JOURNAL OF VIROLOGY, Nov. 1993, p. 6778–6787
0022-538X/93/116778-10$02.00/0

Vol. 67, No. 11

Characterization of Novel Reverse Transcriptase Encoding Human Endogenous Retroviral Sequences Similar to Type A and Type B Retroviruses: Differential Transcription in Normal Human Tissues

PATRIK MEDSTRAND* AND JONAS BLOMBERG

Section of Virology, Department of Medical Microbiology, Sölvegatan 23, S-223 62 Lund, Sweden

Received 10 June 1993/Accepted 11 August 1993

The polymerase chain reaction was used to amplify genomic DNA and reverse-transcribed RNA from human lymphocytes, using primers derived from conserved regions within the retroviral reverse transcriptase. Sequencing of 33 cloned amplification products revealed that a variety of sequences with similarity to mouse mammary tumor virus, mouse intracisternal A particle, and human endogenous retrovirus K10 were detected with this primer pair. The sequences were divided into six subgroups, with a nucleotide sequence dissimilarity of about 25% between the subgroups. Members within five of the subgroups were most closely related to human endogenous retrovirus K10 and mouse mammary tumor virus, whereas sequences of the sixth subgroup also showed similarity to mouse intracisternal A particle. Ten of the sequences had open reading frames with preference for silent mutations at conserved sites. Southern blot analysis showed that some HML (human endogenous MMTV-like) subgroups (HML-4 and HML-5) were present in a few copies (about 5), whereas others (HML-1 to HML-3 and HML-6) were present in at least 10 to 20 copies per genome. Northern (RNA) blot analysis revealed that several of the subgroups are differentially expressed in human normal tissues. A complex pattern of transcripts from about 12 to 1.4 kb was found in several of the tissues tested. However, the most abundant expression was detected in lung (all subgroups), skeletal muscle (HML-4 and HML-5), placenta (HML-2 and HML-5), and kidney (HML-2, HML-3 and HML-5). Expression of reverse transcriptase sequences in human tissues may have biological consequences. The described sequences are similar to elements which cause carcinoma and are immunoregulatory in mice. It remains to be seen whether human sequences also have such functions.

As much as 5 to 10% of the human genome may consist of elements introduced by mechanisms involving reverse transcription (56). Two major classes, the retroposons and the retrotransposons, are identifiable in humans (14, 45, 58). The retroposons include movable elements such as long and short interspersed repeated sequences, whose reverse transcriptase (RT) functions needed for transposition are encoded by the transposons themselves, whereas the sequences of processed pseudogenes and the *Alu* elements do not themselves encode an RT (2, 56). The second class of retroelements, the retrotransposons, have a structural organization similar to that of known infectious proviral retroviruses, that is, long terminal repeats, primer binding sites, and internal coding regions with similarity to *gag*, *pol*, and *env* (11, 27, 28). Their evolutionary history goes back millions of years, as proviral DNAs that are present in humans have been found to be similar to endogenous retroviruses (ERVs) in the primate germ line (4, 31, 49, 53). Exogenous retroviruses cause diseases ranging from neoplasias to immunodeficiencies (54, 55). Many of these viruses are similar to endogenous provirus-like sequences in the human genome which are present in several thousands of copies (5, 7). This finding has led to a search for functions and involvement of human ERVs (HERVs) in human diseases. However, so far no functions have been ascribed to these human elements. Nevertheless, it has been shown that they are not silent components of the genome, since several groups of HERVs are transcriptionally active in both normal and neoplastic cells (reviewed in references 27 and 28; see also references 6, 15, 19, 23, 25, 26, 29, and 35).

An ERV known to cause or be associated with disease is the endogenous mouse mammary tumor virus (MMTV). Inheritance of an MMTV provirus results in a high incidence of mammary carcinoma in certain strains of mice (11). ERVs related to MMTV, HLM-2 (9), NMWV4 (33, 59), HM16 (12), HERV-K (40), and HLM-25 (20) have been identified in human DNA. The complete nucleotide sequence of one full-length provirus of about 9 kb, HERV-K10 (K stands for lysine tRNA primer binding site) has been determined (43). It was found to be relatively uninterrupted by stop codons in the open reading frames (ORFs) for all genes. The *pol* ORF was without stops or frameshifts. Nucleotide sequence data from the MMTV-related HERVs HLM-2 (8), HM16 (12), and NMWV4 (32) indicate that they make up a rather diverse collection of related sequences.

From hybridization data obtained by using an MMTV *gag-pol* probe, Franklin and coworkers (15) identified nine groups of MMTV-related sequences in the human genome. The largest of these groups (64% of the isolated clones) contained HLM-2, HM16, and HERV-K10. Several of these MMTV-related sequences were also shown to be expressed as discrete-sized poly(A)$^+$ RNA transcripts in cultured human cells and placenta. To date, there has been a shortage of nucleotide sequence information needed to classify HML (human endogenous MMTV-like) sequences into distinct subgroups. Detection of new HERV sequences has been based mostly on hybridization with probes several kilobases long. However, degenerate primers derived from evolutionarily con-

* Corresponding author.

TABLE 1. Synthetic oligonucleotides[a]

Probe	Sequence
hml-1	TGGGCAGCTAAAGGAATGGT
hml-2	TATGGCTGGTATAGTAAAGGC
hml-3	ACAGTCTTGCTCAGCTAAGG
hml-4	AAGGGCTGGAACAGTGAAAG
hml-5	TCAACTCTGTCCTCTTCTGC
hml-6	AACAGAAGGCACAGAGAAGG

[a] Antisense stranded. Melting temperature was calculated to be 60°C for all probes.

served regions of the retroviral genome have been used in the polymerase chain reaction (PCR) to amplify sets of related sequences and to detect new endogenous sequences (29, 35, 50). Recently, we described a PCR primer pair which was specified to amplify a 297-bp fragment of the RT region of type A (intracisternal type A particles from hamster and mouse [IAPh {42} and IAPm {36}]), type B (MMTV [37]), and type D (Mason-Pfizer monkey virus [MPMV] [52]) retroviruses. Using this primer pair in amplification on reverse-transcribed RNA, Medstrand et al. detected a rather diverse pattern of sequences with similarity to HERV-K10 and MMTV which were expressed in lymphocytes of blood donors (35). The RT region of *pol* is the most conserved region in the retroviral genome (34). Nucleotide sequence data from this region should make it possible to characterize the extent of sequence variation of the HML family. In this study, we describe HML sequences which fall into six subgroups. Hybridization probes derived from each subgroup were then used to analyze the expression of HMLs in various normal human tissues and to investigate the organization of each subgroup in the human genome.

MATERIALS AND METHODS

DNA preparation and Southern blot analysis. Human peripheral blood mononuclear cells were isolated on metrizoate-Ficoll (Lymphoprep; Nycomed, Oslo, Norway). DNA was prepared as described before (1). DNAs were digested to completion with 10 U of *Eco*RI, *Hin*dIII, and *Xba*I (Boehringer, Mannheim, Germany) per μg of DNA and separated by electrophoresis (0.5 V/cm for 24 h) on 0.6% Tris-borate-EDTA agarose gels. DNA fragments were alkali blotted by using a vacuum blotter (Bio-Rad, Richmond, Calif.) onto nylon membranes (Hybond N^+; Amersham International, Amersham, England). Southern hybridizations were performed at 45 and 50°C. These temperatures were determined in the dot hybridizations (below). No cross-hybridizations between the six probes and the 150 clones were seen.

Membranes were prehybridized for 2 to 4 h in a 10 ml of solution containing 5× SSPE (0.9 M NaCl, 0.05 M NaH_2PO_4, 5 mM Na_2EDTA), 5× Denhardt's solution, 0.5% (wt/vol) sodium dodecyl sulfate (SDS), and 1 mg of heat-denatured herring sperm DNA (1) and then hybridized overnight in the same solution with the addition of 50 ng of 5′-labeled oligonucleotide probes hml-1 to hml-6 (Table 1). Filters were washed twice in a wash solution containing 2× SSPE and 0.1% SDS for 20 min at room temperature then twice in the wash solution at the hybridization temperature. Oligonucleotides (100 ng) were labeled by phosphorylation ([γ-^{32}P]ATP) using T4 polynucleotide kinase (Boehringer). Probes were separated from unincorporated nucleotides in spin columns (Chromaspin-10; Clontech, Palo Alto, Calif.). Southern and Northern (RNA) blots were analyzed by using a Fujix BAS 2000 BioImaging analyzer (Fuji, Tokyo, Japan).

TABLE 2. Homology of primers derived from RT of *pol* to retroviral targets[a]

Target	ABDPOL Sequence								T_d[c] (°C)	ABDPOR Sequence									T_d (°C)	Positions in target[b]	Reference
	S	P,C	W	N	T,S	P	I,V	F	V,I		Q	W	K	V	L	P	Q	G,R	M			
	TCC	CCT	TGG	AAT	ACT	CCT	GTT	TTY	GT		CAR	TGG	AAA	GTT	TTA	CCA	CAA	GGA	ATG			
MMTV	AG-	---	---	---	--G	---	---	--T	--	66	--A	---	---	---	--G	--C	--G	--T	---	66	4247–4544	37
HERV-K10	--A	---	---	T--	---	---	--G	--T	--	66	--G	---	---	-G-	---	--T	--G	---	---	70	4076–4372	43
IAPh	---	--A	---	---	--A	--A	A--	--T	--	64	--G	---	--G	-AC	--G	--G	---	--C	---	64	3476–3772	42
IAPm	--A	---	---	---	--A	--A	A--	--T	--	64	--A	---	--G	-C-	---	---	--G	---	---	68	3224–3520	36
MPMV	--T	--C	---	--C	---	--C	A-A	--T	--	58	--G	---	--G	---	---	---	---	--T	---	72	3396–3692	52
SRV-1	---	---	---	--C	---	--C	A-A	--T	--	64	--G	---	--G	---	---	--G	---	--T	---	72	3408–3704	44
SRV-2	---	--C	---	---	--A	---	A-A	--T	--	64	--A	---	---	-G-	--G	--T	---	--T	---	66	3387–3683	57
SMRV-H	--T	--A	---	---	---	---	A-A	--C	A-	60	--G	---	--G	---	-CT	G--	--G	--C	---	62	3801–4097	39
RSV	AGT	TG-	---	--C	--A	---	--C	--C	--	52	--A	---	--G	-C-	--G	--C	---	--G	---	64	2656–2952	48

[a] Conserved amino acid residues of target sequences are shown above, and target sequences are shown below each primer sequence. Hyphens indicate nucleotides identical to the primers. Both primers are shown in sense orientation. Y = T or C; R = A or G.

[b] Positions of target regions of RT which are expected to be amplified with the primer pair.

[c] T_d, predicted dissociation temperature between primer and target.

TABLE 3. Number of recombinants in each subgroup

Subgroup	No. of recombinants	
	RNA (105)[a]	DNA (45)[a]
HML-1	1	5
HML-2	49	22
HML-3	47	18
HML-4	2	0
HML-5	1	0
HML-6	5	0

[a] Origin (total number of recombinants).

RNA preparation and Northern blot analysis. Peripheral blood mononuclear cells were isolated as described above. Total RNA was extracted by the guanidinium-CsCl method (30) with the addition of a butanol extraction step (1). RNA samples were extensively treated with RNase-free DNase (RQ1 DNase; Promega, Madison, Wis.) and were then controlled to be free of DNA by RT-PCR (35).

Northern blots containing 2 μg of poly(A)$^+$ RNA of eight different normal tissues (Multiple-Tissue Northern blot; Clontech) were used in hybridization with probes hml-1 to hml-6 (Table 1). Hybridizations were performed at 45°C as for Southern blots (see above).

PCR. PCR was performed with *Taq* (*Thermus aquaticus*) DNA polymerase (Amplitaq; Perkin Elmer Cetus, Norwalk, Conn.). PCR after reverse transcription was done as previously described (35). PCR using 0.5 μg of human DNA as the template was performed in a volume of 50 μl containing 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 2.5 mM $MgCl_2$, 0.5 mg of bovine serum albumin per ml, 0.2 mM each deoxynucleoside triphosphate, 100 ng of each primer, and 1 U of Amplitaq. The following temperature profile was used on a DNA Thermal Cycler (Perkin Elmer Cetus). Denaturation at 95°C for 3 min was followed by 30 cycles of 95°C for 30 s, 50°C (annealing) for 30 s, and 72°C (extension) for 1 min and a final extension at 72°C for 5 min. Samples were resolved on Tris-borate-EDTA-based 2% agarose gels containing 0.2 μg of ethidium bromide per ml.

Plasmid amplification on lysed bacteria was performed as for genomic DNA with the addition of 0.2% Tween 20 in each reaction tube.

Cloning and sequence analysis. About 5 ng of the cDNA and DNA amplification products was cloned by using the TA cloning kit (Invitrogen, San Diego, Calif.). Positive bacterial colonies were lysed in a solution (50 μl) containing 1% Nonidet P-40, 20 mM Tris-HCl (pH 8.3), and 2 mM EDTA and incubated at 95°C for 10 min. Centrifuged lysates (2.5 μl) were used in PCR amplification (see above). Amplified plasmid DNAs with inserts of the sizes expected were spotted on nylon membranes (Hybond N$^+$; Amersham) and hybridized with probes hml-1 to hml-6 at 45 and 50°C as described above.

Plasmids were isolated from bacteria by the alkaline lysis method (3, 30). Sequencing reactions were performed by the dideoxynucleotide chain termination method (47), using the Sequenase version 2.0 kit (United States Biochemicals, Cleveland, Ohio). Labeled fragments were separated on 6% denaturing polyacrylamide gels (1) and exposed to X-ray film (SB 5; Eastman Kodak, Rochester, N.Y.). Sequencing was performed on both strands, using M13 primers.

Computer analysis was done with the PC/GENE programs (Intelligenetics, Mountain View, Calif.), and cluster dendrograms were generated by the CLUSTAL programs (kindly provided by D. Higgins, European Molecular Biology Laboratory, Heidelberg, Germany) (17, 18). Cluster dendrograms were derived from percent identities between pairs of sequences (Table 4).

Primers and oligonucleotides. The PCR primer pair ABD POL/ABDPOR has been described previously (35). These primers correspond to conserved regions within RT of type A, type B, type D, and avian type C retroviruses (Table 2). Antisense oligonucleotides hml-1 to hml-6 (Table 1) were derived from conserved regions within but differing between the six HML subgroups (positions of sense strands are underlined in Fig. 1). Oligonucleotides were synthesized by Scandinavian Gene Synthesis (Köping, Sweden).

RESULTS

Identification and sequence determination of genomic and cDNA clones similar to HERV-K10, MMTV, and IAPm. PCR amplification of PBMC cDNA by using a primer pair derived from conserved regions of RT (Table 2), cloning, and nucleotide sequences from two cDNA clones were reported previously (35). The clones were 76 and 98% similar to HERV-K10, indicating that the primer pair detected sequences related to the HERV-K family. This primer pair was subsequently used to amplify genomic DNA, and amplification products of the same size as obtained with cDNA as the template (297 bp) were obtained (data not shown). To determine the nucleotide sequences, PCR amplification products of DNA and cDNA were cloned. A total of 150 clones, 105 derived from cDNA and 45 derived from DNA, were isolated. Five cDNA and ten DNA clones were chosen at random for sequencing, which revealed a variety of sequences related to HERV-K and MMTV. The sequences were at this stage divided into four subgroups based on similarity (see below). Oligonucleotide probes hml-1 to hml-3 and hml-6 (Table 1), derived from sequences of the four subgroups, were hybridized to amplified inserts of the 150 clones which were dot blotted on nylon filters (data not shown). Cloned sequences that were not recognized by any of the probes or hybridized weakly were picked for sequencing. Using this approach, we identified sequences representing two new subgroups, HML-4 and HML-5.

Each of the 150 cloned sequences was ascribed to one of the six subgroups identified (Table 3). Two of the subgroups (HML-2 and HML-3) amplified from both DNA and cDNA made up about 90% of the sequences. Three subgroups

FIG. 1. Nucleotide sequences of cloned PCR amplifications divided into six subgroups. Primer sequences are not shown. All sequences are compared with HERV-K10 (positions 4102 to 4342 [43]). The published sequences of HM16 (12), MMTV (37), IAPm (36), and IAPh (42) were included for comparison. Dashes indicate identities, and dots indicate insertions in the sequences compared with HERV-K10. The origin and length of each sequenced clone are indicated. Sequences differing by three nucleotides or fewer are indicated as similar. The origin and number of sequenced clones similar or identical to a given sequence and, in parentheses, the number of nucleotides differing from that sequence are indicated. Nucleotides encoding a conserved amino acid residue in the exogenous retroviruses (see the legend to Fig. 3) are indicated below each section of the alignment. Sequences doubly underlined correspond to the regions where the oligonucleotide probes were chosen (Table 1) (underlined sequences may not be identical to the corresponding probe, as those were constructed to cover all sequences in a subgroup). The similarity between pairs of sequences is given in Table 4.

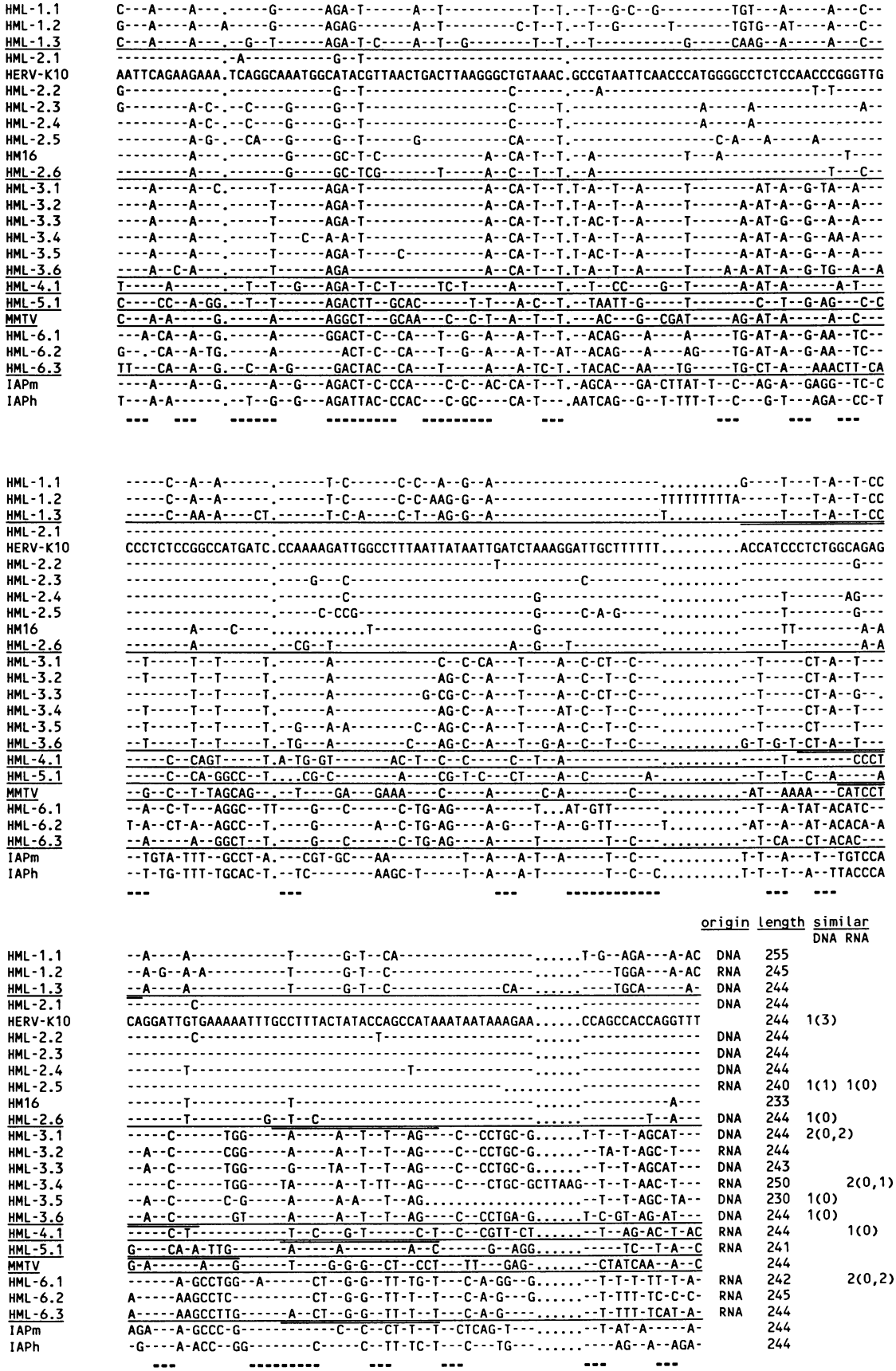
origin length similar
DNA RNA
HML-1.1 DNA 255
HML-1.2 RNA 245
HML-1.3 DNA 244
HML-2.1 DNA 244
HERV-K10 244 1(3)
HML-2.2 DNA 244
HML-2.3 DNA 244
HML-2.4 DNA 244
HML-2.5 RNA 240 1(1) 1(0)
HM16 233
HML-2.6 DNA 244 1(0)
HML-3.1 DNA 244 2(0,2)
HML-3.2 RNA 244
HML-3.3 DNA 243
HML-3.4 RNA 250 2(0,1)
HML-3.5 DNA 230 1(0)
HML-3.6 DNA 244 1(0)
HML-4.1 RNA 244 1(0)
HML-5.1 RNA 241
MMTV 244
HML-6.1 RNA 242 2(0,2)
HML-6.2 RNA 245
HML-6.3 RNA 244
IAPm 244
IAPh 244

TABLE 4. Percent identities between HERV-K10, HM16, MMTV, IAP, and HML sequences[a]

	HML-1			HML-2							HML-3						HML-4	HML-5	MMTV	HML-6			IAPm	IAPh
	1	2	3	1	HERV-K10	2	3	4	5	HM16	6	1	2	3	4	5	6	1	1		1	2	3		
HML-1.1		**92.2**	**88.1**	77.9	77.9	76.2	77.0	77.0	73.3	76.0	77.0	68.4	70.1	67.9	68.0	69.6	66.4	69.7	61.4	63.9	61.6	60.7	59.8	55.3	56.1
HML-1.2	**82.7**		**88.6**	75.8	76.6	74.6	75.4	75.0	71.7	74.2	75.0	66.0	69.3	67.1	65.6	68.7	65.2	67.2	61.0	63.1	62.0	61.6	60.2	54.9	55.3
HML-1.3	**82.7**	**82.7**		76.2	76.2	74.6	76.2	76.2	72.9	75.1	75.0	67.6	71.3	69.1	69.0	70.4	66.8	70.9	63.1	62.3	62.8	61.6	61.1	56.6	57.4
HML-2.1	81.5	77.8	80.2		**98.4**	**96.3**	**93.9**	**93.0**	**89.6**	**88.8**	**87.3**	71.3	72.1	70.8	70.9	72.2	67.2	69.7	66.0	61.5	59.9	58.6	58.6	56.6	57.8
HERV-K10	80.2	77.8	79.0	**96.3**		**95.5**	**93.9**	**93.0**	**89.6**	**88.8**	**87.3**	71.3	72.1	70.8	71.7	72.2	68.0	69.7	65.6	61.5	59.5	59.0	58.2	56.1	57.4
HML-2.2	77.8	74.1	76.5	**93.8**	**92.6**		**92.6**	**91.8**	**88.8**	**86.7**	**86.1**	70.1	70.9	70.4	70.5	71.3	66.0	68.0	64.7	61.1	58.3	57.8	57.8	56.1	57.8
HML-2.3	76.5	72.8	75.3	**91.4**	**90.1**	**87.7**		**95.9**	**91.3**	**88.4**	**86.5**	70.1	70.9	69.5	69.7	71.3	66.0	67.2	66.0	59.0	58.7	57.8	59.0	56.1	54.9
HML-2.4	77.8	74.1	76.5	**91.4**	**90.1**	**87.7**	**95.1**		**92.1**	**89.7**	**87.7**	69.3	70.1	69.1	68.9	70.4	65.2	68.4	63.9	59.4	59.5	57.8	60.2	56.6	56.6
HML-2.5	68.4	65.8	70.9	**83.5**	**82.3**	**81.0**	**82.3**	**81.0**		**86.3**	**83.3**	68.8	69.2	68.3	67.9	68.7	64.2	65.8	62.5	57.9	56.3	54.6	56.7	55.0	54.2
HM16	75.3	72.7	75.3	**88.3**	**87.0**	**84.4**	**85.7**	**87.0**	**83.1**		**92.7**	69.5	70.4	69.1	69.1	67.8	67.0	69.5	65.7	60.1	57.5	58.8	60.1	57.5	56.7
HML-2.6	75.3	72.8	74.1	**86.4**	**86.4**	**84.0**	**84.0**	**85.2**	**77.2**	**87.0**		68.0	68.9	67.9	68.4	69.1	64.3	70.1	64.3	57.8	59.1	59.4	59.0	54.5	56.6
HML-3.1	71.6	66.7	70.4	80.2	79.0	77.8	75.3	75.3	75.9	74.0	70.4		**95.1**	**94.2**	**93.0**	**90.9**	**89.8**	65.6	65.6	58.2	62.0	59.0	61.1	58.2	55.7
HML-3.2	75.3	75.3	79.0	84.0	82.7	80.2	79.0	79.0	75.9	77.9	74.1	**87.7**		**95.1**	**95.5**	**95.2**	**91.0**	66.8	65.1	59.8	63.2	60.7	61.5	59.0	56.1
HML-3.3	71.3	71.3	73.8	78.8	77.5	76.3	73.8	73.8	70.9	72.7	70.0	**88.8**	**91.3**		**92.2**	**91.3**	**88.1**	65.8	64.3	58.4	62.0	60.1	61.3	57.2	55.1
HML-3.4	69.1	66.3	71.6	77.8	77.8	76.5	72.8	72.8	69.6	72.7	70.4	**82.7**	**90.1**	**86.3**		**91.7**	**87.7**	66.0	63.9	57.8	62.0	59.2	60.7	56.6	55.3
HML-3.5	75.0	73.7	77.6	81.6	80.3	78.9	76.3	76.3	71.1	71.1	71.1	**80.3**	**89.5**	**85.5**	**82.9**		**87.8**	66.1	65.2	61.7	62.6	61.3	62.2	60.0	56.5
HML-3.6	65.4	61.7	65.4	69.1	70.4	66.7	64.2	64.2	63.3	67.5	60.5	**77.8**	**81.5**	**76.3**	**74.1**	**75.0**		61.1	61.8	56.6	57.9	54.1	57.4	55.7	52.5
HML-4.1	77.8	72.8	79.0	80.2	79.0	76.5	76.5	77.8	70.9	79.2	77.8	69.1	75.3	71.3	70.4	75.0	61.7		61.0	59.8	56.6	55.7	55.3	51.6	58.2
HML-5.1	58.7	58.7	62.5	67.5	67.5	66.3	65.0	63.8	63.3	64.9	61.3	66.3	66.3	62.5	60.0	65.8	57.5	62.5		57.7	57.7	59.8	57.7	54.8	55.2
MMTV	60.5	58.0	59.3	61.7	61.7	59.3	56.8	58.0	49.4	55.8	54.3	53.1	60.5	56.3	51.9	60.5	50.6	58.0	55.0		57.9	56.1	54.9	60.2	59.8
HML-6.1	53.8	56.3	57.5	55.0	56.3	53.8	52.5	52.5	46.8	51.9	51.3	51.3	56.3	57.5	51.3	55.3	47.5	52.5	55.0	55.0		**85.1**	**79.3**	62.4	59.5
HML-6.2	50.6	51.9	50.6	51.9	53.1	51.9	51.9	51.9	45.6	49.4	49.4	45.7	50.6	51.3	45.7	52.6	39.5	45.7	51.3	50.6	**71.3**		**79.9**	61.1	58.2
HML-6.3	54.3	58.0	55.6	56.8	56.8	55.6	58.0	56.8	50.6	55.8	51.9	51.9	56.8	57.5	50.6	57.9	46.9	51.9	55.0	51.9	**70.0**	**70.4**		60.2	59.4
IAPm	54.3	51.9	53.1	55.6	55.6	54.3	53.1	53.1	49.4	53.2	50.6	50.6	55.6	50.0	46.9	53.9	43.2	53.1	53.8	56.8	55.0	50.6	53.1		70.9
IAPh	53.1	50.6	51.9	58.0	58.0	56.8	54.3	55.6	50.6	54.5	51.9	51.9	56.8	51.3	48.1	55.3	44.4	54.3	52.5	58.0	55.0	50.6	51.9	82.7	

[a] Upper right, percent nucleotide sequence identities; lower left, percent amino acid sequence identities. Sequences within subgroups are shown in boldface.

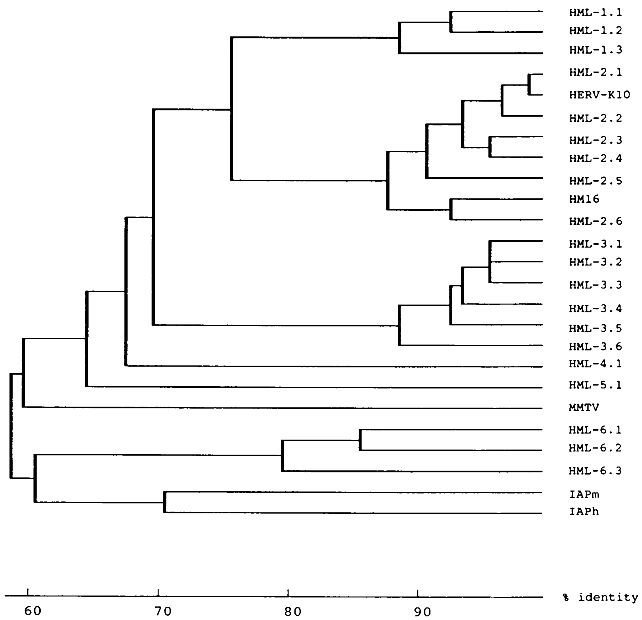

FIG. 2. Cluster dendrogram constructed from nucleotide sequence data (Table 4), indicating that the sequences presented in Fig. 1 are divided into six distinct subgroups.

(HML-4 to HML-6) were represented only in cDNA clones comprising a small proportion of the total, whereas one subgroup (HML-1) was represented in a small proportion of the cDNA and DNA clones.

The sequences are shown in Fig. 1. Members within a subgroup were more than 85% identical (except for HML-2.5 versus HML-2.6 and HML-6.3 versus HML-6.1/HML-6.2), whereas the intersubgroup similarity did not exceed 75% (except for some HML-1 versus HML-2 sequences) (Table 4). All were 54 to 63% similar to MMTV and 52 to 62% to the IAPs. The HML-2 sequences were closely related to HERV-K10 (more than 87%). The sequences of HML-4 and HML-5 were distantly related to the rest (less than 71%), and sequences of the HML-6 subgroup shared more identities with the IAPs (58 to 62%) than with MMTV (55 to 58%). MMTV were found to be most similar to the HML-1 sequences (about 63%).

Sequences of 13 clones were identical or differed by up to three nucleotides (Fig. 1). Sequences differing by only a few bases could be the result of misincorporations by RT or *Taq* polymerase (in the case of the cDNA clones) or by *Taq* polymerase alone (for genomic amplification) (16, 46) and may have been generated from the same target. The previously reported clones 5:30 and 5:28 (35) are closely related to hml-2.1 and hml-3.2, respectively.

A dendrogram based on percent identities (Table 4) is shown in Fig. 2. The sequences of HM16 and HERV-K10 both fall into the HML-2 subgroup. These data are consistent with those obtained by Franklin et al. (15), who found that both HM16 and HERV-K10 were present in the most predominant group (represented by NMWV1). The other sequences of this subgroup and the sequences of the five other clusters have not been reported previously.

Known type A, B, and D retroviruses contain 244 nucleotides over this region of *pol* (Fig. 1). Sequences from each of the subgroups reported here (except HML-4) include members differing from this length. When the nucleotide sequences of the different HML members are translated from reading frames corresponding to ORFs of MMTV and the IAPs (Fig. 3), those (except hml-3.4) differing from the 244-bp length

```
                                                                                                     length frame
HML-1.1   -K------RM----------TR---V------------------M--------...A----A---------V-T-----..S-D--   81  ORF
HML-1.2   -K-/----RA------L-----S--V-X---------------KV--------FFF/-----A-G-K-----V-------..-VD--   83  frameshifts,stop
HML-1.3   -K------RM---------------K--------T-T---X---V-------/..-----A---------V-----NK..-VH--    80  frameshift,stop
HML-2.1   ----X---RM-------------------------------------------...-----------------------..-----   80  stop
HERV-K10  IQKKSGKWHTLTDLRAVNAVIQPMGPLQPGLPSPAMIPKDWPLIIIDLKDCFF...TIPLAEQDCEKFAFTIPAINNKE..PATRF   81  ORF
HML-2.2   --------RM---------I--------S------------------------...----------------L------..-----   81  ORF
HML-2.3   ---T--R-RM-------------K-H----M-------R--------------...-----------------------..-----   81  ORF
HML-2.4   ---T--R-RM-------------K-H---------------------------...----E---F--------------..-----   81  ORF
HML-2.5   ---R-SR-RM------I-------R-I-T---------NAG---------V-...---------------------/...-----   80  frameshift
HM16      ------R-RM------I-------R----V-----T-.../-------------...--S--K--F--------------..-----   77  frameshift
HML-2.6   ------R-RIV--------------------------E-------E-------...-----K--F--L-----------..-----   81  ORF
HML-3.1   -K-N----RM------I-S---------L----------N-----T---S--...----------W-------V--LQ..S-KH-   81  ORF
HML-3.2   -K------RM------I-S------T-------------N----V--------...----------R-------V--LQ..-TK--   81  ORF
HML-3.3   -K------RM------I-SL-----T-------------N---VV-----S--...-----/----W---I---V--LQ..--KH-   80  frameshift
HML-3.4   -K-----RKM------I-S------T---R---------N----V--------...----------W-V---L-V--LQLK--K--   83  ORF
HML-3.5   -K------RM-P----I-SL-----T-------------K--F-V--------...----------Q----K--/.......--K-I   76  frameshift
HML-3.6   -KQ-----R-------I-S-----ET--L--------L-N--S-V---E----...ASS-------S-------V--LK..SGKN-   81  ORF
HML-4.1   --------RM----------L----T---S----S--TE--------------...-----P--F------V--L--VA..--A--   81  ORF
HML-5.1   -P-R----RL-H--C-I--NL-------Q-----TA-.RH--I-VT-------...------E-RVE----------ER..--PX-   79  stop
MMTV      -K------RL-Q-------TMHD--A--------VAV--G-EI-----Q----...N-K-HPE--KR---SV-SP-F-R..--Q--   81  ORF
HML-6.1   -P------GL-H----I--Q-K---A--Q----L-G--R-/---VV---.IV--...--L-HQ--RPG---SV-SV-QR-..-VSC-   80  frameshift
HML-6.2   DP-M----QL-H----IK/Q-K-V-A--Q--S-S-A--R--T-VV-G--EF--/..N---HKK-KPQ---SV-S--QR-..-VS-S   81  frameshifts
HML-6.3   FP----R-RL-H----IHVH-KL--A--KLS----A--R----VV--------...-T--H-K-KPX---SV-S--QR-..-VSH-   80  stop
IAPm      -K------RL-H---PI-EQMNLF--V-R---VLSAL-RG-N-----I------...S---CPR-RPR------S--SD-..-DN--   81  ORF
IAPh      -K------RL-H----I-NQMHLF--V-R---LLSAL-Q--K-----I------...S----PR-RPR------SL-HM-..-E--D   81  ORF
          * * ** *  * **** *      *  *  ***    *        ***** **   * *   *   ***  *  *    *

          * * **  *** *** *       *  * * *     *        *  ****   * *   *   ***  *  *    *  *
MPMV      -K------RL-Q-------TMVL--A--------VA--QGYLK----------...S---HPS-QKR---SL-ST-F--..-MQ--   81  ORF
SMRV-H    -K----S-RL-Q------K-MV---A--------VA--LNYHK-V--------...----HPE-RPY---SV-Q--FQS..-MP-Y   81  ORF
RSV       -R-A--SYRL-H-------KLV-F-AV-Q-A-VLSAL-RG---MVL-------...S-------R-A----L-SV--QA..--R--   81  ORF
```

FIG. 3. Amino acid sequences translated from the nucleotide sequences (Fig. 1). All sequences are compared with HERV-K10 (43). Hyphens indicate identical amino acids, and gaps are shown with dots. Stars correspond to conserved residues (6 of 6 amino acids) of MMTV (37), IAPm (36), IAPh (42), MPMV (52), SMRV-H (39), and RSV (48), whereas asterisks indicate well-conserved residues (19 of 22 amino acids) of HERV-K10, HM16 (12), and the HML sequences. Frameshifts are indicated with a slash; stop codons are shown with an X.

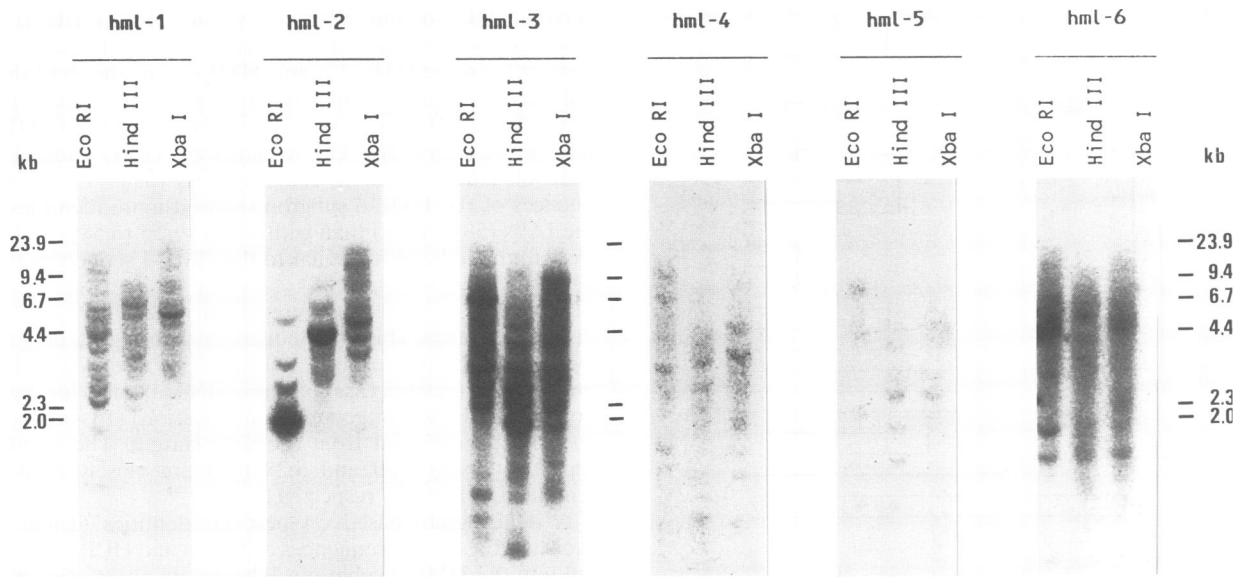

FIG. 4. Southern blot hybridizations of 15 μg of human DNA digested with the restriction enzymes indicated and probed with hml-1 to hml-6 at 50°C.

have ORFs interrupted by frameshifts or by stop codons. Subgroups HML-1 to HML-4 include altogether 10 members with an ORF, and these could thus be part of a functional RT-coding sequence. However, none of the identified members of subgroup HML-5 or HML-6 could encode a functional RT.

The six retroviruses MMTV, IAPm, IAPh, MPMV, squirrel monkey retrovirus H (SMRV-H [39]), and Rous sarcoma virus (RSV [48]) share 31 conserved amino acid positions over this region in RT (indicated with stars in Fig. 3). There is a correlation between conserved residues of the HML sequences (indicated with asterisks in Fig. 3) and those of the infectious viruses (26 of 31 positions). The ratios of nucleotide to amino acid substitution, compared with MMTV, as measured over the conserved 93 nucleotide and 31 amino acid positions were found to be high (10/1 to 20/1 in members of subgroups HML-1 to HML-3) to relatively high (7/1 to 8/1 in members of subgroups HML-4 to HML-6). The ratios of nucleotide to amino acid substitution over all positions (both conserved and nonconserved residues) were lower for all sequences (2/1 to 4/1). At least over this region of RT, the high frequency of synonymous nucleotide substitutions as opposed to replacement nucleotide substitutions suggests a functionally conserved gene.

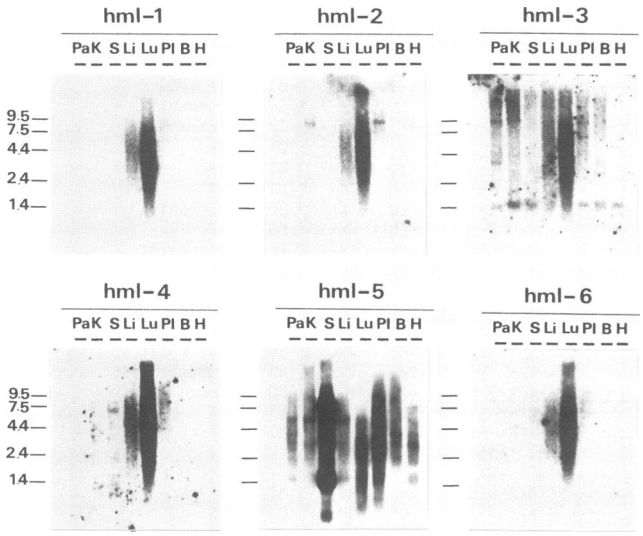

FIG. 5. Multiple-tissue Northern blots (Clontech) with 2 μg of poly(A)$^+$ RNA from pancreas (Pa), kidney (K), skeletal muscle (S), liver (Li), lung (Lu), placenta (Pl), brain (B), and heart (H), probed with hml-1 to hml-6 at 45°C. Sizes (in kilobases) are indicated at the left.

Organization of HML sequences in human DNA. Cellular DNA derived from peripheral blood mononuclear cells was digested with restriction enzymes and analyzed by Southern hybridization with probes hml-1 to hml-6 (Fig. 4). The patterns of hybridization were different for the six probes, and many bands were observed for probes hml-1 to hml-3 and hml-6. The hybridization conditions were determined from the dot blot hybridizations, in which each of the probes hybridized only to clones of one subgroup (Table 3).

The most intense hybridization was obtained with probe hml-3: about 15 to 20 bands were visualized with the restriction enzymes used. Probe hml-1 hybridized to about 15 bands of *Eco*RI-digested DNA, and probe hml-2 hybridized to 10 to 15 restriction fragments in *Hin*dIII and *Xba*I lanes. Fewer bands

TABLE 5. Hybridization patterns of probes hml-1 to hml-6 toward poly(A)$^+$ RNA of normal human tissues

Sizes (kb)[a]	Hybridization pattern[b]							
	Pancreas	Kidney	Skeletal muscle	Liver	Lung	Placenta	Brain	Heart
>9.5	3	3, 5	5	3	2–4, 6			
9–8	3, 5	2, 3, 5	5	4	2–4, 6	2, 5	5	
7.5–7			4, 5	4	1–4, 6	5		5
5	5	5	5	4, 5	1–6	5	5	5
3.5–3	3, 5	3, 5	5	5	1–6	5	5	5
1.4	3, 5	3	3, 5	3, 5	3–5	3, 5	3	3, 5

[a] Approximate sizes of the transcripts.
[b] The numbers 1 to 6 correspond to probes hml-1 to hml-6 (Table 1).

(about five) were observed with probes hml-4 and hml-5. Probe hml-6 hybridized to a large number of fragments.

Probe hml-2 hybridized intensively with *Eco*RI-digested DNA of about 2, 2.7, 3.6, and 4.7 kb. The three smaller fragments recognized by probe hml-2 may be the same as the 2-, 2.9-, and 3.7-kb bands detected by Ono (40) when a HERV-K-derived probe was used for hybridization on *Eco*RI-digested DNA, suggesting that probe hml-2 detected sequences closely related to the HERV-K family (the HML-2 subgroup in Fig. 1).

The number of bands detected in human DNA with use of probes hml-1 and hml-3 to hml-5 agreed with the number of clones isolated from these subgroups (Table 3). However, sequences belonging to subgroup HML-6 were detected in only four of the cDNA clones, whereas a large number of fragments were detected in human DNA. In contrast, subgroup HML-2 contains fewer copies in the human DNA than does subgroup HML-3, but the numbers of clones isolated were approximately the same. A lower degree of homology between the PCR primers and the target regions for some subgroups may account for this discrepancy.

Expression of HML sequences in normal human tissues. Probes hml-1 to hml-6 were hybridized to membrane-bound poly(A)$^+$ RNA of eight different normal tissues. The patterns of hybridization differed among the probes. As shown in Fig. 5, a strong hybridization was obtained for most probes in lung, but hybridizing transcripts were also visualized in skeletal muscle, placenta, pancreas, and kidney for several probes. A variety of transcripts (most frequently of about 12, 8, 7, and 1.4 kb) were seen in one or several tissues. The results are summarized in Table 5.

The strongest hybridization was obtained with probes hml-2 to hml-5. Probe hml-2 hybridized to transcripts of about 8 to 9 kb in kidney, lung, and placenta. More abundant were RNAs of about 2 to 7 kb in lung. The most intense hybridization signals detected by probe hml-3 were obtained from pancreas, kidney, and lung. Transcripts detected by this probe were about 12 kb (pancreas, kidney, liver, and lung), 8 to 9 kb (pancreas, kidney, and lung), 3 to 3.5 kb (pancreas and lung), and 1.4 kb (in all tissues). Transcripts of about 2 to 7.5 kb were visualized in a smear of hybridizing RNAs in liver and lung. Probe hml-4 recognized RNA most abundantly in liver (2 to 9 kb) and lung (about 1.4 to 12 kb), but transcripts of about 7 kb were also detected in skeletal muscle. Probe hml-5 detected RNA transcripts of 7 to 9 kb in pancreas, kidney, skeletal muscle, placenta, brain, and heart, about 5 and 3 to 3.5 kb in all tissues, and about 1.4 kb in all tissues except kidney and brain. The strong Northern hybridization signals obtained with probe hml-5 contrasted the weak signal on Southern blots (cf. Fig. 4).

Probes hml-1 and hml-6 hybridized mainly to lung RNA. For probe hml-1, transcripts of about 1.4 to 7 kb were obtained. Hybridization for probe hml-6 was seen in RNAs at about 2 to 12 kb.

Hybridizations were also performed with the same probes (hml-1 to hml-6) against Clontech Northern blots from another batch (data not shown). The patterns of transcripts detected by the probes were almost the same as those shown in Fig. 5. However, the strong signal seen in lung with probes hml-1 to hml-4 and hml-6 were weaker on these membranes, whereas all probes hybridized to a smear of RNAs of about 4 to 9 kb in the liver tissue lane. Hybridizations were also performed with a β-actin probe (Clontech) and the PCR primers that were used in amplification and isolation of all clones. The sense and antisense PCR primers were end labeled and were hybridized to the membranes. Only the RNA strand-specific (antisense-stranded) primer resulted in hybridization to the immobilized RNA, and transcripts of about 8 kb were seen in all tissues. The sense-stranded primer gave no hybridization (data not shown), which eliminates the possibility that the filters were contaminated with DNA. The β-actin probe (Clontech) recognized 2-kb transcripts at about the same level in all tissues. However, the most intense signals were obtained from placental tissues on all membranes used (data not shown).

DISCUSSION

In this study, the nucleotide sequences of RT-coding sequences from the human genome are reported. All were similar to HERV-K10, MMTV, and IAPm. Primers derived from conserved regions of RT were used for PCR amplification on human lymphocyte DNA and cDNA (Table 2). This was shown to be an effective approach to identify new HML sequences. Six distinct subgroups of related sequences were amplified by this single primer pair. The sequences of HML-1 and HML-3 to HML-6 have not been reported previously. HML-2 contains the described sequences of HM16 and HERV-K10.

Franklin et al. (15) reported that nine groups of MMTV-related sequences are present in the human DNA. Although we screened a total of 150 recombinant clones, we detected only six subgroups. There are several possibilities to account for this. A poor primer-target homology and/or a too stringent annealing temperature between the primer and target sequences could have led to the omission of some sequences. Several groups detected in the study by Franklin et al. (15) were found in only a minority of the recombinants (their groups 8 and 9 were identified in one recombinant each, and groups 6 and 7 were identified in three recombinants each). We may not have screened enough recombinants. Otherwise, some of the groups detected by Franklin and coworkers (15) could have been, by our definition, in the same subgroup.

The six probes (hml-1 to hml-6; Table 1) used for further characterization of the HML sequences were all specified from sequenced PCR-amplified clones. Each of the oligonucleotide probes detected only sequences corresponding to one subgroup, which was determined by dot blot hybridizations in which each of the 150 clones was defined into a specific subgroup (Table 3). Southern and Northern analyses resulted in different patterns of hybridization for the various probes, which indicates that each of them detects a specific set of sequences.

Southern blots revealed that subgroups HML-1 to HML-3 and HML-6 are most frequent in human DNA, whereas the HML-4 and HML-5 sequences are less abundant. All six subgroups are endogenous to human DNA, and together they make up at least 60 to 80 copies per haploid human genome (Fig. 4).

The transcriptional activity of human sequences similar to MMTV was recently studied in cultured breast carcinoma cell lines and placenta (15, 41). The expression of these sequences showed a complex pattern of RNA transcripts in sizes from 1.2 to 12 kb. We have been interested in defining the expression of HML sequences in tissues and cells of healthy individuals. We recently described such expression in lymphocytes of blood donors (35), confirmed by Krieg et al. (23) and Brodsky et al. (6). Because of the large number of HML sequences in the human genome and their expression in normal cells, a definition of the pattern of HML sequences in normal cells is necessary if they are to be studied in human disorders. Probes hml-1 to hml-6 detected RNAs of different sizes and in different tissues (summarized in Table 5). Probes hml-1 and

hml-6 recognized transcripts mainly in lung, whereas the other probes detected RNAs in most tissues tested. However, hml-2 preferentially detected transcripts in kidney, lung, and placenta and hml-3 detected transcripts mainly in pancreas, kidney, and lung, whereas hml-4 hybridized to skeletal muscle, liver, and lung and hml-5 recognized transcripts most strongly in skeletal muscle and placenta. These data show that the sequences of the HML subgroups are differentially expressed in normal tissues. Hormone regulation may be one of several factors influencing the differential expression of these sequences, as has been shown for HERV-K10 (41). Tissue-specific expression of HERVs has been shown for ERV-3 (21) and for NMWV4 and NMWV9 in breast carcinoma cell lines (15). However, it is quite possible that HMLs are expressed in some of the tissues at a level which is not detectable by Northern analysis. It has been demonstrated that RT-PCR is a more sensitive tool for detection of expression than Northern blotting (51), but the advantage of Northern analysis is that information on the size and abundance of transcripts is obtained.

A higher level of expression was detected by several probes in liver on Clontech Northern blots of a batch other than those shown in Fig. 5 (see Results). The level of HERV expression may vary between individuals (23). In addition, the pattern of expression can differ between normal and malignant cells (6). Whether this explains our observed variation is unclear.

Large transcripts (8 to 9 kb) were detected by different probes in all tissues except heart. Transcripts of these sizes may represent full-length endogenous proviral transcripts, as has been found in breast carcinoma cell lines and placental tissue (15, 41). Transcripts of about 12 kb were found in several tissues. These large RNAs could represent full-length retroviral transcripts containing downstream cellular sequences, as has been observed for several ERV transcripts (15, 21, 29, 41). Transcripts of 5, 3.5, 3, and 1.4 kb were detected in several tissues. Small transcripts were also found by Franklin et al. (15) with several of their probes. To our knowledge, there has been no characterization of small transcripts containing *pol* sequences.

The expression of these ERV elements may have several functional implications. Transcripts of defective endogenous sequences might recombine with sequences of exogenous or nondefective endogenous retroviruses, resulting in new infectious or oncogenic recombinants. The ability to code for functional Pol proteins may contribute to transposition of these sequences into new locations in the genome, which may lead to insertional activation of cellular genes. Such events have been observed for IAPm in activation of the c-*mos* proto-oncogene (10) and by L-1 elements in somatic tissues (38) and in the germ line (13, 22).

The biological significance of the expression of HML sequences and HERVs in general is uncertain. Several murine ERVs, which have been subjected to more detailed analysis than have those of humans, have been shown to be involved in mammary tumor development, in autoimmune diseases, and in immune regulation by clonal stimulation or deletion of T cells mediated by murine retroviral proteins (superantigens) (for a review, see reference 24). The observation that nucleotide substitutions at positions corresponding to conserved amino acids of type A, type B, type D, and avian type C retroviruses preferentially were synonymous strengthens the supposition that some of them encode an RT, whose function is beneficial either for the host or for the virus. Data presented here show that MMTV-like sequences are extensively and differentially expressed in tissues of healthy humans and may provide a basis for the search of HERV sequences associated with human disorders.

ACKNOWLEDGMENTS

We thank Alistair Kidd for helpful discussions and support and Mats Lindeskog for valued contributions in this study. We also thank Hans Thornquist, Lena Stensson-Holst, and Gunilla Henningsson, Department of Medical Chemistry IV, Lund, Sweden, for introduction to and help with image analysis, Hugh Connell, Department of Clinical Immunology, Lund, Sweden, for help with Southern blotting, and Elzbieta Vincic for technical assistance.

This work was supported by funds at the Medical Faculty of Lund (grants 107, 151, and 194), the Crafoord and Österlund foundations, the Swedish Medical Research Council (grant B 92-27 X-09528), and the Swedish Society against Cancer (grant 2054-B87-03X).

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, New York.
2. **Baltimore, D.** 1985. Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. Cell **40**:481–482.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Bonner, T. I., C. O'Connell, and M. Cohen.** 1982. Cloned endogenous retroviral sequences from human DNA. Proc. Natl. Acad. Sci. USA **79**:4709–4713.
5. **Brack-Werner, R., C. Leib-Mösch, T. Werner, V. Erfle, and R. Hehlmann.** 1989. Human endogenous retrovirus-like sequences, p. 464–477. *In* R. Neth (ed.), Modern trends in human leukemia VIII. Springer-Verlag, Berlin.
6. **Brodsky, I., B. Foley, D. Haines, J. Johnston, K. Cuddy, and D. Gillespie.** 1993. Expression of HERV-K proviruses in human leukocytes. Blood **81**:2369–2374.
7. **Callahan, R.** 1988. Two families of human endogenous retroviral genomes. Eukaryotic transposable elements as mutagenic agents. Banbury Rep. **30**:91–100.
8. **Callahan, R., I. M. Chiu, J. F. Wong, S. R. Tronick, B. A. Roe, and S. A. Aaronson.** 1985. A new class of endogenous human retroviral genomes. Science **228**:1208–1211.
9. **Callahan, R., W. Drohan, S. Tronick, and J. Schlom.** 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. USA **79**:5503–5507.
10. **Canaani, E., O. Dreagen, A. Khar, G. Rechavi, D. Rain, J. B. Cohen, and D. Givol.** 1983. Activation of the c-*mos* oncogene in a mouse plasmacytoma by insertion of an endogenous intracisternal A-particle genome. Proc. Natl. Acad. Sci. USA **80**:7118–7122.
11. **Coffin, J. M.** 1982. Endogenous viruses, p. 1109–1204. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. **Deen, K. C., and R. W. Sweet.** 1986. Murine mammary tumor virus *pol*-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus *pol* gene. J. Virol. **57**:422–432.
13. **Economou-Pachnis, A., M. A. Loshe, A. V. Furano, and P. N. Tsichlis.** 1985. Insertion of long interspersed repeated elements at the *Igh* (immunoglobulin heavy chain) and *Mlvi-2* (Moloney leukemia virus integration 2) loci of rats. Proc. Natl. Acad. Sci. USA **82**:2857–2861.
14. **Finnegan, D. J.** 1985. Transposable elements in eukaryotes. Int. Rev. Cytol. **93**:281–326.
15. **Franklin, G. C., S. C. Chretien, I. M. Hanson, H. Rochefort, F. E. B. May, and B. R. Westley.** 1988. Expression of human sequences related to those of mouse mammary tumor virus. J. Virol. **62**:1203–1210.
16. **Gelfand, D. H., and T. J. White.** 1990. Thermostable DNA polymerases, p. 129–141. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols—a guide to methods and applications. Academic Press, San Diego, Calif.
17. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73**:237–244.
18. **Higgins, D. G., and P. M. Sharp.** 1989. Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS **2**:151–153.

19. **Hirose, Y., M. Takamatsu, and F. Harada.** 1993. Presence of env genes in members of the RTVL-H family of human endogenous retrovirus-like elements. Virology **192:**52–61.
20. **Horn, T. M., K. Huebner, C. Croce, and R. Callahan.** 1986. Chromosomal locations of members of a family of novel endogenous human retroviral genomes. J. Virol. **58:**955–959.
21. **Kato, N., S. Pfeifer-Ohlsson, M. Kato, E. Larsson, J. Rydnert, R. Ohlsson, and M. Cohen.** 1987. Tissue-specific expression of human provirus ERV3 mRNA in human placenta: two of the three ERV3 mRNA contain human cellular sequences. J. Virol. **61:** 2182–2191.
22. **Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis.** 1988. Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. Nature (London) **332:**164–166.
23. **Krieg, A. M., M. F. Gourley, D. M. Klinman, A. Perl, and A. D. Steinberg.** 1992. Heterogeneous expression and coordinate regulation of endogenous retroviral sequences in human peripheral blood mononuclear cells. AIDS Res. Hum. Retroviruses **12:**1991–1998.
24. **Krieg, A. M., M. F. Gourley, and A. Perl.** 1992. Endogenous retroviruses: potential etiologic agents in autoimmunity. FASEB J. **6:**2537–2544.
25. **La Mantia, G., D. Maglione, G. Pengue, A. Di Cristofano, A. Simeone, L. Lanfrancone, and L. Lania.** 1991. Identification and characterization of novel human endogenous retroviral sequences preferentially expressed in undifferentiated embryonal carcinoma cells. Nucleic Acids Res. **19:**1513–1520.
26. **Lania, L., A. Di Cristofano, M. Strazzullo, G. Pengue, M. Barbara, and G. La Mantia.** 1992. Structural and functional organization of the human endogenous retroviral ERV9 sequences. Virology **191:**464–468.
27. **Larsson, E., N. Kato, and M. Cohen.** 1989. Human endogenous proviruses. Curr. Top. Microbiol. Immunol. **148:**115–132.
28. **Leib-Mösch, C., R. Brack-Werner, T. Werner, M. Bachmann, O. Faff, V. Erfle, and R. Hehlmann.** 1990. Endogenous retroviral elements in human DNA. Cancer Res. **50:**5636–5642.
29. **Lindeskog, M., P. Medstrand, and J. Blomberg.** 1993. Sequence variation of human endogenous retrovirus ERV9-related elements in an *env* region corresponding to an immunosuppressive peptide: transcription in normal and neoplastic cells. J. Virol. **67:**1122–1126.
30. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Mariani-Constantini, R., T. M. Horn, and R. Callahan.** 1989. Ancestry of a human endogenous retrovirus family. J. Virol. **63:**4982–4985.
32. **May, F. E. B., and B. R. Westley.** 1986. Structure of a human retroviral sequence related to mouse mammary tumor virus. J. Virol. **60:**743–749.
33. **May, F. E. B., B. R. Westley, H. Rochefort, E. Buetti, and H. Diggelmann.** 1983. Mouse mammary tumor related sequences in human DNA. Nucleic Acids Res. **11:**4127–4139.
34. **McClure, M. A., M. S. Johnson, D. F. Feng, and R. F. Doolittle.** 1988. Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. Proc. Natl. Sci. USA **85:**2469–2473.
35. **Medstrand, P., M. Lindeskog, and J. Blomberg.** 1992. Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. J. Gen. Virol. **73:**2463–2466.
36. **Mietz, J. A., Z. Grossman, K. K. Lueders, and E. L. Kuff.** 1987. Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. J. Virol. **61:**3020–3029.
37. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. **61:**480–490.
38. **Morse, B., P. G. Rotherg, V. J. South, J. M. Spandorfer, and S. M. Astrin.** 1988. Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma. Nature (London) **333:**87–90.
39. **Oda, T., S. Ikeda, S. Watanabe, M. Hatsushika, K. Akiyama, and F. Mitsunobu.** 1988. Molecular cloning, complete nucleotide sequence, and gene structure of the provirus genome of a retrovirus produced in a human lymphoblastoid cell line. Virology **167:**468–476.
40. **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.
41. **Ono, M., M. Kawakami, and H. Ushikubo.** 1987. Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. J. Virol. **61:**2059–2062.
42. **Ono, M., H. Toh, T. Miyata, and T. Awaya.** 1985. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. J. Virol. **55:**387–394.
43. **Ono, M., T. Yasunaga, T. Miyata, and H. Ushikubo.** 1986. Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J. Virol. **60:**589–598.
44. **Power, M. D., P. A. Marx, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science **231:**1567–1572.
45. **Rogers, J. H.** 1985. The structure and evolution of retroposons. Int. Rev. Cytol. **93:**231–279.
46. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, G. T. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. Science **239:**487–491.
47. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
48. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32:**853–869.
49. **Shih, A., E. E. Coutavas, and M. G. Rush.** 1991. Evolutionary implications of primate endogenous retroviruses. Virology **182:** 495–502.
50. **Shih, A., R. Misra, and M. G. Rush.** 1989. Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: relation to primate retroviruses. J. Virol. **63:**64–75.
51. **Siebert, P. D., S. C. Wright, and J. Larrick.** 1992. Examination of brain-type nitric oxide synthase expression in non-neural tissue by RT-PCR. Clontechniques **7:**1–3.
52. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell **45:**375–385.
53. **Steele, P. E., M. A. Martin, A. B. Rabson, T. Bryan, and S. J. O'Brien.** 1986. Amplification and chromosomal dispersion of human endogenous retroviral sequences. J. Virol. **59:**545–550.
54. **Teich, N., J. Wyke, and P. Kaplan.** 1985. Pathogenesis of retrovirus-induced disease, p. 187–248. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
55. **Teich, N., J. Wyke, T. Mak, A. Bernstein, and W. Hardy.** 1982. Pathogenesis of retrovirus-induced disease, p. 785–998. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
56. **Temin, H. M.** 1985. Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts. Mol. Biol. Evol. **6:**455–468.
57. **Thayer, R. M., M. D. Power, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1987. Sequence relationship of type D retroviruses which cause simian acquired immunodeficiency syndrome. Virology **157:**317–329.
58. **Weiner, A. M., P. L. Deininger, and A. Efstratiadis.** 1986. Nonviral retroposons. Annu. Rev. Biochem. **55:**631–661.
59. **Westley, B., and F. E. B. May.** 1984. The human genome contains multiple sequences of varying homology to mouse mammary tumor virus DNA. Gene **28:**221–227.