4 Ancient Transposable Elements, Processed Pseudogenes, and Endogenous Retroviruses

Adam Pavlicek, PhD and Jerzy Jurka, PhD

CONTENTS

BACKGROUND

The human genome contains a large number of repetitive elements derived from transposable elements (TEs). In addition to active *Alu* and long interspersed element (LINE or L1) interspersed repeats, the human genome comprises a large number of ancient TEs. These include fossil germ-line insertions of DNA transposons, fossil short interspersed elements (SINEs), L2, and L3 LINEs. Processed pseudogenes and human endogenous retroviruses (HERVs) have amplified more recently in evolutionary history and some of them are still well preserved. Copies of some of the recently extinct TEs continue to contribute to genomic rearrangements by homologous recombination. In this chapter, we review ancient SINE and LINE repeats, processed pseudogenes, HERVs, and DNA transposons. We briefly introduce the genomic structure and replication strategy of these elements, their expression competence, and focus on the contribution of these repeats to human diseases. We also discuss some of the TE-derived genes and regulatory elements.

INTRODUCTION

Repetitive elements or repeats are sequences present in multiple copies in the genome but, unlike multigene families, they do not have any clear function in the host. Low-copy repeats generated by large-scale genomic duplications represent a class by themselves, and are discussed in Chapter 5.

From: *Genomic Disorders: The Genomic Basis of Disease*
Edited by: J. R. Lupski and P. Stankiewicz © Humana Press, Totowa, NJ

Table 1
Repetitive Elements in the Human Genome

Type/class		*Superfamily*	*Further division*[a]	*Family*	*Genome (%)*[b]
Tandem repeats	Satellites				Approx 20?
	Telomeric and subtelomeric repeats				<0.01
	Microsatellites and minisatellites				1.4
Interspersed repeats	DNA transposons	Mariner/Tc1		11 families	2.8
		hAT		14 families	2.4
		PiggyBac		1 families	0.02
		MuDr		2 families	0.03
		Harbinger		*HARBI1* gene	—
		P		Single-copy genes	—
	Non-LTR retrotransposons	L1	LINE	L1	17
			SINE	*Alu*	10.5
				SVA (SINE-R)	0.13
			Retropseudogenes	—	0.1–0.3[c]
		CR1	LINE	L2	3.1
				L3	0.31
			SINE	MIR	1
				MIR3	0.29
	LTR retrotransposons	Copia (HERVs)	Class I (gamma retroviruses)	>90 families	2.5
			Class II (beta retroviruses)	11 families	0.5
			Class III (spumaviruses)	>30 families	6
		Gypsy		Single-copy genes	—

This table shows the basic division of human repetitive elements. *See* refs. *3,6,123,128*, and Repbase update *(4)*.
[a]For non-LTR transposons, we also included dependent nonautonomous elements (SINEs, retropseudogenes).
[b]The numbers represent the proportion of detectable repeats in the sequenced genome. Centromeric and heterochromatic satellites are underrepresented in the sequenced regions, but based on reassociation studies the proportion is estimated to be approx 20%.
[c]Depending on the detection method.

In general, repetitive elements are divided into tandem repeats and interspersed repeats (Table 1). Tandem repeats are head-to-tail repetitions of the same sequence motif. Interspersed repeats are active or inactive copies of TEs dispersed throughout the genome. Repetitive elements can be grouped into sets of similar copies, called families or subfamilies. Families of TEs that encode enzymes necessary for their replication are termed autonomous. Nonautonomous elements do not encode all necessary proteins, and their replication (amplification) depends on proteins provided by the autonomous elements. Nonautonomous elements can, thus, be viewed as parasitic elements competing for replication machinery with the autonomous copies. Amplification of both autonomous and nonautonomous elements depends also on additional factors provided by the host cell.

Based on their replication strategy, interspersed repeats are broadly divided into DNA transposons and retrotransposons. The DNA transposons amplify using the host DNA repli-

cation machinery, and their transcripts serve solely as mRNAs participating in translation of transposon-encoded proteins involved in the transposon insertions and excisions. The retrotransposons replicate via an RNA intermediate and, thus, their transcripts serve both as mRNAs for protein translation and as templates for DNA synthesis. Before integration into the genome, retrotransposon RNA must be copied into cDNA using RNA-dependent DNA polymerase, also known as reverse transcriptase (RT) *(1,2)*.

Recognizable copies of all repetitive elements constitute approx 50% of the human genome *(3)*. During the course of the human genome sequencing, more than 600 repeat families and subfamilies have been discovered. All are systematically organized in Repbase Update (RU) (http://www.girinst.org/Repbase_Update.html) *(4)*.

NON-LTR RETROTRANSPOSONS

Non-long terminal repeat (LTR) retrotransposons are the most abundant repetitive elements in human genomic DNA and represent approx one-third of the genome (Table 1). Their DNA copies are co-linear with the RNA transcripts and they lack LTRs present in retroviruses. Autonomous non-LTR retrotransposons are often referred to as LINEs and their nonautonomous counterparts as SINEs. The human genome contains two superfamilies of LINEs: active L1 (LINE1) elements, and extinct families of L2 and L3 elements (Table 1). The latter belong to the CR1 superfamily. Human SINEs are represented by the active *Alu* and SVA repeats retrotransposed by L1 elements, and by the extinct mammalian-wide interspersed repeat (MIR) and MIR3 SINEs that coamplified with L2 and L3 families, respectively. Active L1, *Alu* and SVA non-LTR retrotransposons are described in Chapters 2 and 3. Here, we concentrate on ancient LINE (*see* Ancient LINEs) and SINE (*see* Ancient SINEs) elements. We also review L1-retroposed copies of cellular transcripts known as processed pseudogenes (*see* Processed Pseudogenes).

Ancient LINEs

A typical structure of LINE elements is shown in Fig. 1A. LINEs usually contain two open reading frames referred to as *ORF1* and *ORF2*. The L3 ORF1 protein shares similarity with esterase domains *(5)*. L2 elements apparently lack *ORF1*, although it is possible that the very old age of L2 copies and frequent 5' truncation typical for LINEs prevented reconstruction of a full-length L2 element *(6)*. The ORF2 protein contains the RT and apurinic-apyrimidinic endonuclease enzymatic domains. The transcription of LINE elements starts from a poorly characterized internal promoter for RNA polymerase II (Fig. 1A). After translation, the complex of LINE RNA and protein(s) enters the nucleus, where an endonucleolytic nick at a DNA target serves to prime reverse transcription *(7)*. Target-primed reverse transcription is another feature distinguishing non-LTR from LTR retrotransposons, which normally use cellular tRNA as primers for reverse transcription (*see* Human Endogenous Retroviruses). Unlike in L1 and L1-dependent elements, no target site duplications are created during integration of L2 and L3 elements. Furthermore, L2 and L3 carry microsatellite-like 3' tails instead of the polyA tail found in L1 insertions *(6)*. Based on high sequence diversity, the age of L2 and L3 elements in the human genome is estimated to be approx 200–300 million years, corresponding to the early radiation of reptiles, birds, and mammals. Recognizable L2 and L3 copies, together with their nonautonomous counterparts MIR and MIR3, represent approx 5% of the human genome *(6)*.

The potential of L2 and L3 copies to stimulate genomic rearrangements is very limited owing to their ancient origin, and no such case has been identified in humans so far. The

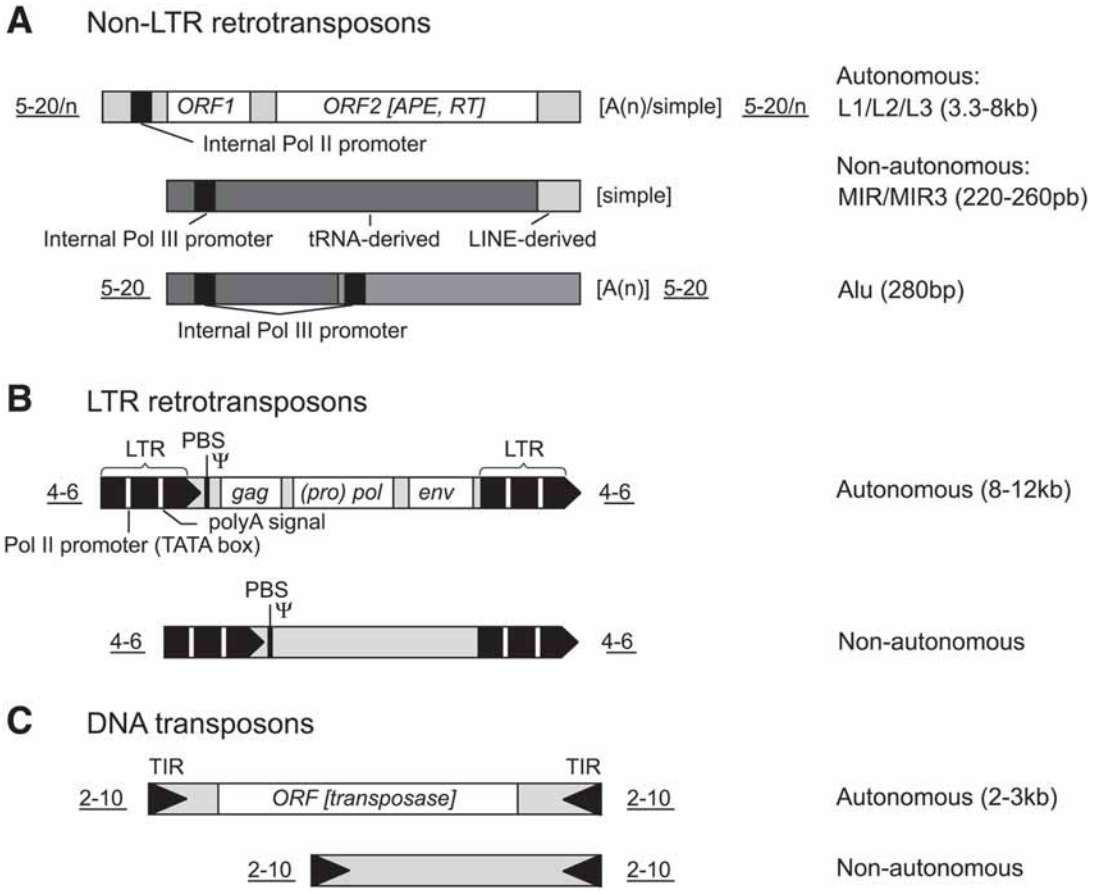

Fig. 1. Structure of human interspersed repeats. The figure shows typical structures of integrated interspersed repeats. (A) Structure of non-long terminal repeats (LTR) retrotransposons. The first bar shows a schematic organization of LINE elements. Human L1 and L3 elements contain two open reading frames (*ORFs*), L2 copies apparently lack *ORF1*. The function of the first *ORF1* is poorly understood. *ORF2* encodes a protein with the apurinic-apyrimidinic endonuclease and reverse transcriptase (RT) enzymatic domains. L1 insertions have a 3' polyA tail and are flanked by variable long target duplications, typically 5- to 20-bp long. L2 and L3 elements lack the target site duplications, and their 3' tails are composed of simple repeat sequences. The second and third schematic bars depict organization of SINEs. L2- and L3-dependent MIR and MIR3 SINEs (middle) consist of two parts. The 5' part (dark gray) is derived from a tRNA gene and harbors an internal pol III promoter. The 3' tails of MIRs are homologous to the 3' end of the corresponding LINE counterparts (light gray). *Alu* elements (bottom) derived from *7SL RNA* genes have a dimeric structure and contain a composite pol III promoter. SINEs share insertional characteristics with their LINE counterparts including the 3' end simple repeats in MIRs. L1-dependent *Alu* repeats contain 3' polyA tails and are flanked by 5- to 20-bp long direct repeats. (B) Structure of LTR retrotransposons. All LTR retrotransposons contain two LTRs, which include a pol II promoter and polyA signal. The internal part of autonomous elements comprises three main open reading frames: *gag*, *pol*, and *env*; in some HERV families *pro* can be separate from *pol*. The 5' part of the internal sequence contains a tRNA primer binding site (PBS) for initiation of reverse transcription and also an encapsidation sequence (Ψ) necessary for incorporation of retroviral RNA molecules into virions. Internal sequences of nonautonomous LTR retrotransposons may or may not share similarity with retroviral ORFs, but they contain all structures required for retroviral replication and reverse transcription such as LTRs, PBS, or the encapsidation signal. Both

proliferation of L2 and L3 elements stopped long before the mammalian radiation, and recombination between highly diverged sequences is very unlikely. No protein-coding gene derived from L2 or L3 elements has been detected.

Ancient SINEs

Successful amplification of SINEs by the LINE machinery requires their transcription to be initiated by internal promoters. Typical SINEs are derived from cellular genes transcribed by RNA polymerase III, containing internal promoters. For example, L1-dependent *Alu* elements are derived from *7SL RNA* genes encoding the RNA scaffold of the signal recognition particle, whereas MIRs are derived from tRNA genes. The 5' parts of MIRs are homologous to tRNA genes, and the 3' terminal portions are homologous to the respective LINE elements on which their proliferation depends (Fig. 1A). MIR shares its 3' terminus with L2 elements, MIR3 with L3 elements *(6,8,9)*. MIR and MIR3 SINEs, like their LINE2 and LINE3 counterparts, are very old and their contribution to genomic rearrangements in the human genome seems to be minimal because no homologous recombination between ancient SINEs associated with human disease has been reported.

Processed Pseudogenes

Eukaryotic genomes contain a large number of pseudogenes *(10)*, which are homologous to known functional genes, but are apparently defective owing to the presence of various mutations, such as truncations, frameshifts, or missense changes. Pseudogenes are sometimes abbreviated using the Greek letter Ψ followed by the symbol of the original functional gene from which the pseudogene was derived. The human genome contains two types of pseudogenes: duplicated pseudogenes and processed pseudogenes generated in germ-line cells. Duplicated pseudogenes resemble normal cellular genes and are created by direct DNA duplications. Processed pseudogenes or retropseudogenes, on the other hand, are structurally distinct. Typically, they are characterized by the lack of both a promoter and introns, and by the acquisition of a polyA-like sequence at their 3' ends *(11,12)*. Because of co-linearity with spliced mRNAs, processed pseudogenes appear to originate by reverse transcription and integration of cellular mRNAs.

The mechanism of amplification of processed pseudogenes represented a long-standing puzzle. The presence of a polyA tail and lack of long terminal repeats indicated involvement of a non-retroviral enzymatic machinery *(12)*. Furthermore, processed pseudogenes share common insertion characteristics, such as the TTAAAA insertion motif shared with LINE-L1 and *Alu* repetitive sequences *(13)*, strongly pointing to active L1 elements as the donors of the RT. Indeed, a series of sophisticated experiments has demonstrated reverse transcription and integration of spliced reporter mRNA by the L1 enzymatic machinery with all hallmarks of processed pseudogenes *(14–18)*. Genomic studies of processed pseudogenes have disclosed additional characteristics of L1-mediated retroposition: integration independent of the chro-

---

the autonomous and nonautonomous elements are flanked by 4- to 6-bp long direct repeats. (C) Structure of DNA transposons. DNA transposons contain terminal inverted repeats (TIRs) of variable sizes (from approx 10 to several hundred nucleotides). Autonomous elements encode the replication enzyme transposase. Nonautonomous elements share TIRs with autonomous copies, which are recognized by transposase. DNA transposons are flanked by 2- to 10-bp long target site duplications, characteristic for each (super)family.

mosomal location of the parent gene *(19,20)*, preferential insertions into GC-poor DNA segments (isochores), dark Giemsa bands *(20–22)*, as well as frequent 5' truncations and inversions *(23–25)*. Therefore, it is generally accepted that human processed pseudogenes are copies of cellular RNA transcripts reverse transcribed and integrated into the genome by L1 machinery. Retroposition of cellular RNAs by other RT-encoding elements such as endogenous retroviruses appears to be rare *(26)*.

The number of processed pseudogenes derived from protein-coding mRNA is estimated to be somewhere between 10,000 and 30,000, depending on detection methods and stringency criteria *(20,27,28)*. Some may be misannoted as functional genes. Moreover, most pseudogenes are 5' truncated and contain only a partial 3' untranslated region and thus are not detectable by standard translated searches *(23)*. The majority of processed pseudogenes preserved in the human genome were created after the split between the rodent and primate lineages. The peak of their generation roughly corresponds to the main period of *Alu* amplification 60–40 million years ago, after which the frequency of retroposition has declined *(19,20,22)*.

The number of processed pseudogenes depends on several properties of mRNA. Short and GC-poor mRNAs tend to produce more pseudogenes *(27)*, although the significance of small size may be artificial *(23)*. The number of pseudogenes is positively correlated with the breadth of expression of the parental gene *(27)* and, as expected, with germline/embryonic cell expression *(20)*, because only germ-line retrotranspositions can be passed onto future generations. The contribution of functional gene groups to the human processed pseudogene population is not random either. Although the number of processed pseudogenes is comparable to the number of human genes, only approx 10% of human genes are represented among processed pseudogene(s), and 30 human genes account for 20% of all human processed pseudogenes *(20)*. Ribosomal protein genes, DNA/RNA binding proteins, receptors, kinases, metabolic enzymes, mitochondrial proteins, and housekeeping genes in general are represented by the highest number of processed pseudogenes *(19,20,22,25,29)*. In summary, it seems that genes expressed in germ-line cells, such as germ-line specific or housekeeping genes, have the highest chance of being retroposed by L1 elements.

In addition to typical transcripts of protein-coding genes, other RNAs can also be retroposed by L1 elements. Processed pseudogenes can be derived from alternatively spliced mRNAs *(24,25,30)*, antisense transcripts *(31)*, or mRNAs derived from other repetitive elements such as endogenous retroviruses *(30,32,33)*. It seems that virtually any RNA including non-coding RNAs can be potentially retroposed by L1 elements. The model of *Alu* and L1 retrotranspositions *(34)* implies two principal requirements for successful retroposition: (1) the RNAs should contain a polyA-like 3' terminal sequence and (2) they should be located close to the newly synthesized L1 proteins (e.g., cytoplasmic localization, close to ribosomes) owing to the L1 *cis* preference. This hints that both pol I and III transcripts may be transposed as well, if they meet the aforementioned conditions. *Alu* elements are a prominent example of polyA-terminated processed pseudogenes derived from a pol III transcribed gene (*7SL RNA*).

Retroposition of processed pseudogenes can serve as a mechanism of gene duplication *(35–37)*. However, unless they acquire a new promoter (most pol II genes have promoters upstream of the transcription start), typical processed pseudogenes cannot be expressed. Expression of recently integrated pseudogenes has been reported in several cases *(38–42)*, but it is not clear whether they encode functional proteins. Nevertheless, there are a few examples of retrogenes (i.e., functional and expressed intronless genes derived from retroposed mRNAs) in the human genome *(35,36)*. For example, retroposition can help to amplify X-linked genes, which have

only one functional copy in somatic cells. X-linked mutations, thus, often result in loss of function and their transfer to autosomes or amplification on X can revert this phenotype. Several such examples of X-to-autosomes or X-to-X amplification of human as well as mouse genes have been reported *(43–54)*. Along the same lines, processed pseudogenes could also provide a new domain (exon) if they insert into introns of other genes. The only well-documented example of a successful exon shuffling in the human genome identified to date is SCAN domain-containing gene 2 (*SCAND2*). This member of the SCAN nuclear protein family gene located on 15q25.2 harbors a C-terminal domain derived from the N-terminal part of *C1orf12/EGLN1* (egl nine homolog 1) located on 1q42.2 *(55)*. Despite the examples of functional retropseudogenes or exon shuffling, however, the vast majority of processed pseudogenes are clearly translationally incompetent (owing to truncations or stop codons), and they appear to follow mutation patterns consistent with neutral evolution *(20,27,28)*. Interestingly, such expressed processed pseudogenes without protein coding capacity can act as functional non-coding RNAs. For example, *Makorin1-p1* is an expressed processed pseudogene regulating the mRNA stability of its progenitor gene *Makorin1*. Disruption in the *Makorin1-p1* pseudogene results in polycystic kidneys and bone deformity in mice, demonstrating its importance for *Makorin1* function *(56)*. Searches for expressed pseudogenes in the human genome indicate that this type of regulation may be more common than previously anticipated *(42)*.

The potential contribution of processed pseudogenes to human diseases is far from being understood. Pseudogene integrations may cause insertional inactivation of genes, but, despite many reports of *Alu* and L1 insertion *(57)*, no processed pseudogene-related insertional mutation has been reported to date. Potential regulatory effects of pseudogene transcripts at the RNA level are intriguing, particularly in the light of early speculations on this subject *(58)*. Finally, it should be pointed out that processed pseudogenes (especially expressed copies) may interfere with the analysis of variation of human genes, because they can be misinterpreted as polymorphism or as mutated alleles of the functional genes.

HUMAN ENDOGENOUS RETROVIRUSES

Retroviruses belong to a broad class of (retro)elements that replicate via an RNA intermediate and include LTRs in their DNA copies. The retroviral life cycle is characterized by reverse transcription of the retroviral RNA genome followed by cDNA integration into the host nuclear DNA, where they can persist in the form of a stable integrated provirus. Retroviral infections of early embryonic and germ-line cells can be inherited by subsequent generations and such ancient proviral relics found in the genome are called ERVs. HERVs resemble well-known exogenous retroviruses and carry typical genes found in infectious retroviruses (reviewed in ref. *59*). The group-specific antigen (*gag*) open reading frame encodes internal structural proteins of the retroviral particle, whereas *pro* and *pol* genes encode enzymes necessary for retroviral amplification (Fig. 1B). Replication-competent retroviruses code these three enzymes: RT *(1,2)*, integrase, which inserts retroviral DNA into host chromosome, and protease, which posttranslationally cleaves Gag, Pol, and Env polyproteins into functional proteins. RT also performs an RNA degradation activity, encoded in a separate domain called RHase H. The last open reading frame *env* encodes surface (envelope) glycoproteins required for attachment of retroviral particles to the cellular receptor and penetration into the cell. The affinity for a particular receptor determines retroviral range of infectivity or tropism for cells and tissues expressing the given receptor. The transmembrane subunit of the Env proteins

contains a so-called immunosuppressive domain, a conserved 17mer with effects on the proliferation and differentiation of lymphocytes *(60,61)*. Interestingly, the immunosuppressive domains of mammalian C-type retroviruses, including some HERV families, are similar to the analogous segments in the envelope glycoprotein of filoviruses (negative-strand RNA viruses) such as Ebola or Marburg viruses *(62)*. HERVs also may contain other auxiliary genes. Several class II and III families code deoxyuridine triphosphatase (dUTPase) (*63,64*; RU). Some HERV-K elements encode a 105 aa functional homolog of the Rev/Rex proteins found in human immunodeficiency virus (HIV)-1 and human T-cell leukemia virus (HTLV), respectively. These proteins mediate nuclear transport of unspliced RNAs via the host factor CRM1 *(65,66)*.

The structure of an integrated DNA provirus is different from the structure of its RNA genome *(59)*. During the process of retroviral reverse transcription two identical copies of part of the retroviral sequence, known as long terminal repeats (LTRs), are generated. LTRs are located at both ends of the provirus and flank the protein-coding internal sequence. LTRs harbor the pol II promoter, enhancers, polyA signal, and other regulatory sequences. Typical HERV LTRs have conserved 5'-TG and 3'-CA termini. During integration, 4- to 6-bp target site duplications (TSDs) of the host DNA are created, which flank the integrated provirus. The length of TSDs is determined by the viral integrase and is characteristic of a given retrovirus or family of HERVs.

Since the discovery of the first human endogenous retrovirus in 1981 *(67)*, more than 400,000 HERV fragments have been found in the human genome, contributing approx 9% of human DNA (Table 1) *(3,6,68)*. HERV sequences were classified using several different systems and the nomenclature used in the literature is rather confusing *(69)*. The most frequently used classification is based on the binding site for the tRNA primer. For example, a recently active family of HERVs uses lysine (K) tRNA molecules to prime reverse transcription and is therefore known as HERVK. Although this classification was sufficient during the early years of HERV research, the system became obsolete after the discovery of many other HERV families in the human genome. Progress in classification based on primer binding sites is hampered by the fact that the same tRNA primer can be shared by unrelated families and further complicated by the existence of chimeric elements composed of segments derived from different families. Such chimeras probably arise by co-packaging of two different retroviral RNA into virion and subsequent template switching between the RNAs during reverse transcription *(70)*. Therefore, a different, more flexible classification system of HERVs is used in the RU database *(4)*. This classification uses a combination of traditional names and numbers. RU families are primarily defined based on a substantial DNA sequence divergence from other retroviral families. This classification together with a list of all known HERV families is extensively reviewed in ref. *6*.

Based on their similarity to exogenous retroviruses, HERV families are grouped into three classes. HERVs with homology to mammalian type C retroviruses (gamma retroviruses), such as murine leukemia virus (MLV), have been placed in class I. Class I represents a highly heterogeneous group of HERVs with many different families *(6)*. This class includes numerous nonautonomous families, some of which constitute a distinct subgroup called MER4. In addition, many class I elements are chimeras composed of segments derived from unrelated retroviruses. All class I families are flanked by 4–5 bp TSDs. Class II consists of HERVs related to mammalian type B and D retroviruses (beta retroviruses) represented by mouse mammary tumor virus. This class is often referred as the HERVK group, because most of the

families use lysine (K) tRNA as the primer for reverse transcription. The HERVK elements are flanked by 6 bp target site duplications. In 1995, a new group (class III) of endogenous retroviruses similar to human foamy viruses (*Spumaviridae*) was discovered *(71)*. HERVL is a prototypic member of this group. With the exception of HERV18 and HERVL66, all other class III families lack *env*-like ORFs *(6)*. In addition to several families with similarity to retroviral proteins, this group contains a large number of nonautonomous elements known as mammalian apparent LTR retrotransposons (MaLRs) and THE1. All class III elements are flanked by 5 bp TSDs. No endogenous counterparts of exogenous lentiviruses such as HIV are known in the human genome. All retroviruses including HERVs belong to the *copia* superfamily of LTR-retrotransposons; no other superfamilies of LTR-retrotransposons have been detected in the human genome. Surprisingly, however, the human genome contains eight cellular genes with significant similarity to gypsy superfamily of LTR retrotransposons, probably derived from very old gypsy elements, whose remnants cannot be otherwise detected in genomic DNA *(6,72–74)*.

Endogenous retroviruses, although still active in mice *(75)*, are nearly extinct in the human genome. However, some class I elements contain potentially functional ORFs (*69,76–80*; RU). Class III elements are found in all placental mammals, and some lineages including simians and mice (but not rats) exhibit sustained activity of these elements *(63,71,81,82)*. In the human genome, on the other hand, class III elements lost their activity long ago and their age ranges from approx 20 to 150 million years *(6)*. Class II embraces the most recent retroviruses found in the human genome (HERVK10), many of which are human-specific *(83,84)*, and the youngest elements appear to be only approximately 1 million years old *(85)*. Only three elements in the human genome seem to harbor all full-length ORFs *(3)*. The most preserved HERVK element potentially capable of reinfections is known to be polymorphic in human populations *(86)*.

The genomic distribution of HERVs is a result of many factors including the integration preference, intensity of recombination, selection, and so on. Although there are no new HERV insertions reported, HERV integration preferences can be approximated from exogenous retroviruses such as HIV and MLV. HIV preferentially integrates into gene- and GC-rich regions *(87,88)*. MLV also targets genes, especially positions around transcription start sites *(89)*. In contrast to these exogenous retroviruses, HERVs are underrepresented in GC-rich regions and especially in genes, probably owing to interference of retroviral polyA signals with gene transcription if they are inserted in the sense orientation relative to transcription *(90)*. Recombination between two direct LTRs of a provirus leave a single copy of the LTR (solo LTR) in the genome *(91)*, and the remaining LTR with the internal sequence is discarded in the form of a circular episomal DNA. Approximately 90% of HERV genomic elements are such solo LTRs *(3)*. Full-length elements are mostly found in regions with low recombination rates, such as AT-rich regions or chromosome Y *(3,30,90,92,93)*. Thus, the HERV distribution is strikingly different from the insertional pattern of exogenous retroviruses and mostly reflects post-insertional processes.

The extent of HERV contribution to human diseases is a subject of considerable debate in the field. On the DNA level, HERV-mediated rearrangements are relatively rare as HERVs are essentially extinct in humans, and no HERV-related insertion is reported in mutational databases. Because of the ancient origin of most HERV copies and the existence of relatively low copy number families, the presence of highly similar, closely spaced HERV copies prone to recombination is infrequent. Indeed, recombination between two HERVs is a sporadic source of genomic instabilities *(6)*. One example involves recurrent deletions of Y-linked azoospermia factor a (*AZFa*) gene by homologous recombination between two HERV15 copies, asso-

ciated with male infertility *(94–97)*. Recombination between solo LTRs has also contributed to allelic variation of the *HLA* locus in the human population *(98,99)*.

Expression of endogenous retroviruses has been implicated in the etiology of various human diseases including cancer and autoimmune disorders (reviewed in refs. *100–103*). We should stress, however, that the potential contribution of HERVs to these diseases is a highly controversial issue *(102,104–108)*, and despite two decades of research on the pathogenic potential of endogenous retroviruses, convincing evidence linking HERVs to diseases is still lacking.

Aside from the negative outcomes of retroviral insertions and recombinations on the host cell, retroviral regulatory elements and proteins could serve as a new source of material for evolutionary experiments. Indeed, there are several examples of HERV-derived promoters, polyA signals and other regulatory signals recruited by the host genome *(6,109–111)*. One famous example represents a HERVE insertion into the promoter of the amylase gene cluster during primate evolution, which stimulated rearrangements of this locus accompanied by the emergence of amylase expression in salivary glands *(112,113)*. Also, genes of protein-coding repetitive elements including HERVs may occasionally be recruited by the host genome as functional cellular genes. Notably, two *env* genes from the HERVW/HERV17 and HERV-FRD/MER50 families seem to serve as functional human genes coding for the syncytin proteins responsible for cell fusion during differentiation of the syncythiotrophoblast in the human placenta *(114–116)*. The extent of such contributions to cellular functions is unclear and the number of unequivocal cases of HERVs benefiting the host is very limited *(6)*.

DNA TRANSPOSONS

Approximately 5% of the human genome is derived from ancient copies of DNA transposons (Table 1) *(6)*. Preserved copies of DNA transposons are flanked by terminal inverted repeats (TIRs) of variable length (Fig. 1C). The main enzyme encoded by autonomous DNA transposons is called transposase. Trasposase has several activities: it specifically binds to TIRs, excises the integrated copy and pastes it to another place in the genome (cut-and-paste process). After insertion, 2- to 10-bp duplications of the target site are created. DNA transposons amplify by co-replicating with the host DNA as they preferentially excise from already replicated DNA and reinsert into nonreplicated segments. This asymmetric replication theoretically increases the number of DNA transposons by a factor of 1.5, though the real efficiency of this process is probably lower.

Human transposons represent ancient genomic fossils hardly recognizable at the DNA level. Consequently, the first evidence disclosing the presence of DNA transposons in the human genome was not obtained until large amounts of human genomic sequence became available *(117)*. To date, six transposon superfamilies have been found in humans: Mariner/Tc1, hAT, piggyBac, MuDR, Harbinger, and P-like elements (Table 1) *(6)*. Without exception, all these families became extinct in the past and no transposition-competent element is preserved in humans.

As in the case of other very old repeats, the pathogenic potential of human DNA transposons is probably very limited. Mariner elements were detected in a large number of genomic duplications including those linked to the Charcot-Marie-Tooth disease type 1A (CMT1A) duplication, and to hereditary neuropathy with liability to pressure palsies (HNPP) deletions *(118,119)*. However, the proposed involvement of the Mariner transposase in generating double stranded breaks stimulating the rearrangements *(118)* seems speculative given the fact that no

translation-competent Mariner copy is found in the human genome. Yet, it should be noted that the human genome contains nearly 30 genes that were derived from DNA transposons and appear to be preserved between mammals and many vertebrates *(6)*. Well known recruited transposases include: RAG1, a part of V(D)J recombinase complex producing functional immunoglobulin and T-cell receptor genes in developing lymphocytes *(120)*, centromere protein CENP-B *(117)*, an the more recently described MER53 *(121)*, Jerky *(122)*, and HARBI1 *(123)*. Interestingly, the majority of known human genes derived from TEs evolved from DNA transposons. This is probably related to the fact that transposases are extremely diverged, as there is practically no detectable similarity between transposases from different superfamilies. Transposases, thus, provide a wide repertoire of sequence variants compared to more conserved retrotransposon proteins *(6)*. Although the function of the transposon-derived genes is mostly unknown, it is likely that they are involved in DNA recognition and rearrangements. Consequently, they may contribute to pathological genomic rearrangements. Notably, it has been shown that RAG1 nicking may stimulate homologous recombination *(124)* and also chromosomal translocation both in vivo and in vitro *(125)*. Finally, human cells seem to be in general permissive to DNA transposons *(126)* and these elements can, thus, be used as vectors for gene therapy *(127)*.

SUMMARY

The human genome has preserved a substantial fraction of ancient DNA transposons, MIR SINEs, L2, and L3 LINEs, that have limited potential to stimulate genomic rearrangements. Some HERVs have amplified more recently and sporadically stimulate genetic instabilities. Several processed pseudogenes, DNA transposons, as well as both ancient and recent LTR-retrotransposons were recruited by the host genome as new genes. Particularly intriguing is the presence of many genes derived from transposases of DNA transposons. Given the endonucleolytic activity of the transposases, it is possible that these recruited genes are involved in DNA processing.

REFERENCES

1. Temin HM, Mizutani S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 1970;226:1211–1213.
2. Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 1970;226: 1209–1211.
3. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
4. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet 2000;16:418–420.
5. Kapitonov VV, Jurka J. The esterase and PHD domains in CR1-like non-LTR retrotransposons. Mol Biol Evol 2003;20:38–46.
6. Kapitonov VV, Pavlicek A, Jurka J. Anthology of Human Repetitive DNA. In: *Encyclopedia of Molecular Cell Biology and Molecular Medicine, vol. 1* (Meyers RA, ed.). Weinheim, Germany: Wiley-VCH 2004; pp 251–306.
7. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993;72:595–605.
8. Smit AF, Riggs AD. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res 1995;23:98–102.
9. Okada N, Hamada M, Ogiwara I, Ohshima K. SINEs and LINEs share common 3' sequences: a review. Gene 1997;205:229–243.

10. Jacq C, Miller JR, Brownlee GG. A pseudogene structure in 5S DNA of Xenopus laevis. Cell 1997;12:109–120.
11. Vanin EF. Processed pseudogenes: characteristics and evolution. Annu Rev Genet 1985;19:253–272.
12. Weiner AM, Deininger PL, Efstratiadis A. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 1986;55:631–661.
13. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA 1997;94:1872–1877.
14. Tchenio T, Segal-Bendirdjian E, Heidmann T. Generation of processed pseudogenes in murine cells. EMBO J 1993;12:1487–1497.
15. Maestre J, Tchenio T, Dhellin O, Heidmann T. mRNA retroposition in human cells: processed pseudogene formation. EMBO J 1995;14:6333–6338.
16. Dhellin O, Maestre J, Heidmann T. Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription. EMBO J 1997;16:6590–6602.
17. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. Nat Genet 2000;24:363–367.
18. Wei W, Gilbert N, Ooi SL, et al. Human L1 retrotransposition: cis preference versus trans complementation. Mol Cell Biol 2001;21:1429–1439.
19. Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. Genome Biol 2003;4:R74.
20. Zhang Z, Harrison PM, Liu Y, Gerstein M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res 2003;13:2541–2558.
21. Pavlicek A, Jabbari K, Paces J, Paces V, Hejnar J, Bernardi G. Similar integration but different stability of Alus and LINEs in the human genome. Gene 2001;276:39–45.
22. Zhang Z, Harrison P, Gerstein M. Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. Genome Res 2002;12:1466–1482.
23. Pavlicek A, Paces J, Zika R, Hejnar J. Length distribution of long interspersed nucleotide elements (LINEs) and processed pseudogenes of human endogenous retroviruses: implications for retrotransposition and pseudogene detection. Gene 2002;300:189–194.
24. Zhang Z, Gerstein M. Identification and characterization of over 100 mitochondrial ribosomal protein pseudogenes in the human genome. Genomics 2003;81:468–480.
25. Strichman-Almashanu LZ, Bustin M, Landsman D. Retroposed copies of the HMG genes: a window to genome dynamics. Genome Res 2003;13:800–812.
26. Jamain S, Girondot M, Leroy P, et al. Transduction of the human gene FAM8A1 by endogenous retrovirus during primate evolution. Genomics 2001;78:38–45.
27. Goncalves I, Duret L, Mouchiroud D. Nature and structure of human genes that generate retropseudogenes. Genome Res 2000;10:672–678.
28. Torrents D, Suyama M, Zdobnov E, Bork P. A genome-wide survey of human pseudogenes. Genome Res 2003;13:2559–2567.
29. Harrison PM, Hegyi H, Balasubramanian S, et al. Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. Genome Res 2002;12:272–280.
30. Pavlicek A, Paces J, ElledерD, Hejnar J. Processed pseudogenes of human endogenous retroviruses generated by LINEs: their integration, stability, and distribution. Genome Res 2002;12:391–399.
31. Ejima Y, Yang L. Trans mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling. Hum Mol Genet 2003;12:1321–1328.
32. Goodchild NL, Freeman JD, Mager DL. Spliced HERV-H endogenous retroviral sequences in human genomic DNA: evidence for amplification via retrotransposition. Virology 1995;206:164–173.
33. Costas J. Characterization of the intragenomic spread of the human endogenous retrovirus family HERV-W. Mol Biol Evol 2002;19:526–533.
34. Boeke JD. LINEs and Alus: the polyA connection. Nat Genet 1997;16:6–7.
35. Brosius J. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 1999;238:115–134.
36. Gentles AJ, Karlin S. Why are human G-protein-coupled receptors predominantly intronless? Trends Genet 1999;15:47–49.
37. Harrison PM, Gerstein M. Studying genomes through the aeons: protein families, pseudogenes and proteome evolution. J Mol Biol 2002;318:1155–1174.

38. McCarrey JR, Kumari M, Aivaliotis MJ, et al. Analysis of the cDNA and encoded protein of the human testis-specific PGK-2 gene. Dev Genet 1996;19:321–332.
39. Fujii GH, Morimoto AM, Berson AE, Bolen JB. Transcriptional analysis of the PTEN/MMAC1 pseudogene, psiPTEN. Oncogene 1999;18:1765–1769.
40. Olsen MA, Schechter LE. Cloning, mRNA localization and evolutionary conservation of a human 5-HT7 receptor pseudogene. Gene 1999;227:63–69.
41. Reyes A, Mezzina M, Gadaleta G. Human mitochondrial transcription factor A (mtTFA): gene structure and characterization of related pseudogenes. Gene 2002;291:223–232.
42. Yano Y, Saito R, Yoshida N, et al. A new role for expressed pseudogenes as ncRNA: regulation of mRNA stability of its homologous coding gene. J Mol Med 2004;82:414–422.
43. McCarrey JR, Thomas K. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature 1987;326:501–505.
44. Ashworth A, Skene B, Swift S, Lovell-Badge R. Zfa is an expressed retroposon derived from an alternative transcript of the Zfx gene. EMBO J 1990:9:1529–1534.
45. Dahl HH, Brown RM, Hutchison WM, Maragos C, Brown GK. A testis-specific form of the human pyruvate dehydrogenase E1 alpha subunit is coded for by an intronless gene on chromosome 4. Genomics 1990;8: 225–232.
46. Mardon G, Luoh SW, Simpson EM, Gill G, Brown LG, Page DC. Mouse Zfx protein is similar to Zfy-2: each contains an acidic activating domain and 13 zinc fingers. Mol Cell Biol 1990;10:681–688.
47. Sargent CA, Young C, Marsh S, Ferguson-Smith MA, Affara NA. The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons. Hum Mol Genet 1994;3:1317–1324.
48. Hendriksen PJ, Hoogerbrugge JW, Baarends WM, et al. Testis-specific expression of a functional retroposon encoding glucose-6-phosphate dehydrogenase in the mouse. Genomics 1997;41:350–359.
49. Sedlacek Z, Munstermann E, Dhorne-Pollet S, et al. Human and mouse XAP-5 and XAP-5-like (X5L) genes: identification of an ancient functional retroposon differentially expressed in testis. Genomics 1999;61: 125–132.
50. Elliott DJ, Venables JP, Newton CS, et al. An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene. Hum Mol Genet 2000;9:2117–2124.
51. Dass B, McMahon KW, Jenkins NA, Gilbert DJ, Copeland NG, MacDonald CC. The gene for a variant form of the polyadenylation protein CstF-64 is on chromosome 19 and is expressed in pachytene spermatocytes in mice. J Biol Chem 2001;276:8044–8050.
52. Wang PJ, Page DC. Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. Hum Mol Genet 2002;11:2341–2346.
53. Emerson JJ, Kaessmann H, Betran E, Long M. Extensive gene traffic on the mammalian X chromosome. Science 2004;303:537–540.
54. Bradley J, Baltus A, Skaletsky H, Royce-Tolland M, Dewar K, Page DC. An X-to-autosome retrogene is required for spermatogenesis in mice. Nat Genet 2004;36:872–876.
55. Dupuy D, Duperat VG, Arveiler B. SCAN domain-containing 2 gene (SCAND2) is a novel nuclear protein derived from the zinc finger family by exon shuffling. Gene 2002;289:1–6.
56. Hirotsune S, Yoshida N, Chen A, et al. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature 2003;423:91–96.
57. Ostertag EM, Kazazian HH Jr. Biology of mammalian L1 retrotransposons. Annu Rev Genet 2001;35: 501–538.
58. Scherrer K. Control of gene expression in animal cells: the cascade regulation hypothesis revisited. Adv Exp Med Biol 1974;44:169–219.
59. Coffin JM, Hughes SH, Varmus HE. *Retroviruses*. Cold Spring Harbor Laboratory Press, New York, NY, 1997.
60. Cianciolo GJ, Copeland TD, Oroszlan S, Snyderman R. Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. Science 1985;230:453–455.
61. Sonigo P, Barker C, Hunter E, Wain-Hobson S. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell 1986;45:375–385.
62. Bukreyev A, Volchkov VE, Blinov VM, Netesov SV. The GP-protein of Marburg virus contains the region similar to the 'immunosuppressive domain' of oncogenic retrovirus P15E proteins. FEBS Lett 1993;323: 183–187.
63. Benit L, Lallemand JB, Casella JF, Philippe H, Heidmann T. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. J Virol 1999;73:3301–3308.

64. Yin H, Medstrand P, Kristofferson A, Dietrich U, Aman P, Blomberg J. Characterization of human MMTV-like (HML) elements similar to a sequence that was highly expressed in a human breast cancer: further definition of the HML-6 group. Virology 1999;256:22–35.
65. Magin C, Lower R, Lower J. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. J Virol 1999;73:9496–9507.
66. Yang J, Bogerd HP, Peng S, Wiegand H, Truant R, Cullen BR. An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein. Proc Natl Acad Sci USA 1999;96:13,404–13,408.
67. Martin MA, Bryan T, Rasheed S, Khan AS. Identification and cloning of endogenous retroviral sequences present in human DNA. Proc Natl Acad Sci USA 1981;78:4892–4896.
68. Paces J, Pavlicek A, Zika R, Kapitonov VV, Jurka J, Paces V. HERVd: the Human Endogenous RetroViruses Database: update. Nucleic Acids Res 2004;32:D50.
69. Wilkinson DA, Mager DL, Leong JC. Endogenous human retroviruses. In: *The Retroviridae* (Levy, J. A., ed.). New York, NY: Plenum Press, 1994; pp 465–535.
70. Zhang J, Temin HM. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science 1993;259:234–238.
71. Cordonnier A, Casella JF, Heidmann T. Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence. J Virol 1995;69:5890–5897.
72. Charlier C, Segers K, Wagenaar D, et al. Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. Genome Res 2001;11:850–862.
73. Ono R, Kobayashi S, Wagatsuma H, et al. A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. Genomics 2001;73:232–237.
74. Lynch C, Tristem M. A co-opted gypsy-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats. Curr Biol 2003;13:1518–1523.
75. Dewannieux M, Dupressoir A, Harper F, Pierron G, Heidmann T. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nat Genet 2004;36:534–539.
76. Cohen M, Powers M, O'Connell C, Kato N. The nucleotide sequence of the env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. Virology 1985;147:449–458.
77. Blond JL, Beseme F, Duret L, et al. Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. J Virol 1999;73:1175–1185.
78. Kjellman C, Sjogren HO, Salford LG, Widegren B. HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. Gene 1999;239:99–107.
79. Lindeskog M, Mager DL, Blomberg J. Isolation of a human endogenous retroviral HERV-H element with an open env reading frame. Virology 1999;258:441–450.
80. Benit L, Dessen P, Heidmann T. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. J Virol 2001;75:11,709–11,719.
81. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520–562.
82. Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 2004;428:493–521.
83. Medstrand P, Mager DL. Human-specific integrations of the HERV-K endogenous retrovirus family. J Virol 1998;72:9782–9787.
84. Barbulescu M, Turner G, Seaman MI, Deinard AS, Kidd KK, Lenz J. Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. Curr Biol 1999;9:861–868.
85. Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N, Meese E. An almost-intact human endogenous retrovirus K on human chromosome 7. Nat Genet 1999;21:257–258.
86. Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J. Insertional polymorphisms of full-length endogenous retroviruses in humans. Curr Biol 2001;11:1531–1535.
87. Elleder D, Pavlicek A, Paces J, Hejnar J. Preferential integration of human immunodeficiency virus type 1 into genes, cytogenetic R bands and GC-rich DNA regions: insight from the human genome sequence. FEBS Lett 2002;517:285–286.
88. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. Cell 2002;110:521–529.

89. Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. Science 2003;300:1749–1751.
90. Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 1999;9:657–663.
91. Mager DL, Goodchild NL. Homologous recombination between the LTRs of a human retrovirus-like element causes a 5-kb deletion in two siblings. Am J Hum Genet 1989;45:848–854.
92. Erlandsson R, Wilson JF, Paabo S. Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. Mol Biol Evol 2000;17:804–812.
93. Medstrand P, van de Lagemaat LN, Mager DL. Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res 2002;12:1483–1495.
94. Blanco P, ShlumukovaM, Sargent CA, Jobling MA, Affara N, Hurles ME. Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. J Med Genet 2000;37:752–758.
95. Sun C, Skaletsky H, Rozen S, et al. Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. Hum Mol Genet 2000;9:2291–2296.
96. Kamp C, Hirschmann P, Voss H, Huellen K, Vogt PH. Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. Hum Mol Genet 2000;9:2563–2572.
97. Kamp C, Huellen K, Fernandes S, et al. High deletion frequency of the complete AZFa sequence in men with Sertoli-cell-only syndrome. Mol Hum Reprod 2001;7:987–994.
98. Kambhu S, Falldorf P, Lee JS. Endogenous retroviral long terminal repeats within the HLA-DQ locus. Proc Natl Acad Sci USA 1990;87:4927–4931.
99. Kulski JK, Gaudieri S, Martin A, Dawkins RL. Coevolution of PERB11 (MIC) and HLA class I genes with HERV-16 and retroelements by extended genomic duplication. J Mol Evol 1999;49:84–97.
100. Nakagawa K, Harrison LC. The potential roles of endogenous retroviruses in autoimmunity. Immunol Rev 1996;152:193–236.
101. Lower R. The pathogenic potential of endogenous retroviruses: facts and fantasies. Trends Microbiol 1999;7:350–356.
102. Nelson PN, Carnegie PR, Martin J, et al. Demystified. Human endogenous retroviruses. Mol Pathol 2003;56:11–18.
103. Perl A. Role of endogenous retroviruses in autoimmune diseases. Rheum Dis Clin North Am 2003;29:123–143.
104. Garson JA, Tuke PW, Giraud P, Paranhos-Baccala G, Perron H. Detection of virion-associated MSRV-RNA in serum of patients with multiple sclerosis. Lancet 1998;351:33.
105. Poser CM. Virion-associated MSRV-RNA in multiple sclerosis. Lancet 1998;351:755.
106. Lower R. Response from Lower. Trends Microbiol 1999;7:431–432.
107. Mager DL. Human endogenous retroviruses and pathogenicity: genomic considerations. Trends Microbiol 1999;7:431.
108. Stoye JP. The pathogenic potential of endogenous retroviruses: a sceptical view. Trends Microbiol 1999;7:430.
109. Sverdlov ED. Retroviruses and primate evolution. Bioessays 2000;22:161–171.
110. Landry JR, Mager DL, Wilhelm BT. Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet 2003;19:640–648.
111. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 2003;19:530–536.
112. Samuelson LC, Wiebauer K, Snow CM, Meisler MH. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. Mol Cell Biol 1990;10:2513–2520.
113. Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. Genes Dev 1992;6:1457–1465.
114. Blond JL, Lavillette D, Cheynet V, et al. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. J Virol 2000;74:3321–3329.
115. Mi S, Lee X, Li X, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 2000;403:785–789.

116. Blaise S, de Parseval N, Benit L, Heidmann T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proc Natl Acad Sci USA 2003;100:13,013–13,018.
117. Smit AF, Riggs AD. Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci USA 1996;93:1443–1448.
118. Reiter LT, Murakami T, Koeuth T, et al. A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. Nat Genet 1996;12:288–297.
119. Reiter LT, Liehr T, Rautenstrauss B, Robertson HM, Lupski JR. Localization of mariner DNA transposons in the human genome by PRINS. Genome Res 1999;9:839–843.
120. Kapitonov VV, Jurka J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol 2005;3:e181.
121. Kapitonov VV, Jurka J. MER53, a non-autonomous DNA transposon associated with a variety of functionally related defense genes in the human genome. DNA Seq 1998;8:277–288.
122. Jurka J, Kapitonov VV. Sectorial mutagenesis by transposable elements. Genetica 1999;107:239–248.
123. Kapitonov VV, Jurka J. Harbinger transposons and an ancient HARBI1 gene derived from a transposase. DNA Cell Biol 2004;23:311–324.
124. Lee GS, Neiditch MB, Salus SS, Roth DB. RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. Cell 2004;117:171–184.
125. Raghavan SC, Swanson PC, Wu X, Hsieh CL, Lieber MR. A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. Nature 2004;428:88–93.
126. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell 1997;91:501–510.
127. Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z, Kay MA. Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system. Nat Genet 2000;25:35–41.
128. Jurka J. Repeats in genomic DNA: mining and meaning. Curr Opin Struct Biol 1998;8:333–337.