GENOMICS 18, 261–269 (1993)

Genomic Distribution and Transcription of Solitary HERV-K LTRs

CHRISTINE LEIB-MÖSCH,*,†,[1] MANUELA HALTMEIER,* THOMAS WERNER,‡ EVA-MARIA GEIGL,†
RUTH BRACK-WERNER,† UTA FRANCKE,§ VOLKER ERFLE,† AND RÜDIGER HEHLMANN*

**III Medizinische Klinik, Klinikum Mannheim der Universität Heidelberg, D-68305 Mannheim, Federal Republic of Germany; †Institut für Molekulare Virologie and ‡Institut für Säugetiergenetik, GSF-Forschungszentrum für Umwelt und Gesundheit, D-85758 Oberschleissheim, Federal Republic of Germany; and §Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, California*

Received March 19, 1993; revised July 15, 1993

The human genome contains a family of endogenous retroviruses, HERV-K, with sequence homology to the B-type mouse mammary tumor virus. We have now identified a single HERV-K LTR within the C-type-related human retroviral element S71. The HERV-K LTR is located in the antisense direction between the S71 *gag* and the *pol* gene, replacing the 5′ half of S71 *pol*. A number of HERV-K LTR-related cDNA clones were detected by screening various human cDNA libraries with an S71 HERV-K LTR probe, indicating abundant transcription of HERV-K-related LTRs in human tissues. Sequence analysis of four cDNA clones revealed LTR sequences with a nucleotide identity of 70 to 90% with HERV-K10 LTR. Some HERV-K-related LTR sequences contain potential short open reading frames. The analyzed cDNA clones do not harbor any retroviral sequences other than those related to HERV-K LTRs. However, most of the solitary LTRs were found to be coexpressed with cellular sequences. Transcription of these LTRs is probably directed by external cellular promoters. We show that HERV-K LTR-like sequences entered the primate genome about 33–40 million years ago. We estimate the human genome to contain about 25,000 copies of HERV-K-related LTRs, which are distributed over most human chromosomes in an irregular manner. © 1993 Academic Press, Inc.

INTRODUCTION

Endogenous retroviruses and retroviral elements have been detected in the DNA of many vertebrate species, including primates. As a rule, they persist as silent retroviral copies in their host cell genome and are transmitted through the germ line as stable Mendelian genes. To date, several distinct families of human endogenous retroviral elements (HERVs) ranging from 1 to 10,000 copies per haploid human genome have been detected (for review see Larson *et al.*, 1989; Leib-Mösch *et al.*, 1990). They are estimated to constitute at least 0.6% of the human genome and are widely distributed over all human chromosomes.

Little is known about the biological function of HERVs. Like other transposable elements, they have probably been dispersed in the genome by intracellular transposition events, thereby acting as a constructive driving force in evolution. In time, some of these elements seem to have acquired cellular functions and may play a role in gene regulation (Kato *et al.*, 1990; Samuelson *et al.*, 1990; Feuchter *et al.*, 1992) Furthermore, gene products of HERVs might be involved in physiological processes such as protection against superinfection by related exogenous retroviruses or modulation of immunological functions (Ciancolo *et al.*, 1985).

One family of HERVs, HERV-K, was characterized by its sequence homology with the B-type mouse mammary tumor virus (MMTV) and Syrian hamster intracisternal A-type particles (IAP) (Ono, 1986; Ono *et al.*, 1986). Members of this family represent full-length proviral genomes of 9.4 kb with LTRs of about 1000 bp at both ends. Thus, of all known retroviral LTRs, HERV-K LTRs rank second in length after the LTRs of MMTV. The HERV-K family of full-length proviruses was found to be present at about 50 copies per human haploid genome.

Another family of HERVs comprising about 15 to 20 copies has been identified by its relationship to the C-type simian sarcoma (associated) virus (SS(A)V) (Leib-Mösch *et al.*, 1986; Brack-Werner *et al.*, 1989; Werner *et al.*, 1990). One member of this family, S71, is a truncated provirus with *gag* and *pol* sequences and a 3′ LTR-like sequence. We now have identified a single HERV-K LTR within the C-type-related human retroviral element S71. Upon screening a number of human cDNA libraries, we have found that HERV-K-related solitary LTRs are abundantly expressed in various human tissues.

MATERIALS AND METHODS

DNA sources. A Southern blot containing *Hin*dIII-digested genomic DNA from various species (Zoo blot) was obtained from Clontech

---

[1] To whom correspondence should be addressed at GSF-Forschungszentrum für Umwelt und Gesundheit, Institut für Molekulare Virologie, Postfach 1129, D-85758 Oberschleissheim, Federal Republic of Germany.

0888-7543/93 $5.00

Laboratories Inc. (Palo Alto, CA). DNAs from a Chinese hamster × human somatic cell hybrid mapping panel were kindly provided by U. Francke (Howard Hughes Medical Institute, Stanford University).

Southern blot hybridization. Genomic DNA was digested with restriction enzymes, and the resulting fragments were separated by electrophoresis in 0.8% agarose gels and transferred to nylon membranes (Zetaprobe, Bio-Rad; or Hybond N+ membrane, Amersham) by the method of Southern (1975). Filters were prehybridized for 2 h at 65°C in 6 × SSC (1× SSC is 0.15 *M* NaCl, 0.015 *M* sodium citrate), 0.05 *M* NaH_2PO_4, pH 6.5, 0.5% SDS, 0.7% powdered milk. Hybridization was carried out for at least 20 h at 65°C in 6× SSC, 0.02 *M* NaH_2PO_4, pH 6.5, 0.5% SDS, 0.7% powdered milk with 1–2 × 10^6 cpm/ml of denatured ^{32}P-labeled DNA. Filters were washed at room temperature for 30 min in 2× SSC, 0.25% powdered milk and at 55°C for 2 h in 2× SSC, 0.1% SDS.

Library screening. Human cDNA libraries HL1008 (placental tissue, 34 weeks old), HL1013 (osteosarcoma cell line MG-63, ATCC CRL 1427), HL1031b (T-cells, PHA-stimulated), HL1003b (normal human temporal cortex tissue), HL1065b (fetal brain from 21-week-old fetus), and HL1049b (glioma cell line HS683, ATCC HTB 138) were purchased from Clontech Laboratories Inc. Phage were screened under low-stringency hybridization conditions as described above using as probe either a 10-kb *Sal*I fragment containing the full-length S71 genome (Leib-Mösch *et al.*, 1986) or a 370-bp *Bgl*II/*Hin*dIII fragment (Fig. 4B) comprising the 3′ half of the U3 and the 5′ half of the R region of S71 HERV-K LTR. Selected clones were plaque purified and inserts were subcloned into pUC vectors.

DNA sequencing and computer analysis. Subcloned cDNAs were sequenced as described (Leib-Mösch *et al.*, 1986) or with the help of a semiautomatic sequencer (ABI 373) by the dideoxy chain termination method (Sanger *et al.*, 1977) using fluorescent sequencing primers and *Taq* DNA polymerase. The nucleotide sequence data are available in the EMBL database under Accession Nos. Z21850 (Os3), Z21851 (Pl64), Z21852 (Pl167), and Z21853 (Pl78). Sequence alignments were carried out on a Vax computer using the GENALIGN program of IntelliGenetics Inc. and the program package for phylogenetic analysis provided by R. F. Doolittle (Feng and Doolittle, 1987).

RESULTS

Insertion of a HERV-K-Related LTR Sequence within the S71 Genome

The C-type-related human retroviral element S71 is a truncated provirus with a complete *gag* gene, the 3′ half of the *pol* gene comprising "tether" sequences, RNase H and integrase, and a 3′ LTR-like sequence (Brack-Werner *et al.*, 1989; Werner *et al.*, 1990). The 5′ half of the *pol* gene normally containing the protease and reverse transcriptase is replaced by a so far unidentified sequence of 1132 bp that we named SNRS for S71 nonretroviral sequence. A search of GenBank and EMBL Databases now revealed that most of the S71 SNRS region is highly homologous to the LTR of the MMTV-related human endogenous retroviral element HERV-K10 (Ono, 1986). The S71 SNRS sequence shows an overall nucleotide identity of 77% with HERV-K10 LTR. The S71 HERV-K LTR-like sequence is inserted in antisense orientation between S71 *gag* and *pol* (Fig. 1). The sequences flanking the HERV-K LTR located 3′ of S71 *gag* and 5′ of S71 *pol*-related sequences comprise 46 and 166 bp, respectively, and show weak homology to 5′ sequences of C-type retroviral proteases.

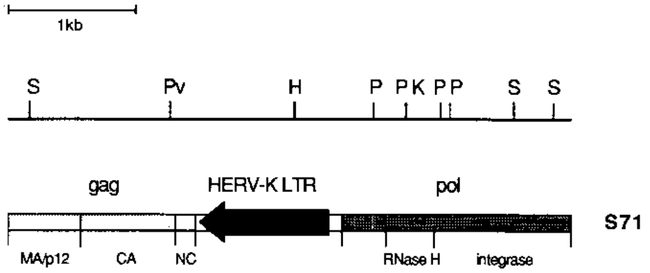


FIG. 1. Structure and restriction map of S71 *gag–pol* region. The 5′–3′ orientation of the inserted HERV-K LTR is indicated by the arrow. Abbreviations: MA, matrix protein p15; CA, capsid protein p30; NC, nucleocapsid protein p10; B, *Bam*HI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; Pv, *Pvu*II; S, *Sst*I.

HERV-K LTR Transcripts in Human Tissues and Cell lines

In the course of the isolation of S71-related transcripts, a cDNA library from a 34-week-old human placenta (Milán, 1986) and a cDNA library from the human osteosarcoma cell line MG-63 were screened with a 10-kb probe containing the full-length S71 genome (Leib-Mösch *et al.*, 1986), under low-stringency hybridization conditions. Using various subgenomic probes of S71 for characterization of the isolated cDNA clones, we identified 12 clones that hybridized only with the S71 HERV-K LTR insert. Some of these clones were analyzed further (Fig. 2). Nucleotide sequencing revealed that the isolated clones Pl64, Pl78, and Pl167 consisted of HERV-K LTR-related sequences and nonretroviral sequences. The latter were located 5′ as well as 3′ of the LTR sequences. Thus, HERV-K LTRs are expressed in the context of cellular sequences. In two cases (Pl64 and Pl167), HERV-K LTR sequences were found to be associated with *Alu* elements. The remaining nonretroviral sequences showed no significant homology to any sequence in the GenBank Database (release 72) or the EMBL Database (release 31).

Molecular clone Pl167 contains a complete HERV-K LTR flanked by cellular sequences at the 5′ and 3′ termini. Consequently, transcription had neither initiated nor terminated within the LTR. Pl64 and Pl78 contain 5′ cellular sequences and a complete U3 region, but lack part of the R region and the complete U5 region at the 3′ ends of the inserts. These cDNAs probably originate from full-length transcripts that were truncated during the cloning procedure, producing artificial 3′ ends. We therefore conclude that the isolated cDNA clones presumably reflect solitary HERV-K LTRs that are not linked to retroviral genes and are transcribed from external cellular transcription units. Os3, the only clone isolated from the human osteosarcoma cDNA library, contains a large deletion of 367 bp that comprises the 3′ half of the U3 region and the 5′ half of the R region. This deletion is not found within the HERV-K LTR-related sequences of the placenta clones.

We further examined the expression of HERV-K LTRs in human tissues and cell lines by low-stringency

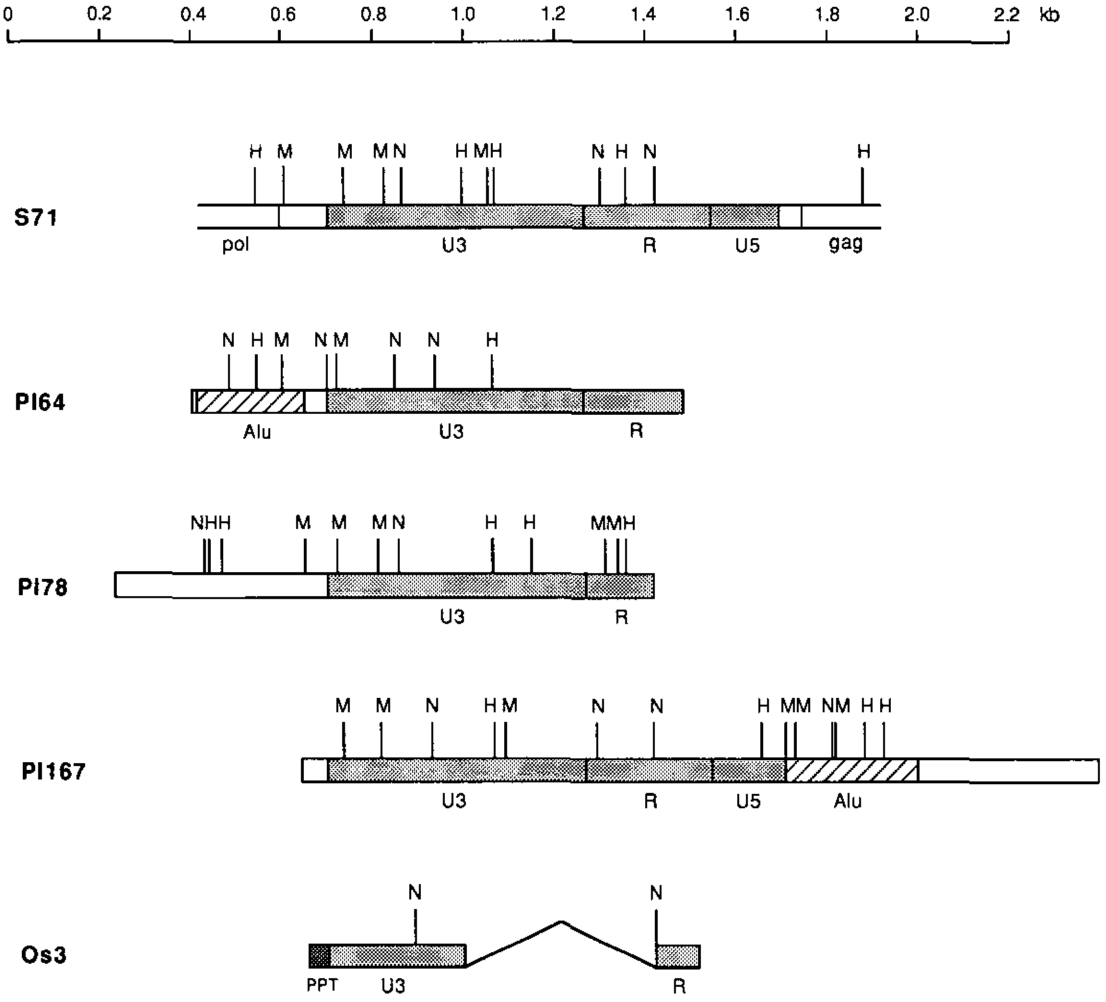


FIG. 2. Structure of the HERV-K LTR-related cDNA clones and comparison with the HERV-K LTR sequence of S71. Stippled regions indicate HERV-K LTR sequences and hatched regions mark *Alu* repeats. Abbreviations: H, *Hae*I; M, *Mae*III; N, *Nsp*cI; PPT, polypurine tract.

hybridization of various human cDNA libraries with a HERV-K-specific probe derived from the U3 and R region of S71 HERV-K LTR (Fig. 4B). Since the cDNA libraries used were made from poly(A) RNA, this method allows a semiquantitative estimate of HERV-K LTR transcription by calculating the percentage of hybridizing plaques in proportion to the total amount of plated phages. All human tissues and cell lines examined contain levels of HERV-K LTR-related transcripts ranging from 0.002 to 0.07% of total mRNA (Table 1). The highest levels of expression were found in human placenta and fetal brain, followed by the human osteosarcoma cell line MG-63 and phytohemagglutinin-stimulated T-cells. A 5- to 10-fold lower expression was found in a human glioma cell line and in normal human cortex. These data indicate that HERV-K LTR sequences are abundantly transcribed in human cells.

Nucleotide Sequence Comparison of HERV-K LTR cDNAs

The HERV-K LTR-like sequence identified within the S71 genome and the isolated cDNA clones were aligned with the prototype HERV-K10 LTR described by Ono (1986) (Fig. 3). The complete HERV-K LTRs of S71 and Pl167 were found to be 978 and 968 bp in length, respectively. The S71 HERV-K LTR is thus 10 bp longer than the LTR of clone Pl167 and the HERV-K10 LTR. All of the LTRs analyzed show the typical features found in HERV-K10 (Ono, 1986). The inverted repeats characteristic of the ends for HERV-K10 LTR are largely conserved. At position 76, a partial glucocorti-

TABLE 1

Expression of HERV-K-Related RNAs in Human Cells

Tissue/cell line	cDNA clones hybridizing with S71 HERV-K LTR (SNRS) (%)
Placenta (34 weeks)	0.07
Osteosarcoma cells (MG-63)	0.03
Fetal brain (21-week-old fetus)	0.015
PHA-stimulated T-cells (from peripheral blood)	0.01
Glioma cells (HS 683)	0.005
Normal human brain (cortex)	0.002

```
                                                                                       GRE   Enhancer
HERV-K10-5'  TGTGGGGAAAAGCAAGAGAGATCAAATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACT--CCATTTTGTTATGTGCTAAGAAAAATTCTT   100
S71 SNRS     ...........TAG..........G.C......----...A...................A..A...--..........C...A......G........
Pl64         .........----..........GG..................................T..--................C...A...........-...
Pl78         .........----...........G.........--....A.........--.A.......A..A...--........A.C...A............G..
Pl167        .....A......A...........G.C.......T.....A.A..................A......--..........C...A...............
Os3          G...........ATT....A...GG..G...G...............GG..G........G.......TTT.........C...A...............

HERV-K10-5'  CTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACGTGTGCTGTGTCAACTCAGGGTTGAATGGATTAAGGGCGGTGC   200
S71 SNRS     ............G..........G.A...C..G.........T.....TG.A..G..A......A...TG.....A....T........T.....T....
Pl64         ............G..................-.......................A..................................A..............TA...
Pl78         ....T.......G..A....C..G.A..T...G.......T.T......A.A.....A........A.TG.A...AA.A.T....C...T.....C....
Pl167        ............G..........G.A...C............T......G.A.....A..G........G.....A....T........T.....TA...
Os3          ........G...C.....G.......................T.........A....A...........C..........A...................

HERV-K10-5'  AG-GATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAA-ACTGCGG   300
S71 SNRS     ..-.............G..A.TG..T..........T.....TGG...--.....TG...T..T........G...................TGC.T...A.
Pl64         .A-.......................T..........T.............A.................................T....-C.......
Pl78         ..-...............CA.TG..T...C......T.....TGG...A.......C...TC.T.C.T...TG.T..A...........GTGC.....A.
Pl167        ..-...................A......C.............TG....A...........................................G.C.......
Os3          .AA.......................................TG..........................G..............................-C...----

HERV-K10-5'  AAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCA-TGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTG   400
S71 SNRS     ......A..............C..........T.............C..........-.....C..C....G...............................
Pl64         ........................................................-.A.......................T...................
Pl78         ..A..................C.........T..G...............C.....C..A..C..C....G.................A.......T..T
Pl167        ........................................................-..........................A...............G..
Os3          ----------------............A...--------------------------------------------------------------------

HERV-K10-5'  ACCGTCCCCCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTC-TTGCAGTTGAGACAAGAGGAAGGCATCT   500
S71 SNRS     .............A.A.....C.............................G..........-..GC...GT.....A..-----..............C
Pl64         .....................CA....T...............................T.............CA....T..........A.................
Pl78         ...A.................A......................---....G.......G...----....T...........T...............
Pl167        .....................C..............................G.........CT.--..--............T...............
Os3          ----------------------------------------------------------------------------------------------------
```

FIG. 3. Multiple alignment of HERV-K LTRs. The HERV-K LTR-related sequences of the cDNA clones Pl64, Pl78, Pl166, and Os3 and the HERV-K LTR inserted in S71 (S71 SNRS) were aligned with the 5′ LTR of the human endogenous retrovirus HERV-K10 (Ono, 1986) with the help of a program package for phylogenetic analysis (Feng and Doolittle, 1987). Only the nucleotides that differ from HERV-K10 LTR are shown for S71 and cDNA HERV-K LTR sequences. The absence of a nucleotide is denoted by a dash. Inverted repeats, glucocorticoid responsive element (GRE), enhancer core sequence, TATAA box, and polyadenylation signal are boxed. U3/R and R/U5 boundaries are indicated according to Ono (1986).

coid-responsive element (TGTTA/CT) is immediately followed by a sequence partially matching to an enhancer core (GTGGA/TA/TA/TG). The putative TATAA box around position 538 differs from the HERV-K10 TATAA box in all cases by a G instead of an A at the second nucleotide. A CAAT-box-like structure is found in HERV-K10, Pl64, and Pl167 about 15 bp upstream of the TATAA box. In clone Os3 the 367-bp deletion includes the promoter sequences. A typical polyadenylation signal, AATAAA, followed by a presumptive polyadenylation site, CA, is well conserved in the full-length LTRs of S71 and Pl167.

The nucleotide identity of the compared LTR sequences ranges between 70 and 90%. Clone Pl167 is most closely related to HERV-K10, showing an overall nucleotide identity of 85%. Within the LTRs, the R and U5 regions show a higher degree of nucleotide sequence variability (20–40%) than the U3 region (10–20%). Furthermore, deletions and insertions occur mainly within the 3′ half of the HERV-K LTRs.

Integration of proviruses into the cellular genome leads to the duplication of a short sequence of the host DNA at the host–proviral junctions (Varmus and Brown, 1989). We have compared the flanking sequences of full-length HERV-K genomes (Ono, 1986) with the flanking sequences of solitary HERV-K-related LTRs. In S71 and Pl167, the LTRs are flanked by 6-bp direct repeats of cellular DNA (GACCTT for S71 and ATCCTG for Pl167) as would be expected after insertion of a provirus. Direct repeats of 4 to 6 nucleotides are typical for integrated proviruses, whereas the flanking direct repeats of SINE elements vary widely in length (Deininger, 1989). Thus, solitary HERV-K LTRs may have been generated by recombination processes between the homologous 5′ and 3′ LTRs of former full-length proviruses, leading to excision of the intervening proviral sequences.

The human osteosarcoma cDNA clone Os3 differs in several structural features from the solitary HERV-K LTRs isolated from human placenta. Clone Os3 begins with a 37-bp stretch related to sequences that are located upstream from the 3′ LTR of HERV-K10 (Fig. 2). These sequences contain a polypurine tract (PPT) that can serve as a primer binding site for plus-strand synthesis during retroviral reverse transcription. Furthermore, clone Os3 does not contain additional cellular sequences. These features as well as the large deletion spanning half of the U3 and R region indicate that Os3 is

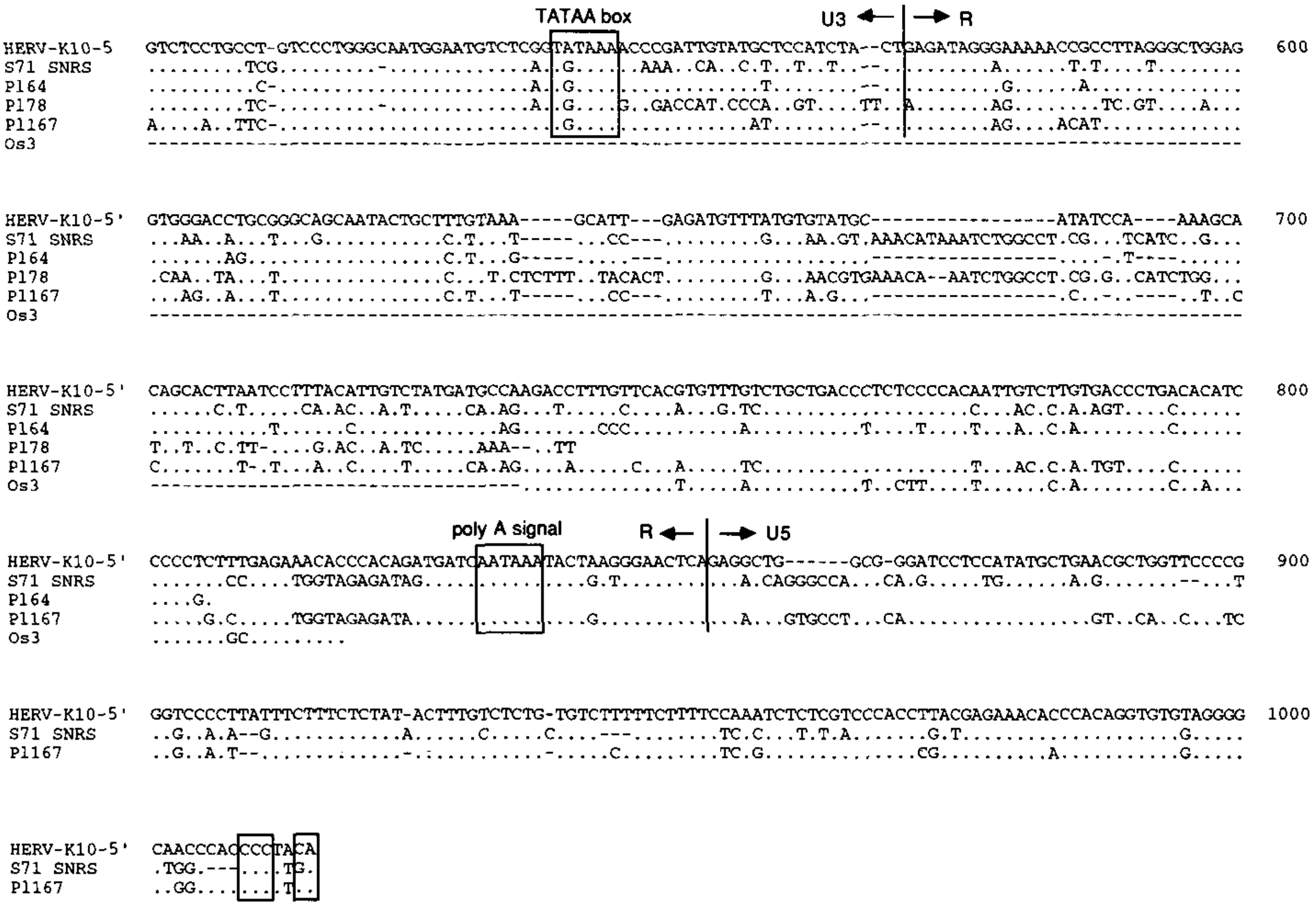


FIG. 3—*Continued*

probably a member of a family of human retroposons termed SINE-R elements that were derived from a HERV-K proviral sequence (Ono *et al.*, 1987). In contrast to SINE-R elements described by Ono, Os3 contains a second deletion of 20 bp located 17 bp upstream from the large deletion (Fig. 3).

The LTRs of MMTVs contain an open reading frame (ORF) of 957 bp (Kennedy *et al.*, 1982) that encodes a potential transmembrane glycoprotein (Korman *et al.*, 1992). We therefore examined the sequenced HERV-K LTRs for potential open reading frames. Within S71 and P178 LTR, we identified an open reading frame starting with a methionine at position 111. These open reading frames comprise 309 and 306 bp, respectively, and therefore could encode polypeptides of 103 and 102 amino acids in length. However, no significant amino acid homology with the MMTV ORF protein was detected.

Presence of HERV-K LTRs in the Eukaryotic Genome

To obtain information on the origin and age of HERV-K LTRs, we hybridized genomic DNA from different species under relaxed hybridization conditions with a HERV-K LTR probe comprising the 3′ half of the U3 and the 5′ half of the R region of S71 HERV-K LTR (Fig. 4). In addition to human DNA, we found HERV-K LTR-related sequences in DNA from rhesus monkeys but not in DNA from nonprimate mammals or avians. We also detected strongly hybridizing bands in DNA from orangutan but not in DNA from marmosets (data not shown). We therefore suggest that HERV-K LTR-like sequences were present in the primate genome at least about 33 million years ago, before the hominoides separated from the Old World monkeys, and probably entered the primate DNA at the earliest about 40 million years ago, after divergence of Old World and New World monkeys.

The high intensity and broad size range of hybridizing fragments indicate a very high copy number of solitary HERV-K LTRs in the primate genome. Because of the high intensity of the bands, it was not possible to estimate the copy number by Southern blot analysis. We therefore determined the number of HERV-K LTRs by screening a human genomic library with the 370-bp *Bgl*II–*Hin*dIII fragment from S71 HERV-K LTR shown in Fig. 4B. This probe spans the large deletion in clone Os3 and therefore does not hybridize with SINE-R elements isolated by Ono *et al.* (1987). Using low-stringency hybridization conditions that should exclude hybridization of sequences with less than 70% homology (Howley *et al.*, 1979), we detected about 140 hybridizing

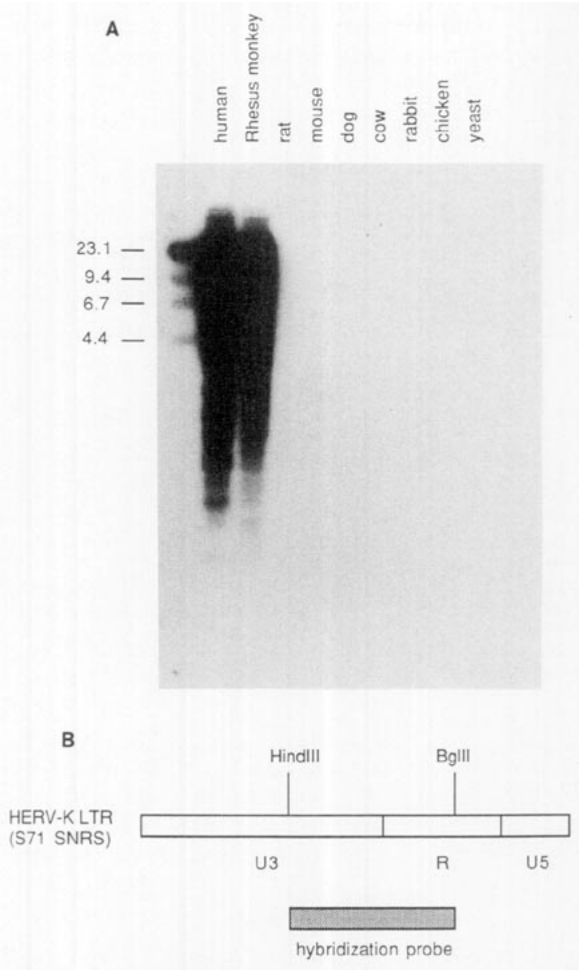
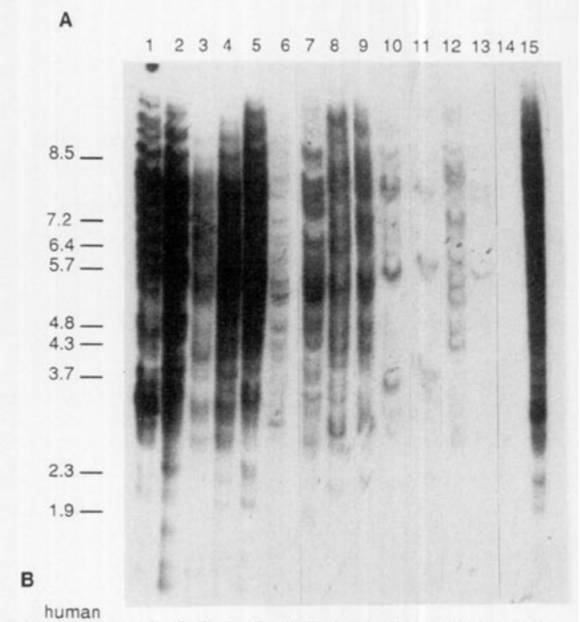


FIG. 4. Detection of HERV-K-related sequences in various species (**A**) *Hin*dIII-digested genomic DNA was hybridized with a 370-bp *Hin*dIII–*Bgl*II fragment from S71 HERV-K LTR (bottom) under low-stringency conditions. Size markers (in kb) are shown for reference. (**B**) S71 HERV-K LTR-specific hybridization probe.

clones in 10,000 plaques. We therefore estimate that the family of solitary HERV-K LTRs comprises about 25,000 related elements per haploid human genome. In contrast, the number of full-length HERV-K elements was calculated to be only about 50 copies in the human genome (Ono, 1986). HERV-K LTRs therefore represent a moderately repetitive sequence that ranges in copy number between SINE elements like *Alu* repeats (ca. 500,000 copies) and low-copy-number full-length proviral elements.

Chromosomal Dispersion of HERV-K LTRs

To examine the chromosomal distribution of HERV-K LTRs, we performed Southern blot hybridization of a panel of human × hamster somatic cell hybrids that contain human chromosomes in different combinations (Fig. 5). Genomic DNA prepared from the cell hybrids was digested with *Hin*dIII and hybridized with the *Bgl*II–*Hin*dIII fragment from S71 under low-stringency conditions. Most cell hybrids showed a very complex hybridization pattern (lanes 1–5, 7–9) as a consequence of the high copy number of HERV-K-related LTRs in the human genome. In some cases, however, we detected only a few distinct bands (lanes 6, 10–13), indicating that the chromosomal distribution of HERV-K LTRs is heterogeneous. Lanes 10, 11, and 13 show an identical hybridization pattern. Considering that strongly hybridizing bands may reflect superpositions of several equally sized fragments, we can estimate that chromosome 18 contains no more than about 6–10 copies and the X chromosome probably does not carry HERV-K LTR-re-



human chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1				+				+							+
2				+	+										+
3	+	+	+	+	+		+		+						+
4				+	+		p								+
5					+	+									+
6				+	+	+	+								+
7							+	p							+
8	+			+		+		+				+			+
9				+	+										+
10							+								+
11		+			+	+	+	p	+			+			+
12				+	+		+		+						+
13	+		+	+	+	+						+			+
14		p	+	+	+			+	+			p			+
15			+	+		+	+								+
16		+	p	+	+										+
17						+		+							+
18		+		+	+	+				+	+		+		+
19				+	+	+		+	+						+
20					+	+	+		+						+
21			+	+		+	+	p							+
22				+		+	+	+							+
X		p	p	p	+		+	+	p	+		p	p		+
Y	+	+										+			+

FIG. 5. Southern blot hybridization of a somatic cell hybrid mapping panel. (**A**) *Hin*dIII-digested DNA from human × Chinese hamster hybrid cell lines and controls was hybridized under low-stringency conditions with the S71 HERV-K LTR probe shown in Fig. 4B. Lanes **1–13**, human × Chinese hamster hybrid cell DNAs; lane **14**, Chinese hamster DNA (380-6); lane **15**, human DNA (LCL 154). (**B**) Human chromosome content. The hybrids in lanes **1–4** and lanes **10–13** are from experiment XII (Francke *et al.*, 1976), in lane **7** from series XVII (Oliver *et al.*, 1978), in lanes **5** and **9** from experiment XVIII (Francke, 1984), and in lanes **6** and **8** from series XXI (Francke and Francke, 1981). Some of the hybrid cell lines contain only defined portions of human chromosomes (p).

lated sequences. This result confirms previous observations using an S71 probe that contained the 5′ part of HERV-K LTR in addition to S71-specific retroviral sequences (Brack-Werner *et al.*, 1989). Very high copy numbers can be assigned to chromosomes 3 (lanes 1–5, 7, 9) and 16 (lanes 2, 4, 5), and relatively low copy numbers probably to chromosomes 8, 11, and 13 (lanes 6, 12).

DISCUSSION

The number of HERV-K proviruses was estimated to be about 50 copies per human haploid genome (Ono, 1986). We have shown in this study that in addition to these full-length retroviral genomes, the human DNA contains several thousand solitary HERV-K-related LTR sequences. These solitary LTRs are widely dispersed in the human genome. The high copy number and dispersion may be explained by three different mechanisms.

One possible mechanism involves two steps. At first, full-length retroviral elements are newly integrated by intracellular retrotransposition events. In the next step, recombination between the homologous 5′ and 3′ LTRs leads to excision of proviral sequences, leaving only a single LTR. Generation of a single LTR by homologous recombination has been described for an element of the RTVL-H family of HERVs (Mager and Goodchild, 1989). This mechanism has further been postulated to have generated a family of solitary LTRs from a human retrotransposon (Paulson *et al.*, 1985). The fact that solitary HERV-K LTRs show target site duplications of 6 bp that are typical in length for integrated proviruses (Varmus and Brown, 1989) supports this hypothesis.

The chromosomal distribution of HERV-K LTRs, however, is unbalanced and some chromosomes seem to be preferred targets, leading to clustering of hundreds to thousands of HERV-K-related LTRs, whereas others carry only a few copies or completely lack HERV-K-related LTRs. This heterogeneity of chromosomal distribution may be explained by a mechanism of propagation that involves amplification of sequences at the DNA level. DNA amplification also offers an alternative explanation for the high ratio of solitary LTRs to full-length HERV-K proviruses. A mechanism of this kind has been shown for the 4-1 family of HERVs. Retroviral elements of this family were found to be amplified together with associated flanking cellular sequences (Steele *et al.*, 1986). Possibly, different mechanisms have contributed independently to the amplification and dispersion of HERV-K-related LTRs within the primate genome.

A third mechanism seems to be responsible for dispersion of HERV-K LTR-derived SINE-R elements (Ono *et al.*, 1987). These elements show the features typical of SINE nonretroviral retroposons: a poly(A) tail at the 3′ end and target site duplications varying between 4 and 14 bp. Although SINE elements do not encode enzymes for reverse transcription and integration like retroviruses, they are also transposed by RNA intermediates (Deininger, 1989). In clone Os3, the presumptive 5′ and 3′ ends of the transcript are missing, probably as a consequence of the cloning procedure. Nevertheless, the remaining sequences of Os3 show a strong structural relationship with SINE-R elements that consists of HERV-K-related 5′ sequences and a large deletion in the U3/R region, indicating that Os3 is a transcription product of a member of this family. Furthermore, solitary retroviral LTRs may be transposed as isolated units by an as yet unknown mechanism, generating the 6-bp target site duplications found in S71 HERV-K LTR and in PI167.

The high copy number of HERV-K-related elements in the genome of humans and monkeys demonstrates that these sequences have been mobile within the primate genome over a long period of time and may still be actively transposed. This is also indicated by the high transcription rate of HERV-K elements in various tissues. Various HERV-K *pol*-related sequences have been amplified by PCR from normal human lymphocytes (Medstrand *et al.*, 1992). We have demonstrated that HERV-K LTR-related sequences are abundantly expressed in every human tissue that we have examined so far. Independent of their copy numbers ranging from 1 to several thousands, all human HERVs analyzed to date have been shown to persist in the primate genome for about 30 to 40 million years (Mariani-Constantini, 1989; Shih *et al.*, 1991; La Mantia *et al.*, 1991). Therefore, the differences in copy number of human endogenous retroviral elements may reflect their transposition activity. The relatively high copy number of HERV-K-related retroviral elements including solitary LTRs (25,000) suggests that these elements may constitute an active family of retroelements, comparable to THE retrotransposons (Sun *et al.*, 1984; Paulson *et al.*, 1985) and nonviral LINE retroposons (Hutchison *et al.*, 1989), which encompass about 40,000 copies (including solitary LTRs) and 100,000 copies per haploid human genome, respectively. The transposition activity of LINE elements in the human genome has recently been demonstrated in several cases. Two *de novo* insertions of LINE-1 sequences into the factor VIII gene were found to cause hemophilia A (Kazazian *et al.*, 1988; Dombroski *et al.*, 1991). A LINE-1 element was further shown to have inserted in the *myc* locus of a human breast carcinoma (Morse *et al.*, 1988). In a colon cancer, a LINE-1 insertion disrupts a putative tumor suppressor gene (Miki *et al.*, 1992). Furthermore, insertion of two solitary HERV-K LTRs within the human major histocombatibility complex locus *HLA-DQ* has been detected in several but not all *DQ* haplotypes (Kambhu *et al.*, 1990).

With the exception of clone Os3, the isolated cDNAs were found to contain nonretroviral cellular sequences 5′ and/or 3′ of the HERV-K LTR sequences. Coexpression of full-length proviruses as well as solitary LTRs with cellular sequences either by read-through transcription of adjacent sequences or by splicing into cellular genes has been observed for the human retroviral element ERV3 (Kato *et al.*, 1987) and several members of the RTVL-H family of HERVs (Wilkinson *et al.*, 1990; Liu and Abraham, 1991). In some cases, regulatory functions were attributed to the retroviral LTRs (Kato *et al.*, 1990; Feuchter and Mager, 1990; Feuchter *et al.*, 1992). In the case of HERV-K-related cDNAs, expression is not directed by the viral promoter located within the U3 region of the HERV-K LTR. We assume that HERV-K LTRs are transcribed by virtue of their insertion into other transcription units. Possibly, some of them are expressed as intron transcripts of precursor RNAs. In clones Pl64 and Pl167, HERV-K LTRs are associated with *Alu* elements, which are often known to be contained in introns or 3′ noncoding regions (Deininger, 1989). Transcripts of the human THE retrotransposon initiated from external promoters have also been observed in human tissue and cell lines (Paulsen *et al.*, 1987).

ACKNOWLEDGMENTS

We thank Margit Jaenicke and Marion Ohlmann for expert technical assistance. We are grateful to Brian Salmons for critically reading the manuscript.

REFERENCES

Brack-Werner, R., Barton, D. E., Werner, T., Foellmer, B. E., Leib-Mösch, C., Francke, U., Erfle, V., and Hehlmann, R. (1989). Human SSAV-related endogenous retroviral element: LTR-like sequence and chromosomal localization to 18q21. *Genomics* **4:** 68–75.

Ciancolo, G. J., Copeland, T. D., Oroszlan, S., and Synderman, R. (1985). Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* **230:** 453–455.

Deininger, P. L. (1989). SINEs: Short interspersed repeated DNA elements in higher eucaryotes. *In* "Mobile DNA" (D. E. Berg and M. M. Howe, Eds.), pp. 619–636, American Society of Microbiology, Washington, DC.

Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F., and Kazazian H. H., Jr. (1991). Isolation of an active human transposable element. *Science* **254:** 1805–1808.

Feng, D.-F., and Doolittle, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25:** 351–360.

Feuchter, A., and Mager, D. (1990). Functional heterogeneity of a large family of human LTR-like promoters and enhancers. *Nucleic Acids Res.* **18:** 1261–1270.

Feuchter, A., Freeman, D., and Mager, D. L. (1992). Strategy for detecting cellular transcripts promoted by human endogenous long terminal repeats: Identification of a novel gene (CDC4L) with homology to yeast CDC4. *Genomics* **13:** 1237–1246.

Francke, U. (1984). Random X inactivation resulting in mosaic nullisomy of region Xp21.1–p21.3 associated with heterozygosity for ornithine transcarbamylase deficiency and for chronic granulomatous disease. *Cytogenet. Cell Genet.* **38:** 1298–1307.

Francke, U., Busby, N., Shaw, D., Hansen, S., Hansen, S., and Brown, M. G. (1976). Intrachromosomal gene mapping in man: Assignment of nucleoside phosphorylase to region 14cen–14q21 by interspecific hybridization of cells with a t(X;14)(p22;q21) translocation. *Somatic Cell Genet.* **2:** 27–40.

Francke, U., and Francke, B. (1981). Requirement of the human chromosome 11 long arm for replication of herpes simplex virus type 1 in nonpermissive Chinese hamster × human diploid fibroblast hybrids. *Somatic Cell Genet.* **7:** 171–191.

Howley, P. M., Israel, M. A., Law, M.-F., and Martin, M. A. (1979). A rapid method for detection and mapping homology between heterologous DNAs. *J. Biol. Chem.* **244:** 4876–4883.

Hutchison, C. A., III, Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, M. H. (1989). LINE and related retroposons: Long interspersed repeated sequences in the eucaryotic genome. *In* "Mobile DNA" (D. E. Berg and M. M. Howe, Eds.), pp. 619–636, American Society of Microbiology, Washington, DC.

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:** 4927–4931.

Kato, N., Pfeifer-Ohlsson, S., Kato, M., Larsson, E., Rydnert, J., Ohlsson, R., and Cohen, M. (1987). Tissue-specific expression of human provirus ERV3 mRNA in human placenta: Two of the three ERV3 mRNAs contain human cellular sequences. *J. Virol.* **61:** 2182–2191.

Kato, N., Shimotono, K., VanLeeuwen, D., and Cohen, M. (1990). Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger protein related to krüppel. *Mol. Cell. Biol.* **10:** 4401–4405.

Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332:** 164–166.

Kennedy, N., Knedlitscheck, G., Groner, B., Hynes, N. E., Herrlich, P., Michalides, R., and van Ooyen, A. J. J. (1982). Long terminal repeats of endogenous mouse mammary tumour virus contain a long open reading frame which extends into adjacent sequences. *Nature* **295:** 622–624.

Korman, A. J., Bourgarel, P., Meo, T., and Rieckhof, G. E. (1992). The mouse mammary tumour virus long terminal repeat encodes a type II transmembrane glycoprotein. *EMBO J.* **11:** 1901–1905.

La Mantia, G., Maglione, D., Pengue, G., Di Cristofano, A., Simeone, A., Lanfrancone, L., and Lania, L. (1991). Identification and characterization of novel human endogenous retroviral sequences preferentially expressed in undifferentiated embrional carcinoma cells. *Nucleic Acids Res.* **19:** 1513–1520.

Larsson, E., Kato, N., and Cohen, M. (1989). Human endogenous proviruses. *Curr. Top. Microbiol. Immunol.* **148:** 115–132.

Leib-Mösch, C., Brack, R., Werner, T., Erfle, V., and Hehlmann, R. (1986). Isolation of an SSAV-related endogenous sequence from human DNA. *Virology* **155:** 666–667.

Leib-Mösch, C., Brack-Werner, R., Werner, T., Bachmann, M., Faff, O., Erfle, V., and Hehlmann, R. (1990). Endogenous retroviral elements in human DNA. *Cancer Res.* **50:** 5636s–5642s.

Liu, A. Y., and Abraham, B. A. (1991). Subtractive cloning of a hybrid human endogenous retrovirus and calbindin gene in the prostate cell line PC3. *Cancer Res.* **51:** 4107–4110.

Mager, D. L., and Goodchild, N. L. (1989). Homologous recombination between the LTRs of a human retrovirus-like element causes a 5-kb deletion in two siblings. *Am. J. Hum. Genet.* **45:** 848–854.

Mariani-Constantini, R., Horn, T. M., and Callahan, R. (1989). Ancestry of a human endogenous retrovirus family. *J. Virol.* **63:** 4982–4985.

Medstrand, P., Lindeskog, M., and Blomberg, J. (1992). Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. *J. Gen. Virol.* **73:** 2463–2466.

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

Milán, J. L. (1986). Molecular cloning and sequence analysis of human placental alkaline phosphatase. *J. Biol. Chem.* **261:** 3112–3115.

Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M., and Astrin, S. M. (1988). Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* **333:** 87–90.

Oliver, N., Francke, U., and Pellegrino, M. A. (1978). Regional assignment of genes for mannose phosphate isomerase, pyruvate kinase-3, and β2-microglobulin expression on human chromosome 15 by hybridization of cells from a t(15;22)(q14;q13.3) translocation carrier. *Cytogenet. Cell Genet.* **22:** 506–510.

Ono, M. (1986). Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J. Virol.* **58:** 937–944.

Ono, M., Kawakami, M., and Takezawa, T. (1987). A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids. Res.* **15:** 8725–8737.

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.

Paulson, K. E., Deka, N., Schmid, C. W., Misra, R., Schindler, C. W., Rush, M. G., Kadyk, L., and Leinwand, L. A. (1985). Transposon-like element in human DNA. *Nature* **316:** 359–361.

Paulson, K. E., Matera, A. G., Deka, N., and Schmid, C. W. (1987). Transcription of a human transposon-like sequence is usually directed by other promoters. *Nucleic Acids Res.* **15:** 5199–5215.

Samuelson, L. C., Wiebauer, K., Snow, C. M., and Meisler, M. H. (1990). Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylas genes from a single gene during primate evolution. *Mol. Cell. Biol.* **10:** 2513–2520.

Sanger, F., Nicklein, S., and Coulson, A. (1977). DNA sequencing with chain termination inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Shih, A., Coutavas, E. E., and Rush, M. G. (1991). Evolutionary implications of primate endogenous retroviruses. *Virology* **182:** 495–502.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments seperated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.

Steele, P. E., Martin, M. A., Rabson, A. B., Bryan, T., and O'Brien, S. J. (1986). Amplification and chromosomal dispersion of human endogenous retroviral sequences. *J. Virol.* **59:** 545–550.

Sun, L., Paulson, K. E., Schmid, C. W., Kadyk, L., and Leinwand, L. (1984). Non-*Alu* family interspersed repeats in human DNA and their transcriptional activity. *Nucleic Acids Res.* **12:** 2669–2690.

Varmus, H., and Brown, P. (1989). Retroviruses. *In* "Mobile DNA" (D. E. Berg and M. M. Howe, Eds.), pp. 53–108, American Society of Microbiology, Washington, DC.

Werner, T., Brack-Werner, R., Leib-Mösch, C., Backhaus, H., Erfle, V., and Hehlmann, R., (1990). S71 is a phylogenetic distinct human endogenous retroviral element with structural and sequence homology to simian sarcoma virus (SSV). *Virology* **155:** 225–238.

Wilkinson, D. A., Freeman, D., Goodchild, N. L., Kelleher, C. A., and Mager, D. L. (1990). Autonomous expression of RTVL-H endogenous retroviruslike elements in human cells. *J. Virol.* **64:** 2157–2167.