MINIREVIEW Genomics & Informatics Vol. 3(4) 133-141, December 2005

Genomic Features of Retroelements and Implications for Human Disease

Heui-Soo Kim*

Department of Biology, College of Natural Sciences, Pusan National University, Busan 609-735, Korea

Abstract

Most of the endogenous retroviral genes integrated into the primate genome after the split of New World monkeys in the Oligocene era, approximately 33 million years ago. Because they can change the structure of adjacent genes and move between and within chromosomes they may play important roles in evolutionas well as in many kinds of disease and the creation of genetic polymorphism. Comparative analysis of HERVs (human endogenous retroviruses) and their LTR (long terminal repeat) elements in the primate genomes will help us to understand the possible impact of HERV elements in the evolution and phylogeny of primates. For example, HERV-K LTR and SINE-R elements have been identified that have been subject to recent change in the course of primate evolution. They are specific elements to the human genome and could be related to biological function. The HERV-M element is related to the superfamily of HERV-K and is integrated into the periphilin gene as the truncated form, 5'LTR-*gag*-*pol*-3'LTR. PCR and RT-PCR approaches indicated that the insertion of various retrotransposable elements in a common ancestor genome may make different transcript variants in different primate species. Examination of the HERV-W elementrevealed that *env* fragments were detected on human chromosomes 1, 3-7, 12, 14, 17, 20, and X, whilst the *pol* fragments were detected on human chromosomes 2-8, 10-15, 20, 21, X, and Y. Bioinformatic blast search showed that almost full-length of the HERV-W family was identified on human chromosomes 1-8, 11-15, 17, 18, 21, and X. Expression analysis of HERV-W genes (*gag*, *pol*, and *env*) in human tissues by RT-PCR indicated that *gag* and *pol* were expressed in specific tissues, whilst *env* was constituitively expressed in all tissues examined. DNA sequence based phylogenetic analysis indicated that the *gag*, *pol* and *env* genes have evolved independently during primate evolution. It will thus be of considerable interest to expand the current HERV gene information of various primates and disease tissues.

Keywords: HERV elements, diseases, gene information, evolution, primates

*Corresponding author: Email khs307@pusan.ac.kr, Tel +82-51-510-2259, Fax +82-51-581-2962
Accepted 21 November 2005

Introduction

Human endogenous retroviruses (HERVs) and long terminal repeat (LTR) like elements are dispersed over 8% of the whole human genome (International Human Genome Sequencing Consortium, 2001). There are at least 22 distinct HERV families within the human genome (Tristem, 2000). HERVs originate by germ-cell infection of the exogenous retrovirus during primate evolution (Lower *et al.*, 1996). Full length retroviral sequences may interact with cellular oncogenes (Varmus, 1982) and retroviral LTR elements have the capacity to exert a regulatory influence as promoters and enhancers of cellular genes (Akopov *et al.*, 1998; Baust *et al.*, 2000). Elucidation of the HERV elements should provide information about fundamental cellular activities and the pathogenesis of multifactorial diseases such as cancer and autoimmune disease (Nakagawa *et al.*,1997; Mayer, 2001).

The HERV-K family of human endogenous retroviral sequences was originally cloned from Syrian hamster intra-cisternal A type particles (Ono *et al.*, 1986), has homology to mouse mammary tumour virus, and includes sequences that are expressed in normal placenta and leukemic cells (Simon *et al.*,1994). HERV-K elements probably entered the primate genome after the split of New World monkeys in the Oligocene era, 33-40 million years ago (Steinhuber *et al.*, 1995). It is estimated that there are now approximately 25,000 copies of HERV-K LTRs in the human genome, and they have randomly transposed across the chromosomes in the course of human evolution (Leib-Mösch *et al.*, 1993). The possibility that some of this increase has occurred recently was suggested by the presence of a functional integrase in HERV-K sequences (Kitamura *et al.*, 1996). The ratio of synonymous to nonsynonymous substitutions suggests that evolutionary selection has recently been or is still operating (Zsiros *et al.*, 1998). Medstrand and Mager (1998) have demonstrated that some clusters of

HERV-K LTRs show low (average 1.5%) divergence; in comparisons across primate species, members of their cluster 9 are specific to the *Homo sapiens*. HERV-K LTRs bind host cell nuclear proteins and have the potential to activate neighboring genes (Akopov *et al.*, 1998). A possible role of HERV-K elements in disease has been considered in relation to insulin dependent diabetes mellitus (Conrad *et al.*, 1997), seminoma (Sauter *et al.*, 1995) and HERV-K-T47D mammary carcinoma (Seifarth *et al.*, 1998). Recently, HERV-M (HERV-K22), the oldest beta retrovirus, was identified in the human genome. The HERV-M proviral sequences were detected in hominoids, Old and New World monkeys, but not in prosimians. Compared to the other HERV-K families, most of their structural genes (*gag* and *env*) were severely truncated (Lavie *et al.*,2004).

SINE-R retroposons are derived from the endogenous retrovirus HERV-K family. SINE-R11, 14, 19 were isolated by colony blot hybridization using the 3′ long terminal repeat and small upstream flanking regions of HERV-K10 as probe (Ono *et al.*, 1987). These elements entered the genome of hominoid primates after the split of Old World monkeys in the Oligocene period (Kim *et al.*, 1999a,c,d). The SINE-R.C2 element was discovered in relation to a variable number of tandem repeat (VNTR) sequence within the third intron of the gene for C2, the second component of complement, located in class III of the major histocompatibility complex on the short arm of human chromosome 6 (Carroll *et al.*, 1984; Zhu *et al.*, 1992). By Southern blot analysis, SINE-R.C2 has been found to be confined to *Homo sapiens* (Zhu *et al.*,1994). Within the Xq21.3 block of homology to Yp11.2 that was created in the course of hominid evolution (Laval *et al.*, 1998), two retroposons (HS307 and HS408) with 92 to 96% homology to SINE-R.C2 were reported (Kim *et al.*, 1999c). SINE elements have generally arisen either from 7SL RNA or from tRNA, have been amplified many times in the genome, and have become fixed at certain stages of evolution. They share three prime sequences with LINE elements, which probably have an earlier evolutionary origin, and upon which they may depend for the reverse transcriptase activity necessary for retroposition (Okada *et al.*, 1997).

A new HERV-W family has been identified by successive overlapping cDNA clones from multiple sclerosis virion (Komurian-Pradel *et al.*,1999). The HERV-W has been one of the most widespread investigated HERV elements, since the isolation of a HERV-W-related retrovirus (MSRV) from retroviral particles produced by cell cultures from patients with multiple sclerosis (Perron *et al.*, 1997). The 7q21-22 region of human chromosome 7 has been identified as potentially involved in genetic susceptibility to multiple sclerosis (Charmley *et al.*,1991), contains a RGH proviral copy (HERV-H/RGH) and a complete HERV-W proviral copy (HERV-W/7q). The *env* gene of HERV-W/7q encods a protein expressed in fetal tissues and in placenta preferentially (Alliel *et al.*, 1998). Recently, the HERV-W *env* protein, syncytin, was found to be involved in placental morphogenesis (Mi *et al.*, 2000). The major sites of syncytin expression were placental syncytiotrophoblasts and multinucleated cells that originated from fetal trophoblasts. Here I review gene information and disease implication of the retroelements, HERV-K, HERV-M, SINE-R, and HERV-W, in humans and primates.

HERV-K LTR and SINE-R elements and implications for human diseases

The long terminal repeat (LTR) elements of the human endogenous retrovirus HERV-K have been found to be present in hominoids and Old World monkeys (Zhu *et al.*, 1994), whilst HERV-K LTR-like elements were identified in New World monkeys (Kim *et al.*, 1999a,e). In the human genome, several thousand copies of HERV-K LTR elements are found with an essentially random distribution. Some of the LTR elements have already been identified using human monochromosome and genomic DNAs (Kim and Crow, 1999; Kim *et al.*, 2000), whilst additional human-specific HERV-K LTR elements were identified from the GenBank database (Medstrand and Mager, 1998) (AC002350, AC002400, AC002508, etc.). All of these LTR elements and the new HERV-K LTR elements that were transcribed in human brain tissue were used to construct a phylogenetic tree (Fig. 1). The HERV-K LTR elements were mainly divided into two groups through evolutionary divergence, although note that this is only weakly supported. The human-specific HERV-K LTR element (AL034407) has proliferated within human chromosome region Xq26 recently (Kim *et al.*, 2000). Similarly, other HERV-K LTR elements (AC002350, AC002400, AC002508, L47334, U47924, Z80898) are only detected in the human genome (Medstand and Mager, 1998). Some HERV-K LTR elements (HKL-B7, HKL-B8, HKL-B10), identified in human brain cDNA, were closely related to human-specific HERV-K LTR elements, indicating that they may be related to brain function or neuropsychiatric disease.

SINE-R retroposons, derived from the HERV-K LTR elements (Ono *et al.*, 1987; Zhu *et al.*, 1992, 1994), have been found to be hominoid-specific by PCR analysis (Kim *et al.*, 1999a,c). By BLAST search, cDNA prepared from the brain of a patient with schizophrenia (GenBank, accession no. AA772777) was found to have sequences

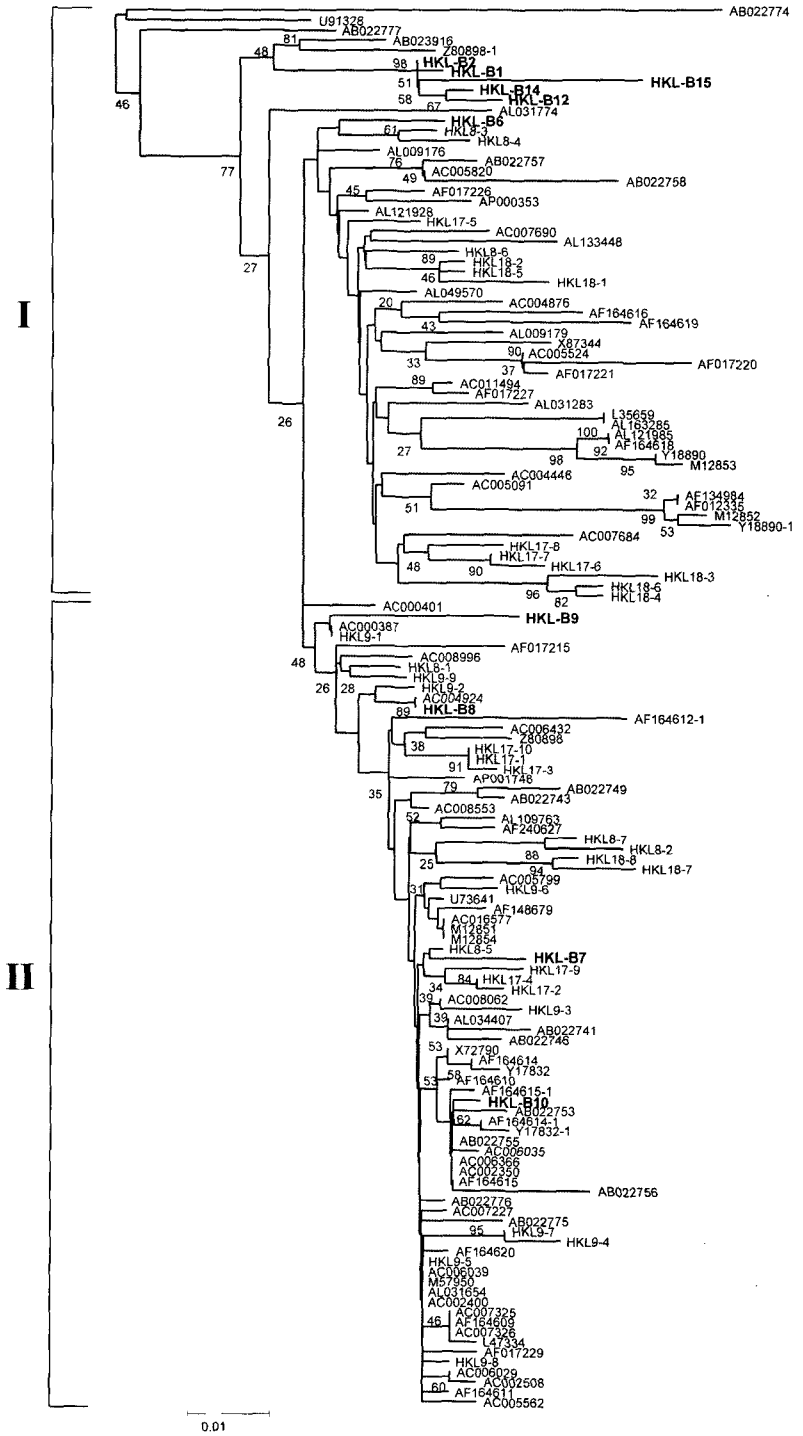

Fig. 1. Phylogenetic tree obtained by neighbor-joining method for the HERV-K LTR family. Branch lengths are proportional to the distances between the taxa. The values at branch-points indicate the percentage support for a particular node after 100 bootstrap replications. The accession numbers of HERV-K LTR elements were obtained from the GenBank database. The HERV-K LTR elements (HKL-B1, HKL-B2, HKL-B6, HKL-B7, HKL-B8, HKL-B9, HKL-B10, HKL-B12, HKL-B14, HKL-B15) presented as bold letters were derived from human brain cDNA library. Those LTR elements were amplified by the specific primer pair HS45 (5′-GTATTGTCCAAGGTTTCTCCC-3′, bases 8541-8561) and DS14 (5′-GTGCTGTGCTTTTGGATATGC-3′, bases 8859-8879) from the human endogenous retroviral element, HERV-K10 (GenBank, accession no. M14123).

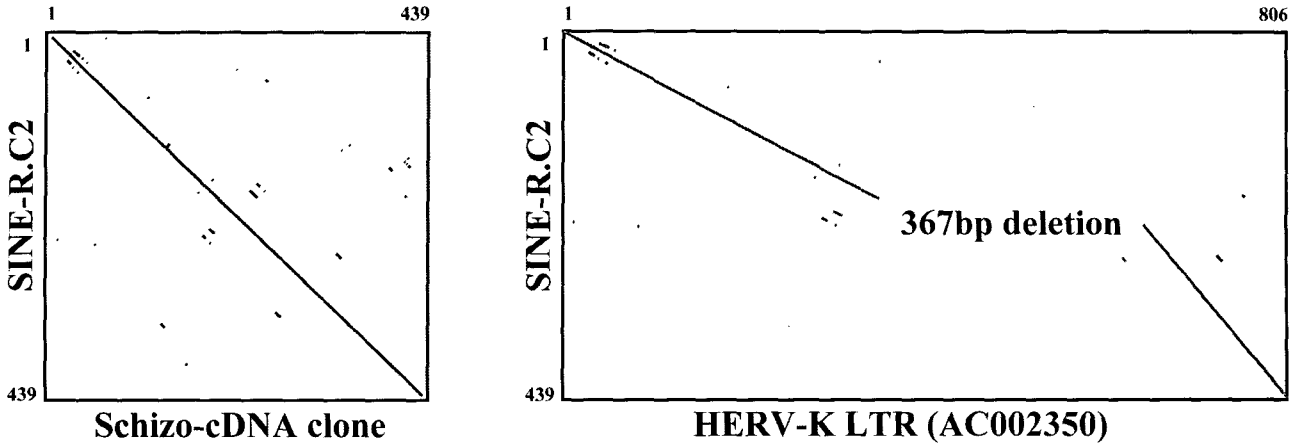

Fig. 2. Dot plot matrix of SINE-R retroposon relative to sequences from the schizophrenic brain cDNA (schizo-cDNA) clone and HERV-K LTR element (GenBank, accession no. AC002350)

containing a high degree of homology with the SINE-R11, 14, 19 (Ono *et al.*, 1987), and with SINE-R.C2 (Zhu *et al.*, 1994). HS307 and HS408 retroposons are located in Xq21.3 region of the human genome, transposed to the Yp11 after the separation of the chimpanzee and human lineages (Lambson *et al.*,1992). Both HS307 and HS408 retroposons show a high degree of sequence similarity with the schizophrenic brain cDNA derived from the postmortem brain tissue from the frontal cortex of an individual suffering from schizophrenia who committed suicide at the age of 34 years (Yee *et al.*,1998). Interestingly, Laval *et al.*(1998) showed evidence of linkage of both psychosis and degree of handedness (an index of cerebral asymmetry) to Xq21.3 region. Therefore, in order to explore the retroviral elements in Xq21.3 region, a YAC clone panel derived from the Xq21.3 region was examined and four HERV-K LTR elements found and characterized (Kim and Crow, 2001). Two of those elements (K-X10-5 and K-X13-1) have a high degree of sequence similarity to the human specific HERV-K LTRs. Elements of the SINE-R.C2 class have similarities to the clusters of the HERV-K LTR family (Fig. 2). SINE-R retroposons have a polypurine tract that serves as a primer binding site for plus-strand DNA synthesis from retrovirally mediated reverse transcription (Ono *et al.*,1987). In addition, they have a hormone responsive element and enhancer core of the HERV-K LTR element although the promoter is deleted. These SINE-R retroposons and HERV-K LTR elements could contribute to the structural change or genetic variation associated with neuropsychiatric disease and primate evolution. Akopov *et al.* (1998) noted that HERV-K sequences have the capacity to modify the expression of neighboring genes, and suggested that such modifications may have been acquired in the course of human evolution. The HERV-K-T47D-related LTR element mediates polyadenylation of cellular transcripts (Baust *et al.*, 2000). Two HERV-K LTR elements have been detected in the human histocompatibility complex locus HLA-DQ (Kambhu *et al.*, 1990). One LTR element (DQ-LTR3) of the HERV-K family at the HLA-DQB1 locus has been associated with rheumatoid arthritis (Seidl *et al.*, 1999). An element described as almost identical to SINE-R.C2 derived from the HERV-K LTR element is the likely cause of Fukuyama-type muscular dystrophy (Kobayashi *et al.*, 1998). Clearly, these HERV-K LTR and SINE-R elements deserve further investigation as potential leads to various human diseases.

Molecular characterization of HERV-M and *PPHLN1* hybrid transcripts

The HERV superfamily is one of the well-known repeat elements in the human genome, and the various families were phylogenetically characterized after completion of the human genome project (Tristem, 2000). The HERV-M family (HERV-K22) was recently identified and the genomic structure, localization, and phylogenetic relationship of HERV-M elements are reviewed elsewhere (Lavie *et al.*, 2004). However, the HERV-M located within the *PPHLN1* gene has a disrupted open reading frame yet seems to overcome the situation of functionlessness in the human genome with a gene-like transcription strategy during primate evolution. Generally, HERV families do not use the donor and acceptor sites like a eukaryoticcell, their PBS (promoter binding site) and polyadenylation signals are used during

their transcription. Phylogenetic analysis suggests that more than 45 million years was sufficient for changing of the HERV-M from a viral to cellular gene with the accumulation of mutations.

The HERV-M could make various hybrid transcripts for the *PPHLN1* gene in coding regions with an alternative splicing pattern like a normal eukaryotic gene transcript by data base screening (EST and Refseq mRNA data base) with a HERV-M consensus sequence. The HERV-M sequence was expressed in human brain, prostate, testis, heart, kidney, liver, lung, placenta, skeletal muscle, spleen, thymus, and uterus, but not in kidney, liver or lung. However, RT-PCR amplification suggested it was not detected in as many of the tested tissues in the crab-eating and squirrel monkey being absent in testis, kidney, liver, lung, and spleen (Huh *et al.*, 2005).

Insertion and expression events of the HERV-M elements seem to occur independently. Although the insertion of HERV-M happened before the divergence of New and Old World monkeys, expression of HERV-M in the *PPHLN1* gene has been found only in humans and crab-eating and squirrel monkeys. After the integration of HERV-M, a series of repeat element insertions may affect the instability of the primate genome, and subsequently the gene transcription mechanisms. Although the exact mechanism of the *PPHLN1* gene transcription related to HERV-M is not yet clear, elucidation of such events are likely to provide be a good research model of HERV exonization.

The *PPHLN1* gene harbors different repeat elements of HERV-M, *Alu*sq, and L1ME4a within the human chromosome 12q12 region. The insertion event of HERV-M occurred in New World monkeys approximately 50 million years ago, and *Alu*sq and L1ME4a were inserted into the mammalian genome before the divergence of simian and prosimian lineages about 60-70 million years ago. These data indicated that lineage specific duplication events could have occurred independently.

Bioinformatic analysis and genomic features of the HERV-W family

The HERV-W family has been one of the most broadly investigated HERVs during the last few years. It consists of an estimated 30 to 100 provirus copies per haploid human genome and is phylogenetically related to ERV-9 and RTLV-H families. HERV-W gag and *pol* open reading frames (ORFs) of the family have acquired frameshifts and stop codons by deletions/ insertions, whereas a complete *env* ORF encoding an envelope was found on the cl.PH74 placental cDNA clone and on the genomic RG083M05 clone (Blond *et al.*, 1999). HERV-W may have a chimeric genome structure since the murine type C and simian type D retroviruses show strong homology with the *pol* and *env* genes, respectively. The HERV-W on chromosome 7 has 100% identity to syncytin cDNA, which is expressed specifically in the syncytiotrophoblast layer of the placenta (Alliel *et al.*, 1998; Mi *et al.*, 2000). This observation implicates the potential HERV-W family involvment in human placental development. That product of the HERV-W *env* gene is a highly fusogenic membrane glycoprotein that induces the formation of syncytia on interaction with the type D mammalian retrovirus receptor (Blond *et al.*, 2000) supports this notion as well as suggesting a direct role in human trophoblast cell fusion and differentiation (Frendo *et al.*, 2003). Recently, the HERV-W gene product was shown to block infection by exogenous retroviruses, suggesting that expressed HERV genes could have beneficialfunctions to the host (Ponferrada *et al.*, 2003). The possibility that they have relevance to pathogenic capacity could be subjected to change in primate evolution and could be associated with variation between different tissues and gene expression.

The HERV-W families are related to multiple sclerosis-associated retrovirus (MSRV) sequences identified in particles recovered from monocyte cultures from patients with multiple sclerosis.Retroviral sequences belonging to HERV-W family were also identified in brain of individual with schizophrenia, increasing the possibility that HERV-W family could be involved in human brain disease (Karlsson *et al.*, 2001). Such HERV-W env families are located on human chromosomes 1, 3, 4, 5, 6, 7, 12, 14, 17, 20, and X (Kim and Lee, 2001), while HERV-W pol families are located on human chromsomes 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 20, 21, X, and Y (Kim, 2001). Within the HERV-W family, *env* and *pol* genes showed a high degree of nucleotide sequence similarity (91.6-99.6%), but translation of these sequences revealed the presence of both frameshifts and premature termination codons caused by deletion/ insertionor point mutations in some clones (Kim *et al.*, 1999b). Outside of humans, the *pol* and *env* gene sequences of the HERV-W family were detected by PCR in hominoids and Old World monkeys, but not in New World monkeys. The HERV-W family has been phylogenetically analyzed from the databases and divided into three different subfamilies of the human genome derived HERV-W group (Costas, 2002). The average divergence between sequences for each of the subfamilies indicated that most of the HERV-W elements were inserted within the genome during a short period of evolutionary time.

To date, BLAST search of the DDBJ/EMBL/GenBank

databases for HERV-W family reveals 140 sequences representing 39 HERV-W proviruses, 40 full-length HERV-W retroposons, and 61 truncated HERV-W retrosequences (Costas, 2002). The number of identified HERV-W related fragments comprises of at least 70 copies for *gag* and 30 copies for *env* per haploid genome, and is correlated with the increase in complexity from *env* to *gag* and pro regions as previously described (Voisset *et al.*, 2000). From the 15 HERV-W *env* groups found within the human genome (Kim and Lee, 2001),one of them, clone W-7-1 from chromosome 7, showed identical sequences with those of brain, testis, kidney, lung, placenta, skeletal muscle, and uterus of normal tissues (WE1-5, WE3-8, WE5-4, WE7-2, WE8-1, WE9-3, WE10-14), suggesting that they are actively expressed in various normal tissues (Yi *et al.*, 2004). Other HERV families are expressed in normal tissues and cancer cells. For example, HERV-E family on human chromosome 17q11 was expressed specifically in human pancreas and thyroid gland with two major transcripts, 3.3 and 4.1kb in size (Shiroma *et al.*, 2001). HERV-H *env* transcripts have been detected in both T-cell leukemia cell lines and lymphocytes from healthy blood donors by RT-PCR (Lindeskog and Blomberg, 1997). Recently, the tissue-specific expression of two HERVs transcripts (HERV-K *cORF/rec* and HERV-R *env*) was found in normal human tissues (kidney, tongue, heart, liver, central nervous system) during embryogenesis, suggesting a possible role in the development and differentiation of human tissues (Andersson *et al.*, 2002). In a comparative analysis between normal and cancer cells, the HERV-W gag gene was expressed in normal brain tissue, whereas no expression but not in PFSK-1 cancer cells derived from brain tissue. The HERV-W *pol* gene was also expressed in prostate tissue, but not in PC3 cancer cells derived from prostate tissue (Yi *et al.*, 2004). In the case of the HERV-E *env* gene, it was actively expressed in prostate carcinoma tissues and cell lines, but not in normal prostate tissues and cells as detected by RT-PCR, RNA *in situ* hybridization, and Northern blot hybridization (Wang-Johanning *et al.*, 2003). No expression of the HERV-E *env* gene was detected in breast tissues, but in contrast, HERV-K *env* transcripts were detected in most breast cancer cell lines and many breast tumor tissues (Wang-Johanning *et al.*, 2002). Such HERV-K *gag* transcripts were found in gonocytes of all genadoblastomas and in neoplastic germ cell tumours whilst the transcriptional activity of the HERV-K *gag* gene was 5- to 10-fold higher in blood cells of leukemia patients than in normal cells (Depil *et al.*, 2002).

Whilst products of the HERV-W *env* gene are implicated in a direct role in human trophoblast cell fusion and differentiation (Frendo *et al.*, 2003) and, via blocking the infection by an exogenous retrovirus, mayhave beneficial function to the host (Ponferrada, 2003), the relationship of HERV elements in general and human diseases remains contraversial. Reports of detection of various HERV mRNA, HERV proteins, and even HERV particles in human diseases, especially cancer, are becoming numerous, but remain correlative. The HERV-W *env* gene sequences have been characterizedin some human cancer cell lines to examine the role of HERV in human cancer (Yi *et al.*, 2002) and to investigate the comparative analysis of HERV-W families between humans and monkeys to understand their tissue distribution, evolution, and phylogenetic relationships (Yi *et al.*, 2003). In bioinformatic analysis, a BLASTN query on the EST (expressed sequence tag) database with HERV-W showed multiple transcripts in various human tissues such as bladder, brain, breast, colon, hypothalamus, kidney, liver, lung, ovary, parathyroid, placenta, prostate, skin, stomach, testis, and uterus (data not shown), supporting that HERV-W families could have a biological role in multiple human tissues. Taken together, these data suggest that some recently proliferated and expressed HERV-W families are active in the genome of tissues and cancer cells, and that they therefore could have a functional role in human tissues and cancer cells.

Acknowledgement

This study was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (A050337).

References

Akopov, S.B., Nikolaev, L.G., Khil, P.P., Lebedev, Y.B., and Sverdlov, E.D. (1998). Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. *FEBS Lett.* 421, 229-233.

Alliel, P.M., Perin, J.P., Pierig, R., Nussbaum, J.L., Menard, A., and Rieger, F. (1998). Endogenous retroviruses and multiple sclerosis. II. HERV-7q. *C. R. Acad. Sci. III* 321, 857-863.

Andersson, A.C., Venables, P.J., Tonjes, R.R., Scherer, J., Eriksson, L., and Larsson, E. (2002). Developmental expression of HERV-R (ERV3) and HERV-K in human tissue. *Virology* 297, 220-225.

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

Blond, J.L., Beseme, F., Duret, L., Bouton, O., Bedin, F., Perron, H., Mandrand, B., and Mallet, F. (1999). Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. *J. Virol.* 73, 1175-1185.

Carroll, M.C., Campbell, R.D., Bentley, D.R., and Porter, R.R. (1984). A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. *Nature* 307, 237-241.

Charmley, P., Beall, S.S., Concannon, P., Hood, L., and Gatti, R.A. (1991). Further localization of a multiple sclerosis susceptibility gene on chromosome 7q using a new T cell receptor beta-chain DNA polymorphism. *J. Neuroimmunol.* 32, 231-240.

Conrad, B., Weissmahr, R.N., Boni, J., Arcari, R., Schupbach, J., and Mach, B. (1997). A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell 90,* 303-313.

Costas, J. (2002). Characterization of the intragenomic spread of the human endogenous retrovirus family HERV-W. *Mol. Biol. Evol.* 19, 526-533.

Depil, S., Roche, C., Dussart, P., and Prin, L. (2002). Expression of a human endogenous retrovirus, HERV-K, in the blood cells of leukemia patients. *Leukemia* 16, 254-259.

Frendo, J.L., Olivier, D., Cheynet, V., Blond, J.L., Bouton, O., Vidaud, M., Rabreau, M., Evain-Brion, D., and Mallet, F. (2003). Direct involvement of HERV-W Env glycoprotein in human trophoblast cell fusion and differentiation. *Mol. Cell. Biol.* 23, 3566-3574.

Huh, J.W., Kim, T.H., Yi, J.M., and Kim, H.S. (2005). Molecular evolution of the periphilin gene in relation to human endogenous retrovirus M element. *J. Mol. Evol.*, in press.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Karlsson, H., Bachmann, S., Schroder, J., McArthur, J., Torrey, E.F., and Yolken, R.H. (2001). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proc. Natl. Acad. Sci. USA* 98, 4634-4639.

Kambhu, S., Falldorf, P., and Lee, J.S. (1990). Endogenous retroviral long terminal repeats within the HLA-DQ locus. *Proc. Natl. Acad. Sci. USA* 87, 4927-4934.

Kim, H.S. (2001). Sequence and phylogeny of HERV-W *pol* fragments. *AIDS Res. Hum. Retroviruses* 17, 1665-1671.

Kim, H.S., Choi, J.Y., Lee, W.H., Jang, K.L., Park, W.H., Moon, D.H., Takenaka, O., and Hyun, B.H. 2000. Phylogenetic analysis of HERV-K LTR family in human chromosome Xq26 and New World monkeys. *J. Life. Sci.* 10, 32-36.

Kim, H.S. and Crow, T.J. (1999). Presence and phylogeneticanalysis of HERV-K LTR on human X and Y chromosomes: evidence for recent proliferation. *Genes Genet. Syst.* 74, 267-270.

Kim, H.S. and Crow, T.J. (2001). Cloning and phylogeny of endogenous retroviral elements belonging to the HERV-K LTR in cDNA library of human fetal brain and Xq21.3 region linked to psychosis. *Korean J. Genet.* 23, 129-134.

Kim, H.S. and Lee, W.H. (2001). Human endogenous retrovirus HERV-W family: chromosomal localization, identification, and phylogeny. *AIDS Res. Hum. Retroviruses* 17, 643-648.

Kim, H.S., Takenaka, O., and Crow, T.J. (1999a). Cloning and nucleotide sequence of retroposons specific to hominoid primates derived from an endogenous retrovirus (HERV-K). *AIDS Res. Hum. Retroviruses* 15, 595-601.

Kim, H.S., Takenaka, O., and Crow, T.J. (1999b). Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. *J. Gen. Virol.* 80, 2613-2619.

Kim, H.S., Wadekar, R.V., Takenaka, O., Winstanley, C., Mitsunaga, F., Kageyama, T., Hyun, B.H., and Crow, T.J. (1999c). SINE-R.C2 (a *Homo sapiens* specific retroposon) is homologous to cDNA from post-mortem brain in schizophrenia and to two loci in the Xq21.3/Yp block linked to handedness and psychosis. *Am. J. Med. Genet.* (*Neuropsychiat. Genet.*) 88, 560-566.

Kim, H.S., Wadekar, R.V., Takenaka, O., Hyun, B.H., and Crow, T.J. (1999d). Phylogenetic analysis of a retroposon family in African great apes. *J. Mol. Evol.* 49, 699-702.

Kim, H.S., Wadekar, R.V., Takenaka, O., Hyun, B.H., and Crow, T.J. (1999e). Phylogenetic analysis of HERV-K LTR-like elements in primates: presence in some New World monkeys and evidence of recent parallel evolution in these species and in *Homo sapiens*. *Arch. Virol.* 144, 2035-2040.

Kitamura, Y., Ayakuwa, T., Ishikawa, T., Kanda, T., and Yoshiike, K. (1996). Human endogenous retrovirus K10 encodes a functional integrase. *J. Virol.* 70, 3302-3306.

Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Lida, E., Nomura, Y., Segawa, M., Yoshioka, M., Saito, K., Osawa, M., Hamano, K., Sakakihara, Y., Nonaka, I., Nakagome, Y., Kanazawa, I., Nakamura, Y., Tokunaga, K., and Toda, T. (1998). An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* 394, 388-392.

Komurian-Pradel, F., Paranhos-Baccala, G., Bedin, F., Ounanian-Paraz, A., Sodoyer, M., Ott. C., Rajoharison, A., Garcia, E., Mallet, F., Mandrand, B., and Perron, H.

(1999). Molecular cloning and characterization of MSRV-related sequences associated with retrovirus-like particles. *Virology* 260, 1-9.

Lambson, B., Affara, N.A., Mitchell, M., and Ferguson-Smith, M.A. (1992). Evolution of DNA sequence homologies between the sex chromosomes in primate species. *Genomics* 14, 1032-1042.

Laval, S.H., Dann, J., Butler, R.J., Loftus, J., Rue, J., Leask, S.J., Bass, N., Comazzi, M., Vita, A., Nanko, S., Shaw, S., Peterson, P., Shields, G., Smith, A.B., Stewart, J., DeLisi, L.E., and Crow, T.J. (1998). Evidence for linkage to psychosis and cerebral asymmetry (relative hand skill) on the X chromosome. *Am. J. Med. Genet.* (*Neuropsychiat. Genet.*) 81, 420-427.

Lavie, L., Medstrand, P., Schempp, W., Meese, E., and Mayer, J. (2004). Human endogenous retrovirus family HERV-K(HML-5): status, evolution, and reconstruction of an ancient beta-retrovirus in the human genome. *J. Virol.* 78, 8788-8798.

Leib-Mösch, C., Haltmeier, M., Werner, T., Geigl, E.M., Brack-Werner, R., Francke, U., Erfle, V., and Hehlmann, R. (1993). Genomic distribution and transcription of solitary HERV-K LTRs. *Genomics* 18, 261-269.

Lindeskog, M. and Blomberg, J. (1997). Spliced human endogenous retroviral HERV-H *env* transcripts in T-cell leukaemia cell lines and normal leukocytes: alternative splicing pattern of HERV-H transcripts. *J. Gen. Virol.* 78, 2575-2585.

Lower, R., Lower, J., and Kurth, R. (1996). The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci. USA* 93, 5177-5184.

Mayer, J. (2001). Status of HERV in human cells: expression and coding capacity of human proviruses. *Development. Biol.* 106, 439-441.

Medstand, P. and Mager, D. L. (1998). Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* 72, 9782-9787.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., Keith, J.C., and McCoy, J.M. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785-789.

Nakagawa, K., Brusic, V., McColl, G., and Harrison, L.C. (1997). Direct evidence for the expression of multiple endogenous retroviruses in the synovial compartment in rheumatoid arthritis. *Arthritis and Rheumatism* 40, 627-638.

Okada, N., Hamada, M., Ogiwara, I., and Ohshima, K. (1997). SINEs and LINEs share common 3' sequences: a review. *Gene* 205, 229-243.

Ono, M., Yasunaga, T., Miyata, T., and Ushikubo, H. (1986). Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *J. Virol.* 60, 589-598.

Ono, M., Kawakami, M., and Takezawa, T. (1987). A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res.* 21, 8725-8737.

Perron, H., Garson, J.A., Bedin, F., Beseme, F., Paranhos-Baccala, G., Komurian-Pradel, F., Mallet, F., Tuke, P.W., Voisset, C., Blond, J.L., Lalande, B., Seigneurin, J.M., and Mandrand, B. (1997). Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. The Collaborative Research Group on Multiple Sclerosis. *Proc. Natl. Acad. Sci. USA* 94, 7583-7588.

Ponferrada, V.G., Mauck, B.S., and Wooley, D.P. (2003). The envelope glycoprotein of human endogenous retrovirus HERV-W induces cellular resistance to spleen necrosis virus. *Arch. Virol.* 148, 659-675.

Sauter, M., Schommer, S., Kremmer, E., Remberger, K., Dolken, G., Lemm, L., Buck, M., Best, B., Neumann-Haefelin, N., and Mueller-Lantzsch, M. (1995). Human endogenous retrovirus K10: expression of gag protein and detection of antibodies in patients with seminomas. *J. Virol.* 69, 414-421.

Seidl, C., Donner, H., Petershofen, E., Usadel, K.H., Seifried, E., Kaltwasser, J.P., and Badenhoop, K. (1999). An endogenous retroviral long terminal repeat at the HLA-DQB1 gene locus confers susceptibility to rheumatoid arthritis. *Hum. Immunol.* 60, 63-68.

Seifarth, W., Baust, C., Murr, A., Skadny, H., Krieg-Schneider, F., Blusch, J., Werner, T., Hehlmann, R., and Leib-Mosch, C. (1998). Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. *J. Virol.* 72, 8384-8391.

Shiroma, T., Sugimoto, J., Oda, T., Jinno, Y., and Kanaya, F. (2001). Search for active endogenous retroviruses: identification and characterization of a HERV-E gene that is expressed in the pancreas and thyroid. *J. Hum. Genet.* 46, 619-625.

Simon, M., Haltmeier, M., Papakonstantinou, G., Werner, T., Hehlmann, R., and Leib-Mosch, C. (1994). Transcription of HERV-K-related LTRs in human placenta and leukemic cells. *Leukemia* S1, 12-17.

Steinhuber, S., Brack, M., Husmann, G., Schwelberger, H.,Dierich, M.P., and Vogetseder, W. (1995). Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates. *Hum. Genet.* 96, 188-192.

Tristem, M. (2000). Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.* 74, 3715-3730.

Varmus, H. E. (1982). Form and function of retroviral proviruses. *Science* 216, 812-820.

Voisset, C., Bouton, O., Bedin, F., Duret, L., Mandrand, B., Mallet, F., and Paranhos-Baccala, P. (2000). Chromosomal distribution and coding capacity of the human endogenous retrovirus HERV-W family. *AIDS Res. Hum. Retroviruses* 16, 731-740.

Wang-Johanning, F., Frost, A.R., Jian, B., Epp, L., Lu, D.W., and Johanning,G.L. (2003). Quantitation of HERV-K *env* gene expression and splicing in human breast cancer. *Oncogene* 22, 1528-1535.

Wang-Johanning, F., Frost, A.R., Johanning, G.L., Khazaeli, M.B., LoBuglio, A.F., Shaw, D.R., and Strong, T.V. (2002). Expression of human endogenous retrovirus k envelope transcripts in human breast cancer. *Clinical Cancer Res.* 7, 1553-1560.

Yee, F., Johnston, N.L., Leister, F., Li, S., Ross, C.A., Torrey, E.F., Yolken, R.H., and Stanley Neuropathology Consortium. (1998). Differential expression of viral and virus-associated RNA transcripts of the brains of individuals with schizophrenia and bipolar disorder. *Schzophr. Res.* 29, 92-92.

Yi, J.M., Takenaka, O., and Kim, H.S. (2003). Molecular characterization and phylogenetic relationship of HERV-W family in the *Macaca fuscata*. *Arch. Virol.* 148, 1613-1622.

Yi, J.M., Kim, H.M., Lee, W.H., and Kim, H.S. (2002). Molecular cloning and phylogenetic analysis of new human endogenous retrovirus HERV-W family in cancer cells. *Curr. Microbiol.* 44, 216-220.

Yi, J.M., Kim, H.M., and Kim, H.S. (2004). Expression of the human endogenous retrovirus HERV-W family in various human tissues and cancer cells. *J. Gen. Virol.* 85, 1203-1210.

Zhu, Z.B., Hsieh, S.L., Bentley, D.R., Campbell, R.D., and Volanakis, J.E. (1992). A variable number of tandem repeats locus within the human complement C2 gene is associated with a retroposon derived from a human endogenous retrovirus. *J. Exp. Med.* 175, 1783-1787.

Zhu, Z. B., Jian, B., and Volanakis, J.E. (1994). Ancestry of SINE-R.C2 a human specific retroposon. *Hum. Genet.* 93, 545-551.

Zsiros, J., Jebbink, M.F., Lukashov, V.V., Voute, P.A., and Berkhout, B. (1998). Evolutionary relationships within a subgroup of HERV-K related human endogenous retroviruses. *J. Gen. Virol.* 79, 61-70