Functional Analysis of the Endogenous Retroviral Promoter of the Human Endothelin B Receptor Gene

Josette-Renée Landry and Dixie L. Mager*

*Terry Fox Laboratory, British Columbia Cancer Agency, and Department of Medical Genetics,
University of British Columbia, Vancouver, British Columbia, Canada*

We previously reported that the long terminal repeats (LTRs) of retroviral elements belonging to the HERV-E family contribute to the expression of the human apolipoprotein C1 (*APOC1*) and endothelin B receptor (*EDNRB*) genes by providing alternative promoters. While both LTRs were shown to promote transcription in vivo and in vitro, their respective activity and tissue specificity appeared to differ even though they shared a high degree of sequence identity. In the present study, we further characterized the promoter of the *EDNRB* LTR and delineated the regions and motifs required for strong activity. We confirmed the placenta-restricted expression of the LTR by transient transfections and quantitative real-time PCR and determined that the retroviral promoter contributes significantly to the level of *EDNRB* transcripts in placenta, where chimeric mRNAs were found to represent 15% of overall *EDNRB* mRNAs. Transient transfection of 5′ deletion constructs in cells of placental origin identified a motif, named LPE1, between positions 111 and 122 of the *EDNRB* LTR necessary for transcriptional activity. Removal of this region, which contains a putative SP1 binding site, abolished promoter activity. A second enhancing region resides between positions 175 and 215 of the LTR and was termed LPE2. Interestingly, this section contained three binding sites that were not present in the *APOC1* LTR due to minor nucleotide differences. The predicted motifs in the *EDNRB* LTR were found to likely act in symbiosis as modifications to any of the three sites reduced transcription by one-third while alterations to all three eliminated promoter activity. The results from this study illustrate how slight variations in transcriptional regulatory sequences can have a profound effect on promoter activity and demonstrate the complex regulatory effects of human endogenous retrovirus elements on human gene expression.

Endogenous retroviruses (ERVs) are important constituents of mammalian genomes. In humans, retroviral-like elements and long terminal repeat (LTR)-containing sequences account for approximately 8% of the genome (9) while in mice they comprise nearly 10% of nuclear DNA (27). Most ERVs are believed to be derived from germ line infections by ancient retroviruses and are broadly grouped into three classes based on the sequence similarity of their *pol* regions to those of exogenous retroviruses. Class II, for example, includes sequences related to mammalian betaretroviruses, such as the mouse mammary tumor virus. Human endogenous retroviruses (HERVs), which are reviewed elsewhere (11, 28), are further categorized with respect to the tRNA believed to be used for priming in the reverse transcription of the RNA virion into DNA. Therefore, a retroviral element that carries a primer binding site homologous to the 3′ end of a glutamic acid tRNA is called HERV-E.

While full-length elements resemble exogenous retroviruses in genomic structure, the vast majority, if not all, of HERVs are noninfectious today as they are no longer able to code for retroviral proteins. However, several HERVs have retained functional transcriptional elements, such as promoter and enhancer sequences, within their LTRs (18). Some of these retroviral regulatory elements have evolved a cellular function by contributing to the expression of nearby genes (2, 6, 10, 12, 19,

22). HERV-E elements, more specifically, have been shown to participate in the regulation of several human genes by donating a tissue-specific enhancer to the salivary amylase gene (17, 22) and contributing an alternative promoter to the *MID1* gene (10). In addition, the LTR of an HERV-E element was recently shown by our group to participate in the transcription of the endothelin receptor type B mRNAs by acting as an alternative promoter (12).

Endothelin receptor type B is a G protein-coupled, seven-transmembrane receptor encoded by the *EDNRB* gene on chromosome 13 (1). It is one of the two receptors by which the potent vasoactive effects of endothelins are mediated (16). In addition, mutations in *EDNRB* are likely implicated in the etiology of Hirschsprung disease (3, 4), a multigenic congenital disorder characterized by the absence of ganglions cells along a segment of the intestine (OMIM 142623). The *EDNRB* gene appears to be alternatively spliced, as several transcript variants have been reported in the literature (7, 23). At least four 5′ isoforms have been described for this gene which are likely derived from a common promoter region as their transcription initiation sites clustered within a 1-kb region (Fig. 1). Our group has also recently characterized an additional 5′ variant which does not originate from this shared promoter but instead initiates 57.5 kb further upstream from an HERV-E LTR (12) (Fig. 1). This retroviral element contributes to the expression of the *EDNRB* gene by providing a placenta-specific promoter and possibly an enhancer (12).

Our group has also shown that a different HERV-E LTR alternatively promotes a gene involved in lipid metabolism, *APOC1*, in several tissues (12). Although the *APOC1*- and

* Corresponding author. Mailing address: Terry Fox Laboratory, B. C. Cancer Agency, 601 W. 10th Ave., Vancouver, British Columbia V5Z 1L3, Canada. Phone: (604) 877-6070, ext. 3185. Fax: (604) 877-0712. E-mail: dixie@interchange.ubc.ca.
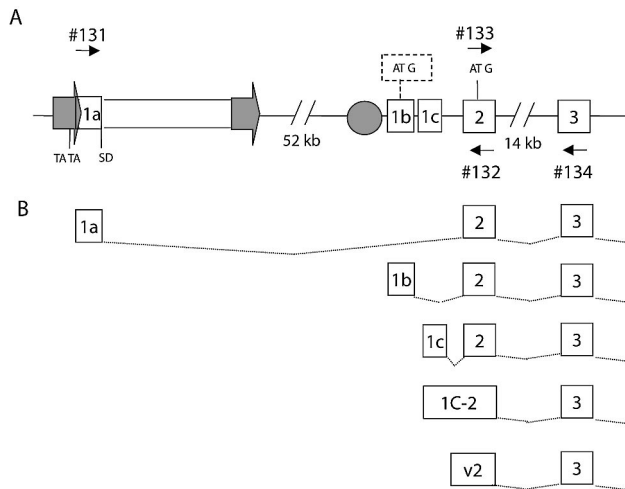
A

B

FIG. 1. Schematic representation of the human *EDNRB* locus. (A) Genomic organization of *EDNRB*, where the position of the HERV-E retroviral element, depicted as a rectangle flanked by two arrows, is shown with respect to the second exon of *EDNRB*. The location of the putative TATA box is indicated in the LTR (arrow) and the splice donor site is indicated (SD). The HERV-E element is approximately 6 kb in length and resides 52 kb upstream of the native promoter region, which is depicted as a circle. The transcription of four variant 5′ untranslated region isoforms appear to initiate within 1 kb downstream of the native promoter. The alternative transcripts possess identical coding regions starting with the ATG indicated in exon 2, with the exception of the isoform 1b, which contains an upstream ATG (in a dashed box) The small numbered arrows represent oligonucleotides used for real-time PCR. (B) Illustration of the alternative 5′ transcript forms.

EDNRB-associated HERV-E LTRs share a high sequence identity, their promoter activity appeared to vary in both strength and tissue specificity (12). We hypothesized that sequence differences between the LTRs resulted in the presence of specific transcription factor binding sites which, for the *EDNRB*-associated LTR, conferred strong placenta-restricted transcriptional activity. This study confirms the tissue-specificity of both LTRs and dissects the regions in the *EDNRB* LTR that contribute to high placental expression.

MATERIALS AND METHODS

Reverse transcription and real-time PCR. Total RNA from human adult and fetal tissues was purchased from BD BioSciences Clontech. Following the elimination of remaining genomic DNA with DNase (Gibco BRL), first-strand cDNA was synthesized as previously described (13) by using random primers, SuperScript II reverse transcriptase (Gibco BRL), and 10 μg of RNA in a total volume of 250 μl.

Real-time PCR was performed on a 1/100 volume (2.5 μl) of cDNA with 25 μl of 2× Sybr Green PCR master mix (PE Applied Biosystems) and the following amplification conditions: 30 s at 95°C, 30 s at 55°C, and 30 s at 72°C for 35 cycles on a Bio-Rad iCycler.

For *APOC1*, primers 129 and 130 (located in exons 3 and 4, respectively) were used to amplify all *APOC1* transcripts and primers 127 and 128 (located in the retroviral first exon, 1R, and exon 3, respectively) were used to amplify only *APOC1* transcripts containing the HERV-E sequence. For *EDNRB*, primers 133 and 134 (located in exons 2 and 3, respectively) were used to amplify all *EDNRB* transcripts and primers 131 and 132 (located in the retroviral first exon, 1R, and exon 2, respectively) were used to amplify only *EDNRB* transcripts containing the HERV-E sequence. Finally, primers 60 and 61 were used to amplify all glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcripts. All primer pairs used for real-time PCR are listed in Table 1 and were designed in accordance with the PE Applied Biosystems guidelines to ensure that the amplification efficiencies of the different primer pairs were close to equal. For that purpose, all primer pairs amplified sequences of similar lengths (~100 bp). In addition, validation experiments were conducted to confirm that the amplification efficiencies of the products were very similar. Dissociation curves were run to detect nonspecific amplification, and it was determined that single products were amplified in each reaction. The relative quantitation of *APOC1* and *EDNRB* chimeric and total expression was calculated by using the comparative threshold cycle method (PE Applied Biosystems user bulletin no. 2, ABI PRISM 7700 sequence detection system), and the levels of *EDNRB*- and *APOC1*-amplified transcripts were normalized to the levels of *GAPDH* obtained. For *EDNRB*, results were expressed relative to the level of *EDNRB* transcripts in testis tissue, which was given an arbitrary value of 1 as it contained the least amount of *EDNRB* mRNA in the various tissues tested. For *APOC1*, results were expressed relative to the level in skeletal muscles.

Sequence analysis. Pairwise alignment of the *APOC1* and *EDNRB* 5′ LTRs was performed by ClustalX, version 1.8 (21), and displayed by using Genedoc, version 2.6 (14). Putative transcription factor binding sites in the retroviral promoters were predicted by using Alibaba, version 2.1, and the Transcription Element Search System (http://www.cbil.upenn.edu/tess/index.html).

Plasmid constructions. The retroviral promoter constructs were designed by cloning the 5′ LTR of the *EDNRB* and *APOC1* HERV-E elements into the *Kpn*I/*Bgl*II sites of the pGL3 basic luciferase vector (Promega). The 462-bp LTRs were amplified from genomic DNA with primers 16 and 17 followed by a nested PCR with primers 18 and 19 for the *EDNRB* LTR and primers 13 and 14 for the *APOC1* LTR. The region amplified and cloned for the *EDNRB* 5′ LTR represents positions 59303 to 58842 of accession number AL139002.5 while the *APOC1* 5′ LTR is present at positions 26537 to 26992 of accession number AF050154.1. All oligonucleotides used in plasmid construction are listed in Table 1.

Progressive 5′ deletion constructs of the retroviral promoter were generated by amplifying the *EDNRB* LTR with the following primers: oligonucleotides 75 and 19 for the fragment from positions 241 to 462 of the LTR, oligonucleotides 78 and 19 for positions 191 to 462, oligonucleotides 79 and 19 for positions 131 to 462, oligonucleotides 109 and 19 for positions 122 to 46, oligonucleotides 86 and 19 for positions 111 to 462, and oligonucleotides 80 and 19 for positions 97 to 462. The resulting LTR sections were then cloned in the *Kpn*I-*Bgl*II site of pGL3B.

For the mutation constructs Mut A, B, C, and D, in vitro mutagenesis was performed by amplifying positions 111 to 462 of the *EDNRB* LTR with oligonucleotide 19 and the following mutating oligonucleotides: oligonucleotide 108 to generate the mutations T124A, C125A, T126A, G127A, and G128A (Mut B); oligonucleotide 110 for C118A, C119A, and G121A (Mut B); oligonucleotide 111 for C115A and C117A (Mut C); and oligonucleotide 112 for G111A, C112A, and C113A (Mut D). The mutated LTR fragments, representing positions 111 to 462, were then inserted in the multicloning site of the luciferase plasmid pGL3B.

The hybrid *APOC1* and *EDNRB* LTR constructs were generated by digesting the LTRs cloned in pGL3B (see above) with a common restriction enzyme that cuts both LTRs once at the same position. The restriction enzymes *Pfl*mI (restriction site present at position 118 of the LTRs), *Sfc*I (position 168), *Sph*I (position 247), and *Sst*I (position 358) were used to cleave the LTRs into 2 segments. Following digestion, restriction fragments from both LTRs were electrophoresed on a 1.5% agarose gel and purified by using Qiaex II gel extraction kits (Qiagen) and the 5′ fragments of one LTR were ligated to the 3′ segments of the other. The resulting hybrid LTRs were then cloned in the *Kpn*I-*Bgl*II site of pGL3B.

For the mutation constructs Mut E through K, in vitro mutagenesis was performed by amplifying the 3′ part of the *EDNRB* LTR (position 159 to 462) with oligonucleotide 14 and the following mutating oligonucleotides: oligonucleotide 146 to generate mutation C177A (Mut E); oligonucleotide 147 for G190A (Mut F); oligonucleotide 148 for T209C, G213A, and C215T (Mut G); oligonucleotide 149 for C177A, G190A, T209C, G213A, and C215T (Mut H); oligonucleotide 145 for T176A, C177A, C178A, and G180A (Mut I); oligonucleotide 144 for T185A, G186A, T187A, G190A, and G191T (Mut J); and oligonucleotide 143 for T202A, G203A, T209A, and C210A (Mut K). Following amplification, the 3′ fragments of the LTR were digested with *Sfc*I and *Bgl*II and purified by using Qiaex PCR purification kits (Qiagen). Nonmutated 5′ sections of the *EDNRB* LTR were also digested by using *Kpn*I and *Sfc*I and purified as described above. The mutated 3′ parts of the LTRs were then ligated to 5′ fragments of the *EDNRB* LTR and cloned in the *Kpn*I-*Bgl*II site of pGL3B.

Cell culture and transient transfections. The human choriocarcinoma Jeg-3 cell line was maintained in RPMI medium supplemented with 5% fetal calf serum and antibiotics. The human colon cell line DLD-1 was cultured in alpha minimal essential medium supplemented with 10% fetal calf serum and antibiotics. The human glioma U87, lung carcinoma A549, liver carcinoma HepG2,

TABLE 1. Oligonucleotides used for constructs and real-time study

Oligonucleotide	Sequence[a] (5′–3′)	Description
13	cggggtaccTAAGGGAGGAGACCACCCCT	APOC1 LTR *Kpn*I
14	gaagatctTGTAGCAGGAGGAGCCGCAG	APOC1 LTR *Bgl*II
16	AACATCCTCTGTCTCTCC	EDNRB LTR flanking
17	CCAGTTCCTTCCCAGTGT	EDNRB LTR flanking
18	cggggtaccTAAGGGAGGATACCACC	EDNRB LTR *Kpn*I
19	gaagatctTGTAGCAGGACAAGCTGC	EDNRB LTR *Bgl*II
60	GCCCAGGATGCCCTTGA	GAPDH
61	GTGTCCCCACTGCCAAC	GAPDH
75	cggggtaccGTGCATGCAGCCCCCAGT	EDNRB LTR Del 1
78	cggggtaccGAAAAGCACTGTGAAAATTCC	EDNRB LTR Del 2
79	cggggtaccGTTAAAGATCGACCCCTGAC	EDNRB LTR Del 3
80	cggggtaccGCAGACAGTCTGGTGCCA	EDNRB LTR Del 4
86	cggggtaccGCCACACCCTGGGTCTGGTAG	EDNRB LTR Del 3.5
108	cggggtaccGCCACACCCTGGGAAAAATAGTTAAAGATCGACCCCTG	EDNRB LTR Mut4 (A)
109	cggggtaccGGTCTGGTAGTTAAAGATC	EDNRB LTR Del 3.2
110	cggggtaccGCCACACAATAGGTCTGGTAGTT	EDNRB LTR Mut5 (B)
111	cggggtaccGCCAAAACCTGGGTCTGGTA	EDNRB LTR Mut6 (C)
112	cggggtaccAAAACACCCTGGGTCTGGTA	EDNRB LTR Mut7 (D)
127	TCTGAGGACCCCACAGAGT	APOC1 Ex1S
128	GATCGACAGAACCACCACC	APOC1 Ex3A
129	GGTGGTGGTTCTGTCGATC	APOC1 Ex3S
130	CAGTGTGTTTCCAAACTCCTT	APOC1 Ex4A
131	GGGAAGGAACTGGTACTTGG	EDNRB Ex1S
132	ACTTGGAGGCGGCTGCATG	EDNRB Ex2A
133	GACCTGCTGCACATCGTCAT	EDNRB Ex2S
134	CAGCTTACACATCTCAGCTCC	EDNRB Ex3A
143	GGTTATGTTATCTATAGATTCCAGACATTGTATGGAAAAGCACTGAAAAAATAACTGTCCTG	EDNRB LTR Mut8 (K)
144	GGTTATGTTATCTATAGATTCCAGACATAAAATATAAAAGCACTGTG	EDNRB LTR Mut9 (J)
145	GGTTATGTTATCTATAGATAAAAAACATTGTATGGAAA	EDNRB LTR Mut10 (I)
146	GGTTATGTTATCTATAGATTACAGACATTG	EDNRB LTR Mut12 (E)
147	GGTTATGTTATCTATAGATTCCAGACATTGTATAGAAAAGCACTG	EDNRB LTR Mut13 (F)
148	GGTTATGTTATCTATAGATTCCAGACATTGTATGGAAAAGCACTGTGAAAATCCCTATTCTGTTCTGTTC	EDNRB LTR Mut14 (G)
149	GGTTATGTTATCTATAGATTACAGACATTGTATAGAAAAGCACTGTGAAAATCCCTATTCTGTTCTGTTC	EDNRB LTR Mut15 (H)

[a] Lowercase letters represent linker sequence used to facilitate cloning.

and embryonic kidney 293 cell lines were maintained in Dulbecco's minimal essential medium supplemented with 10% fetal calf serum and antibiotics. Cells were seeded 24 h prior to transfection in six-well plates at a density of 2×10^5 cells/well. Monolayers of U87, A549, Jeg-3, and 293 cells were cotransfected with 1.8 μg of plasmid DNA and 200 ng of the *Renilla* luciferase vector pRL-TK by using 7 μl of Lipofectamine (Life Technologies). DLD-1 cells were cotransfected similarly by using 4 μl of Lipofectamine (Life Technologies) and 6 μl of Plus Reagent (Life Technologies). HepG2 cells were cotransfected with 1.5 μg of plasmid DNA and 50 ng of the luciferase vector pRL-TK by using calcium phosphate (Cellphect) as described by the supplier. All cells were washed 24 h following transfection in phosphate-buffered saline and harvested in 500 μl of 1× passive lysis buffer (Promega). Firefly and luciferase activities were measured by using the dual-luciferase reporter assay system (Promega). The data were normalized to the internal *Renilla* luciferase control and expressed with respect to pGL3B (basic promoterless vector).

RESULTS

Expression pattern and contribution of *APOC1* and *EDNRB* chimeric transcripts. To further characterize the abundance and tissue specificity of the chimeric mRNAs, we performed real-time PCR on cDNAs from various tissues. The hybrid transcript forms were amplified with primers specific to the retroviral isoforms while total *APOC1* and *EDNRB* transcripts were amplified using primers located in invariant exons found in all transcript forms (*EDNRB* primer locations are shown in Fig. 1). We determined the contribution of the LTR in driving the transcription of the *APOC1* and *EDNRB* genes by calculating the percentage of overall transcripts that contained the retroviral first exon and by evaluating the overall abundance of transcripts when normalized to *GAPDH* levels. As shown in Fig. 2, the EDNRB retroviral isoform had a very low relative abundance (less than 1%) in all tissues tested, with the exception of testis and placenta tissues. While less than 5% of *EDNRB* mRNAs were chimeric in testis tissue, 15% possessed a retroviral first exon in placenta. As overall *EDNRB* transcripts appeared to be abundant in placenta tissue, this suggests that high levels of *EDNRB* chimeric transcripts are present in this tissue.

In contrast, the overall levels of *APOC1* appeared to be very low in all tissues tested, with the exception of liver and fetal liver tissues (data not shown). While the chimeric forms of *APOC1* were found to represent over 10% of overall *APOC1* transcripts in several tissues, the LTR did not significantly contribute to the hepatic levels of *APOC1* (data not shown).

Transcriptional activity of the retroviral promoters. Results from past experiments in which the *APOC1* and *EDNRB* LTR reporter constructs had been tested in liver and placenta cell lines, respectively, indicated that the *APOC1* LTR was a weak promoter while the *EDNRB* LTR had very strong transcriptional activity in the cell line tested (12). Because of the new knowledge regarding the tissues in which the relative abundance of the chimeric transcripts was high, we decided to
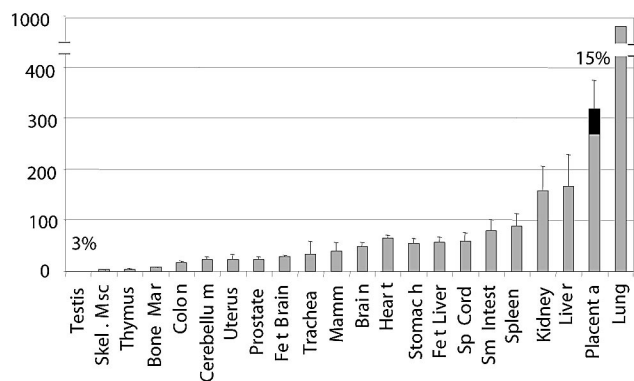
FIG. 2. Proportion of *EDNRB* transcripts contributed by the LTRs. The relative abundance of chimeric *EDNRB* transcripts compared to overall mRNA levels of *EDNRB* is shown. Total cDNAs from various human tissues were subjected to real-time PCR with primers (primer positions are given in Fig. 1 and Materials and Methods) that amplified either all *EDNRB* transcripts or only those with a retroviral first exon (chimeric). Total *EDNRB* levels, normalized to *GAPDH* levels, are depicted by grey bars + standard deviations. The black portions represent the percentages (written above the bars) of overall *EDNRB* mRNAs that possess retroviral first exons. Values are plotted for tissues in which more than 1% of *EDNRB* mRNAs are chimeric. Skel. Msc, skeletal muscle; Mar, marrow; Fet, fetal; Mamm, mammary; Sp, spinal; Sm Intest, small intestine.

further analyze the promoter activity of the retroviral LTRs. We transiently transfected the same luciferase plasmids in which the LTRs had been inserted upstream of the reporter gene (12) into various cell lines, including those derived from tissues where the levels of retroviral isoforms were elevated. As shown in Fig. 3, the *APOC1* LTR was determined to have weak promoter activity in colon, brain, lung, liver, kidney, and placental cell lines. On the other hand, the transcriptional activity of the *EDNRB* LTR was found to be very high in a placental cell line but nearly absent in other cell types. Similarly, retroviral *EDNRB* isoforms were shown to represent a high proportion of total transcripts in placenta tissue. The disparity in promoter activity and tissue specificity between the two LTRs is striking as the two LTRs belong to the same group of endogenous retroviruses, the HERV-E family, and are 85% identical. Figure 4 shows a comparison of the sequence of the 5′ LTR of the HERV-E elements in the *APOC1* and *EDNRB* locus. Since the LTRs were similar in sequence but not in transcriptional activity, we decided to dissect the *EDNRB* retroviral promoter in order to identify regions that conferred high promoter activity in placental cells.

Mapping of the EDNRB LTR promoter. To begin the characterization of the retroviral promoter, we generated a series of 5′ *EDNRB* LTR deletion luciferase constructs. Transfections of these plasmids were carried out in Jeg-3 cells as the LTR promoter appeared to be particularly well utilized in this cell line. Our deletion analysis, shown in Fig. 5, demonstrated that a region present between positions 111 and 122 of the *EDNRB* LTR was necessary for high promoter activity. While a construct containing positions 111 to 462 of the *EDNRB* LTR resulted in a relative luciferase level of nearly 300, the activity obtained with a similar plasmid containing positions 122 to 462 was reduced to a level of less than 5. Therefore, the removal of

the 11 bp from 111 to 122 bp resulted in a 60-fold reduction in promoter activity.

In silico analysis of this small region revealed that a potential Sp1 binding site was situated between positions 111 and 120. The importance of the predicted Sp1 binding site was confirmed by site-directed mutagenesis (Fig. 6). While mutations of several nucleotides between positions 124 and 130 (Mut A) did not significantly reduce the reporter activity, the mutagenesis of 2 or 3 nucleotides between positions 111 and 122 (Mut B, C, and D) severely decreased the promoter activity. Combined, the results obtained by the mutation and deletion analysis indicate that the nucleotide sequence between positions 111 and 122 of the *EDNRB* LTR is required for strong promoter activity in placental cells. Although transcription factor binding site software identified a putative Sp1 binding site in this motif, the identity of the protein could not be confirmed by an electrophoretic mobility shift assay with an Sp1 antibody. While specific binding of this region with a Jeg-3 nuclear extract was competed by an Sp1 consensus oligonucleotide, it could be not supershifted by an Sp1 antibody (results not shown). The functional domain isolated is therefore likely bound by other members of the Sp1 superfamily. Since the identity of the transcription factor was not resolved, the motif present between positions 111 and 122 of the *EDNRB* LTR will be referred to as LPE1 for LTR placental enhancer 1.

Hybrid construct experiments. While the region from position 111 to 122 of the *EDNRB* LTR appears to be indispensable for high activity in Jeg-3 cells, this retroviral segment cannot be sufficient for high activity, as an identical motif is present in the same location in the *APOC1* LTR (Fig. 4). To further delineate additional regions important for high pro-
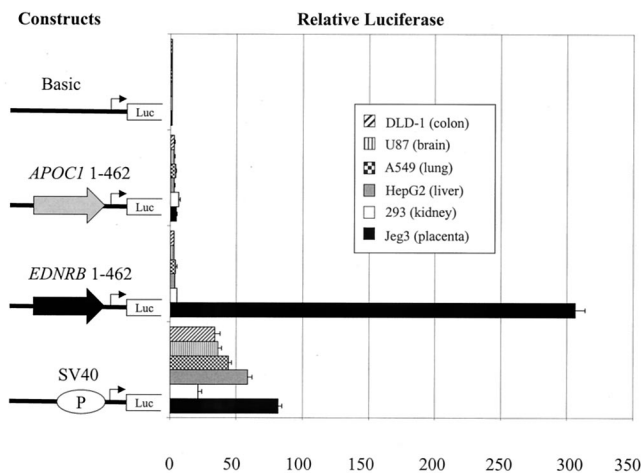

FIG. 3. Promoter activity of the *APOC1* and *EDNRB* LTRs. Representation of the retroviral promoter constructs in which the *APOC1*- or *EDNRB*-associated LTRs were inserted upstream of the promoterless pGL3B vector and transiently transfected into the DLD-1, U87, A549, HepG2, 293, and Jeg-3 cell lines. The basic pGL3B vector and the simian virus 40 (SV40) promoter pGL3p plasmid were also transfected in the above cell lines. The luciferase activities obtained with each plasmid were corrected for transfection efficiency with the *Renilla* luciferase pRL-TK plasmid and are presented as increases (*n*-fold) over the activity of the basic (pGL3B) vector, which was assigned a value of 1. Each bar gives the mean of relative luciferase activity from at least 2 experiments ± standard deviation.

```
                    *        20          *        40          *        60          *        80
EDNRB : TAAGGGAGGAT ACCACCCCTCATATTGTCTTATGCCCAATTTCTGCCTC TG AAGAAAGAAT AAGTAAAAACTAAAAGGCA
APOC1 : TAAGGGAGGAG ACCACCCCTCATATTGTCTTATGCCCAATTTCTGCCTC CA AAGAAAGAAA AAGTAAAAACTAAAAGGCA

                                                    LPE1
                    *        100         *        120          *       140          *       160
EDNRB : GAAATGAAATCCAC TGGCAGACAG T CT GGT GCCACACCCTGGG T CTGGT A GTTAAAGAT C GACCCCTGACCTAA CC GGTT
APOC1 : GAAATGAAATCCAC AA GCAGACAG CCC G-C GCCACACCCTGGG C CTGGT G GTTAAAGAT T GACCCCTGACCTAA T CCGTT

                                                    LPE2
                                    A           B             C
                    *        180         *       200          *       220          *       240
EDNRB : AT GTTATCTATAGATT C CAGACATTGTAT G GAAAAGCACTGTGAAAAT T CCT GT C CTGTT C TGTTC T GATCT G ATTACC A
APOC1 : AG GTTATCTATAGATT A CAGACATTGTAT A GAAAAGCACTGTGAAAAT C CCT ATT CTGTT T TGTTC C GATCT A ATTACC G

                    *        260         *       280          *       300          *       320
EDNRB : GTGCATGCAGCCCCCAGT T ATGT ACCTG CTGCTTG CTCAATC AATCAC AACCCT T TCACG CAGACCC CCTTAGAG CTGTG
APOC1 : GTGCATGCAGCCCCCAGT C ACGCAT CCC CTGCTTG TTCAATC GATCAC GACCCT C TCACG TGC ACCCA CTTAGAG TTGTG

                    *        340         *       360          *       380          *       400
EDNRB : AGCCC TTAAAA GG GAT AGGA ATTGCTCACTC A GAGAGCTC AGCTCTTGAGACAGGA GTCTTGCC GAT GCTT CT GGC CGAA
APOC1 : AGCCC TTAAAA GG AAC AGGA GATTGCTCACTC G GGGAGCTC GGCTCTTGAGACAGGA ATCTTGCC CATT CCC C- C AACGAA

                    *        420         *       440          *       460
EDNRB : TAAACCT CTTC T TT CTTT AAT TC GG TA TC C GAGGAATTTTGTCTGC AGCT TG TCCTGCTACA : 462
APOC1 : TAAACC C CTTC C TT ---- AAC TC A CCGTC T GAGGAATTTTGTCTGC GGCT CC TCCTGCTACA : 456
```
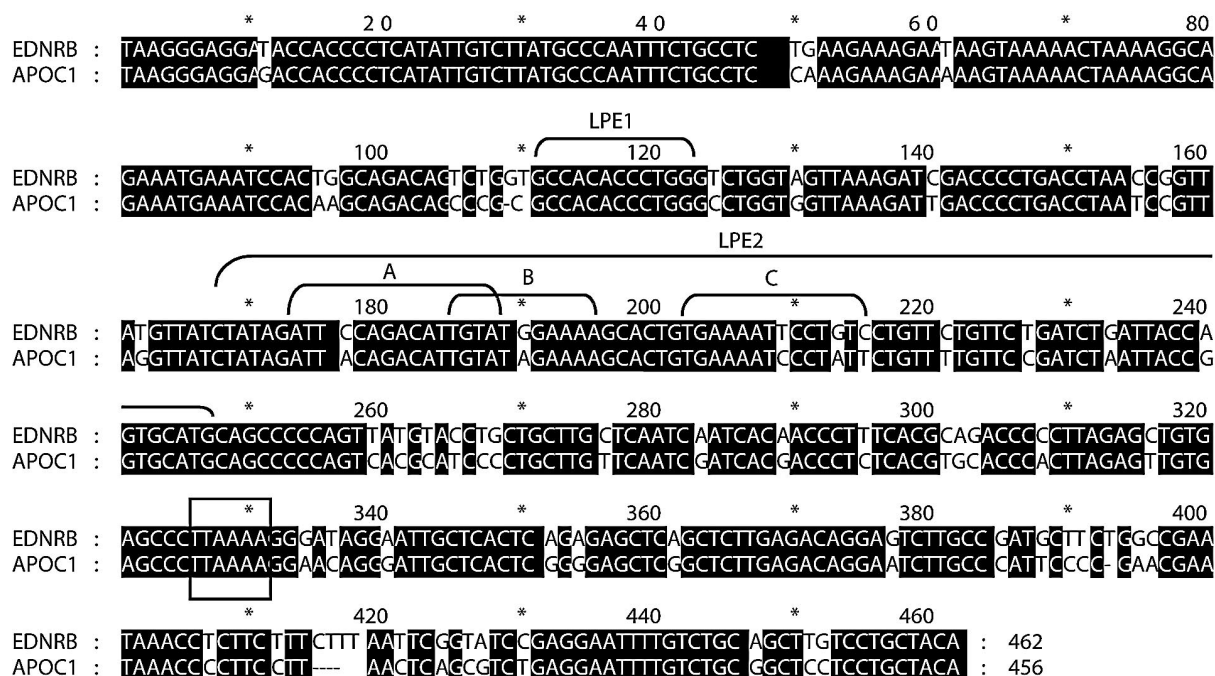
FIG. 4. Sequence comparison of the retroviral promoters of the *EDNRB* and *APOC1* genes. Pairwise alignment of the 5′ LTRs of the *APOC1*- and *EDNRB*-associated HERV-E elements is shown. The shaded regions represent nucleotides that are identical between the two LTRs. Nucleotide numbering starts from the first position of the LTR. The putative TATA boxes and the identified placental enhancers LPE1 and LPE2 are indicated.

moter activity in the *EDNRB* LTR, 3′ deletion experiments could not be performed, as the likely TATA box is present toward the 3′ end of the LTR. Instead we resorted to dissecting the important domains required for promoter activity by creating hybrid LTRs by using the sequence of the *APOC1* and *EDNRB* LTRs. In the first set of hybrid constructs, as shown in top half of Fig. 7, 3′ sections of various lengths of the *EDNRB* LTR were replaced with the corresponding sequences from the *APOC1* LTR. Transient transfection of these constructs in Jeg-3 cells followed by luciferase assays indicated that the second half of the *EDNRB* LTR, from positions 247 to 462, could be replaced by the *APOC1* section without any reduction in transcriptional activity. However, further replacements, such as from positions 168 to 462, completely abolished the promoter activity, suggesting that another region critical for high transcription was present between positions 168 and 247. The importance of this region was confirmed by complementary constructs in which 3′ segments of the *APOC1* LTR were replaced with the appropriate *EDNRB* sections. As shown in lower part of Fig. 7, substituting the 3′ half of the *APOC1* LTR, from position 247 to 462, with the corresponding *EDNRB* LTR sequence did not result in any improvement in transcriptional activity while swapping an additional 79 bp, from position 168 to 247, increased the promoter activity to levels obtained with the full-length *EDNRB* LTR.

Site-directed mutagenesis of the *EDNRB* LTR. We performed sequence analysis with different transcription factor binding site software programs to identify possible *trans* elements that may bind this second functional retroviral segment of the *EDNRB* LTR between positions 168 and 247, which we have called LPE2. Using this approach, three candidate binding sites were identified for which transcription factors had

been previously found to be involved in placental expression; the heterodimer E47/Thing1 (8), Oct-1 (5, 25), and NF-κB (24). These motifs present within LPE2 will be referred to as A, B, and C, as the identities of the proteins that bound to
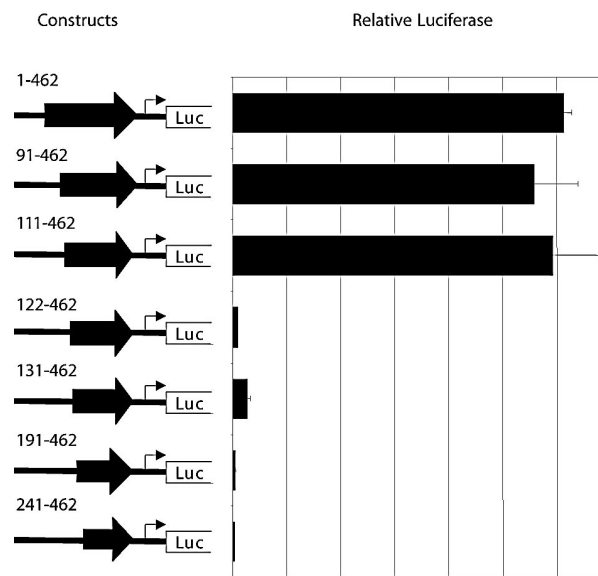

FIG. 5. Effects of 5′ deletions on the transcriptional activity of the *EDNRB* LTR. Representation of the 5′ LTR deletion plasmids transiently transfected in the Jeg-3 choriocarcinoma cell line. All constructs contain variable lengths of the LTR. The name of each plasmid indicates the positions of the LTR included. Plasmid 1-462 corresponds to the entire LTR and is identical to the LTR promoter construct shown in Fig. 3. Results are illustrated as described in the legend to Fig. 3.
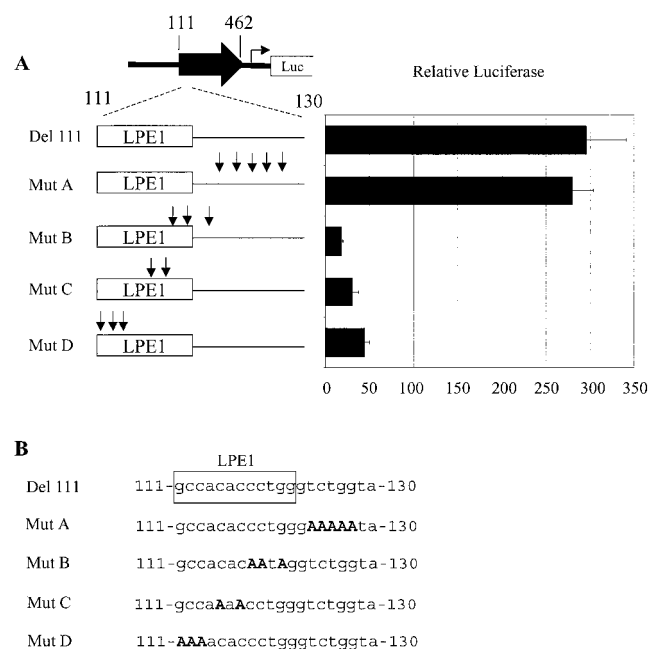
FIG. 6. Confirmation of a *cis* element between positions 111 and 122 of the *EDNRB* retroviral promoter. (A) Mutational analysis of positions 111 to 130 of the *EDNRB* LTR in Jeg-3 cells. All constructs contain positions 111 to 462 of the LTR. While the first plasmid (Del 111) does not have any alterations, the Mut constructs have been mutated at the positions indicated by the arrows. The motif present between positions 111 and 122 of the *EDNRB* LTR is referred to as LPE1. The constructs are not to scale. (B) The sequence of nucleotides 111 to 130 in each construct is indicated, and the LPE1 motif is boxed. Mutations are shown in bold capital letters.

them were not confirmed. Interestingly, for each of the putative binding sites identified in the *EDNRB* LTR between positions 168 and 247, nucleotide differences exist between the *APOC1* and *EDNRB* LTRs; one in motif A, another in B, and three in site C (Fig. 4). To investigate a possible role for these binding sites and to determine whether the sequence variation between the *APOC1* and *EDNRB* LTRs at these sites resulted in the promoter strength disparity, we modified the *EDNRB* LTR by mutating the positions that differed in the putative transcription factor binding sites between the two to the sequence present in the *APOC1* LTR. Individually, the changes made in each motif to mimic the *APOC1* LTR sequence only reduced the reporter activity by half (Fig. 8A, Mut E, F, and G). However, in combination (Mut H), the *EDNRB*-to-*APOC1* changes resulted in the ablation of the promoter. Interestingly, the complete removal of any of the putative transcription factor binding sites, by mutation of several nucleotides, also significantly reduced the promoter activity (Fig. 8B). These results suggest that all three binding sites are important and that the proteins which bind to them act in combination. While the transcription factors might have bound to the *APOC1* sequence if only one of three sites differed with respect to the *EDNRB* LTR, the amalgamation of the variant sites likely results in a severely reduced binding efficiency which then leads to the decrease in promoter activity.

DISCUSSION

Regulatory sequences participating in placenta-restricted transcription have been identified for a number of genes (2, 5, 25, 26, 29). In some cases, the proteins that confer tissue specificity appear to be ubiquitously expressed transcription factors. For example, important motifs in the placenta-specific promoter of the human gonadotropin-releasing hormone receptor gene (*GNRH*) have been found to interact with the common *trans* factors, Oct-1, CRE, GATA, and AP1 (5). In other reported placenta-specific elements, the transcription factors involved are preferentially expressed in the placenta. A novel placenta-restricted *trans* factor was shown to bind a placenta-specific element, with the sequence CATGGCCTGAA CTAGTTTT, in the enhancer of the human leukemia inhibitory factor receptor (*LIFR*) gene (25). Other placental DNA binding proteins which have been identified are the trophoblast-specific element (TSE) binding protein that recognizes the core sequence RNCCTNNRG in the TSE of the aromatase cytochrome P450 gene (30) and the protein hGCMa that binds to a second element, TSE2 (CATAAGACCCTCATTCCAGA GG), in the human aromatase gene (29). This last tissue-restricted protein, hGCMa, has also been shown to recognize the PLE1 (placental leptin enhancer) element (CAGTACCC
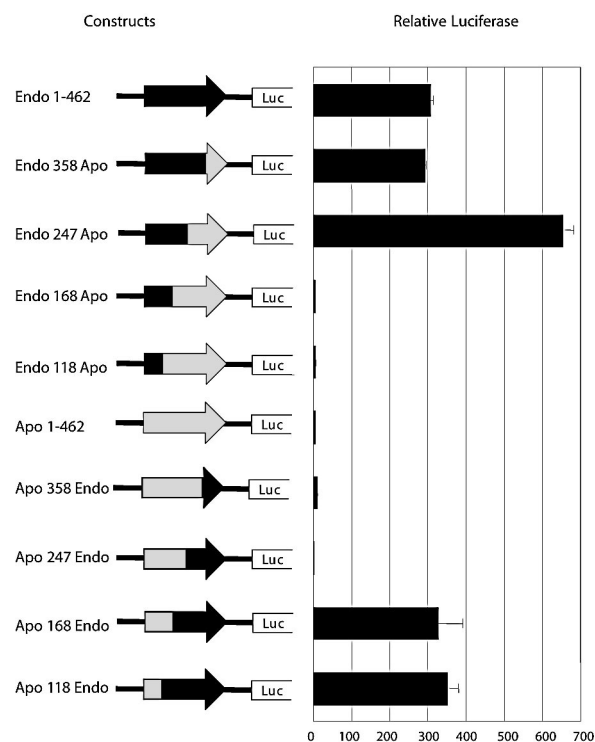

FIG. 7. Fusion study of the *APOC1* and *EDNRB* LTRs. Schematics of the hybrid *APOC1-EDNRB* constructs transfected in Jeg-3 cells to delimit the regions necessary for high promoter activity are shown. The fusion constructs were designed by using common restriction sites present in both LTRs and are named based on the position at which the LTR section was swapped. For example, the construct Endo 358 Apo contains positions 1 to 358 of the *EDNRB*-associated LTR followed by positions 359 to 456 of the *APOC1* LTR. The black sections of the arrows (LTRs) represent sequences from the *EDNRB* LTR while the grey regions are from the *APOC1* LTR.
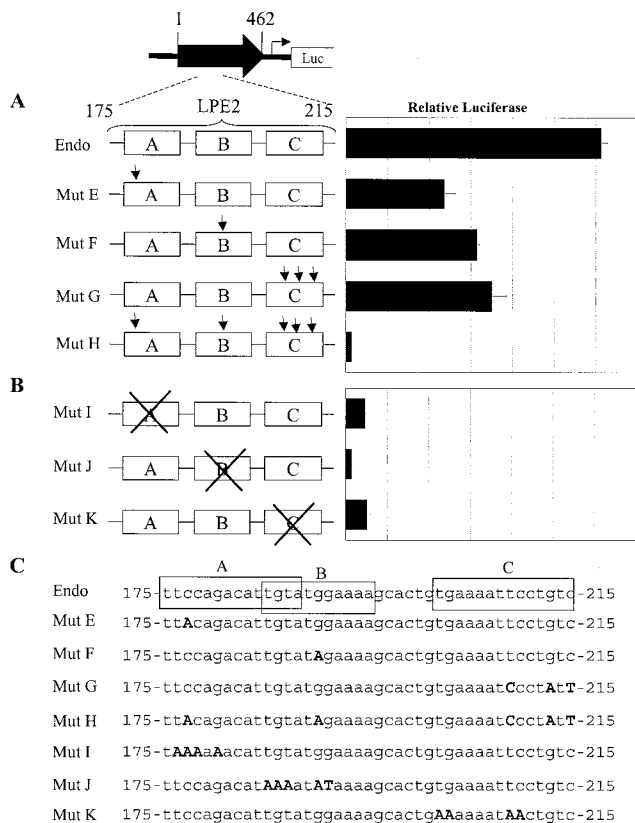
FIG. 8. Mutational analysis of positions 175 to 215 of the *EDNRB* retroviral promoter. All constructs with mutations between positions 175 and 215 contain the full-length *EDNRB* LTR and were transfected in Jeg-3 cells. (A) In constructs Mut E, F, G, and H, the arrows represent nucleotides that have been replaced so as to have the same sequence as the *APOC1* LTR. (B) In constructs Mut I, J, and K, X designates the putative transcription factor binding sites that have been removed by 3 to 5 mutations. (C) The sequence of nucleotides 175 to 215 in each construct is indicated, and the LPE2 motifs A, B, and C are boxed. Mutated nucleotides are shown in bold capital letters.

TCAGGCTTACTAGGGTGGTGAAAAACTC) in the placental promoter of the leptin gene (2, 29). Interestingly, the identified *cis* element PLE1 and another placental transcription factor binding site, PLE3 (CCTGGTAAATTTGTGGT CAGACCAGTTTTCTGCTCT), were shown to reside within a retroviral HERV-K LTR (2). The protein hGCMa was also recently found to activate the expression of HERV-W-encoded syncytin in placental cells (31).

Other retroviral sequences have also been shown to play important roles in regulating the placental expression of human genes. An HERV-E element was determined by our group to act as an alternative placental promoter for the *MID1* gene (10). Another member of the HERV-E family has also been suggested to contribute to the placenta-specific transcription of the human pleiotrophin (*PTN*) gene (19). An Sp1 binding site in the retroviral *PTN* enhancer was found to be essential for the placental expression of this promoter (20).

We now report that we have isolated two critical regions in the retroviral promoter of the human *EDNRB* gene that are necessary for strong placental transcriptional activity. The first identified motif, which we have named LPE1, is present be-tween positions 111 and 122 of the LTR. Transfection experiments with deletion and mutation constructs indicate that the LPE1 region is essential for promoter activity and stimulates transcription 60-fold. However, LPE1 is not sufficient to confer strong placenta-specific transcriptional activity as another HERV-E LTR, associated with *APOC1* transcription, contains the LPE1 motif but represents a weak promoter in placenta tissue. Interestingly, like the HERV-E placental enhancer of the *PTN* gene, LPE1 was also predicted to contain an Sp1 binding site. Results from mobility shift assays suggest that, while Sp1 protein does not appear to bind LPE1, Sp1-related proteins likely interact with LPE1.

A second positive regulatory element, LPE2, was mapped between positions 168 and 247 of the *EDNRB* LTR. This region appeared to be adequate for high activity in placental cells, as replacing the corresponding segment in the *APOC1* LTR with LPE2 increased the promoter strength from nearly zero to levels on a par with the *EDNRB* LTR. LPE2 was predicted to contain three binding sites, referred to as A, B, and C, for proteins which had been shown to participate in the placental transcription of genes. These were the heterodimer E47/Thing1 and the Oct-1 and NF-κB transcription factors, respectively. The putative Thing1 binding site, which is also known as Hand1, was specially interesting, as this protein has been demonstrated to have a tissue-restricted expression pattern in placenta and heart tissues and had been shown to be important in placentation (15). A putative site for the Hand1/ E47 heterodimer was found in the LPE2 region of the *EDNRB* LTR between positions 173 and 188. Although mutations confirmed the importance of this site, we were unable to confirm binding of Hand1 or E47 to LPE2 by electrophoretic mobility shift assays, as the addition of antibodies to either of these two proteins did not result in supershifts (results not shown). It is possible that novel proteins might interact with LPE2 to confer strong placental-restricted expression.

The LPE1 and LPE2 regions of the *EDNRB* LTR appear to be conserved, which supports an important role for these *cis* elements. An identical LPE1 motif was present in the *EDNRB* LTRs of gorillas, chimpanzees, orangutans, gibbons, and baboons (results not shown). With the exception of that of the chimpanzee, which had one nucleotide difference, the sequence of the LPE2 motif was also identical in the above species across 55 bp of the motif, between positions 171 and 226 (results not shown). To determine whether other HERV-E retroviruses besides the *EDNRB*-associated element possessed the LPE1 and LPE2 elements, we analyzed LTRs derived from an HERV-E phylogenetic study (J.-R. Landry, unpublished data). A survey of the 60 HERV-E LTRs with the highest identity to *EDNRB*, which included the *APOC1* LTR, found LPE1 sequences at the same position in 20 retroviral elements. The first binding site in LPE2, A, was found in 5 HERV-E elements while the second, B, was present in 10 HERV-E LTRs, including 3 of the 5 which also possessed binding site A. Finally, only one HERV-E LTR contained motif C in the same position as in the *EDNRB* LTR, but it did not have either of the A or B binding sites. The LPE1 *cis* element therefore appears to be abundant in HERV-E retroviral elements while the complete LPE2 motif could not be identified in any other LTR.

The *MID1* HERV-E LTR, which had previously been found

by our group to contain a placenta-restricted promoter (10), was not one of the 60 nearest neighbors of the *EDNRB* LTR analyzed, as it belonged to another subfamily (Landry, unpublished). We therefore searched for LPE1 and LPE2 motifs in this functional retroviral promoter, but neither were present. It is likely that the *MID1*-associated HERV-E contributes to placental expression by utilizing different *cis* and *trans* elements.

In summary, we have confirmed the placental specificity of the retroviral promoter of the human *EDNRB* gene and characterized motifs important in its tissue-restricted expression. We have shown that the identified LPE1 and LPE2 regions of the HERV-E element are critical for the strong placental transcriptional activity of the *EDNRB* LTR. Our results illustrate the complexity and diversity of mechanisms by which endogenous retroviral sequences can contribute to the transcription of human genes.

ACKNOWLEDGMENTS

REFERENCES

1. **Arai, H., K. Nakao, K. Takaya, K. Hosoda, Y. Ogawa, S. Nakanishi, and H. Imura.** 1993. The human endothelin-B receptor gene; structural organization and chromosomal assignment. J. Biol. Chem. **268:**3463–3470.
2. **Bi, S., O. Gavrilova, D.-W. Gong, M. M. Mason, and M. Reitman.** 1997. Identification of a placental enhancer for the human leptin gene. J. Biol. Chem. **272:**30583–30588.
3. **Carrasquillo, M. M., A. S. MsCallion, E. G. Puffenberger, C. S. Kashuk, N. Nouri, and A. Chakravarti.** 2002. Genome-wide association study and mouse model identify interaction between RET and EDNRB pathways in Hirschsprung disease. Nat. Genet. **32:**237–243.
4. **Chakravarti, A.** 1996. Endothelin receptor-mediated signaling in Hirschsprung disease. Hum. Mol. Genet. **5:**303–307.
5. **Cheng, K. W., B. K. Chow, and P. C. Leung.** 2001. Functional mapping of a placenta-specific upstream promoter for human gonadotropin-releasing hormone receptor gene. Endocrinology **142:**1506–1516.
6. **Di Cristofano, A. D., M. Strazzullo, L. Longo, and G. La Mantia.** 1995. Characterization and genomic mapping of the ZNF80 locus: expression of this zinc-finger gene is driven by a solitary LTR or ERV9 endogenous retroviral family. Nucleic Acids Res. **23:**2823–2830.
7. **Elshourbagy, N. A., J. E. Adamou, A. W. Gagnon, H.-L. Wu, M. Pullen, and P. Nambi.** 1996. Molecular characterization of a novel human endothelin receptor splice variant. J. Biol. Chem. **271:**25300–25307.
8. **Knofler, M., G. Meinhardt, S. Bauer, T. Loregger, R. Vasicek, D. J. Bloor, S. J. Kimber, and P. Husslein.** 2002. Human Hand1 basic helix-loop-helix (bHLH) protein: extra-embryonic expression pattern, interaction partners and identification of its transcription repressor domain. Biochem. J. **361:**641–651.
9. **Lander, E. S., L. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al.** 2001. Initial sequencing and analysis of the human genome. Nature **409:**860–921.
10. **Landry, J.-R., A. Rouhi, P. Medstrand, and D. L. Mager.** 2002. The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter. Mol. Biol. Evol. **19:**1934–1942.
11. **Lower, R., J. Lower, and R. Kurth.** 1996. The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. Proc. Natl. Acad. Sci. USA **93:**5177–5184.
12. **Medstrand, P., J. R. Landry, and D. L. Mager.** 2001. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J. Biol. Chem. **276:**1896–1903.
13. **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.
14. **Nicholas, K. B., H. B. J. Nicholas, and D. W. I. Deerfield.** 1997. GeneDoc: analysis and visualization of genetic variation. EMBNEW News **4:**14.
15. **Riley, P., L. Anson-Cartwright, and J. C. Cross.** 1998. The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. Nat. Genet. **18:**271–275.
16. **Sakurai, T., M. Yanagisawa, Y. Takuwa, H. Miyazaki, S. Kimura, K. Goto, and T. Masaki.** 1990. Cloning of a cDNA encoding a non-isopeptide-selective subtype of the endothelin receptor. Nature **348:**732–735.
17. **Samuelson, L. C., K. Wiebauer, C. M. Snow, and M. H. Meisler.** 1990. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. Mol. Cell. Biol. **10:**2513–2520.
18. **Schon, U., W. Seifarth, C. Baust, C. Hohenadl, V. Erfle, and C. Leib-Mosch.** 2001. Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats. Virology **279:**280–291.
19. **Schulte, A. M., S. Lai, A. Kurtz, F. Czubayko, A. T. Riegel, and A. Wellstein.** 1996. Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. Proc. Natl. Acad. Sci. USA **93:**14759–14764.
20. **Schulte, A. M., C. Malerczyk, R. Cabal-Manzano, J. J. Gajarsa, H. J. List, A. T. Riegel, and A. Wellstein.** 2000. Influence of the human endogenous retrovirus-like element HERV-E.PTN on the expression of growth factor pleiotrophin: a critical role of a retroviral Sp1-binding site. Oncogene **19:**3988–3998.
21. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
22. **Ting, C. N., M. P. Rosenberg, C. M. Snow, L. C. Samuelson, and M. H. Meisler.** 1992. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. Genes Dev. **6:**1457–1465.
23. **Tsutsumi, M., G. Liang, and P. A. Jones.** 1999. Novel endothelin B receptor transcripts with the potential of generating a new receptor. Gene **228:**43–49.
24. **Wang, C. D., G. D. Chang, Y. K. Lee, and H. Chen.** 2001. A functional composite cis-element for NF kappa b and RBJ kappa in the rat pregnancy-specific glycoprotein gene. Biol. Reprod. **65:**1437–1443.
25. **Wang, Z., and S. Melmed.** 1998. Functional map of a placenta-specific enhancer of the human leukemia inhibitory factor receptor gene. J. Biol. Chem. **273:**26069–26077.
26. **Watanabe, K., C. A. Kessler, C. J. Bachurski, Y. Kanda, B. D. Richardson, J. Stanek, S. Handwerger, and A. K. Brar.** 2001. Identification of a decidua-specific enhancer on the human prolactin gene with two critical activator protein 1 (AP-1) binding sites. Mol. Endocrinol. **15:**638–653.
27. **Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, A. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, et al.** 2002. Initial sequencing and comparative analysis of the mouse genome. Nature **420:**520–562.
28. **Wilkinson, D. A., D. L. Mager, and J.-A. C. Leong.** 1994. Endogenous human retroviruses, p. 465–535. *In* J. A. Levy (ed.), The Retroviridae. Plenum Press, New York, N.Y.
29. **Yamada, K., H. Ogawa, S. I. Honda, N. Harada, and T. Okazaki.** 1999. A GCM motif protein is involved in placenta-specific expression of human aromatase gene. J. Biol. Chem. **45:**32279–32286.
30. **Yamada, K., H. Ogawa, S. I. Honda, N. Harada, and T. Okazaki.** 1995. Regulation of placenta-specific expression of the aromatase cytochrome P-450 gene. J. Biol. Chem. **270:**25064–25069.
31. **Yu, C., K. Shen, M. Lin, P. Chen, C. Lin, G. D. Chang, and H. Chen.** 2002. GCMa regulates the syncytin-mediated trophoblastic fusion. J. Biol. Chem. **277:**50062–50068.