



HAL
open science

Impact of transposable elements on the human genome

Jean-Marc Deragon, Pierre Capy

► **To cite this version:**

Jean-Marc Deragon, Pierre Capy. Impact of transposable elements on the human genome. *Annals of Medicine*, 2009, 32 (4), pp.264-273. 10.3109/07853890009011771 . hal-04821526

HAL Id: hal-04821526

<https://hal.science/hal-04821526v1>

Submitted on 18 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Impact of transposable elements on the human genome

Jean Marc Deragon & Pierre Capy

To cite this article: Jean Marc Deragon & Pierre Capy (2000) Impact of transposable elements on the human genome, *Annals of Medicine*, 32:4, 264-273, DOI: [10.3109/07853890009011771](https://doi.org/10.3109/07853890009011771)

To link to this article: <https://doi.org/10.3109/07853890009011771>



Published online: 08 Jul 2009.



Submit your article to this journal [↗](#)



Article views: 520



View related articles [↗](#)

Impact of transposable elements on the human genome

Jean Marc Deragon¹ and Pierre Capy²

Presence of transposable elements (TEs) in the human genome has profound effects on genome function, structure and evolution. TE mobility and inter-TE recombination are the origin of a large spectrum of mutations and genome reorganization leading to diseases. From the data provided by the Human Genome Project and from information on the detection and dynamics of TEs within and between species acquired during the last two decades, we now know that these elements are not only involved in mutagenesis but can also participate in many cellular functions including recombination, gene regulation, protein-coding RNA messages and, possibly, cellular stress response and centromere function. TEs also promote a general genome shuffling process that has been important for the evolution of several gene families and for the development of new regulatory pathways.

Key words: diseases; human genome; transposable elements.

Ann Med 2000; 32: 264–273.

Introduction

Transposable elements (TEs) are defined as DNA sequences able to move from one genomic position to another in a replicative or nonreplicative process. In denaturation–reassociation experiments they represent a large part of the middle repetitive fraction of the genome (1). Most TEs encode protein(s) required for their mobility and use the host cellular machinery for their transcription and translation. TEs are present at different levels in all genomes from prokaryotes to vertebrates. For instance, in the yeast *Saccharomyces cerevisiae* Ty elements account for 3.1% of the total genome (2), while in the maize *Zea mays* this proportion is more than 50% (3). In humans, the latest estimates suggest that more than 40% of the

genome consists of TE sequences or is directly derived from them (4).

The impact of TEs on the host genome is not neutral. At the end of the 1970s and the beginning of the 1980s, they were generally considered as ‘selfish genes’, suggesting that these elements ‘spread by forming additional copies of themselves within the host genome and did not contribute to the phenotype’ (5). Later, it was assumed that ‘selfish genes may not have effects on phenotypes, but many of them do’ (6). Since the end of the 1980s this vision has changed slightly. Today TEs are considered as potentially useful for the host genome. More precisely, while the mobility of TEs can be highly deleterious for the host, TEs can also be beneficial generating new genetic structures and regulation pathways that can favour a better adaptation to environmental conditions.

In this respect, it is also interesting to mention that recent research has pointed out the effect of environmental stress on the activity and transposition of some TEs. For example, in bacteria stress can induce the activity of several insertion sequences (7), and in plants the *Tnt1* elements can be activated by various types of biotic and abiotic stress (8). Thus, as McClintock wrote in 1984 (9): ‘These elements are probably involved in the genome response to environ-

From the ¹Biomove, Centre National de la Recherche Scientifique - UMR 6547, University Blaise-Pascal Aubière; and the ²Laboratory of Populations, Genetics and Evolution, CNRS - UPR Gif-sur-Yvette, France.

Correspondence: Pierre Capy, MD, Laboratoire Populations, Génétique et Evolution, CNRS - UPR 9034, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France. E-mail: capy@pge.cnrs-gif.fr, Fax: +33 1 69070421.

mental challenges'. According to this view, although the mobility of TEs is a risk for an individual as they can be deleterious, at the population level they can be useful. In other words, they can be seen as 'useful parasites'.

Different classes of TEs in the human genome

In the human genome, two classes of elements are found (Fig 1): class I elements that use an RNA intermediate and a reverse transcriptase, and class II elements that use a DNA intermediate and a transposase. The most abundant class I elements are those belonging to the long interspersed elements (LINEs) and the short interspersed elements (SINEs). The former encode a reverse transcriptase while the latter probably use the proteins produced by the former. The abundance of these elements is relatively high as more than 10 000 copies of LINEs and 500 000 copies of SINEs, in particular *Alu* elements, are found in each genome (10). Class II elements are less abundant, and most of them seem to be members of the *mariner-Tc1* superfamily. However, it has been estimated that

more than 100 000 degenerate copies of short DNA elements (180–1200 bp) with short inverted terminal repeats (14–25 bp) and with 2–8-bp target sites are present in the human genome (11). Most of these were considered as fossils. For a recent and detailed description of human TEs see (4).

Retroviruses are repeated sequences closely related to transposable elements. Evolutionary relationships between long terminal repeat (LTR) retrotransposons and retroviruses have been suspected since the end of the 1970s (12), and more recently these relationships have been clearly included in evolutionary models of transposable elements (13, 14). Several endogenous retroviruses have been described in the human genome (15). One of the main differences between these elements and the LTR retrotransposons is the presence of an envelope gene in retroviruses. In the present review our attention will focus mainly on transposable elements rather than on retroviruses.

Transposable elements as mutagenic agents

TEs can generate mutations either by *de novo* insertions or by DNA rearrangements. Mutations

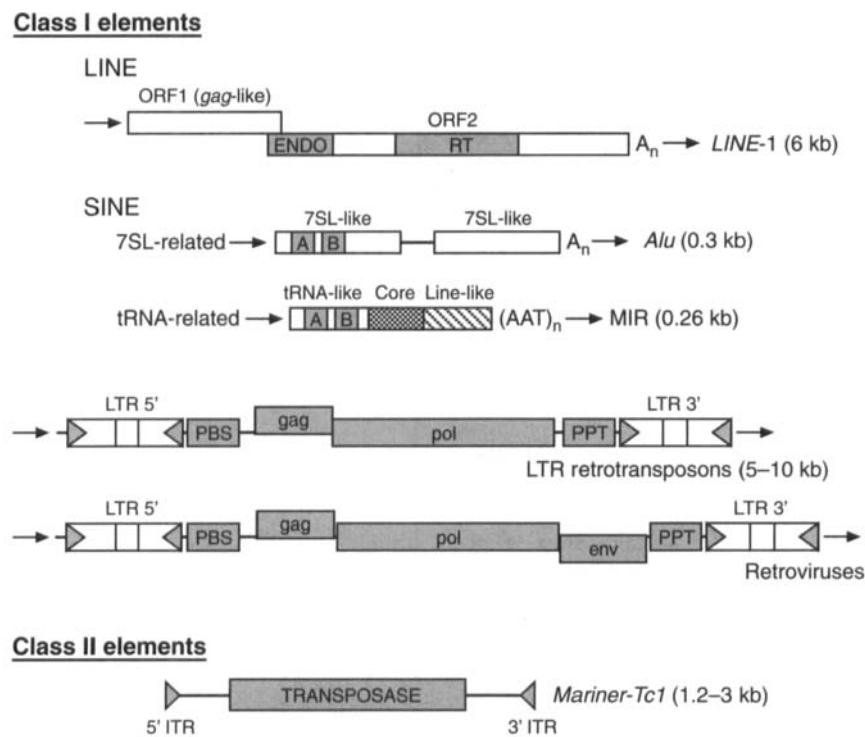


Figure 1. Structure of the main classes of transposable elements found in the human genome. A and B, RNA polymerase III promoters; Core, conserved region of unknown function; ENDO, endonuclease; gag, capsid gene; ITR, inverted terminal repeat; LINE, long interspersed element; LTR, long terminal repeat; pol, polyprotein including reverse transcriptase, integrase, ribonuclease H and protease domains; PBS, primer binding site; PPT, polypurine tract; RT, reverse transcriptase; SINE, short interspersed element; transposase, protein required for the excision and insertion of class II elements.

occurring in germ cells may cause hereditary disorders while somatic mutagenic events can be involved in oncogenesis. Although the frequency of mutations that can be attributed to TEs has not been determined empirically, it can be estimated on the basis of available sequence data. Nearly 1 in 600 mutations in humans are estimated to arise directly from the integration of TEs (16), and about 0.1% of known genetic disorders can be explained by this mechanism (10). Table 1 lists diseases that can be explained by TE insertions. Mutations resulting from homologous recombinations between TE repeats are more frequent, and it is estimated that another 0.3% of known human genetic diseases can be explained by this mechanism (10). Table 2 lists diseases resulting from TE-mediated recombinations. It is likely that these frequencies are underestimates of the true contribution of TEs to diseases, as many mutation detection strategies are designed to identify point mutations and may overlook insertions. Also, many TE insertions and/or TE-induced recombinations may be lethal during embryogenesis.

De novo insertion of TEs can disturb gene ex-

pression in many ways. The major impact comes from the insertion of TEs in introns that disrupt splicing, causing exon skipping and intron sliding (20). TEs can also alter gene expression by integrating in the promoter region or can directly disrupt reading frames by integrating in exons. An unequal homologous recombination between two TEs located in different introns of a gene can lead to deletions and duplications of gene sequences. Similar effects can result if TE sequences recombine with non-TE sequences in nonhomologous recombination events. TE can also be implicated in ectopic recombination events generating large chromosomal rearrangements including chromosome translocation (see Table 2).

Two other mutagenic mechanisms linked to TE insertions are emerging from recent experimental results. The first is the capacity for SINE retroposons to be used as nucleation centres for *de novo* methylation (36–38). During or after integration, SINEs are directly targeted by methylases, and this methylation can be propagated into flanking genomic sequences creating distal epigenetic modifications. Gene regulation can therefore be modified by these

Table 1. Diseases generated by the presence of transposable elements (TEs) in gene coding regions.

Disease	Gene affected	TE implicated	Reference
Acholinesterasaemia	Cholinesterase	Alu	(10*)
Acute intermittent porphyria	PBGD	Alu	(17)
Alport's syndrome	COL4A3	Alu	(18*)
Apert's syndrome†	FGFR2	Alu	(10*)
Branchio-oto-renal syndrome	EYA1	Alu	(10*)
Breast cancer	c-myc	LINE-1	(18*)
Breast cancer	BRCA2	Alu	(18*)
B-cell lymphoma	Mlvi-2	Alu	(18*)
Duchenne muscular dystrophy†	Dystrophin	LINE-1	(18*)
Fukuyama-type congenital muscular dystrophy	FCMD	SINE-R.C2	(19)
Glycerol kinase deficiency	GK	Alu	(10*)
Heart disease (risk factor)	ACE	Alu	(10*)
Haemophilia A†	Factor VIII	LINE-1	(18*)
Haemophilia B	Factor XI	Alu	(18*)
Hereditary desmoid disease†	APC	Alu	(10*)
	APC	LINE-1	(18*)
Hypocalcaemic hypercalcaemia and neonatal severe hyperparathyroidism	CaR	Alu	(10*)
Hereditary angioneurotic oedema type I	C1 inhibitor	Alu	(20*)
Neurofibromatosis type I	NF1	Alu	(18*)
Gyrate atrophy of the choroid and retina	Ornithine δ -aminotransferase	Alu	(18*)
Mucopolysaccharidosis type VII	Beta-glucuronidase	Alu	(21)
Ovarian carcinoma	PROGINS	Alu	(10*)
Retinitis pigmentosa	RET3C11	Alu	(22)
Rheumatoid arthritis (susceptibility to)	HLA-DQB1	HERV-K	(23)
X-linked immunodeficiency (XSCID)	IL2RG	Alu	(10*)
X-linked agammaglobulinaemia	Btk	Alu	(10*)

*Including references therein

†More than one independent event

HERV-K, human endogenous retrovirus; LINE, long interspersed element; SINE, short interspersed element.

Table 2. Diseases resulting from transposable element (TE)-mediated recombination.

Disease	Locus	TE implicated	Consequence	Reference
Acute myeloid leukaemia†	ALL-1	Alu	Duplication	(10*–20*)
Adenosine deaminase (-) severe combined immunodeficiency	Adenosine deaminase	Alu	Deletion	(20*)
Alports's syndrome–diffuse leiomyomatosis†	Collagen type IV (COL4A5-6)	LINE-1	Deletion	(24)
Breast cancer	BRCA1	Alu	Deletion	(10*–25)
C3 deficiency	C3	Alu	Deletion	(10*)
Charcot-Marie Tooth disease type 1A	CMT1A	Mariner	Deletion/duplication	(26)
Chronic granulomatous disease	CYBB	LINE-1	Deletion	(27)
Chronic myeloid leukaemia	BCR/ABL	Alu	Translocation	(28)
Duchenne muscular dystrophy	Dystrophin	Alu	Duplication	(20*)
Ehler–Danlos syndrome	Lysyl hydroxylase	Alu	Duplication	(20*)
Ewing sarcoma	EWS/FLI1	Alu	Translocation	(29)
Fabry's disease	α -galactosidase A	Alu	Deletion	(20*)
Familial hypercholesterolaemia	Low-density lipoprotein receptor	Alu	Duplication	(20*)
Familial hypercholesterolaemia†	Low-density lipoprotein receptor	Alu	Deletion	(10*–20*)
Fanconi's anaemia	FAA	Alu	Deletion	(30)
Glanzmann's thrombasthenia	Platelet fibrinogen receptor	Alu	Inversion/deletion	(10*)
Glanzmann's thrombasthenia	Glycoprotein IIIa (beta3)	Alu	Deletion	(30)
Glioma brain tumours (association to)	Retinoblastoma gene (RB)	Alu	Deletion	(32)
Glycogen storage disease type II (GSDII)	Lysosomal acid alpha-glucosidase	Alu	Deletion	(33)
Glycogen storage disease type II	Phosphorylase kinase (PHKB)	LINE-1	Deletion	(34)
Hereditary angioneurotic oedema type I†	C1 inhibitor	Alu	10*–20*	
Hereditary angioneurotic oedema type I	C1 inhibitor	Alu	Duplication	(20*)
Hereditary nonpolyposis colorectal cancer	MLH1	Alu	Deletion	(10*)
Hereditary persistence of fetal haemoglobin	β -globin cluster	Alu	Deletion	(20*)
Hypo- β -lipoproteinaemia	Apolipoprotein B	Alu	Deletion	(10*)
Insulin-independent diabetes mellitus	Ins Rec β	Alu	Deletion	(20*)
Krabbe's disease	GALC	Alu	Deletion	(35)
Lesch–Nyhan syndrome	Hypoxanthine phosphoribosyl transferase	Alu	Duplication	(20*)
Lipoprotein lipase deficiency	Lipoprotein lipase	Alu	Duplication	(20*)
Maple syrup urine disease	Branched chain acyltransferase	Alu	Deletion	(20*)
Meningioma	c-sis proto-oncogene	Alu	Deletion	(20*)
Mucopolysaccharidosis type IVA	GALNS	Alu	Deletion	(10*)
Sandhoff's disease	β -hexosaminidase, β -subunit	Alu	Deletion	(20*)
Tay–Sach's disease	β -hexosaminidase, α -subunit	Alu	Deletion	(20*)
Thrombophilia	Antithrombin	Alu	Deletion	(20*)
$\gamma\delta\beta$ -thalassaemia	β -globin cluster	Alu	Deletion	(20*)
$\beta\delta$ -thalassaemia	β -globin cluster	Alu	Deletion	(10*–20*)
α -thalassaemia†	α -globin cluster	Alu	Deletion	(10*–20*)
von Willebrand's disease type II	von Willebrand factor	Alu	Deletion	(10*)
XX male	XY chromosome	Alu	Translocation	(10*)

*Including references therein

†More than one independent event

LINE, long interspersed element.

epigenetic modifications. The other mechanism involves the possibility that SINE insertions can modify chromatin organization and gene expression domains. Human SINEs have been shown to associate preferentially with matrix attachment regions (MARs) (39, 40). MARs are genomic DNA sequences with an affinity for a set of insoluble proteins constituting the nuclear skeleton (matrix). According to a current model, the final compaction accommodating the linear DNA molecule inside the nucleus is achieved by the folding of the DNA–histone solenoids into loops of various sizes. The base of such putative structural loops is attached via MARs to the nuclear matrix.

Thus, MARs are genomic elements that can potentially delimit individual structural (gene-expressing) units of chromatin. Changes in MAR activity following SINE integration, in either positive or negative direction, might be of significance for the structure and/or function of the genome, and a single event could modify the expression of a large number of genes. Recently, a plant SINE element, which presents many similarities to human SINEs, was shown to integrate into MARs and to modify MAR activity *in vitro* (AP Tikhonov, C Tatout, L Lavie, JL Bennetzen, Z Avramova and JM Deragon, unpublished observations, 2000). It remains to be determined whether

human SINEs can also modify MAR activity upon integration.

Impact of TE RNA or TE proteins on cell biology

The purpose of TE RNA and proteins is to promote transposition. However, in several cases TE RNA or TE proteins have been implicated in physiological functions independent of the transposition process. One example is the putative role of *Alu* RNA on cell physiology (reviewed in (42)). Viral infection, inhibition of translational elongation and heat shock all increase *Alu* RNA abundance. As *Alu* RNA is a potent inhibitor of the double-stranded-RNA-dependent kinase (PKR), which is known to regulate translation (through the α -subunit of the eukaryotic initiation factor 2 (eIF2 α) phosphorylation), transcription (through nuclear factor κ B (NF κ B) phosphorylation) and tumour necrosis factor (TNF)- α -induced apoptosis (42), the over-expression of *Alu* RNA could have a profound effect on cell physiology. A second example of a physiological impact of TE RNA is the implication of a small (1.7-kb) *LINE-1* transcript in human cell proliferation (43). In agreement with a role for *LINE-1* in cell proliferation, an augmentation of *LINE-1* RNA and/or of the *LINE-1* p40 protein has

been observed in several cases of breast and urothelial carcinomas (44, 45). Human endogenous retrovirus products have also been implicated as mediators of normal biological processes, such as cellular differentiation or cell fusion, and in the development of diseases, including autoimmune disease and tumour development (reviewed in (46)). Concerning the *mariner* DNA element, transcripts are abundant in sclerotic epileptic hippocampi, while no *mariner*-specific RNA can be detected in nonsclerotic hippocampi (47). It is suggested that *mariner* activity could establish an epileptogenic substrate in the hippocampus.

Regulatory regions provided by TE sequences

Promoter regions of eukaryotic genes are organized in short protein-binding DNA motifs that direct the assembly of multiprotein complexes directing transcription. TEs are known to contain high-affinity binding sites for transcription factors. These binding sites are necessary for the transcription of the element itself but can also influence the transcriptional regulation of nearby genes. Therefore, following *de novo* integrations, TEs can create new regulatory functions that can be domesticated by the host. Examples of domestication are presented in Table 3. The *Alu* element is particularly rich in transcription factor

Table 3. Transposable elements (TEs) implicated in gene regulatory functions.

Locus	TE implicated	Function	Reference
Apolipoprotein B RNA editing deaminase gene	Alu	Promoter	(47)
BRCA-1	Alu	Carries an oestrogen receptor-dependent enhancer	(48*)
Calbindin D28K	RTVL-H	Promoter	(49*)
CDC4L	RTVL-H	Promoter	(49*)
CD8	Alu	Polyadenylation signal	(49*)
cDNA 7, cDNA γ	THE-1	Polyadenylation signal	(49*)
cH-6	RTVL-H	Polyadenylation signal	(49*)
cH-7	RTVL-H	Polyadenylation signal	(49*)
cPB-3	RTVL-H	Polyadenylation signal	(49*)
Gamma chain of Fc and T-cell receptors	Alu	Carries + and - promoter response elements	(48*)
β -globin locus	ERV-9	Carries enhancer linked to locus control region	(50)
HHLA2 (immunoglobuline-like)	HERV-H	Polyadenylation signal	(51)
HHLA3	HERV-H	Polyadenylation signal	(51)
Inosine monophosphate dehydrogenase type 1 53	Alu	Promoter	(52)
Keratin k18	Alu	Carries retinoic acid receptor binding sites	(48*)
Myeloperoxidase	Alu	Carries retinoic acid and thyroid hormone receptor binding site	(48*)
Phospholipase A2-like (PLA2L)	HERV-H	Promoter	(49*)
PLT	RTVL-H	Polyadenylation signal	(49*)
Proliferating-cell nucleolar protein gene	Alu	Carries promoter response elements	(53)
Salivary amylase	HERV-E	Promoter	(49*)
Wilms' tumour gene	Alu	Acts as silencer	(48*)

*Including references therein

ERV-1, HERV-H and RTVL-H are endogenous retroviruses.

THE-1, transposon-like human element.

binding sites (55, 56) and is used as a promoter element of many genes, including genes responding to retinoic acid or hormones (56) (Table 3). Human endogenous retroviruses (HERV, RTVL) also possess strong transcriptional signals that carry promoter functions for genes. In addition, endogenous retroviruses provide the polyadenylation signal for many genes (Table 3). Thus, TEs function by disseminating regulatory cassettes throughout the genome. These cassettes are raw materials that were used to select new regulatory pathways during evolution.

TE sequences as part of mRNA

Insertion of TEs in gene coding regions may contribute to protein evolution if the new combinations generated are either advantageous or neutral. In a large number of cases, pieces of TEs found in exons are translated in functional proteins (see Table 4). Splicing-

Table 4. Recruitment of transposable elements (TEs) into protein-coding RNA messages.

Protein	TE in coding region	Reference
A4 amyloid peptide	Alu	(20*)
AD7C-NTP p41	Alu	(56)
Anion exchanger (AE1)	MIR	(57*)
Arginine vasopressin receptor type II	MIR	(57*)
Ataxia-telangiectasia-predicted protein (ATM)	MIR	(58)
Biliary glycoprotein	Alu	(20*)
Cathepsin B	Alu	(59)
Colony-stimulating factor 1 receptor (CSF1R)	MIR	(57*)
Complement C5	Alu	(20*)
c-rel phosphoprotein	Alu	(20*)
Cytokine receptor family (CRF2)	Alu	(60)
Decay accelerating factor	Alu	(20*)
Dopamine- β -hydroxylase	MIR	(57*)
HLA-DR- β 1	Alu	(20*)
Hydroxyindole-O-methyltransferase (HIOMT)	LINE-1	(61)
Immunoglobuline	THE I	(62)
Insuline-like growth factor I (IGF-1)	MIR	(57*)
Integrine β 1	Alu	(20*)
Lectine-like membrane protein	Alu	(20*)
Lens fibre cell membrane protein (mp19)	MIR	(57*)
Monoamine oxidase B (MAOB)	MIR	(57*)
Properdin	MIR	(57*)
RNA-specific editase 1 (hRED1)	Alu	(63)
Serine/threonine kinase	Alu	(20*)
Transaldolase	TARE	(64)
Urokinase-type plasminogen receptor	MIR	(57*)
Vasopressin receptor V2	MIR	(57*)
Zinc finger protein ZNF2	MIR	(57*)

*Including references therein

LINE, long interspersed element; MIR, mammalian interspersed repeat; TARE, transaldolase-associated repetitive element; THE-1, transposon-like human element.

mediated insertion of intronic elements is the major mechanism by which TE segments are introduced into mRNA (66). TE sequences are also raw genetic material from which new gene coding combinations are derived. TE sequences are also present in 3'- and 5'-untranslated regions (UTRs) of mRNAs. For example, it is estimated that 5% of fully spliced human mRNAs contain *Alu* sequences (67). The presence of TE elements in UTRs of human transcripts may influence their regulation, as these regions are known to be involved in the control of RNA stability and translation. An example of this type of regulation has been found in rodents in which B1 and ID SINE retroposons, when present in 3'-UTRs, are a part of the regulatory elements that ensure a coordinated expression of a set of mRNAs (68).

Role of TE in gene duplication events

The evolution of certain gene families can be explained by TE-mediated duplications or deletions. The glycophorin gene family is composed of three paralogous copies. A three-step scenario was proposed to explain its expansion in the human genome. The first step, a duplication, is followed by an unequal homologous *Alu-Alu* recombination followed by a second duplication (20). *Alu* elements were also involved in the duplication of the α 1-acid glycoprotein gene (69) and in the duplication of the *opsin* gene (70). *Alu*, *LINE-1* and endogenous retroviruses also participated in the evolution and diversity of duplicated *HLA-B* and *HLA-C* genes (71). Several duplicated genes located in the class I region of the major histocompatibility complex (MHC) and potentially implicated in susceptibility to several diseases also appear to have been generated through retroelement recombination (72, 73).

Exon shuffling: a byproduct of TE transposition and possible general mechanism for the evolution of new genes

SINEs and LINEs use reverse transcriptase in their replication cycle. While the 5' end of the template RNA used in that process is easily defined by the transcription initiation site of the element, its 3' end will be determined by an external terminator (SINEs) or by a RNA maturation event directed by a polyadenylation signal (LINEs). As the terminator is provided by the genomic site, the length of SINE RNA will systematically include a variable amount of sequences copied from the integration locus. As for the *LINE-1* element, the polyadenylation signal has been shown to be very weak, and sequences derived from the 3' flanks can also be included in the *LINE-1*

RNA. The possibility to reverse transcribe and integrate not only retrotransposon sequences but also variable amount of sequences from the integration site may represent a general mechanism for the evolution of new genetic combinations. Recently, a *LINE-1* element was shown to be capable of transposing a gene placed in its 3' flanking regions (74). This process, called exon shuffling, could have contributed to the evolution of new genes. It remains to be determined what the frequency of these events is, but taking into account the high number of SINEs and LINEs in the human genome, a fairly low probability of exon shuffling could still have a significant impact on gene evolution.

TE and centromere function

The *LINE-1* element has been proposed to play a role in the evolution of centromere function (75). The DNA sequences necessary and sufficient to form an active human centromere are not known, but α -satellite repeats are considered to play an essential role in that function (76). The amount of alphoid sequences between homologues can be highly variable as, for example, two chromosome 21 homologues were shown to possess 420 kb and 5.6 Mb of alphoid sequences at their centromere, respectively (77). The length of alphoid sequences in the centromere of one chromosome can increase following unequal cross-overs with its homologue or after a massive *de novo* amplification process (ie 'saltatory amplification'). As a consequence, the human centromeric regions could increase indefinitely in size without a mechanism to limit their expansion. The *LINE-1* retrotransposon is known to insert frequently in alphoid satellite sequences (78).

A recent study of the centromeric and pericentromeric regions of chromosome 5 (79) has shown that many *LINE-1* elements were present within the α -satellite at the boundary between alphoid and nonalphoid sequences on both sides of the centromere. However, *LINE-1* elements were excluded from the central region of the centromere that would be composed of uninterrupted α -repeats. Surprisingly, pericentromeric *LINE-1* are highly polymorphic in human populations (presence/absence polymorphism), and several independent *LINE-1* insertions can be defined as all present/all absent haplotypes (74, 79). These results suggest that intrachromosomal recombination involving pericentromeric *LINE-1* elements on each side of the centromere is used as a mechanism to loop out alphoid sequences (uninterrupted and *LINE-1* containing alpha repeats) in compensation to α -repeats expansion processes. In that case, *LINE-1* would have been domesticated to limit human centromere obesity. However, *LINE-1*

could also have a damaging effect as its accumulation in centromeric regions could limit the length of uninterrupted α -repeats needed for correct functioning of centromeres. *LINE-1* retrotransposons would therefore have both potential damaging and repairing effects on human centromeres.

One characteristic of the evolution of centromeric repeats is that, like most tandemly repeated DNA families, they undergo repeated rounds of recombination (mainly intrachromosomally) leading to homogenization of repeats within and between chromosomes (ie concerted evolution). The centromere protein B (CENP-B) is a highly conserved protein that can bind to a degenerate motif (ie the CENP-B box) found in most centromeric satellite sequences. *CENP-B* gene probably evolved from a transposase gene as the protein is homologous to the transposase of the *Tigger* transposon and as the CENP-B box is homologous to the inverted repeats of the same transposon (80). Based on these similarities, CENP-B, by its putative nicking activity, could promote homologous recombination reactions that lead to the homogenization of satellite repeats. Therefore centromere evolution has been highly influenced by a protein that most likely evolved from a transposable element-coding region.

TEs as new generation molecular markers

The presence or absence of retroelements (ie SINEs, LINEs, retrotransposons) at different loci has been used recently to evaluate the ancestral relationships of host species. In contrast to classical phylogenetic tools, where many characters are needed to construct a significant phylogenetic tree, a few consistent retroelement insertions are sufficient to infer filiations (reviewed in (81, 82)). Retroelement insertional variability can be weighted very highly for several reasons. Retroelements have very low insertion site specificity (83, 84) and can insert essentially anywhere in the genome so that the probability of insertions at identical positions in two distinct lineages (ie convergent insertions) is highly unlikely. Retroelement insertions are very stable despite their repetitive nature. Unlike other markers where the ancestral state is ambiguous, the ancestral state of a retroelement insertion is known and always represents the absence of the element at a given locus. In humans, polymorphic insertions of the SINE *Alu* have been used to support the African origin of man (85) and the Asian origin of native American populations (86). *Alu* insertions (for a recent example see (87)) as well as endogenous retroviral sequences (88) are also useful in the evaluation of genetic relations between man and higher primates. The *LINE-1* element has also been used recently to study rodent evolution (89) and is

now being developed to study also human population dynamics. Finally, *Alu* elements can be used as multiple-loci polymorphic DNA markers for linkage mapping of the human genome (90).

Conclusion

As shown in this review, TEs have a profound impact on the human genome. This impact includes structural rearrangements, mutational insertions and interaction with the host cellular machinery. While TE-induced variations are generally deleterious to the host, as reported in Tables 1 and 2, some insertions are advantageous and have probably been selected for (Table 3). Therefore, these elements are probably not

as 'selfish' as proposed in the 1980s. The analysis of TE dynamics and evolution in several organisms suggests the existence of a coevolution process between hosts and TEs. Indeed, a genome needs to be plastic to solve the challenges of the environment rapidly. In other words, a genome without any TEs probably represents an evolutionary 'cul-de-sac'. Therefore, even if these elements may have a deleterious impact on the host genomes at the individual level, they can also be seen as a useful genetic load at the population level. Finally, the understanding of the dynamics of such elements in a genome, and in particular, in the human genome, will probably be useful to our understanding of several diseases and how to treat them.

References

1. Britten RJ, Kohne DE. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* 1968; 161: 529–40.
2. Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 1998; 8: 464–78.
3. SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, et al. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 1996; 274: 765–8.
4. Smit AFA. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* 1999; 9: 657–63.
5. Dawkins R. *The extended phenotype*. Oxford, UK. Oxford University Press; 1982.
6. Dawkins R. *The selfish gene*. Oxford, UK. Oxford University Press; 1976.
7. Hall BG. Mobile elements as activators of cryptic genes in *E. coli*. *Genetica* 2000; (in press).
8. Grandbastien MA. Activation of plant retrotransposons under stress conditions. *Trends Plants Sci* 1998; 3: 181–7.
9. McClintock B. The significance of responses of the genome to challenge. *Science* 1984; 226: 792–801.
10. Deininger PL, Batzer MA. *Alu* repeats and human disease. *Mol Genet Metab* 1999; 67: 183–93.
11. Smit AF, Riggs AD. *Tiggers* and other DNA transposon fossils in the human genome. *Proc Natl Acad Sci U S A* 1996; 93: 1443–8.
12. Temin HM. Origin of retroviruses from cellular moveable genetic elements. *Cell* 1980; 21: 599–600.
13. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 1990; 9: 3353–62.
14. Lerat E, Brunet F, Bazin C, Capy P. Is the evolution of transposable elements modular? *Genetica* 2000, (in press).
15. Friedlander A, Patarca R. Endogenous proviruses. *Crit Rev Oncog* 1999; 10: 129–59.
16. Kazanian HH Jr. An estimated frequency of endogenous insertional mutations in humans. *Nat Genet* 1999; 22: 130.
17. Mustajoki S, Ahola H, Mustajoki P, Kauppinen R. Insertion of *Alu* element responsible for acute intermittent porphyria. *Hum Mutat* 1999; 13: 431–8.
18. Miki Y. Retrotransposal integration of mobile genetic elements in human diseases. *J Hum Genet* 1998; 43: 77–84.
19. Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-Ida E, Nomura Y, et al. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* 1998; 394: 388–91.
20. Labuda D, Zietkiewick E, Mitchell GA. *Alu* element as a source of genomic variation: deleterious effects and evolutionary novelties. In: Maraia RJ, ed. *The impact of short interspersed elements (SINEs) on the host genome*. Austin, TX: RG Landes Company, Springer; 1995: 1–24.
21. Vervoort R, Gitzelmann R, Lissens W, Liebaers I. A mutation (IVS8+0.6kbpdelTC) creating a new donor splice site activates a cryptic exon in an *Alu*-element in intron 8 of the human β -glucuronidase gene. *Hum Genet* 1998; 103: 686–93.
22. den Hollander AI, ten Brink JB, de Kok YJ, van Soest S, van den Born LI, van Driel MA, et al. Mutations in a human homologue of drosophila crumbs cause retinitis pigmentosa (RP12). *Nat Genet* 1999; 23: 217–21.
23. Seidl C, Donner H, Petershofen E, Usadel KH, Seifried E, Kaltwasser JP, et al. An endogenous retroviral long terminal repeat at the HLA-DQB1 gene locus confers susceptibility to rheumatoid arthritis. *Hum Immunol* 1999; 60: 63–8.
24. Segal Y, Peissel B, Renieri A, de Marchi M, Ballabio A, Pei Y, et al. LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis. *Am J Hum Genet* 1999; 64: 62–9.
25. Montagna M, Santacatterina M, Torri A, Menin C, Zullato D, Chieco-Bianchi L, et al. Identification of a 3 kb *Alu*-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families. *Oncogene* 1999; 18: 4160–5.
26. Reiter LT, Murakami T, Koeuth T, Pentao L, Muzny DM, Gibbs RA, et al. A recombination hot spot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. *Nat Genet* 1996; 12: 288–97.
27. Kumatori A, Faizunnessa NN, Suzuki S, Moriuchi T, Kurozumi H, Nakamura M. Nonhomologous recombination between the cytochrome b558 heavy chain gene (CYBB) and LINE-1 causes an X-linked chronic granulomatous disease.

- Genomics*; 1998; 53: 123–8.
28. Jeffs AR, Benjes SM, Smith TL, Sowerby SJ, Morris CM. The BCR gene recombines preferentially with Alu elements in complex BCR-ABL translocation of chronic myeloid leukaemia. *Hum Mol Genet* 1998; 7: 767–76.
 29. Obata K, Hiraga H, Nojima T, Yoshida MC, Abe S. Molecular characterization of the genomic breakpoint junction in a t(11;22) translocation in Ewing sarcoma. *Genes Chromosomes Cancer* 1999; 25: 6–15.
 30. Centra M, Memeo E, d'Apolito M, Savino M, Ianzano L, Notarangelo A, et al. Fine-exon-intron structure of the Fanconi anemia group A (FAA) gene and characterization of two genomic deletions. *Genomics* 1998; 51: 463–7.
 31. Rosenberg N, Yatuv R, Orion Y, Zivelin A, Dardik R, Peretz H, et al. Glanzmann thrombasthenia caused by an 11.2-kb deletion in the glycoprotein IIIa (beta3) is a second mutation in Iraqi Jews that stemmed from a distinct founder. *Blood* 1997; 89: 3654–62.
 32. Rothberg PG, Ponnuru S, Baker D, Bradley JF, Freeman AI, Cibis GW, et al. A deletion polymorphism due to Alu-Alu recombination in intron 2 of the retinoblastoma gene: association with human gliomas. *Mol Carcinog* 1997; 19: 69–73.
 33. Huie ML, Shanske AL, Kasper JS, Marion RW, Hirschhorn R. A large Alu-mediated deletion, identified by PCR, as the molecular basis for glycogen storage disease type II (GSDII). *Hum Genet* 1999; 104: 94–8.
 34. Burwinkel B, Kilmann MW. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J Mol Biol* 1998; 277: 513–7.
 35. Luzzi P, Rafi MA, Wenger DA. Characterization of the large deletion in the GALC gene found in patients with Krabbe disease. *Hum Mol Genet* 1995; 4: 2335–8.
 36. Magewu AN, Jones PA. Ubiquitous and tenacious methylation of the CpG site in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer. *Mol Cell Biol* 1994; 14: 4225–32.
 37. Graff JR, Herman JG, Myöhänen S, Baylin SB, Vertino PM. Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in *de novo* methylation. *J Biol Chem* 1997; 272: 22322–9.
 38. Arnaud P, Goubely C, Péliissier T, Deragon JM. SINE retrotransposons can be used *in vivo* as nucleation centers for *de novo* methylation. *Mol Cell Biol* 2000, (in press).
 39. Romig H, Ruff J, Fackelmeyer F, Patil M, Richter A. Characterization of two intronic nuclear matrix attachment regions in the human DNA topoisomerase I gene. *Eur J Biochem* 1994; 221: 411–9.
 40. Bruni R, Argentini C, D'Ugo E, Ciccaglione A, Rapicetta M. Recurrence of WHV integration in the *b3n* locus in woodchuck hepatocellular carcinoma. *Virology* 1995; 214: 229–34.
 41. Schmid CW. Does SINE evolution preclude Alu function? *Nucl Acids Res* 1998; 26: 4541–50.
 42. Kuo KW, Sheu HM, Huang YS, Leung WC. Expression of transposon LINE-1 is relatively human-specific and function of the transcripts may be proliferation-essential. *Biochem Biophys Res Commun* 1998; 253: 566–70.
 43. Asch HL, Eliacin E, Fanning TG, Connolly JL, Bratthauer G, Asch B. Comparative expression of the LINE p40 protein in human breast carcinomas and normal breast tissues. *Oncol Res* 1996; 8: 239–47.
 44. Florl AR, Löwer R, Schmitz-Dräger BJ, Schulz WA. DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. *Br J Cancer* 1999; 80: 1312–21.
 45. Larsson E, Andersson G. Beneficial role of human endogenous retroviruses: facts and hypotheses. *Scand J Immunol* 1998; 48: 329–38.
 46. Xie H, Brines ML, de Lanerolle NC. Transcripts of the transposon mariner are present in epileptic brain. *Epilepsy Res* 1998; 32: 140–53.
 47. Fujino T, Navaratnam N, Scott J. Human apolipoprotein B RNA editing deaminase gene (APOBEC1). *Genomics* 1998; 47: 266–75.
 48. Brittens RJ. Mobile elements inserted in the distant past have taken on important functions. *Gene* 1997; 205: 177–82.
 49. Brosius J, Tiedge H. Reverse transcriptase: mediator of genomic plasticity. *Virus Genes* 1996; 11: 163–79.
 50. Long Q, Bengra C, Li C, Kutlar F, Tuan D. A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human β -globin locus control region. *Genomics* 1998; 54: 542–55.
 51. Mager DL, Hunter DG, Schertzer M, Freeman JD. Endogenous retroviruses provide the primary polyadenylation signal for two new human genes. *Genomics* 1999; 59: 255–63.
 52. Gu JJ, Sychala J, Mitchell BS. Regulation of the human inosine monophosphate dehydrogenase type I gene. Utilization of alternative promoters. *J Biol Chem* 1997; 272: 4458–66.
 53. Lowe D, Henning D, Reddy R. An RNA polymerase III transcription unit located in the upstream control regions of the human proliferating-cell nucleolar protein p120 is transcribed *in vitro* and *in vivo*. *Eur J Biochem* 1999; 266: 112–22.
 54. Kazakov VI, Tomilin NV. Increased concentration of some transcription factor binding sites in human retrotransposons of the Alu family. *Genetica* 1996; 97: 15–22.
 55. Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, Tomilin N. Association of some potential hormone response elements in human genes with the alu family repeats. *Gene* 1999; 239: 341–9.
 56. Monte SM, Ghanbari K, Frey WH, Beheshti I, Averbach P, Hauser SL, et al. Characterization of the AD7C-NTP cDNA expression in Alzheimer's disease and measurement of a 41-kD protein in cerebrospinal fluid. *J Clin Invest* 1997; 100: 3093–104.
 57. Tulko J, Korotkov EV, Phoenix DA. MIRs are present in coding regions of human genes. *DNA Seq* 1997; 8: 31–8.
 58. Platzer M, Rotman G, Bauer D, Uziel T, Savitsky K, Bar-Shira A, et al. Ataxia-telangiectasia locus: sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene. *Genome Res* 1997; 7: 592–605.
 59. Berquin IM, Ahram M, Sloane BF. Exon 2 of human cathepsin B derives from an Alu element. *FEBS Lett* 1997; 419: 121–3.
 60. Mullersman JE, Pfeffer LM. An Alu cassette in the cytoplasmic domain of an interferon receptor subunit. *J Interferon Cytokine Res* 1995; 15: 815–7.
 61. Donohue SJ, Roseboom PH, Illnerova H, Weller JL, Klein DC. Human hydroxyindole-O-methyltransferase: presence of LINE-1 fragment in a cDNA clone and pineal mRNA. *DNA Cell Biol* 1993; 12: 715–27.
 62. Hakim I, Amariglio N, Grossman Z, Simoni-Brok F, Ohno S, Rechavi G. The genome of the THE I transposable repetitive elements is composed of a basic motif homologous to an ancestral immunoglobulin gene sequence. *Proc Natl Acad Sci U S A* 1994; 91: 7967–9.
 63. Gerber A, O'Connell MA, Keller W. Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. *RNA* 1997; 3: 453–63.
 64. Banki K, Halladay D, Perl A. Cloning and expression of the human gene for transaldolase. A novel highly repetitive element constitutes an integral part of the coding sequence. *J Biol Chem* 1994; 269: 2847–51.
 65. Makalowski W, Mitchell GA, Labuda D. Alu sequences in

- the coding regions of mRNA: a source of protein variability. *Trends Genet* 1994; 10: 188–93.
66. Yulug IG, Yulug A, Fisher EMC. The frequency and position of Alu repeats in cDNAs, as determined by database searching. *Genomics* 1995; 27: 544–8.
 67. Vidal F, Mougneau E, Glaichenhaus N, Vaigot P, Darmon M, Cuzin F. Coordinated posttranscriptional control of gene expression by modular elements including Alu-like repetitive sequences. *Proc Natl Acad Sci U S A* 1993; 90: 208–12.
 68. Merritt CM, Easteal S, Board PG. Evolution of human α 1-acid glycoprotein genes and surrounding Alu repeats. *Genomics* 1990; 6: 659–65.
 69. Dulai KS, von Dornum M, Mollon JD, Hunt DM. The evolution of trichromatic color vision by opsin gene duplication in new world and old world primates. *Genome Res* 1999; 9: 629–38.
 70. Kulski JK, Gaudieri S, Bellgard M, Balmer L, Giles K, Inoko H, et al. The evolution of MCH diversity by segmental duplication and transposition of retroelements. *J Mol Evol* 1997; 45: 599–609.
 71. Gaudieri S, Leelayuwat C, Townsend DC, Kulski JK, Dawkins RL. Genomic characterization of the region between HLA-B and TNF: implication for the evolution of multicopy gene families. *J Mol Evol* 1997; 44: S147–54.
 72. Kulski JK, Gaudieri S, Martin A, Dawkins RL. Coevolution of PERB11 and HLA class I genes with HERV-16 and retroelements by extended genomic duplication. *J Mol Evol* 1999; 49: 84–97.
 73. Moran JV, DeBerardinis RJ, Kazazian HH. Exon shuffling by LINE-1 retrotransposition. *Science* 1999; 283: 1530–4.
 74. Laurent AM, Puechberty J, Roizès G. Hypothesis: for the worst and for the best, LINE-1Hs retrotransposons actively participate in the evolution of the human centromeric alphoid sequences. *Chrom Res* 1999; 7: 305–17.
 75. Choo KH. Centromere DNA dynamics: latent centromeres and neocentromere formation. *Am J Hum Genet* 1997; 61: 1225–33.
 76. Marçais B, Bellis M, Gérard A, Pagès M, Boublik Y, Roizès G. Structural organization and polymorphism of the alpha satellite DNA sequences of chromosome 13 and 21 as revealed by pulse field electrophoresis. *Hum Genet* 1991; 86: 311–6.
 77. Laurent AM, Puechberty J, Prades C, Gimenez S, Roizès G. Site-specific retrotransposition of LINE-1 elements within human alphoid satellite sequences. *Genomics* 1997; 46: 127–32.
 78. Puechberty J, Laurent AM, Gimenez S, Billault A, Brun-Laurent MA, Calenda A, et al. Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: recombination across Scen. *Genomics* 1999; 56: 274–87.
 79. Kipling D, Warburton PE. Centromeres, CENP-B and Tigger too. *Trends Genet* 1997; 13: 141–4.
 80. Cook JM, Tristem M. SINEs of a times-transposable element as clade markers for their hosts. *Trends Ecol Evol* 1997; 12: 295–7.
 81. Hillis DM. SINEs of the perfect character. *Proc Natl Acad Sci U S A* 1999; 96: 9979–81.
 82. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc Natl Acad Sci U S A* 1997; 94: 1872–7.
 83. Feng Q, Moran JV, Kazazian HH, Boeke JD. Human LINE-1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 1996; 87: 905–16.
 84. Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, et al. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci U S A* 1994; 91: 12288–92.
 85. Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, et al. Polymorphic Alu insertions and the Asian origin of Native American populations. *Hum Biol* 1998; 70: 23–39.
 86. Hamdi H, Nishio H, Zielinski R, Dugaiczak A. Origin and phylogenetic distribution of Alu DNA repeats: irreversible events in the evolution of primates. *J Mol Biol* 1999; 289: 861–71.
 87. Johnson WE, Coffin JM. Constructing primate phylogenies from ancient retrovirus sequences. *Proc Natl Acad Sci U S A* 1999; 96: 10254–60.
 88. Verneau O, Catzeflis F, Furano A. Determining and dating recent rodent speciation events by using LINE-1 retrotransposons. *Proc Natl Acad Sci U S A* 1998; 95: 11284–9.
 89. Zietkiewicz E, Labuda M, Sinnott D, Glorieux FH, Labuda D. Linkage mapping by simultaneous screening of multiple polymorphic loci using Alu oligonucleotide-directed PCR. *Proc Natl Acad Sci U S A* 1992; 89: 8448–51.