© 1991 Oxford University Press Nucleic Acids Research, Vol. 19, No. 7 1513

Identification and characterization of novel human endogenous retroviral sequences prefentially expressed in undifferentiated embryonal carcinoma cells

Girolama La Mantia, Domenico Maglione, Gina Pengue, Antonio Di Cristofano, Antonio Simeone[1], Luisa Lanfrancone[2] and Luigi Lania*

Department of Genetics, General and Molecular Biology, University of Naples, Via Mezzocannone 8, 80124 Naples, [1]International Institute of Genetics and Biophysics, CNR, Naples and [2]Institute of Medical Clinics I, University of Perugia, 06100 Perugia, Italy

Received December 31, 1990; Revised and Accepted March 13, 1991 EMBL accession nos X57147, X57148

ABSTRACT

A novel endogenous retroviral sequence (ERV-9) has been isolated from a human embryonal carcinoma cDNA library by hybridization to a probe containing a recently described human repetitive element. DNA sequence analysis of the 4kb cDNA insert (pHE.1) revealed the presence of ORFs potentially coding for putative retrovirus-related *gag, pol* and *env* proteins. Northern blot and RNase protection experiments showed that RNA homologous to the pHE.1 insert is detected only in embryonal carcinoma cells as a 8 kb mRNA, and its expression is negatively regulated during retinoic acid induced differentiation of the human teratocarcinoma cell line NT2/D1. Using a *pol* specific probe we have isolated a genomic locus containing the ERV-9 sequences. Characterization by restriction enzyme analysis and DNA sequencing allowed us to define LTR-like sequences, that are composed by a complex array of subrepetitive elements. In addition we show that ERV-9 LTR sequences are capable to drive expression of linked CAT gene in a cell specific manner as LTR promoter activity has been detected only in NT2/D1 cells.

INTRODUCTION

Analysis of gene regulation during early mammalian embryogenesis has been greatly aided by the availability of the embryonal carcinoma (EC) cell system, and by the ability to clone genes that are differentially expressed during development. The human EC cells NT2/D1 represent a subclone of the teratocarcinoma cell line TERA-2, which irreversibly differentiates into multiple cell types, upon addition of retinoic acid (RA) to the culture medium (1).

The regulation of several cloned genes during retinoic acid (RA) induced differentiation has been reported. For example the expression of the transcription factor AP-2 and some homeobox genes has been recently shown to be enhanced upon differentiation of NT2/D1 cells (2,3,4), whereas the expression of the transcription factor Oct-3 (5,6) and the K-fgf proto-oncogene (7) is restricted to the undifferentiated embryonal carcinoma cells. In addition, Pol II and Pol III transcripts derived from repeated sequences are often much more prevalent in EC cells than after differentiation (8, 9). Moreover, we have recently identified cDNA clones containing a novel class of repetitive sequences that are predominantly expressed in undifferentiated NT2/D1 cells, and whose expression is negatively regulated upon RA induced differentiation. Nucleotide sequencing of the cloned cDNAs indicated that the transcriptional units are composed by a complex array of repeated DNA elements with no obvious open reading frame (10).

To further define the nature of these transcriptional units we have analyzed additional cDNA clones. The characterization of the longest clone, pHE.1, revealed the presence of repeated elements in the 3′ terminal portion; moreover, sequences containing ORFs potentially coding for putative retroviral proteins *gag*, *pol* and *env*, extend to the 5′ portion of pHE.1 cDNA clone. However, clone pHE.1 contains nucleotide substitutions and deletions that introduce stop codons or changes in the reading frames. The pHE.1 sequences differ from the reported human endogenous retroviral sequences (11–19), defining a new family of endogenous retroviral sequences that we have named ERV-9. Using the *pol* region as probe we have isolated genomic clones, and here we report the characterization of one member of the ERV-9 family. Restriction enzymes analysis and DNA sequencing revealed the presence of long LTR-like sequences that are composed by a complex array of the sub-repetitive elements present also in previously described cDNAs (10).

Human genome contains multiple copies of endogenous retroviral sequences (HER) related to mouse and primate retroviral sequences. Several human endogenous provirus homologous to mammalian B, or C retrovirus, or unidentifiable retroviral proviruses have been isolated and characterized

* To whom correspondence should be addressed

(11–19). Despite the abundant occurrence of HER elements their functional significance is yet not known. In murine cells transposition of endogenous retrovirus-like elements have been implicated in both gene inactivation (by insertion mutagenesis) (20,21) and in gene activation (22,23). Rearrangements involving HER members may contribute to alterations in human gene expression.

Transcriptional studies indicate that at least one member of the the ERV-9 family is transcribed in the embryonal carcinoma cell line NT2/D1, and its expression is negatively regulated during in vitro induced differentiation. Using sensitive RNase protection experiments we fail to observe detectable ERV-9 expression in a variety of cell lines of different embryological derivation. In addition we have shown that ERV-9 LTR sequences are capable to drive expression of linked CAT gene and the promoter activity has only been detected in NT2/D1 cells. These findings suggest that the tissue-specific expression of ERV-9 sequences might be due to regulatory sequences present within the LTR sequences.

MATERIALS AND METHODS

Cell Cultures

The cloned human EC cell line NT2/D1 was mantained in high glucose Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. Differentiation was induced by seeding cells at 5×10^5 per 75 cm^2 flask in 10^{-5} M retinoic acid (all-trans, Sigma Chemical Co.). Cells were refeed every 4 days with RA-containing medium and analyzed after 14 days as previously described (3–4). Briefly the EC stem cell markers SSEA-3, SSEA-4 and TRA1-60, and two antigens ME311 and A2B5 which are expressed in EC-derived differentiated cells, were analysed using the corresponding monoclonal antibodies. Undifferentiated cells, as monitored by immunochemical staining, decrease steadily after the addition of RA and are <2% after 7 days of treatment. The cell lines used in transfection experiments and for the preparation of RNA were obtained from the American Type Cell Collection.

DNA and RNA analysis

Total RNA was extracted by the guanidine-isothiocyanate technique, selected for poly $(A)^+$ by one passage on a oligo(dT) cellulose column according to standard procedures, run in 3 μg aliquots on 1% agarose-formaldheyde gel and transferred to nitrocellulose filter. The blot was hybridized to DNA probe labelled by random priming to a specific activity of 5×10^8 cpm/μg in a buffer containing 50% formamide, 5×SSPE, 5×Denhardt's, 0.1% SDS and 100 μg/ml salmon sperm DNA. The filters were washed under stringent conditions (0.1×SSC, 0.2% SDS at 65°C for 2 h) and exposed to Kodak XAR-5 film for 3 days. RNase protection analysis was performed on 20 μg of total RNA. The *pol* specific probe BglII-BamHI (nt. 1390–1600) and a 84 bp long probe from the 3′-end non coding region of the human β-actin gene (4, 24) were subcloned in pGEM vectors (Promega Biotec.) and antisense strand RNA probes were synthesized with Sp6 polymerase. RNase digestion and electrophoresis on 7% urea/polyacrylamide gels were carried out as described (4). Southern blot analysis was performed using human DNA isolated from peripheral blood lymphocytes. The blot was hybridized to the *pol* specific Bam HI fragment from nt. 1190 to nt. 1600 and the probe was labelled by the multiprime labelling kit (Amersham) to a specific activity of 5×10^8 cpm/μg. The filter was washed under stringent conditions (0.1×SSC, 0.2% SDS at 65°C for 60 min) and exposed to Kodak XAR-5 film for 2 days.

DNA library screening and DNA sequencing

The cDNA library from undifferentiated human teratocarcinoma cell line NT2/D1 was kindly provided by J. Skowronsky. 3×10^5 plaques were screened with the 3′-end portion of the cDNA pTR2 (10) under stringent conditions. The longest cDNA clone (pHE.1) was further analyzed after subcloning in pGEM plasmids. The DNA sequences were determined using progressive deletions made by the ExoIII-S1 nucleases, using the Erase-a-Base System (Promega, Biotec). DNA sequencing was determined by the dideoxynucleotide chain termination method using the Sequenase kit (United States Biochemical Co.). The sequences of both strands were determined. Nucleotide sequencing data were analyzed using the MicroGenie Sequence Analysis Program software (Beckman). Genomic clones were isolated after screening with a *pol* specific probe (BglII-BamHI nt. 1390–1600) a lambda Fix human lung fibroblast genomic library (Stratagene). 10^5 plaques were screened and 20 randomly choosen positive plaques were isolated. The recombinant phage λFix1.1 was further analyzed by restriction enzymes analysis and DNA sequencing of relevant regions was performed after subcloning of appropriate fragments in pGEM-3 plasmid.

Plasmid constructions and CAT assays

The expression vectors pRSV-CAT, and p8-CAT-0 have been previously described (25,26). The pR1-CAT was constructed by inserting the PvuII-HindIII fragment present in the 5′ LTR sequences of the genomic clone λFix 1.1 in the Sma-HindIII sites of p8-CAT-0. The pHE.1 cDNA 3′ LTR was first isolated as a PvuII-EcoRI fragment and cloned in pGEM-3; then, since this LTR had functioned to polyadenylate transcripts, it terminated in a poly A tail which was removed by exonuclease III, and subcloned in pGEM-3 plasmid. The 3′ LTR sequences were then isolated by digestion with PvuII and a second enzyme (HindIII) specific of pGEM-3 plasmid and this fragment was inserted in both orientation in the HindIII site of p8-CAT-0 via HindIII linkers. The two CAT plasmids were named pHE1-CAT (5′-3′ orientation) and pHE2-CAT (3′-5′ orientation). DNA transfection was performed by calcium phosphate precipitation using subconfluent cell cultures. Cell extracts were isolated 44 hours after transfection and assayed for CAT activity using 200μg of total proteins extract. The products of reaction were chromatographed by ascending TLC and quantitated by scintillation counting. Experiments were repeated 5 to 8 times and average percentage conversion values determined.

RESULTS

Sequence analysis of pHE.1 cDNA clone

In the course of characterizing human cDNA clones containing repeated sequences we have screened the NT2/D1 cDNA library with a probe corresponding to the repeated elements present in the previously described clone pTR2 (10). The longest isolated cDNA clone, pHE.1, has been further analyzed and the entire nucleotide sequence is reported in Fig 1. The portion of pHE.1 from nt 3314 to the end of the clone is identical to the 3′-end portion of the cDNA clone pTR2 (10), and contains both a

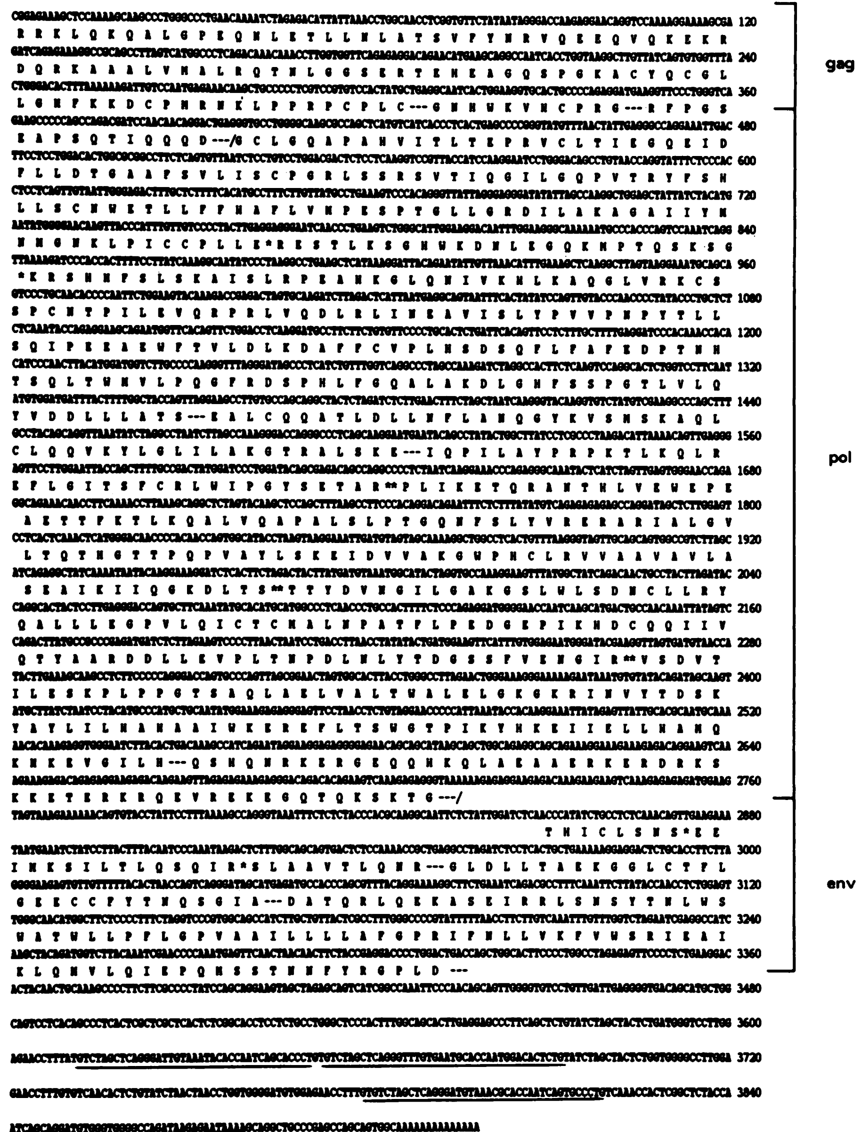

Figure 1. Nucleotide and putative amino acids sequences of pHE.1 cDNA clone. The regions of homology to *gag*, *pol* and *env* proteins are indicated. Dashed lines indicate termination codons. Adjustments for nucleotide deletions were made to mantain a reading frame possessing colinear amino acid homology with known *gag*, *pol* and *env* proteins; these are indicated by asterisks. Assignment of putative *gag*, *pol* and *env* regions is based upon alignment of the deduced amino acids with those of other known retroviruses as discussed in the text. The 41bp repeated element is underlined. On the bottom is reported a physical map of clone pHE.1 with primary restriction sites.

polyadenylation signal and a poly A tail. This fragment does not appear to code for proteins due to the presence of numerous termination codons.

Computer-assisted comparison of the pHE.1 sequences to all sequences contained in the GenBank and EMBL databases indicated various extents of homologies to retroviral regions coding for the *gag*, *pol* and *env* proteins, and it was evident that the pHE.1 sequences are part of a novel human endogenous retroviral family, that we have named ERV-9, accordingly to the Human Gene Mapping nomenclature.

To determine a possible functional relationship between pHE.1 sequences and the human retroviruses, the predicted amino acids sequences were analyzed. The assignment of putative retroviral protein coding regions has been based upon alignment of deduced amino acids sequences with those of MoMLV, AKV, and both human endogenous and exogenous retroviruses. Termination codons or point deletions causing reading frame shifts are present in pHE.1 clone. Adjustments for nucleotide deletions were made to mantain a reading frame possessing colinear amino acid homology with human retroviruses. The sequences present in the Eco R1-Sma1 fragment (nt. 1–444 Fig 1) are potentially capable to code for the amino terminal portion of a putative *gag* protein.

	A		B		C		D		E
HE.1	EAEWFTVLDLKDAFFCV	[20]	QLTWMVLPQGFRDSPHIFGQALAKDLG	[8]	VLQYVDDLLLA	[19]	NQGYKVSMSKAQ	[5]	VKYLGLILAKG
HERVC	EDSWFTCLDLKDAFFSI	[23]	QYTWTQLPQGFKNSPTIFGEALARDLQ	[10]	LLQYVDDLLLG	[19]	TVGIRCPRKKAQ	[5]	VCYLGFTIQQC
HERVK	KDWPLIIIDLKDAFFTI	[24]	RFQWKDLPQGMLNSPTICQTFVGRALQ	[10]	IIHYIDDILCA	[19]	NAGLAIASDKIQ	[4]	FHYLGMQIENR
HTLVII	ALPHLQTIDLTDAFFQI	[24]	RYAWTVLPQGFKNSPTLFEQQLAAVLN	[10]	IVQYNDDILLA	[19]	THGLPISQEKTQ	[5]	IRFLGQVISPN
HIV2	KKRRITVLDVGDAYFSI	[24]	RYIYKSLPQGWKGSPAIFQHTMRQVLE	[10]	IIQYMDDILIA	[20]	GLGFSTPDEKFQ	[4]	YHWMGYELWPT
HIV1	KKKSVTVLDVGDAYFSV	[24]	RYQYNVLPQGWKGSPAIFQSSMTKILE	[10]	IYQYMDDLYVG	[20]	RWGLTTPDKKHQ	[4]	FLWMGYELHPD
ATLV	TLAHLQTIDLRDAFFQI	[24]	RYAWKVLPQGFKNSPTLFEMQLAHILQ	[10]	ILQYMDDILLA	[19]	SHGLPVSENKTQ	[5]	IKFLGQIISPN

Figure 2. Alignments of conserved motifs within RNA-dependent polymerases and a large domain from the pHE.1 putative *pol* region HERVC, human endogenous retrovirus C (15); HERVK, human endogenous retrovirus K (11); HTLVII, human T-cell leukaemia type II, (30); HIV2, human immunodeficiency type 2 (31); HIV1, human immunodeficiency type 1 (32); ATLV, Human adult T-cell leukaemia (33). Invariant amino acids are showed by vertical bars.

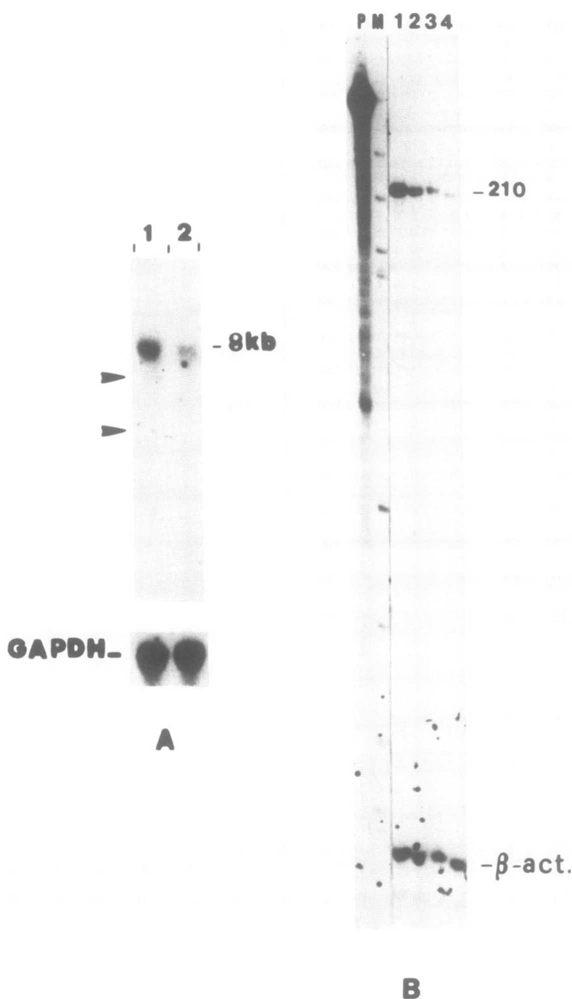

Figure 3. Analysis of the pHE.1-related transcript in NT2/D1 cells before and after RA-induced differentiation. Panel A) Northern blot analysis of pHE.1 transcript in NT2/D1 cells. 3 μg of poly A^+ RNA samples from undifferentiated NT2/D1 cells (lane 1) and from NT2/D1 cells grown for 7 days in the presence of retinoic acid (lane 2), were hybridized to a *pol* probe as described in the text. The filter was re-hybridized to a GAPDH probe for normalization. Arrows indicated the position of rRNAs. Panel B) RNase protection analysis of pHE.1 sequences using 20μg of total RNA from NT2/D1 cells, uninduced (lane 1), and at various times after RA induction: lane 2 (3 days), lane 3 (5 days), lane 4 (7 days). Expression of β-actin is shown as internal control. Antisense strand RNA probes were synthesized and used as described under Material and Methods section.

The primary translation product of the *gag* gene is a polyprotein eventually cleaved into virions core structural proteins (27). One of these is a nucleic acid binding protein (NBP) structurally characterized by the presence of one or multiple adjacent copies of the amino acid motif $CX_2CX_4HX_4C$ (28). A similar motif is present in the amino terminal part of the *gag* protein in pHE1 (nt. 223 to nt. 264). The bona-fide *gag-pol* junction in clone pHE.1 could be identified by the presence of the nearly identical deduced amino acids Asp-End-Gly-Leu-Gly in both pHE.1 and the reported human retroviral sequences HRVC (15).

The DNA sequences from nt.450 to nt.2715, (Fig.1) appear to code for a putative *pol* protein although termination codons have been found. In Fig 2 is reported the homology of a large domain of 210 amino acids between pHE.1 and human retroviruses *pol* proteins. The 210 amino acids domain is centred on the six motifs that have been recently reported to be conserved with the same linear arrangement in the *pol* proteins encoded by mammalian retroviruses (29), and it has been proposed that this domain is a prerequisite 'polymerase module' of a RNA dependent polymerase. The 210 amino acids long domain of the pHE.1 *pol* protein contains the appropriate amino acids residues to fit with those identified as invariant residues in viral reverse transcriptases (29) (Fig 2).

Between nt.2848 and nt.3320 (Fig.1) an open reading frame potentially coding for an *env* protein has been found, although two stop codons are present. The putative pHE.1 *env* protein displays an evident (45%) homology to the human retroviruses *env* proteins. The sequences from nt. 2715 to nt. 2848 did not show any significant homology to retroviral sequences.

Transcriptional analysis of the human endogenous retroviral sequences

Treatment of NT2/D1 cells with retinoic acid results in the loss of the undifferentiated phenotype, and the cells go through differentiation into specific cell types, including neurons. Transcription of the pHE.1 corresponding locus was studied by Northern blot analysis of Poly-A^+ RNA samples derived from NT2/D1 cells either untreated or seven days after RA exposure. The Northern blot was hybridized to the 410 bp Bam HI fragment located within the *pol* region (Fig.1). As shown in Fig. 3 (panel A), the *pol* probe clearly detected a 8 kb message in undifferentiated NT2/D1 cells, and RA treatment of NT2/D1 cells results in a lower expression. We studied the time course of pHE.1 expression during NT2/D1 differentiation, by RNAse protection assay performed on total RNA from cells at 0, 3, 5 and 7 days after exposure to retinoic acid. The RNAs were

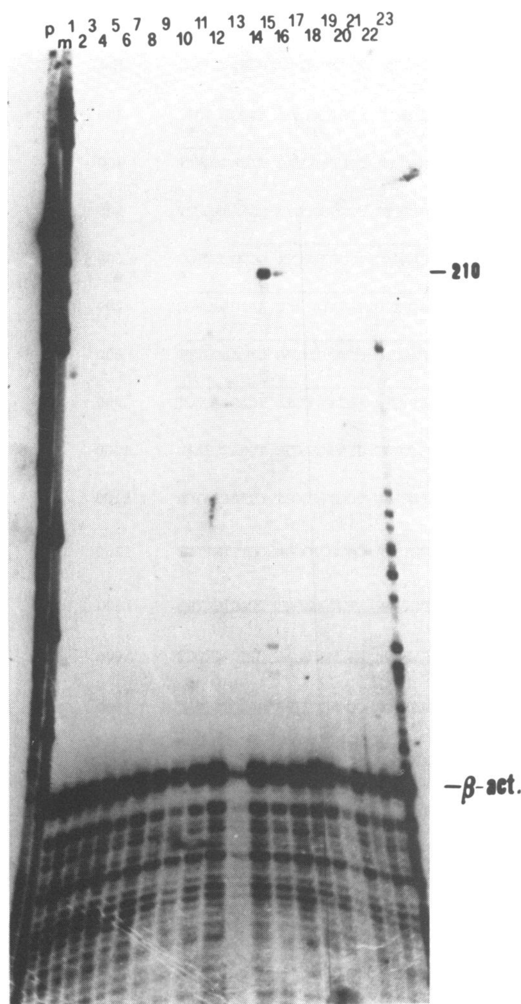

Figure 4. Expression of pHE.1 sequences in various human cell lines. 20 μg of total RNA were hybridized to both pHE.1 *pol* specific RNA probe and β-actin probe as described in the text.The RNA samples are: (1) Daudi, Burkitt lymphoma,; (2) CB33, lymphoblastoid cells, (3) HL60; (4) U937 cells; (5) KG1. AML cells; (6) FRO, T-ALL cells; (7) PEER, T-ALL cells; (8) MOLT-4, T-ALL cells; (9) 5637, bladder carcinoma; (10) MTA Ca-2, pancreatic carcinoma; (11) glioblastoma cells;((12) WI-38, embryonal lung cells; (14) NT2/D1 cells; (15) RA-induced (7 days) NT2/D1 cells; (13) Calu-1, lung carcinoma; (16) Hep3B, epatoma cells; (17) HepG2, epatoma cells; (18) Hela cells; (19) SK-N-MC, neuroblastoma; (20) HEL, erytroleukemia cells; (22) HTB, breast carcinoma; (23) placenta. The RNase protected fragments are detected in NT2/D1 cells (lanes 14 and 15). Expression of b-actin is shown as an internal control.

hybridized to a 210 bp *pol* specific probe together with a 3′-end β-actin probe as internal control (Fig. 3, panel B). The results clearly indicated that the pHE.1 transcript is negatively regulated during retinoic acid induced differentiation.

Examination of the pHE.1 transcript in various human cell lines of different embryological derivation by RNase protection experiments revealed the presence of ERV-9 RNA expression in NT2/D1 cells whereas the RA-treated NT2/D1 cells synthesize a relatively low, although detectable, level of ERV-9 RNA (Fig 4. lanes 14 and 15,respectively). Using RNA derived from placenta (Fig. 4 lane 23) we observed many protected fragments of different sizes all smaller than the expected 210 bp RNase protected fragment. Since it has been previously reported the expression of endogenous retroviral sequences in placenta

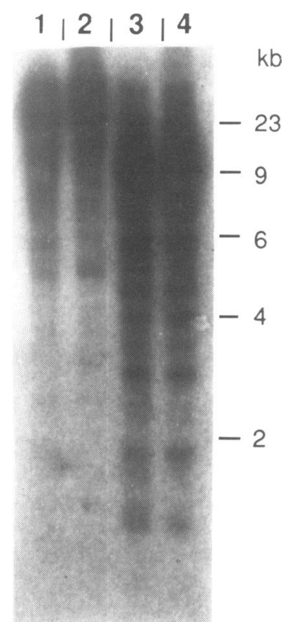

Figure 5. Southern blot hybdridization of ERV-9 sequences in human DNA. Human DNA was isolated from two individuals and digested with HindIII (lanes 1 and 2) and EcoR1 (lanes 3 and 4). The blot was hybridized to the *pol* specific probe (BamHI fragment from nt. 1190 to nt 1600 of pHE.1) as described in the text.

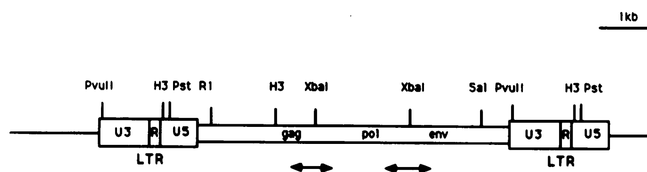

Figure 6. Restriction enzyme mapping of the proviral ERV-9 sequences in λFix1.1. The arrows correspond to sequenced regions of *gag*, *pol* and *env* regions.

(34,35), it is possible that the smaller protected RNase fragments are due to the expression of pHE.1 related retroviral sequences.

The experimental data reported in Figs. 3 and 4 suggest that the pHE.1 ERV-9 sequences are preferentially expressed in undifferentiated cells and their expression is negatively regulated upon RA induced cell differentiation.

Copy number and organization of ERV-9 sequences in the human genome

To analyse the organization and copy number of the ERV-9 sequences present in the human genome we performed Southern blot hybridization experiments. High molecular weight DNA from two individuals was digested with restriction enzymes and hybridized to a probe derived from the putative *pol* region (Fig.5). Under high stringency hybridization conditions 30 to 40 bands were detected, therefore we estimate the presence of about 35–40 different ERV-9 loci per haploid human genome. A comparable estimate of the ERV-9 copy number was obtained by dot blot and genomic library screening (data not shown). To better define the structure of ERV-9 sequences we have isolated genomic recombinant phages using as a probe a *pol* specific fragment isolated from pHE.1. In Fig. 6 is reported the restriction enzyme

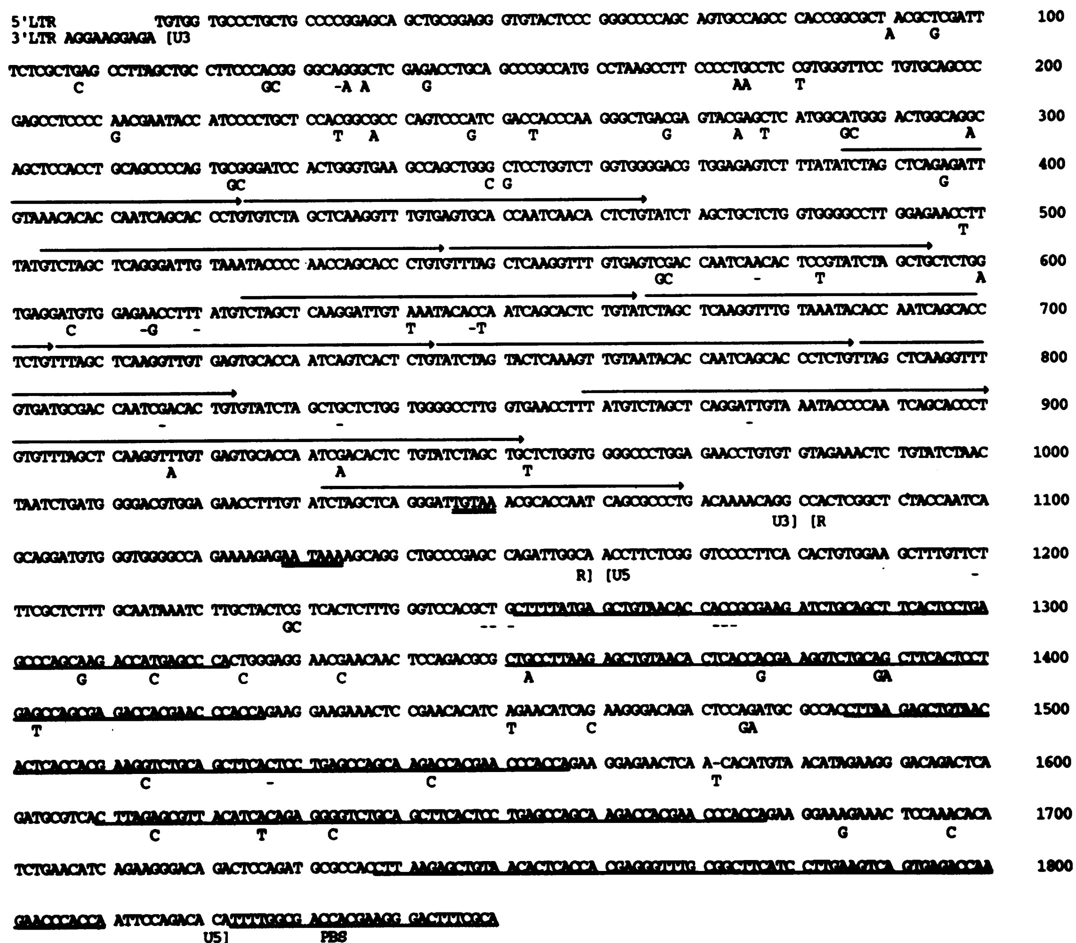

Figure 7. Nucleotide sequence of LTR's.The sequence begins with the polypurine tract which precedes the 3′ LTR and ends with a putative tRNA binding site just beyond the 5′ LTR. The 3′ LTR sequence is shown only when it differs from the 5′ LTR. Dashes indicate deletions. The 72 bp B elements are underlined; the 41 bp E elements are indicated by an arrowed overlining. Putative TATA and polyadenylation signal sequences are double underlined.

mapping of one phage, λFix1.1. The regions hybridizing to *gag*, *pol* and *env* probes are flanked by 1.8 kb long direct repeats whose sequences are reported in Fig.7. 5′ and 3′ 1.8 kb repeated sequences show more than 90 percent base matching and they display several characteristic features of mammalian LTR sequences (reviewed in 27, 36, 37). These include a 4-base inverted repeat bordering the two LTRs (TGTG and CACA); a putative TATA box (TGTAA) at nt. 1046 within the U3 region followed by the transcriptional start site, usually GC (at nt. 1080), which defines the beginning of the R region. The R region contains the polyadenylation signal at nt. 1129 followed, 24bp downstream, by the polyadenylation site CA. The remaining LTR sequences constitute the U5 region of about 670 bp. Finally a polypurine tract precedes the 3′ LTR and a putative tRNA primer binding site follows the 5′ LTR. The PBS sequence present in ERV-9 is highly homologous to the corresponding sequences in ERV-3. (38). The 1.8kb long LTRs sequences reveal a complex structure composed by tandemly repeated sub-elements. The E element is 41bp long and it is repeated 12 times within the U3 region, whereas the B element is 72bp long and it is repeated 4 times in the U5 region. Similar elements have been found in the 3′ and 5′ termini of cDNAs previously described (10), and in the 3′-end of pHE.1 cDNA. Moreover, preliminary analysis of additional genomic clones indicated that ERV-9 LTR sequences are heterogeneous in length, and the length variation is due to the number of tandemly repeated E and B sub-elements (La Mantia et al., manuscript in preparation and ref 10).

ERV-9 LTR promoter activity

To examine the ERV-9 LTRs functional capacities we have made use of the system in which the bacterial CAT gene can be placed under the transcriptional control of heterologous promoter sequences, and the synthesis of bacterial enzyme assayed in transfected cells (25). As the viral transcriptional regulatory elements are located within the U3 region, and the transcription start site defines the beginning of R region, fragments containing the U3 and part of the R element from both the genomic 5′ LTR present in clone λFix1.1 and the 3′-end portion of pHE.1, were cloned in p8-CAT-O vector. The construction of the CAT plasmids used in this analysis is described in the Materials and Methods section. To determine the range and relative promoter activity the plasmid constructs were transfected on HeLa, WI-38 and NT2/D1 cells. The vectors p8-CAT-0 and pRSV-CAT were used as negative and positive controls, respectively. A typical CAT assay is shown in Fig 8. The results clearly indicate that both genomic and cDNA derived LTRs fragments are capable to direct CAT expression only in NT2/D1 cells, although the individual promoter strength is different. The promoter activity

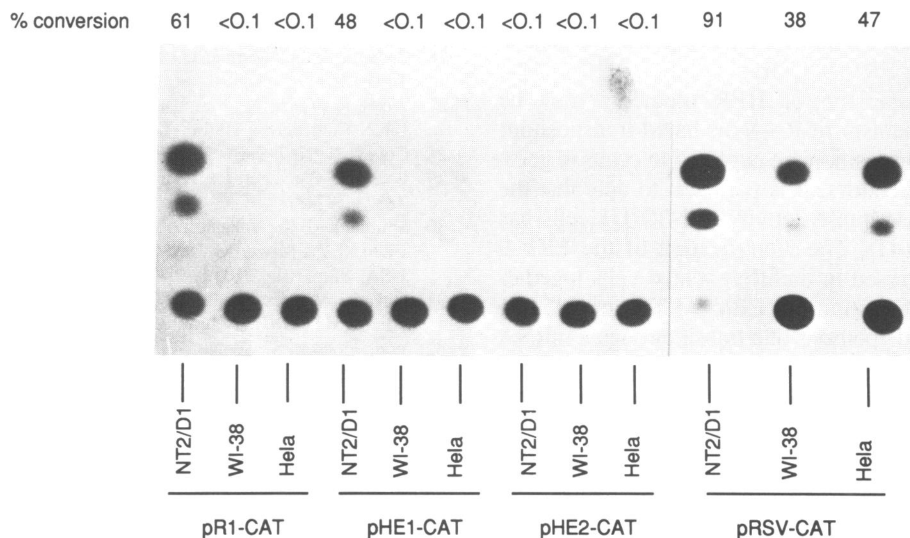

Figure 8. Assay of ERV-9 LTR promoter activity in different human cell lines. The pHE.1-CAT clone contains the 3′ end portion of pHE.1 cDNA clone in the 5′-3′ orientation, the pR1-CAT contains the 5′ LTR sequences derived from the genomic lFix 1.1 clone. Cell extracts were prepared 48 h post transfection with the indicated plasmid and assayed as described (25). 200μg of cell extracts were used for the transfection with pHEl-CAT, pHE2-CAT and pR1-CAT, conversely 30μg of extracts were used for pRSV-CAT transfected cells. The CAT activity was measured by conversion of ^{14}C chloramphenicol to its acetylated forms.

is orientation (5′-3′) dependent, as the pHE2-CAT is ineffective in all cell lines tested. These data suggest that the tissue specific ERV-9 transcription might results from the activity of the ERV-9 LTR promoter rather than being a consequence of fortuitous integrations of an ERV-9 element in a position adjacent to an embryonic specific gene.

Inspection of the LTR regulatory region does not reveal the presence of consensus sequences for any known binding site of embryonic transcription factors, such as Oct3 (5,6).

DISCUSSION

The characterization of the cDNA clone pHE.1 by comparative sequence analysis indicated that the pHE.1 clone contains sequences potentially capable to code for putative *gag*, *pol* and *env* proteins, thus showing that the cloned sequences correspond to a genomic locus of a novel human endogenous retrovirus that we named ERV-9. Moreover the ERV-9 sequences present in pHE.1 clone are associated to a repetitive element located at the 3′-terminal portion of the clone (10). Analysis of a genomic clone indicated that ERV-9 sequences are flanked by repetitive elements at both ends, thus resembling a LTR-like structure. To further define the genomic structure of ERV-9 sequences we are currently analyzing the structure of several genomic loci that have been isolated by hybridization to a *pol* specific probe (manuscript in preparation). Southern and dot blots experiments indicate that there are about 35–40 different ERV-9 loci per haploid human genome, therefore the ERV-9 sequences present in the pHE.1 clone are members of a repeated family, which is conserved in primates but not in rodents (10, and data not shown). We cannot exclude that other members of this polymorphic family may encode for functional products.

The ERV-9 LTRs sequences have a complex structure composed by tandemly repeated sub-elements. A 41 bp long repeated element (E element) is located within the U3 region, whereas a 72 bp long repeat is present in the U5 region. While structurally similar to known LTRs, the ERV-9 LTRs sequences do not resemble other known LTR sequences, such as the LTRs present in the previously described HERs (11–19). Analysis of additional genomic and cDNA clones indicated that ERV-9 LTRs sequences are heterogeneous in length, and the length variation is due to the number of tandemly repeated E and B sub-elements (La Mantia et al., manuscript in preparation and ref 10).

In this paper we have showed that ERV-9 sequences are preferentially expressed in undifferentiated NT2/D1 cells, as no detectable expression of ERV-9 sequences was found in various cell lines of different embryological derivation, and the ERV-9 expression is negatively regulated during in vitro differentiation of NT2/D1 cells induced with retinoic acid. In addition we have shown that ERV-9 LTR sequences are capable to regulate expression of linked CAT gene only in NT2/D1 cells, indicating that the cell-specific expression of ERV-9 sequences might be determined autonomously by LTR ERV-9 regulatory elements.

Human cells contain a complex variety of endogenous retroviral sequences (HER), that have been isolated by hybridization to known retroviral sequences, accidentally found in flanking regions of other genes, or by hybridization to the 3′-end terminus of tRNA (11–19). To date the function of the HERs is unknown and conclusive evidences for a direct role of HERs in human diseases have been lacking. Nevertheless, the HERs do not appear to be silent components of the human genome since some of them are transcriptionally active. Polyadenylated HER RNAs have been detected in human placenta, breast carcinoma, and colon carcinoma cells (34,35). Moreover the use of cross-reactive antibodies indicates that HER sequences may be expressed at the protein level (39, 40, 41). Recently, testicular teratocarcinoma cell lines have been shown to release C-type retroviral particles (42). Rearrangement of members of endogenous retrovirus-like elements in mice have been implicated in both activation and inactivation of gene expression (20–23). The myelomonocytic murine leukemia cell line WEHI-3B consitutively produces the growth factor Interleukin-3 due to an IAP insertion 5′ of the gene (22), and transcription has been shown to originate within the LTR. To verify the possibility that rearrangements or activation

of ERV-9 sequences may contribute to alterations in gene expression, we are currently screening human genome for genetic rearrangement involving ERV-9 LTRs.

Amplification and dispersion of HER members may be obtained through a mechanism of RNA-mediated transpositon, and amplification and transposition are expected to occur in germ line cells or in their progenitors. It is pertinent to note that the presence of reverse transcriptase activity in NT2/D1 cells has recently been reported (43). The identification of the ERV-9 mRNA specifically expressed in undifferentiated cells together with the specific promoter activity of ERV-9 LTRs in NT2/D1 cells may substantiate the hypothesis of a transit through a mRNA intermediate.

ACKNOWLEDGEMENTS

We would like to thank R. Terracciano for skilful technical help, and E. Boncinelli and P.P. Di Nocera for helpful discussions. This work was paid for by grants from the Italian Association for Cancer Research (AIRC) to L.L., MPI (40%) to G.L.M. and from grant n° 5206072 from Progetto AIDS. G.P. is recipient of an AIRC fellowship.

REFERENCES

1. Andrews, P.W. (1984) Dev. Biol.,103, 285−293.
2. Luscher, B., Mitchell, P. J., Williams, T. and Tjian, R. (1989) Genes Dev., 3, 1507−1517.
3. Mavilio, F., Simeone, A., Boncinelli, E., and Andrews, P.W. (1988) Differentiation, 37, 73−79.
4. Simeone, A., Acampora, D., Arcioni, L., Anrews, P.W., Boncinelli, E.,and Mavilio F. (1990) Nature, 346, 763−766.
5. Okamoto, K., Hiroshi, O., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H (1990) Cell 60, 461−472.
6. Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier,F., Rigby, P.W., and Staudt, L.M. (1990) Nature 345, 686−692.
7. Curatola, A.M., and Basilico C. (1990) Mol.Cell Biol,10, 2475−2484.
8. Skowronski, J., and Singer, M.F. (1985) Proc. Natl. Acad. Sci. USA, 82, 6050−6054.
9. White, R.J., Scott, D, and Rigby, P.W. (1989) Cell, 59, 1081−1092.
10. La Mantia, G., Pengue, G., Maglione, D., Pannuti, A., Pascucci, A., and Lania, L. (1989) Nucleic Acids Res., 17, 5913−5922
11. Callahan, R., Chiu, I.-M., Wong,J.F.H., Tronick,S.R., Roe, B.A., Aaronson, S.A., and Schlom, J. (1985) Science, 228, 1208−1211.
12. Ono, M., Yasunaga, T., Miyata, T., and Ushikubo, H. (1986) J.Virol., 60, 589−598.
13. Bonner, T.I., O'Connel, C., and Cohen, M. (1982) Proc. Natl. Acad. Sci. USA, 79, 4709−4713.
14. Martin, M. A., Bryan, T., Rasheed, S., and Khan, A. S. (1981). Proc. Natl. Acad. Sci. USA,78,4892−4896.
15. Repaske, R., Steele, P. E., O'Neill, R. R., Rabson, A. B., and Martin, M. A. (1985). J.Virol., 54, 764−772.
16. Mager, D. L., and Hentorn, P. S. (1984). Proc. Natl. Acad. Sci. USA, 81, 7510−7514.
17. Kroger,G., and Horak,I. (1987). J.Virol., 61, 2071−2075.
18. Perl, A., Rosenblatt, J. D., Chen, I.S.I., Di Vincenzo, J. P., Bever, R., Poiesz, J., and Abraham, G. N. (1989). Nucleic Acids Res., 17, 6841−6854.
19. Leib-Mosch, C., Brack, R., Werner, T., Erfle, V., and Hehlmann, R. (1986). Virology, 155, 666−677.
20. Hawley, R.G., Shulman, M.J., Murialdo, H., Gibson, D.M., and Hozumi, N. (1982) Proc. Natl. Acad. Sci. USA 79, 7425−7429.
21. Min Man, Y., Delius, H., and Leader, D.P. (1987) Nucleic Acids Res. 15, 3291−3304.
22. Ymer, S., Tucker, W.Q.J., Sanderson, C.J., Hapel,A.J., CXampbell, H.D., and Young, I.G. (1985) Nature 317, 225−258.
23. Kongsuwan, K., Allen, J., and Adams, J.M. (1989) Nucleic Acids Res. 17, 1881−1892.
24. Ponte, P., Gunning, P., Blau, H., and Kedes,L. (1983). Mol. Cell Biol. 3, 1783−1791.
25. Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982) Mol. Cell. Biol. 2, 1044−1051.
26. Pannuti, A., La Mantia, G., and Lania, L. (1987) Nucleic Acids Res. 15, 1595−1613.
27. Weiss, R., Teich, N., Varmus, H., and Coffin, J. (1985). RNA tumor viruses. Ed.2. Cold Spring Harbor Laboratory Press, New York.
28. Covey, S. N. (1986). Nucleic Acids Res., 14, 623−633.
29. Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989). EMBO J., 8, 3867−3874.
30. Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Chen, S.Y., Miwa, M., and Sugimura, T. (1985). Proc. Natl. Acad. Sci. USA, 82, 3101−3105.
31. Guyader,M., Emerman,M., Sonigo,P., Clavel,F., Montagnier,L., and Alizon,M. (1987). Nature, 326, 662−669.
32. Wain-Hobson, S., Sonigo, P., Danaos, O., Cole, S., and Alizon, M. (1985). Cell, 40, 9−17.
33. Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). Proc. Natl. Acad. Sci. USA, 80, 3618−3622.
34. Rabson, A. B., Hamagishi, Y., Steele, P. E., Tykocinski, M., and Martin, M.A. (1985). J.Virol., 56, 176−182.
35. Kato,N., Pfeifer-Ohlsson,S., Kato,M., Larsson,E., Rydnert,J., Ohlsson,R., and Cohen,M. J. (1987). J. Virol., 61, 2182.2191.
36. Varmus, H.E. (1982) Science 216, 812−820
37. Chen, H.R., and Barker, W.C. (1984) Nucleic Acids Res. 12, 1767−1778.
38. O'Connel, C. D., and Cohen, M. Science 1984) 226, 1204−1206
39. Jerabek,L.B., Mellors,R.C., Elkon,K.B., and Mellors,J.W. (1984). Proc. Nat. Acad. Sci. USA, 81, 6501−6505.
40. Suni, J., Narvanen, A., Wahlstrom, T., Aho, M., Pakkanen, R., Vaheri, A., Copeland, T., Cohen, M., and Oroszlan, S. (1984). Proc. Acad. Natl. Sci. USA, 81, 6197−6201.
41. Derks,J.P.A., Hofmans,L., Bruning,H.W., and Van Rood,J.J. (1982). Cancer Res. 42, 681−686.
42. Lower, J., Wondrak, E. M., and Kurth, R. (1987). J. Gen. Virol. 68, 2807−2815.
43. Deragon, J-M., Sinnett, D., and Labuda, D. (1990). EMBO J. 9, 3363−3368.