Gene, 155 (1995) 247–252

GENE 08695

Cloning and characterization of the endogenous retroviral-tRNA^{Glu} multigene family from human genomes of different racial backgrounds

(Proviral structure; transduced sequence; truncated structure; restriction-fragment length polymorphism)

Kai-Wun Yeh[a,b], Wen K. Yang[a], Huey-Chung Huang[c], Ya-Nan Feng[a], Jau-Ching Liu[a], Felicia Y.-H. Wu[a] and Cheng-Wen Wu[a]

[a]*Division of Cancer Research, Institute of Biomedical Sciences (IBMS), Academia Sinica, Taipei 115, Taiwan;* [b]*Department of Botany, National Taiwan University, Taipei 107, Taiwan; and* [c]*Institute of Biochemistry, National Taiwan University, Taipei 100, Taiwan*

Received by A.-M. Skalka: 22 April 1994; Revised/Accepted: 7 October 1994; Received at publishers: 8 December 1994

SUMMARY

An 8.3-kb human endogenous retroviral-tRNA^{Glu} (HERV-E)-encoding cDNA clone and a 1.5-kb genomic clone were isolated from a Chinese-derived cervical cancer cell line, CC7T, and their sequences determined. The former is a full-length endogenous retroviral cDNA containing corresponding u_5-*gag-pol-env*-u_3-*r* regions. The latter is a partial retroviral DNA segment, covering the *gag* and *pol* genes. Analysis of normal human DNA by Southern blot hybridization with three specific HERV-E molecular DNA probes revealed complex restriction-fragment length polymorphisms (RFLP), implying that the human genome contains diverse proviral structures and dispersed integration sites. The complex patterns were virtually identical between DNAs from African-Americans, Asians and Caucasians, with only a few minor variations. The data suggest that these proviral sequences were mostly incorporated into the human genome before racial divergence and, hence, may serve as markers for distinct chromosomal sites.

INTRODUCTION

The human genome contains various retroviral multigene families that possess the characteristic provirus structure (Larsson et al., 1989). These human endogenous retrovirus (HERV) genes generally carry a putative primer binding site (*pbs*) which is complementary to the 3′ 18 nt of a distinct tRNA. The genes are designated according to which tRNA they bind; thus, *pbs* of HERV-E binds tRNA^{Glu} (Repaske et al., 1985), while *pbs* of HERV-H, -I and -P, respectively, bind tRNAs for His, Ile, and Pro (Maeda, 1985; Harada et al., 1987; Kroger and Horak, 1987; Mager and Freeman, 1987). It is unknown whether these HERV genes can be expressed as viruses or retrotranspose to new genomic sites. Recently, certain ERV genes located at defined chromosomal sites in certain mouse strains have been shown to have distinctive genetic effects. For example, the characteristic coat color *dilute* of DBA/2 mice (Jenkins et al., 1981) and the *hairless* trait of HRS/J mice (Stoye et al., 1988) are the result of the germline insertion of a murine-leukemia-related endogenous retroviral gene. Also, 'superantigens' of the mouse histocompatibility complex involved in T-cell development of immune self-tolerance are encoded by endogenous mammary tumor virus genes (Dyson et al., 1991; Woodland et al., 1991). As nothing is known about these aspects of ERV genes in humans, a primary question is whether or not persons of different racial background carry different HERV loci. Thus, we chose the HERV-E multigene family for this study, which

Correspondence to: Dr. C.-W. Wu, Institute of Biomedical Sciences (IBMS), Academia Sinica, Taipei 115, Taiwan. Tel. (886-2) 789-9106; Fax (886-2) 782-5573; e-mail: bmkenwu@ccvax.sinica.edu.tw

Abbreviations: bp, base pair(s); CC, cervical cancer; cDNA, complementary DNA; HERV, human endogenous retrovirus or retroviral gene(s); kb, kilobase(s) or 1000 bp; LTR, long terminal repeat; nt, nucleotide(s); ORF, open reading frame; *pbs*, primer-binding site(s); RFLP, restriction-fragment length polymorphism.

SSDI 0378-1119(94)00906-6

was originally isolated from human genome by use of homological sequence of murine leukemia virus (Martin et al., 1985). The aim of present study was to isolate several HERV-E proviral sequences from Chinese origin, then to examine the pattern of integrated HERV-E proviral sequences in different human racial genomes.

EXPERIMENTAL AND DISCUSSION

(a) Molecular cloning, sequencing and transcription of HERV-E genes

A Chinese-derived cervical cancer cell line, CC7T, was grown to confluence in Dulbecco-modified Eagle medium containing 10% fetal bovine serum, in an atmosphere of 5% CO_2/95% air at 37°C. cDNA libraries were established in pUC19 from poly(A)$^+$RNA. The libraries (10^6 clones) were screened by hybridization with a HERV-E 4-1 clone DNA. Strongly positive clones were selected, one of which, designated as CC-3-1, was isolated and characterized in detail. The nt sequence of 8.3-kb cDNA showed 97% identity to the corresponding u_5-*gag*-*pol*-*env*-u_3-regions of the HERV-E 4-1 clone (Repaske et al., 1985). The putative LTR u_5 of the CC-3-1 clone encompasses the first 112 bp, while the u_3 and polyadenylation site, respectively, begin at nt 7932 and 8327. When compared to the 4-1 clone, the CC-3-1 clone showed 13 single nt substitutions (eight in *gag*, four in *pol* and one in *env*), four single-nt insertions (two in *gag*, and one in *pol* and *env*, respectively) and one deletion of nt 1019 in *gag*. The three nt, which were absent in the 4-1 clone (*pol* region) but were needed to maintain a proper reading frame for retroviral *pol* proteins, were also absent in the CC-3-1 sequences. Thus, analysis of the possible ORFs and the stop codon contents revealed that the HERV-E CC-3-1 proviral genome was likely defective in the 3′ (one third) portion of *pol* and parts of the *gag* region, but its *env* genes and the 5′ portion of *pol* gene might remain intact (Fig. 1B). In addition, genomic libraries of *Eco*RI-restricted CC7T cellular DNA constructed in λgt10 phage were screened by using 4-1 clone DNA probe; a 1.5-kb genomic DNA fragment clone, designated HRC, was obtained. The complete nt sequence of HRC clone shows 78% identity to the corresponding *gag*-*pol* region of HERV 4-1 clone, suggesting considerable genetic alterations. In addition to 3 single nt insertions at 3′-end *gag* region, there are also 12 1–5-nt deletions and three insertions occurring in the *pol* region. Aside from the stop codon (TGA) that presumably seperates *gag* and *pol* genes (nt 219 to 221), an additional five stop codons are found in *pol* region, which are not present in clone 4-1 (Fig. 1B, HRC). An interesting feature of this analysis is the presence of nearly identical deduced aa (Asp-End-Gly-Gly) at the putative *gag*-*pol* junction in both the HRC and the 4-1 clones. This implies that HRC possibly belongs to the same endogenous retroviral DNA family as the clone 4-1 (Repaske et al., 1985). The restriction maps of these DNA clones in comparison with HERV 4-1 clone are shown in Fig. 1A.

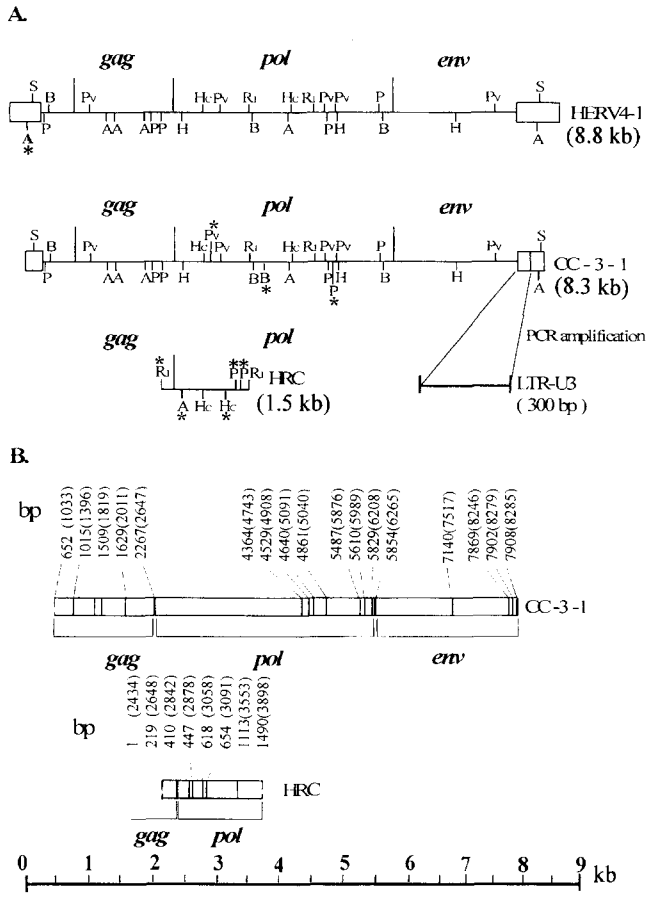

Fig. 1. Physical maps (panel **A**) and ORFs (panel **B**) of CC-3-1, HRC and LTR-u_3 (LTR-U3) DNA clones, compared with that of HERV-E 4-1 clones (Repaske et al., 1985). Primary restriction sites within the clones are indicated. Assignment of putative *gag*, *pol* and *env* genes regions, as well as reading frames is based on the data of Repaske et al. (1985). CC-3-1 (accession No. M74509) and HRC (accession No. M96062) clones, as described in section **a**, were isolated from cDNA libraries and genomic libraries, respectively. LTR-u_3 clone was amplified from the LTR-u_3 region of CC-3-1 by PCR, then subcloned in pGEM3Z vector. In panel **B**, a stop codon in the ORF is depicted as a solid line and designated by the first nt of the triplet from 5′ end or (in parentheses) the corresponding sequence numbers of the 4-1 clone. A, *Acc*I; B, *Bam*HI; H, *Hin*dIII; HC; *Hin*cII; P, *Pst*I; Pv; *Pvu*II; R, *Eco*RI; S, *Sac*I. Asterisks (*) indicate restriction sites which differ among HERV 4-1, CC-3-1 and HRC clones.

Northern blot hybridization analysis of CC7T poly(A)$^+$RNA with CC-3-1 cDNA probe clearly demonstrated that HERV-E genes expressed in these cultured cervical cancer cells (Fig. 2A). In addition to the full-length 8.3-kb gene, one 6.3- and one 1.6- kb HERV-E-related RNA transcripts could be detected. A survey of HERV-E-related gene transcripts in some other human

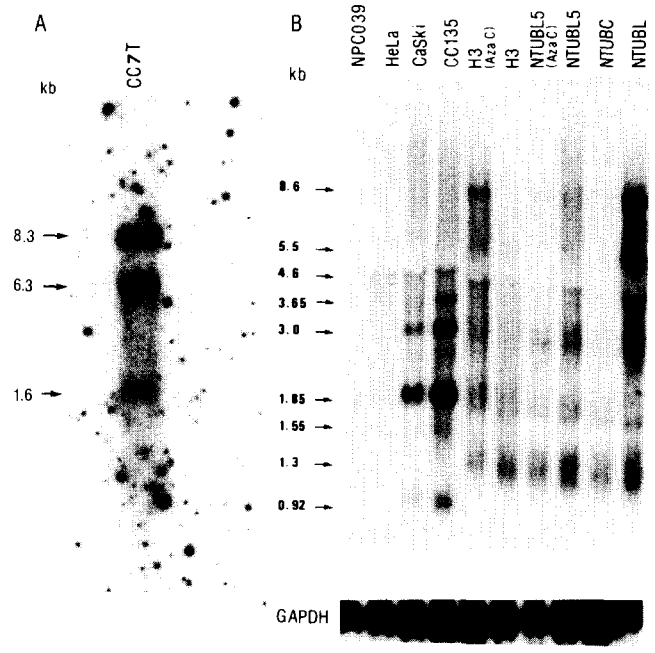

Fig. 2. Northern blot analysis of HERV-E transcripts in **(A)** CC7T cervical cancer cell line and **(B)** other human cancer cells, including four hepatoma cell lines, NTUBL, NTUBC, NTUBL5 (an isolated cell clone of NTUBL) and H3; three cervical cancer cell lines, CC-135, CaSki and HeLa; and a nasopharyngeal cancer cell line NP039. Using the procedure of Birnboim (1988), 15 μg poly(A)$^+$RNA from these cultured cancer cells was isolated and separated by electrophoresis in a 1.2% agarose gel; it was blotted and hybridized with ^{32}P-labeled CC-3-1 probe using standard conditions (Sambrook et al., 1989). After washing, the membrane was autoradiographed for 36 h. AzaC indicates that the cell culture was treated with 0.7 mM azacytidine for 24 h prior to RNA extraction. GAPDH indicates the glyceraldehyde-3-phosphate dehydrogenase gene used as a probe to perform an internal control in Northern blot.

cancer cell lines showed considerable variation in the gel pattern (Fig. 2B). Also, treatment with 5-azacytidine, that presumably can cause DNA hypomethylation and is known to induce endogenous retroviral gene expression (Groudine et al., 1981), was found to increase the HERV-E transcripts in some cancer cell lines (e.g., H3 hepatoma cells) but not in others (e.g., NTUBL5 hepatoma cells). Furthermore, we could not detect the presence of HERV-E transcripts in RNA preparations from human normal fibroblast culture cell lines (data not shown). This implies that the mechanism for regulating endogenous retroviral genes is complicated, as reported by others (Larsson et al., 1989; Löwer et al., 1993).

(b) Complexity of HERV-E RFLP

In order to distinguish the HERV-E proviral structures in human genome, a series of Southern gel blot analyses were performed. The normal human DNA samples that we prepared invariably gave complicated patterns of *Eco*RI-, *Bam*HI- and *Kpn*I-cleaved DNA fragments in Southern blot analysis with the HERV-E full-length CC-3-1 cDNA probe or the *gag-pol* HRC probe, as previously observed (Martin et al., 1981; Steele et al., 1984). Since only a single *Sac*I site is found at the u_3-*r* junctional region of each LTR in the full-length 4-1 genomic clone (Repaske et al., 1985) and at the same region of the CC-3-1 cDNA clone (Fig. 1), an individual HERV-E locus carrying the same *Sac*I site is expected to yield three *Sac*I fragments, i.e., the 5′ flanking-u_3 (LTR), proviral *r*-u_5-[*gag-pol-env*]-u_3 and the 3′ *r*-u_5 (LTR)-flanking sequences, the first two of which can be recognized by

TABLE I

*Sac*I-cleaved HERV-E fragments characterized by length and differential hybridization with CC-3-1, u_3 and HRC sequence probes

Band size (kb)	Probe[a]			
	CC-3-1	u_3	HRC	Group[b]
	Autoradiographic intensity[c]			
20	−	+	−	V
17	+	+	+	I
15	++	++	+	I
14	+	+	+	I
12.8	+++	+++	++	I
11	+	−	+	IV
9.8	+	++	+	I
8.6	+	−	−	I
8.3	++	++	++	II
7.5	+	−	++	IV
7.25	+	−	−	IV
7.1	++	−	+	IV
6.1	++	+++	+++	III-b
5.5	++	−	++/+	IV
5.1	++	+	+	III-b
4.7	++	−	+	IV
4.1	+/++	+++/++	+++	III-b
3.7	+/++	++	−	III-a
3.4	+/++	+	−	III-a
3.0	+	+++	+/++	III-b
2.9	+	++	−	III-a
2.8	+	+	−	III-a
2.45	+/±	+/++	++	III-b
2.2	+	+	++	III-b
2.1	+	+	−	III-a
1.96	++	++/+++	+	III-b
1.5	+	+	++	III-b
1.3	±/−	+	−	V
1.2	±/−	+	−	V
0.9	++	+++	−	III-a
0.7	−	+	−	V
0.56	−	+	−	V

[a] Probes CC-3-1, u_3 and HRC were shown in Fig. 1.

[b] Five groups of HERV-E sequences were interpreted as following: I, transduced fragments or sequence that lost the LTR *Sac*I site; II, full-length proviral structure; III, truncated proviral structures with deleted *gag-pol* region (IIIa) or others (IIIb); IV, sequence with rearrangement and loss; V, solitary LTR.

[c] Autoradiographic intensity of the gel bands is graded from the highly positive (+++) to the apparently negative (−), with a slash (/) to indicate the intermediate intensity.

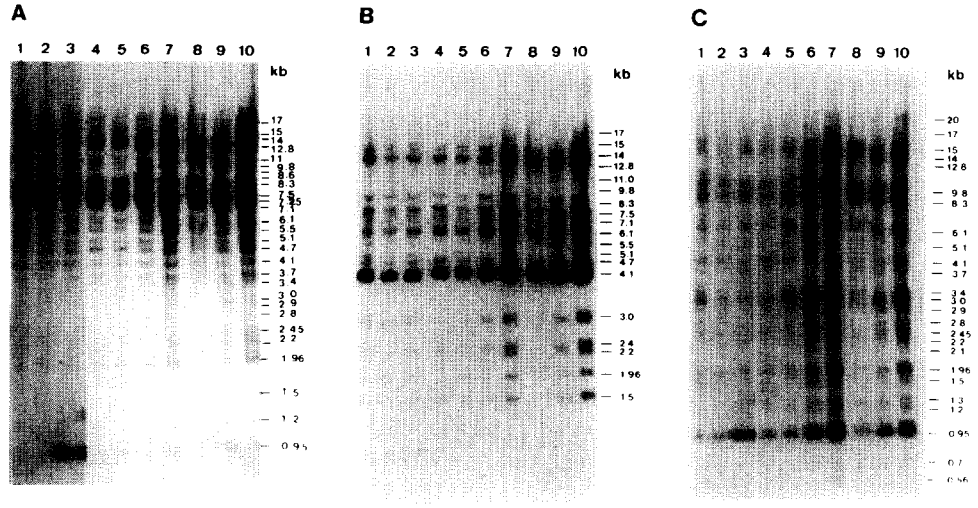

Fig. 3. *Sac*I-cleaved HERV-E fragment patterns of genomic DNA prepared from three African-Americans (lanes **1–3**), four Asians (lanes **4–7**) and three Caucasians (lanes **8–10**). All the genomic DNAs were isolated from fresh mononuclear white blood cell samples by ficol-hypaque density centrifugation. DNA of each sample (10 μg) was digested with 4 units enzyme per μg DNA for 12 h, under reaction conditions suggested by respective suppliers. The cleaved DNA was electrophoresed in a 0.6% agarose gel, blotted, and hybridized with probe CC-3-1 (panel **A**), HRC *gag-pol* probe (panel **B**) and u_3 (panel **C**), respectively. High-stringency conditions (0.1% SSC/0.1% SDS, 65°C) were used in washing steps. After hybridization with one probe and autoradiography, the DNA membrane was boiled in 10 mM Tris·Cl/10 mM EDTA buffer (pH 7.6) for 3–5 min to remove the probe, and then hybridized with another labeled probe. All DNA membranes were examined sequentially with the three specific probes.

using CC-3-1, HRC and the LTR-u_3 as the probes (Fig. 3). Such analysis revealed at least 30 different *Sac*I fragments, which could be divided into 5 groups (I–V) according to their different length and differential hybridization with the three probes (Table I). Repeated Southern blot analyses under a variety of electrophoretic conditions yielded essentially the same results, although some electrophoretic bands apparently contained overlapping or co-migrating heterogeneous fragments, sometimes with strong bands masking the closely migrating weak bands. Different HERV-E *Sac*I fragments showed different autoradiographic intensity, suggesting possible variation in copy number or degree of sequence homology (Table I). In particular, the 8.3-kb *Sac*I band (group II) was not quantitatively prominent, indicating that full-length proviruses of the 4-1/CC-3-1 type are not in the majority among the population of HERV-E gene. The complex qualitative and quantitative patterns of the *Sac*I restriction fragments clearly show the highly diverse character of the HERV-E multigene family in the normal human genome.

(c) Similarity of the complex HERV-E RFLP in different human races

Fig. 4 shows a comparison of the *Pst*I fragment patterns of HERV-E from three African Americans, six Asians and five Caucasians. Despite the complexity, the patterns were remarkably similar both qualitatively and quantitatively among these individuals; nor difference was detected between male and female suggesting the absence of Y-chromosome-associated HERV-E elements.

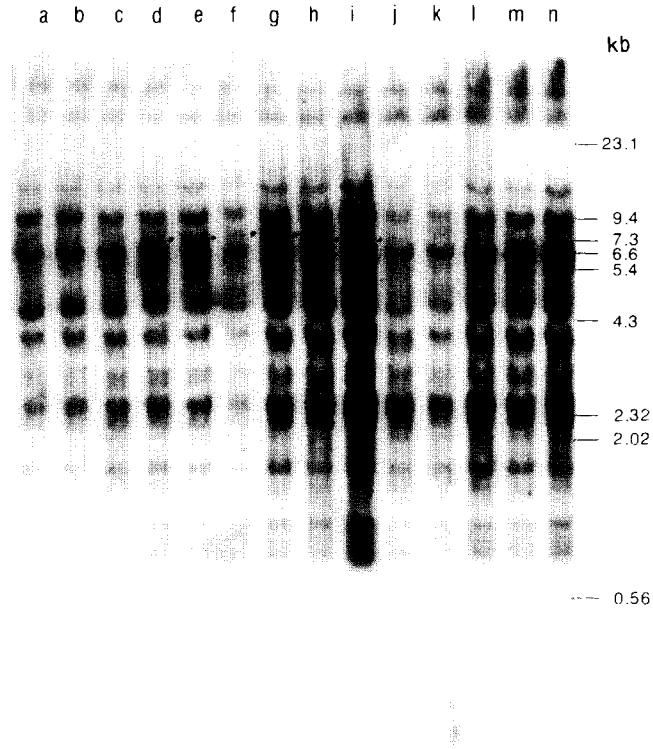

Fig. 4. Southern blot analysis of *Pst*I-cleaved HERV-E fragments detected by LTR-u_3 probe in DNA preparations of three African-Americans (lanes **a–c**), six Asians (lanes **d–i**) and five Caucasians (lanes **j–n**). A 7.3-kb fragment detected in the DNA of Asians is indicated by a dot on the right side of the lanes. DNA preparation and hybridization were performed as described in the legend to Fig. 3.

When these DNA preparations were digested with *Eco*RI (data not shown), again complex and yet nearly identical fragment patterns of HERV-E were observed regardless

of the racial or sex origin. Minor differences in the DNA of some individuals were noted but were mostly not reproducible in repeated enzyme digestion and Southern blot analyses of the same DNA preparations. One exception was a 7.3-kb u_3-containing *Pst*I fragment that was consistently detected in all samples from Asian individuals but not (or hardly) in the others. This apparent Asian-specific *Pst*I HERV-E fragment was likely due to restriction fragment polymorphism in the flanking cellular sequence, rather than the proviral structure, as no corresponding specific *Sac*I fragment was detected in these Asian DNA samples. By contrast, some marked changes in HERV-E restriction fragment patterns were found in DNA preparations of some human cancer cells, such as a Chinese patient-derived hepatoma-cellular carcinoma cell lines H1A, that also showed marked karyotypic changes. It implies that HERV-E restriction fragment changes could be due to chromosomal alterations in these cancer cells (data not shown).

Although our initial aim was to demonstrate differences, we instead observed remarkable similarity of the HERV-E multi-genes in normal cellular DNA prepared from humans of various racial backgrounds. This is in contrast to the murine leukemia virus-related proviral genes in the mouse genome, which vary considerably among different mouse strains (Wejman et al., 1984; Ch'ang et al., 1989; Frankel et al., 1989). The present results are consistent with two speculations, namely, that most (if not all) HERV-E genes were incorporated into the human genome early on in human evolution (Martin et al., 1981), and that the integrated proviral structures and their adjacent cellular sequences largely have not changed during human racial divergence (Steel et al., 1984). Concerning the latter, the apparent minor variations in HERV-E such as those found in the DNA samples from Asian individuals should be confirmed. Finally, since HERV-E genes possess marked sequence diversity and yet little racial variation, they might serve as useful markers for distinct chromosomal locations. A recent study (Nakamura et al., 1991) showed that HERV-E sequences detectable by an LTR probe could be segregated with human chromosomes in various human-rodent somatic cell hybrids. Further study is required to determine whether particular HERV-E loci on segregated human chromosomes can be distinguished by the *Sac*I fragment patterns revealed by our specific molecular probes.

ACKNOWLEDGEMENTS

We thank Dr. Arnold B. Rabson (NIH) for providing the original HERV-E 4-1 clone, Dr. H.T. Ng (Veterans General Hospital of Taipei) for giving the Chinese cervical cancer cell line CC135; and Dr. D.S. Chen (NTU Hospital) for providing the initial Chinese hepatoma cell line H3. This research was supported by Academia Sinica and National Science Council of the Republic of China (NSC82-0203-B001-03), and in part (W.K.Y.) by US-NIEHS (YO1-ES-4-0118). We also express our appreciation to Ms. Bon Bon Hsiao for typing the manuscript and to Dr. Catherine Fletcher of the Institute of Biomedical Sciences for reading the manuscript.

REFERENCES

Birnboim, H.C.: Rapid extraction of high molecular weight RNA from cultured cells and granulocytes for Northern analysis. Nucleic Acids Res. 16 (1988) 1487–1497.

Ch'ang, L.Y., Yang, W.K., Myer, F.E., Koh, C.K. and Boone, L.R.: Specific sequence deletion in two classes of murine leukemia virus-related provirus in the mouse genome. Virology 168 (1989) 245–252.

Chao, K.C., Ng, H.T., Chang, C.C., Hu, C.P. and Chang, C.M.: Properties of ectopic proteins synthesis in vitro by human cervical carcinoma cells. Asia-Oceania J. Obstet. Gynecol. 13 (1987) 221–226.

Dyson, P.J., Knight, A.M., Fairchild, S., Simpson, E. and Tomonari, K.: Genes encoding ligands for deletion of Vb 11 T cells cosegregate with mammary tumor virus genomes. Nature 349 (1991) 531–532.

Frankel, W.N., Stoye, J.P., Taylor, B.A. and Coffin, J.M.: Genetic identification of endogenous polytropic provirus by using recombinant inbred mice. J. Virol. 63 (1989) 3810–3821.

Groudine, M., Eisenman, R. and Weintraub, H.: Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. Nature 292 (1981) 311–317

Harada, F, Tsukada, N. and Kato, N.: Isolation of three kinds of human endogenous retrovirus-like sequences using tRNA as a probe. Nucleic Acids Res. 15 (1987) 9153–9162.

Jenkins, N.A., Copeland, N.G., Taylor, B.A. and Lee, B.K.: Diluted coat color mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. Nature 293 (1981) 370–374.

Kröger, B. and Horak, I.: Isolation of novel human retrovirus-related sequences by hybridization to synthetic oligonucleotides complementary to the tRNA primer-binding site. J. Virol. 61 (1987) 2071–2075.

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

Löwer, R., Löwer, J., Tondera-Koch, C. and Kurth, R.: A general method for the identification of transcribed retrovirus sequences (R-U_5 PCR) reveals the expression of the human endogenous retrovirus loci HERV-H and HERV-K in tetratocarcinoma cells. Virology 192 (1993) 501–511.

Maeda, N.: Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair: the haptoglobin-related gene contains a retrovirus-like element. J. Biol. Chem. 260 (1985) 6698–6709.

Mager, D. and Freeman, J.: Human endogenous retrovirus genome with type C *pol* sequence and *gag* segment related to human T-cell lymphotropic viruses. J. Virol. 61 (1987) 4060–4066.

Martin, M.A., Boyan, T., Rasheed, S. and Khan, A.S.: Identification and cloning of endogenous retroviral sequences present in human DNA. Proc. Natl. Acad. Sci. USA 78 (1981) 4892–4896.

Nakamura, N., Sugino, H., Takahara, K., Jin, C., Fukushige, S. and Matsubara, K.: Endogenous retroviral LTR DNA sequences as

markers for individual chromosomes. Cytogenet. Cell Genet. 57 (1991) 18–22.

Repaske, R., Steele, P., O'Neill, R., Rabson, A. and Martin, M.: Nucleotide sequence of a full-length human endogenous retroviral segment. J. Virol. 54 (1985) 764–772.

Sambrook, J., Fritsch, E.F. and Maniatis, T.: Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Steele, P., Rabson, A., Bryan, T. and Martin, M.: Distinctive termini characterize two families of human endogenous retroviral sequence. Science 225 (1984) 943–947.

Stoye, J.P., Fenner, S., Greenoak, G.E., Moran, C. and Coffin, J.M.: Role of endogenous retrovirus as mutagens: the *hairless* mutation of mice. Cell 54 (1988) 383–391.

Wejman, J.C., Taylor, B.A., Jenkin, N.A. and Copeland, N.G.: Endogenous xenotropic murine leukemia virus-related sequences map to chromosomal regions encoding mouse lymphocyte antigen. J. Virol. 50 (1984) 237–247.

Woodland, D.L., Happ, M.P., Gollob, K.J. and Palmer, E.: An endogenous retrovirus mediating deletion of αβ T cells. Nature 349 (1991) 529–530.