J Hum Genet (1998) 43:77–84

MINIREVIEW

Yoshio Miki

Retrotransposal integration of mobile genetic elements in human diseases

Received: February 5, 1998 / Accepted: March 4, 1998

Abstract Approximately one-third of the mammalian genome is composed of highly repeated DNA sequences, of which the two major families, the long and short interspersed nucleotide elements (LINEs and SINEs), are represented in humans by *L1* and *Alu* elements respectively. Both types of element are considered to be retrotransposable and to play significant roles in genomic function and evolution. The majority of inserted elements are truncated and often rearranged relative to full-length elements; usually, such retrotransposed sequences are flanked by target-site duplications of various lengths and contain 3′ polyA tracts, common characteristics of retrotransposal integration. Retrotransposal integrations of *Alu* and *L1* sequences into biologically important genes appear to play significant roles in some human diseases. Most of the inserted sequences that cause human diseases seem to belong to one or a few subsets of each type of retrotransposon, suggesting that only a few active elements can function as templates for retrotransposition. Integrations observed in oncogenes and in tumor suppressor genes may participate in carcinogenesis by altering the activity of the affected genes. The exact mechanism of these events is unclear; however, retrotransposal integration may be a general mechanism of mutation in humans.

Key words Retrotransposon · Repeated DNA sequence · Mobile genetic element · *L-1* sequence · *Alu* sequence · Insertion mutation · Human disease

Y. Miki
Department of Molecular Diagnosis, The Cancer Institute, Japanese Foundation for Cancer Research, 1-37-1 Kami-ikebukuro, Toshima-ku, Tokyo 170, Japan
Tel. and Fax +81-3-5394-3926
e-mail: yosmiki@ims.u-tokyo.ac.jp

Introduction

The mammalian genome contains several families of repeated DNA sequences, some of which appear to be transposable elements. Retrotransposal integrations of sequences such as *Alu* and *L-1* into biologically important genes appear to play significant roles in some human genetic diseases. Retrotransposons are a specific group of movable genetic elements, structurally and functionally related to retroviruses, that transpose by way of an RNA intermediate; i.e., the element is transcribed to RNA, reverse-transcribed to DNA, and reintegrated elsewhere in the genome.

Eukaryotic retrotransposons are of two general types, Class I and Class II (Fanning and Singer 1987). The coding regions of Class I elements, such as *Ty* in yeast or *copia* in *Drosophila*, are flanked by long terminal repeats (LTRs) and contain reverse transcriptase- and/or integrase-related sequences that possibly initiate their own retrotransposition. There is no apparent DNA sequence specificity for host integration sites although some preference has been noted (Wilson and Young 1975). Class I retrotransposal elements are very similar to retroviruses, but the viruses possess envelope-protein genes to help them move from one cell to another. Reverse transcription of retroviruses and LTR-retrotransposons is an evolutionarily conserved process (Telesnitsky and Goff 1993). During integration, characteristic sequence changes occur in both the viral DNA and the host DNA: two nucleotides are deleted at each end of the viral DNA, and four to six base pairs of host DNA are duplicated at the integration site.

Class II retrotransposons do not contain LTRs, have polyadenylic acid (polyA) or A-rich sequences at the 3′ ends, and are flanked by target-site duplications of various lengths. In addition, these elements have at least one open reading frame (ORF) that is similar to the reverse transcriptases of retroviruses and class I retrotransposal elements (Xiong and Eickbush 1990; Doolittle et al. 1989). Class II elements can be subdivided into sequence-specific and non-sequence-specific groups. The insertion mecha-

nism of the R2 from *Bombyx mori* and the mitochondrial group II introns *al1* and *al2* from *Saccharomyces cerevisiae* implicate target site-primed reverse transcription at a sequence-specific single-strand break (Luan and Eickbush 1995; Moran et al. 1995). The most numerous transposable genetic elements in mammals are the short and long interspersed nucleotide elements (SINEs and LINEs), represented in the human genome by *Alu* and *L1*. Class II retrotransposons, such as *L1*, and other elements, such as *Alu* and processed pseudogenes, are considered to move by retrotransposition and lack LTRs and target-site duplications of specific lengths; however, Class II retrotransposons may encode their own reverse transcriptases. This characteristic differentiates them from *Alu* sequences and processed pseudogenes (Temin 1985). Evidence is convincing that both *Alu* and *L1* have been spread by insertion throughout the human genome over time through retrotransposition, wherein new copies of active elements can be reinserted into the genome at new locations. Depending on its genomic location, the new element may disrupt the normal expression of a gene and result in disease.

Insertion mutations of *L1* elements in human diseases

L1, a common family of repetitive sequences in the human genome, is considered to be a transposable element that can be transcribed into RNA, transcribed reversely into cDNA, and then reintegrated into genomic DNA. About 10^5 copies of *L1* sequence are present in all mammalian genomes, comprising roughly 5% of genomic DNA (Fanning and Singer 1987). They are up to 7 kilobases long, flanked by target-site duplications that vary from 5 to 15 bases. Unlike retroviral transposons, they carry no long terminal repeats (LTRs), but A-rich regions, usually preceded by polyadenylation signals, are present at their 3′ ends. A consensus human *L1* sequence constructed from genomic *L1* sequences (Scott et al. 1987) comprises 6.0 kb and contains a 5′ untranslated region (5'UTR) with an internal promoter (Swergold 1990), two long open reading frames (ORF-1 and ORF-2), and a 3'UTR. ORF-1 encodes a 40-kDa RNA-binding protein, but its precise function is unknown (Holmes et al. 1992). ORF-2 contains a region which is highly homologous to retroviral reverse transcriptase and RNase H (Hattori et al. 1989; Loeb et al. 1986; Xiong and Eickbush 1990). These features indicate that *L1* elements may be "nonviral retrotransposons" (Schwarz-Sommer et al. 1987), which transpose though an RNA intermediate with reverse transcriptase activity, in a mechanism similar to that postulated for generating pseudogenes (Sharp 1983). However, because most *L1* elements are heterogeneously truncated from their 5′ ends (up to 95% in mammals; Schwarz-Sommer et al. 1987) and some contain internal rearrangements, it is extremely difficult to identify a structure or sequence which may still function as a transposon.

Kazazian et al. (1988) documented evidence that some *L1* elements in the human genome are retrotransposable, when they detected de novo insertion of a truncated *L1* element into exon 14 of the Factor VIII gene on the X chromosome in two patients with hemophilia A. Each of these insertions contained a 3′ portion of the *L1* sequence; one was composed of a 3.8-kb truncated *L1* sequence followed by a 57-bp polyA tract, creating an A-rich target-site duplication of at least 12 base pairs of the Factor VIII sequence. The other insertion was composed of a 2.3-kb 3′ end of the *L1* sequence, flanked by an A-rich target-site duplication of at least 13 nucleotides. However, this *L1* sequence contained an internal rearrangement. On the basis of its unique 3′ trailer sequence, the inserted element was closely related to members of the *Ta* subset of the *L1* family (Skowronski and Singer 1986). The full-length precursor (LRE1) to one of the factor VIII insertions was identified on chromosome 22 (Dombroski et al. 1991) and shown to contain the two ORFs predicted by the *L1* consensus sequence. ORF1 encoded a protein of unknown function in a transient expression assay (Holmes et al. 1992); ORF2 encoded a protein with reverse transcriptase activity in a yeast assay system (Mathias et al. 1991).

Two groups have reported insertions of *L1* sequence into the dystrophin gene that resulted in Duchenne muscular dystrophy (DMD). An insertion identified in exon 48 of the dystrophin gene by Holmes et al. (1994) comprised a total of 1968 bp, including 1401 bp of truncated and rearranged *L1* element followed by a 37-bp polyA tail; 530 bp of non-*L1* sequence flanked the 3′ end of the element. A perfect 12- to 15-bp target-site duplication of sequence from exon 48 surrounded the entire insertion. Since the 3′ non-*L1* sequence of this insertion showed no obvious homologies with other known sequences, it was designated a "unique sequence component" (USC). Analysis of genomic DNA from normal individuals revealed that the USC was adjacent to the progenitor *L1* element in the genome. It appears that read-through transcription resulted in an *L1*-USC transcript capable of retrotransposition. The *L1* sequence inserted into this patient's dystrophin gene (dystrophin *L1*) was shown to belong to the *Ta* subset of the *L1* family, and therefore was closely related to the *L1* sequence inserted into the Factor VIII gene mentioned before (Factor VIII *L1*). However, since the full-length progenitor of the Factor VIII *L1*, *LRE1*, differed in eight nucleotides from the dystrophin *L1*, *LRE1* was not the precursor in this case. *LRE2*, the actual progenitor, was identified on chromosome 1q using the USC of dystrophin *L1*.

Narita et al. (1993) reported evidence of retrotransposition of *L1* into the dystrophin gene of two Japanese brothers with DMD. In these patients the dystrophin gene was disrupted by insertion of a truncated *L1* sequence into exon 44. The 606–608 bp *L1* fragment was identical to the inverse complement of the 3′ portion of the *Ta* subset of *L1* elements (Skowronski and Singer 1986). The insertion lay within an A-T rich region in the dystrophin gene, and its second ORF was conserved; however, no target-site duplication was created and two nucleotides were deleted from the target site in the exon. Exon 44 of the transcript was skipped during splicing, an event that would shift the coding frame and lead to early termination of translation.

We ourselves reported disruption of the *APC* gene in a colon carcinoma, caused by a somatic insertion of an *L1* element into the last exon of this gene (Miki et al. 1992). The 750-bp inserted sequence contained a 3′ portion of the *L1* consensus sequence; the first fifth of the insertion was almost identical to the inverse component of the *L1* consensus, and the central part of the insertion was highly homologous to the 3′ portion of the *L1* consensus. The last fifth was composed of a polyadenylation signal and a subsequent 180-base polyA tract. Furthermore, since we observed an AT-rich, 8-bp target-site duplication without deletion of any *APC* sequence, retrotransposal insertion of an active *L1* sequence was suspected as the cause of this somatic mutational event.

Morse et al. (1988) reported yet another somatic insertion of an *L1* sequence into a human gene. In that case, the rearranged *L1* element was inserted between coding exons 2 and 3 of *c-myc* in a human breast carcinoma. This evidence suggested that somatic rearrangement caused by the *L1* insertion, by activating the *c-myc* protooncogene, represented an important contributor to breast carcinogenesis in this patient.

Insertion mutations of *Alu* elements in human diseases

Alu elements constitute a specifically human family of interspersed repetitive sequences. They are mobile elements, with copy numbers in excess of 500,000 within the human genome (Deininger et al. 1981). Mobilization of *Alu* elements is thought to occur through RNA polymerase III-derived transcription in a retrotransposition process (Rogers 1983). Consensus *Alu* sequences are approximately 280 bp long (Fuhrman et al. 1981; Willis 1993), and consist of two similar but distinct monomers linked by an oligo-d(A) tract of variable length. *Alu* elements, which are derived from structural RNA (Ullu and Tschudi 1984), are flanked by short direct repeats of host DNA at integration sites in the genome.

We have reported a case of retrotransposal integration of an *Alu* element into a breast cancer susceptibility gene, *BRCA2*, of a breast-cancer patient (Miki et al. 1996). In this novel germline event, insertion of the *Alu* element into exon 22 resulted in alternative splicing that skipped this exon. The presence of a 64-bp polyA tract and an 8-bp target-site duplication, which is a common characteristic of retrotransposal integration of mobile elements into staggered single-strand nicks (Deininger 1989), implied that retrotransposal insertion of a transcriptionally active *Alu* element had caused an inactivating mutation. Furthermore, the sequence corresponding to the insertion site was AT-rich, in keeping with the hypothesis that *Alu* elements preferentially integrate into AT-rich regions (Daniels and Deininger 1985). The sequence inserted into *BRCA2* in this patient was highly homologous to the consensus sequence of the conserved *Alu* subfamily (Fig. 1) (Matera et al. 1990a); however, the middle portion of the insertion was identical to a sequence of the *HS Alu* subfamily, which is considered to be of relatively recent evolutionary origin (Batzer and Deininger 1991). Hence, the origin of the inserted *Alu* element is not clear. In any event, the skipping of exon 22 in the *BRCA2* transcript would shift the coding frame and lead to early termination of translation.

Four other groups have reported disease-causing inactivation of human genes due to retrotransposal integration of *Alu* elements. In the first example, a *de novo* insertion of an *Alu* sequence was found within exon 5 of the factor IX gene in a patient with severe hemophilia B (Vidaud et al. 1993). The inserted *Alu* element was 322 bp long, lay in the sense direction, and interrupted the reading frame at glutamic acid 96 of the mature factor IX protein, resulting in an in-frame stop codon within the inserted sequence. This alteration of reading frame probably caused the disease. The inserted sequence differed from the *HS Alu* family only by one additional adenine residue, was flanked by perfect 15-bp duplications of the target-site sequence, and contained a pure polyA tract of at least 78 residues at the 3′ end. The direct repeats were not A-T rich; however, the sequence surrounding the repeats consisted predominantly of A and T residues. Furthermore, the sequence of the direct repeats, 5′-GAN-3′, is a highly specific target site for insertion of elements from the HS subfamily (Matera et al. 1990b).

Muratani et al. (1991) identified insertion of an *Alu* element that disrupted exon 2 of the *ChE* gene in a patient with acholinesterasemia. The inserted element was 342 bp long, including a polyA tract of 38 bp, and was flanked by 15-bp target-site duplications at both junctions; no mutation was present in the original *ChE* gene sequence. This inserted element showed 93% homology with the evolutionarily most recent human *Alu* consensus sequence (Britten et al. 1988). These results suggested that the *Alu* insertion in this patient might represent a retrotransposal mechanism. The insertion had occurred in both alleles of the proband and was inherited in the patient's family. Two of the four patients studied in this family were homozygous for the *ChE* mutation and possessed no serum ChE activity; the two heterozygous patients had half the normal level of ChE activity. Insertion of the *Alu* element appeared to lead to loss of function of the *ChE* gene.

Wallace et al. (1991) reported a *de novo* insertion of an *Alu* repetitive element into an intron of the *NF1* gene, resulting in alternative splicing that skipped a downstream exon and consequently shifted the reading frame. The 320-bp *Alu* element was inserted 44 bp upstream of exon 6, terminating in exon 7, and lay in the opposite orientation to that of the *NF1* gene. It included a polyA tract and was flanked by 3–13-bp direct repeats (precise lengths were not determined due to the polyA tract). The integration site included an A/T stretch of 26 bp. This *Alu* sequence was most highly homologous to the subfamily known as *HS-1*, which is considered to be transcriptionally active and the most recent *Alu* subgroup known to be capable of expansion in the genome.

Familial hypocalciuric hypercalcemia (FHH) is an autosomal dominant condition characterized by elevation of serum calcium, hypocalciuria, and nonsuppressed parathyroid

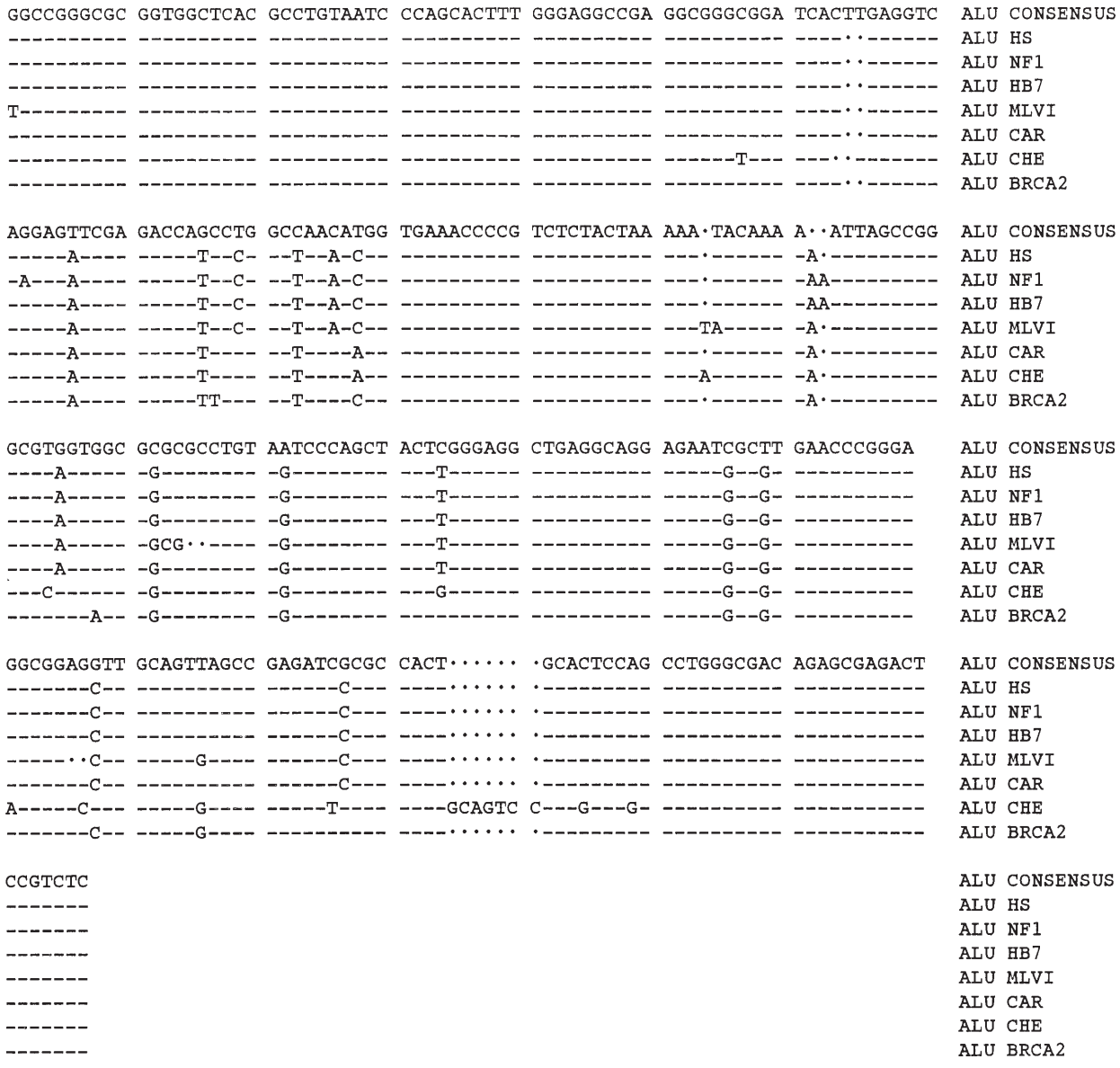

Fig. 1. Comparison of *Alu* insertions in reported clinical cases with *Alu* and *HS Alu* consensus sequences. The *Alu* consensus in the *first line* is based on 168 human Alu sequences (Quentin 1988); the *AluHS* consensus is derived from the recently formed subfamily (Batzer and Deininger 1991). Sequences for the *Alu* elements inserted into genes encoding neurofibromin (ALU NF1), factor IX (ALU HB7), Mlvi-2 (ALU MLVI), calcium-sensing receptor (ALU CAR), cholinesterase (ALU CHE), and BRCA2 product (ALU BRCA2) are shown *below* the consensus sequences. Identical nucleotides are indicated by *dashes*, absent nucleotides by *dots*

hormone (PTH). Neonatal severe hyperparathyroidism (NSHPT) is considered an autosomal-recessive disorder characterized by marked elevation in serum calcium and PTH levels, skeletal demineralization, and parathyroid cellular hyperplasia that can be lethal without parathyroidectomy. Although the clinical features of these disorders are different, inactivating mutations of the calcium-sensing receptor (*CASR*) gene at 3q13.3–21 have been detected in some families that exhibit both FHH and NSHPT, suggesting a molecular relationship between the two conditions. Janicic et al. (1995) identified insertion of an *Alu* element within exon 7 of the *CASR* gene in FHH and NSHPT patients from the same family. The inserted *Alu* element was 383 bp long, contained an exceptionally long polyA tract (92–94 bp), and was in opposite orientation relative to the *CASR* gene. The insertion created a direct repeat of a 15-bp sequence, without any deletion of *CASR*. Individuals with FHH in that family were heterozygous, and some cases of NSHPT were homozygous for the insertion mutation.

The *Mlvi-2* locus, containing a putative oncogene involved in the induction of thymic lymphomas in rats (Tsichlis et al. 1983), was originally defined as one of at least five common regions for proviral integration of Moloney murine leukemia virus in this species. In a human B-cell lymphoma, one allele of the human homologue of the *Mlvi-2* locus was found to be rearranged due to the insertion of an *Alu* element (Economou-Pachis and Tsichlis 1985). The

inserted *Alu* sequence had a polyA tail at the 3′ end and was flanked by an 8-bp direct repeat. Since no normal tissue was available from the patient with this lymphoma, whether the insertion was a somatic or germline event could not be determined.

Evidence for at least one more *Alu* retrotransposition causing inactivation of a human gene was reported in two patients with Huntington's disease (HD), a late-onset autosomal dominant neuropsychiatric disorder. *Alu* insertions were identified in genomic DNA which showed complete cosegregation with the disease in both patients' families, but insertions were not found in 1000 unrelated HD controls (Goldberg et al. 1993). However, as the sites of insertion were not within the HD gene itself (Huntington's Disease Collaborative Research Group 1993) it is not clear whether the observed insertion was a direct cause of the disease in these two families.

Other pathogenic mechanisms involving *Alu* elements

Reintegrated *Alu* sequences can play roles in human genetic diseases in three ways: (1) by de novo retrotransposition of an *Alu* sequence into genes as already described; (2) by splice-mediated insertion of an *Alu* sequence into mRNA; or (3) by homologous recombination that produces deletions and/or chromosomal rearrangements.

Gyrate atrophy of the choroid and retina (GA) is an autosomal recessive chorioretinal degeneration caused by deficiency of ornithine d-aminotransferase (OAT). In studies of OAT mutations, an allele whose mature mRNA showed a 142-nucleotide insertion at the junction of sequences from exons 3 and 4 was discovered (Mitchell et al. 1991). The inserted sequences corresponded to the complement of the right half of an *Alu* element; a single *Alu* element normally is present in OAT intron 3 (between exons 3 and 4) oriented in the direction opposite to OAT. A cytosine to guanine transversion of the *Alu* in intron 3 had created a new donor splice site, activating a cryptic acceptor splice site at its 5′ end and resulting in splice-mediated insertion of an *Alu* fragment into the mature ornithine d-aminotransferase mRNA.

Alport syndrome is a mainly X-linked hereditary disease of basement membranes characterized by progressive renal failure, deafness, and ocular lesions. The *a3*(IV) and *a4*(IV) collagen genes are involved in the less frequent autosomal recessive form of this disorder. Knebelmann et al. (1995) detected a mutant allele of *COL4A3* whose transcripts were disrupted by a 74-bp insertion at the junction of exons 4 or 5 and 6. This insertion derived from an antisense *Alu* element in intron 5, which activated a cryptic acceptor splice site within the *Alu* sequence. This mutation always segregated with disease in the family. Similar donor site-creating mutations have been found in intron 2 of *β-globin* (Treisman et al. 1983; Chen et al. 1984; Dobkin and Bank 1985) and in the human proto-oncogene *c-rel* (Brownell et al. 1989).

Deletions of genes due to homologous recombination between two *Alu* elements have been detected in genes encoding the LDL receptor (Lehrman et al. 1985), β-hexosaminidase α-chain (Myerowitz and Hogikyan 1987), α-globin (Nicholls et al. 1987), and adenosine deaminase (ADA) (Markert et al. 1988). Additionally, a sex-chromosome rearrangement in a human XX male, caused by *Alu-Alu* recombination, was reported (Rouyer et al. 1987).

Insertions of transposons into genes of other mammalian species

An intracisternal A particle (*IAP*), a member of an abundant family of murine transposable elements, is an endogenous retroviral-like DNA element in mice. Insertion of an *IAP* was found within the coding region of the oncogene *c-mos* in mouse plasmacytoma (Rechavi et al. 1982); the rearranged *c-mos* gene in this tumor was actively transcribed and possessed transforming activity. Furthermore, a somatic insertion of a transposable element showing approximately 60% homology to *L1* elements was identified in the 5′ flanking region of the *c-myc* gene in a canine transmissible venereal tumor (Katzir et al. 1985). Target-site duplication and a polyA tail suggested that the insertion had occurred as a retrotransposal event in this dog.

In mice, a dominant mutation of the *Fu* gene is related to a kinky tail, and homozygotes for mutated *Fu* alleles exhibit deafness, other neurological symptoms, and urogenital defects. Vasicek et al. (1997) detected two dominant mutations that resulted from insertions of *IAP* retrotransposons into the murine *Fu* gene; both led to complete or partial absence of wild-type products.

The *reeler* mouse shows neurological symptoms that include ataxia and tremors. In this mutant strain, Takahashi et al. (1996) found that exon-skipping of the *reeler* gene caused a 220-bp deletion in the transcript that disrupted the normal reeler product. The skipped exon contained an inserted element composed of the full-length murine *L1* sequence.

Mice homozygous for the spastic mutation (*spa*) suffer from a complex motor disorder. The glycine receptor β-subunit gene (*Glrb*) maps to the same region of mouse chromosome 3 as *spa*, and *Glrb* mRNA is markedly reduced in brains of spa mice. Two groups of investigators have documented an insertion of a 7.1-kb *L1* element in intron 6 of *Glrb* that caused exon skipping (Kingsmore et al. 1994; Mulhardt et al. 1994).

Hereditary renal carcinoma in the Eker rat is an example of a Mendelian dominant predisposition to a specific cancer, and the Eker mutation is tightly linked to the tuberous sclerosis (*Tsc2*) gene. Kobayashi et al. (1995) showed that the *Tsc2* gene in the Eker rat was disrupted by germline insertion of an approximately 5-kb DNA fragment thought to be an *L1* sequence, resulting in aberrant RNA expression.

Discussion

Table 1 summarizes reports of mammalian genes whose inactivation by retrotransposal integration has led to disease. Of the seven cases on the list with *L1* retrotransposal insertion, five were germline and the others were somatic events. Both somatic insertions of *L1* elements involved inversion of part of the *L1* sequence, while four of the six mammalian germline *L1* insertions discussed [four in humans, two in mice; *spa* and *reeler* – the details of the human *L1* insertion described by Bakker and Omenn (unpublished) are not known] had the expected collinear organization. Moreover, six of the seven *L1* insertions listed in Table 1 are members of the *Ta* subset, suggesting that this subset predominates among active elements. However, each of those six insertions differed in sequence and in truncation or rearrangement pattern. Four were rearranged relative to the *L1* consensus sequence, three of them in two blocks; in all three of those cases the first block was inverted. The fourth rearrangement consisted of three blocks, with the first and third being inverted. The sequence inserted into a dystrophin gene in the Japanese family cited earlier, which was one of the *L1* insertions without rearrangement, corresponded to the inverse complement of the 5′-truncated *L1* sequence (Narita et al. 1993); this insertion lacked a target-site duplication, and two nucleotides of the dystrophin sequence were deleted.

The *L1* element is common in all mammalian genomes; however, active *L1* elements that can transpose appear to be rare. Two specific human *L1* elements (*L1.2* and *LRE2*) are the likely precursors of disease-producing insertions. *L1.2* and *LRE2* autonomously retrotranspose into chromosomal DNA at high frequencies, and that efficiency of retrotransposition requires the presence of conserved regions of the ORF1 and ORF2 products. However, the actual number of *L1* elements that are capable of retrotransposition is still unknown. Recently, Sassaman et al. (1997) isolated 13 full-length elements, including three *L1*s capable of retrotransposition, by means of a selective screening strategy to enrich for active *L1* sequences. On the basis of their data, they estimated that the average diploid human genome contains 30–60 active *L1* elements.

In recent years a number of different functions have been described for *Alu* elements, and it has been learned that *Alu* elements can play several roles in human genetic disease. As already mentioned, homologous recombination between *Alu* elements resulting in chromosomal rearrangements, splice-mediated insertion mutations involving intronic antisense *Alu* sequences, or de novo retrotransposition of *Alu* sequences into genes can inactivate critical processes. The frequency or significance of mutations resulting from each type of event is unknown. This review has highlighted retrotransposal insertions of *Alu* sequence known to be factors in several different human diseases. Figure 1 compares the sequences of six inserted *Alu* sequences with the *Alu* consensus and the *Alu HS*-family sequence. Although each of the six insertions differs in sequence, four (ALU NF1, ALU HB7, ALU MLVI, and ALU CAR) are thought to belong to the *HS Alu* family, which is the most recent subfamily of transposable *Alu* sequences. The *Alu* sequence within the cholinesterase gene (ALU CHE) showed a high degree of homology with the human *Alu* consensus sequence of the most recent evolutionary branch (class VI). The inserted *Alu* element in the *BRCA2* gene showed a high degree of homology with the conserved *Alu* subfamily, which is considered to be older than the *HS Alu* subfamily.

The majority of *Alu* retrotranspositions have been completed; the process continues only in the youngest subfami-

Table 1 Inactivation of human genes due to retrotransposal integration

Gene	Element	Insertion	Disease	Reference
Factor VIII	L1	Germline	Hemophilia A	Kazazian et al. (1988)
	L1	Germline		
Dystrophin	L1	Germline	Duchenne muscular dystrophy	Narita et al. (1993)
	L1	Germline		Holmes et al. (1994)
	L1	Germline		Bakker and Omenn (unpublished)
APC	L1	Somatic	Familial polyposis coli	Miki et al. (1992)
c-myc	L1	Somatic	Breast cancer	Morse et al. (1988)
NF1	Alu	Germline	Neurofibromatosis type 1	Wallace et al. (1991)
Factor IX	Alu	Germline	Hemophilia B	Vidaud et al. (1993)
Cholinesterase	Alu	Germline	Acholinesterasemia	Muratani et al. (1991)
Calcium-sensing receptor	Alu	Germline	Neonatal severe hyperparathyroidism (NSHPT)	Janicic et al. (1995)
			Familial hypocalciuric hypercalcemia (FHH)	
BRCA2	Alu	Germline	Breast cancer	Miki et al. (1996)
Mlvi-2 locus	Alu	unknown	B cell lymphoma	Economou-Pachnis et al. (1985)
5′ flanking of Huntington gene[a]	Alu	Germline	Huntington's disease	Goldberg et al. (1993)
Ornithine δ-aminotransferase[b]	Alu	Germline	Gyrate atrophy of the choroid and retina	Mitchell et al. (1991)
COL4A3[b]	Alu	Germline	Autosomal recessive Alport syndrome	Knebelmann et al. (1995)

[a] The direct association of the disease with the insertion event is not clear.
[b] Splice-mediated insertion of an Alu sequence.

lies, and is still a rare event. Because most *Alu* elements scattered in the human genome are truncated at their 5′-ends, they are considered to have lost their ability to expand further. In addition, most of the *Alu* sequences involved in reported insertion mutations belong to the *Alu HS* subfamily. These observations support the "master gene" model of *Alu* evolution; i.e., only a few "master" *Alu* elements are able to produce new retrotransposons (Deininger et al. 1992).

The exact mechanism of retrotransposition and the biological function(s) of mobile genetic elements are still far from clear. However, we have considerable evidence that repetitive elements such as *L1* and *Alu* are capable of transposing through retrotranspositional events, and in some cases such events have resulted in human disease states. Insertion of retrotransposons in fact may be a general mechanism of mutation in humans, and it is likely that additional reports of its contribution to human disease will be forthcoming. It will be of interest to determine the frequency with which transposable-element-mediated inactivation of human genes occurs in somatic or germline cells, and discover whether inherited or environmental factors influence that frequency.

References

Batzer MA, Deininger PL (1991) A human-specific subfamily of *Alu* sequence. Genomics 9: 481–487

Britten RJ, Baron WF, Stout DB, Davidson EH (1988) Sources and evolution of human repeated *Alu* sequences. Proc Natl Acad Sci USA 85: 4770–4774

Brownell E, Mittereder N, Rice NR (1989) A human rel proto-oncogene cDNA containing an *Alu* fragment as a potential coding exon. Oncogene 4: 935–942

Chen TC, Orkin SH, Antonarakis SE, Potter MJ, Sexton JP, Markham AF, Giardina PJ, Li A, Kazazian HH Jr (1984) β-Thalassemia in Chinese: use of in vivo RNA analysis and oligonucleotide hybridization in systemic characterization of molecular defects. Proc Natl Acad Sci USA 81: 2821–2825

Daniels GR, Deininger PL(1985) Integration site preferences of the *Alu* family and similar repetitive DNA sequences. Nucleic Acids Res 13: 8939–8954

Deininger PL (1989) SINEs, short interspersed repeated DNA elements in higher eucaryotes. In: Howe M, Berg D (eds) Mobile DNA. ASM Press, Washington DC, pp 67–78

Deininger PL, Jolly DJ, Rubin CM, Friedman T, Schmid CW (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J Mol Biol 151: 17–23

Deininger PL, Batzer MA, Hutchison CA, Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8: 307–311

Dobkin C, Bank A (1985) Reversibility of IVS2 missplicing in a mutant human beta-globin gene. J Biol Chem 260: 16332–16337

Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr (1991) Isolation of an active human transposable element. Science 254: 1805–1808

Doolittle RF, Feng DF, Johnson MS, McClure MA (1989) Origins and evolutionary relationships of retroviruses. Q Rev Biol 64: 1–30

Economou-Pachnis A, Tsichlis PN (1985) Insertion of *Alu* SINE in the human homologue of the *Mlvi-2* locus. Nucleic Acids Res 13: 8379–8387

Fanning TG, Singer MF (1987) LINE-1:a mammalian transposable element. Biochem Biophys Acta 910: 203–212

Fuhrman SA, Deininger PL, LaPorte P, Friedmann T, Geiduschek EP (1981) Analysis of transcription of the human *Alu* family ubiquitous repeating element by eukaryotic RNA polymerase III. Nucleic Acids Res 9: 6439–6456

Goldberg YP, Rommens JM, Andrew SE, Hutchinson GB, Lin B, Theilmann J, Graham R, Graves ML, Starr E, McDonald H, Nasir J, Schappert K, Kalchman MA, Clarke LA, Hayden MR (1993) Identification of an *Alu* retrotransposition event in close proximity to a strong candidate gene for Huntington's disease. Nature 362: 370–373

Hattori M, Kuhara S, Takenaka O, Sakaki Y (1989) L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse-transcriptase-related protein. Nature 321: 625–628

Holmes SE, Singer MF, Swergold GD (1992) Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. J Biol Chem 267: 19765–19768

Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH Jr (1994) A new retrotransposable human L1 element from the *LRE2* locus on chromosome 1q produces a chimaeric insertion. Nat Genet 7: 143–148

Huntington's disease collaborative research group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosome. Cell 72: 971–983

Janicic N, Pausova Z, Cole DEC, Hendy GN (1995) Insertion of an Alu sequence in the Ca^{2+}-sensing receptor gene in familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. Am J Hum Genet 56: 880–886

Katzir N, Rechavi G, Cohen JB, Unger T, Simoni S, Segal S, Cohen D, Givol D (1985) "Retrotransposon" insertion into the cellular oncogene *c-myc* in canine transmissible venereal tumor. Proc Natl Acad Sci USA 82: 1054–1058

Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332: 164–166

Kingsmore SF, Giros B, Suh D, Bieniarz M, Caron MG, Seldin MF (1994) Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. Nat Genet 7: 136–142

Knebelmann B, Forestier L, Drouot L, Quinones S, Chuet C, Benessy F, Antignac SJ (1995) Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. Hum Mol Genet 4: 675–679

Kobayashi T, Hirayama Y, Kobayashi E, Kubo Y, Hino O (1995) A germline insertion in the tuberous sclerosis (*Tsc2*) gene gives rise to the Eker rat model of dominantly inherited cancer. Nat Genet 9: 70–74

Lehrman MA, Schneider WJ, Sudhof TC, Brown MS, Goldstein JL, Russell DW (1985) Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science 227: 140–146

Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB, Edgell MH, Hutchison CA III (1986) The sequence of a large L1 Md element reveals a tandemly repeated 5′ end and several features found in retrotransposon. Mol Cell Biol 6: 168–182

Luan DD, Eickbush TH (1995) RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol Cell Biol 15: 3882–3891

Markert ML, Hutton JJ, Wiginton DA, States JC, Kaufman RE (1988) Adenosine deaminase (ADA) deficiency due to deletion of the *ADA* gene promoter and first exon by homologous recombination between two Alu elements. J Clin Invest 81: 1323–1327

Matera AG, Hellmann U, Hintz MF, Schmid CW (1990a) Recently transposed Alu repeats result from multiple source genes. Nucleic Acids Res 18: 6019–6023

Matera AG, Hellmann U, Schmid CW (1990b) A transpositionally and transcriptionally competent Alu subfamily. Mol Cell Biol 10: 5424–5432

Mathias SL, Scott AF, Kazazian HHJr , Boeke J, Gabriel A (1991) Reverse transcriptase encoded by a human transposable element. Science 254: 1808–1810

Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y (1992) Disruption of the *APC* gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res 52: 643–645

Miki Y, Katagiri T, Kasumi F, Yoshimoto T, Nakamura Y (1996) Mutation analysis in the *BRCA2* gene in primary breast cancers. Nat Genet 13: 245–247

Mitchell GA, Labuda D, Fontaine G, Saudubray JM, Bonnefont JP, Lyonnet S, Brody LC, Steel G, Obie C, Valle D (1991) Splice-mediated insertion of an Alu sequence inactivates ornithine d-aminotransferase: A role for Alu elements in human mutation. Proc Natl Acad Sci USA 88: 815–819

Moran JV, Zimmerly S, Eskes R, Kennell JC, Lambowitz AM, Butow RA, Perlman PS (1995) Mobile group II introns of yeast mitochondrial DNA are site-specific retroelements. Mol Cell Biol 15: 2829–2838

Morse B, Rothberg G, South VJ, Spandorfer JM, Astrin SM (1988) Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma. Nature 333: 87–90

Mulhardt C, Fischer M, Gass P, Simon-Chazottes D, Guenet JL, Kuhse J, Betz H, Becker CM (1994) The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused intronic insertion of L1 element. Neuron 13: 1003–1015

Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K (1991) Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. Proc Natl Acad Sci USA 88: 11315–11319

Myerowitz R, Hogikyan ND (1987) A deletion involving Alu sequences in the beta-hexosaminidase alpha-chain gene of French Canadians with Tay-Sachs disease. J Biol Chem 262: 15386–15399

Narita N, Nishio H, Kitoh Y, Ishikawa Yuka, Ishikawa Yuki, Minami R, Nakamura H, Matsuo M (1993) Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. J Clin Invest 91: 1862–1867

Nicholls RD, Fischel-Ghodsian N, Higgs DR (1987) Recombination at the human alpha-globin gene cluster: sequence features and topological constraints. Cell 49: 369–378

Quentin Y (1988) The Alu family developed through successive waves of fixation closely connected with primate lineage history. J Mol Evol 27: 194–202

Rechavi G, Givol D, Canaani E (1982) Activation of a cellular oncogene by DNA rearrangement: possible involvement of an IS-like element. Nature 300: 607–611

Rogers J (1983) Retroposons defined. Nature 301: 460

Rouyer F, Simmler MC, Weissenbach J (1987) A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination. Cell 51: 417–425

Sassaman DM, Dombroski BA, Mran JV, Kimberland ML, Naas TP, DeBerardinis R, Gabriel A, Swergold GD, Kazazian HH Jr (1997) Many human L1 elements are capable of retrotransposition. Nat Genet 16: 37–43

Schwarz-Sommer Z, Leclercq L, Gobel E, Saedler H (1987) Cin4, an insert altering the structure of the *Alu* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J 6: 3873–3880

Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, Cooley T, Heath P, Smith KO, Margolet I (1987) Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics 1: 113–125

Sharp PA (1983) Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature 301: 471–472

Skowronski J, Singer M (1986) The abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. Cold Spring Harbor Symp Quant Biol 51: 457–464

Swergold GD (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol 10: 6718–6729

Takahashi T, Ohsumi T, Kuromitsu J, Shibata K, Sasaki N, Okazaki Y, Shibata H, Sato S, Yoshiki A, Kusakabe M, Muramatsu M, Ueki M, Okuda K, Hayashizaki Y (1996) Dysfunction of the Orleans *reeler* gene arising from exon skipping due to transposition of a full-length copy of an active L1 sequence into the skipped exon. Hum Mol Genet 5: 989–993

Telesnitsky A, Goff SP (1993) Strong-stop strand transfer during reverse transcription. In: Skalka AM, Goff SP (eds) Strong-stop strand transfer during reverse transcription. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 49–83

Temin HM (1985) Reverse transcription in the eukaryotic genome: Retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. Mol Biol Evol 2: 455–468

Treisman R, Orkin SH, Maniatis T (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature 302: 591–596

Tsichlis PN, Strauss PG, Hu LF (1983) A common region for proviral DNA integration in MoMuLV-induced rat thymic lymphomas. Nature 302: 445–449

Ullu E, Tschudi C (1984) Alu sequences are processed 7SL RNA gene. Nature 312: 171–172

Vasicek TJ, Zeng L, Guan XJ, Zhang T, Costantini F, Tilghman SM (1997) Two dominant mutations in the mouse *Fused* gene are the result of transposon insertions. Genetics 147: 777–786

Vidaud D, Vidaud M, Bahnak B, Siguret V, Sanchez SG, Laurian Y, Meyer D, Goossens M, Lavergne JM (1993) Haemophilia B due to a de novo insertion of a human-specific Alu subfamily member within the coding region of the factor IX gene. Eur J Hum Genet 1: 30–36

Wallace MR, Anderson LB, Saulino AM, Gregory PE, Glover TW, Collins FS (1991) A de novo Alu insertion results in neurofibromatosis type 1. Nature 353: 864–866

Willis IM (1993) RNA polymerase III. Genes, factors and transcriptional specificity. Eur J Biochem 212: 1–11

Wilson GA, Young FE (1975) Isolation of a sequence specific endonuclease (*Bam*I) from *Bacillus amyloliquefaciens* H. J Mol Biol 97: 123–125

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequence. EMBO J 9: 3353–3362