

# Expression and phylogenetic analyses of human endogenous retrovirus HC2 belonging to the HERV-T family in human tissues and cancer cells

Joo-Mi Yi · Heui-Soo Kim

Received: 26 September 2006 / Accepted: 26 December 2006 / Published online: 3 February 2007  
© The Japan Society of Human Genetics and Springer 2007

**Abstract** Recently, a new HERV-T family, representative of the HERV-S71 and HERV-HC2 family, was identified using a screen for envelope genes and a computer-assisted database search. Here, we investigate expression of *pol* fragments of HERV-HC2 belonging to the HERV-T family in various human tissues and cancer cells. The *pol* gene was expressed in nearly all human tissues examined and in all cancer cell lines. Expression analyses suggest that the *pol* gene of HERV-HC2 family is more actively transcribed in human cancer cells than in normal tissues, suggesting a functional role during carcinogenesis. Phylogenetic analysis of the HERV-HC2 *pol* family revealed three groups (I, II, and III) generated through evolutionary divergence during primate evolution, indicating that they were integrated into primate genomes approximately 56 million years (MY) ago and have evolved at a rate of 0.2% nucleotide differences per MY. Our data might contribute to an understanding of the information on the transcriptional and pathological potential of the HERV-T family in human disease, including cancer.

**Keywords** HERV-HC2 · Expression · Phylogeny · Human tissues · Cancer cells · Evolution

J.-M. Yi  
Cancer Biology Division, The Sidney Kimmel  
Comprehensive Cancer Center, Johns Hopkins University,  
Baltimore, MD 21231, USA

J.-M. Yi · H.-S. Kim (✉)  
Division of Biological Sciences, College of Natural Sciences,  
Pusan National University, Pusan 609-735, South Korea  
e-mail: khs307@pusan.ac.kr

## Introduction

Approximately 8% of the human genome is comprised of human endogenous retroviruses (HERVs) and other long terminal repeat (LTR)-like elements (International Human Genome Sequencing Consortium 2001). These sequences can be easily recognized because they consist of at least three genes—*gag* (coding structural proteins), *pol* (viral enzymes), and *env* (surface envelope proteins)—as well as LTRs. HERVs have been subjected to many amplification and transposition events during evolution resulting in a widespread distribution of complete or partial retroviral sequences. HERVs represent the remnants of ancestral retroviral infections that were integrated into the germline of primates 35–45 million years (MY) ago (Sverdlov 2000). It was recently reported that, using phylogenetic analyses based on either the highly conserved reverse transcriptase (RT) domain of the *pol* gene or the transmembrane (TM) moiety of the envelope gene, at least 26 HERV families with distinct retroviral lineages reside within the human genome (Tristem 2000; Benit et al. 2001). Subsequent retrotransposition event and recombination mechanisms amplified these sequences to a high degree during primate evolution. Such retroviral elements have been described in primates and humans (Kabat et al. 1996; Kim and Crow 1999; Shi et al. 1991). Several HERV families with very different copy numbers (1–1,000 copies) and genomic integration have been identified. These different copy numbers could represent either multiple integration events or proviruses that have been amplified by retrotransposition following integration.

It has been suggested that HERVs have played a role in influencing the functional organization of the

human genome (Baltimore 1985). The open reading frames of many extinct HERVs are interrupted by premature stop codons that interrupt the “read through” of genetic information. However, the structural genes of some HERV families are expressed preferentially in human placenta (Kjellman et al. 1999; Mi et al. 2000; Venables et al. 1995) and several cancer cell lines (Armbruster et al. 2002; Lower et al. 1993; Yi et al. 2004). Expression of HERVs could be influenced by the outcome of infections in different ways that can be either beneficial or detrimental to the host. Full-length retroviral sequences may interact with cellular oncogenes (Varmus 1982) and retroviral LTR sequences have the capacity to exert a regulatory influence as promoters and enhancers of cellular genes (Leib-Mosch et al. 1993). A small number of these sequences have acquired a role in regulating gene expression, and some of these may be related to differences between individuals, and to expression of disease (Sverdlov 1998).

The murine leukemia virus (MuLV)-related endogenous retrovirus HC2 was isolated by high-stringency hybridization with the 847 bp *pol* gene. It was found to be more closely related to the human endogenous retrovirus (HERV) S71 than to other MuLV-related retroviruses such as gibbon ape leukemia virus (GaLV), feline leukemia virus (FeLV), baboon endogenous virus (BaEV), and simian sarcoma virus/simian associated sarcoma virus (SSV/SSAV) (Kabát et al. 1996). The genome organization of S71, which is localized on chromosome 18q21 (Brack-Werner et al. 1989), is truncated in the *pol* gene and it carries a 920-bp insertion of a solitary HERV-K LTR element (Haltmeier et al. 1995). The deletion within the *pol* region spans the 3' domain of the protease gene, the complete reverse transcriptase gene, and the 5' end of the tether (Werner et al. 1990). The S71-related retroviral elements pCRTK1 and pCRTK6, which contain intact protease and reverse transcriptase domains that are missing in S71, were identified by polymerase chain reaction (PCR) amplification. Over the past decades a considerable number of studies have classified HERVs using different methodologies. In general, HERV families are often named according to the type of tRNA used by their primer binding site (PBS), which is located downstream of the 5' LTR. For example, the HERV-W, H, and F are primed by the tryptophan, histidine, and phenylalanine, respectively. Despite previous characterization, PBS data were not available for the HERV-S71 family due to the absence of a 5' LTR (Tristem 2000). Recently, a new HERV-T family, HERV-S71 (renamed by Bénit et al. 2001), was identified from human sequences in the databases. Based

on phylogenetic analysis, the HERV-T family is most closely related to the HERV-R family (Benit et al. 2001) and consists of 100 copies within the human genome (de Parseval et al. 2003). Previous studies suggest that the human endogenous retrovirus HERV-HC2 is closely related to HERV-S71, and to the retroviral fragments pCRTK1 and pCRTK6 (Werner et al. 1990; Haltmeier et al. 1995), suggesting that they create a single phylogenetic group. HERV-HC2 is an incomplete provirus containing intact *gag* and *pol* genes and a 3' LTR, but missing the 5' LTR and *env* gene (Kabát et al. 1996). HERV-HC2 is present as a single copy, located at chromosome 10q26, within the haploid human genome (Kim and Crow 1998). We previously reported that HERV-HC2 belongs to HERV-T family, and comprises four members in the human genome (Kim and Crow 1999) and 12 members in the primate genome (Kim et al. 1999a). Additionally, we also reported that the HERV-HC2 *pol* gene is expressed in several tissues of the squirrel monkey (Yi et al. 2003). This information led us to examine expression of HERV-HC2 in normal human tissues and cancer cells.

## Materials and methods

### Cell culture

Human cancer cells (RT4, PFSK-1, BT-474, HCT-116, TE-1, UO-31, Jurkat, HepG2, A549, MCF7, OVCAR-3, MIA-PaCa-2, PC3, LOX-IMVI, AZ521, 2F7, U-937, C-33A) were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum, 2 mM glutamine, 1 mM non-essential amino acids, 1 mM sodium pyruvate, 100 U/ml penicillin, 0.1 mg/ml streptomycin at 37°C, in a 5% CO<sub>2</sub> incubator.

### Isolation of total RNA, and RT-PCR analysis for HERV-HC2 belong to the HERV-T family

Total RNA from human cancer cell lines was isolated using a High Pure RNA isolation kit (Roche, Indianapolis, IN) following the protocol provided by the manufacturer. Human tissues (brain, prostate, testis, heart, kidney, liver, lung, placenta, skeletal muscle, spleen, thymus, uterus) were purchased from Roche. RNA was treated with 1 U DNase I (RNase-free) at 25°C for 20 min to remove contaminating DNA. Using the pure 1 µg/µl mRNA only, expression patterns of the HERV-HC2 *pol* gene were examined by the Titan One Tube RT-PCR System (Roche). RT-PCR was performed in a 25 µl volume. We performed PCR

amplification without a reverse transcription (RT) reaction with pure mRNA samples, to verify that mRNA samples prepared from the human tissues and cancer cells did not contain genomic DNA. Using an RT-PCR approach, novel 339-bp *pol* fragments of the HERV-HC2 family were amplified by primer pairs HS307 (5'-TG GGTAAAGCTCCTTCCTTTAG-3', bases 5,128–5,149) and DS706 (5'-ATGACATATGAGGTCCTTTTC-3', bases 5,444–5,465) from HERV-HC2 (GenBank, accession no. Z70664). The RT-PCR conditions followed were the standard protocol of the Titan One Tube RT-PCR System, with an annealing temperature of 50°C. As a standard control, G3PDH was amplified using the primers GPH-S (5'-CAAAGTTGTCATGGATGACC-3', bases 31,721–31,740) and GPH-AS (5'-CCATGGAGAAGGCTGGGG-3', bases 31,898–31,915) from human G3PDH (GenBank accession no. AC068657).

#### Molecular cloning of RT-PCR products, sequencing, and data analyses

RT-PCR products were separated on a 1.8% agarose gel, purified with the QIAEX II gel extraction kit (Qiagen, Hilden, Germany) and cloned into the pGEM-T-easy vector (Promega, Madison, WI). The cloned DNA was isolated by the alkali lysis method using the High Pure plasmid isolation kit (Roche). Individual plasmid DNA was digested by the restriction enzyme *EcoRI*. Positive samples were subjected to sequence analyses on both strands with T7 and M13 reverse primers using an automated DNA sequencer (Model 373A) and the DyeDeoxy terminator kit (Applied Biosystems, Foster City, CA). Nucleotide sequence analyses were performed using the GAP, PILEUP, and PRETTY from the GCG software (<http://www.accelrys.com/products/gcg/>). The neighbor-joining tree was obtained using the MEGA2 program (Kumar et al. 2001). Bootstrap evaluation of the branching patterns was performed with 100 replications. Distances were estimated by the Kimura two-parameter method in the MEGA2 program to determine sequence divergence (Kimura 1980). Nucleotide sequences of HERV-HC2 family were retrieved from the GenBank database with the aid of the BLAST network server (Altschul et al. 1997).

#### Nucleotide sequence accession numbers

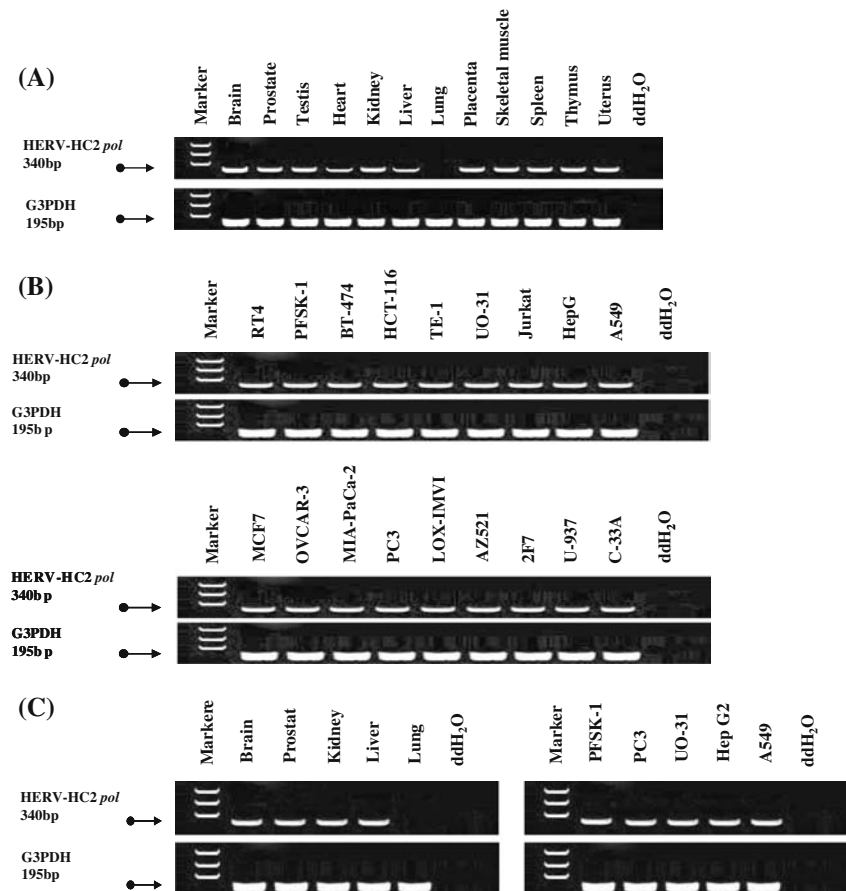
The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers: AB167270–AB167348.

## Results and discussion

Several studies on HERV transcription have been performed by Northern blot analyses (Lower et al. 1996), indicating HERV sequences are expressed preferentially in the placenta. The best example of an HERV with a known function is HERV-W, which is involved in placenta development. Particularly, the high-level expression observed in the placenta, of three endogenous coding *env* genes namely, HERV-R, HERV-W, and the newly identified HERV-FRD, have been implicated in two major physiological processes in human placental tissue. We have already used RT-PCR to examine the expression of HERV-HC2 *pol* fragments in mRNA from New World monkey tissues, which were found to have a more specific expression pattern than that detected in squirrel monkeys (Yi et al. 2003). This information motivated us to examine Old World monkey tissues (manuscript in preparation), human tissues, and abnormal tissues for comparison. Therefore, in this study, RT-PCR and sequencing was used to analyse expression of the *pol* fragment of HERV-HC2 belong to the HERV-T family in 12 human tissues and 18 human cancer cell lines. Expression of the *pol* gene was detected in all human tissues examined (brain, prostate, testis, heart, kidney, liver, placenta, skeletal muscle, spleen, thymus, and uterus) except for lung. It was also detected in all cancer cell lines examined (RT4, PFSK-1, BT-474, HCT-116, TE-1, UO-31, Jurkat, HepG2, A549, MCF7, OVCAR-3, MIA-PaCa-2, PC3, LOX-IMVI, AZ521, 2F7, U-937, and C-33A) (Fig. 1a, b). Thus, the HERV-HC2 *pol* gene may play a biological role in human tissues and cancer cells. These data suggest that the coding capacity of HERV-HC2 was acquired during hominid evolution. Recently, HERV-W structural genes (*gag/pollenv*) were found to be expressed in multiple human tissues and cancer cell lines, implying a potential transcriptional role for the HERV-W family in human cancer genomes (Yi et al. 2004). In addition, coding envelope genes of HERV-K, T, W, FRD, and R have been detected in the majority of human normal tissues (de Parseval et al. 2003).

As expected, control PCRs carried out without the RT step never resulted in any amplification. The RT-PCR products were cloned and sequenced. The clones of the expressed HERV-HC2 *pol* sequences are described in Table 1: 29 clones from different human tissues and 50 clones from human cancer cells, respectively, were sequenced. The HERV-HC2 *pol* fragment from human tissues and cancer cell lines showed a high degree of nucleotide sequence similarity (79.1–99.0%) to that of HERV-HC2 families from

**Fig. 1** Reverse transcription-polymerase chain reaction (RT-PCR) analysis of mRNA for the expression of human endogenous retrovirus (HERV)-HC2 *pol* fragments from various **a** human tissues and **b** cancer cells. **c** Comparative expression pattern between human normal tissues and cancer cells. *G3PDH* Internal positive control



human chromosomes retrieved from the NCBI database (AL662890 on chromosome 6, AC007276 on chromosome 7, AL117337 and AC073174 on chromosome 10, AC009533 on chromosome 12, AC116553 on chromosome 16, AC092364 on chromosome 19, AC011749 on chromosome Y and M32788 on chromosome 18) and our previous data from a human monochromosomal panel [AB016525 (HC2-3) on chromosome 3, AB016526 (HC2-7) on chromosome 7, AB016527 (HC2-10) on chromosome 10 and AB016528 (HC2-16) on chromosome 16]. Examination of inferred amino acid sequences revealed that 20/29 clones from human tissues and 28/50 clones from human cancer cell lines (highlighted in Table 1) had no translation interruptions resulting in point mutations or deletion/insertions. Alignment of the putative amino acid sequences of the *pol* fragments of the HERV-HC2 family from various human tissues and cancer cell lines together with our previous data (Kim and Crow 1999) (Fig. 2) revealed features that could be of great importance in understanding their functional roles. The percentage identity of 98 amino acid sequences of HERV-HC2 *pol* fragments ranged from 79.2 to 99.9% (Tables 2, 3). We also analyzed synonymous

and non-synonymous substitutions within the HERV-HC2 family in order to discover the evolutionary forces at work. In *pol* fragments from human tissues and human cancer cells, mean synonymous substitution ( $K_s$ ) ranged from 0 to 18.6%, whereas mean non-synonymous substitution ( $K_a$ ) ranged from 0 to 22.8%; the value of  $K_a/K_s$  ranged from 0 to 3.38. In terms of the  $K_a/K_s$  ratio, 53.2% of the values in pairwise comparisons were  $<1$  (data not shown). The data suggest that negative selection is acting on these sequences of the HERV-HC2 family.

The *pol* gene sequences of the HERV-HC2 family have been detected in hominoids, Old World monkeys and New World monkeys by PCR amplification (Kim et al. 1999a), in distinction to HERV-W *pol* gene sequences, which have been detected in hominoids and Old World monkeys, but not in New World monkeys (Kim et al. 1999b). To understand the evolutionary relationships within the HERV-HC2 family in human tissues and cancer cells, we performed phylogenetic analyses with these clones. Using all HERV-HC2 *pol* members including our previous data from a human monochromosomal DNA panel and the GenBank database, a neighbor-joining tree was constructed. The

**Table 1** Human endogenous retrovirus (HERV) HC2- *pol* gene sequences from various human tissues and cancer cell lines

Human tissues	Clones
Brain	HC21-1, HC21-2, HC21-3, HC21-4 <sup>a</sup> , HC21-7
Prostate	HC22-1, HC22-3 <sup>a</sup> , HC22-5 <sup>a</sup>
Testis	HC23-1 <sup>a</sup> , HC23-2 <sup>a</sup> , HC23-3 <sup>a</sup> , HC23-5 <sup>a</sup> , HC23-6, HC23-7 <sup>a</sup> , HC23-8 <sup>a</sup>
Heart	HC24-1 <sup>a</sup> , HC24-2 <sup>a</sup>
Kidney	HC25-1, HC25-2, HC25-6 <sup>a</sup>
Liver	HC26-1, HC26-2 <sup>a</sup> , HC26-3 <sup>a</sup>
Lung	–
Placenta	HC28-1 <sup>a</sup> , HC28-3 <sup>a</sup>
Skeletal muscle	HC29-2 <sup>a</sup>
Spleen	HC210-1 <sup>a</sup>
Thymus	HC211-1 <sup>a</sup>
Uterus	HC212-1 <sup>a</sup>

Cancer cell line	Organs	Clones
RT4	Bladder, urinary, carcinoma	CHC21-3, CHC21-4
PFSK-1	Brain, primitive neuroectodermal tumor brain	CHC22-2 <sup>a</sup> , CHC22-5 <sup>a</sup>
BT-474	Ductal, carcinoma, breast	CHC23-1 <sup>a</sup> , CHC23-2 <sup>a</sup>
HCT-116	Colon, carcinoma	CHC24-1 <sup>a</sup> , CHC24-2 <sup>a</sup> , CHC24-3, CHC24-5
TE-1	Esophagus	CHC25-1, CHC25-2 <sup>a</sup> , CHC25-3 <sup>a</sup> , CHC25-4 <sup>a</sup>
UO-31	Kidney	CHC26-1 <sup>a</sup> , CHC26-2 <sup>a</sup> , CHC26-4 <sup>a</sup>
Jurkat	Acute T cell leukemia	CHC27-1 <sup>a</sup>
Hep G2	Liver, hepatoblastoma	CHC28-1 <sup>a</sup> , CHC28-3 <sup>a</sup>
A549	Lung, carcinoma	CHC29-1 <sup>a</sup> , CHC29-4 <sup>a</sup>
MCF7	Mammary gland, adenocarcinoma	CHC210-1, CHC210-2, CHC210-3, CHC210-4, CHC210-5
OVCAR-3	Ovary, adenocarcinoma	CHC211-1 <sup>a</sup> , CHC211-5 <sup>a</sup>
MIA-PaCa-2	Pancreas, carcinoma	CHC212-1, CHC212-2, CHC212-5
PC3	Prostate, adenocarcinoma	CHC213-1 <sup>a</sup> , CHC213-2 <sup>a</sup>
LOX-IMVI	Skin	CHC214-1, CHC214-2 <sup>a</sup>
AZ521	Stomach	CHC215-1, CHC215-2, CHC215-4, CHC215-5
2F7	Transformed B cell lymphoma	CHC216-1, CHC216-2 <sup>a</sup>
U-937	Histiocytic lymphoma	CHC217-1, CHC217-2 <sup>a</sup> , CHC217-3, CHC217-4 <sup>a</sup> , CHC217-5 <sup>a</sup>
C-33A	Cervix, carcinoma	CHC218-1, CHC218-2 <sup>a</sup> , CHC218-3 <sup>a</sup>

<sup>a</sup> Pure peptide sequences without frameshifts and stop codons

tree showed that HERV-HC2 members have been continuously amplified and scattered by duplication events during primate evolution, clustering randomly between clones from multiple human normal tissues and cancer cells (Fig. 3). As shown in Fig. 3, HERV-HC2 family members fell into three distinct groups (I, II, III) through evolutionary divergence, indicating that the HERV-HC2 *pol* family has been amplified at least three times after the original integration into the genome. We used the *pol* gene sequence of HERV-S71 (M32788), which is the HERV-T family, as an out-group. On this tree, we found interesting features within groups T-1, T-2, T-3, T-4, T-5, T-6, and T-7. Among clones of the HERV-HC2 family from human tissues and cancer cell lines, 100% nucleotide sequence similarity to each other was discovered. Strikingly, clone HC23-1 from testis of normal human tissue was completely identical to AC009533 from human chromosome 12 in group T-1. In a similar manner, HERV-HC2 families from different tissues (brain, prostate, testis, kidney, liver, placenta, spleen, thymus, and

uterus) and cancer cells (PFSK-1, TE-1, Jurkat, Hep G2, OVCAR-3, MIA-PaCa-2, PC3, LOX-IMVI, 2F7, and U-937) belonging to the T-3, T-4, T-5, and T-7 families showed completely identical sequences to AC007276 from human chromosome 7, AL117337, and AC073174 from human chromosome 10 and AB016525 (HC2-3) from human chromosome 3, respectively.

On the assumption that the dendrogram in Fig. 3 reflects the phylogenetic relationships of the HERV-HC2 family, we computed the pairwise divergence for the three groups as group I = 12.7%, group II = 18.8%, and group III = 22.5%, employing the Kimura two-parameter method (Kimura 1980). We then estimated the divergence times of the three groups as 31 MY (group I), 47 MY (group II), and 56 MY (group III), using the average evolutionary rate of 0.2% per million years (Anderssen et al. 1997). This is in agreement with the result of our PCR analysis, which showed that HERV-HC2 *pol* gene sequences were present in hominoids, Old World monkeys, and New World monkeys (Kim et al. 1999), suggesting that



**Fig. 2** Amino acid sequence alignment of the HERV-HC2 *pol* fragments without nonsense or frameshift mutations. Consensus sequences are shown on the top row. Dashes indicate residues identical to the consensus sequences and dots indicates gaps introduced to maximize the alignments

Consensus	LLRVRCTPYG	AGFSPFEIMY	GRAPPILLKL	RDTHLAKISQ	ANLLQYVQSL
1	1				50
HC2-3	--L-----	X-----	-----	-----	-----
HC2-7	-----K-	-----	-----P--	K-----E--	-----L--
AC007276	-----W-	-----	-----P--	-----E--	-----L--
AC073174	-----CQ	-----	-----P--	-----S-E--	-----L--
AC116553	-----W-	-----	--L--P--	-----E--	-----LW--
AC009533	P-----R	-R-----	R-----P--	-----E--	-----L--
HC21-4	--L-----	-----	-----	-----	-----
HC22-3	-----W-	-----	-----P--	-----E--	-----L--
HC22-5	-----W-	-----	-----P--	-----E--	-----L--
HC23-1	P-----R	-R-----	R-----P--	-----E--	-----L--
HC23-2	--L-----	-----	-----	-----	-----
HC23-3	--L-----	-----	-----	-----	-----
HC23-5	--L-----	-----V-	-----	-----	-----
HC23-7	-----CQ	-----	-----P--	-----S-E--	-----L--
HC23-8	-----R	-R-----	R-----P--	-----E--	-----L--
HC24-1	--L-----	-----	-----	-----R	-----
HC24-2	--L-----	-----	-----	-----	-----
HC25-6	-----CQ	-----	-----P--	-----S-E--	-----L--
HC26-2	-----Q	-----	-----P--	K-----E--	-----L--
HC26-3	--L-----	-----	-----	-----	-----
HC28-1	-----Q	-----	-----P--	K-----E--	-----L--
HC28-3	--L-----	-----	-----	-----	-----
HC29-2	--L-----	-----	-----	-----	-----
HC210-1	--L-----	-----	-----	-----	-----
HC211-1	--L-----	-----	-----	-----	-----
HC212-1	--L-----	-----	-----	-----	-----
CHC22-2	--L-----	-----	-----	-----	-----
CHC22-5	-----W	--L-----	--VL--P--	--AQ-----	T-----L--P
CHC23-1	--L-----	-----	-----	-----	-----
CHC23-2	--L-----	-----	-----	-----	-----
CHC24-1	--K-----Q	-N-----	KK-----P--	-----E--	-----
CHC24-2	--L-----	-----	-----	--I-----	-----
CHC25-2	--L-----	-----	-----	-----	-----
CHC25-3	-----W	-----	-----P--	-----E--	-----
CHC25-4	-----Q	-----	-----	-----	-----
CHC26-1	-----Q	-----	-----P--	K-----E--	-----L--
CHC26-2	-----Q	-----	-----P--	-----E--	-----L--
CHC26-4	-----W	-----	-----P--	-----E--	-----L--
CHC27-1	--L-----	-----	-----	-----	-----
CHC28-1	--L-----	-----	-----	-----	-----
CHC28-3	--L-----	-----	-----	-----	-----
CHC29-1	-----Q	-----	-----P--	K-----E--	-----L--
CHC29-4	-----Q	-----K--	-----P--	K-----E--	-----L--
CHC211-1	--L-----	-----	-----	-----	-----
CHC211-5	-----Y-T-W	-----	-----P--	-----E--	-----L--
CHC213-1	-----W	-----	-----	-----	-----
CHC213-2	--L-----	-----	-----	-----	-----
CHC214-2	--L-----	-----	-----	-----	-----
CHC216-2	--L-----	-----	-----	-----	-----
CHC217-2	-----W	-----	-----P--	-----E--	-----L--
CHC217-4	--L-----	-----	-----	-----	-----
CHC217-5	-----W	-----	-----P--	-----E--	-----L--
CHC218-2	--K-----	-----	-----	-----	-----
CHC218-3	--K-----Q	-----	--L--P--	-----E--	-----

the HERV-T family was integrated into the primate genome before the divergence of New World monkeys and prosimians (Fig. 4). Taken together, our data support the notion that the HERV-HC2 *pol* elements were inserted into the primate genome about 56 MY ago, suggesting that an evolutionary rate of 0.2% per MY was available to the HERV-HC2 *pol* family during primate evolution. Accordingly, groups I, II, and III proliferated about 31, 47, and 56 MY ago, respectively.

There have been numerous reports that several HERV sequences are expressed in various tissues, indicating that they could have a biological role. The syncytin gene (HERV-W *env* gene) is one well-known functional HERV family member that is actively

expressed in human placenta and testis, with two transcripts of 4 and 8 kb in Northern blot analysis (Mi et al. 2000). Syncytin could mediate placental cytotrophoblast fusion connected to placental development. HERVs have also been frequently proposed to play a role in the etiology of chronic diseases such as cancer, autoimmunity and neuropsychiatric disease (Lower 1999). The growing number of studies on HERV transcription has revealed preferential expression of HERV families both in normal human tissues and abnormal tissues such as tumors. In comparative analyses between normal tissues and cancer cells, the HERV-HC2 *pol* gene was expressed in A549 lung cancer cells whereas no expression appeared in normal lung tissue

**Fig. 2** continued

Consensus	QQVQDIIQPL	VWGAHPNPVP	DQMGPCHSFQ	PGDLVYVKKF	QKEGLTLA
51					98
HC2-3	-----	-----	-----	-----	-----
HC2-7	-R-----L-	-R---S---	--T-----	L-----	--K---P-
AC007276	---RV-----	-----S--A-	--SE---F--	-----	-----P-
AC073174	-----	-R-----	--T-----	Q-----R-	-----P-
AC116553	---R-----	-R-----	--T-----	-----	E-----PS
AC009533	-----	-R-----	--T-L-----	-----	--KKKKDS
HC21-4	-----	-----	-----	-----	-----
HC22-3	---RV-----	-----S--A-	--SE---F--	-----	-----P-
HC22-5	---RV-----	-----S--A-	--SE---F--	-----	-----P-
HC23-1	-----	-R-----	--T-L-----	-----	--KKKKDS
HC23-2	-----	-----	-----	-----	-----
HC23-3	-----	-----	-----	-----	-----
HC23-5	-----	-----	-----	-----	-----
HC23-7	-----	-R-----	--T-----	Q-----	-----P-
HC23-8	-----	-R-----	-----	-----	-----
HC24-1	-----	-----	-----	-----	-----
HC24-2	-----	-----	-R-----	-----	-----
HC25-6	-----	-R-----	--T-----	Q-----R-	-----P-
HC26-2	-R-----L-	-R--Q-S---	--T-----	L-----	--K---P-
HC26-3	-----	-----	-----	-----	-----
HC28-1	-R-----L-	-R--Q-S---	--T-----	L-----	--K---P-
HC28-3	-----	-----	-----	-----	-----
HC29-2	-----	-----	-----	-----	-----
HC210-1	-----	-----	-----	-----	-----
HC211-1	-----	-----	-----	-----	-----
HC212-1	-----	-----	-----	-----	-----
CHC22-2	-----	-----	-----	-----	-----
CHC22-5	-----L--	-Q-T-----	--T-S-----	-----	---A--P-
CHC23-1	-----	-----	-----	-----	-----
CHC23-2	-----	-----	-----	-----	-RK-PHTV
CHC24-1	-----	-----	-----	-----	-----
CHC24-2	-----	-----	-----	-----	-----
CHC25-2	-----	-----	-----	-----	-----
CHC25-3	-----	-----	-----	-----	-----
CHC25-4	-----	-----	-----	-----	-----
CHC26-1	-R-----L-	-R--Q-S---	--T-----	L-----	--K---P-
CHC26-2	---RV-----	-----S--A-	--SE---L--	-----	-----P-
CHC26-4	-R-----L-	-----S--A-	--SE---F--	-----	-----P-
CHC27-1	-----	-----	-----	-----	-----
CHC28-1	-----	-----	-----	-----	-----
CHC28-3	-----	-----	-----	-----	-----L-R
CHC29-1	-R-----L-	-R---S---	--T-----	L-----	--K---P-
CHC29-4	-R-----L-	-R--Q-S---	--T-----	L-----	--K---P-
CHC211-1	-----	-----	-----	-----	-----
CHC211-5	---RV-----	-----S--A-	--SE---F--	-----	-----P-
CHC213-1	-----	-----	-----	-----	-----P-
CHC213-2	-----	-----	-----	-----	-----
CHC214-2	-----	-----	-----	-----	-----
CHC216-2	-----	-----	-----	-----	-----
CHC217-2	---RV-----	-----S--A-	--SE---F--	-----	-----P-
CHC217-4	-----	-----	-----	-----	-----
CHC217-5	---RV-----	-----S--A-	--TKS-----	-----	-EDS-L-
CHC218-2	-----	-----	-----	-----	-----
CHC218-3	-----	-----	-----	-----	-----

(Fig. 1c). To date, using RT-PCR, RNA in situ hybridization and Northern blot analysis, only the HERV-E *env* gene has been shown to be actively expressed in prostate carcinoma tissues and cell lines, but not in normal prostate tissues and cells (Wang-Johanning et al. 2003). No expression of the HERV-E *env* gene was detected in breast tissue. On the contrary, HERV-K *env* transcripts have been detected in most breast cancer cell lines and in many breast tumor tissues (Wang-Johanning et al. 2001). More strikingly, screening human sequence databases for endogenous retroviral elements with coding envelope genes has characterized 16 candidate genes. As part of a study

aimed at understanding the potential biological function of these latter sequences, a previously uncharacterized fusogenic envelope gene belonging to the HERV-FRD family was identified using a cell–cell fusion assay, suggesting that, like HERV-W, such sequences could have fusogenic properties in placenta formation (Blaise et al. 2004). Furthermore, a systematic screen of the expression level of 16 coding *env* genes present in the human genome was performed in a series of 19 healthy human tissues. The newly identified HERV-T *env* gene was generally transcribed in the majority of the tissues; specifically, high-level expression was detected in the thyroid, implying that this

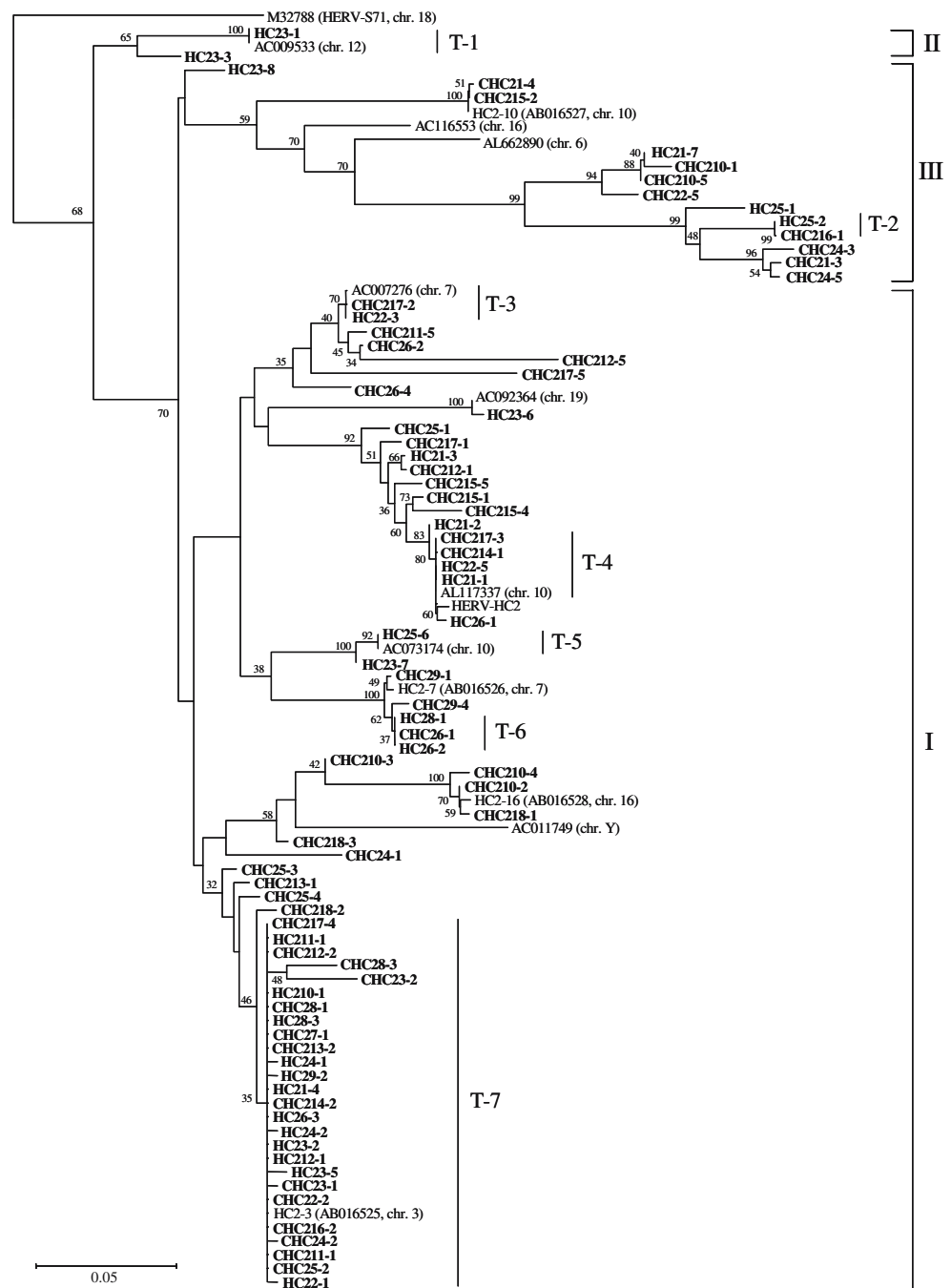
**Table 2** Percentage identity of amino acid sequences of *HC2-pol* fragment from various human tissues

HC2	1	2	3	4	5	6	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
1	HC2-3	-																									
2	HC2-7	85.4	-																								
3	HC2-16	85.4	87.8	-																							
4	AC007276	86.5	84.7	85.7	-																						
5	AC073174	86.5	88.8	90.8	85.7	-																					
6	AC116553	84.4	84.7	86.7	86.7	85.7	-																				
7	AC009533	87.9	90.2	89.0	86.8	90.1	87.9	-																			
8	HC21-4	99.0	86.5	86.5	87.5	87.5	85.4	89.0	-																		
9	HC22-3	86.5	84.7	85.7	100	85.7	86.7	86.8	87.5	100	-																
10	HC23-5	86.5	84.7	85.7	100	85.7	86.7	86.8	87.5	100	-																
11	HC23-1	87.9	90.2	89.0	86.8	90.1	87.9	100	89.0	86.8	86.8	-															
12	HC23-2	99.0	86.5	86.5	87.5	87.5	85.4	89.0	100.0	87.5	87.5	89.0	-														
13	HC23-3	95.7	86.0	87.0	84.8	85.9	83.7	90.7	96.7	84.8	84.8	91.0	96.7	-													
14	HC23-5	98.0	85.4	85.4	86.5	86.5	84.4	87.9	99.0	86.5	86.5	87.9	99.0	95.7	-												
15	HC23-7	87.5	89.8	91.8	86.7	99.0	86.7	91.2	88.5	86.7	86.7	91.2	88.5	87.0	87.5	-											
16	HC23-8	90.8	89.6	88.5	88.5	90.6	88.5	97.8	91.8	88.5	88.5	97.8	91.8	88.0	90.8	91.7	-										
17	HC24-1	98.0	85.4	85.4	86.5	86.5	84.4	87.9	99.0	86.5	86.5	87.9	99.0	95.7	98.0	87.5	90.8	-									
18	HC24-2	98.0	85.4	85.4	86.5	86.5	84.4	87.9	99.0	86.5	86.5	87.9	99.0	95.7	98.0	87.5	90.8	98.0	-								
19	HC25-6	86.5	88.8	90.8	85.7	100	85.7	90.1	87.5	85.7	85.7	90.1	87.5	85.9	86.5	99.0	90.6	86.5	86.5	-							
20	HC26-2	84.4	98.0	87.8	83.7	88.8	83.7	89.1	85.4	83.7	83.7	89.1	85.4	84.9	84.4	89.8	88.5	84.4	84.4	88.8	-						
21	HC26-3	99.0	86.5	86.5	87.5	85.4	89.0	100	87.5	87.5	87.5	89.0	100	96.7	99.0	88.5	91.8	99.0	87.5	85.4	85.4	-					
22	HC28-1	84.4	98.0	87.8	83.7	88.8	83.7	89.1	85.4	83.7	83.7	89.1	85.4	84.9	84.4	89.8	88.5	84.4	84.4	88.8	100.0	85.4	-				
23	HC28-3	99.0	86.5	86.5	87.5	87.5	85.4	89.0	100	87.5	87.5	89.0	100	96.7	99.0	88.5	91.8	99.0	87.5	85.4	100	85.4	-				
24	HC29-2	99.0	86.5	86.5	87.5	87.5	85.4	89.0	100	87.5	87.5	89.0	100	96.7	99.0	88.5	91.8	99.0	87.5	85.4	100	85.4	100	-			
25	HC210-1	99.0	86.5	86.5	87.5	87.5	85.4	89.0	100	87.5	87.5	89.0	100	96.7	99.0	88.5	91.8	99.0	87.5	85.4	100	85.4	100	100	-		
26	HC211-1	99.0	86.5	86.5	87.5	87.5	85.4	89.0	100	87.5	87.5	89.0	100	96.7	99.0	88.5	91.8	99.0	87.5	85.4	100	85.4	100	100	100	-	
27	HC212-1	99.0	86.5	86.5	87.5	87.5	85.4	89.0	100	87.5	87.5	89.0	100	96.7	99.0	88.5	91.8	99.0	87.5	85.4	100	85.4	100	100	100	100	-





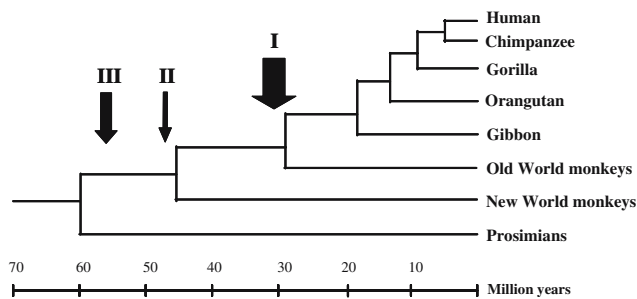
**Fig. 3** Phylogenetic tree for *pol* fragments of HERV-HC2 from human tissues and cancer cell lines. The tree was constructed using the neighbor-joining method. Branch lengths are proportional to the distances between the taxa. The values at the branch points indicate the percentage support for a particular node after 100 bootstrap replicates were performed. HC2-clones are derived from human tissues and CHC2-clones are derived from cancer cells



phenomenon might be relevant to the hormone-producing status of the thyroid gland and to specific sequences in the HERV LTR (de Parseval et al. 2003). It has been strongly suggested that enhancer and promoter elements in retroviral LTRs influence the transcription of neighboring genes (Kowalski et al. 1999).

In summary, we examined the expression pattern and relationship of HERV-HC2 belonging to the HERV-T family in human tissues and human cancer cells. With our previous data from various primates, we already had information that HERV-HC2 family is

present in hominoids, as well as in both Old- and New-World monkeys. Based on the average divergence of the HERV-HC2 family, these sequences have evolved at a rate of 0.2% nucleotide differences per MY during primate evolution, suggesting that they were integrated into primate genomes approximately 56 MY ago. Although the functional and biological roles of HERV-T family members remain obscure, our data provide useful information for studies on the transcriptional potential of the HERV-T family in the human genome as well as pathological knowledge of the HERV-T



**Fig. 4** Putative integration times of HERV-HC2 belonging to the HERV-T family during primate evolution based on the data presented in Fig. 3. An average evolutionary rate of 0.2% per million years was used. Evolutionary estimation of the HERV-HC2 family was based on the average divergence of members of each subfamily from their respective consensus sequences. The evolutionary tree was modified from Lebedev et al. (2000), Kim et al. (1999), and Anderssen et al. (1997)

family in human cancers. In further studies, it will be worthwhile investigating the relationship between expression of HERV-T and neighboring genes, as well as the gene function of the HERV-T family.

**Acknowledgments** We thank Dr. H.M. Kim for providing cancer cell lines and K.E. Schuebel for critical reading, advice, and encouragement. This study was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (A050593).

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Anderssen S, Sjøttem E, Svineng G, Johansen T (1997) Comparative analyses of LTRs of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset, African green monkey, and man. *Virology* 234:14–30
- Armbruster V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, Roemer K, Mueller-Lantzsch N (2002) A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res* 8:1800–1807
- Baltimore D (1985) Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* 40:481–482
- Benit L, Dessen P, Heidmann T (2001) Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *J Virol* 75:11709–11719
- Blaise S, Ruggieri A, Dewannieux M, Cosset FL, Heidmann T (2004) Identification of an envelope protein from the FRD family of human endogenous retroviruses (HERV-FRD) conferring infectivity and functional conservation among simians. *J Virol* 78:1050–1054
- Brack-Werner R, Barton DE, Werner T, Foellmer BE, Leib-Mosch C, Francke U, Erfle V, Hehlmann R (1989) Human SSAV-related endogenous retroviral element: LTR-like sequence and chromosomal localization to 18q21. *Genomics* 4:68–75
- de Parseval N, Lazar V, Casella JF, Benit L, Heidmann T (2003) Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins. *J Virol* 77:10414–10422
- Haltmeier M, Seifarth W, Blusch J, Erfle V, Hehlmann R, Leib-Mosch C (1995) Identification of S71-related human endogenous retroviral sequences with full-length *pol* genes. *Virology* 209:550–560
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Kabat P, Tristem M, Opavsky R, Pastorek J (1996) Human endogenous retrovirus HC2 is a new member of the S71 retroviral subgroup with a full-length *pol* gene. *Virology* 226:83–94
- Kim HS, Crow TJ (1998) Human endogenous retrovirus HC2. Map position 10q26. *Chromosome Res* 6:579
- Kim HS, Crow TJ (1999) Isolation of novel human endogenous retrovirus HC2-like elements in human chromosomes. *AIDS Res Hum Retroviruses* 15:299–302
- Kim HS, Takenaka O, Crow TJ (1999a) Identification of human endogenous retrovirus HC2-like elements in primates. *Folia Primatol* 70:295–297
- Kim HS, Takenaka O, Crow TJ (1999b) Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. *J Gen Virol* 80:2613–2619
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Kjellman C, Sjogren HO, Salford LG, Widegren B (1999) HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. *Gene* 239:99–107
- Kowalski PE, Freeman JD, Mager DL (1999) Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes. *Genomics* 57:371–379
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Lebedev YB, Belonovitch OS, Zybrowa NV, Khil PP, Kurdyukov SG, Vinogradova TV, Hunsmann G, Sverdlov ED (2000) Differences in HERV-K LTR insertions in orthologous loci of humans and great apes. *Gene* 247:265–277
- Leib-Mosch C, Haltmeier M, Werner T, Geigl EM, Brack-Werner R, Francke U, Erfle V, Hehlmann R (1993) Genomic distribution and transcription of solitary HERV-K LTRs. *Genomics* 18:261–269
- Lower R (1999) The pathogenic potential of endogenous retroviruses: facts and fantasies. *Trends Microbiol* 7:350–356
- Lower R, Lower J, 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
- Lower R, Boller K, Hasenmaier B, Korbmacher C, Muller-Lantzsch N, Lower J, Kurth R (1993) Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proc Natl Acad Sci USA* 90:4480–4484
- Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC, McCoy M (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785–789
- Shi A, Coutavas EE, Rush MG (1991) Evolutionary implications of primate endogenous retroviruses. *Virology* 182:495–502
- Sverdlov ED (1998) Perpetually mobile footprints of ancient infections in human genome. *FEBS Lett* 428:1–6

- Sverdlov ED (2000) Retroviruses and primate evolution. *Bioessays* 22:161–171
- 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 HE (1982) Form and function of retroviral proviruses. *Science* 216:812–820
- Venables PJ, Brookes SM, Griffiths D, Weiss RA, Boyd MT (1995) Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function. *Virology* 211:589–592
- Wang-Johanning F, Frost AR, Jian B, Azerou R, Lu DW, Chen DT, Johanning GL (2003) Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma. *Cancer* 98:187–197
- Wang-Johanning F, Frost AR, Johanning GL, Khazaeli MB, LoBuglio AF, Shaw DR, Strong TV (2001) Expression of human endogenous retrovirus k envelope transcripts in human breast cancer. *Clin Cancer Res* 7:1553–1560
- Werner T, Brack-Werner R, Leib-Mosch C, Backhaus H, Erfle V, Hehlmann R (1990) S71 is a phylogenetically distinct human endogenous retroviral element with structural and sequence homology to simian sarcoma virus (SSV). *Virology* 174:225–238
- Yi JM, Kim TH, Huh JI, Nakamura S, Takenaka O, Lee WH, Hyun BH, Kim HS (2003) Endogenous retrovirus HC2 *pol* fragments in the squirrel monkey: expression, evolution, and phylogeny. *Arch Virol* 148:2257–2265
- Yi JM, Kim HM, Kim HS (2004) Expression of the human endogenous retrovirus HERV-W family in various human tissues and cancer cells. *J Gen Virol* 85:1203–1210