Germ Cell Expression of an Isolated Human Endogenous Retroviral Long Terminal Repeat of the HERV-K/HTDV Family in Transgenic Mice

ARMELLE E. CASAU,1 JOE E. VAUGHAN,2 GUILLERMINA LOZANO,2 AND ARNOLD J. LEVINE3*

Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544; Department of Molecular Genetics, University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030; and The Rockefeller University, New York, New York 10021-6399

Received 3 June 1999/Accepted 8 September 1999

Endogenous retroviruses are present in multiple copies dispersed throughout the genomes of host species (9, 47). These retroelements have been vertically transmitted through the germ line as Mendelian genes for millions of years and may represent the endogenous counterparts of exogenous retroviruses that infected ancestral species long ago. It is estimated that up to 10% of the human genome is composed of sequences that were reverse transcribed and that up to 1% of the genome is made up of endogenous retroviruses that are very similar to exogenous retroviruses both in sequence and in structure (3, 5, 47). Both the prevalence and maintenance of these elements suggest that they may play a role in the biology of the host species.

Endogenous retroelements may affect their hosts via multiple mechanisms (46, 47): (i) they can act in trans by expressing viral gene products that could interfere with or modulate cellular activities (20); (ii) they can act in cis by activating neighboring cellular genes with their regulatory sequences (11) or by retrotransposing and leading to insertional mutagenesis (31, 32); (iii) they can promote genomic plasticity as well as contribute to allelic variations (42, 48); (iv) and they have been shown, in some cases, to potentially play a role in autoimmune diseases such as glomerulonephritis (18, 37, 46).

One human endogenous retrovirus of interest is HERV-K (human endogenous retrovirus K), where the K denotes the lysine tRNA used by this element as a primer for reverse transcriptase (35). An HERV-K family member, HERV-K10, was first isolated by using a Syrian hamster intracisternal A particle (IAP) pol probe and was shown to have some homology to the mouse mammary tumor virus (MMTV) env gene (34). HERV-K elements entered the primate lineage after the divergence of New World monkeys and Old World monkeys more than 30 million years ago and underwent amplifications as well genomic rearrangements (40). Human-specific integration events have been reported and indicate relatively recent HERV-K activity (30).

HERV-K is present in about 30 to 50 copies in the human genome (34), along with an estimated 10,000 solitary long terminal repeats (LTRs) (21). Unlike many other HERV family members, some of the HERV-K proviruses have retained open reading frames for their viral genes (13, 17, 25, 27, 28, 33, 44). Although no one provirus has been demonstrated to produce the HERV-K viral particles, an almost intact provirus with all open reading frames has been found on chromosome 7 (29). Noninfectious particles derived from as of yet unidentified HERV-K proviruses have been observed in testicular teratocarcinoma cell lines and tumors, as well as in the placenta, but not in the vast majority of other tumor types or cell lines (4, 8, 24, 39).

Testicular teratocarcinomas are germ cell tumors (GCTs) that are composed of an undifferentiated embryonal carcinoma (EC) stem cell population and a variety of differentiated cell populations (1, 2). The human teratocarcinoma-derived viruses (HTDVs), which are expressed in these tumors, are arrested at late budding stages and lack the electron-lucent space between the envelope and the core shell indicative of mature particles (4). These particles seem to no longer be observed in the differentiated derivatives of GCTs such as differentiated embryonal carcinoma cells, cultures of yolk sac carcinomas,
and teratomas (8), although it has also been suggested that HTDVs are actually budding from the trophoblastic component of differentiating teratocarcinoma cell cultures (22, 24).

High levels of expression of HERV-K members appear to be restricted to GCTs (including testicular teratocarcinoma cell lines) and their testicular precursor lesions (14, 15, 23). Mature and immature teratomas and spermatocytic seminomas with no embryonal carcinoma cell component show no HERV-K expression (15). Furthermore, many other tissues, tumor types, and cell lines do not demonstrate detectable levels of HERV-K expression (15, 23). Occasionally, a high level of expression has been found in chronic myeloid leukemias (6), leukocytes (7), placenta (38), peripheral blood mononuclear cells, and brain tissues of healthy persons and multiple sclerosis patients (36).

The significance of these sporadic and pecuilar expression patterns remains elusive.

In this study, an active HERV-K LTR has been isolated and used to study the expression patterns of the viral promoter-enhancer(s) in order to determine when the transcriptional regulatory elements of a typical HERV-K virus direct gene expression. The isolated HERV-K LTR was demonstrated to drive high-level expression of a reporter gene in human and murine teratocarcinoma cell lines but not in the differentiated counterparts or in various other tumor cell lines tested. To study the expression pattern of this viral LTR in vivo, transgenic mice harboring similarly active LTRs might be developmentally regulated, being active primarily in adult gonocytes and in undifferentiated GCTs, but not in their differentiated counterparts nor in a variety of other tissue and tumor types.

MATERIALS AND METHODS

Sequences. The accession number for the 5′ LTR of HERV-K10 is M12851. The accession number for HERV-K LTR3 is AF148679. The BLAST program (28) was used to find homologous sequences.

PCR kit with oligo(dT) primers as specified by the vendor (Perkin Elmer). RT reactions were performed by standard methods with 10 to 20 ng of total RNA or 10 μg of poly(A) mRNA. Formaldehyde-agarse gels (1.2% formaldehyde) were electrophoresed overnight, stained with ethidium bromide to examine 18S/28S hybridization to the same solution once at room temperature. The slides were washed three times with 1× SSC–0.1% SDS for 30 min.

Northern blots. Total RNA was extracted from frozen mouse testis, brain, and teratomas (8), although it has also been suggested that HTDVs are actually budding from the trophoblastic component of differentiating teratocarcinoma cell cultures (22, 24).

High levels of expression of HERV-K members appear to be restricted to GCTs (including testicular teratocarcinoma cell lines) and their testicular precursor lesions (14, 15, 23). Mature and immature teratomas and spermatocytic seminomas with no embryonal carcinoma cell component show no HERV-K expression (15). Furthermore, many other tissues, tumor types, and cell lines do not demonstrate detectable levels of HERV-K expression (15, 23). Occasionally, a high level of expression has been found in chronic myeloid leukemias (6), leukocytes (7), placenta (38), peripheral blood mononuclear cells, and brain tissues of healthy persons and multiple sclerosis patients (36).

The significance of these sporadic and pecuilar expression patterns remains elusive.

In this study, an active HERV-K LTR has been isolated and used to study the expression patterns of the viral promoter-enhancer(s) in order to determine when the transcriptional regulatory elements of a typical HERV-K virus direct gene expression. The isolated HERV-K LTR was demonstrated to drive high-level expression of a reporter gene in human and murine teratocarcinoma cell lines but not in the differentiated counterparts or in various other tumor cell lines tested. To study the expression pattern of this viral LTR in vivo, transgenic mice harboring similarly active LTRs might be developmentally regulated, being active primarily in adult gonocytes and in undifferentiated GCTs, but not in their differentiated counterparts nor in a variety of other tissue and tumor types.
<table>
<thead>
<tr>
<th>KE3</th>
<th>1</th>
<th>tggtaactggtcgtgttaga agacaa-tagcatacaggagac tccatggtttgtattat aaaaatctctgcttga</th>
<th>K10</th>
<th>78</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>gattcactttcctgccctgcagctc tgggaaatagtcgcttcgtaac gctgaagaggtgat</td>
<td>400</td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>aaggctggcagtacgtgc tttggtaaacctatgtctgtaa ggaacagcatcctccactaatc</td>
<td>560</td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>tcaagtaccacggaggacaac aaagtgcagagggcgagggac gaccctgtctaggaagggga aggtatatcgcaagtttttc</td>
<td>640</td>
<td></td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>ccacctagatgattaagaa ttaggcctgtcgagagaaagggc aagaacctctgccttccc</td>
<td>720</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>tgtgtcagggaggattga gaagagggtsagatctct tgcaggtacacagagggaa gcctgtctctcctcggt</td>
<td>800</td>
<td></td>
<td></td>
</tr>
<tr>
<td>480</td>
<td>cctggagacattggatct ctgTATAAAAacccgtttagttgtc tgcctcactcactgatag</td>
<td>880</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 560 | ggagggagacccctgggca gcaattactctttttaagggc atttagagtttttgattatatgc acttatccaaagggcagcc | |}

FIG. 1. Nucleotide sequence homology between HERV-K LTRE3 (KE3: nt 1 to 921) and HERV-K10′ LTR (K10: nt 78 to 999). Identities are depicted as dots. The TATA box and polyadenylation signal are capitalized. The potential glucocorticoid/enhancer site is underlined.

Since the LTR PCR assay would probably favor selection of the 10,000 solitary LTRs rather than the 30 to 50 full-length proviruses, a human genomic library was screened with an HERV-K gag probe to isolate LTRs associated with 1 of the 50 or so full-length HERV-K proviruses in the genome. Of the 10 isolated lambda clones, 6 yielded PCR products with LTR specific primers. These LTRs were cloned into the luciferase reporter vector used previously and transfected in the various cell lines described above. Two of the six LTRs behaved similarly to the previously isolated LTRE3 (60 to 72% of LTRE3 activation), whereas the other four seemed to be relatively inactive in the in vitro assay described above (4 to 17% of LTRE3 activation) (data not shown).

HERV-K LTR expression is restricted to teratocarcinoma cell lines. It has been previously observed that HERV-K expression (as determined by mRNA and protein expression as well as particle formation) is highest in testicular teratocarcinomas as well as in the cell lines derived from these tumors and is either undetectable or present at lower levels in a number of normal tissues and tumor types. Results with the isolated LTRE3 corroborate this conclusion. The enhancer(s) in LTRE3 increased luciferase expression relative to promoter alone from 10- to 70-fold in three human GCT cell lines (Tera 1, Tera 2, and 1218E) and in two murine GCT cell lines (F9 and P19) (Fig. 2A), but it did not enhance luciferase expression by more than 2-fold in several non-GCT human cell lines including a transformed embryonic kidney cell line, one osteogenic sarcoma, a lung adenocarcinoma, three breast carcinomas, and a colon adenocarcinoma (293, SaOs2, H1299, MDA231, MDA435, T47D, and SW480, respectively) (Fig. 2B).

RA differentiation of a teratocarcinoma cell line leads to a decrease in LTR expression. Previous observations (8) have suggested the possibility that HERV-K LTR expression is dependent on the state of differentiation of GCTs. EC cells are the undifferentiated, rapidly dividing stem cell population of teratocarcinomas. The murine EC cell line F9 has been successfully used as an in vitro model to study the regulation of differentiation. These EC cells do not differentiate spontaneously but can be induced to do so upon treatment with the acidic form of vitamin A, retinoic acid (RA) (41). These EC cells can differentiate into primitive endoderm cells (when incubated for 30 min at room temperature. The sections were again washed three times for 5 min with PBSBT. Mounting solution (30% glycerol, 70% PBS) was used to mount the slides with coverslips.

**RESULTS**

**Isolation of active HERV-K LTRs.** To isolate HERV-K LTRs from human genomic DNA, primers were designed from the published HERV-K10 LTR sequences (34). LTR PCR products from a teratocarcinoma cell line (Tera 1) were cloned upstream of a luciferase reporter gene vector (PGL2-Promot-er), which includes its own SV40 early promoter, to identify LTRs harboring functional enhancers (data not shown). It had been previously shown (by investigating endogenous viral transcripts by Northern and RT-PCR analyses) that HERV-K family members can transcribe their viral genes as well as some adjacent cellular genes to high levels in testicular teratocarcinoma cell lines (GCT cell lines) but only to lower or non-detectable levels in most other tumors and cell lines tested to date (23, 25, 26, 45). However, no detailed studies investigating the expression patterns with isolated active LTRs have been done.

To assess the general activity of the isolated LTRs, the cloned LTRs, driving the expression of the luciferase reporter gene were transfected into GCT cell lines, Tera and Tera 2, and into non-GCT cell lines, 293 (transformed embryonic kidney cell line) and H1299 (lung adenocarcinoma). Of 33 different LTR clones isolated by PCR selection, 18 were inactive in both GCT and non-GCT cell lines while 13 clones enhanced luciferase levels only a few-fold over promoter alone in GCT cell lines (data not shown). Two LTR clones were found to be highly active in GCT cell lines and not in non-GCT cell lines. The most active clone (LTRE3), with 10- to 90-fold activation in various GCT cell lines, was chosen for further study. The sequence of LTRE3 compared to that of HERV-K10′ LTR5′ is shown in Fig. 1. Both LTRs have a consensus TATA box and a polyadenylation signal, as well as a putative glucocorticoid/enhancer site. LTRE3 was shown to respond to the synthetic glucocorticoid hormone dexamethasone in GCT cells (fourfold activation upon addition of 10 mM dexamethasone) (data not shown).
treated with RA), into visceral endoderm (upon aggregation and treatment with RA), and into parietal endoderm (with RA and dibutyryl cyclic AMP) (41).

This in vitro system was used to determine whether HERV-K expression is affected by the degree of differentiation in a GCT cell line. LTRE3 was cloned upstream of a bacterial lacZ expression plasmid which is devoid of any promoter sequences to test whether the LTR was able to act as its own promoter, using its intact TATA box and other regulatory elements (Fig. 1). F9 cells were induced to differentiate by adding 10^{-6} M RA to the medium for 1 week. Cells treated in this manner exhibited the characteristic morphological changes associated with differentiation (see Fig. 3B, panel 4). F9 cells and RA-differentiated F9 cells were transiently transfected either with a β-gal expression vector devoid of a promoter (lacZ vector) or one driven by LTRE3 (LTRE3-lacZ vector). A constitutively expressed luciferase vector (PGL2-control vector) was also cotransfected to determine the efficiency of transfection in each cell population and normalize the results. LTRE3-driven lacZ expression in undifferentiated F9 cells (LTRE3-lacZ vector) was 30-fold higher than that of the basal levels of lacZ activity (lacZ vector) (Fig. 3A). However, LTRE3-driven lacZ expression in RA-differentiated F9 cells (LTRE3-lacZ vector) was reduced to fivefold compared to expression of the lacZ gene without the LTRE3 (Fig. 3A). Furthermore, LTRE3 expression was very low (less than two-fold higher than basal levels) in the fully differentiated murine embryonic fibroblast cell line BALB/c 3T3 (data not shown).

Additionally, X-Gal staining of pooled F9 clones stably transfected with LTRE3-lacZ and a neo vector (used to select for stably transfected cells), before and after RA differentiation, revealed a strong staining before differentiation and a lack of staining after differentiation (Fig. 3B). These observations demonstrate that HERV-K LTRE3 expression is sensitive to the degree of cell differentiation and hence may be developmentally regulated.

**Generation of transgenic mice harboring the LTRE3-lacZ sequence.** Since LTRE3 was shown to be active in murine cells (Fig. 3), it was postulated that this human LTR could be tested for tissue specificity in its ability to drive transcription in a transgenic mouse system. The same construct used in the RA differentiation experiment in Fig. 3 was used to create transgenic mice harboring the human HERV-K LTRE3 driving the expression of the bacterial lacZ gene. Four mice were shown to be transgenic for the human endogenous viral LTR and the bacterial lacZ gene by both Southern blot and PCR analyses of tail DNA (data not shown). The four independent transgenic lines (Tg2, Tg4, Tg7, and Tg9) were then crossed to 129Sv mice to generate mouse lines and homozygotes of all four independent transgenic lines.

**Transgene expression is restricted mainly to the more undifferentiated spermatocytes of adult testes.** Total RNA was extracted from mouse tissues (brain, liver, lung, thymus, heart, intestines, testis, ovaries, uterus, skeletal muscle, kidney, and spleen) and used for RT-PCR as well as Northern blot analysis. Of the four transgenic lines obtained, two (Tg7 and Tg9) were shown to express the lacZ gene in a few tissues and to various degrees. Tg7 and Tg9 were both found to express lacZ in adult testes when analyzed by RT-PCR (Fig. 4A). The control experiment with no reverse transcriptase enzyme (--RT) showed no bands even when the RT-PCR gel was assayed by Southern blotting and long exposures were taken. This demonstrates that there is no DNA contamination in the RT-PCR reactions. The ROSAβ-geo26 strain (R+) was used as a positive control since it constitutively expresses lacZ in most of its organs and tissues. The 129Sv strain (S−) was used as a negative control since it does not contain any lacZ sequences. Northern blot analysis of total RNA from the two expressing transgenic lines revealed that Tg7 expresses discrete sized transcripts in adult testes while Tg9 does not (expression was detectable only by RT-PCR analysis), reflecting differences in the levels of expression between the two independently derived transgenic lines. The lacZ transgenic line ROSAβ-geo26 and the nontransgenic line 129Sv were used to control for the specificity of the lacZ probe (Fig. 4A and C). mRNA was also isolated from both
transgenic lines but, as with Northern blot analysis with total RNA, no detectable expression was seen for Tg9 (Fig. 4D).

RT-PCR analysis of other mouse tissues revealed that lacZ expression driven by the HERV-K LTR is indeed limited primarily to the testes (Fig. 5). RT-PCR analysis was carried out in multiple experiments, each done in triplicate to minimize the nonquantitative nature of PCR amplifications. In addition to the positive signals in the testes, only brain tissues generated positive signals by RT-PCR (consistently high in Tg9 and lower in Tg7). Tissues such as kidney, liver, and thymus at times produced faint bands, but these RT-PCR signals were not reproducibly consistent. The other tissues did not generate reproducible signals by RT-PCR (Fig. 5) and did not generate signals by Northern blot analysis (data not shown).

To determine the cell types in the seminiferous tubules of adult testes that are responsible for HERV-K LTR expression, frozen testes sections were used for immunofluorescence studies with a β-gal polyclonal antibody specific for the bacterial LacZ protein product. The strongest staining was observed in the testes of the positive control ROSAβ-geo26 mice (ROSA129SvJ) (Fig. 6A). This staining was restricted primarily to the more differentiated cell types of the seminiferous tubules facing the interior of the lumen. As negative controls, testes sections from the ROSA129SvJ line were also stained with primary antibody only (anti-β-gal) (Fig. 6B) and secondary antibody only (Texas Red) (Fig. 6C). Compared to the nontransgenic 129Sv sections (Fig. 6D to F), in which faint background staining was visible, sections of testes from Tg7 mice (Fig. 6G to I)
exhibited a more intense staining, which appeared to be restricted primarily to the more undifferentiated spermatocytes in the seminiferous tubules.

DISCUSSION

This family of viruses is thought to be the most active HERV group in the human genome due to its high levels of expression in some tumor cell lines and its ability to code for functional proteins that can give rise to particles, albeit immature noninfectious ones. These proviruses have been maintained in the human genome for more than 30 million years. Although these proviruses may be regarded as vestiges of the past, they remain active, with the potential to impact differentiated or pathological states. It remains possible that the expression of at least one of the HERV-K viral genes or some cellular genes driven by the retroviral LTRs has been selected for and that this is the reason why some have not been mutated out of existence.

It was previously observed that HERV-K proviruses potentially express their viral genes in GCTs such as testicular teratocarcinomas. HERV-K expression has also been found in some other tissue and tumor types, but it has been sporadic and has been present at much lower levels. No thorough studies had been undertaken to determine the expression patterns dictated by isolated active HERV-K LTRs. To address how this family of virus regulates the expression of its viral genes or neighboring cellular genes, it was necessary to isolate and characterize an active LTR.

The results in this paper demonstrate that an isolated active HERV-K LTR confers the same types of expression patterns that had been observed for the viral genes. An active HERV-K LTR3 was able to drive the expression of reporter genes in testicular teratocarcinoma cell lines but not in a variety of other cell lines including a transformed kidney cell line, breast carcinomas, an osteosarcoma, and lung and colon adenocarcinomas. Furthermore, this LTR was downregulated upon differentiation of a teratocarcinoma cell line, indicating that it might be developmentally regulated. A transgenic mouse model was used to determine that expression in the adult tissues was limited to the testes and, to a lesser extent, the brain. Although the expression of some HERV-K elements was previously noted in the placenta (39), no expression was found in the placenta of near-term LTR3 transgenic pups by RT-PCR (data not shown). The strong possibility remains that there are some HERV-K LTRs which generate other expression patterns (such as placenta expression or differentiated trophoderm expression) due to sequence variations that are still uncharacterized.

All expression patterns were noted in two independent transgenic lines and were stable over all generations tested. A few other tissues did, in some RT-PCR experiments, generate faint signals, but the irregularity of these signals and their low intensities suggest that they might be considered low basal levels of expression. In contrast, RT-PCR-detected expression in testis was very reproducible and always generated strong signals in Tg7, and the expression was also seen by Northern blot analysis. Tg9 expressed the reporter gene to lower levels in
preferentially expressed in germ cells, as noted by its high levels of expression in undifferentiated spermatocytes and testicular teratocarcinomas. Since LTRs drive gene expression by interacting with a variety of ubiquitous and cell-type-specific transcription factors, additional HERV-K LTR mutational and deletion studies to identify proteins that are involved in the regulation of proliferation and differentiation of germ cells and other stem cells are under way.

Testicular teratocarcinomas are the primary malignancy that have long been associated with high levels of HERV-K expression as well as particle formation. However, the significance of these findings with respect to HERV-K gene expression (i.e., whether it is involved in or merely a consequence of the development of these tumors) remains unclear. The data presented in this paper support the previous observations that HERV-K proviruses are expressed primarily in GCTs containing EC components. We contributes the novel finding that the differentiation of a GCT leads to a dramatic decrease in HERV-K LTR expression. Furthermore, the LTR is also found to be preferentially expressed in the more undifferentiated spermatocytes of adult testes. Taken together, it can now be concluded that active HERV-K members can be expressed not only in GCTs but also in healthy adult germ cells. Experiments to cross the LTRE3 transgenic mice with strains of mice that have a high predisposition to testicular cancer are under way. This should be useful in determining the extent to which this virus is expressed as the cancer develops.

ACKNOWLEDGMENTS

We thank A. Teresky for help in the initial setup of the mouse colonies. We are indebted to J. Moran, B. Prabhakaran, P. Robinson, J. Levorsee, M. Shin, and C. Unrine for helpful suggestions and protocols for the analysis of the transgenic mice. We also thank A. Muller for helpful discussions and critical reading of the manuscript. This work was supported by cancer training grant 5T32 CA09528-14 from the National Institutes of Health.

REFERENCES