CHAPTER THREE

Unique Functions of Repetitive Transcriptomes

Gerald G. Schumann,[*] Elena V. Gogvadze,[†] Mizuko Osanai-Futahashi,[‡] Azusa Kuroki,[‡] Carsten Münk,[§] Haruko Fujiwara,[‡] Zoltan Ivics,[¶,‖] *and* Anton A. Buzdin[†]

Contents

[*] Paul-Ehrlich-Institut, Federal Institute for Vaccines and Biomedicines, Langen, Germany
[†] Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia
[‡] Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan
[§] Clinic for Gastroenterology, Hepatology and Infectiology, Medical Faculty, Heinrich-Heine-University, Düsseldorf, Germany
[¶] Max Delbruck Center for Molecular Medicine, Berlin, Germany
[‖] University of Debrecen, Debrecen, Hungary

International Review of Cell and Molecular Biology, Volume 285
ISSN 1937-6448, DOI: 10.1016/S1937-6448(10)85003-8

Abstract

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. They generally encode "selfish" proteins necessary for the proliferation of transposable elements (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. 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. In addition, genomic repeats include regulatory sequences like promoters, enhancers, splice sites, polyadenylation signals, and insulators, which actively reshape cellular transcriptomes. TE expression is tightly controlled by the host cells, and some mechanisms of this regulation were recently decoded. Finally, capacity of TEs to proliferate in the host genome led to the development of multiple biotechnological applications.

Key Words: Repetitive sequences, Transposable elements, Retrotransposons, APOBEC 3 proteins, RNA interference, Gene delivery
Genome evolution.

1. Introduction

The eukaryotic genome is a complex and dynamic structure. Only about 3% of the mammalian genome is composed of protein-coding sequences compared to ~50% constituted by transposable elements (TEs). Transposable or mobile genetic elements are DNA sequences that are able to jump into new locations within genomes (Bohne et al., 2008). They can

reach very high copy numbers and represent the major fraction of eukaryotic genomes. Since their initial discovery in the maize genome by Barbara McClintock in 1956 (McClintock, 1956), mobile elements have been found in genomes of almost all organisms. They constitute more than 50% of the maize genome (Wessler, 2006), 22% of the *Drosophila* genome (Kapitonov and Jurka, 2003), and 42% of human DNA (Lander et al., 2001). Initially considered as "junk" DNA or genomic parasites, mobile elements are now suggested to be "functional genome reshapers," which are able to alter gene expression and promote genome evolution (Beauregard et al., 2008; Goodier and Kazazian, 2008; Han and Boeke, 2005).

TEs can be grouped in two major classes (Kazazian, 2004). Class II elements or DNA transposons comprise about 3% of the human genome and most move by a so-called cut-and-paste mechanism. No currently active DNA transposons have been identified in mammals to date (Bohne et al., 2008). Class I elements are termed retrotransposons or retroelements (REs). They move by a "copy-and-paste" mechanism involving reverse transcription of an RNA intermediate and insertion of its cDNA copy at a new site in the host genome. This process is termed retrotransposition. Retrotransposons can be grouped into two major subclasses (Kazazian, 2004). Retroviral-like or long terminal repeat (LTR) retrotransposons include endogenous retroviruses (ERVs), which are relics of past rounds of germline infection by exogenous retroviruses that lost their ability to reinfect and became trapped in the genome because they harbor inactivating mutations that render them replication defective. These elements undergo reverse transcription in virus-like particles (VLPs) by a complex multistep process. LTR-containing REs account for ~10% of the mammalian genomes and their life cycle includes the formation of VLPs that, in several instances—but not systematically—can remain strictly intracellular as observed for the well-characterized murine intracisternal A-particle (IAP) and MusD elements (the so-called intracellularized ERVs; Dewannieux et al., 2004; Ribet et al., 2008), or that can bud at the cell membrane to replicate via an extracellular infection cycle as observed for the recently identified murine intracisternal A-particle-related envelope-encoding element (IAPE; Ribet et al. 2008) and for the 'reconstituted' infectious, human progenitor of the HERV-K(HML2) family members (Dewannieux et al. 2006; Lee and Bienasz, 2007).

The second major subclass comprises the strictly intracellular non-LTR retrotransposons and is represented in the mammalian genome by long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and processed pseudogenes accounting for ~30% of each mammalian genome. Only primate genomes harbor the fourth group of non-LTR retrotransposons termed SVA (SINE–variable number of tandem repeats–Alu-like). The transposition process for non-LTR retrotransposons is fundamentally different from the process observed for LTR

retrotransposons. RNA copies of non-LTR retrotransposons become part of a ribonucleoprotein (RNP) complex and are thought to be carried back into the nucleus where their reverse transcription and integration occur in a single step on the genomic target DNA itself (Goodier and Kazazian, 2008). Major groups of the LTR- and non-LTR retrotransposons are schematized in Fig. 3.1.

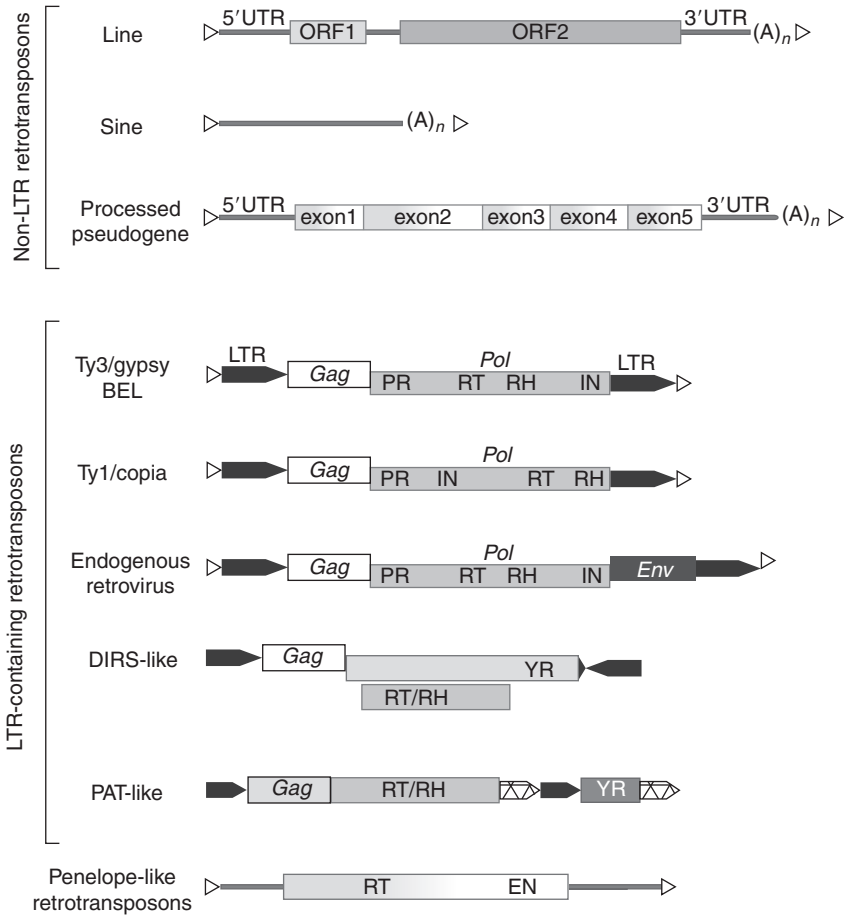


Figure 3.1 Schematic representation of the different types of retrotransposons. White triangles, short direct repeats (target site duplications); UTR, untranslated region; ORF, open reading frame; LTR, long terminal repeat; PR, protease; RT, reverse transcriptase; RH, ribonuclease H; IN, integrase; Env, envelope; YR, tyrosine recombinase; EN, endonuclease.

2. Eukaryotic Retrotransposons

In this section, we focus on class I TEs (retrotransposons) because they generally constitute significant proportions of higher eukaryotic DNAs and are the only group of TEs actively proliferating in the mammalian genomes.

2.1. LINE retrotransposons

LINEs are termed autonomous because they are coding for the protein machinery that is required for their mobilization. They are widely distributed in eukaryotes. About 21% of the human genome is covered by elements that belong to the families LINE1, LINE2, or CR1/LINE3 (Lander et al., 2001). LINE2 and CR1/LINE3 represent ancient inactive fossils that constitute ~3% and ~0.3% of the human genome, respectively, and 0.4% and 0.05% of the mouse genome, accordingly (Gentles et al., 2007). In spite of the low-copy number of LINE2 and LINE3 sequences, their presence may be valuable for the host. For example, a LINE-2 fragment was shown to be a potent T-cell-specific silencer regulatory sequence (Donnelly et al., 1999).

The LINE-1 (L1) family is covering about 500,000 L1 copies occupying ~18% of the haploid genome. L1 elements represent the only family of autonomous non-LTR retrotransposons harboring functional elements that are currently expanding in humans (Goodier and Kazazian, 2008). However, only 80–100 elements are functional and retrotransposition-competent (Brouha et al., 2003). A human full-length L1 is 6 kb long and has a 900-nt 5′-untranslated region (UTR) that functions as an RNA polymerase II internal promoter, two open reading frames (ORF1 and ORF2), a short 3′-UTR, and a poly(A) tail. ORF1 encodes a nucleic acid-binding protein that lacks sequence similarity with any other known protein (Goodier et al., 2007; Han and Boeke, 2005). The ORF2 protein contains endonuclease (EN) and reverse transcriptase (RT) activities as well as a Cys-rich domain, and all three domains are absolutely essential for retrotransposition (Moran et al., 1996). Usually, L1 sequences are flanked by short direct repeats called target site duplications (TSDs) (Fig. 3.1).

L1 retrotransposition is thought to occur by a mechanism termed target-primed reverse transcription (TPRT) (Luan et al., 1993). During TPRT, L1EN recognizes and cleaves the DNA consensus target sequence 5′-TTTT/AA-3′ which means that there are a multitude of potential genomic L1 integration sites (Feng et al., 1996; Jurka, 1997). Due to the *cis*-preference of the L1-encoded protein machinery for its own mRNA, L1 mobilizes preferentially itself (Wei et al., 2001). However, in very rare cases, L1s are able to mobilize *Alu* (Dewannieux et al., 2003) and SVA RNAs (Raiz and Schumann, unpublished data) as well as cellular mRNAs whose retrotransposition results in pseudogene formation (Esnault et al., 2000).

2.2. SINE retrotransposons

SINEs are reiterated, short (80–500 bp) sequences, comprise about 12% of the human genome, and do not code for proteins (Kramerov and Vassetzky, 2005). SINEs harbor an internal promoter, are pol III-transcribed, and possess at their 3′-end a pA-rich tail (Fig. 3.1) Most SINEs within a given family are full-length and are flanked by TSDs of varying length. Structural similarities between LINEs and SINEs suggested early that the LINE-encoded protein machinery is responsible for SINE mobilization. SINEs "hijack" the RT encoded by an autonomous non-LTR retrotransposon for their own mobilization. It is generally accepted that LINEs are used as a source of RT for SINE proliferation (Eickbush, 1992).

In the human genome, SINEs are represented by two major families termed MIR (mammalian-wide interspersed repeats) and *Alu*. MIR elements are tRNA-like SINEs that include ~470,000 copies constituting about 2% of the human genome; while *Alu*s are 7SL RNA-derived elements, include $\sim 1.1 \times 10^6$ elements occupying 10% of the genome (Kramerov and Vassetzky, 2005; Lander et al., 2001). *Alu* elements are the most abundant repeats in the human genome. The major burst of *Alu* retrotransposition took place 50–60 million years ago (mya) and has since dropped to a frequency of one new transposition event in every 20–125 births (Cordaux et al., 2006; Shen et al., 1991).

2.3. SVA elements

SVA elements are primate-specific nonautonomous non-LTR retrotransposons which originated <25 mya and represent the youngest retrotransposon family in primates. So far their copy number has increased to ~3000 in the human genome (Ostertag et al., 2003; Wang et al., 2005). SVA elements stand out from the group of human non-LTR retrotransposons due to their composite structure including modules derived from other primate repetitive elements. Starting at the 5′-end, a full-length SVA element is composed of a $(CCCTCT)_n$ hexamer repeat region; an *Alu*-like region consisting of three antisense *Alu* fragments adjacent to an additional sequence of unknown origin; a variable number of tandem repeat (VNTR) region which is made up of copies of a 36- to 42-bp sequence or of a 49- to 51-bp sequence (Ostertag et al., 2003); and a short interspersed element of retroviral origin (SINE-R) region. The latter is derived from the 3′-end of the *env* gene and the 3′-LTR of the ERV HERV-K10. A poly(A) tail is positioned downstream of the predicted conserved polyadenylation signal AATAAA (Ostertag et al., 2003). Considering the number of disease-causing insertions relative to their overall copy number, SVA elements are thought to represent a highly active retrotransposon family in humans (Ostertag et al., 2003; Wang et al., 2005).

The origin of SVA elements can be traced back to the beginnings of hominid primate evolution, only ~18–25 mya. Their very young evolutionary age represents a unique opportunity to study the entire evolutionary history of a human retrotransposon. In addition, SVA elements may be valuable as markers for primate or human phylogenetic and population genetic studies, as has been the case for *Alu* elements (Bamshad et al., 2003; Watkins et al., 2003; Xing et al., 2007).

2.4. Processed pseudogenes

Processed pseudogenes are found in most mammalian genomes and their structure is that of an integrated cDNA copy of a cellular mRNA: They do contain introns, have lost the untranscribed part of the promoter, end with a polyA tail, and are flanked by TSDs (Fig. 3.1) (Brosius, 1999b; Esnault et al., 2000; Weiner et al., 1986). Similar to Alu retrotransposition, processed pseudogenes were demonstrated to be generated by *trans*-mobilization of cellular mRNAs by the protein machinery encoded by intact LINE retrotransposons (Esnault et al., 2000). In most cases, processed pseudogenes are not functional because they do not include complete promoters and/or because of the accumulation of mutations which occur in the absence of any selection pressure. On rare occasions, processed pseudogenes are functional due to the fortuitous presence of a promoter upstream of the insertion site and the conservation of an intact ORF with a new expression pattern. Generally, there are 1–10 (in some cases up to 100) processed pseudogenes for each human gene (Brosius, 1999b).

2.5. LTR retrotransposons and ERVs

The group is also termed LTR-containing REs and combines LTR retrotransposons and ERVs which are all flanked by LTRs. All LTR-containing elements comprise about 8% of the human genome (Bannert and Kurth, 2004). The organization of LTR retrotransposons is similar to that of retroviruses except for the absence of the *env* (envelope) gene in all LTR retrotransposons (Eickbush and Jamburuthugoda, 2008). LTR-containing REs include a *gag* gene, coding for a structural protein with nucleic acid-binding activity, and a *pol* gene which encodes a polyprotein with protease, RT, RNaseH, and integrase activities. There are three distinct lineages of LTR retrotransposons in vertebrates: the Ty1/copia, Ty3/gypsy, and BEL (Fig. 3.1).

ERVs are relics of past rounds of germline infection by exogenous retroviruses that lost their ability to reinfect and became trapped in the genome because they harbor inactivating mutations that render them replication defective. They were found in all vertebrate genomes and constitute about 8% of the human genomic DNA (Lander et al., 2001). Most ERV

sequences have undergone extensive deletions and mutations, therefore becoming transpositionally deficient and transcriptionally silent (Sverdlov, 2000). Moreover, the majority of ERVs reside in the genome in the form of solitary LTRs, arisen most probably due to homologous recombination between two LTRs of a full-length element.

Tyrosine-recombinase encoding retrotransposons (or YR-retrotransposons) represent an additional group of LTR-containing REs (Poulter and Goodwin, 2005). These elements have structures quite distinct from the REs described above. The major difference is that YR-retrotransposons do not code for an integrase, but for a tyrosine recombinase (YR) instead (Fig. 3.1). The first element of this group was identified in the slime mold, *Dictyostelium discoideum*, and was called DIRS (Cappello et al., 1985). Later on, related elements were found in the genomes of numerous fungi, plants, and animals (Goodwin and Poulter, 2004). All these elements could be divided into two basic groups: DIRS-like elements which are flanked by inverted repeats and contain an internal complementary region, and elements of the PAT and Ngaro type which have split direct repeats (Fig. 3.1). The unusual structure of the terminal repeats of the YR-retrotransposons was suggested to be required for their replication via a free circular intermediate (Cappello et al., 1985; Goodwin and Poulter, 2004). The circular intermediate is believed to be integrated into the genome by site-specific recombination without the formation of TSDs. The human genome contains a DNA sequence similar to a large fragment of a DIRS1-like recombinase gene. However, no full-length mammalian DIRS-like elements have been found to date (Poulter and Goodwin, 2005).

2.6. Penelope-like elements

PLEs constitute a novel class of eukaryotic REs that are distinct from both non-LTR and LTR retrotransposons (Evgen'ev and Arkhipova, 2005) (Fig. 3.1). They were first discovered in *Drosophila virilis* as elements responsible for the hybrid dysgenesis syndrome, and characterized by simultaneous mobilization of several unrelated TE families in the progeny of dysgenic crosses. PLEs were further found in genome databases of various eukaryotes (Gladyshev and Arkhipova, 2007). They have a rather complex and highly variable organization. These elements were shown to contain an internal promoter (Schostak et al., 2008) and one ORF coding for RT and EN activities that differ from the corresponding proteins of LTR-containing and/or non-LTR retrotransposons (Evgen'ev and Arkhipova, 2005). The PLE EN belongs to the URI protein family, which includes, *inter alia*, catalytic modules of the GIY-YIG ENs of group I introns, as well as bacterial UvrC DNA repair proteins. The RT of PLEs mostly resembles the RT domain of telomerase. Both RT and EN domains encoded

by *D. virilis* Penelope are functionally active, but the mechanism of their transposition remains unclear.

3. Mechanisms of Intracellular Defense Against TEs

TEs have played an important role in evolution and speciation. However, mobilization of these elements can also be deleterious to the host and can result in various genetic disorders and cancer. Given these various deleterious effects, it is not surprising that the cell has generated multiple mechanisms controlling their proliferation. To limit the negative effects of retrotransposition, several strategies have been adopted to restrict mobility and potentially deleterious consequences of uncontrolled retrotransposition. Such host-encoded strategies include DNA methylation, RNA interference (RNAi), and inhibition of retrotransposition by the activity of members of the APOBEC (named after apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1, APOBEC1) protein family which comprises 11 closely related DNA or RNA cytidine deaminases.

Epigenetic modifications controlling the activity of TEs were initially reported more than 24 years ago (Chandler and Walbot, 1986). Since then many genes involved in epigenetic silencing of TEs (including DNA methyltransferases and demethylases, histone modifying enzymes, chromatin remodeling enzymes, and genes involved in small RNA metabolism) were characterized. Defects in different components of silencing mechanisms were shown to increase transposition events (Weil and Martienssen, 2008). Most of the methylated cytosines in mammalian genomes reside in repetitive elements and it has been proposed that DNA methylation evolved primarily to suppress the activity of TEs and to protect the host cell (Yoder et al., 1997). Hypomethylation of REs was demonstrated to be associated with genomic instability in cancer (Daskalos et al., 2009).

On the one hand, chromatin condensation may suppress the activity of REs. On the other hand, DNA methylation, initiated within RE, may spread to the surrounding genomic regions and, hence, suppress their functional activity. Spreading of CpG methylation from SINEs into flanking genomic regions was suggested to create distal epigenetic modifications in plants (Arnaud et al., 2000). Human Alu elements were proposed as potential *de novo* methylation centers implicated in tumor suppressor gene silencing in neoplasia (Graff et al., 1997). Recent studies have shown the involvement of RNAi-related mechanisms in the control of TE activities, in particular, in DNA methylation of TE sequences and in the formation of heterochromatin. Plants, yeasts, and animals use different strategies to detect transposons and to generate small RNAs against them (Girard and Hannon, 2008; Slotkin and Martienssen, 2007).

3.1. Impact of AID on retrotransposition

The activation-induced cytidine deaminase (AID) gene is evolutionary quite old because it is part of the genomes of vertebrates down to jawless vertebrates (Rogozin et al., 2007). AID deaminates cytidines to uridines (C to U) in single-stranded DNA and is required for antibody maturation involving somatic hypermutation at the immunoglobulin variable regions, class switch recombination at the switch regions, and Ig gene conversion in some species. In general, the enzyme is mainly expressed in the B-cell compartment. In the mouse, AID is detectable in the spleen, ovaries, and oocytes (MacDuff et al., 2009; Morgan et al., 2004) and is moderately expressed in the heart (MacDuff et al., 2009) and in murine embryonic stem (ES) and germ cells (Morgan et al., 2004). Contrasting to the mouse, humans express AID only in B-cells and testes (MacDuff et al., 2009; Schreck et al., 2006).

L1 retrotransposition reporter assays performed separately in HEK293 cells in the presence of overexpressed HA-tagged AID proteins from different species demonstrated that human L1 retrotransposition is inhibited by ~24–68% by AID from human, mouse, rat, chicken, pufferfish, and zebrafish, while porcine AID restricted L1 activity by even ~90% (MacDuff et al., 2009). Interestingly, neither mutations in the catalytically active site nor mutations in the predicted DNA-binding site of human AID had any consequences for the L1-inhibiting effect of AID. This is indicating that AID-mediated inhibition of L1 is both cytidine deaminase- and DNA-binding-independent (MacDuff et al., 2009). In an earlier report, inhibition of L1 retrotransposition activity by human AID could not be demonstrated (Niewiadomska et al., 2007).

In the presence of overexpressed wild-type AID proteins from multiple species or the catalytically inactive mutant AID-E58Q, retrotransposition of the mouse LTR retrotransposon MusD was inhibited by only 13–50% in HeLa cells (Esnault et al., 2006; MacDuff et al., 2009). Under such experimental conditions, AID induced a small number of mutations in *de novo* MusD retrotransposition events (Esnault et al., 2006). In contrast to the MusD element, the yeast LTR retrotransposon Ty1 was not affected by coexpression of AIDs in yeast cells (MacDuff et al., 2009).

It was reported that neither APOBEC1 nor APOBEC2 had any effect on MusD retrotransposition in HeLa cells (Chen et al., 2006; Esnault et al., 2006). Also, L1- and IAP retrotransposition were not impaired by A2 (Chen et al., 2006; Niewiadomska et al., 2007).

3.2. APOBEC3 proteins

The APOBEC3 (A3; apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3) proteins are Zn^{2+}-dependent DNA cytidine deaminases, which were discovered to constitute a defensive network of proteins

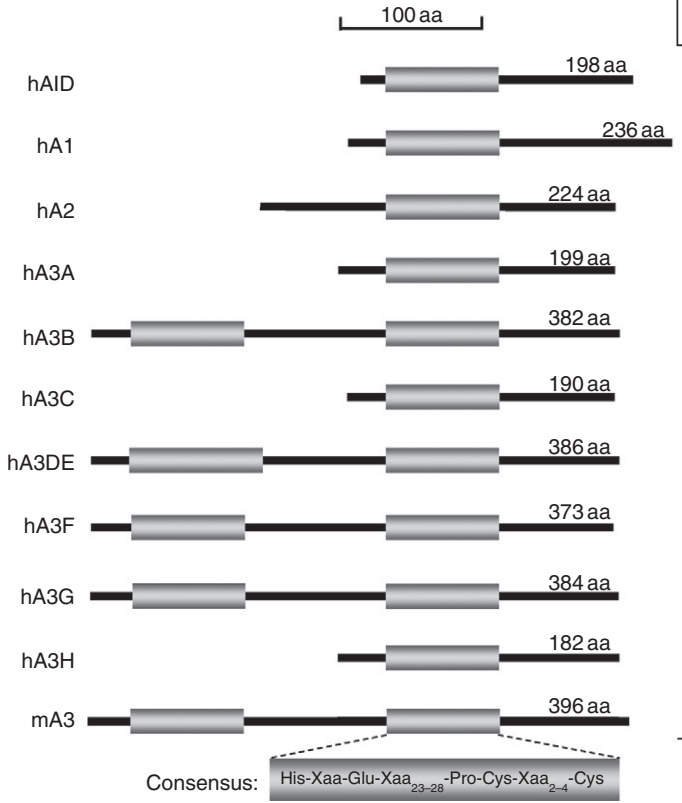
that restrict the replication of retroviruses (Bishop et al., 2004; Chiu and Greene, 2008) and an incredible range of mobile genetic elements (Chiu and Greene, 2008; Goila-Gaur and Strebel, 2008; Hultquist and Harris, 2009; Malim and Emerman, 2008). The A3 genes are only present in placental mammals (LaRue et al., 2008; Munk et al., 2008). Phylogenetic studies have indicated that the first A3 gene(s) arose from an AID-like ancestral gene through a series of duplication and diversification events (Conticello, 2008; Conticello et al., 2007; LaRue et al., 2008). The APOBEC3 gene family has proliferated during mammalian speciation and many members of which exhibit signs of positive (diversifying) selection in the primate and felid lineages and accordingly are highly polymorphic (Kidd et al., 2007; Munk et al., 2008; OhAinle et al., 2006, 2008; Sawyer et al., 2004; Zhang and Webb, 2004).

The seven A3 genes are positioned in tandem on human chromosome 22: A3A, A3B, A3C, A3D (formerly A3DE), A3F, A3G, and A3H (Jarmuz et al., 2002; OhAinle et al., 2006). A defining feature of each A3 gene is that it encodes a protein with one or two conserved zinc (Z)-coordinating deaminase domains. Each Z domain belongs to one of three distinct phylogenetic groups: Z1 (A3A and the C-terminal halves of A3B and A3G), Z2 (A3C, both domains of A3D and A3F, and the N-terminal halves of A3B and A3G), and Z3 (A3H) (LaRue et al., 2009). Based on the relatedness of these Z domains, the human A3 repertoire appears to be the result of a minimum of eight unequal crossing-over recombination events, which mostly occurred during the radiation of primates (LaRue et al., 2008; Munk et al., 2008). The net result is that the human A3 mRNAs share considerable identity, ranging from 30% to nearly 100% (Refsland et al., 2010).

The members of the human A3 protein family differ from each other with respect to their intracellular localization after overexpression (Fig. 3.2). While A3B is found exclusively in the nucleus, A3A is predominantly located to the nucleus. A3C is equally distributed and A3H is located in the cytoplasm and in nucleoli, while A3F, A3D, and A3G appeared exclusively in the cytoplasmic compartment (Bogerd et al., 2006a; Kinomoto et al., 2007; Muckenfuss et al., 2006; Stenglein and Harris, 2006; Zielonka et al., 2009).

3.2.1. APOBEC3 deaminases as inhibitors of LTR retrotransposons

APOBEC3 proteins do in fact function as inhibitors of LTR retrotransposons. Human A3A, A3B, A3C, A3F, A3G, and mA3 all effectively restrict mouse IAP and MusD elements (Bogerd et al., 2006a; Chen et al., 2006; Esnault et al., 2005, 2006), whereas hA3C, hA3F, and hA3G inhibit retrotransposition of the yeast Ty1 element (Dutko et al., 2005; Schumacher et al., 2005). APOBEC3 proteins exert dual inhibitory effects on these ERVs, involving both a decrease in the number of transposed cDNA copies and extensive editing of the transposed copies (Esnault et al., 2005, 2006).



Subcellular localization	Inhibitory effect of APOBEC proteins on					
	L1	Alu	HERV-K	IAP/IAPE	MusD	Ty1
N/C	$+^{1}$ $-_{1}$	/	$+^{1,7}$	/	$+^{1,3}$	$\pm^{1,3}$
N/C	/	/	/	/	$-_{3}$	/
n.d.	$-_{2,4}$	/	/	$-_{4}$	$-_{3,4}$	/
N/C	$++^{1,2,4,5,11,12}$ $++^{13,15,16}$	$++^{5,12}$	$++^{9,10,19}$	$++^{4,5,8}$	$++^{4}$	/
N	$++^{1,2,4,5,11,13}$ $++^{14,15,16}$	$++^{5}$	$++^{9,10,19}$	$++^{4,5,8}$	$++^{4}$ $-_{3}$	/
N/C	$++^{2,4,5,11,15}$ $++^{16}$	$+^{5}$	$-_{9,10,19}$	$++^{4}$	$++^{4}$ $-_{3}$	$++^{4}$
N/C	$+^{2,16}$	$+^{12}$	$++^{10,19}$ $-_{9}$	/	/	/
C	$++^{1,2,4,11,14}$ $++^{15,16}$ $-_{5,17}$	$++^{11}$	$++^{9,10,19}$	$++^{4}$	$++^{3,4}$	$++^{6,7}$
C	$++^{1,2,11,14,16}$ $-_{4,5,15,17,18,20}$	$+^{11,12,17}$ $+^{21}$	$+^{10}$ $-_{9,19}$	$++^{4,8,10,18}$	$++^{4,3,18}$	$++^{6,7}$
N/C	$+^{12,16}$ $-_{15,11}$	$+^{12}$	$+^{10}$ $-_{9}$	/	/	/
C	$+^{13}$	/	$-_{10,19}$	$++^{4,10,18}$	$++^{4,13,18}$	$++^{6}$

These effects are reminiscent of the dual effects of hA3G in HIV-1 replication. In the mouse genome, many preexisting retrotransposon sequences bear mutations consistent with APOBEC3-mediated deamination (Esnault et al., 2005, 2006). Interestingly, hA3A effectively inhibits IAP and MusD retrotransposition through a novel deamination-independent mechanism (Bogerd et al., 2006a).

One human ERV [HERV-K(HML-2)], which has replicated in humans for the past few million years but is now thought to be extinct, was recently reconstituted in a functional form in two separate laboratories by aligning human-specific proviruses and synthesizing a pseudo-ancestral proviral construct termed HERV-K_{CON} and Phoenix, respectively (Dewannieux et al., 2006; Lee and Bieniasz, 2007). Of all tested human A3 proteins, only hA3A, hA3B, and hA3F were shown to be intrinsically capable of mutating and inhibiting infection by HERV-K $_{CON}$ in cell culture by up to 80% while hA3G led only to a marginal restriction of infection (Lee et al., 2008b). Differing from these results, Esnault and coworkers reported that the infectious HERV-K and murine IAPE elements (Ribet et al., 2008) are both restricted efficiently by the murine A3 protein and by hA3G in an *ex vivo* assay for infectivity (Esnault et al., 2008). They also demonstrated that hA3A, hA3B, hA3D, and hA3F restrict infectivity of HERV-K but do not affect IAPE infectivity. The same report presented evidence of strand-specific G-to-A editing of both proviruses (Esnault et al., 2008). For HERV-K, G-to-A editing was not observed with hA3A (Esnault et al., 2008). *In silico* analysis of the naturally occurring genomic copies of the corresponding endogenous elements performed on the mouse and human genomes discloses "traces" of A3-editing, with the specific signature of the murine A3 and human A3G enzymes, respectively, and to a variable extent depending on the family member (Esnault et al., 2006).

There is striking evidence indicating that two HERV-K(HML-2) proviruses that are fixed in the modern human genome (HERV-K60 and HERV-KI) were subjected to hypermutation by A3G (Lee et al., 2008b).

Figure 3.2 Members of the APOBEC protein family and their effects on the activity of retroelements. The domain organization of human and murine APOBEC proteins is depicted. Gray bars represent CDA motifs in each protein. The consensus amino acid sequence of the catalytic domains and the numbers of amino acids (aa) that compose the proteins are shown. Nuclear (N) and cytoplasmic (C) localization of the proteins as well as publications reporting major (++), modest (+), minimal (+/−), and no (−) effects are indicated. /, not determined (n.d.). 1, MacDuff et al. (2009); 2, Niewiadomska et al. (2007); 3, Esnault et al. (2006); 4, Chen et al. (2006); 5, Bogerd et al. (2006b); 6, Dutko et al. (2005); 7, Schumacher et al. (2005); 8, Bogerd et al. (2006a); 9, Lee et al. (2008b); 10, Esnault et al. (2008); 11, Khatua et al. (2010); 12, Tan et al. (2009); 13, Lovsin and Peterlin (2009); 14, Stenglein and Harris (2006); 15, Muckenfuss et al. (2006); 16, Kinomoto et al. (2007); 17, Hulme et al. (2007); 18, Esnault et al. (2005); 19, Turelli et al. (2004); 20, Chiu et al. (2006).

These are rare examples for the antiretroviral effects of A3G in the setting of natural human infection, whose consequences have been fossilized in human DNA.

3.2.2. APOBEC3 deaminases as inhibitors of L1 retrotransposition

It was reported concordantly that A3A and A3B are potent inhibitors of human L1 retrotransposition causing a reduction of retrotransposition frequency by 85–99% and 75–90%, respectively, while A3C-mediated inhibition is less pronounced (40–75%) (Bogerd et al., 2006b; Chen et al., 2006; Khatua et al., 2010; Kinomoto et al., 2007; Lovsin and Peterlin, 2009; MacDuff et al., 2009; Muckenfuss et al., 2006; Niewiadomska et al., 2007; Stenglein and Harris, 2006; Tan et al., 2009). A3F-mediated L1 inhibition was shown to range from 66% to 85% (Chen et al., 2006; Khatua et al., 2010; Kinomoto et al., 2007; MacDuff et al., 2009; Muckenfuss et al., 2006; Niewiadomska et al., 2007; Stenglein and Harris, 2006), but these findings were questioned in two recent reports (Bogerd et al., 2006b; Hulme et al., 2007). The effect of A3G on L1 retrotransposition is more controversial because results presented in one-half of all studies argue against any A3G-mediated inhibitory effect (Bogerd et al., 2006b; Chen et al., 2006; Esnault et al., 2005; Hulme et al., 2007; Muckenfuss et al., 2006; Turelli et al., 2004) while the other half demonstrates L1 restriction by A3G by ~30–90% (Khatua et al., 2010; Kinomoto et al., 2007; MacDuff et al., 2009; Niewiadomska et al., 2007). In the case of A3D, both minor inhibition by 35–45% (Kinomoto et al., 2007) and major inhibition by ~95% (Niewiadomska et al., 2007) were referred. Previous reports have shown that A3H restricts L1 activity by ~50% (Kinomoto et al., 2007) or not at all (Khatua et al., 2010; Muckenfuss et al., 2006). This was explained by the existence of human Z2-type cytidine deaminase A3H variants that have varying intrinsic abilities to restrict REs. It was found that in contrast to A3H, the variant A3H-Var is a highly effective inhibitor of L1 retrotransposition being almost as potent as A3A (OhAinle et al., 2008; Tan et al., 2009). The frequency of this A3H variant allele, A3H-Var, is the highest among sub-Saharan Africans and is significantly lower in Asian and European populations (Tan et al., 2009). Only little human L1 restriction by 32% is mediated by mA3 (Lovsin and Peterlin, 2009).

The mechanisms responsible for A3-mediated L1 inhibition are unclear to date. So far, there is no direct evidence for L1 inhibition by cytidine deamination of L1 cDNA or any kind of editing of L1 nucleic acid sequences, strongly suggesting that cytidine deaminase-independent mechanisms are involved in A3-mediated L1 restriction.

Niewiadomska and coauthors reported that A3A was associated with L1 RNA in high-molecular mass (HMM) complexes that presumably contain L1 RNPs. Consistently, A3A–HMM complexes were destroyed by RNase treatment (Niewiadomska et al., 2007). An interaction between L1 ORF1

protein (L1 ORF1p) and A3A could not be demonstrated (Lovsin and Peterlin, 2009). A catalytically active cytidine deaminase domain (CDD) was shown to be essential for the interaction of A3A with L1 RNA and for the ability of A3A to inhibit L1 retrotransposition, even though no G-to-A hypermutations were detectable in L1 *de novo* retrotransposition events that occurred in the presence of A3A. It was concluded that A3A might indirectly interfere with L1 metabolism, probably by binding L1 RNA. A3A may also interfere with L1 reverse transcription/integration, similar to A3G-mediated restriction of Vif-deficient HIV-1 (Bishop et al., 2004). Alternatively, the association of A3A or other hA3 proteins with the L1 RNP could impede the intracellular movement of the L1 RNP and therefore its retrotransposition (Niewiadomska et al., 2007).

Interaction between L1 RNA and hA3 proteins is also supported by the fact that it was not possible to date to demonstrate any direct interaction between hA3 proteins and L1 ORF1p in the absence of RNA (Lovsin and Peterlin, 2009). Data were presented indicating that A3B and mA3 bind to L1 ORF1p via an RNA bridge (Lovsin and Peterlin, 2009).

3.2.3. APOBEC3 deaminases as inhibitors of *Alu* retrotransposition

Retrotransposition of *Alu* elements is mediated by the L1 ORF2 protein (L1 ORF2p) which has RT and EN activities but does not require the RNA-binding L1 ORF1p (Babushok et al., 2007; Dewannieux et al., 2003). Transient expression of hA3A, hA3B, hA3D, or hA3G was shown to restrict *Alu* retrotransposition frequency by 75–98% regardless of whether L1 ORF2p alone or both L1 ORF1p and L1 ORF2p were coexpressed (Chiu et al., 2006; Hulme et al., 2007; Khatua et al., 2010; Tan et al., 2009). A3C and A3H restrict *Alu* retrotransposition by only 50–70% and ~65%, respectively (Bogerd et al., 2006b; Tan et al., 2009).

Several groups have reported that A3G is localized to P-bodies and stress granules which are sites of mRNA storage and metabolism, raising the question of whether P-bodies and/or stress granules play a role in A3G-mediated *Alu* retrotransposition. A3G-mediated inhibition of *Alu* retrotransposition is thought to result from *Alu* RNA sequestration by A3G in cytoplasmic HMM complexes, particularly Staufen-containing RNA granules, denying these REs access to the nuclear L1 machinery. These effects appear to explain how hA3G interdicts the *Alu* retrotransposition cycle (Chiu et al., 2006; Hulme et al., 2007). This inhibitory mechanism does not involve editing of the *Alu* RNA and also differs from hA3A- and hA3B-mediated inhibition of *Alu* retrotransposition which is based on the alteration of the activity of the L1 machinery in the nucleus by the A3 protein.

A3A and A3H which also act as potent *Alu* inhibitors have been reported to be localized to both cytoplasm and nucleus. Unlike A3G, both A3A and A3H interacted poorly with *Alu* RNAs. However, A3H

associates with HMM complexes while A3A interacts poorly with P-bodies and mRNA-containing HMM complexes (Niewiadomska et al., 2007), suggesting that A3A-mediated suppression of *Alu* retrotransposition is not linked to P-bodies.

Altogether, these results suggest that different A3 proteins may have evolved distinct inhibitory mechanisms against *Alu* REs. It is reasonable to hypothesize that A3 cytidine deaminases such as A3H and A3A have evolved to inhibit *Alu* mobilization by interfering with components of the L1 machinery and/or host factors that are required for *Alu* retrotransposition (Tan et al., 2009).

Recently, exosomes secreted by $CD4^+$ H9 T cells were reported to encapsidate A3G and A3F and to inhibit *Alu* retrotransposition by 92–96%, being almost equally potent inhibitors as A3G and A3F themselves (Khatua et al., 2010). Exosomes secreted by mature monocyte-derived dendritic cells (M-DC) that expressed A3G inhibited *Alu* retrotransposition by 55%. The data indicated that the inhibitory effect of exosomes against *Alu* mobilization is caused, at least in part, by the presence of encapsidated A3G. H9 exosomes that were originally found to encapsidate A3G and inhibit HIV-1 replication (Khatua et al., 2010) had a strong inhibitory effect against L1. They were also found to encapsidate mRNAs coding for A3C, A3F, and A3G. Since A3G mRNA isolated from exosomes was shown to be functional and supports protein synthesis in an *in vitro* translation system, the authors speculated that transfer of functional A3 proteins and corresponding mRNAs modulate or enable cells to resist invading or endogenous REs. It was suggested that exosomes with encapsidated A3 proteins may serve to fine tune the response against transiently expressed REs in germ cells, during early stages of human embryogenesis, or even in somatic cells (Khatua et al., 2010).

3.2.4. A3 expression profile in human tissues

The expression profile of each of the seven human A3 genes was determined by RT-PCR, quantitative RT-PCR, and Northern blot analysis (Koning et al., 2009; Refsland et al., 2010; Schumann, 2007). The expressed A3 repertoire was profiled in 25 distinct human tissues, common T-cell lines, a variety of primary hematopoietic cell types, tumors, and tumor cell lines (Koning et al., 2009; Refsland et al., 2010; Schumann, 2007). It was demonstrated that multiple A3 genes are expressed constitutively in most types of cells and tissues, and that distinct A3 genes are induced upon T-cell activation and interferon treatment (Refsland et al., 2010). The relatively high expression levels of A3 proteins in human testis and ovary (hA3G, hA3F, and hA3C) (Jarmuz et al., 2002; Koning et al., 2009; OhAinle et al., 2006) and in ES cells (hA3B) (Bogerd et al., 2006a) point to a physiologically relevant role for these DNA deaminases in the protection of these cells from the potentially deleterious effects of

endogenous RE mobilization. Brain tissue is exceptional because it expresses virtually no A3 proteins (Refsland et al., 2010). This is consistent with the observation that endogenous L1 elements retrotranspose in neural progenitor cells (Coufal et al., 2009). More generally, nearly every cell type and tissue expresses multiple A3s, consistent with a model in which parasitic elements must evolve ways to cope with a constitutive set of restriction factors that can be further fortified by transcriptional induction.

3.3. Evidence for ADAR editing of *Alu* elements

RNAs in higher eukaryotes can be subjected to a posttranscriptional modification called RNA editing by adenosine deaminases acting on RNA (ADARs). This process involves modification of individual adenosine bases to inosine in RNA molecules. Inosine acts as guanosine during translation, and A-to-I conversion in coding sequences leads to amino acid changes, alterations of transcriptional start and stop codons, as well as RNA splice sites. When comparing genomic with cDNA sequences, edited sites are identified by A-to-G transitions because inosine base pairs with cytosine and, therefore, is replaced by guanosine during reverse transcription and PCR amplification.

RNA editing patterns characteristic of ADAR enzymes have been detected in several viral RNAs, including those of measles virus, influenza virus, hepatitis delta virus, and hepatitis C virus. To date, there is no evidence for any ADAR-mediated modulation of the activity of TEs from tissue culture experiments. However, *in vivo* findings described below indicate an intimate relationship between ADARs and retrotransposons.

The total number of currently known A-to-I edited genes in mammals is small. ADAR editing is functionally crucial for the expression of some neurotransmitter receptors in the brain, and ADAR1-deficient mice show embryonic lethality. Usually, both the edited and the unedited versions of the RNA and/or protein coexist in the same cell. ADARs recognize and edit through their interaction with the complete or incomplete dsRNA structure formed between the edited site which is located within an exon sequence and its complementary sequence usually located in an intron sequence. Currently, it is not possible to predict if and to what extent a given RNA molecule is a substrate for A-to-I editing *in vivo*. Three distinct ADAR genes have been identified in mammals. ADAR1 is ubiquitously expressed in mammalian tissues; ADAR2 expression levels are highest in the brain. ADAR3, exclusively expressed in the brain has so far not shown any catalytic activity using synthetic dsRNA or known ADAR targets.

Because the p150 isoform of ADAR1 is interferon inducible and upregulated in immune cells during inflammation, it is likely that ADAR1 is important for the cellular resistance to pathogens. Furthermore, ADAR1

is involved in the RNAi pathway and is known to alter both the targeting and the processing of microRNAs (miRNA).

Since 2004 several thousands of edited human mRNAs were identified (Athanasiadis et al., 2004). Clusters of A-to-G discrepancies in the cDNAs were found to be the result of RNA editing involving intermolecular pairs of inverted *Alu* repeat sequences (Athanasiadis et al., 2004). It was suggested that the vast majority of primary human gene transcripts are subject to A-to-I RNA editing (Athanasiadis et al., 2004; Blow et al., 2004; Kim et al., 2004).

Signs of ADAR editing were detected in initial screens of small numbers of cDNAs of genes as well as in subsequent larger computational surveys (~2000 different genes) of the majority of transcripts. In most of these cases, the location of the A-to-G cluster coincides with the position of a repetitive element, such as *Alu* and L1 present in introns or UTRs in the cDNAs (Athanasiadis et al., 2004; Blow et al., 2004; Kim et al., 2004; Levanon et al., 2004). These findings suggested that repetitive elements, such as *Alu* sequences, in RNAs might be frequent targets of ADAR editing which presumably requires the intramolecular pairing of two oppositely oriented base pairing repeat elements within the RNA molecule (Athanasiadis et al., 2004; Blow et al., 2004). It has been shown that many hyperedited, inosine-containing RNAs are restrained in the nucleus by a protein complex containing the inosine-binding protein $p54^{nrb}$ (also known as NONO), PSF, and matrin3 (Zhang and Carmichael, 2001).

In view of the widespread editing of *Alu* sequences, this offers an intriguing mechanism to mark nonstandard transcripts and preclude aberrantly spliced mRNAs and repetitive elements containing RNAs from exiting the nucleus (Athanasiadis et al., 2004; Kim et al., 2004). In mice, the ADAR editing sites are mainly found in B1 and B2 SINEs, in L1 and MaLR LTR sequences (Neeman et al., 2006; Riedmann et al., 2008).

3.4. piRNAs and PIWI proteins as regulators of mammalian retrotransposon activity

DNA methylation and RNAi (Carmell and Hannon, 2004) are independent pathways that are restricting retrotransposons and can combine to form a powerful and redundant additional mechanism for keeping retrotransposons in check. This mechanism utilizes small noncoding RNAs (ncRNAs) that guide the effector complex, which is including members of the PIWI/ARGONAUTE protein family to degrade and/or suppress target mRNAs encoded by LTR- and non-LTR retrotransposons.

Members of the evolutionarily conserved PIWI/ARGONAUTE protein family are only expressed in germ cells and are key players in RNA silencing (Hutvagner and Simard, 2008). The PIWI/ARGONAUTE protein family can be subdivided into AGO and PIWI subfamilies. AGO proteins bind to small interfering RNAs (siRNAs) and miRNAs, and

have been shown to play crucial roles in the siRNA and miRNA pathways in many tissues. PIWI proteins bind to a novel class of germ cell-specific ncRNAs called PIWI-interacting RNAs (piRNAs) (Aravin et al., 2006; Girard and Hannon, 2008; Lau et al., 2006; Watanabe et al., 2006) and have diverse functions in germline development and gametogenesis (Cox et al., 1998). Several lines of evidence indicate that PIWI proteins lead to epigenetic repression of retrotransposon-encoding regions presumably through piRNAs (Thomson and Lin, 2009).

Mammalian PIWI family genes, including the three mouse PIWI homologs, Miwi, Mili, and Miwi2 (Carmell and Hannon, 2004; Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004), show germ cell-specific expression and are essential for spermatogenesis (Lin, 2007; Peters and Meister, 2007; Siomi and Kuramochi-Miyagawa, 2009). The phenotypes of Mili and Miwi2 gene-targeted mice were essentially the same and showed male sterility due to apoptosis of the germ cells at early pachytene phase (Carmell et al., 2007; Kuramochi-Miyagawa et al., 2004). MILI, which is expressed in primordial germ cells (PGCs) at embryonic day 12.5 to round spermatids, binds to 26- to 27-nucleotide (nt) piRNAs (Aravin et al., 2006; Kuramochi-Miyagawa et al., 2008). MIWI2, which is expressed in fetal gonocytes from embryonic day 15.5 until soon after birth, binds to 28- to 29-nt piRNAs (Aravin and Bourc'his, 2008; Aravin and Hannon, 2008; Kuramochi-Miyagawa et al., 2010; Thomson and Lin, 2009). About 25% of the piRNAs at the fetal stage were derived from LTR retrotransposon/ERV (MER, ERVK, ERVL, ERV1, MaLR) and non-LTR retrotransposon sequences (L1, SINEs), and the production of piRNA was markedly impaired in MILI- and MIWI2-deficient mice (Kuramochi-Miyagawa et al., 2008). MILI and MIWI2 have been implicated in the repression of LTR retrotransposon IAP and non-LTR retrotransposon L1, with methylation of the L1 5′-UTR being reduced in newborn mice defective in these proteins (Aravin et al., 2007a; Carmell et al., 2007). These data suggest that MILI and MIWI2 are involved in piRNA production in the fetal male gonads, and that the piRNA production would play some important role(s) in gene silencing of retrotransposons via DNA methylation. MILI is central to the primary processing of sense piRNAs from retrotransposon mRNAs and other cellular transcripts (Aravin and Hannon, 2008; Aravin et al., 2007b; Kuramochi-Miyagawa et al., 2008). Then, the primary piRNAs guide the production of secondary piRNAs, which are loaded onto MIWI2, from mostly antisense RNAs transcribed from retrotransposons and other genomic elements (Aravin and Hannon, 2008). The loss of Mili leads to a gross reduction in total piRNAs and those loaded onto MIWI2.

piRNAs are processed from long precursors encoded by large primary transcripts. Often piRNAs cluster to arrays that appear to be bidirectionally transcribed, while less often are primary transcripts derived from one strand. The clusters are transcribed as long transcripts that go through primary

processing to give rise to mature piRNAs, but the mechanism of primary processing is not understood. piRNAs themselves may have a role in retrotransposon repression, as the deletion of a small piRNA cluster in mice leads to increased retrotransposon activity, consistent with a role for piRNAs in retrotransposon regulation (Xu et al., 2008).

To date, two additional host-encoded factors were demonstrated to be involved in the PIWI-mediated restriction of L1 retrotransposons: MVH (mouse vasa homolog) is the homolog of the VASA protein which is an evolutionarily conserved RNA helicase essential for germ cell development in *Drosophila*. Expression of the murine MVH is restricted to the germ cell lineage and can be observed in male germ cells from embryonic day 10.5 to round spermatids (Toyooka et al., 2000), which covers the period of *de novo* DNA methylation of retrotransposons. Since MILI and MIWI were found to bind to MVH (Kuramochi-Miyagawa et al., 2004), it was postulated that MVH plays some role(s) in piRNA production and subsequent DNA methylation of retrotransposons. Indeed, essential roles for MVH in *de novo* methylation of retrotransposons could be confirmed, and it was demonstrated that MVH is an essential factor in the piRNA processing pathway (Kuramochi-Miyagawa et al., 2010). Also, TDRD1, a tudor-domain-containing protein, associates with MILI, participates in the PIWI pathway to suppress retrotransposons (Reuter et al., 2009; Wang et al., 2009) and is essential for retrotransposon silencing and male meiosis in mice. During male germ cell development, Tdrd9 participates in ensuring a proper piRNA profile and in establishing DNA methylation of L1. The TDRD9 protein forms a discrete subcellular compartment with MIWI2 under the control of *Mili* in fetal prospermatogonia (Shoji et al., 2009).

4. The Use of Transposable Elements in Biotechnology and in Fundamental Studies

4.1. DNA transposons as genetic tools

Class II TEs or DNA transposons are discrete pieces of DNA with the ability to change their positions within the genome via a "cut-and-paste" mechanism called transposition. In nature, these elements exist as single units containing the transposase gene flanked by inverted terminal repeats (ITRs) that carry transposase-binding sites (Fig. 3.3A). However, under laboratory conditions, it is possible to use transposons as bicomponent systems, in which virtually any DNA sequence of interest can be placed between the transposon ITRs and mobilized by *trans*-supplementing the transposase in form of an expression plasmid (Fig. 3.3B) or mRNA synthesized *in vitro*. In the transposition process, the transposase enzyme mediates the excision of the element from its donor plasmid, followed by

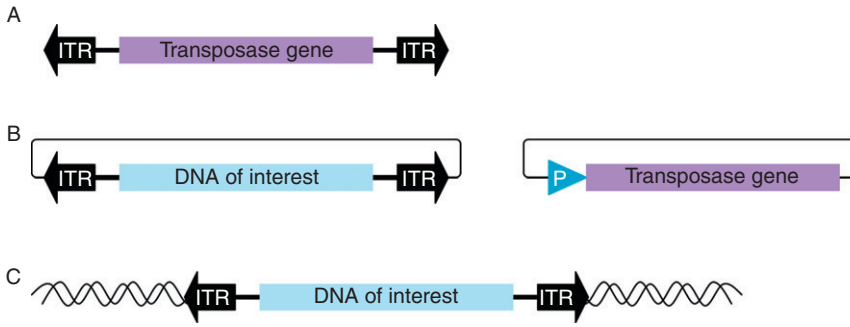


Figure 3.3 General organization of class II transposable elements and mechanism of transposition. (A) Autonomous transposable elements consist of inverted terminal repeats (ITR; black arrows) that flank the transposase gene. (B) Bicomponent transposon vector system for delivering transgenes that are maintained in plasmids. One component contains a DNA of interest between the transposon ITRs carried by a plasmid vector, while the other component is a transposase expression plasmid. Blue triangle represents the promoter (P) driving expression of the transposase. (C) The transposon carrying a DNA of interest is integrated at a chromosomal donor site.

reintegration of the transposon into a chromosomal locus (Fig. 3.3C). This feature makes transposons natural and easily controllable DNA delivery vehicles that can be used as tools for versatile applications ranging from somatic and germline transgenesis to functional genomics and gene therapy (Fig. 3.4).

4.1.1. The *Sleeping Beauty* transposon system

Even though DNA transposons have been extensively harnessed as tools for genome manipulation in invertebrates (Cooley et al., 1988; Thibault et al., 2004; Zwaal et al., 1993), there was no known transposon that was active enough to be used for such purposes in vertebrates. In 1997, Ivics et al. succeeded to engineer the *Sleeping Beauty* (*SB*) transposon system by molecular reconstruction of an ancient, inactive Tc1/*mariner*-type transposon found in several fish genomes (Ivics et al., 1997). This newly reactivated element allowed highly efficient transposition-mediated gene transfer in major vertebrate model species without the potential risk of cross mobilization of endogenous transposon copies in host genomes. Indeed, *SB* has been successfully used as a tool for genetic modifications of a wide variety of vertebrate cell lines and species including humans (Ivics et al., 2009; Mates et al., 2007, 2009).

However, although the resurrected *SB* element was active enough to be mobilized in vertebrate cells, its transpositional activity still presented a

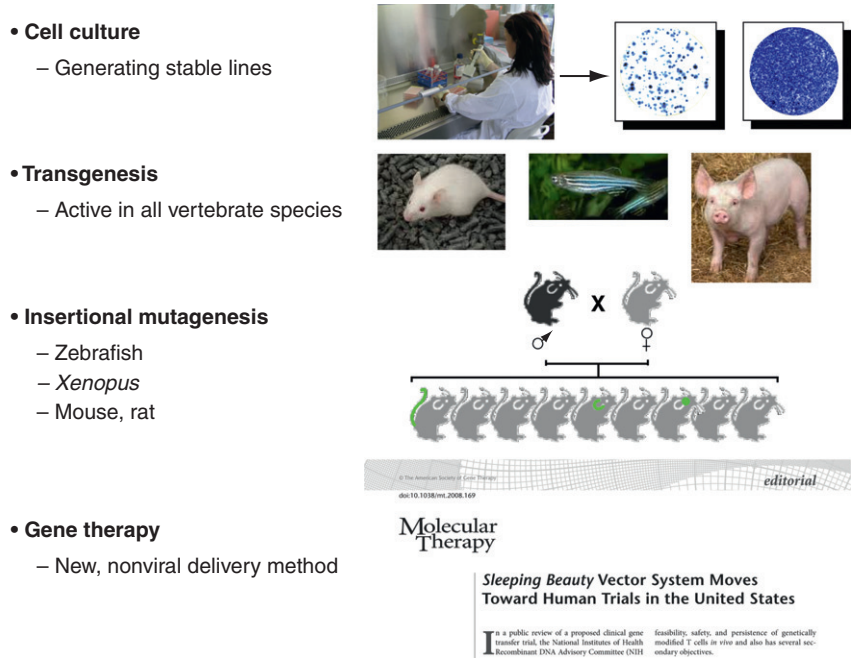


Figure 3.4 Broad applicability of *Sleeping Beauty* in vertebrate genetics.

bottleneck for some applications. Requirements for transfection of primary cells and other hard-to-transfect cell types or for remobilization of transposons from chromosomally resident single-copy donor sites demanded an enzyme with more robust activity. In the past years, significant efforts have been put into enhancing *SB*'s transpositional efficiency and engineering hyperactive versions by mutagenizing and modifying the transposon ITRs and the transposase-coding region (Baus et al., 2005; Cui et al., 2002; Geurts et al., 2003; Mates et al., 2009; Vigdal et al., 2002; Wilson et al., 2005; Zayed et al., 2004). These endeavors yielded a novel hyperactive *SB* transposase (referred to as *SB100X*) (Mates et al., 2009) that is up to 100 times more active than the originally reconstructed *SB* enzyme with its efficiency in transgene delivery reaching those of viral vectors.

4.1.2. Transposons and functional genomics

The postgenomic era presented the scientific community with the new challenge of functional annotation of every gene and identification of elaborate genetic networks. Diverse methods have been employed to address this task, including mutational analysis that proved to be one of the most direct ways to decipher gene functions. There are versatile

strategies for creating mutations, including insertional mutagenesis by discrete pieces of foreign DNA that has the advantage that the inserted DNA fragment can serve as a molecular tag that allows rapid, usually PCR-based, identification of the mutated allele. Since the function of the gene in which the insertion has occurred is often disturbed, such loss-of-function insertional mutagenesis is frequently followed by functional analysis of mutant phenotypes. In many instances, retroviral vectors were utilized to introduce mutagenic cassettes into genomes, but their chromosomal insertion bias does not allow full coverage of genes. The random integration pattern of the *SB* transposon combined with its ability to efficiently integrate versatile transgene cassettes into chromosomes established this system as a highly useful tool for insertional mutagenesis in both ES cells (Kokubu et al., 2009; Liang et al., 2009) as well as in somatic (Collier et al., 2005; Dupuy et al., 2005) and germline tissues (Carlson et al., 2003; Dupuy et al., 2001; Fischer et al., 2001; Geurts et al., 2006b; Horie et al., 2001; Kitada et al., 2007; Lu et al., 2007; Roberg-Perez et al., 2003) in animal models.

Insertional mutagenesis can be applied in cultured, germline-competent stem cells including ES and spermatogonial stem (SS) cells. One advantage of this approach lies in the possibility to perform preselection of modified ES cell clones before generating mutant animals as well as in the possibility to differentiate selected clones into many different tissue types *in vitro*. It is possible to perform large-scale, *SB*-based, insertional mutagenesis screens in ES and SS cells by simply transfecting or electroporating transposon donor and transposase expression plasmids into the cells. The amounts of the delivered plasmids can be adjusted for obtaining the desired insertion frequencies per cell. In addition, *SB* transposons can also be remobilized from chromosomally resident loci and reintegrated somewhere else in the genome by transiently providing the transposase source; such excision–reintegration events can be monitored by using double selection systems, in which excision results in activation of the first and reintegration in activation of the second selection marker (Luo et al., 1998).

Since several aspects of physiology in rats have evolved to be more similar to humans than that of mice, it would be highly desirable to link the rat into the process of annotating the human genome with function. However, the lack of technology for generating defined mutants in the rat genome has hindered the identification of causative relationships between genes and disease phenotypes. As an important step toward this goal, an approach of establishing *SB* transposon-mediated insertional mutagenesis in rat SS cells was recently reported (Izsvak et al., 2010). *SB* transposition can be used to tag and simultaneously mutate thousands of genes in culture, by taking advantage of gene trap cassettes. Importantly, culture conditions maintain the potential of genetically manipulated SS cells to produce viable sperm cells. The spermatogonial clones were transplanted to repopulate the

testes of sterilized, wild-type recipient male rats. The stem cell genome is then passed on to transgenic offspring upon crossing the recipient males with wild-type females. Although transposition events in a given target gene occur by chance, the tissue culture conditions allow screening for a large number of events. Transposition-mediated gene insertion and cell culture conditions thus allow generation of libraries of gene knockouts in rat SS cells. This technology has the potential to develop powerful genomic tools for the rat, offering the opportunity to create a bridge between physiology and genomics.

Another method, in which TEs are utilized for insertional mutagenesis in animal models, employs a "jumpstarter and mutator" scheme (Carlson et al., 2003; Dupuy et al., 2001; Horie et al., 2001). In this arrangement, mutator transgenic lines carry *SB* transposon-based gene-trapping vectors, while a jumpstarter line expresses the transposase preferably in the male germ line (Fischer et al., 2001; Horie et al., 2003). Crossing of the two lines results in transposition in the germline of the F1 double-transgenic males, which are then repeatedly crossed with wild-type females to segregate the transposition events that occurred in their sperm cells to separate F2 animals (Fig. 3.4). In the mouse system, a single sperm cell of an F1 male contains, on average, two transposon insertions (Dupuy et al., 2001), and up to 90% of the F2 progeny can carry transposon insertions (Horie et al., 2001). The applicability of this approach has been demonstrated by the identification of mouse genes with either ubiquitous or tissue-specific expression patterns (Carlson et al., 2003; Geurts et al., 2006a; Horie et al., 2003; Yae et al., 2006). Recently, a similar system for *SB* insertional mutagenesis was also established in rats (Kitada et al., 2007; Lu et al., 2007).

4.1.3. Transposons as vectors for gene therapy

Considerable effort has been devoted to the development of gene delivery strategies for the treatment of inherited and acquired disorders in humans. A desirable gene therapy approach should (i) achieve delivery of therapeutic genes at high efficiency specifically into relevant cells, (ii) be adaptable to changing needs in terms of vector design, (iii) minimize the risk of genotoxicity, and (iv) be cost-effective.

Adapting viruses for gene transfer is a popular approach; for example, γ-retroviral and lentiviral vectors are efficient at integrating foreign DNA into the chromosomes of transduced cells and have enormous potential for lifelong gene expression. A major concern of using retroviral vectors is the potential for mutagenic effects at the sites of genomic integration (Baum et al., 2004; Hacein-Bey-Abina et al., 2003, 2008). Indeed, insertional mutagenesis has been observed in clinical trials using a retroviral vector for gene therapy of X-linked severe combined immunodeficiency (SCID-X1) (Hacein-Bey-Abina et al., 2003, 2008; Thrasher et al., 2006). The clinical use of retroviral vectors can be curtailed due to the limited size of the

payload, as multiple or large transgenes compromise the efficiency of viral reverse transcription and packaging. Finally, regulatory issues and high costs associated with manufacture of clinical-grade retrovirus hamper their widespread translation into clinical practice. An ideal therapeutic vector would combine the favorable attributes of integrating viral vectors (i.e., stable chromosomal insertion) while significantly reducing the potential for adverse events. Transposons could potentially offer such an alternative (Fig. 3.4).

The advantage of *SB* transposon-based gene delivery is that, due to stable genomic insertion of expression cassettes, it can lead to both long-term and efficient transgene expression in preclinical animal models (Ivics and Izsvak, 2006). Thus, the *SB* plasmid-based transposon system combines the advantages of viral vectors with those of naked DNA molecules. However, in contrast to viral vectors, transposon vectors can be maintained and propagated as plasmid DNA that makes them simple and inexpensive to manufacture—an important issue regarding the implementation of future clinical trials. The further advantages of the *SB* system include its reduced immunogenicity, no strict limitation of the size of expression cassettes (Zayed et al., 2004), and improved safety/toxicity profiles (Ivics et al., 2007; Mates et al., 2009; Moldt et al., 2007; Walisko et al., 2008). Since the transposition mechanism does not involve reverse transcription, DNA-based transposon vectors are not prone to incorporate mutations and can tolerate larger and more complex transgenes, including those containing repeated DNA motifs. Moreover, the use of *SB*-based gene delivery eliminates the risk of rearrangements of the expression cassette that, as part of a transposing unit of DNA, integrates into chromosomal DNA in an intact form (Ivics and Izsvak, 2006). In comparison to retroviral systems, the *SB* vectors have an inherently low enhancer/promoter activity (Moldt et al., 2007; Walisko et al., 2008). Inserting insulator sequences flanking the transcription units of the cargo to prevent accidental *trans*-activation of promoters of neighboring genes further increased the safety features of the *SB* system (Walisko et al., 2008). Notably, the transposase can be provided as mRNA, thereby reducing the risk of "re-hopping" of the transposon-based vector (Wilber et al., 2006). Chromosomal integration of *SB* transposons is precise and no *SB*-associated adverse effects have been reported (Fernando and Fletcher, 2006; Ivics and Izsvak, 2006). The past couple of years have seen a steady growth in interest in applying the *SB* system for the treatment of a number of conditions including hemophilia A and B (Liu et al., 2006; Ohlfest et al., 2005b), junctional epidermolysis bullosa (Ortiz-Urda et al., 2002), tyrosinemia I (Montini et al., 2002), glioblastoma (Ohlfest et al., 2005a), Huntington disease (Graepler et al., 2005), and type 1 diabetes (He et al., 2004). In addition, important steps have been made toward *SB*-mediated gene transfer in the lung for potential therapy of alpha-1-antitrypsin deficiency, cystic fibrosis, and a variety of cardiovascular diseases (Belur et al., 2003; Liu et al., 2004). Thus, the establishment of

nonviral, integrating vectors generated considerable interest in developing efficient and safe vectors for human gene therapy (Essner et al., 2005; Hackett et al., 2005; Izsvak and Ivics, 2004).

The first clinical application of the *SB* system (Fig. 3.4) will be tested using autologous T cells genetically modified to redirect specificity for B-lineage malignancies (Williams, 2008). Lymphocytes are a suitable initial platform for testing new gene transfer systems as there have been hundreds of infusions of clinical-grade T cells genetically modified using viral and nonviral approaches without apparent genotoxicity (Bonini et al., 2003). The *SB* transposon to be introduced in the first-in-human application carries a chimeric antigen receptor (CAR) to render the T cells specifically cytotoxic toward $CD19^+$ lymphoid tumors (Huang et al., 2008; Tuteja et al., 2001). The advantage of using the *SB* system for the genetic modification of T cells includes the reduced cost associated with manufacturing clinical-grade DNA plasmids compared with recombinant viral vectors. This is important when one considers that trials infusing CAR^+ T cells are only now beginning to demonstrate antitumor effects (Pule et al., 2008; Till et al., 2004). The higher enzymatic activity of *SB100X* might enable to achieve integration efficiencies comparable to that of retroviral vectors for next-generation trials.

The next phase of preclinical research will focus on further refinement in large animal models to undertake *SB*-mediated transposition *in vivo* and improving the safety profile of *SB* vectors by target-selected transgene integration into genomic "safe harbors." While it remains to be seen whether the first clinical application of the *SB* system will result in an antitumor effect, this trial will help validate the safety of this approach and allow investigators to revisit the design of DNA vectors in general to help improve the therapeutic effect in subsequent next-generation trials.

4.2. Retrotransposons as genetic tools

4.2.1. General introduction

DNA transposons, the mobile elements that move via a "cut-and-paste" mechanism, have been used in various types of biotechnology. However, substantial genome projects have revealed that the genomic proportion of retrotransposons, the mobile elements that move by a "copy-and-paste" mechanism, often exceeds that of DNA transposons (Table 3.1). The recent availability of genome information and the advances in understanding of retrotransposition mechanisms allow us to utilize retrotransposons for biotechnological applications such as genetic markers, insertional mutagenesis, and gene delivery vectors.

Retrotransposons are widespread in metazoan genomes. Most integrate into random sites of the host genome, but some have a sequence preference. A few subclades of LINEs show highly sequence-specific integration into

Table 3.1 Comparison of contents of transposable elements in the genome

	Retrotransposons			
	LINEs (%)	SINEs (%)	LTRs (%)	DNA transposons (%)
Human (*Homo sapiens*)	21.1	13.6	8.6	3.0
Mouse (*Mus musculus*)	19.5	8.0	10.4	0.9
Dog (*Canis familiaris*)	19.2	10.8	3.9	2.0
Chicken (*Gallus gallus*)	6.5	⋆	1.3	0.8
Fruit fly (*Drosophila melanogaster*)	1.3	⋆	3.3	0.5
Silkworm (*Bombyx mori*)	13.8	12.8	1.7	3.0
Mosquito (*Aedes aegypti*)	14.4	1.9	10.5	3.0
Rice (*Oryza sativa*)	1.2	0.1	5.4	9.3

⋆, not stated. The data sources are as follows: human, mouse, and dog, Lindblad-Toh et al. (2005); chicken, International Chicken Genome Sequencing Consortium (2004); fruit fly, Bergman et al. (2006); silkworm, Osanai-Futahashi et al. (2008); mosquito, Nene et al. (2007); rice, Yu et al. (2002).

the genome. In this chapter, we introduce retrotransposons whose transposition mechanisms have been well studied in various organisms and discuss retrotransposition assay systems and their availability for biotechnological tools, focusing especially on the sequence specificity used for site-specific gene delivery.

Several genetic elements such as homing endonuclease genes (HEGs) (Edgell, 2009; Marcaida et al., 2010; Paques and Duchateau, 2007; Redondo et al., 2008; Remy et al., 2010) and engineered zinc-finger nucleases (ZFNs) (Beumer et al., 2008; Geurts et al., 2009; Le Provost et al., 2010; Shukla et al., 2009; Townsend et al., 2009) have been toward the practical use for the site-specific gene delivery or transgenic tools in various organisms. In this chapter, however, we do not deal these subjects because they are not retroelements. Mobile group II introns, retroTEs found in bacterial and organellar genomes, are also hoped to be used for site-specific DNA integration in gene targeting (Cui and Davis, 2007; Lambowitz and Zimmerly, 2004; Mastroianni et al., 2008; Yao et al., 2005; Zhuang et al., 2009). The group II introns recognize their target DNA sites mainly by base paring of the intron RNA, and thus it is possible to design the target site simply by modifying the RNA. Such programmable gene-targeting vectors named "targetrons" from *Lactococcus lactis* L1.LtrB introns have been used in various bacteria. Recently, the group II intron knockout technology is also expanding in application for eukaryote systems. There are several detailed reviews for group II introns toward application (Cui and Davis, 2007; Lambowitz and Zimmerly, 2004) and thus we focus other REs in this chapter.

4.2.2. Retrotransposons in insects, focusing on sequence-specific integration

4.2.2.1. Retrotransposons in Drosophila Although many putative retrotransposons have been identified in the genomes of *Drosophila* and other insects, few of them have been shown to have retrotransposition activity. An LTR retrotransposon in *Drosophila melanogaster*, gypsy, has been shown to integrate to the *ovo* gene in a site-specific manner (Mevel-Ninio et al., 1989). Its retrotransposition activity was shown by a mating experiment (Mevel-Ninio et al., 1989), a microinjection experiment of egg plasm (Kim et al., 1994) or a feeding experiment (Song et al., 1994). Gypsy integrates into the *ovo* locus at a high frequency that is determined by multiple DNA-binding sites within the gene (Labrador and Corces, 2001). At the end of the gypsy unit, there is a short insulator sequence (Gdula et al., 1996) that has been used in various experiments (Gerasimova and Corces, 2001). Importantly, gypsy works in other organisms including *Saccharomyces cerevisiae* (Donze and Kamakaka, 2001) and mice (Yao et al., 2003). Recently, the importance of the relative orientation of two gypsy sequences was shown for enhancer blocking activity (Labrador et al., 2008). Gypsy was established as a useful general genetic tool for several organisms. Another LTR retrotransposon in *D. melanogaster*, ZAM, seems to have a sequence preference for its integration. The recombinant protein of EN domain recognizes CGCGCG within the *white* gene (Faye et al., 2008), although its retrotransposition activity was not directly shown.

A randomly integrated LINE, I factor, was originally found in the I-R hybrid dysgenesis of *D. melanogaster* (Bucheton, 1990). The first direct evidence of LINE transposition intermediating RNA was shown with I factor (Jensen and Heidmann, 1991; Pelisson et al., 1991). Various markers were tagged with I factor to investigate its retrotransposition activity. It shows similarity to the mammalian L1 family, is transcribed from an internal promoter by RNA polymerase II, is capped at the 5′-end, polyadenylated, and prefers A-rich regions for integration. Most studies have been done *in vivo*, crossing R females and I males, with a few *in vitro* studies in cell culture (Jensen et al., 1994). The features of such experiments limited the application of I factor to *Drosophila*; however, the experiments showed the mechanism of retrotransposon silencing by disappearing female sterility through several generations (Chambeyron and Bucheton, 2005).

4.2.2.2. Target-specific LINEs in insects It is of interest that many LINEs found in insects have specific target sequences in the host genome. The existence of R1 and R2, inserted in the specific sites of 28S rDNA, was originally suggested in *D. melanogaster* (Roiha et al., 1981) and the silkworm *Bombyx mori* (Fujiwara et al., 1984), and was later confirmed as a LINE (Burke et al., 1987; Xiong and Eickbush, 1988). In particular, fine studies of

R2 using *in vitro* (Luan and Eickbush, 1995; Luan et al., 1993) and *in vivo* retrotransposition assays (Eickbush et al., 2000) clarified general aspects of the TPRT mechanisms that are peculiar to LINE. More recently, several R elements (R4 (Burke et al., 1995), R5 (Burke et al., 2003), R6 (Kojima and Fujiwara, 2003), R7 (Kojima and Fujiwara, 2003), R8 (Kojima et al., 2006), and RT (Besansky et al., 1992)) that insert into different sites of rDNA than do R1 and R2, telomeric repeat-specific LINEs, SART (Takahashi et al., 1997), and TRAS (Kubo et al., 2001; Okazaki et al., 1995) and other several target-specific LINEs that integrate into microsatellite or other repetitive sequences in the genome, have been found, mainly in insect genomes (Kojima and Fujiwara, 2004).

Of the 16 clades of LINE known to date, early branched groups occupy five clades including the well-studied R2 clade. While many early branched LINEs are target site specific, not all members have this specificity. Early branched LINEs encode only one ORF including a restriction endonuclease-like endonuclease (RLE) close to the C-terminus (Yang et al., 1999) (Fig. 3.5A). This RLE is involved in a sequence-specific digestion of the target, but its functional role is not certain. However, the N-terminal domain of R2 from *B. mori* genome has been shown to bind target DNA substrate *in vitro* (Christensen et al., 2005).

The rest of the LINE clades are categorized as recently branched LINEs, which normally encode two ORFs and an AP endonuclease-like endonuclease (APE) at the N-terminus of ORF2 (Fig. 3.5A). Most are not target specific, but the R1 and Tx clades include target-specific elements (Kojima and Fujiwara, 2004, 2005a). The major domain involved in the target specificity of recently branched LINEs is thought to be APE, because it cuts the target DNA in a sequence-specific manner *in vitro* (shown for R1 (Feng et al., 1998) and TRAS1 (Anzai et al., 2001) elements). The comparison of the crystal structure of the isolated APE domains from target-specific LINEs TRAS1 and R1 and the nonspecific LINE human L1 clarify the putative amino acids involved in target-specific recognition and digestion (Maita et al., 2004, 2007). Regions other than the APE domain also may play a role in target-specific integration. The ORF1 (Gag) proteins of the telomere-specific LINEs TART, HeT-A (Rashkova et al., 2002), and SART1 (Matsumoto et al., 2004, 2006) show a punctate localization in nuclei, which may be involved in recruiting the LINE unit to the telomere region. In addition, read-through transcripts of R1 are necessary for its precise integration (Anzai et al., 2005).

4.2.2.3. Target-specific retrotransposition systems and their application In most transposition systems, plasmid DNA is injected into embryos or transfected into cells. For target-specific LINEs from insects, however, a novel retrotransposition system using baculovirus AcNPV has been developed (Takahashi and Fujiwara, 2002) (Fig. 3.5B).

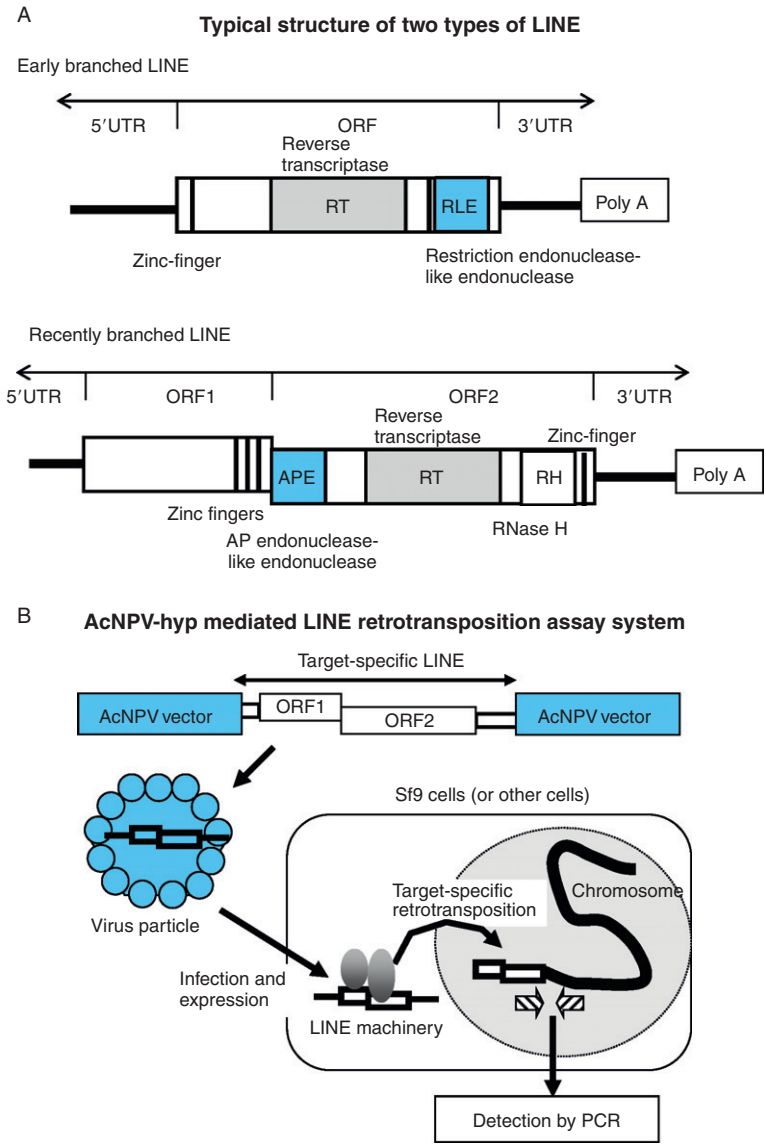


Figure 3.5 Structure- and virus-mediated retrotransposition assay of target-specific LINE. (A) Schematic structures of two types of LINE are shown. Early branched LINE has only one open reading frame (ORF) and recently branched LINE two ORFs. There are some target-specific LINEs in both types. Vertical lines in ORF represent zinc-finger motifs. (B) Target-specific LINE recombinated in the baculovirus (AcNPV) vector can infect Sf9 (insect) cells or other cells and be expressed effectively. Target-specific LINE integrates into the specific sequence of the host chromosome and its integration can be easily detected by PCR.

The telomere-specific LINEs TRAS and SART, when recombined in the AcNPV vector, can integrate into their respective specific targets when they are infected in Sf9 insect cells, and their target can be altered by swapping APEs between them (Takahashi and Fujiwara, 2002). This system is also available for other target-specific LINEs, 28S rDNA-specific LINEs, R1, R2 (Fujimoto et al., 2004), and R7 and R8 (unpublished data). Many target-specific LINEs recognize the 3′-UTR of LINE–RNA during the initial step of the TPRT process. A sequence such as a fluorescent protein reporter with an added 3′-UTR sequence can be retrotransposed in a target-specific manner with a helper LINE by transcomplementation, a useful feature for a gene delivery tool (Osanai et al., 2004).

This baculovirus-mediated retrotransposition system, which enables very high expression of the LINE machinery, also has the advantage that it is easily converted to an *in vitro* (Matsumoto et al., 2006) retrotransposition assay. The purified retrotransposition machinery for SART, after being expressed with AcNPV vector in Sf9 cells, can integrate specifically *in vitro* at (TTAGG) telomeric repeats. The *in vitro* assay contributes to understanding of not only the mechanism of target-specific retrotransposition but also the structural features of LINEs. The baculovirus AcNPV, which is derived from a moth, has a wide host range and can infect a broad range of species including human cells. Thus, the AcNPV-mediated system is thought to be available *in vivo* in various species. It has been shown that AcNPV recombined with R1 and SART1 can integrate at specific target sites in the genome of various organs when injected into larva of *B. mori* (Kawashima et al., 2007), and that this gene transfer system is also useful in the honeybee (Ando et al., 2007). This AcNPV-mediated LINE system is also applicable for human and fish cell lines as a target-specific gene delivery tool (Kawashima, unpublished data).

4.2.3. Retrotransposons in fish

4.2.3.1. Randomly integrated LINEs and SINEs in fish Several LINEs have been found in fish genomes and some of them have been shown to have retrotransposition activity. UnaL2, found in the eel genome, was shown to move in human HeLa cells by being tagged with an inverted *neo* gene divided by an intron (Kajikawa and Okada, 2002). UnaL2 integrates randomly into the genome. It also recognizes 3′-UTR sequences of specific "passenger" SINEs like UnaSINE1 and UnaSINE2 and mobilizes them by transcomplementation (Kajikawa and Okada, 2002; Kajikawa et al., 2005), which is a useful feature for a gene delivery tool. ZfL2-1 and ZfL2-2, found in the zebrafish genome, also can retrotranspose in HeLa cells (Sugano et al., 2006; Tamura et al., 2007). ZfL2-1 has two ORFs and ZfL2-2 has one ORF but both belong to the same clade L2. These LINEs have the ability to mobilize in zebrafish *in vivo*.

4.2.3.2. Target-specific LINEs in fish, good candidates for gene therapy tools We have recently identified a 28S rDNA-specific LINE R2Ol from the genome of the medaka fish, *Oryzias latipes* (Kojima and Fujiwara, 2005b), which was inserted at the same site as R2 elements in insects. We have established a sequence-specific retrotransposition assay system for R2Ol in zebrafish embryos by microinjecting mRNA transcribed *in vitro*. R2Ol shows effective retrotransposition activity of around 50%, and we obtained transgenic F1 fish retaining integrated R2Ol (Yatabe, manuscript in preparation). This *in vivo* retrotransposition assay in zebrafish has advantages for investigating LINE mechanisms through development or through several generations. We also succeeded in retrotransposing R2Ol into a precise site of 28S rDNA of several human cell lines (Kawashima, manuscript in preparation), suggesting that R2Ol is a good candidate for a gene therapy tool that avoids undesirable integration into the human genome.

4.2.4. Mammalian retrotransposons and their application as genetic tools

4.2.4.1. Mammalian retrotransposons In many mammals, retrotransposons account for a notable portion of the genome. Genome sequencing has revealed that there are many inter- and intraspecies polymorphisms in retrotransposon insertion. The *de novo* insertion of a retrotransposon is stable compared to that of a DNA transposon, which enables us to investigate molecular phylogeny by analysis of retrotransposon insertions. Advances in understanding of the retrotransposition mechanisms of L1 and intracisternal A particle (IAP) elements allow the exploration of the possibility of utilizing these elements in insertional mutagenesis and gene delivery.

4.2.4.2. Genetic markers Retrotransposons have been used as genetic markers for nearly two decades. SINE insertions have been used to analyze phylogenetic relationships in fish (Murata et al., 1993; Takahashi et al., 2001), the evolution of whales (Nikaido et al., 1999; Shimamura et al., 1997), and the evolutionary history of humans and primates (Batzer et al., 1994; Li et al., 2009; Perna et al., 1992; Stoneking et al., 1997). The progress of genome sequencing in mammals is accelerating the use of retrotransposons as intraspecies genetic markers. In humans, some retrotransposons have been inserted so recently that they are polymorphic for presence or absence among populations and individuals. These polymorphisms are also being used as forensic tools (Ray et al., 2007). In dogs, it is reported that more than 10,000 loci are bimorphic for SINE insertions, which may be used as genotyping markers or for identifying the ancestral relationships between dog breeds (Wang and Kirkness, 2005). Active retrotransposons are also reported in the bovine genome (Adelson et al., 2009), so retrotransposons are promising as genetic markers in breed identification of cattle. Interspecies identification can also be conducted by detecting retrotransposon sequences:

for instance, it can be identified whether the meat in sausages is derived from beef, pork, chicken, or ruminant materials (Ray et al., 2007).

4.2.4.3. Mutagenesis Attempts at using retrotransposons in random mutagenesis have been made in mammals. Insertional mutagenesis using retrotransposons has several advantages over methods based on chemicals such as ENU (*N*-ethyl-*N*-nitrosourea). One advantage is that while ENU creates point mutations that often do not affect gene function, retrotransposons can be used to deliver a gene trap that efficiently splices into genes and disrupts their function. Another is that the mutated genes can be cloned more easily when retrotransposons are used compared with when ENU is used.

Retrotransposon-based mutagenesis is also likely to have some advantages over DNA transposon-based mutagenesis. DNA transposons have a tendency to insert novel copies close to their original genomic localization, so-called local hopping effect (Fischer et al., 2001; Guimond et al., 2003; Horie et al., 2003; Keng et al., 2005; Scali et al., 2007; Wang et al., 2008), but retrotransposons such as L1 do not have this tendency and insert randomly into the genome (An et al., 2006; Babushok et al., 2006). In addition, retrotransposon insertion appears to be more stable than DNA transposon insertion. Retrotransposon insertions are fixed in the genome, whereas the inserted DNA transposon has a possibility of being excised if a DNA transposase gene exists in the host genome.

Intense studies of the application of human and mouse L1 retrotransposons as insertional mutagenesis tools (Ostertag et al., 2007; Rangwala and Kazazian, 2009) have recently been made. L1 is a LINE-type retrotransposon that inserts randomly into the genome without apparent locus specificity. L1 retrotransposition has been detected in cultured cells of human (Moran et al., 1996), mouse (Moran et al., 1996), rat (Kirilyuk et al., 2008; Muotri et al., 2005), hamster (Morrish et al., 2007), and chicken (Suzuki et al., 2009; Wallace et al., 2008) and also in transgenic mice (Kano et al., 2009). In a cultured cell assay, L1-expressing plasmids were transfected into the cultured cells. The rate of L1 retrotransposition varied from 0.04% to 10% of cells containing the L1 plasmid, depending on the kind of L1 element used (Han and Boeke, 2004; Moran et al., 1996). The codon-optimized mouse L1 has the highest retrotransposition rate so far detected (Han and Boeke, 2004).

Transgenic mice have been generated for studying L1 retrotransposition *in vivo* (An et al., 2006, 2008; Babushok et al., 2006; Muotri et al., 2005; Ostertag et al., 2002; Prak et al., 2003). These transgenic mice contain the L1 retrotransposon that has a gene cassette for detecting *de novo* retrotransposition inside the L1 3′-UTR (L1/transgene). In many transgenic lines, the gene cassette is designed so that the EGFP gene will be expressed when L1 retrotransposition occurs. At the cell culture level, a gene cassette for gene-trapping purposes that consists of a bidirectional poly(A) signal

sandwiched between two oppositely oriented splice acceptors was effective (An et al., 2006).

Transcription of the L1/transgene in transgenic animals is driven by the endogenous promoter in the L1 5′-UTR (Muotri et al., 2005; Ostertag et al., 2002) or a heterologous promoter such as mouse RNA polymerase II large subunit promoter (Ostertag et al., 2002; Prak et al., 2003), the Hsp70-2 promoter (Babushok et al., 2006), or the CAG promoter (An et al., 2006, 2008). The use of a heterologous promoter increases the frequency of retrotransposition events (Ostertag et al., 2002). The highest germline retrotransposition frequency so far reported is that of the codon-optimized mouse L1, ORFeus. In a transgenic line that has 10 copies of the donor ORFeus/transgene (An et al., 2006), the germline insertion frequency was 33%, while a recently generated transgenic line that has a single copy of ORFeus/transgene had a germline insertion frequency of 9.4% (An et al., 2008).

In transgenic mice carrying the L1/transgene, *de novo* insertions have been detected in all chromosomes (An et al., 2006). This contrasts strikingly with DNA transposons that have a tendency to insert close to their original genomic localization, a process called "local hopping" (Fischer et al., 2001; Horie et al., 2003; Keng et al., 2005; Wang et al., 2008).

Tissue-specific activation systems for the L1 transgene using Cre/loxP have been developed, and specific activation in germline and pancreas has been reported (An et al., 2008). The tissue-specific activation system in combination with a gene trap cassette containing strong splice acceptors is useful for somatic mutagenesis. One application of L1-based somatic mutagenesis in mice may be a screening system for tumor suppressors and oncogenes (Ostertag et al., 2007).

In addition to the L1 retrotransposon, there is also the possibility of using the IAP retrotransposon for insertional mutagenesis. IAP is a mouse LTR retrotransposon, and a cell culture assay has been established for examining IAP retrotransposition. IAP retrotransposition is detected in ~0.2–1.0% of the cells transfected with an IAP construct (Dewannieux et al., 2004; Horie et al., 2007; Saito et al., 2008).

4.2.4.4. Gene delivery vectors There are also studies using L1 for a gene delivery vector in combination with helper-dependent adenoviruses (Kubo et al., 2006; Soifer and Kasahara, 2004; Soifer et al., 2001). Adenovirus lacks the machinery for efficient integration into host chromosomes, and rarely integrates into the genome, but efficiently infects many cell types. The helper-dependent adenovirus is engineered to lack all the coding sequences that could be toxic or immunogenic to the host (Kochanek et al., 1996; Mitani et al., 1995; Schiedner et al., 1998).

There are several advantages in using L1–adenovirus rather than retrovirus in gene delivery. Retrovirus integrates preferentially into active genes, but L1 integrations are random, so there is a smaller possibility of disrupting

other genes. Retroviral integration always includes the LTR sequences, which have strong promoter activity and have a risk of activating genes flanking the insertion site, but L1 usually loses its endogenous promoter during integration.

In the L1–adenovirus hybrid vector, the adenovirus delivers the L1/transgene element into the cells, and the L1 integrates the transgene into the genome. Helper-dependent adenovirus does not propagate in the infected cells. By the use of helper-dependent adenovirus, L1/transgene retrotransposition frequencies of up to 91% of infected cells have been observed (Kubo et al., 2006). This is markedly greater than the retrotransposition efficiency achieved by direct plasmid transfection. The maximum retrotransposition efficiency achieved by direct plasmid transfection is about one cell per 10 transfected cells using the highly active ORFeus construct (Han and Boeke, 2004).

In the traditional plasmid transfection method, most of the L1 retrotransposition reported was in transformed or immortalized cells. In the L1/adenovirus system, the high transduction ability of the adenovirus allows L1/transgene retrotransposition in differentiated human primary somatic cells, including dermal fibroblasts and hepatocytes, and furthermore, in nondividing cells arrested in the G1/S cells, suggesting the potential for utilization in gene therapy.

Besides GFP expression cassettes, short hairpin RNA (shRNA) expression cassettes have been delivered by L1 to the genomes of cultured cells (Yang et al., 2005). Although in this experiment L1/shRNA was transduced to cells by the traditional plasmid transfection method, it was demonstrated that both exogenous (GFP) and endogenous (GADPH) gene expression could be reduced by L1-transmitted shRNA. One copy of retrotransposed L1 with a GFP shRNA cassette was sufficient to reduce GFP fluorescence by 90%. This report suggests that the L1-based RNAi system is promising as a stable gene-silencing system in human cells.

5. Domestication of Mobile DNA by the Host Genomes

5.1. Genomic repeats as transcriptional promoters

Recent whole-genome analysis revealed that about 25% of all human promoters contain REs in their sequence (van de Lagemaat et al., 2003). Moreover, 7–10% of experimentally characterized transcription factor-binding sites (TFBS) were shown to be derived from repetitive sequences including simple sequence repeats and TEs (Polavarapu et al., 2008). TFBS that originated from repeats evolve more rapidly than nonrepetitive TFBS but still show signs of sequence conservation on functionally critical bases.

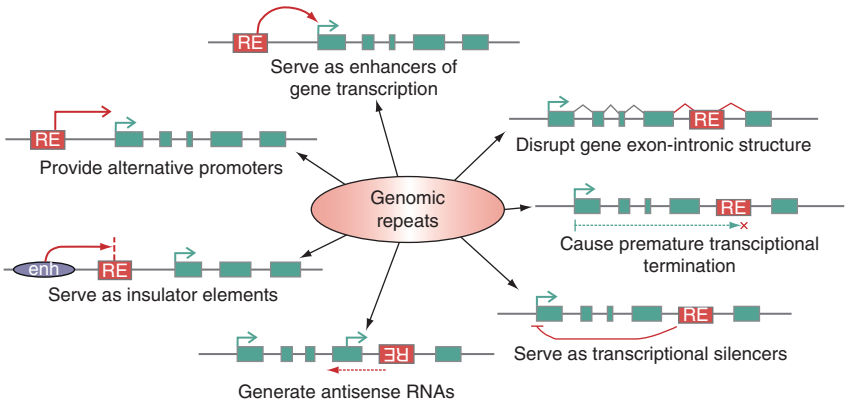


Figure 3.6 Genomic repeats influence on transcription of the host genes. Red boxes, retroelements; green boxes, gene exons; green arrow, gene transcriptional start site; purple oval, enhancer element.

Such rapidly evolving TFBS are likely to direct species-specific regulation of gene expression, thus participating in evolutionary process (Fig. 3.6).

In the majority of examples reported to date, REs act as alternative promoters. REs can either influence the level of a corresponding RNA transcription or change the tissue specificity of its expression. For example, LTR integration into *CYP19* gene, encoding for aromatase P450, the key enzyme in estrogen biosynthesis, led to the formation of alternative promoter located 100 kb upstream of the coding region (van de Lagemaat et al., 2003). This event resulted in the primate-specific transcription of *CYP19* in the syncytiotrophoblast layer of the placenta. Placental-specific expression might play an important role in controlling estrogen levels during pregnancy. Placental-specific transcription driven from endogenous retroviral promoters was also shown for *Mid1* gene linked with inheritable Opitz syndrome (Landry et al., 2002), endothelin B receptor (Medstrand et al., 2001), and insulin-like growth factor INSL4 (Bieche et al., 2003).

Solitary ERV-L LTR was shown to promote *β3GAL-T5* transcription in various tissues, being especially active in colon, where it is responsible for the majority of gene transcripts (Dunn et al., 2005). *β3GAL-T5* is involved in the synthesis of type 1 carbohydrate chains in gastrointestinal and pancreatic tissues. Interestingly, murine *β3GAL-T5* gene is also expressed primarily in colon, despite the absence of an orthologous LTR in the mouse genome. It is likely that in humans the LTR adopted the function of an ancestral mammalian promoter active in colon (Dunn et al., 2005). An interesting example of gene transcriptional regulation by LTR was shown for *NAIP* (*BIRC1*) gene coding for neuronal apoptosis inhibitory protein (Romanish et al., 2007). Although human and rodent *NAIP*

promoter regions share no similarity, in both cases LTR serve as an alternative promoter. Thus, two different LTR retrotransposons were recruited independently in primate and rodent genomes for the gene transcriptional regulation.

REs may also represent the only known promoter for some human genes. For example, the only apparent promoter of the liver-specific *BAAT* gene recently implicated in familial hypercholanemia is an ancient LTR in human but not in mouse (Carlton et al., 2003). Antisense L1 and Alu sequences were shown to act as the unique promoter for *HYAL-4* gene, necessary for hyaluronan catabolism (van de Lagemaat et al., 2003).

The application of novel high-throughput techniques such as cap analysis of gene expression (CAGE) and paired-end ditag (PET) sequencing revealed 51,197 ERV-derived promoter sequences. In 1743 cases, ERVs were located in gene proximal or 5′-UTRs. In all, 114 ERV-derived transcription start sites can be demonstrated to drive transcription of 97 human genes, producing chimeric transcripts initiated within LTR and read-through into known gene sequences (Conley et al., 2008b).

Recently, we found that at least 50% of human-specific HERV-K LTRs possess promoter activity and the level of their expression ranges from ~0.001 to ~3% of the beta-actin gene transcriptional level (Buzdin et al., 2006a,b). We have also shown that 5′-proviral LTR is more transcriptionally active than 3′-proviral or solitary LTRs and that the relative content of promoter-active LTRs in gene-rich regions is significantly higher than that in gene-poor loci.

5.2. REs as enhancers for host cell gene transcription

One of the first striking reports of the involvement of REs in tissue-specific gene transcriptional regulation was for the human amylase locus (Meisler and Ting, 1993). In humans, amylase is produced in pancreas and in salivary glands. Human amylase locus includes two genes of pancreatic amylase (*AMY2A* and *AMY2B*) and three genes of salivary amylase (*AMY1A*, *AMY1B*, and *AMY1C*). The latter three genes are likely products of a recent triplication, because in the chimpanzee genome there is only one gene for *AMY1*. Exon-intronic structures of these genes are identical, except for an additional untranslated exon at the 5′-terminus of the salivary amylase genes. Moreover, all genes for salivary amylase contain a full-length insert of HERV-E upstream their transcription start site. It was hypothesized that the insertion of full-length ERV activated a cryptic promoter that drives the transcription of amylase in salivary glands. When there is a solitary LTR instead of full-length HERV-E provirus, cryptic promoter cannot be activated and the gene is expressed only in pancreas.

There are several other well-supported examples of LTR involvement in gene regulation. 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 might be responsible for controlling expression of this cluster in erythroid cells (Long et al., 1998). It was suggested that the enhancer effect might be caused by LTR-initiated transcription driven in the direction of associated gene promoter (Ling et al., 2002, 2003). Another example is the mouse *Slp* (sex-limited protein) gene. ERV-located upstream of the *Slp* in antisense orientation was shown to direct androgen-specific expression of this gene in males (Loreni et al., 1988).

LINEs and SINEs can also serve as transcriptional enhancers. The enhancer of human apoliprotein A was shown to reside within LINE element (Yang et al., 1998). Alu sequence is a part of enhancer element located in the last intron of the human CD8 alpha gene (Hambor et al., 1993). Expression of this gene is restricted to cells of lymphoid lineage and is developmentally regulated during thymopoesis. A CORE-SINE RE (ancient tRNA-derived SINE with a conserved core sequence) was found to represent a neuronal enhancer for the *POMC* (proopiomelanocortin) gene (Santangelo et al., 2007). *POMC* encodes a prohormone that gives rise to several bioactive peptides that participate in the stress response, skin and hair pigmentation, analgesia, and the regulation of food intake and energy balance. CORE-SINE was shown to be responsible for the expression of *POMC* in ventral hypothalamic neurons. Recently, AmnSINEs (a new SINE family identified in the genomes of Amniota) have been shown to act as distal transcriptional enhancers for *FGF8* (fibroblast growth factor 8) and *SATB2* genes in developing mouse forebrain (Sasaki et al., 2008).

5.3. REs as providers of new splice sites for the host genes

Apart from the modulation of transcription, REs can also regulate splicing of pre-mRNA. An outstanding role here belongs to SINE elements, namely Alu retrotransposons in the case of human transcriptome. In a genome-wide comparison of the genomes of human and mouse, a total of 3,932,058 and 3,122,416 TEs have been identified in human and mouse, respectively. Interestingly, 60% of transposons in both human and mouse are located in intronic sequences, even though introns comprise only 24% of the human genome (Sela et al., 2007). All families of transposons in both human and mouse can "exonize," that is, be included in the exons of mature mRNA. Transposons 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 RE 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 (1824 exons compared with 506 exons, respectively) (Sela et al., 2007). Alus are the most abundant repetitive elements in the human genome. The major burst of *Alu* retroposition took place 50–60 mya and has since dropped to a frequency of one new retroposition for every 20–125

new births (Batzer et al., 1993; Cordaux et al., 2006). Alus are presented by more than 1.1 million of copies (Chen et al., 2009), and over 0.5 million of them reside in introns of human protein-coding genes (Levy et al., 2008).

Alu elements have several sequence motifs resembling consensus splice sites in both sense and antisense orientations (Gotea and Makalowski, 2006), and the insertion of Alu elements into intronic regions may introduce new exons into existing functional genes. It has became a current opinion now that it is the exonization of Alu elements that plays a crucial role in birth of new exons in primate genomes (Corvelo and Eyras, 2008; Lin et al., 2008). Most Alu-derived exons are short (median length of the 330 exons is 121 nucleotides) (Lin et al., 2008). Almost all Alu-derived exons are alternatively spliced, and the vast majority of these exons have low transcript inclusion levels (are included only in the minor transcript isoforms). However, younger Alu-derived exons have weaker splice sites and lower absolute values for the relative abundance of putative splicing regulators between exonic and adjacent intronic regions. This relative abundance was shown to increase with exon age, leading to higher exon inclusion (Corvelo and Eyras, 2008). Furthermore, using exon array data of 330 Alu-derived exons in 11 human tissues and detailed RT-PCR analyses of 38 exons, it has been demonstrated that some Alu-derived exons are constitutively spliced in a broad range of human tissues, and some display strong tissue-specific switch in their transcript inclusion levels. Most of these latter exons were derived from ancient Alu elements in the genome (Lin et al., 2008). This is probably due to the fact that exons derived from older Alu elements had more evolutionary time to accumulate nucleotide substitutions that supported exon inclusion in the transcript products (Lin et al., 2008).

Alu consists of two monomeres derived each from a truncated copy of 7SL RNA involved in protein sorting. In rodent genomes, there are also multiple copies of 7SL RNA-derived short retrotransposons, but of a monomeric structure. Why there are so many alternatively spliced Alus compared to rodent retrotransposons? The most probable explanation is its dimeric organization (Gal-Mark et al., 2008). Alus are composed of two related but distinct monomers, left and right arms. Most exonizations occur in right arms of antisense Alu elements. Without the left arm, exonization of the right arm shifts from alternative to constitutive splicing. This eliminates the evolutionary conserved isoform of the host gene and thus may be selected against. The insertion of the left arm downstream of a constitutively spliced non-Alu exon shifts splicing from constitutive to alternative. Although the two arms are highly similar, the left arm is characterized by weaker splicing signals and lower exonic splicing regulatory densities. Mutations that improve these potential splice signals activate exonization and shift splicing from the right to the left arm. Interplay between two or more putative splice signals renders the intronic left arm with a pseudoexon function. Thus, the dimeric form of the Alu element fortuitously provides it

with an evolutionary advantage, allowing enrichment of the primate transcriptome without compromising its original repertoire (Gal-Mark et al., 2008).

Overall, Alu-derived exons had significantly weaker splicing signals compared to nonrepetitive constitutively spliced exons and typical cassette exons (other alternatively spliced exons). This is most probably due to a lower density of exonic splicing regulatory elements in Alu-derived exons. Alu-derived exons had much higher evolutionary rates during primate evolution, compared to constitutive exons and cassette exons. For exons present in both human and chimpanzee genomes, the overall nucleotide substitution rate of Alu-derived exons was 1.34% compared to 0.73% for cassette exons and 0.52% for constitutive exons. Similarly, between human and orangutan genomes, the overall nucleotide substitution rates of Alu-derived exons, cassette exons, and constitutive exons were 3.69%, 1.81%, and 1.31%, respectively. However, at least six Alu-containing exons (in genes *FAM55C*, *NLRP1*, *ZNF611*, *ADAL*, *RPP38*, and *RSPH10B*) are constitutively spliced in human tissues (Lin et al., 2008; Makalowski et al., 1994; Sorek et al., 2002). In addition, Alu sequence provided a donor splice site to one of the constitutive exons of the human gene encoding survivin (a member of the apoptosis inhibitor family that is overexpressed in many malignancies) (Mola et al., 2007).

There is also an excess of Alu-derived internal exons in the 5′-UTRs of the genes compared to the 3′-UTRs. This phenomenon likely reflects stronger purifying selection pressure against exon creation in 3′-UTR because such exons may trigger mRNA nonsense-mediated decay (Lin et al., 2008). In addition, there is an "exclusion zone" in intron sequences flanking exons, where insertion of *Alu* elements is presumably under purifying selection. The length of this "exclusion zone" is similar to that of the human–mouse conserved sequences flanking alternatively spliced exons (~80–150 nucleotides) (Lev-Maor et al., 2008). 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 (Sakate et al., 2007). In both 5′- and 3′-UTR sequences, presence of an Alu element may be important for posttranscriptional regulation of gene expression, for example, by affecting protein translation, alternative splicing, and mRNA stability (Hasler et al., 2007).

Interestingly, Alu exonization might have played a 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 (Lev-Maor et al., 2008). The second example 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 Alu RE into 92 bp-long *CMP* exon, disrupted normal ORF for this enzyme and resulted in the lack of *N*-glycolyl neuraminic acid (Neu5Gc) on a surface of cell membranes (Chou et al., 1998; Irie et al., 1998). 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, for example, brain organogenesis. A subset of other Alu-derived exons, especially those derived from more ancient Alu elements in the genome, might have contributed to functional novelties during evolution of many other primates as well. Some novel polymorphic Alu inserts interfere with the normal pre-mRNA splicing by providing additional splicing enhancers, thus causing inheritable diseases (Gu et al., 2007).

Noteworthy, canine SINE element SINEC_Cf is likely to play a major role in the evolution of dog genomes nowadays (Wang and Kirkness, 2005). Canine genomes harbor a high frequency of alleles that seem to differ only by the absence or presence of a SINEC_Cf repeat. Comparison of an individual dog (a poodle) DNA with a draft genome sequence of a distinct dog (a boxer) has revealed the chromosomal coordinates for $>10{,}000$ loci that are bimorphic for SINEC_Cf insertions. Further analysis of SINE insertion sites from the genomes of nine additional dogs indicates an additional 10,000 bimorphic loci could be readily identified in the general dog population. Approximately, half of all annotated canine genes contain SINEC_Cf repeats. When transcribed in the antisense orientation, they provide splice acceptor sites that can result in incorporation of novel exons. The high frequency of bimorphic SINE insertions in the dog population is predicted to provide numerous examples of allele-specific transcription patterns that may be valuable for the study of differential gene expression among the dog breeds (Wang and Kirkness, 2005).

LINE elements may be involved in constitutive or alternative splicing of cellular RNAs too, although with relatively lower frequencies. For example, mammalian L1 elements contain numerous functional internal splice sites that generate a variety of processed L1 transcripts (most probably useless for the L1 retrotransposition) and also contribute to the generation of hybrid transcripts between L1 elements and host genes. Interestingly, L1 splicing is delayed during the course of L1 expression (Belancio et al., 2008b). This delay in L1 splicing may also serve to protect host genes from the excessive burden of L1 interference with their normal expression via aberrant splicing (Belancio et al., 2008b). However, an increased ratio of constitutively spliced L1s relatively to alternatively spliced ones has been reported

compared to Alu elements. Proportion of L1 elements in gene introns is significantly lower than the one of Alu repeats, although both retrotransposons utilize the same retrotranspositional mechanism (Buzdin, 2004). This bias is probably due to a purifying selection acting against accumulation of L1s in genes. In other vertebrate genomes, LINEs also have been reported to generate new chimeric spliced mRNA variants for the host functional genes, for example, in zebrafish (Tamura et al., 2007) or in pig cells (Sironen et al., 2007).

LTR retrotransposons also may contribute to a diversity of alternatively spliced RNAs (van de Lagemaat et al., 2006). For example, in the case of human gene *VEGFR-3/FLT4* for endothelial angeogenesis controlling receptor, two different isoforms of this protein are encoded by the same gene. Polypeptide encoded by the shorter transcript lacks 65 C-terminal aminoacids. The short *VEGFR-3* transcript is formed because of the use of a noncanonical acceptor splice site within the endogenous retroviral sequence located between the exons 1 and 2. These different forms of *VEGFR-3* gene product probably have different biological functions (Hughes, 2001).

Apart from animal DNA, retrotransposons comprise a significant fraction of plant genomes and are likely involved in gene regulation there too, although the effects of retrotransposon insertions in plants are not well understood. For example, one-sixth of rice genes is associated with retrotransposons, with insertions either in the gene itself or within its putative promoter region. Among genes with inserts in the promoter region, the likelihood of the gene being expressed was shown to be directly proportional to the distance of the retrotransposon from the translation start site. In addition, retrotransposon insertions in the transcribed region of the gene were found to be positively correlated with the presence of alternative splicing forms. Some of the retrotransposons that are embedded in cDNA contribute splice sites and give rise to novel exons (Krom et al., 2008).

Yet, the effect of intronic repeats on splicing of the flanking exons is largely unknown. Importantly, more Alus flank alternatively spliced exons than constitutively spliced ones. This implies that *Alu* insertions may change the mode of splicing of the flanking exons; this is especially notable for those exons that have changed their mode of splicing from constitutive to alternative during primate genome evolution (Lev-Maor et al., 2008). Lev-Maor and colleagues demonstrated experimentally that two *Alu* elements that were inserted into an intron in opposite orientation undergo base pairing, as evident by RNA editing, and affect the splicing patterns of a downstream exon, shifting it from constitutive to alternative. It may also be possible that formation of a long and stable double-stranded structure in the upstream intron, especially near the splice site, reduces the ability of the splicing machinery to properly recognize the downstream exon, leading to slower splicing kinetics or suboptimal exon selection and, thus, to intron retention or exon skipping (Lev-Maor et al., 2008).

Smalheiser and Torvik (2005) showed that a few mammalian miRNA precursors are derived from intronic insertion of two adjacent LINE retrotransposons in opposite orientation, creating a hairpin structure that serves as miRNA precursor. Some other elements have an intrinsic hairpin structure and/or serve as miRNA precursors when inserted into transcriptionally active genomic regions (Hernandez-Pinzon et al., 2009; Piriyapongsa and Jordan, 2007). Many of the newly identified piRNAs are derived from retrotransposons and play a role in transposon silencing in zebrafish germ cells (Houwing et al., 2007; Levy et al., 2008). In order to catalog the data on TEs that may have an impact in gene regulation and functioning, a comprehensive database termed "TranspoGene database" has been constructed that covers genomes of seven species: human, mouse, chicken, zebrafish, fruit fly, nematode, and sea squirt (Levy et al., 2008). The database includes information about repeat localization relative to gene: proximal promoter TEs, exonized TEs (insertion within an intron that led to exon creation), exonic TEs (insertion into an existing exon), or intronic TEs. A variant of this database termed "microTranspoGene" collects the data on human, mouse, zebrafish, and nematode TE-derived miRNAs (Levy et al., 2008).

Overall, the proportion of proteins with retrotransposon-encoded fragments (~0.1%), although probably underestimated, is much less than what the data at transcript level suggest (~4%). In all cases, the RE cassettes are most frequently derived from older REs, in line with the hypothesis that incorporation of TE fragments into functional proteins requires long evolutionary periods. The role of evolutionary recent REs is probably limited to regulatory functions (Gotea and Makalowski, 2006).

5.4. REs as sources of novel 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 polyadenylated with the remarkable exception of two polyadenylated polymerase III-transcribed RNAs (Borodulina and Kramerov, 2008). 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. This is probably the right explanation for the clearly seen strong negative selective pressure on such elements oriented in the same transcriptional direction as the enclosing gene (Buzdin, 2007; Cutter et al., 2005; van de Lagemaat

et al., 2006; Wheelan et al., 2005; Zemojtel et al., 2007). Indeed, all protein-coding intronic REs (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, nonautonomous REs like Alu do not 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 (Roy-Engel et al., 2005).

Even in the antisense direction relatively to enclosing genes, many retrotransposons provide poly(A) signals that may dynamically modify 3′-ends of genes through evolution. For example, in breast cancer cell line T47D, there were identified four mRNAs polyadenylated at the sequence of HERV-K retroviral LTR (Baust et al., 2000). Transcripts of gene *NSBP1* can be alternatively polyadenylated at the retroviral sequences located in the 3′-UTR of that gene (King and Francomano, 2001). 5′-LTR of the retrovirus HERV-F may function as the alternative polyadenylation site for gene *ZNF195* (Kjellman et al., 1999). Human genes *HHLA2* and *HHLA3* utilize HERV-H LTRs as the major polyadenylation signals. In the baboon genome, orthologous loci lack retroviral inserts and these genes recruit other polyadenylation sites (Mager et al., 1999).

Interestingly, REs are mostly associated with nonconserved poly(A) sites (Lee et al., 2008a). Of the 1.1 million of human Alu retrotransposons, about 10,000 are inserted in the 3′-UTRs of protein-coding genes and 1% of these (107 events) are active as poly(A) sites (Chen et al., 2009). Alu inserts usually represent weak or cryptic poly(A) signals, but often constitute the major or the unique poly(A) site in a gene. Strikingly, although Alus in 3′-UTR are indifferently inserted in the forward or reverse direction, 99% of polyadenylation-active Alu sequences are forward oriented (Chen et al., 2009).

Recently, it was estimated that ~8% of all mammalian poly(A) sites are associated with TEs (Lee et al., 2008a). Interestingly, human poly(A) sites that are not conserved in mouse were found to be associated with TEs to a much greater extent than the conserved ones. This result suggests the involvement of TEs in creation or modulation of poly(A) sites in evolution.

5.5. REs as transcriptional silencers

Some retrotransposons are known to function as transcriptional silencers by downregulating transcription of the enclosing genes. For example, one out of 44 Alu repeats located in human *GH* locus encoding for human growth hormone genes *hGH-1* and *hGH-2* harbors a regulatory element that most probably acts by decreasing the rate of promoter-associated histone acetylation, resulting in a significant decrease of RNA polymerase II recruitment to the promoter. This silencer likely provides for regulatory control of *hGH* gene expression in pituitary cells (Trujillo et al., 2006).

Expression of the tumor suppressor protein BRCA2 is tightly regulated throughout development. Sharan et al. identified a transcriptional silencer at the distal part of the human *BRCA2* gene promoter. This silencer was involved in the tissue-specific negative regulation of *BRCA2* expression in breast cell lines. The former mapped 221-base pair long silencer region was also a part of a full-length Alu element (Sharan et al., 1999).

Another example is the transcriptional regulation of a human gene *Hpr* for haptoglobin-related protein. *Hpr* sequence is 92% identical to haptoglobin gene *HP* (Maeda, 1985; Smith et al., 1995). Both genes are transcribed at the higher levels in liver. *Hpr* promoter is stronger than *HP* promoter (Oliviero et al., 1987), but the concentration of *Hpr* liver transcripts is ~17-fold lower than the one of *HP* mRNA (Hatada et al., 2003). The major distinction between these genes is the endogenous retroviral sequence RTVL-Ia in the intron of *Hpr* (Maeda and Kim, 1990). RTVL-Ia fragment has demonstrated significant silencer activity in a series of luciferase transient transfection experiments (Hatada et al., 2003). The mechanism of the negative *Hpr* regulation by the RTVL-Ia ERV is not clear, but the authors propose that this effect is due to an aberrant splicing of the *Hpr* transcript with the retroviral sequences. This hypothesis was supported by the identification of the corresponding abnormal transcripts (Hatada et al., 2003).

5.6. REs as antisense regulators of the host gene transcription

It has been demonstrated that RE inserts in gene introns are preferentially fixed in the antisense orientation relatively to enclosing gene transcriptional direction (Medstrand et al., 2002; van de Lagemaat et al., 2006). Therefore, promoters of intronic retrotransposons 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 (Copeland et al., 2007; Domansky et al., 2000; Dunn et al., 2006; Feuchter and Mager, 1990; Huh et al., 2008; Matlik et al., 2006), and even downstream insertions of these elements relatively to genes may result in production of the antisense RNA. These complementary RNAs may alter functional host gene expression. The possibility of retrotransposon involvement in antisense regulation of gene expression was suggested few years ago (Mack et al., 2004). Retroposition likely accounts for the origin of a significant number of functional sense–antisense pairs in eukaryotic genomes (Galante et al., 2007). Recently, applied CAGE technology identified 48,718 human gene antisense transcriptional start sites within TEs (Conley et al., 2008a).

Gogvadze et al. found 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 (Gogvadze and Buzdin, 2009;

Gogvadze et al., 2009). 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. 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 (Kovalskaya et al., 2006). 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. Similarly, intronically located 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 (Kashkush and Khasdan, 2007).

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* (Galante et al., 2007). 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 nonfunctional RNAs containing multiple premature transcription termination codons. Normally, such RNAs are immediately degraded in the cytoplasm by nonsense-mediated decay machinery (Fasken and Corbett, 2005). Alternatively, antisense transcript base pairing to the target RNA can lead to its rapid enzymatic degradation directly in the nucleus.

D. melanogaster genome has no active copies of telomerase gene. Remarkably, transcription of *Drosophila* retrotransposons HeT-A, TART, and TAHRE that have an important function of maintaining *D. melanogaster* telomere lengths instead of telomerase is tightly regulated by a specialized RNAi mechanism. This mechanism acts through so-called repeat-associated short interfering (rasi)RNAs. Telomeric retrotransposons are bidirectionally transcribed, and the antisense transcription in ovaries is regulated by a promoter localized within its 3′-UTR. The expression of antisense transcripts of telomeric elements is regulated by the RNA silencing machinery, suggesting rasiRNA-mediated interplay between sense and antisense transcripts in the cell (Shpiz et al., 2009). In the genome of yeasts, Ty1 retrotransposon is most likely regulated by the antisense transcripts

encompassing its 5′-LTR, that mediate RNA-dependent gene silencing and repress Ty1 mobility. This Ty1 regulatory RNA was shown to repress Ty1 transcription and transposition in *trans* by acting on the *de novo* transcribed Ty1 RNA (Berretta et al., 2008).

Smalheiser and Torvik (2005) showed that some mammalian miRNA precursors are derived from intronic insertion of two adjacent LINE retrotransposons oriented opposite to each other. Some other elements have an intrinsic hairpin structure and/or serve as miRNA precursors when inserted into transcriptionally active genomic regions (Hernandez-Pinzon et al., 2009; Piriyapongsa and Jordan, 2007).

5.7. REs as insulator elements

The temporal and spatial regulation of gene expression is linked to the establishment of functional chromatin domains. Several lines of evidence have been provided recently that retrotransposons can serve *in vivo* as insulator sequences that distinguish blocks of active and transcriptionally silent chromatin. For example, a B2 SINE element located in the murine growth hormone locus is required for the correct spatio-temporal activation of that gene. This repeat serves as a boundary to block the influence of repressive chromatin modifications by generating short transcripts, which are necessary and sufficient to enable gene activation (Lunyak et al., 2007). Mammalian LINE elements are frequently found within matrix attachment regions (MARs) (Akopov et al., 2006; Purbowasito et al., 2004). Some *Drosophila* LTR retrotransposons have insulator activity and may block the activity of transcriptional enhancer elements when located between enhancer and promoter (Dorsett, 1993; Kostyuchenko et al., 2008). For example, in some fruitfly lineages, there is an insert of LTR retrotransposon *gypsy* into the 5′-region of the gene *yellow* that is responsible for the pigmentation of cuticula. Upstream of the *gypsy* element there are two enhancer elements that account for the transcription of *yellow* in different tissues; another enhancer that is responsible for the *yellow* expression in cilia is located downstream. In the lineage y^2, *gypsy* insertion between the promoter and two upstream enhancers blocks these enhancers and downregulates *yellow* in the corresponding tissues, but the *yellow* expression in cilia remains unaffected (Dorsett, 1993).

5.8. REs as regulators of translation

Although REs have been found in UTRs of many functional cellular genes, effect of their presence on the translational regulation of gene expression is still poorly investigated. Among the few known examples, there is human zinc-finger gene *ZNF177*, which incorporates Alu and L1 segments into the 5′-UTR of transcripts. The presence of the Alu and L1 segments which

form one 5′-UTR exon modifies gene expression on the protein level by decreasing translation efficiency. Interestingly, the same Alu and L1 repeats in the 5′-UTR of *ZNF177* exert a positive transcriptional enhancer effect, but repress translation (Landry et al., 2001). Approximately 4% of human 5′-UTRs harbor Alu sequences, indicating that the expression of many genes might be influenced by Alu repeats (Landry et al., 2001). In the mouse genome, there is a SINE retrotransposon-derived gene for neuronal dendrite-specific BC1 RNA. This small, nonprotein-coding RNA is thought to somehow regulate translation in dendritic microdomains. However, the mechanism of such a regulation remains a mystery, and further efforts are needed to investigate this phenomenon (Khanam et al., 2007).

6. Retrotransposons as Drivers of Mammalian Genome Evolution

6.1. REs generate new REs

RE integrations into the genome can cause multiple effects and, among them, they may lead to the formation of novel REs, as in case of SVA elements. SVA is a composite element consisting of four parts: hexamer repeats $(CCCTCT)_n$, Alu, 15–23 tandemly repeated sequences (VNTR), and SINE-R (SVA = SINE-R + VNTR + Alu) (Ostertag et al., 2003; Wang et al., 2005). These elements are nonautonomous and are mobilized by L1-encoded proteins in *trans*. The SVA family that is thought to be the youngest genus of primate REs is presented by ~3000 copies in the modern human genome (Ostertag et al., 2003). The first SVA probably appeared due to the integration of several elements into the same genomic locus (Wang et al., 2005). SVAs are flanked by TSDs, terminate in a poly(A) tail, and are occasionally truncated and inverted during their integration into the genome. SVA remain active in the human DNA. Several genetic diseases have been reported to be due to SVA insertions (Hancks and Kazazian, 2010). However, their impact in human genome diversity is not restricted to insertion mutagenesis. Evolution of this complex retrotransposon is still going on, first, via quantitative and qualitative changes in tandem repeats, oligomerization, and acquisition of new sequences. This acquisition of genomic sequences by SVA elements may occur in the middle part of an SVA (e.g., due to pseudogene insertion into SVA element), or on SVA termini.

Recently, a novel human-specific family of TEs that consists of fused copies of the CpG island containing first exon of gene *MAST2* and retrotransposon SVA was discovered (Bantysh and Buzdin, 2009; Damert et al., 2009; Hancks and Kazazian, 2010; Hancks et al., 2009). A mechanism proposed for generation of this family comprises an aberrant splicing event. After the divergence of human and chimpanzee ancestor lineages,

retrotransposon SVA has inserted into the first intron of gene *MAST2* in the sense orientation. Due to splicing of an aberrant RNA driven by *MAST2* promoter, but terminally processed using SVA polyadenylation signal, the first exon of *MAST2* has fused to a spliced 3′-terminal fragment of SVA retrotransposon. The above ancestor CpG–SVA element due to retrotranspositions of its own copies has formed a novel family presented in the human genome by 76 members. Recruitment of a *MAST2* CpG island was probably beneficial to the hybrid retrotransposon as a positive transcriptional regulator.

Furthermore, it is speculated that LTR-containing retrotransposons and SINEs themselves represent chimeric elements (Kramerov and Vassetzky, 2001, 2005; Malik and Eickbush, 2001; Ohshima et al., 1996). A phylogenetic analysis of the ribonuclease H domain revealed that LTR-containing REs might have been formed as a fusion between DNA transposon and non-LTR retrotransposon (Malik and Eickbush, 2001). tRNA-derived SINEs likely descended from retroviral strong-stop DNAs (Ohshima et al., 1996). 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 in the genome, namely, it seems very likely that they recruited the enzymatic machinery from LINEs through a common "tail" sequence (Ohshima et al., 1996).

6.2. REs and recombination events

Recombination is a powerful factor of evolution that produces genetic variability by using already existing blocks of biological information (Makalowski, 2000). Because of their high copy number and sequence similarity, REs are the 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 (Boissinot et al., 2006; Song and Boissinot, 2007). Recombination causes genetic rearrangements that can be deleterious, advantageous, or null.

There are numerous reported cases of human diseases caused by recombination between REs. For example, glycogen storage disease (Burwinkel and Kilimann, 1998), Alport syndrome (Segal et al., 1999) as a result of recombination between L1 elements and complete germ cell aplasia due to

recombination between HERV-I (Kamp et al., 2000). Alu elements were implicated in almost 50 disease-causing recombination events (Belancio et al., 2008a; Xing et al., 2009).

Apart from deleterious effects, recombination between REs can also have positive consequences. For example, human glycophorin gene family evolved through several duplication steps that involved recombination between Alu elements (Makalowski, 2000). Furthermore, Alu-derived ectopic recombination generated 492 human-specific deletions, the distribution of which is biased toward gene-rich regions of the genome (Sen et al., 2006). About 60% of Alu recombination-mediated deletions were shown to be located in genes and, in at least three cases, exons have been deleted in human genes relative to their chimpanzee orthologs. Finally, L1s were shown to join DNA breaks by inserting into the genome through EN-independent pathway, thus participating in DNA double-strand breaks repair (Morrish et al., 2002).

6.3. Transduction of flanking sequences

The ability to transduce 3′-flanking DNA to new genomic loci was firstly shown for L1 elements (Goodier et al., 2000; Moran et al., 1999; Pickeral et al., 2000). 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 (Ostertag et al., 2003; Wang et al., 2005). Moreover, SVA-mediated transduction can serve as a previously uncharacterized mechanism for gene duplication and the creation of new gene families (Xing et al., 2006).

In 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 more than 1500 bp. Probably, the most striking example of this phenomenon is the transduction of a whole gene *AMAC* (acyl-malonyl condensing enzyme 1) in the great ape genomes (Xing et al., 2006). Due to SVA 3′-transduction, human genome has three functional 1.2 kb-long copies of *AMAC* gene, and at least two of them are transcribed in different human tissues.

Another kind of transduction results in attaching of new sequences to the 5′-end of an SVA. RE 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 RE. 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 in a new RE genomic insert carrying 5′-transduced part (Brosius, 1999a).

6.4. Formation of processed pseudogenes

Genomes of all higher eukaryotes contain pseudogenes. These elements normally do not contain introns, end in a poly(A) tail, and are flanked by short direct repeats. Such pseudogenes are referred to as processed pseudogenes (Weiner et al., 1986) and are believed to be produced by the action of LINE retrotransposons (Esnault et al., 2000).

As long as RNA polymerase II-transcribed genes generally lack any promoter sequence in their RNA, processed pseudogenes were classically thought to be transcriptionally silent. Indeed, there were not so many reported cases of active pseudogenes that happened to integrate within an existing transcription unit and gave rise to a novel gene or a novel transcriptional pattern of the existing ones. These include *jingway* element of *Drosophila yakuba* and *D. melanogaster* formed due to integration of alcohol dehydrogenase pseudogene into *yellow-emperor* gene (Long et al., 1999), mouse *PMSE2b* retrogene inserted into the L1 sequence under the control of LINE promoter (Zaiss and Kloetzel, 1999), mouse *PHGP* pseudogene, which is expressed from its 5′-adjacent sequence in a tissue-specific manner (Boschan et al., 2002), *TRIMCyp* gene of owl monkey, formed by retrotransposition of cyclophilin A transcript to intron 7 of TRIM5 ubiquitin ligase and shown to confer HIV-1 resistance in owl monkey (Babushok et al., 2007), and several others. However, recent genome-wide analysis of EST databases as well as transcriptional analyses of individual pseudogenes have revealed that up to a third of processed pseudogenes are transcribed, most of them specifically in testes (Babushok et al., 2007; Vinckenbosch et al., 2006). In humans, >1000 pseudogene transcripts were detected and the number of functionally active pseudogenes was estimated to be ~ 120 (Vinckenbosch et al., 2006). Interestingly, a striking predominance of autosomal retrogenes, which are copies of X-linked parental genes, was shown. These autosomal substitutes probably sustain essential functions during male X chromosome inactivation in the process of spermatogenesis (Babushok et al., 2007; Vinckenbosch et al., 2006).

6.5. Chimeric retrogene formation during reverse transcription

Apart from RE retrotransposition and formation of pseudogenes, RT is also able to change templates during cDNA synthesis. This feature of RT is well known for retroviruses. The RT jumps from one place of the template to another are necessary for the synthesis of retroviral LTRs (Temin, 1993).

Template switches can also occur during LINE-directed reverse transcription. Recently, bipartite and tripartite chimeric retrogenes were found in three mammalian and in one fungal genomes. A total of 82, 116, 66, and 31 elements were found in human, mouse, rat, and rice blast fungus *Magnaporthe grisea* DNAs, respectively (Buzdin et al., 2003, 2007; Fudal et al., 2005; Gogvadze et al., 2007). These elements are composed of DNA copies of different 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, 7SL RNA, and Alu retroposon. The chimeras have the following common features: (i) 5′-parts are full-length copies of cellular RNAs, whereas 3′-parts are 5′-truncated copies of the corresponding RNAs (mostly LINEs); (ii) both parts are directly joined with the same transcriptional orientation; (iii) chimeras have a poly(A) tail at their 3′-end, and (iv) chimeras are flanked by short direct repeats. The last structural feature demonstrates that these elements were transposed as bipartite DNA copies. The simultaneous integration of both parts of these chimeras was confirmed by the data obtained from PCR-based multispecies insertion polymorphism assay (Buzdin et al., 2003). The chimeras were formed by a template switch during LINE reverse transcription. This mechanism was further supported by the direct analysis of LINE retrotranspositions *in vitro* and *in vivo* (Babushok et al., 2006; Gilbert et al., 2005). The presence of structurally similar chimeric elements in evolutionary distinct organism shows that template switching during LINE reverse transcription represents an evolutionary conserved mechanism of genome rearrangement. Moreover, many of the chimeras can be considered as new genes, as they were shown to be transcribed, some of them in a tissue-specific manner (Buzdin et al., 2003; Gogvadze et al., 2007).

Except generating chimeric retrogenes, template switches during LINE reverse transcription could give rise to chimeric SINE elements (Nishihara et al., 2006) and to mosaic rodent L1 structures (Brosius, 1999a; Hayward et al., 1997). Evolution of certain LINE families might also involve change of a template during reverse transcription, resulting in fusion of the 3′-part of a LINE to a new sequence, as suggested by the observation that the 5′-UTRs of human, mouse, rat, and rabbit L1 families share no considerable sequence identity (Furano, 2000).

7. Concluding Remarks

In this chapter, we have tried to put together the major findings on the impact of TEs in both functioning of eukaryotic cells and in development of modern biotechnology. About 1000 papers on eukaryotic TEs were appearing annually during the past decade, and the total number of publications on the TEs is close to 20,000 for all years. Therefore, a lot of information left beyond the frameworks of this chapter. This ensures also that when this chapter will be published, many novel interesting and/or important related cases will be known. Moreover, the ongoing progress in sequencing technologies gives a realistic promise that not only a qualitative, but also an integrated quantitative figure of the TE impact on the eukaryotic organisms functioning in health and disease will become available in the nearest future.

ACKNOWLEDGMENTS

E. V. G. and A. A. B. were supported by the Russian Foundation for Basic Research grants 09-04-12302 and 10-04-00593-a, by the President of the Russian Federation grant MD-2010, and by the Program "Molecular and Cellular Biology" of the Presidium of the Russian Academy of Sciences. C. M. is grateful for constant support by Dieter Häussinger and the Heinz-Ansmann Foundation for AIDS Research. G. G. S was supported by grant DA 545/2-1 of the *Deutsche Forschungsgemeinschaft*. H. F. was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan and the Program for Promotion of Basic Research Activities for Innovative Bioscience (PROBRAIN).

REFERENCES

Adelson, D.L., Raison, J.M., Edgar, R.C., 2009. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. Proc. Natl. Acad. Sci. USA 106, 12855–12860.

Akopov, S.B., Ruda, V.M., Batrak, V.V., Vetchinova, A.S., Chernov, I.P., Nikolaev, L.G., et al., 2006. Identification, genome mapping, and CTCF binding of potential insulators within the FXYD5-COX7A1 locus of human chromosome 19q13.12. Mamm. Genome 17, 1042–1049.

An, W., Han, J.S., Wheelan, S.J., Davis, E.S., Coombes, C.E., Ye, P., et al., 2006. Active retrotransposition by a synthetic L1 element in mice. Proc. Natl. Acad. Sci. USA 103, 18662–18667.

An, W., Han, J.S., Schrum, C.M., Maitra, A., Koentgen, F., Boeke, J.D., 2008. Conditional activation of a single-copy L1 transgene in mice by Cre. Genesis 46, 373–383.

Ando, T., Fujiyuki, T., Kawashima, T., Morioka, M., Kubo, T., Fujiwara, H., 2007. In vivo gene transfer into the honeybee using a nucleopolyhedrovirus vector. Biochem. Biophys. Res. Commun. 352, 335–340.

Anzai, T., Takahashi, H., Fujiwara, H., 2001. Sequence-specific recognition and cleavage of telomeric repeat (TTAGG)(*n*) by endonuclease of non-long terminal repeat retrotransposon TRAS1. Mol. Cell. Biol. 21, 100–108.

Anzai, T., Osanai, M., Hamada, M., Fujiwara, H., 2005. Functional roles of 3′-terminal structures of template RNA during in vivo retrotransposition of non-LTR retrotransposon, R1Bm. Nucleic Acids Res. 33, 1993–2002.

Aravin, A.A., Bourc'his, D., 2008. Small RNA guides for de novo DNA methylation in mammalian germ cells. Genes Dev. 22, 970–975.

Aravin, A.A., Hannon, G.J., 2008. Small RNA silencing pathways in germ and stem cells. Cold Spring Harb. Symp. Quant. Biol. 73, 283–290.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., et al., 2006. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442, 203–207.

Aravin, A.A., Hannon, G.J., Brennecke, J., 2007a. The Piwi–piRNA pathway provides an adaptive defense in the transposon arms race. Science 318, 761–764.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., Hannon, G.J., 2007b. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 316, 744–747.

Arnaud, P., Goubely, C., Pelissier, T., Deragon, J.M., 2000. SINE retroposons can be used in vivo as nucleation centers for de novo methylation. Mol. Cell. Biol. 20, 3434–3441.

Athanasiadis, A., Rich, A., Maas, S., 2004. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2, e391.

Babushok, D.V., Ostertag, E.M., Courtney, C.E., Choi, J.M., Kazazian Jr., H.H., 2006. L1 integration in a transgenic mouse model. Genome Res. 16, 240–250.

Babushok, D.V., Ostertag, E.M., Kazazian Jr., H.H., 2007. Current topics in genome evolution: molecular mechanisms of new gene formation. Cell. Mol. Life Sci. 64, 542–554.

Bamshad, M.J., Wooding, S., Watkins, W.S., Ostler, C.T., Batzer, M.A., Jorde, L.B., 2003. Human population genetic structure and inference of group membership. Am. J. Hum. Genet. 72, 578–589.

Bannert, N., Kurth, R., 2004. Retroelements and the human genome: new perspectives on an old relation. Proc. Natl. Acad. Sci. USA 101 (Suppl. 2), 14572–14579.

Bantysh, O.B., Buzdin, A.A., 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.

Batzer, M.A., Schmid, C.W., Deininger, P.L., 1993. Evolutionary analyses of repetitive DNA sequences. Methods Enzymol. 224, 213–232.

Batzer, M.A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D.H., Shaikh, T.H., et al., 1994. African origin of human-specific polymorphic Alu insertions. Proc. Natl. Acad. Sci. USA 91, 12288–12292.

Baum, C., von Kalle, C., Staal, F.J., Li, Z., Fehse, B., Schmidt, M., et al., 2004. Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. Mol. Ther. 9, 5–13.

Baus, J., Liu, L., Heggestad, A.D., Sanz, S., Fletcher, B.S., 2005. Hyperactive transposase mutants of the Sleeping Beauty transposon. Mol. Ther. 12, 1148–1156.

Baust, C., Seifarth, W., Germaier, H., Hehlmann, R., Leib Mosch, C., 2000. HERV-K-T47D-Related long terminal repeats mediate polyadenylation of cellular transcripts. Genomics 66, 98–103.

Beauregard, A., Curcio, M.J., Belfort, M., 2008. The take and give between retrotransposable elements and their hosts. Annu. Rev. Genet. 42, 587–617.

Belancio, V.P., Hedges, D.J., Deininger, P., 2008a. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res. 18, 343–358.

Belancio, V.P., Roy-Engel, A.M., Deininger, P., 2008b. The impact of multiple splice sites in human L1 elements. Gene 411, 38–45.

Belur, L.R., Frandsen, J.L., Dupuy, A.J., Ingbar, D.H., Largaespada, D.A., Hackett, P.B., et al., 2003. Gene insertion and long-term expression in lung mediated by the Sleeping Beauty transposon system. Mol. Ther. 8, 501–507.

Bergman, C.M., Quesneville, H., Anxolabehere, D., Ashburner, M., 2006. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. Genome Biol. 7, R112.

Berretta, J., Pinskaya, M., Morillon, A., 2008. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. Genes Dev. 22, 615–626.

Besansky, N.J., Paskewitz, S.M., Hamm, D.M., Collins, F.H., 1992. Distinct families of site-specific retrotransposons occupy identical positions in the rRNA genes of *Anopheles gambiae*. Mol. Cell. Biol. 12, 5102–5110.

Beumer, K.J., Trautman, J.K., Bozas, A., Liu, J.L., Rutter, J., Gall, J.G., et al., 2008. Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. Proc. Natl. Acad. Sci. USA 105, 19821–19826.

Bieche, I., Laurent, A., Laurendeau, I., Duret, L., Giovangrandi, Y., Frendo, J.L., et al., 2003. Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element. Biol. Reprod. 68, 1422–1429.

Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.J., Malim, M.H., 2004. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr. Biol. 14, 1392–1396.

Blow, M., Futreal, P.A., Wooster, R., Stratton, M.R., 2004. A survey of RNA editing in human brain. Genome Res. 14, 2379–2387.

Bogerd, H.P., Wiegand, H.L., Doehle, B.P., Lueders, K.K., Cullen, B.R., 2006a. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res. 34, 89–95.

Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., et al., 2006b. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc. Natl. Acad. Sci. USA 103, 8780–8785.

Bohne, A., Brunet, F., Galiana-Arnoux, D., Schultheis, C., Volff, J.N., 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. Chromosome Res. 16, 203–215.

Boissinot, S., Davis, J., Entezam, A., Petrov, D., Furano, A.V., 2006. Fitness cost of LINE-1 (L1) activity in humans. Proc. Natl. Acad. Sci. USA 103, 9590–9594.

Bonini, C., Grez, M., Traversari, C., Ciceri, F., Marktel, S., Ferrari, G., et al., 2003. Safety of retroviral gene marking with a truncated NGF receptor. Nat. Med. 9, 367–369.

Borodulina, O.R., Kramerov, D.A., 2008. Transcripts synthesized by RNA polymerase III can be polyadenylated in an AAUAAA-dependent manner. RNA 14, 1865–1873.

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.

Brosius, J., 1999a. Genomes were forged by massive bombardments with retroelements and retrosequences. Genetica 107, 209–238.

Brosius, J., 1999b. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 238, 115–134.

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.

Bucheton, A., 1990. I transposable elements and I-R hybrid dysgenesis in *Drosophila*. Trends Genet. 6, 16–21.

Burke, W.D., Calalang, C.C., Eickbush, T.H., 1987. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. Mol. Cell. Biol. 7, 2221–2230.

Burke, W.D., Muller, F., Eickbush, T.H., 1995. R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. Nucleic Acids Res. 23, 4628–4634.

Burke, W.D., Singh, D., Eickbush, T.H., 2003. R5 retrotransposons insert into a family of infrequently transcribed 28S rRNA genes of planaria. Mol. Biol. Evol. 20, 1260–1270.
Burwinkel, B., 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.
Buzdin, A.A., 2004. Retroelements and formation of chimeric retrogenes. Cell. Mol. Life Sci. 61, 2046–2059.
Buzdin, A., 2007. Human-specific endogenous retroviruses. ScientificWorldJournal 7, 1848–1868.
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.
Buzdin, A., Kovalskaya-Alexandrova, E., Gogvadze, E., Sverdlov, E., 2006a. 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.
Buzdin, A., Kovalskaya-Alexandrova, E., Gogvadze, E., Sverdlov, E., 2006b. GREM, a technique for genome-wide isolation and quantitative analysis of promoter active repeats. Nucleic Acids Res. 34, e67.
Buzdin, A., Gogvadze, E., Lebrun, M.H., 2007. Chimeric retrogenes suggest a role for the nucleolus in LINE amplification. FEBS Lett. 581, 2877–2882.
Cappello, J., Handelsman, K., Lodish, H.F., 1985. Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. Cell 43, 105–115.
Carlson, C.M., Dupuy, A.J., Fritz, S., Roberg-Perez, K.J., Fletcher, C.F., Largaespada, D.A., 2003. Transposon mutagenesis of the mouse germline. Genetics 165, 243–256.
Carlton, V.E., Harris, B.Z., Puffenberger, E.G., Batta, A.K., Knisely, A.S., Robinson, D.L., et al., 2003. Complex inheritance of familial hypercholanemia with associated mutations in TJP2 and BAAT. Nat. Genet. 34, 91–96.
Carmell, M.A., Hannon, G.J., 2004. RNase III enzymes and the initiation of gene silencing. Nat. Struct. Mol. Biol. 11, 214–218.
Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., et al., 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev. Cell 12, 503–514.
Chambeyron, S., Bucheton, A., 2005. I elements in *Drosophila*: in vivo retrotransposition and regulation. Cytogenet. Genome Res. 110, 215–222.
Chandler, V.L., Walbot, V., 1986. DNA modification of a maize transposable element correlates with loss of activity. Proc. Natl. Acad. Sci. USA 83, 1767–1771.
Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., et al., 2006. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. Curr. Biol. 16, 480–485.
Chen, C., Ara, T., Gautheret, D., 2009. Using Alu elements as polyadenylation sites: a case of retroposon exaptation. Mol. Biol. Evol. 26, 327–334.
Chiu, Y.L., Greene, W.C., 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu. Rev. Immunol. 26, 317–353.
Chiu, Y.L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., et al., 2006. High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proc. Natl. Acad. Sci. USA 103, 15588–15593.
Chou, H.H., Takematsu, H., Diaz, S., Iber, J., Nickerson, E., Wright, K.L., et al., 1998. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. Proc. Natl. Acad. Sci. USA 95, 11751–11756.

Christensen, S.M., Bibillo, A., Eickbush, T.H., 2005. Role of the *Bombyx mori* R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. Nucleic Acids Res. 33, 6461–6468.

Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., Largaespada, D.A., 2005. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature 436, 272–276.

Conley, A.B., Miller, W.J., Jordan, I.K., 2008a. Human cis natural antisense transcripts initiated by transposable elements. Trends Genet. 24, 53–56.

Conley, A.B., Piriyapongsa, J., Jordan, I.K., 2008b. Retroviral promoters in the human genome. Bioinformatics 24, 1563–1567.

Conticello, S.G., 2008. The AID/APOBEC family of nucleic acid mutators. Genome Biol. 9, 229.

Conticello, S.G., Langlois, M.A., Yang, Z., Neuberger, M.S., 2007. DNA deamination in immunity: AID in the context of its APOBEC relatives. Adv. Immunol. 94, 37–73.

Cooley, L., Kelley, R., Spradling, A., 1988. Insertional mutagenesis of the *Drosophila* genome with single P elements. Science 239, 1121–1128.

Copeland, C.S., Mann, V.H., Brindley, P.J., 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.

Cordaux, R., Hedges, D.J., Herke, S.W., Batzer, M.A., 2006. Estimating the retrotransposition rate of human Alu elements. Gene 373, 134–137.

Corvelo, A., Eyras, E., 2008. Exon creation and establishment in human genes. Genome Biol. 9, R141.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., et al., 2009. L1 retrotransposition in human neural progenitor cells. Nature 460, 1127–1131.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., Lin, H., 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev. 12, 3715–3727.

Cui, X., Davis, G., 2007. Mobile group II intron targeting: applications in prokaryotes and perspectives in eukaryotes. Front. Biosci. 12, 4972–4985.

Cui, Z., Geurts, A.M., Liu, G., Kaufman, C.D., Hackett, P.B., 2002. Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon. J. Mol. Biol. 318, 1221–1235.

Cutter, A.D., Good, J.M., Pappas, C.T., Saunders, M.A., Starrett, D.M., Wheeler, T.J., 2005. Transposable element orientation bias in the *Drosophila melanogaster* genome. J. Mol. Evol. 61, 733–741.

Damert, A., Raiz, J., Horn, A.V., Lower, J., Wang, H., Xing, J., et al., 2009. 5′-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome Res. 19, 1992–2008.

Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., et al., 2009. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. Int. J. Cancer 124, 81–87.

Deng, W., Lin, H., 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Dev. Cell 2, 819–830.

Dewannieux, M., Esnault, C., Heidmann, T., 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. 35, 41–48.

Dewannieux, M., Dupressoir, A., Harper, F., Pierron, G., Heidmann, T., 2004. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nat. Genet. 36, 534–539.

Dewannieux, M., Harper, F., Richaud, A., Letzelter, C., Ribet, D., Pierron, G., et al., 2006. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Res. 16, 1548–1556.

Domansky, A.N., Kopantzev, E.P., Snezhkov, E.V., Lebedev, Y.B., Leib-Mosch, C., Sverdlov, E.D., 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.

Donnelly, S.R., Hawkins, T.E., Moss, S.E., 1999. A conserved nuclear element with a role in mammalian gene regulation. Hum. Mol. Genet. 8, 1723–1728.

Donze, D., Kamakaka, R.T., 2001. RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in *Saccharomyces cerevisiae*. EMBO J. 20, 520–531.

Dorsett, D., 1993. Distance-independent inactivation of an enhancer by the suppressor of Hairy-wing DNA-binding protein of *Drosophila*. Genetics 134, 1135–1144.

Dunn, C.A., van de Lagemaat, L.N., Baillie, G.J., Mager, D.L., 2005. Endogenous retrovirus long terminal repeats as ready-to-use mobile promoters: the case of primate beta3-GAL-T5. Gene 364, 2–12.

Dunn, C.A., Romanish, M.T., Gutierrez, L.E., van de Lagemaat, L.N., Mager, D.L., 2006. Transcription of two human genes from a bidirectional endogenous retrovirus promoter. Gene 366, 335–342.

Dupuy, A.J., Fritz, S., Largaespada, D.A., 2001. Transposition and gene disruption in the male germline of the mouse. Genesis 30, 82–88.

Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., Jenkins, N.A., 2005. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature 436, 221–226.

Dutko, J.A., Schafer, A., Kenny, A.E., Cullen, B.R., Curcio, M.J., 2005. Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. Curr. Biol. 15, 661–666.

Edgell, D.R., 2009. Selfish DNA: homing endonucleases find a home. Curr. Biol. 19, R115–R117.

Eickbush, T.H., 1992. Transposing without ends: the non-LTR retrotransposable elements. New Biol. 4, 430–440.

Eickbush, T.H., Jamburuthugoda, V.K., 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res. 134, 221–234.

Eickbush, D.G., Luan, D.D., Eickbush, T.H., 2000. Integration of *Bombyx mori* R2 sequences into the 28S ribosomal RNA genes of *Drosophila melanogaster*. Mol. Cell. Biol. 20, 213–223.

Esnault, C., Maestre, J., Heidmann, T., 2000. Human LINE retrotransposons generate processed pseudogenes. Nat. Genet. 24, 363–367.

Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A.J., et al., 2005. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature 433, 430–433.

Esnault, C., Millet, J., Schwartz, O., Heidmann, T., 2006. Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. Nucleic Acids Res. 34, 1522–1531.

Esnault, C., Priet, S., Ribet, D., Heidmann, O., Heidmann, T., 2008. Restriction by APOBEC3 proteins of endogenous retroviruses with an extracellular life cycle: ex vivo effects and in vivo "traces" on the murine IAPE and human HERV-K elements. Retrovirology 5, 75.

Essner, J.J., McIvor, R.S., Hackett, P.B., 2005. Awakening gene therapy with Sleeping Beauty transposons. Curr. Opin. Pharmacol. 5, 513–519.

Evgen'ev, M.B., Arkhipova, I.R., 2005. Penelope-like elements—a new class of retroelements: distribution, function and possible evolutionary significance. Cytogenet. Genome Res. 110, 510–521.

Fasken, M.B., Corbett, A.H., 2005. Process or perish: quality control in mRNA biogenesis. Nat. Struct. Mol. Biol. 12, 482–488.

Faye, B., Arnaud, F., Peyretaillade, E., Brasset, E., Dastugue, B., Vaury, C., 2008. Functional characteristics of a highly specific integrase encoded by an LTR-retrotransposon. PLoS ONE 3, e3185.

Feng, Q., Moran, J.V., Kazazian Jr., H.H., Boeke, J.D., 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87, 905–916.

Feng, Q., Schumann, G., Boeke, J.D., 1998. Retrotransposon R1Bm endonuclease cleaves the target sequence. Proc. Natl. Acad. Sci. USA 95, 2083–2088.

Fernando, S., Fletcher, B.S., 2006. Sleeping beauty transposon-mediated nonviral gene therapy. BioDrugs 20, 219–229.

Feuchter, A., Mager, D., 1990. Functional heterogeneity of a large family of human LTR-like promoters and enhancers. Nucleic Acids Res. 18, 1261–1270.

Fischer, S.E., Wienholds, E., Plasterk, R.H., 2001. Regulated transposition of a fish transposon in the mouse germ line. Proc. Natl. Acad. Sci. USA 98, 6759–6764.

Fudal, I., Bohnert, H.U., Tharreau, D., Lebrun, M.H., 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.

Fujimoto, H., Hirukawa, Y., Tani, H., Matsuura, Y., Hashido, K., Tsuchida, K., et al., 2004. Integration of the 5′ end of the retrotransposon, R2Bm, can be complemented by homologous recombination. Nucleic Acids Res. 32, 1555–1565.

Fujiwara, H., Ogura, T., Takada, N., Miyajima, N., Ishikawa, H., Maekawa, H., 1984. Introns and their flanking sequences of *Bombyx mori* rDNA. Nucleic Acids Res. 12, 6861–6869.

Furano, A.V., 2000. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog. Nucleic Acid Res. Mol. Biol. 64, 255–294.

Galante, P.A., Vidal, D.O., de Souza, J.E., Camargo, A.A., de Souza, S.J., 2007. Sense-antisense pairs in mammals: functional and evolutionary considerations. Genome Biol. 8, R40.

Gal-Mark, N., Schwartz, S., Ast, G., 2008. Alternative splicing of Alu exons—two arms are better than one. Nucleic Acids Res. 36, 2012–2023.

Gdula, D.A., Gerasimova, T.I., Corces, V.G., 1996. Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. Proc. Natl. Acad. Sci. USA 93, 9378–9383.

Gentles, A.J., Wakefield, M.J., Kohany, O., Gu, W., Batzer, M.A., Pollock, D.D., et al., 2007. Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. Genome Res. 17, 992–1004.

Gerasimova, T.I., Corces, V.G., 2001. Chromatin insulators and boundaries: effects on transcription and nuclear organization. Annu. Rev. Genet. 35, 193–208.

Geurts, A.M., Yang, Y., Clark, K.J., Liu, G., Cui, Z., Dupuy, A.J., et al., 2003. Gene transfer into genomes of human cells by the sleeping beauty transposon system. Mol. Ther. 8, 108–117.

Geurts, A.M., Collier, L.S., Geurts, J.L., Oseth, L.L., Bell, M.L., Mu, D., et al., 2006a. Gene mutations and genomic rearrangements in the mouse as a result of transposon mobilization from chromosomal concatemers. PLoS Genet. 2, e156.

Geurts, A.M., Wilber, A., Carlson, C.M., Lobitz, P.D., Clark, K.J., Hackett, P.B., et al., 2006b. Conditional gene expression in the mouse using a Sleeping Beauty gene-trap transposon. BMC Biotechnol. 6, 30.

Geurts, A.M., Cost, G.J., Freyvert, Y., Zeitler, B., Miller, J.C., Choi, V.M., et al., 2009. Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325, 433.

Gilbert, N., Lutz, S., Morrish, T.A., Moran, J.V., 2005. Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol. Cell. Biol. 25, 7780–7795.

Girard, A., Hannon, G.J., 2008. Conserved themes in small-RNA-mediated transposon control. Trends Cell Biol. 18, 136–148.

Gladyshev, E.A., Arkhipova, I.R., 2007. Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. Proc. Natl. Acad. Sci. USA 104, 9352–9357.

Gogvadze, E., Buzdin, A., 2009. Retroelements and their impact on genome evolution and functioning. Cell. Mol. Life Sci. 66, 3727–3742.

Gogvadze, E., Barbisan, C., Lebrun, M.H., 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.

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.

Goila-Gaur, R., Strebel, K., 2008. HIV-1 Vif, APOBEC, and intrinsic immunity. Retrovirology 5, 51.

Goodier, J.L., Kazazian Jr., H.H., 2008. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 135, 23–35.

Goodier, J.L., Ostertag, E.M., Kazazian Jr., H.H., 2000. Transduction of 3′-flanking sequences is common in L1 retrotransposition. Hum. Mol. Genet. 9, 653–657.

Goodier, J.L., Zhang, L., Vetter, M.R., Kazazian Jr., H.H., 2007. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. Mol. Cell. Biol. 27, 6469–6483.

Goodwin, T.J., Poulter, R.T., 2004. A new group of tyrosine recombinase-encoding retrotransposons. Mol. Biol. Evol. 21, 746–759.

Gotea, V., Makalowski, W., 2006. Do transposable elements really contribute to proteomes? Trends Genet. 22, 260–267.

Graepler, F., Lemken, M.L., Wybranietz, W.A., Schmidt, U., Smirnow, I., Gross, C.D., et al., 2005. Bifunctional chimeric SuperCD suicide gene—YCD: YUPRT fusion is highly effective in a rat hepatoma model. World J. Gastroenterol. 11, 6910–6919.

Graff, J.R., Herman, J.G., Myohanen, S., Baylin, S.B., Vertino, P.M., 1997. Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methylation. J. Biol. Chem. 272, 22322–22329.

Gu, Y., Kodama, H., Watanabe, S., Kikuchi, N., Ishitsuka, I., Ozawa, H., et al., 2007. The first reported case of Menkes disease caused by an Alu insertion mutation. Brain Dev. 29, 105–108.

Guimond, N., Bideshi, D.K., Pinkerton, A.C., Atkinson, P.W., O'Brochta, D.A., 2003. Patterns of hermes transposition in *Drosophila melanogaster*. Mol. Genet. Genomics 268, 779–790.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., et al., 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 302, 415–419.

Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., et al., 2008. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. 118, 3132–3142.

Hackett, P.B., Ekker, S.C., Largaespada, D.A., McIvor, R.S., 2005. Sleeping beauty transposon-mediated gene therapy for prolonged expression. Adv. Genet. 54, 189–232.

Hambor, J.E., Mennone, J., Coon, M.E., Hanke, J.H., 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.

Han, J.S., Boeke, J.D., 2004. A highly active synthetic mammalian retrotransposon. Nature 429, 314–318.

Han, J.S., Boeke, J.D., 2005. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? Bioessays 27, 775–784.

Hancks, D.C., Kazazian, Jr., H.H., 2010. SVA retrotransposons: evolution and genetic instability. Semin. Cancer Biol. 20, 234-245.
Hancks, D.C., Ewing, A.D., Chen, J.E., Tokunaga, K., Kazazian Jr., H.H., 2009. Exon-trapping mediated by the human retrotransposon SVA. Genome Res. 19, 1983–1991.
Hasler, J., Samuelsson, T., Strub, K., 2007. Useful 'junk': Alu RNAs in the human transcriptome. Cell. Mol. Life Sci. 64, 1793–1800.
Hatada, S., Grant, D.J., Maeda, N., 2003. An intronic endogenous retrovirus-like sequence attenuates human haptoglobin-related gene expression in an orientation-dependent manner. Gene 319, 55–63.
Hayward, B.E., Zavanelli, M., Furano, A.V., 1997. Recombination creates novel L1 (LINE-1) elements in *Rattus norvegicus*. Genetics 146, 641–654.
He, C.X., Shi, D., Wu, W.J., Ding, Y.F., Feng, D.M., Lu, B., et al., 2004. Insulin expression in livers of diabetic mice mediated by hydrodynamics-based administration. World J. Gastroenterol. 10, 567–572.
Hernandez-Pinzon, I., de Jesus, E., Santiago, N., Casacuberta, J.M., 2009. The frequent transcriptional readthrough of the tobacco Tnt1 retrotransposon and its possible implications for the control of resistance genes. J. Mol. Evol. 68, 269–278.
Horie, K., Kuroiwa, A., Ikawa, M., Okabe, M., Kondoh, G., Matsuda, Y., et al., 2001. Efficient chromosomal transposition of a Tc1/mariner-like transposon Sleeping Beauty in mice. Proc. Natl. Acad. Sci. USA 98, 9191–9196.
Horie, K., Yusa, K., Yae, K., Odajima, J., Fischer, S.E., Keng, V.W., et al., 2003. Characterization of Sleeping Beauty transposition and its application to genetic screening in mice. Mol. Cell. Biol. 23, 9189–9207.
Horie, K., Saito, E.S., Keng, V.W., Ikeda, R., Ishihara, H., Takeda, J., 2007. Retrotransposons influence the mouse transcriptome: implication for the divergence of genetic traits. Genetics 176, 815–827.
Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., et al., 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 129, 69–82.
Huang, X., Guo, H., Kang, J., Choi, S., Zhou, T.C., Tammana, S., et al., 2008. Sleeping Beauty transposon-mediated engineering of human primary T cells for therapy of CD19+ lymphoid malignancies. Mol. Ther. 16, 580–589.
Hughes, D.C., 2001. Alternative splicing of the human VEGFGR-3/FLT4 gene as a consequence of an integrated human endogenous retrovirus. J. Mol. Evol. 53, 77–79.
Huh, J.W., Kim, D.S., Kang, D.W., Ha, H.S., Ahn, K., Noh, Y.N., et al., 2008. Transcriptional regulation of GSDML gene by antisense-oriented HERV-H LTR element. Arch. Virol. 153, 1201–1205.
Hulme, A.E., Bogerd, H.P., Cullen, B.R., Moran, J.V., 2007. Selective inhibition of Alu retrotransposition by APOBEC3G. Gene 390, 199–205.
Hultquist, J.F., Harris, R.S., 2009. Leveraging APOBEC3 proteins to alter the HIV mutation rate and combat AIDS. Future Virol. 4, 605.
Hutvagner, G., Simard, M.J., 2008. Argonaute proteins: key players in RNA silencing. Nat. Rev. Mol. Cell Biol. 9, 22–32.
International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.
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.
Ivics, Z., Izsvak, Z., 2006. Transposons for gene therapy!. Curr. Gene Ther. 6, 593–607.
Ivics, Z., Hackett, P.B., Plasterk, R.H., Izsvak, Z., 1997. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 91, 501–510.

Ivics, Z., Katzer, A., Stuwe, E.E., Fiedler, D., Knespel, S., Izsvak, Z., 2007. Targeted Sleeping Beauty transposition in human cells. Mol. Ther. 15, 1137–1144.

Ivics, Z., Li, M.A., Mates, L., Boeke, J.D., Nagy, A., Bradley, A., et al., 2009. Transposon-mediated genome manipulation in vertebrates. Nat. Methods 6, 415–422.

Izsvak, Z., Ivics, Z., 2004. Sleeping beauty transposition: biology and applications for molecular therapy. Mol. Ther. 9, 147–156.

Izsvak, Z., Frohlich, J., Grabundzija, I., Shirley, J.R., Powell, H.M., Chapman, K.M., et al., 2010. Generating knockout rats by transposon mutagenesis in spermatogonial stem cells. Nat. Methods 7, 443–445.

Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., et al., 2002. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics 79, 285–296.

Jensen, S., Heidmann, T., 1991. An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element. EMBO J. 10, 1927–1937.

Jensen, S., Cavarec, L., Dhellin, O., Heidmann, T., 1994. Retrotransposition of a marked *Drosophila* line-like I element in cells in culture. Nucleic Acids Res. 22, 1484–1488.

Jurka, J., 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Natl. Acad. Sci. USA 94, 1872–1877.

Kajikawa, M., Okada, N., 2002. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 111, 433–444.

Kajikawa, M., Ichiyanagi, K., Tanaka, N., Okada, N., 2005. Isolation and characterization of active LINE and SINEs from the eel. Mol. Biol. Evol. 22, 673–682.

Kamp, C., Hirschmann, P., Voss, H., Huellen, K., Vogt, P.H., 2000. Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. Hum. Mol. Genet. 9, 2563–2572.

Kano, H., Godoy, I., Courtney, C., Vetter, M.R., Gerton, G.L., Ostertag, E.M., et al., 2009. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev. 23, 1303–1312.

Kapitonov, V.V., Jurka, J., 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc. Natl. Acad. Sci. USA 100, 6569–6574.

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.

Kawashima, T., Osanai, M., Futahashi, R., Kojima, T., Fujiwara, H., 2007. A novel target-specific gene delivery system combining baculovirus and sequence-specific long interspersed nuclear elements. Virus Res. 127, 49–60.

Kazazian Jr., H.H., 2004. Mobile elements: drivers of genome evolution. Science 303, 1626–1632.

Keng, V.W., Yae, K., Hayakawa, T., Mizuno, S., Uno, Y., Yusa, K., et al., 2005. Region-specific saturation germline mutagenesis in mice using the Sleeping Beauty transposon system. Nat. Methods 2, 763–769.

Khanam, T., Raabe, C.A., Kiefmann, M., Handel, S., Skryabin, B.V., Brosius, J., 2007. Can ID repetitive elements serve as cis-acting dendritic targeting elements? An in vivo study. PLoS ONE 2, e961.

Khatua, A.K., Taylor, H.E., Hildreth, J.E., Popik, W., 2010. Inhibition of LINE-1 and Alu retrotransposition by exosomes encapsidating APOBEC3G and APOBEC3F. Virology 400, 68–75.

Kidd, J.M., Newman, T.L., Tuzun, E., Kaul, R., Eichler, E.E., 2007. Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet. 3, e63.

Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Purd'homme, N., Bucheton, A., 1994. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 91, 1285–1289.

Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., et al., 2004. Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res. 14, 1719–1725.

King, L.M., Francomano, C.A., 2001. Characterization of a human gene encoding nucleosomal binding protein NSBP1. Genomics 71, 163–173.

Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., et al., 2007. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res. 35, 2955–2964.

Kirilyuk, A., Tolstonog, G.V., Damert, A., Held, U., Hahn, S., Lower, R., et al., 2008. Functional endogenous LINE-1 retrotransposons are expressed and mobilized in rat chloroleukemia cells. Nucleic Acids Res. 36, 648–665.

Kitada, K., Ishishita, S., Tosaka, K., Takahashi, R., Ueda, M., Keng, V.W., et al., 2007. Transposon-tagged mutagenesis in the rat. Nat. Methods 4, 131–133.

Kjellman, C., Sjogren, H.O., Salford, L.G., Widegren, B., 1999. HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. Gene 239, 99–107.

Kochanek, S., Clemens, P.R., Mitani, K., Chen, H.H., Chan, S., Caskey, C.T., 1996. A new adenoviral vector: replacement of all viral coding sequences with 28 kb of DNA independently expressing both full-length dystrophin and beta-galactosidase. Proc. Natl. Acad. Sci. USA 93, 5731–5736.

Kojima, K.K., Fujiwara, H., 2003. Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol. Biol. Evol. 20, 351–361.

Kojima, K.K., Fujiwara, H., 2004. Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol. Biol. Evol. 21, 207–217.

Kojima, K.K., Fujiwara, H., 2005a. An extraordinary retrotransposon family encoding dual endonucleases. Genome Res. 15, 1106–1117.

Kojima, K.K., Fujiwara, H., 2005b. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. Mol. Biol. Evol. 22, 2157–2165.

Kojima, K.K., Kuma, K., Toh, H., Fujiwara, H., 2006. Identification of rDNA-specific non-LTR retrotransposons in Cnidaria. Mol. Biol. Evol. 23, 1984–1993.

Kokubu, C., Horie, K., Abe, K., Ikeda, R., Mizuno, S., Uno, Y., et al., 2009. A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice. Nat. Genet. 41, 946–952.

Koning, F.A., Newman, E.N., Kim, E.Y., Kunstman, K.J., Wolinsky, S.M., Malim, M.H., 2009. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. J. Virol. 83, 9474–9485.

Kostyuchenko, M.V., Savitskaya, E.E., Volkov, I.A., Golovnin, A.K., Georgiev, P.G., 2008. Study of functional interaction between three copies of the insulator from the MDG4 transposable element in the model system of the miniwhite gene of *Drosophila melanogaster*. Dokl. Biochem. Biophys. 421, 239–243.

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.

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

Kramerov, D.A., Vassetzky, N.S., 2005. Short retroposons in eukaryotic genomes. Int. Rev. Cytol. 247, 165–221.

Krom, N., Recla, J., Ramakrishna, W., 2008. Analysis of genes associated with retrotransposons in the rice genome. Genetica 134, 297–310.

Kubo, Y., Okazaki, S., Anzai, T., Fujiwara, H., 2001. Structural and phylogenetic analysis of TRAS, telomeric repeat-specific non-LTR retrotransposon families in Lepidopteran insects. Mol. Biol. Evol. 18, 848–857.

Kubo, S., Seleme, M.C., Soifer, H.S., Perez, J.L., Moran, J.V., Kazazian Jr., H.H., et al., 2006. L1 retrotransposition in nondividing and primary human somatic cells. Proc. Natl. Acad. Sci. USA 103, 8036–8041.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., et al., 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development 131, 839–849.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., et al., 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 22, 908–917.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Takamatsu, K., Chuma, S., Kojima-Kita, K., et al., 2010. MVH in piRNA processing and gene silencing of retrotransposons. Genes Dev. 24, 887–892.

Labrador, M., Corces, V.G., 2001. Protein determinants of insertional specificity for the *Drosophila* gypsy retrovirus. Genetics 158, 1101–1110.

Labrador, M., Sha, K., Li, A., Corces, V.G., 2008. Insulator and Ovo proteins determine the frequency and specificity of insertion of the gypsy retrotransposon in *Drosophila melanogaster*. Genetics 180, 1367–1378.

Lambowitz, A.M., Zimmerly, S., 2004. Mobile group II introns. Annu. Rev. Genet. 38, 1–35.

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.

Landry, J.R., Medstrand, P., 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.

Landry, J.R., Rouhi, A., Medstrand, P., Mager, D.L., 2002. The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter. Mol. Biol. Evol. 19, 1934–1942.

LaRue, R.S., Jonsson, S.R., Silverstein, K.A., Lajoie, M., Bertrand, D., El-Mabrouk, N., et al., 2008. The artiodactyl APOBEC3 innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. BMC Mol. Biol. 9, 104.

LaRue, R.S., Andresdottir, V., Blanchard, Y., Conticello, S.G., Derse, D., Emerman, M., et al., 2009. Guidelines for naming nonprimate APOBEC3 genes and proteins. J. Virol. 83, 494–497.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., et al., 2006. Characterization of the piRNA complex from rat testes. Science 313, 363–367.

Le Provost, F., Lillico, S., Passet, B., Young, R., Whitelaw, B., Vilotte, J.L., 2010. Zinc finger nuclease technology heralds a new era in mammalian transgenesis. Trends Biotechnol. 28, 134–141.

Lee, Y.N., Bieniasz, P.D., 2007. Reconstitution of an infectious human endogenous retrovirus. PLoS Pathog. 3, e10.

Lee, J.Y., Ji, Z., Tian, B., 2008a. 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.

Lee, Y.N., Malim, M.H., Bieniasz, P.D., 2008b. Hypermutation of an ancient human retrovirus by APOBEC3G. J. Virol. 82, 8762–8770.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., et al., 2004. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotechnol. 22, 1001–1005.

Lev-Maor, G., Ram, O., Kim, E., Sela, N., Goren, A., Levanon, E.Y., et al., 2008. Intronic Alus influence alternative splicing. PLoS Genet. 4, e1000204.

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.

Li, J., Han, K., Xing, J., Kim, H.S., Rogers, J., Ryder, O.A., et al., 2009. Phylogeny of the macaques (Cercopithecidae: Macaca) based on Alu elements. Gene 448, 242–249.

Liang, Q., Kong, J., Stalker, J., Bradley, A., 2009. Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. Genesis 47, 404–408.

Lin, H., 2007. piRNAs in the germ line. Science 316, 397.

Lin, L., Shen, S., Tye, A., Cai, J.J., Jiang, P., Davidson, B.L., et al., 2008. Diverse splicing patterns of exonized Alu elements in human tissues. PLoS Genet. 4, e1000225.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., et al., 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803–819.

Ling, J., Pi, W., Bollag, R., Zeng, S., Keskintepe, M., Saliman, H., et al., 2002. The solitary long terminal repeats of ERV-9 endogenous retrovirus are conserved during primate evolution and possess enhancer activities in embryonic and hematopoietic cells. J. Virol. 76, 2410–2423.

Ling, J., Pi, W., Yu, X., Bengra, C., Long, Q., Jin, H., et al., 2003. The ERV-9 LTR enhancer is not blocked by the HS5 insulator and synthesizes through the HS5 site noncoding, long RNAs that regulate LTR enhancer function. Nucleic Acids Res. 31, 4582–4596.

Liu, L., Sanz, S., Heggestad, A.D., Antharam, V., Notterpek, L., Fletcher, B.S., 2004. Endothelial targeting of the Sleeping Beauty transposon within lung. Mol. Ther. 10, 97–105.

Liu, L., Mah, C., Fletcher, B.S., 2006. Sustained FVIII expression and phenotypic correction of hemophilia A in neonatal mice using an endothelial-targeted sleeping beauty transposon. Mol. Ther. 13, 1006–1015.

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.

Long, M., Wang, W., Zhang, J., 1999. Origin of new genes and source for N-terminal domain of the chimerical gene, jingwei, in *Drosophila*. Gene 238, 135–141.

Loreni, F., Stavenhagen, J., Kalff, M., Robins, D.M., 1988. A complex androgen-responsive enhancer resides 2 kilobases upstream of the mouse Slp gene. Mol. Cell. Biol. 8, 2350–2360.

Lovsin, N., Peterlin, B.M., 2009. APOBEC3 proteins inhibit LINE-1 retrotransposition in the absence of ORF1p binding. Ann. NY Acad. Sci. 1178, 268–275.

Lu, B., Geurts, A.M., Poirier, C., Petit, D.C., Harrison, W., Overbeek, P.A., et al., 2007. Generation of rat mutants using a coat color-tagged Sleeping Beauty transposon system. Mamm. Genome 18, 338–346.

Luan, D.D., Eickbush, T.H., 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol. Cell. Biol. 15, 3882–3891.

Luan, D.D., Korman, M.H., Jakubczak, J.L., Eickbush, T.H., 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72, 595–605.

Lunyak, V.V., Prefontaine, G.G., Nunez, E., Cramer, T., Ju, B.G., Ohgi, K.A., et al., 2007. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science 317, 248–251.

Luo, G., Ivics, Z., Izsvák, Zs., Bradley, A., 1998. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA 95, 10769–10773.

MacDuff, D.A., Demorest, Z.L., Harris, R.S., 2009. AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. Nucleic Acids Res. 37, 1854–1867.

Mack, M., Bender, K., Schneider, P.M., 2004. Detection of retroviral antisense transcripts and promoter activity of the HERV-K(C4) insertion in the MHC class III region. Immunogenetics 56, 321–332.

Maeda, N., 1985. Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair. The haptoglobin-related gene contains a retrovirus-like element. J. Biol. Chem. 260, 6698–6709.

Maeda, N., Kim, H.S., 1990. Three independent insertions of retrovirus-like sequences in the haptoglobin gene cluster of primates. Genomics 8, 671–683.

Mager, D.L., Hunter, D.G., Schertzer, M., Freeman, J.D., 1999. Endogenous retroviruses provide the primary polyadenylation signal for two new human genes. Genomics 59, 255–263.

Maita, N., Anzai, T., Aoyagi, H., Mizuno, H., Fujiwara, H., 2004. Crystal structure of the endonuclease domain encoded by the telomere-specific long interspersed nuclear element, TRAS1. J. Biol. Chem. 279, 41067–41076.

Maita, N., Aoyagi, H., Osanai, M., Shirakawa, M., Fujiwara, H., 2007. Characterization of the sequence specificity of the R1Bm endonuclease domain by structural and biochemical studies. Nucleic Acids Res. 35, 3918–3927.

Makalowski, W., 2000. Genomic scrap yard: how genomes utilize all that junk. Gene 259, 61–67.

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

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.

Malim, M.H., Emerman, M., 2008. HIV-1 accessory proteins—ensuring viral survival in a hostile environment. Cell Host Microbe 3, 388–398.

Marcaida, M.J., Munoz, I.G., Blanco, F.J., Prieto, J., Montoya, G., 2010. Homing endonucleases: from basics to therapeutic applications. Cell Mol. Life Sci. 67, 727–748.

Mastroianni, M., Watanabe, K., White, T.B., Zhuang, F., Vernon, J., Matsuura, M., et al., 2008. Group II intron-based gene targeting reactions in eukaryotes. PLoS ONE 3, e3121.

Mates, L., Izsvak, Z., Ivics, Z., 2007. Technology transfer from worms and flies to vertebrates: transposition-based genome manipulations and their future perspectives. Genome Biol. 8 (Suppl. 1), S1.

Mates, L., Chuah, M.K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., et al., 2009. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat. Genet. 41, 753–761.

Matlik, K., Redik, K., Speek, M., 2006. L1 antisense promoter drives tissue-specific transcription of human genes. J. Biomed. Biotechnol. 2006, 71753.

Matsumoto, T., Takahashi, H., Fujiwara, H., 2004. Targeted nuclear import of open reading frame 1 protein is required for in vivo retrotransposition of a telomere-specific non-long terminal repeat retrotransposon, SART1. Mol. Cell. Biol. 24, 105–122.

Matsumoto, T., Hamada, M., Osanai, M., Fujiwara, H., 2006. Essential domains for ribonucleoprotein complex formation required for retrotransposition of telomere-specific non-long terminal repeat retrotransposon SART1. Mol. Cell. Biol. 26, 5168–5179.
McClintock, B., 1956. Controlling elements and the gene. Cold Spring Harb. Symp. Quant. Biol. 21, 197–216.
Medstrand, P., Landry, J.R., Mager, D.L., 2001. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J. Biol. Chem. 276, 1896–1903.
Medstrand, P., van de Lagemaat, L.N., Mager, D.L., 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res. 12, 1483–1495.
Meisler, M.H., Ting, C.N., 1993. The remarkable evolutionary history of the human amylase genes. Crit. Rev. Oral Biol. Med. 4, 503–509.
Mevel-Ninio, M., Mariol, M.C., Gans, M., 1989. Mobilization of the gypsy and copia retrotransposons in *Drosophila melanogaster* induces reversion of the ovo dominant female-sterile mutations: molecular analysis of revertant alleles. EMBO J. 8, 1549–1558.
Mitani, K., Graham, F.L., Caskey, C.T., Kochanek, S., 1995. Rescue, propagation, and partial purification of a helper virus-dependent adenovirus vector. Proc. Natl. Acad. Sci. USA 92, 3854–3858.
Mola, G., Vela, E., Fernandez-Figueras, M.T., Isamat, M., Munoz-Marmol, A.M., 2007. Exonization of Alu-generated splice variants in the survivin gene of human and non-human primates. J. Mol. Biol. 366, 1055–1063.
Moldt, B., Yant, S.R., Andersen, P.R., Kay, M.A., Mikkelsen, J.G., 2007. Cis-acting gene regulatory activities in the terminal regions of sleeping beauty DNA transposon-based vectors. Hum. Gene Ther. 18, 1193–1204.
Montini, E., Held, P.K., Noll, M., Morcinek, N., Al-Dhalimy, M., Finegold, M., et al., 2002. In vivo correction of murine tyrosinemia type I by DNA-mediated transposition. Mol. Ther. 6, 759–769.
Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., Kazazian Jr., H.H., 1996. High frequency retrotransposition in cultured mammalian cells. Cell 87, 917–927.
Moran, J.V., DeBerardinis, R.J., Kazazian Jr., H.H., 1999. Exon shuffling by L1 retrotransposition. Science 283, 1530–1534.
Morgan, H.D., Dean, W., Coker, H.A., Reik, W., Petersen-Mahrt, S.K., 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J. Biol. Chem. 279, 52353–52360.
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.
Morrish, T.A., Garcia-Perez, J.L., Stamato, T.D., Taccioli, G.E., Sekiguchi, J., Moran, J.V., 2007. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. Nature 446, 208–212.
Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., et al., 2006. APOBEC3 proteins inhibit human LINE-1 retrotransposition. J. Biol. Chem. 281, 22161–22172.
Munk, C., Beck, T., Zielonka, J., Hotz-Wagenblatt, A., Chareza, S., Battenberg, M., et al., 2008. Functions, structure, and read-through alternative splicing of feline APOBEC3 genes. Genome Biol. 9, R48.
Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., Gage, F.H., 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature 435, 903–910.

Murata, S., Takasaki, N., Saitoh, M., Okada, N., 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc. Natl. Acad. Sci. USA 90, 6995–6999.

Neeman, Y., Levanon, E.Y., Jantsch, M.F., Eisenberg, E., 2006. RNA editing level in the mouse is determined by the genomic repeat repertoire. RNA 12, 1802–1809.

Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., et al., 2007. Genome sequence of Aedes aegypti, a major arbovirus vector. Science 316, 1718–1723.

Niewiadomska, A.M., Tian, C., Tan, L., Wang, T., Sarkis, P.T., Yu, X.F., 2007. Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. J. Virol. 81, 9577–9583.

Nikaido, M., Rooney, A.P., Okada, N., 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA 96, 10261–10266.

Nishihara, H., Smit, A.F., Okada, N., 2006. Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res. 16, 864–874.

OhAinle, M., Kerns, J.A., Malik, H.S., Emerman, M., 2006. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. J. Virol. 80, 3853–3862.

OhAinle, M., Kerns, J.A., Li, M.M., Malik, H.S., Emerman, M., 2008. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell Host Microbe 4, 249–259.

Ohlfest, J.R., Demorest, Z.L., Motooka, Y., Vengco, I., Oh, S., Chen, E., et al., 2005a. Combinatorial antiangiogenic gene therapy by nonviral gene transfer using the sleeping beauty transposon causes tumor regression and improves survival in mice bearing intracranial human glioblastoma. Mol. Ther. 12, 778–788.

Ohlfest, J.R., Frandsen, J.L., Fritz, S., Lobitz, P.D., Perkinson, S.G., Clark, K.J., et al., 2005b. Phenotypic correction and long-term expression of factor VIII in hemophilic mice by immunotolerization and nonviral gene transfer using the Sleeping Beauty transposon system. Blood 105, 2691–2698.

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.

Okazaki, S., Ishikawa, H., Fujiwara, H., 1995. Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. Mol. Cell. Biol. 15, 4545–4552.

Oliviero, S., Morrone, G., Cortese, R., 1987. The human haptoglobin gene: transcriptional regulation during development and acute phase induction. EMBO J. 6, 1905–1912.

Ortiz-Urda, S., Thyagarajan, B., Keene, D.R., Lin, Q., Fang, M., Calos, M.P., et al., 2002. Stable nonviral genetic correction of inherited human skin disease. Nat. Med. 8, 1166–1170.

Osanai, M., Takahashi, H., Kojima, K.K., Hamada, M., Fujiwara, H., 2004. Essential motifs in the 3′ untranslated region required for retrotransposition and the precise start of reverse transcription in non-long-terminal-repeat retrotransposon SART1. Mol. Cell. Biol. 24, 7902–7913.

Osanai-Futahashi, M., Suetsugu, Y., Mita, K., Fujiwara, H., 2008. Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 38, 1046–1057.

Ostertag, E.M., DeBerardinis, R.J., Goodier, J.L., Zhang, Y., Yang, N., Gerton, G.L., et al., 2002. A mouse model of human L1 retrotransposition. Nat. Genet. 32, 655–660.

Ostertag, E.M., Goodier, J.L., Zhang, Y., Kazazian Jr., H.H., 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. Am. J. Hum. Genet. 73, 1444–1451.

Ostertag, E.M., Madison, B.B., Kano, H., 2007. Mutagenesis in rodents using the L1 retrotransposon. Genome Biol. 8 (Suppl. 1), S16.
Paques, F., Duchateau, P., 2007. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr. Gene Ther. 7, 49–66.
Pelisson, A., Finnegan, D.J., Bucheton, A., 1991. Evidence for retrotransposition of the I factor, a LINE element of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 88, 4907–4910.
Perna, N.T., Batzer, M.A., Deininger, P.L., Stoneking, M., 1992. Alu insertion polymorphism: a new type of marker for human population studies. Hum. Biol. 64, 641–648.
Peters, L., Meister, G., 2007. Argonaute proteins: mediators of RNA silencing. Mol. Cell 26, 611–623.
Pickeral, O.K., Makalowski, W., Boguski, M.S., Boeke, J.D., 2000. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. Genome Res. 10, 411–415.
Piriyapongsa, J., Jordan, I.K., 2007. A family of human microRNA genes from miniature inverted-repeat transposable elements. PLoS ONE 2, e203.
Polavarapu, N., Marino-Ramirez, L., Landsman, D., McDonald, J.F., Jordan, I.K., 2008. Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. BMC Genomics 9, 226.
Poulter, R.T., Goodwin, T.J., 2005. DIRS-1 and the other tyrosine recombinase retrotransposons. Cytogenet. Genome Res. 110, 575–588.
Prak, E.T., Dodson, A.W., Farkash, E.A., Kazazian Jr., H.H., 2003. Tracking an embryonic L1 retrotransposition event. Proc. Natl. Acad. Sci. USA 100, 1832–1837.
Pule, M.A., Savoldo, B., Myers, G.D., Rossig, C., Russell, H.V., Dotti, G., et al., 2008. Virus-specific T cells engineered to coexpress tumor-specific receptors: persistence and antitumor activity in individuals with neuroblastoma. Nat. Med. 14, 1264–7120.
Purbowasito, W., Suda, C., Yokomine, T., Zubair, M., Sado, T., Tsutsui, K., et al., 2004. Large-scale identification and mapping of nuclear matrix-attachment regions in the distal imprinted domain of mouse chromosome 7. DNA Res. 11, 391–407.
Rangwala, S.H., Kazazian Jr., H.H., 2009. The L1 retrotransposition assay: a retrospective and toolkit. Methods 49, 219–226.
Rashkova, S., Karam, S.E., Kellum, R., Pardue, M.L., 2002. Gag proteins of the two *Drosophila* telomeric retrotransposons are targeted to chromosome ends. J. Cell Biol. 159, 397–402.
Ray, D.A., Walker, J.A., Batzer, M.A., 2007. Mobile element-based forensic genomics. Mutat. Res. 616, 24–33.
Redondo, P., Prieto, J., Munoz, I.G., Alibes, A., Stricher, F., Serrano, L., et al., 2008. Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. Nature 456, 107–111.
Refsland, E.W., Stenglein, M.D., Shindo, K., Albin, J.S., Brown, W.L., Harris, R.S., 2010. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic Acids Res. 38 (13), 4274–4284.
Remy, S., Tesson, L., Menoret, S., Usal, C., Scharenberg, A.M., Anegon, I., 2010. Zinc-finger nucleases: a powerful tool for genetic engineering of animals. Transgenic Res. 19, 363–371.
Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., Pillai, R.S., 2009. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nat. Struct. Mol. Biol. 16, 639–646.
Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., Heidmann, T., 2008. An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. Genome Res. 18, 597–609.
Riedmann, E.M., Schopoff, S., Hartner, J.C., Jantsch, M.F., 2008. Specificity of ADAR-mediated RNA editing in newly identified targets. RNA 14, 1110–1118.

Roberg-Perez, K., Carlson, C.M., Largaespada, D.A., 2003. MTID: a database of Sleeping Beauty transposon insertions in mice. Nucleic Acids Res. 31, 78–81.

Rogozin, I.B., Iyer, L.M., Liang, L., Glazko, G.V., Liston, V.G., Pavlov, Y.I., et al., 2007. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat. Immunol. 8, 647–656.

Roiha, H., Miller, J.R., Woods, L.C., Glover, D.M., 1981. Arrangements and rearrangements of sequences flanking the two types of rDNA insertion in *D. melanogaster*. Nature 290, 749–753.

Romanish, M.T., Lock, W.M., van de Lagemaat, L.N., Dunn, C.A., Mager, D.L., 2007. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. PLoS Genet. 3, e10.

Roy-Engel, A.M., El-Sawy, M., Farooq, L., Odom, G.L., Perepelitsa-Belancio, V., Bruch, H., et al., 2005. Human retroelements may introduce intragenic polyadenylation signals. Cytogenet. Genome Res. 110, 365–371.

Saito, E.S., Keng, V.W., Takeda, J., Horie, K., 2008. Translation from nonautonomous type IAP retrotransposon is a critical determinant of transposition activity: implication for retrotransposon-mediated genome evolution. Genome Res. 18, 859–868.

Sakate, R., Suto, Y., Imanishi, T., Tanoue, T., Hida, M., Hayasaka, I., et al., 2007. Mapping of chimpanzee full-length cDNAs onto the human genome unveils large potential divergence of the transcriptome. Gene 399, 1–10.

Santangelo, A.M., de Souza, F.S., Franchini, L.F., Bumaschny, V.F., Low, M.J., Rubinstein, M., 2007. Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. PLoS Genet. 3, 1813–1826.

Sasaki, T., Nishihara, H., Hirakawa, M., Fujimura, K., Tanaka, M., Kokubo, N., et al., 2008. Possible involvement of SINEs in mammalian-specific brain formation. Proc. Natl. Acad. Sci. USA 105, 4220–4225.

Sawyer, S.L., Emerman, M., Malik, H.S., 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol. 2, E275.

Scali, C., Nolan, T., Sharakhov, I., Sharakhova, M., Crisanti, A., Catteruccia, F., 2007. Post-integration behavior of a Minos transposon in the malaria mosquito *Anopheles stephensi*. Mol. Genet. Genomics 278, 575–584.

Schiedner, G., Morral, N., Parks, R.J., Wu, Y., Koopmans, S.C., Langston, C., et al., 1998. Genomic DNA transfer with a high-capacity adenovirus vector results in improved in vivo gene expression and decreased toxicity. Nat. Genet. 18, 180–183.

Schostak, N., Pyatkov, K., Zelentsova, E., Arkhipova, I., Shagin, D., Shagina, I., et al., 2008. Molecular dissection of Penelope transposable element regulatory machinery. Nucleic Acids Res. 36, 2522–2529.

Schreck, S., Buettner, M., Kremmer, E., Bogdan, M., Herbst, H., Niedobitek, G., 2006. Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. J. Pathol. 210, 26–31.

Schumacher, A.J., Nissley, D.V., Harris, R.S., 2005. APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. Proc. Natl. Acad. Sci. USA 102, 9854–9859.

Schumann, G.G., 2007. APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem. Soc. Trans. 35, 637–642.

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.

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.

Sen, S.K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P.A., et al., 2006. Human genomic deletions mediated by recombination between Alu elements. Am. J. Hum. Genet. 79, 41–53.

Sharan, C., Hamilton, N.M., Parl, A.K., Singh, P.K., Chaudhuri, G., 1999. Identification and characterization of a transcriptional silencer upstream of the human BRCA2 gene. Biochem. Biophys. Res. Commun. 265, 285–290.

Shen, M.R., Batzer, M.A., Deininger, P.L., 1991. Evolution of the master Alu gene(s). J. Mol. Evol. 33, 311–320.

Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., et al., 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388, 666–670.

Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., et al., 2009. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. Dev. Cell 17, 775–787.

Shpiz, S., Kwon, D., Rozovsky, Y., Kalmykova, A., 2009. rasiRNA pathway controls antisense expression of *Drosophila* telomeric retrotransposons in the nucleus. Nucleic Acids Res. 37, 268–278.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., et al., 2009. Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459, 437–441.

Siomi, M.C., Kuramochi-Miyagawa, S., 2009. RNA silencing in germlines—exquisite collaboration of Argonaute proteins with small RNAs for germline survival. Curr. Opin. Cell Biol. 21, 426–434.

Sironen, A., Vilkki, J., Bendixen, C., Thomsen, B., 2007. Infertile Finnish Yorkshire boars carry a full-length LINE-1 retrotransposon within the KPL2 gene. Mol. Genet. Genomics 278, 385–391.

Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8, 272–285.

Smalheiser, N.R., Torvik, V.I., 2005. Mammalian microRNAs derived from genomic repeats. Trends Genet. 21, 322–326.

Smith, A.B., Esko, J.D., Hajduk, S.L., 1995. Killing of trypanosomes by the human haptoglobin-related protein. Science 268, 284–286.

Soifer, H.S., Kasahara, N., 2004. Retrotransposon-adenovirus hybrid vectors: efficient delivery and stable integration of transgenes via a two-stage mechanism. Curr. Gene Ther. 4, 373–384.

Soifer, H., Higo, C., Kazazian Jr., H.H., Moran, J.V., Mitani, K., Kasahara, N., 2001. Stable integration of transgenes delivered by a retrotransposon-adenovirus hybrid vector. Hum. Gene Ther. 12, 1417–1428.

Song, M., Boissinot, S., 2007. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. Gene 390, 206–213.

Song, S.U., Gerasimova, T., Kurkulos, M., Boeke, J.D., Corces, V.G., 1994. An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. Genes Dev. 8, 2046–2057.

Sorek, R., Ast, G., Graur, D., 2002. Alu-containing exons are alternatively spliced. Genome Res. 12, 1060–1067.

Stenglein, M.D., Harris, R.S., 2006. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J. Biol. Chem. 281, 16837–16841.

Stoneking, M., Fontius, J.J., Clifford, S.L., Soodyall, H., Arcot, S.S., Saha, N., et al., 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. 7, 1061–1071.

Sugano, T., Kajikawa, M., Okada, N., 2006. Isolation and characterization of retrotransposition-competent LINEs from zebrafish. Gene 365, 74–82.
Suzuki, J., Yamaguchi, K., Kajikawa, M., Ichiyanagi, K., Adachi, N., Koyama, H., et al., 2009. Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. PLoS Genet. 5, e1000461.
Sverdlov, E.D., 2000. Retroviruses and primate evolution. Bioessays 22, 161–171.
Takahashi, H., Fujiwara, H., 2002. Transplantation of target site specificity by swapping the endonuclease domains of two LINEs. EMBO J. 21, 408–417.
Takahashi, H., Okazaki, S., Fujiwara, H., 1997. A new family of site-specific retrotransposons, SART1, is inserted into telomeric repeats of the silkworm, *Bombyx mori*. Nucleic Acids Res. 25, 1578–1584.
Takahashi, K., Terai, Y., Nishida, M., Okada, N., 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Mol. Biol. Evol. 18, 2057–2066.
Tamura, M., Kajikawa, M., Okada, N., 2007. Functional splice sites in a zebrafish LINE and their influence on zebrafish gene expression. Gene 390, 221–231.
Tan, L., Sarkis, P.T., Wang, T., Tian, C., Yu, X.F., 2009. Sole copy of Z2-type human cytidine deaminase APOBEC3H has inhibitory activity against retrotransposons and HIV-1. FASEB J. 23, 279–287.
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.
Thibault, S.T., Singer, M.A., Miyazaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M., et al., 2004. A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. Nat. Genet. 36, 283–287.
Thomson, T., Lin, H., 2009. The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. Annu. Rev. Cell Dev. Biol. 25, 355–376.
Thrasher, A.J., Gaspar, H.B., Baum, C., Modlich, U., Schambach, A., Candotti, F., et al., 2006. Gene therapy: X-SCID transgene leukaemogenicity. Nature 443, E5–E6, discussion E6–E7.
Till, B.J., Burtner, C., Comai, L., Henikoff, S., 2004. Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res. 32, 2632–2641.
Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K., et al., 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459, 442–445.
Toyooka, Y., Tsunekawa, N., Takahashi, Y., Matsui, Y., Satoh, M., Noce, T., 2000. Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. Mech. Dev. 93, 139–149.
Trujillo, M.A., Sakagashira, M., Eberhardt, N.L., 2006. The human growth hormone gene contains a silencer embedded within an Alu repeat in the 3′-flanking region. Mol. Endocrinol. 20, 2559–2575.
Turelli, P., Vianin, S., Trono, D., 2004. The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay. J. Biol. Chem. 279, 43371–43373.
Tuteja, N., Singh, M.B., Misra, M.K., Bhalla, P.L., Tuteja, R., 2001. Molecular mechanisms of DNA damage and repair: progress in plants. Crit. Rev. Biochem. Mol. Biol. 36, 337–397.
van de Lagemaat, L.N., Landry, J.R., Mager, D.L., Medstrand, P., 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet. 19, 530–536.
van de Lagemaat, L.N., Medstrand, P., Mager, D.L., 2006. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. Genome Biol. 7, R86.

Vigdal, T.J., Kaufman, C.D., Izsvak, Z., Voytas, D.F., Ivics, Z., 2002. Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. J. Mol. Biol. 323, 441–452.
Vinckenbosch, N., Dupanloup, I., Kaessmann, H., 2006. Evolutionary fate of retroposed gene copies in the human genome. Proc. Natl. Acad. Sci. USA 103, 3220–3225.
Walisko, O., Schorn, A., Rolfs, F., Devaraj, A., Miskey, C., Izsvák, Z. and Ivics, Z. 2008. Transcriptional activities of the Sleeping Beauty transposon and shielding its genetic cargo with insulators. Mol. Ther. 16, 359–369.
Wallace, N.A., Belancio, V.P., Deininger, P.L., 2008. L1 mobile element expression causes multiple types of toxicity. Gene 419, 75–81.
Wang, W., Kirkness, E.F., 2005. Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Res. 15, 1798–1808.
Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A., et al., 2005. SVA elements: a hominid-specific retroposon family. J. Mol. Biol. 354, 994–1007.
Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., et al., 2008. Chromosomal transposition of PiggyBac in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA 105, 9290–9295.
Wang, J., Saxe, J.P., Tanaka, T., Chuma, S., Lin, H., 2009. Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. Curr. Biol. 19, 640–644.
Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., et al., 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev. 20, 1732–1743.
Watkins, W.S., Rogers, A.R., Ostler, C.T., Wooding, S., Bamshad, M.J., Brassington, A. M., et al., 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res 13, 1607–1618.
Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., et al., 2001. Human L1 retrotransposition: cis preference versus trans complementation. Mol. Cell. Biol. 21, 1429–1439.
Weil, C., Martienssen, R., 2008. Epigenetic interactions between transposons and genes: lessons from plants. Curr. Opin. Genet. Dev. 18, 188–192.
Weiner, A.M., Deininger, P.L., Efstratiadis, A., 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. 55, 631–661.
Wessler, S.R., 2006. Transposable elements and the evolution of eukaryotic genomes. Proc. Natl. Acad. Sci. USA 103, 17600–17601.
Wheelan, S.J., Aizawa, Y., Han, J.S., Boeke, J.D., 2005. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. Genome Res. 15, 1073–1078.
Wilber, A., Frandsen, J.L., Geurts, J.L., Largaespada, D.A., Hackett, P.B., McIvor, R.S., 2006. RNA as a source of transposase for sleeping beauty-mediated gene insertion and expression in somatic cells and tissues. Mol. Ther. 13, 625–630.
Williams, D.A., 2008. Sleeping beauty vector system moves toward human trials in the United States. Mol. Ther. 16, 1515–1516.
Wilson, M.H., Kaminski, J.M., George Jr., A.L., 2005. Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. FEBS Lett. 579, 6205–6209.
Xing, J., Wang, H., Belancio, V.P., Cordaux, R., Deininger, P.L., Batzer, M.A., 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc. Natl. Acad. Sci. USA 103, 17608–17613.
Xing, J., Wang, H., Zhang, Y., Ray, D.A., Tosi, A.J., Disotell, T.R., Batzer, M.A., 2007. A mobile element-based evolutionary history of guenons (tribe Cercopithecini). BMC Biol 5, 5.

Xing, J., Zhang, Y., Han, K., Salem, A.H., Sen, S.K., Huff, C.D., et al., 2009. Mobile elements create structural variation: analysis of a complete human genome. Genome Res. 19, 1516–1526.

Xiong, Y., Eickbush, T.H., 1988. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol. Cell. Biol. 8, 114–123.

Xu, M., You, Y., Hunsicker, P., Hori, T., Small, C., Griswold, M.D., Hecht, N.B., 2008. Mice deficient for a small cluster of Piwi-interacting RNAs implicate Piwi-interacting RNAs in transposon control. Biol Reprod 79, 51–57.

Yae, K., Keng, V.W., Koike, M., Yusa, K., Kouno, M., Uno, Y., et al., 2006. Sleeping beauty transposon-based phenotypic analysis of mice: lack of Arpc3 results in defective trophoblast outgrowth. Mol. Cell. Biol. 26, 6185–6196.

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.

Yang, J., Bogerd, H.P., Peng, S., Wiegand, H., Truant, R., Cullen, B.R., 1999. An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein. Proc. Natl. Acad. Sci. USA 96, 13404–13408.

Yang, N., Zhang, L., Kazazian Jr., H.H., 2005. L1 retrotransposon-mediated stable gene silencing. Nucleic Acids Res. 33, e57.

Yao, S., Osborne, C.S., Bharadwaj, R.R., Pasceri, P., Sukonnik, T., Pannell, D., et al., 2003. Retrovirus silencer blocking by the cHS4 insulator is CTCF independent. Nucleic Acids Res. 31, 5317–5323.

Yao, J., Zhong, J., Lambowitz, A.M., 2005. Gene targeting using randomly inserted group II introns (targetrons) recovered from an *Escherichia coli* gene disruption library. Nucleic Acids Res. 33, 3351–3362.

Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13, 335–340.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296, 79–92.

Zaiss, D.M., 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.

Zayed, H., Izsvak, Z., Walisko, O., Ivics, Z., 2004. Development of hyperactive sleeping beauty transposon vectors by mutational analysis. Mol. Ther. 9, 292–304.

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.

Zhang, Z., Carmichael, G.G., 2001. The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell 106, 465–475.

Zhang, J., Webb, D.M., 2004. Rapid evolution of primate antiviral enzyme APOBEC3G. Hum. Mol. Genet. 13, 1785–1791.

Zhuang, F., Mastroianni, M., White, T.B., Lambowitz, A.M., 2009. Linear group II intron RNAs can retrohome in eukaryotes and may use nonhomologous end-joining for cDNA ligation. Proc. Natl. Acad. Sci. USA 106, 18189–18194.

Zielonka, J., Bravo, I.G., Marino, D., Conrad, E., Perkovic, M., Battenberg, M., et al., 2009. Restriction of equine infectious anemia virus by equine APOBEC3 cytidine deaminases. J. Virol. 83, 7547–7559.

Zwaal, R.R., Broeks, A., van Meurs, J., Groenen, J.T., Plasterk, R.H., 1993. Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. Proc. Natl. Acad. Sci. USA 90, 7431–7435.