bidirectional promoter [27, 34, 35, 39, 63, 94], and even downstream insertions of these elements relatively to genes may result in production of the antisense RNAs. Recently applied CAGE technology identified 48,718 human gene antisense transcriptional start sites within transposable elements [23].

One possible mechanism of the antisense regulation on the pre-mRNA level is connected with the generation of alternatively spliced mRNAs. It has been shown previously that antisense transcripts can inhibit splicing of pre-mRNA *in vitro* and *in vivo* [44]. The possible mechanism involves pairing of antisense transcript and a sense target RNA with the formation of double-stranded RNA that could induce the spliceosome to skip the paired region, thus forming an alternatively spliced transcript. This would result in the formation of non-functional RNAs containing multiple premature transcription termination codons. Normally, such RNAs are immediately degraded in the cytoplasm by nonsense-mediated decay machinery [38]. Alternatively, antisense transcript basepairing to the target RNA can lead to its rapid enzymatic degradation directly in the nucleus.

23.1.2.6 TEs as Recombination Agents

Recombination is a powerful factor of evolution that produces genetic variability by using reshuffling of already existing blocks of biological information [90]. Because of their high copy number and sequence similarity, TEs are the ideal substrates for illegitimate homologous recombination, also called ectopic recombination. The chance that an ectopic recombination will occur depends on the number of homologous sequences and on the length of the elements [6, 121]. Recombination causes genetic rearrangements that can be deleterious, advantageous or null. Alu-derived ectopic recombination generated 492 human-specific deletions, the distribution of which is biased towards gene-rich regions of the

genome [120]. Finally, L1s were shown to join DNA breaks by inserting into the genome through endonuclease-independent pathway, thus participating in DNA double-strand breaks repair [100].

23.1.2.7 TE-Transduction of the Flanking Sequences

The ability to transduce 3′-flanking DNA to new genomic loci was firstly shown for the L1 elements [53, 99, 108]. L1s have a rather weak polyadenylation signal; therefore, RNA polymerase sometimes gets through it and terminates an RNA synthesis on any polyadenylation site located downstream. It was estimated that ~20% of all L1 inserts contain transduced DNA at the 3′ends. The length of these sequences varies from few bases to over 1 kb. Taken together, such transduced DNA makes up ~0.6–1% of the human genome. Therefore, L1-mediated transductions have the potential to shuffle exons and regulatory sequences to new genomic sites.

Recently it was shown that SVA elements are also able to transduce downstream sequence and it was estimated that about 10% of human SVA elements were involved in DNA transduction events [107, 133]. Moreover, SVA-mediated transduction can serve as a previously uncharacterized mechanism for gene duplication [138]. It the latter case new sequences may appear either on the 5′- or on the 3′ terminus of an SVA (5′ and 3′ SVA transduction, respectively). 3′ Transduction mechanism is similar to that proposed for L1 retrotransposon. The size of genomic sequence transferred in such a way may differ from several base pairs to over 1.500 bp. The most striking example is the transduction of a whole gene *AMAC* (acyl-malonyl condensing enzyme 1) in the great ape genomes [138]. Due to SVA 3′ transduction, human genome has two additional copies of *AMAC*.

Another kind of transduction results in attaching of new sequences to the 5′ end of an SVA. TE transcription initiation may proceed from any promoter located upstream in the genomic sequence. In this case termination of transcription and RNA processing usually occur using normal polyadenylation signal of a TE. This results in a mature RNA having on its 5′ end an additional copy of flanking genomic sequence and a copy of RE at its 3′ end. Subsequent reverse transcription and integration into the genome of a nascent cDNA result to a new RE genomic insert carrying 5′ transduced part [8].

23.2 Results

23.2.1 Discovery of New Human TE Families

23.2.1.1 RNA Recombination-Derived TEs

A typical LINE element encodes two proteins: ORF1p that is a RNA binding protein which likely helps reverse transcription as a nucleic acid chaperone [93], and ORF2p, the reverse transcriptase and the endonuclease [71]. Due to a ‘cis-preference’,

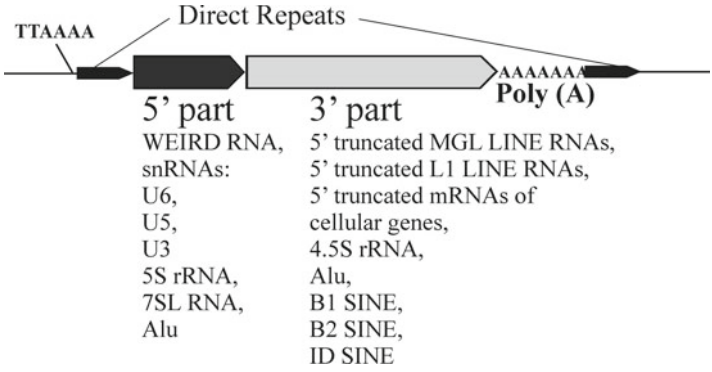

Fig. 23.3 Schematic representation of the bipartite chimeric retrogenes. Inserts are flanked by 10–20 bp long genomic direct repeats

the enzymatic machinery of a retrotransposition-competent LINE predominantly transposes its own copies [135]. However, LINEs are also able to mediate the transposition of other sequences, mostly non autonomous elements termed SINEs, but also cDNAs originating from different cellular RNAs, leading to the formation of processed pseudogenes [32]. Recently, we have shown that LINEs are involved in the formation of bi- and tripartite chimeric retrogenes during reverse transcription in many genomes including human and fungi [13, 14, 16, 41, 50]. Bipartite chimeric retrogenes with an unusual structure were identified in three mammalian and in one fungal genomes (Fig. 23.3).

A total of 82, 116, 66 and 31 elements were found in human, mouse, rat and rice blast fungus *Magnaporthe grisea* DNAs, respectively [13, 14, 41, 50]. These elements are composed of DNA copies from cellular transcripts either directly fused to each other or more frequently fused to the 3′ part of a LINE retrotransposon. The various cellular transcripts found in these chimeras correspond to messenger RNAs, ribosomal RNAs, small nuclear RNAs, and 7SL RNA.

The chimeras have the following common features: (1) 5′-parts are full-length copies of cellular RNAs; (2) 3′-parts are 5′-truncated copies of the corresponding RNAs (mostly LINEs); (3) sites of these truncations occur at random in the corresponding RNA; (4) both parts are directly joined with the same transcriptional orientation; (5) chimeras have a poly (A) tail at their 3′ end, and (6) chimeras are flanked by short direct repeats.

The last structural feature demonstrate that these elements were transposed as bipartite DNA copies. Indeed, mammalian chimeras carried at their 5′ ends a T_2A_4 hexanucleotide or its variants [13, 14, 50] that correspond to the T_2A_4 genomic site used by LINEs to initiate reverse transcription on oligo (A) motifs and separate newly inserted DNA by short tandem repeats [67]. The simultaneous integration of both parts of these chimeras was further supported by the data came from PCR-based evolutionary insertion polymorphism assay [13, 14].

This suggests that these bipartite elements are generated by a specific active mechanism. It frequently combines functional cellular transcripts that have nothing in common with transposable elements [19]. Many of the chimeras can be considered

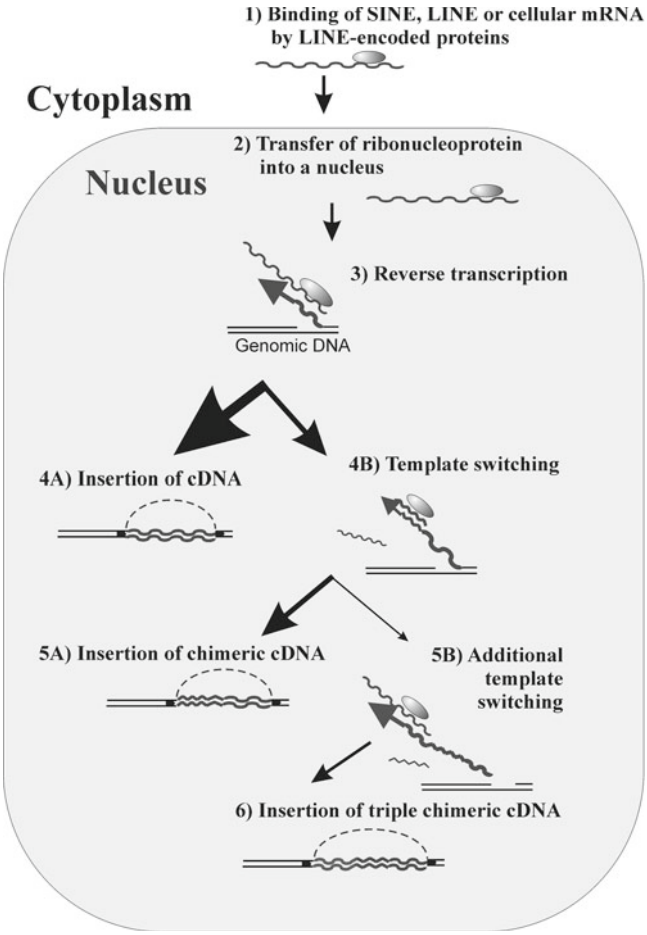

Fig. 23.4 Mechanism for the chimeras' formation using LINE enzymatic machinery. (*Step 1*) LINE pre-integration complex binds LINE, SINE or RNA in the cytoplasm. (*Step 2*) The resulting ribonucleoprotein is transferred to the nucleus. (*Step 3*) Reverse transcription of the bound RNA primed by a genomic DNA single-stranded break (target site primed reverse transcription). (*Step 4A*) Successful integration of the reverse transcribed cDNA copy into the genomic DNA. (*Step 4B*) Switch of templates on another RNA during the reverse transcription. (*Step 5A*) Integration of the chimera formed into genomic DNA. (*Step 5B*) The second template switch to another RNA with subsequent DNA reparation mediates formation of a tripartite chimeric retrogene insertion flanked by short direct repeats. The normal LINE integration pathway is: steps (*1*), (*2*), (*3*), (*4A*)

as new genes, as they were shown to be transcribed, some of them in a tissue-specific manner [14, 48, 50, 51]. Later on, in the mammalian and fungal genomes we found also the tripartite chimeras of a similar structure [51]. We further proposed that these chimeric retrogenes were generated through a mechanism involving RNA recombination during the reverse transcription of cellular RNAs (Fig. 23.4). This model includes a switch from the nascent cDNA serving as template for the reverse

transcription of the 3′ part of the chimera to another RNA template corresponding to the 5′ part, followed by the chimera integration into the host genome [10].

Although RT main enzymatic activity is the continuous synthesis of the cDNA on RNA template, RT is able to switch templates during reverse transcription. For example, in retroviruses, RT jumps from one site of the RNA template to another site, are necessary for the synthesis of LTRs. Moreover, as retroviral particles usually contain two independent RNA molecules [126], the high template switch frequency significantly increases the retroviral diversity through recombination between these RNAs [69]. These recombination events most probably account for the mosaic structure of most retroviruses [66, 125].

This model for the chimera formation was further supported by results obtained with human L1 LINE element using an elegant experimental system of retrotransposition *in vitro* [46]. The authors managed to characterize 100 *de novo* retrotransposition events in HeLa cells. Importantly, one insert (1%) represented a newly formed chimera similar to those we identified in human genome, consisting of a full length U6 snRNA fused to a 5′ truncated L1. Similar results were obtained *in vivo* with a transgenic mouse model for L1 retrotransposition by Babushok and coauthors that characterized 33 novel retrotransposition events. 13% of these events likely result from template switching during reverse transcription [1]. Interestingly, it has been recently postulated that RT template jumps from LINE RNA to host genomic DNA might facilitate integration and, thus, could be normally required for successful LINE retrotransposition [1, 4].

Besides generating chimeric retrogenes, template switching events during LINE reverse transcription could give rise to chimeric SINE elements [104] and to mosaic rodent L1 structures, likely resulting from RNA recombination between L1 templates [8, 61]. Evolution of certain LINE families might also involve RNA-RNA recombination, resulting in the fusion of the 3′ part of a LINE to a new sequence at their 5′ end, as suggested by the observation that the 5′-untranslated regions of human, murine, rat and rabbit L1 families are not homologous to each other [42]. Interestingly, RT encoded by another member of LINE superfamily – R2 from arthropods, was documented to jump from one template to another *in vitro*, with R2-R2 chimeras being formed [4].

Furthermore, it is speculated that LTR-containing retrotransposons and SINEs themselves represent chimeric elements [75, 76, 92, 105]. A phylogenetic analysis of the ribonuclease H domain revealed that LTR-containing retroelements might have been formed as a fusion between DNA transposon and non-LTR retrotransposon [92]. tRNA-derived SINEs likely descended from retroviral strong-stop DNAs [105]. They consist of two regions: a conservative, including a tRNA promoter and a core domain, and a variable one similar to 3′-terminal sequence of different LINE families. The core domain of tRNA-like SINEs has conservative regions similar to fragments of lysine tRNA-primed retroviral LTRs. On the basis of these structural peculiarities it was suggested that tRNA-derived SINEs emerged due to the integration of retroviral strong-stop DNA into the LINE 3′-terminal part. The RE formed could be transcribed by RNA polymerase III and spread through the genome. Such a mechanism of SINE formation could also explain how these elements can transpose

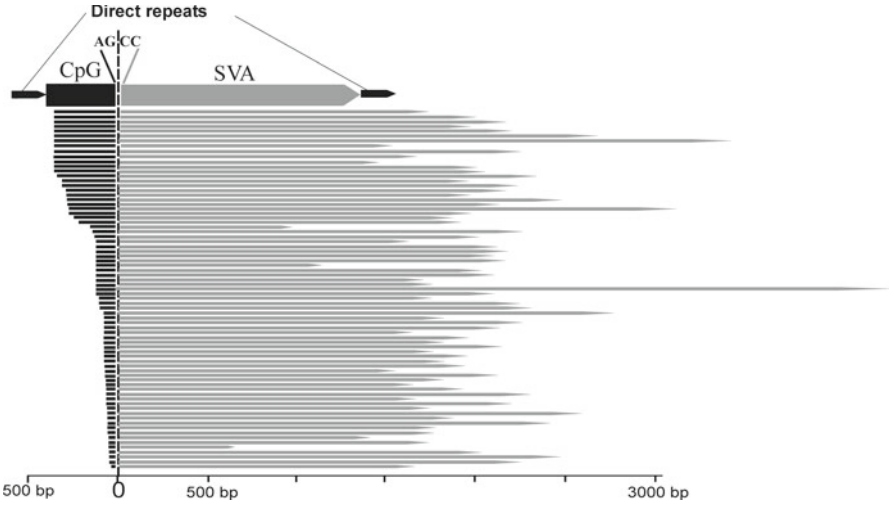

Fig. 23.5 Structure of chimeric CpG-SVA retrotransposons. CpG-SVA Inserts are flanked by direct repeats. Lengths of 5′ terminal (exonic) part vary from 35 to 383 bp, lengths of 3′ (SVA-derived) part vary from 662 to 4255 bp. 5′ Terminal parts are homologous to the first exon of *MAST2* gene, 3′ terminal parts – to SVA retrotransposon. Junction point between the two parts is identical in all CpG-SVA elements (canonical splice acceptor site AG from the side of exonic part and non-canonical splice donor site CC from the side of SVA). All SVA fragments start from the position 396 of the SVA consensus sequence

in the genome. Namely, it seems very likely that they recruited the enzymatic machinery from LINEs through a common "tail" sequence [105].

23.2.1.2 Human-Specific Hybrid Family CpG-SVA

Detailed structural analysis of the human specific SVA retrotransposons revealed 76 elements of an unusual structure. At the 5′ termini these elements carried copies of the first exon of *MAST2* gene, whereas at the 3′ end – SVA retrotransposon sequences. The border between exonic and SVA parts was located exactly between canonical acceptor splice site AG from exonic part and non-canonical donor splice-site CC from SVA-part (396 position in the SVA consensus sequence). Lengths of both parts of chimeric elements significantly varied: from 35 to 383 bp for the 5′-terminal part and from 662 to 4,255 bp for the 3′ terminal part. The border between the two parts was constant in all the chimeras (Fig. 23.5). On the 3′ terminus, the chimeras harbored a poly (A) sequence of variable length. These bipartite elements were flanked by 12–18 bp long direct repeats. In one case the length of direct repeats was unusually big (131 bp). Presence of the direct repeats surrounding chimeric inserts suggest implication of L1 retrotranspositional machinery in their formation, whereas poly (A) sequence indicates that retrotransposed RNA was transcribed by RNA polymerase II. The identified family of chimeric REs was called "CpG-SVA"

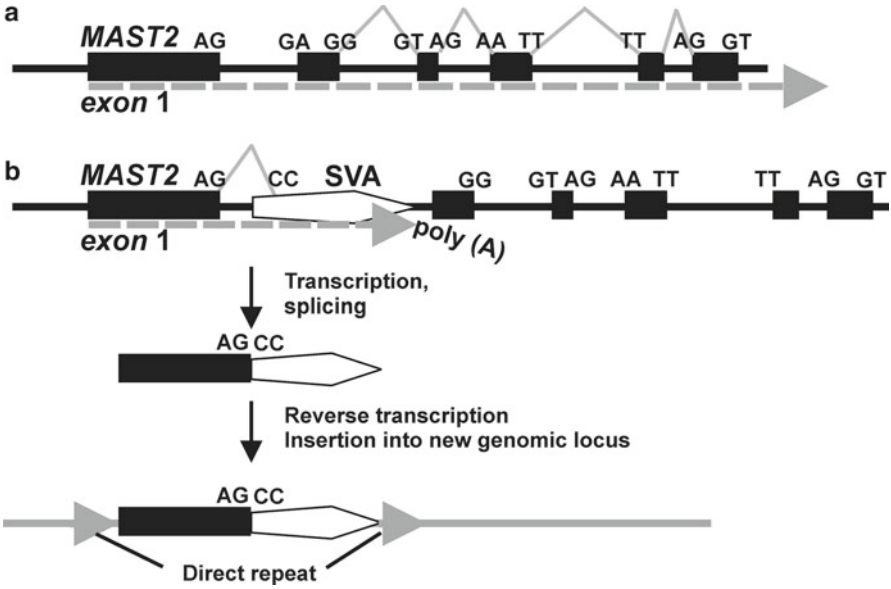

Fig. 23.6 Proposed mechanism of CpG-SVA family formation. (**a**) – schematic representation of genomic locus comprising human gene *MAST2*. *Dotted arrow* designates transcriptional direction, exons and splice sites are shown. (**b**) – Insert of an SVA retrotransposon in the sense orientation has changed gene exon-intronic structure and gave rise to aberrantly spliced mRNA polyadenylated at SVA sequence. Copy of this mRNA has inserted into a new locus of human genome and gave rise to CpG-SVA family that continued proliferation in human DNA. However, the ancestral allele of *MAST2* gene comprising SVA insert was lost due to the negative selection

because its 5′ terminal part complementary to the first exon of *MAST2* gene included a CpG island sequence. CpG-SVA elements were found only in human genomic DNA, whereas separately both SVA retrotransposons and *MAST2* exon sequence exist in the genomes of all great apes. Therefore, CpG-SVA may be regarded as a new human specific family of retrotransposons [2].

Two other papers describing the same family of hybrid retrotransposons (CpG-SVA) have been simultaneously published, where this family was termed either “MAST2-SVA” [59] or “SVA-F1” [30].

Basing on the structural features of the identified CpG-SVA family members, we purposed a mechanism for their formation (Fig. 23.6). At the first stage, SVA retrotransposon most probably has inserted into the first intron of *MAST2* gene in the sense orientation. After that there was formed an aberrant RNA driven by *MAST2* promoter and terminally processed using SVA polyadenylation signal. This RNA was further spliced which resulted in a fusion of the first exon of *MAST2* with a 3′-terminal fragment of an SVA (starting from 393 nucleotide of the SVA consensus sequence). This spliced chimeric RNA was then reverse transcribed by the L1 retrotranspositional machinery followed by integration of a nascent cDNA into the genome. This resulted to emerging of the master copy of CpG-SVA inserted into

human DNA and flanked by direct repeats. The newly Inserted CpG-SVA element appeared to be transcriptionally active, possible due to the enclosed CpG-islet, and gave rise to a new family of REs.

This hypothesis is supported by the observation that there is the canonical *MAST2* gene acceptor splice site AG of on the border between the *MAST2*- and SVA-derived fragments. The putative donor splice site CC within an SVA is not canonical, what may be explained by the peculiarities of *MAST2* exon-intronic structure where non-canonical splice sites form the majority (Fig. 23.6).

Interestingly, at present there is no fixed SVA insert into *MAST2* gene intron in the human genome. Apparently, an ancestral allele containing the above SVA element in gene intron was eliminated by the negative selection as it could not provide functional *MAST2* mRNA formation because of the aberrant splicing of transcripts and/or preliminary polyadenilation on the SVA sequence.

We have found among the CpG-SVA elements several cases of 5′ and 3′ transduction of unrelated genomic DNA, proven by the mapping of the enclosing direct repeats. As in the classical 3′ transduction mechanism, it is likely that the downstream genomic fragments were captured due to "getting through" of SVA polyadenilation signals by the RNA polymerase II complex with the subsequent termination on any downstream sequence. In case of 5′ CpG-SVA transduction, there was apparently transcription of CpG-SVA elements initiated from upstream genomic promoters. Overall, we identified 18 and 11 cases of the 5′ and 3′ CpG-SVA transductions, respectively. The size of transferred genomic sequence differed from 8 to 854 bp for 5′- and from 141 to 734 bp for 3′ transduction events. Remarkably, four CpG-SVA elements contained both 5′ and 3′ transduced sequences. These four elements were highly identical and consisted of 364 bp long *MAST2* exon and 2,143–3,361 bp long SVA sequences. SVA length variations were caused by the Instability its internal satellite repeat modules. The double transducer CpG-SVAs were flanked by Alu sequence (member of evolutionally ancient AluSc family) at the 5′-termini and by the 400 bp long sequence including evolutionally ancient AluSp element at the 3′ ends. These structure similarities evidence common ancestry of these four elements from a single progenitor CpG-SVA element.

Once the exonic parts of the chimeras varied in length, but not in their primary structure, the SVA-derived parts had very different both lengths and primary structure. In the SVA parts there were different genetic changes like insertions, deletions, duplications, quantitative changes in tandem repeat composition and even insertions of retrotransposons. Together with the presence of transduced genomic sequences, this enabled us to construct phylogenetic tree for the members of CpG-SVA family to trace their reciprocal neighborhood. According to the primary structure similarity, CpG-SVA elements were grouped into three major branches (Fig. 23.3). Interestingly, although there was a kind of correlation between the size of "exonic" part and sequence localization on the tree, all three above brunches contained elements having exonic parts of very different lengths. There was also no connection between the position on a tree and lengths of the SVA parts. In several cases different tree brunches were including elements with the exactly same lengths of exonic part. For example, brunch 2 contained one CpG-SVA element with 364 bp long exonic part,

whereas brunch 1 – five such elements. Exonic parts of seven elements from brunch 2 and of one element from brunch 3 were 148 bp long. There were also similar coincidences for the lengths 64, 76, 88 and 361 bp. These coincidences of exonic part sizes evidence that there were multiple independent events when CpG-SVA elements with identical exonic parts were formed.

The observed peculiarities of distribution of lengths of CpG-SVA exonic parts may be explained by the following factors: (1) there could be multiple functional transcription start sites within CpG-SVA, or (2) in some cases reverse transcription of the CpG-SVA RNA could terminate before the complete copying of the template has finished. The resulting shortened CpG-SVA inserts could, in turn, generate new elements having even shorter exonic parts, etc.

What are the functions of exonic part of CpG-SVA? Considering that (1) the first exon of *MAST2* gene includes CpG island, (2) CpG islands usually play major roles in gene transcriptional regulation, and (3) *MAST2* is strongly upregulated in testis, It can be hypothesized that the exonic part provides increased transcription of CpG-SVA family members in testis. This may be beneficial for the CpG-SVA family as it facilitates fixation of new inserts in the genome. To be fixed, RE insertion must occur into germ line cells, *e.g*. those localized in testis. Indeed, in terms of proliferation in the genome, the evolutionary young family CpG-SVA should be considered as very successful one: offsprings of only one among more than 1,000 SVA copies that resided in human DNA at that time (i.e. < 0.1%) have generated 76 new fixed inserts (~9% of all 860 human specified SVA elements) [2]. Experimental investigation of this hypothesis will be a matter of our further studies.

23.2.2 Functional Characterization of a Family of Human Specific Endogenous Retroviruses HERV-K(HML-2)

23.2.2.1 Identification of Human Specific Promoters

Promoter activity of human specific LTRs was investigated in both *in vitro* and *in vivo* assays. In transient transfection experiments with the luciferase or GFP reporter genes, the same human specific element from contig L47334 displayed very low promoter activity in three of the ten cell lines tested, moderate activity (10–20% of the SV40 early promoter) was observed in six cell lines and, finally, the maximal value of ~100% of SV40 promoter activity was obtained in Tera-1 cells, similarly to the above enhancer activity tests [34]. In the experiments by Lavie et al. [79], five human specific proviral 5′ LTRs have demonstrated the promoter strengths as high as 5–15% of the cytomegalovirus (CMV) promoter activity in Tera-1 cells (AP000776 – 15% of CMV promoter expression, AC025757 – 9%, AC072054 – 8%, AC025420 – 6% and AL590785 – 5%). The authors have demonstrated that the promoter activities of these elements directly depend on the methylation status of their CpG dinucleotides. Interestingly, the same five LTRs were strongly transcriptionally repressed in T47D cells [79].

In *in vivo* experiments, 5′ RACE (rapid amplification of cDNA ends) – based mapping of transcriptional start sites for five actively transcribed human specific LTRs provided evidence for the presence of two functional promoter regions within the LTR sequence [74]. Both promoters possess TATA box motif and other upstream regulatory sequences. The first promoter was the canonical element located in the LTR U3 region, whereas the second one was mapped in the very 3′ terminus of the LTR R region. Both promoters appeared to be active in solitary LTRs and in full-length proviruses. Surprisingly, this second non-canonical element was even more active than the classical U3-located retroviral promoter. Therefore, the R region is excluded from most transcripts initiated on LTRs, whereas a classical retroviral life cycle model implies that the transcription is driven from between the LTR U3 and R elements (first promoter), and the R transcript is a 5′-terminal component of the newly synthesized proviral RNA. Such a mode of proviral DNA transcription is a basis of the life cycle that provides the possibility of template jumps during proviral RNA reverse transcription. A shift of the transcriptional start site can be explained by the presence of at least two alternative promoters within the LTR, one of which is normally used for viral gene expression, and the other for transcription of retrotransposition-competent copies of the integrated provirus. The latter type of transcripts is supposed to be far less abundant, what basically corresponds to the above observations. It should be mentioned that alternative promoters with unknown functions were found earlier for many other retrotransposons [10, 31, 103].

Recently, we performed the comprehensive study of the expression of human specific LTRs *in vivo* in human germ-line tissue (testicular parenchyma) and in the corresponding tumor (seminoma) sampled from the same patient [17]. These were chosen because of markedly high endogenous retroviral transcriptional activity in germ-line cells, which is most probably needed to make *de novo* retroviral integrations inheritable [88, 112]. To this end, a new experimental technique that makes it possible to detect repetitive element own promoter activity has been developed [18]. This technique, termed GREM (genomic repeat expression monitor), combines the advantages of 5′-RACE and nucleic acid hybridization techniques. GREM is based on hybridization of total pools of cDNA 5′ terminal parts to genome wide pools of repetitive elements flanking DNA, followed by selective PCR amplification of the resulting hybrid cDNA-genome duplexes. A library of cDNA/genomic DNA hybrid molecules obtained in such a way can be used as a set of tags for individual transcriptionally active repetitive elements [18]. The method is both quantitative and qualitative, as the number of tags is proportional to the content of mRNA driven from the corresponding promoter active repetitive element. The GREM outcome was a set of amplified cDNA/genomic DNA heteroduplexes, below referred to as Expressed LTR Tags (ELTs), which were further cloned and sequenced. This study was the first detailed characterization of the functional promoters provided by a particular group of genomic repetitive elements. The data obtained in such a way suggest that at least 45% of human specific LTRs possessed promoter activity, and a total of 60 new human promoters have been identified. Individual LTRs were expressed at markedly different levels ranging from ~0.001% to ~3% of the housekeeping beta-actin gene transcript level. Although HS elements formed several subclusters on a phylogenetic tree [15, 95], no clear correlation between LTR

primary structure and transcriptional activity was found. In contrast, the LTR status (solitary, 5′ or 3′ proviral) was an important factor affecting LTR activity: promoter strengths of solitary and 3′ proviral LTRs were almost identical in both tissues, whereas 5′ proviral LTRs displayed higher promoter activity (~2-fold and ~5-fold greater in testicular parenchyma and seminoma, respectively). These data suggest that a proviral sequence harbors some yet unknown downstream regulatory elements that provide significantly higher 5′ LTR expression, especially in seminoma [17, 18]. Another important factor affecting promoter activity was the LTR distance from genes: the relative content of promoter-active LTRs in gene-rich regions was significantly higher than in gene-poor genomic loci.

The data obtained suggest also a selective suppression of transcription in both tissues for proviral 3′ LTRs located in gene introns. Such a transcriptional suppression might be aimed at silencing of the proviral gene expression in gene-rich regions. In testicular parenchyma, the promoter strength of intronically located solitary LTRs was also significantly decreased. This may suggest yet unknown mechanism(s) for selective suppression of "extra" promoters generated due to mutations or viral integrations and located within gene introns or very closely to genes. Such a mechanism might minimize possible destructive effects of undesirable transcription. Many transcriptionally competent LTRs were mapped near known human genes, and as many as 86–90% of all genes located in close proximity to promoter active LTRs are known to be transcribed in testis. However, in general no clear-cut correlation was observed between transcriptional activities of genes and closely located LTRs [17]. Overall, LTRs provided at least 60 functional human specific promoters for host non-repetitive DNA, that are transcribed at different levels ranging from ~0.001% to ~3% of beta-actin transcript level.

23.2.2.2 Antisense Regulation of Functional Genes by the Human Specific HERV-K(HML-2) Elements

Later on, we reported the first evidence for the human specific antisense regulation of gene activity occurring due to promoter activity of HERV-K(HML-2) endogenous retroviral inserts [49, 52]. Human-specific LTRs located in the introns of genes *SLC4A8* (for sodium bicarbonate cotransporter) and *IFT172* (for intraflagellar transport protein 172) *in vivo* generate transcripts that are complementary to exons within the corresponding mRNAs in a variety of human tissues (Fig. 23.7). As shown by using 5′RACE technique (rapid amplification of cDNA ends), in both cases the LTR-promoted transcription starts within the same position of the LTR consensus sequence, which coincides with the previously reported HERV-K (HML-2) LTR transcriptional start site [74].

The effect of the antisense transcript overexpression on the mRNA level of the corresponding genes was investigated using quantitative real-time RT-PCR. Almost fourfold increase in *SLC-AS* expression led to 3.9-fold decrease of *SLC4A8* mRNA level, and overexpression of *IFT-AS* transcript 2.9-fold reduced the level of *IFT172* mRNA. In all cases the level of the antisense RNAs in the transfected cells was close to or lower than in many human tissues [52]. Similarly, intronically-located

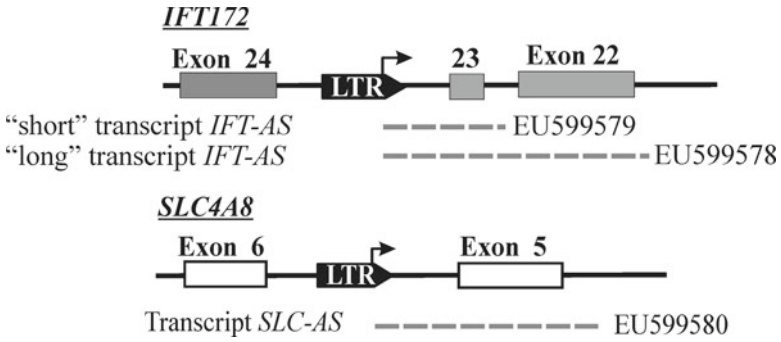

Fig. 23.7 Types of the antisense transcripts found and their corresponding accession numbers

representatives of an LTR retrotransposon family from rice genome called Dasheng likely regulate tissue-specific expression of several adjacent functional genes *via* antisense transcripts driven by the LTRs [70].

23.2.2.3 Enhancer Activity of Human Specific HERV-K(HML-2)

HERV-K (HML-2) LTR sequence harbors a complete set of regulatory elements required for regulation of the retroviral transcription *in vivo* and include a functional enhancer element including multiple transcription factor binding sites. HERV-K (HML-2) LTR enhancer activity was extensively studied *in vitro*, mostly for non-human specific members [33, 64, 115], with the only exception of the human specific solitary LTR from the genomic contig L47334 [115]. In transient transfection experiments on a panel of 10 mammalian cell lines, this LTR has demonstrated enhancer activity only in Tera-1 human testicular embryonal carcinoma cells (thus showing ~8-fold increase in luciferase expression, as compared to control plasmid lacking the enhancer element) [115].

In our recent studies, we found that ~ one-third of all human specific HERV-K (HML-2) LTRs are located in the close gene neighborhood. Nine such elements reside in the upstream regions of known human genes, close to transcriptional start site (*i.e.* at the distance less than 5 kb). In our experiments, three elements over nine have shown a strong enhancer effect in cell culture tests (up to ninefold increase in transcriptional activity). However, only one element, located upstream human gene *PRODH*, has demonstrated a correlation between the enhancer activities *in vitro* and *in vivo*. In the case of two other elements, the LTR inserts were deeply methylated in all the investigated tissues. In contrast, the LTR from *PRODH* region was mostly unmethylated in genomic DNAs of human brain and spinal cord. Our further studies revealed that the LTR enhancer activity is fully regulated by the methylation: the higher is the level of the methylation, the weaker is the LTR enhancer activity, and vice versa. Importantly, *PRODH* promoter is unmethylated in all the tissues, and this gene is transcribed predominantly in the central nervous system (CNS). In the

experiments with the mouse brain progenitor cells, we have shown that *in vitro* the LTR insert has a strong enhancer activity on the *PRODH* promoter, thus fivefold increasing transcriptional level of a reporter gene. Furthermore, we have identified a family of transcriptional factors SRY/SOX, that are the most likely candidates for being the LTR activity mediators in germ cells and in the CNS.

PRODH encodes a CNS-specific isoform of the proline dehydrogenase. Unlike the liver isoform, *PRODH* is involved not only in the proline catabolism, but, mostly, in the synthesis of neuromediators like dopamine, GABA, aspartate and glutamate. The deficiency in PRODH activity causes first-type hyperprolinemia, that is frequently linked with severe cognitive disorders and CNS malfunctions, and, in several documented cases, with schizophrenia. Due to its important functions, *PRODH* expression is tightly regulated in human brains, and its expression profile in humans has little in common with the rodent ones. It should be noted, that *PRODH* promoter sequence itself is highly conserved among the mammalians, and the major structural distinction of the *PRODH* upstream region in human, rat, mouse and chimpanzee is the presence of the HERV-K (HML-2) LTR insert in human.

23.3 Concluding Remarks

Thus, the detailed analysis of a small fraction of human specific transposable elements revealed that they may regulate our genes by acting both in *cis* (as promoters and enhancers) and in *trans* (as antisense regulators and RNA recombination hotspots). At least three genes have been identified that are the subjects of a human-specific regulation by the TEs. Considering that only a relatively small portion of the human specific TEs was thoroughly analyzed to the date (~2% of all human specific TEs), one can expect that in the future the detailed genome-wide functional characterization of all human-specific TEs will make it possible to identify tens- or hundreds of genes having unique for human expression profiles. This knowledge, hopefully, at least partly will help us to answer the question "What makes us humans?" … On the molecular level, of course.

Acknowledgments The authors were sponsored by the Russian Foundation for Basic Research grants 09-04-12302 & 10-04-00593-a, by the President of the Russian Federation grant MD-480.2010.4 and by the Program "Molecular and Cellular Biology" of the Presidium of the Russian Academy of Sciences. A.Buzdin presentation was supported by the NATO fellowship.

References

1. Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH Jr (2006) L1 integration in a transgenic mouse model. Genome Res 16:240–250
2. Bantysh OB, Buzdin AA (2009) Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA. Biochemistry (Mosc) 74:1393–1399

3. Batzer MA, Schmid CW, Deininger PL (1993) Evolutionary analyses of repetitive DNA sequences. Methods Enzymol 224:213–232
4. Bibillo A, Eickbush TH (2004) End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. J Biol Chem 279:14945–14953
5. Boissinot S, Chevret P, Furano AV (2000) L1 (LINE-1) retrotransposon evolution and amplification in recent human history. Mol Biol Evol 17:915–928
6. Boissinot S, Davis J, Entezam A, Petrov D, Furano AV (2006) Fitness cost of LINE-1 (L1) activity in humans. Proc Natl Acad Sci USA 103:9590–9594
7. Borodulina OR, Kramerov DA (2008) Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. RNA 14:1865–1873
8. Brosius J (1999) Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica 107:209–238
9. Brosius J (1999) RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 238:115–134
10. Buzdin AA (2004) Retroelements and formation of chimeric retrogenes. Cell Mol Life Sci 61:2046–2059
11. Buzdin A (2006) Transposable elements and their use for target site specific gene delivery. Curr Pharmacogenomics 4:1–8
12. Buzdin A (2007) Human-specific endogenous retroviruses. Sci World J 7:1848–1868
13. 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
14. Buzdin A, Gogvadze E, Kovalskaya E, Volchkov P, Ustyugova S, Illarionova A, Fushan A, Vinogradova T, Sverdlov E (2003) The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination. Nucleic Acids Res 31: 4385–4390
15. Buzdin A, Ustyugova S, Khodosevich K, Mamedov I, Lebedev Y, Hunsmann G, Sverdlov E (2003) Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages(small star, filled). Genomics 81:149–156
16. Buzdin A, Vinogradova T, Lebedev Y, Sverdlov E (2005) Genome-wide experimental identification and functional analysis of human specific retroelements. Cytogenet Genome Res 110:468–474
17. Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E (2006) At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. J Virol 80:10752–10762
18. Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E (2006) GREM, a technique for genome-wide isolation and quantitative analysis of promoter active repeats. Nucleic Acids Res 34:e67
19. Buzdin A, Gogvadze E, Lebrun MH (2007) Chimeric retrogenes suggest a role for the nucleolus in LINE amplification. FEBS Lett 581:2877–2882
20. Chen FC, Chen CJ, Li WH, Chuang TJ (2007) Human-specific insertions and deletions inferred from mammalian genome sequences. Genome Res 17:16–22
21. Chen C, Ara T, Gautheret D (2009) Using Alu elements as polyadenylation sites: a case of retroposon exaptation. Mol Biol Evol 26:327–334
22. Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, Wright KL, Muchmore EA, Nelson DL, Warren ST, Varki A (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. Proc Natl Acad Sci USA 95:11751–11756
23. Conley AB, Miller WJ, Jordan IK (2008) Human cis natural antisense transcripts initiated by transposable elements. Trends Genet 24:53–56
24. Conley AB, Piriyapongsa J, Jordan IK (2008) Retroviral promoters in the human genome. Bioinformatics 24:1563–1567
25. consortium, C. s. a. a (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87

26. Cooper GM, Brudno M, Green ED, Batzoglou S, Sidow A (2003) Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. Genome Res 13:813–820
27. Copeland CS, Mann VH, Brindley PJ (2007) Both sense and antisense strands of the LTR of the Schistosoma mansoni Pao-like retrotransposon Sinbad drive luciferase expression. Mol Genet Genomics 277:161–170
28. Cordaux R, Hedges DJ, Herke SW, Batzer MA (2006) Estimating the retrotransposition rate of human Alu elements. Gene 373:134–137
29. Cutter AD, Good JM, Pappas CT, Saunders MA, Starrett DM, Wheeler TJ (2005) Transposable element orientation bias in the Drosophila melanogaster genome. J Mol Evol 61:733–741
30. Damert A, Raiz J, Horn AV, Lower J, Wang H, Xing J, Batzer MA, Lower R, Schumann GG (2009) 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome Res 19:1992–2008
31. Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr (2003) Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 13:651–658
32. Dewannieux M, Esnault C, Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 35:41–48
33. Domanskii AN, Akopov SB, Lebedev Iu B, Nikolaev LG, Sverdlov ED (2002) Enhancer activity of solitary long terminal repeat of the human endogenous retrovirus of the HERV-K family. Bioorg Khim 28:341–345
34. Domansky AN, Kopantzev EP, Snezhkov EV, Lebedev YB, Leib-Mosch C, Sverdlov ED (2000) Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. FEBS Lett 472:191–195
35. Dunn CA, Romanish MT, Gutierrez LE, van de Lagemaat LN, Mager DL (2006) Transcription of two human genes from a bidirectional endogenous retrovirus promoter. Gene 366:335–342
36. Ebersberger I, Metzler D, Schwarz C, Paabo S (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. Am J Hum Genet 70:1490–1497
37. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Paabo S (2002) Intra- and interspecific variation in primate gene expression patterns. Science 296:340–343
38. Fasken MB, Corbett AH (2005) Process or perish: quality control in mRNA biogenesis. Nat Struct Mol Biol 12:482–488
39. Feuchter A, Mager D (1990) Functional heterogeneity of a large family of human LTR-like promoters and enhancers. Nucleic Acids Res 18:1261–1270
40. Fu N, Drinnenberg I, Kelso J, Wu JR, Paabo S, Zeng R, Khaitovich P (2007) Comparison of protein and mRNA expression evolution in humans and chimpanzees. PLoS One 2:e216
41. Fudal I, Bohnert HU, Tharreau D, Lebrun MH (2005) Transposition of MINE, a composite retrotransposon, in the avirulence gene ACE1 of the rice blast fungus Magnaporthe grisea. Fungal Genet Biol 42:761–772
42. Furano AV (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol 64:255–294
43. Gagneux P, Varki A (2001) Genetic differences between humans and great apes. Mol Phylogenet Evol 18:2–13
44. Galante PA, Vidal DO, de Souza JE, Camargo AA, de Souza SJ (2007) Sense-antisense pairs in mammals: functional and evolutionary considerations. Genome Biol 8:R40
45. Gilad Y, Man O, Paabo S, Lancet D (2003) Human specific loss of olfactory receptor genes. Proc Natl Acad Sci USA 100:3324–3327
46. Gilbert N, Lutz S, Morrish TA, Moran JV (2005) Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol 25:7780–7795
47. Glaser B, Yen PH, Schempp W (1998) Fibre-fluorescence in situ hybridization unravels apparently seven DAZ genes or pseudogenes clustered within a Y-chromosome region frequently deleted in azoospermic males. Chromosome Res 6:481–486
48. Gogvadze EV, Buzdin AA (2005) New mechanism of retrogene formation in mammalian genomes: in vivo recombination during RNA reverse transcription. Mol Biol (Mosk) 39:364–373

49. Gogvadze E, Buzdin A (2009) Retroelements and their impact on genome evolution and functioning. Cell Mol Life Sci 66:3727–3742
50. Gogvadze EV, Buzdin AA, Sverdlov ED (2005) Multiple template switches on LINE-directed reverse transcription: the most probable formation mechanism for the double and triple chimeric retroelements in mammals. Bioorg Khim 31:82–89
51. Gogvadze E, Barbisan C, Lebrun MH, Buzdin A (2007) Tripartite chimeric pseudogene from the genome of rice blast fungus Magnaporthe grisea suggests double template jumps during long interspersed nuclear element (LINE) reverse transcription. BMC Genomics 8:360
52. Gogvadze E, Stukacheva E, Buzdin A, Sverdlov E (2009) Human specific modulation of transcriptional activity provided by endogenous retroviral inserts. J Virol 83:6098–6105
53. Goodier JL, Ostertag EM, Kazazian HH Jr (2000) Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum Mol Genet 9:653–657
54. Goodman M (1999) The genomic record of Humankind's evolutionary roots. Am J Hum Genet 64:31–39
55. Gotea V, Makalowski W (2006) Do transposable elements really contribute to proteomes? Trends Genet 22:260–267
56. Hahn Y, Lee B (2006) Human-specific nonsense mutations identified by genome sequence comparisons. Hum Genet 119:169–178
57. Haile-Selassie Y (2001) Late Miocene hominids from the Middle Awash, Ethiopia. Nature 412:178–181
58. Hambor JE, Mennone J, Coon ME, Hanke JH, Kavathas P (1993) Identification and characterization of an Alu-containing, T-cell-specific enhancer located in the last intron of the human CD8 alpha gene. Mol Cell Biol 13:7056–7070
59. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH Jr (2009) Exon-trapping mediated by the human retrotransposon SVA. Genome Res 19:1983–1991
60. Hasler J, Samuelsson T, Strub K (2007) Useful 'junk': Alu RNAs in the human transcriptome. Cell Mol Life Sci 64:1793–1800
61. Hayward BE, Zavanelli M, Furano AV (1997) Recombination creates novel L1 (LINE-1) elements in Rattus norvegicus. Genetics 146:641–654
62. Horvath JE, Viggiano L, Loftus BJ, Adams MD, Archidiacono N, Rocchi M, Eichler EE (2000) Molecular structure and evolution of an alpha satellite/non-alpha satellite junction at 16p11. Hum Mol Genet 9:113–123
63. Huh JW, Kim DS, Kang DW, Ha HS, Ahn K, Noh YN, Min DS, Chang KT, Kim HS (2008) Transcriptional regulation of GSDML gene by antisense-oriented HERV-H LTR element. Arch Virol 153:1201–1205
64. Illarionova AE, Vinogradova TV, Sverdlov ED (2007) Only those genes of the KIAA1245 gene subfamily that contain HERV(K) LTRs in their introns are transcriptionally active. Virology 358:39–47
65. Irie A, Koyama S, Kozutsumi Y, Kawasaki T, Suzuki A (1998) The molecular basis for the absence of N-glycolylneuraminic acid in humans. J Biol Chem 273:15866–15871
66. Jamain S, Girondot M, Leroy P, Clergue M, Quach H, Fellous M, Bourgeron T (2001) Transduction of the human gene FAM8A1 by endogenous retrovirus during primate evolution. Genomics 78:38–45
67. Jurka J (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA 94:1872–1877
68. Kakuo S, Asaoka K, Ide T (1999) Human is a unique species among primates in terms of telomere length. Biochem Biophys Res Commun 263:308–314
69. Kandel ES, Nudler E (2002) Template switching by RNA polymerase II in vivo. Evidence and implications from a retroviral system. Mol Cell 10:1495–1502
70. Kashkush K, Khasdan V (2007) Large-scale survey of cytosine methylation of retrotransposons and the impact of readout transcription from long terminal repeats on expression of adjacent rice genes. Genetics 177:1975–1985
71. Kazazian HH Jr (2004) Mobile elements: drivers of genome evolution. Science 303:1626–1632

72. Kehrer-Sawatzki H, Cooper DN (2007) Structural divergence between the human and chimpanzee genomes. Hum Genet 120:759–778
73. Kjellman C, Sjogren HO, Salford LG, Widegren B (1999) HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. Gene 239:99–107
74. Kovalskaya E, Buzdin A, Gogvadze E, Vinogradova T, Sverdlov E (2006) Functional human endogenous retroviral LTR transcription start sites are located between the R and U5 regions. Virology 346:373–378
75. Kramerov DA, Vassetzky NS (2001) Structure and origin of a novel dimeric retroposon B1-diD. J Mol Evol 52:137–143
76. Kramerov DA, Vassetzky NS (2005) Short retroposons in eukaryotic genomes. Int Rev Cytol 247:165–221
77. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
78. Lane RP, Cutforth T, Young J, Athanasiou M, Friedman C, Rowen L, Evans G, Axel R, Hood L, Trask BJ (2001) Genomic analysis of orthologous mouse and human olfactory receptor loci. Proc Natl Acad Sci USA 98:7390–7395
79. Lavie L, Kitova M, Maldener E, Meese E, Mayer J (2005) CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). J Virol 79:876–883
80. Lebedev YB, Amosova AL, Mamedov IZ, Fisunov GY, Sverdlov ED (2007) Most recent AluY insertions in human gene introns reduce the content of the primary transcripts in a cell type specific manner. Gene 390:122–129
81. Lee J, Cordaux R, Han K, Wang J, Hedges DJ, Liang P, Batzer MA (2007) Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. Gene 390:18–27
82. Lee JY, Ji Z, Tian B (2008) Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. Nucleic Acids Res 36:5581–5590
83. Lev-Maor G, Ram O, Kim E, Sela N, Goren A, Levanon EY, Ast G (2008) Intronic Alus influence alternative splicing. PLoS Genet 4:e1000204
84. Levy A, Sela N, Ast G (2008) TranspoGene and microTranspoGene: transposed elements influence on the transcriptome of seven vertebrates and invertebrates. Nucleic Acids Res 36:D47–D52
85. Lin L, Shen S, Tye A, Cai JJ, Jiang P, Davidson BL, Xing Y (2008) Diverse splicing patterns of exonized Alu elements in human tissues. PLoS Genet 4:e1000225
86. Livak KJ, Rogers J, Lichter JB (1995) Variability of dopamine D4 receptor (DRD4) gene sequence within and among nonhuman primate species. Proc Natl Acad Sci USA 92:427–431
87. Long Q, Bengra C, Li C, Kutlar F, Tuan D (1998) A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta- globin locus control region. Genomics 54:542–555
88. Lower R, Lower J, Kurth R (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proc Natl Acad Sci USA 93:5177–5184
89. Mager DL, Hunter DG, Schertzer M, Freeman JD (1999) Endogenous retroviruses provide the primary polyadenylation signal for two new human genes. Genomics 59:255–263
90. Makalowski W (2000) Genomic scrap yard: how genomes utilize all that junk. Gene 259:61–67
91. Makalowski W, Mitchell GA, Labuda D (1994) Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet 10:188–193
92. Malik HS, Eickbush TH (2001) Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res 11:1187–1197

93. Martin SL (2006) The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. J Biomed Biotechnol 2006:45621
94. Matlik K, Redik K, Speek M (2006) L1 antisense promoter drives tissue-specific transcription of human genes. J Biomed Biotechnol 2006:71753
95. Medstrand P, Mager DL (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. J Virol 72:9782–9787
96. Medstrand P, van de Lagemaat LN, Mager DL (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res 12:1483–1495
97. Meneveri R, Agresti A, Rocchi M, Marozzi A, Ginelli E (1995) Analysis of GC-rich repetitive nucleotide sequences in great apes. J Mol Evol 40:405–412
98. Mills RE, Bennett EA, Iskow RC, Luttig CT, Tsui C, Pittard WS, Devine SE (2006) Recently mobilized transposons in the human and chimpanzee genomes. Am J Hum Genet 78:671–679
99. Moran JV, DeBerardinis RJ, Kazazian HH Jr (1999) Exon shuffling by L1 retrotransposition. Science 283:1530–1534
100. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet 31:159–165
101. Nadezhdin EV, Vinogradova TV, Sverdlov ED (2001) Interspecies subtractive hybridization of cDNA from human and chimpanzee brains. Dokl Biochem Biophys 381:415–418
102. Nickerson E, Nelson DL (1998) Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. Genomics 50:368–372
103. Nigumann P, Redik K, Matlik K, Speek M (2002) Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 79:628–634
104. Nishihara H, Smit AF, Okada N (2006) Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res 16:864–874
105. Ohshima K, Hamada M, Terai Y, 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
106. Osada N, Wu CI (2005) Inferring the mode of speciation from genomic data: a study of the great apes. Genetics 169:259–264
107. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr (2003) SVA elements are nonautonomous retrotransposons that cause disease in humans. Am J Hum Genet 73:1444–1451
108. Pickeral OK, Makalowski W, Boguski MS, Boeke JD (2000) Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res 10:411–415
109. Polavarapu N, Marino-Ramirez L, Landsman D, McDonald JF, Jordan IK (2008) Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. BMC Genomics 9:226
110. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, Kern AD, Dehay C, Igel H, Ares M Jr, Vanderhaeghen P, Haussler D (2006) An RNA gene expressed during cortical development evolved rapidly in humans. Nature 443:167–172
111. Prabhakar S, Noonan JP, Paabo S, Rubin EM (2006) Accelerated evolution of conserved noncoding sequences in humans. Science 314:786
112. Prudhomme S, Bonnaud B, Mallet F (2005) Endogenous retroviruses and animal reproduction. Cytogenet Genome Res 110:353–364
113. Roy-Engel AM, El-Sawy M, Farooq L, Odom GL, Perepelitsa-Belancio V, Bruch H, Oyeniran OO, Deininger PL (2005) Human retroelements may introduce intragenic polyadenylation signals. Cytogenet Genome Res 110:365–371
114. Royle NJ, Baird DM, Jeffreys AJ (1994) A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. Nat Genet 6:52–56
115. Ruda VM, Akopov SB, Trubetskoy DO, Manuylov NL, Vetchinova AS, Zavalova LL, Nikolaev LG, Sverdlov ED (2004) Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. Virus Res 104:11–16

116. Ruvolo M (2004) Comparative primate genomics: the year of the chimpanzee. Curr Opin Genet Dev 14:650–656
117. Sakaki Y, Watanabe H, Taylor T, Hattori M, Fujiyama A, Toyoda A, Kuroki Y, Itoh T, Saitou N, Oota S, Kim CG, Kitano T, Lehrach H, Yaspo ML et al (2003) Human versus chimpanzee chromosome-wide sequence comparison and its evolutionary implication. Cold Spring Harb Symp Quant Biol 68:455–460
118. Sakate R, Suto Y, Imanishi T, Tanoue T, Hida M, Hayasaka I, Kusuda J, Gojobori T, Hashimoto K, Hirai M (2007) Mapping of chimpanzee full-length cDNAs onto the human genome unveils large potential divergence of the transcriptome. Gene 399:1–10
119. Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G (2007) Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. Genome Biol 8:R127
120. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA (2006) Human genomic deletions mediated by recombination between Alu elements. Am J Hum Genet 79:41–53
121. Song M, Boissinot S (2007) Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. Gene 390:206–213
122. Sorek R, Ast G, Graur D (2002) Alu-containing exons are alternatively spliced. Genome Res 12:1060–1067
123. Sverdlov ED (1998) Perpetually mobile footprints of ancient infections in human genome. FEBS Lett 428:1–6
124. Sverdlov ED (2000) Retroviruses and primate evolution. Bioessays 22:161–171
125. Swanstrom R, Parker RC, Varmus HE, Bishop JM (1983) Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. Proc Natl Acad Sci USA 80:2519–2523
126. Temin HM (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc Natl Acad Sci USA 90:6900–6903
127. Trask BJ, Friedman C, Martin-Gallardo A, Rowen L, Akinbami C, Blankenship J, Collins C, Giorgi D, Iadonato S, Johnson F, Kuo WL, Massa H, Morrish T, Naylor S, Nguyen OT, Rouquier S, Smith T, Wong DJ, Youngblom J, van den Engh G (1998) Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. Hum Mol Genet 7:13–26
128. Ustiugova SV, Amosova AL, Lebedev Iu B, Sverdlov ED (2006) A tissue-specific decrease in the pre-mRNA level of L1- and alu-containing alleles of human genes. Bioorg Khim 32:103–106
129. Ustyugova SV, Lebedev YB, Sverdlov ED (2006) Long L1 insertions in human gene introns specifically reduce the content of corresponding primary transcripts. Genetica 128:261–272
130. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 19:530–536
131. van de Lagemaat LN, Medstrand P, Mager DL (2006) Multiple effects govern endogenous retrovirus survival patterns in human gene introns. Genome Biol 7:R86
132. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM et al (2001) The sequence of the human genome. Science 291:1304–1351
133. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA (2005) SVA elements: a hominid-specific retroposon family. J Mol Biol 354:994–1007
134. Wang X, Grus WE, Zhang J (2006) Gene losses during human origins. PLoS Biol 4:e52
135. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD, Moran JV (2001) Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol 21:1429–1439
136. Wessler SR (1998) Transposable elements and the evolution of gene expression. Symp Soc Exp Biol 51:115–122
137. Wheelan SJ, Aizawa Y, Han JS, Boeke JD (2005) Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. Genome Res 15:1073–1078

138. Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci USA 103:17608–17613
139. Yang Z, Boffelli D, Boonmark N, Schwartz K, Lawn R (1998) Apolipoprotein(a) gene enhancer resides within a LINE element. J Biol Chem 273:891–897
140. Yu H, Jiang H, Zhou Q, Yang J, Cun Y, Su B, Xiao C, Wang W (2006) Origination and evolution of a human-specific transmembrane protein gene, c1orf37-dup. Hum Mol Genet 15:1870–1875
141. Yunis JJ, Prakash O (1982) The origin of man: a chromosomal pictorial legacy. Science 215:1525–1530
142. Zemojtel T, Penzkofer T, Schultz J, Dandekar T, Badge R, Vingron M (2007) Exonization of active mouse L1s: a driver of transcriptome evolution? BMC Genomics 8:392
143. Zimonjic DB, Kelley MJ, Rubin JS, Aaronson SA, Popescu NC (1997) Fluorescence in situ hybridization analysis of keratinocyte growth factor gene amplification and dispersion in evolution of great apes and humans. Proc Natl Acad Sci USA 94:11461–11465