BIOLOGY OF REPRODUCTION **68,** 1422–1429 (2003)
Published online before print 27 November 2002.
DOI 10.1095/biolreprod.102.010322

Placenta-Specific *INSL4* Expression Is Mediated by a Human Endogenous Retrovirus Element[1]

Ivan Bièche,[3,4,6] Anne Laurent,[3,4] Ingrid Laurendeau,[4] Laurent Duret,[7] Yves Giovangrandi,[4] Jean-Louis Frendo,[5] Martine Olivi,[4] Jean-Luc Fausser,[8] Danièle Evain-Brion,[5] and Michel Vidaud[2,4]

Laboratoire de Génétique Moléculaire,[4] and INSERM U427,[5] Faculté des Sciences Pharmaceutiques et Biologiques, Université René Descartes—Paris V, F-75006 Paris, France
Laboratoire d'Oncogénétique,[6] Centre René Huguenin, F-92211 St-Cloud, France
Laboratoire BGBP-UMR CNRS 5558,[7] Université Claude Bernard, Lyon 1, F-69622 Villeurbanne Cedex, France
Institut d'Embryologie EA3428,[8] F-67085 Strasbourg, France

ABSTRACT

The human insulin-family genes regulate cell growth, metabolism, and tissue-specific functions. Among these different members, only *INSL4* gene shows a predominant placenta-specific expression. Here, we show that the human *INSL4* gene is tightly clustered with three other members of the human insulin superfamily (*RLN1, RLN2,* and *INSL6*) within a 176-kilobase genomic segment on chromosome region 9p23.3–p24.1. We also report evidence that *INSL4* is probably the only insulin-like growth factor gene to be primate-specific. We identified an unexpected human endogenous retrovirus (HERV) element inserted into the human *INSL4* promoter with a sequence similar to that of *env* gene, flanked by two long terminal repeats(LTRs). The emergence of *INSL4* gene and genomic insertion of HERV appear to have occurred after the divergence of New World and Old World monkeys (~45 million years ago). Transient transfection experiments showed that the placenta-specific expression of *INSL4* is mediated by the 3′ LTR of the HERV element, and that the latter may have a major role in *INSL4* up-regulation during human cytotrophoblast differentiation into syncytiotrophoblast. Finally, we identified an *INSL4* alternatively spliced mRNA species that encodes putative novel INSL4-like peptides. These data support the view that ancient retroviral infection may have been a major event in primate evolution, especially in the functional evolution of the human placenta.

gene regulation, insulin, placenta, pregnancy, relaxin

INTRODUCTION

Insulin and insulin-like growth factors belong to a family of polypeptides that are essential for the proper regulation of physiological processes, including cell growth, cell differentiation, development, and energy metabolism [1]. The human insulin family members include vertebrate insulin (*INS*); insulin-like growth factors I and II (*IGF-I* and *IGF-II*); relaxins (*RLN1* and *RLN2*); Leydig cell insulin-like peptide (*INSL3*); early placenta insulin-like peptide (*INSL4*); and three recently discovered insulin-like hormones, *INSL5* [2], *INSL6* [3], and *RLN3* [4].

We have previously shown that *RLN1*, *RLN2*, and *INSL4* map to chromosome region 9p23.3–p24.1 [5]. *INLS6* has also been mapped using a human-rodent radiation hybrid panel to chromosome arm 9p24 [3], whereas the other insulin-like genes (*INS*, *IGF-I*, *IGF-II*, *INSL3*, *INSL5*, and *RNL3*) are located on other chromosomes. These findings suggest that *RLN1*, *RLN2*, *INSL4*, and *INSL6* may be clustered in the same chromosome region and may have arisen through a mechanism of *cis* local duplication [6]. These four genes are expressed in tissue-specific fashion: *RLN1* and *RLN2* are primarily expressed in the prostate [7], *INSL6* in the testis [3], and *INSL4* in the placenta [8].

INSL4 was identified by screening a human cytotrophoblast-subtracted cDNA library, and its expression was found to be highly placenta-specific [8]. *INSL4* is expressed more strongly during the first trimester of pregnancy [8] and in the differentiated syncytiotrophoblast than in cytotrophoblast cells [9], suggesting that *INSL4* may be important in human placental morphogenesis.

To further characterize the placenta-specific expression of *INSL4* relative to the other three insulin-like genes located in chromosome region 9p23.3–p24.1 (*RLN1*, *RLN2*, and *INSL6*), we characterized the genomic organization of *INSL4*, *INSL6*, *RLN1*, and *RLN2* at this chromosome region. We identified a human endogenous retrovirus (HERV) element inserted in the *INSL4* promoter, and examined the significance of its 3′ long terminal repeat (LTR) in the placenta-specific expression of *INSL4*. We also discuss the evolutionary history of the insulin-like gene emergences and the HERV element integration in chromosome region 9p24.

MATERIALS AND METHODS

Samples

Term placentas were obtained from healthy mothers with uncomplicated pregnancies following elective caesarean deliveries. Villous tissue was dissected free of membranes and vessels, rinsed, and minced in Ca^{2+}-free and Mg^{2+}-free Hanks balanced salt solution for cytotrophoblast cell isolation and culture.

Total RNA from normal human tissues (brain, thymus, placenta, liver, testis, breast, prostate, stomach, colon, skin, leukocytes, adrenal gland, pancreas, and uterus) and genomic DNA from *Macaca mulatta* (rhesus monkey) were purchased from Clontech (Palo Alto, CA).

Cercopithecus aethiops (Africa green monkey) cell line (kidney COS-1 cell line) was obtained from the American Tissue Type Culture Collection (Manassas, VA). *Eulemur macaco* (black lemur) and *Calimico goeldii*

[1]This work was supported by the Ministère de l'Enseignement Supérieur et de la Recherche.
[2]Correspondence: Michel Vidaud, Laboratoire de Génétique Moléculaire—UPRES EA 3618, Faculté des Sciences Pharmaceutiques et Biologiques, Université René Descartes—Paris V, 4 Avenue de l'Observatoire, F-75006 Paris; France. FAX: 33 1 44 07 17 54; e-mail: mvidaud@teaser.fr
[3]These authors contributed equally to this work.

Received: 14 August 2002.
First decision: 5 September 2002.
Accepted: 30 October 2002.

ISSN: 0006-3363. http://www.biolreprod.org

TABLE 1. Oligonucleotide primer sequences used.

Genes	Oligonucleotide	Sequence
RPLPO	Upper primer	5′-GGC GAC CTG GAA GTC CAA CT-3′
	Lower primer	5′-CCA TCA GCA CCA CAG CCT TC-3′
INSL4	Upper primer	5′-CCC ATG CCT GAG AAG ACA TTC-3′
	Lower primer	5′-TGT CGT ACC TAA GGC TTG TCC A-3′
INSL6	Upper primer	5′-TTC CCA ACT GTT TAC TGC TTC TTC-3′
	Lower primer	5′-CGA AGC CTA CAG CCC ATA CC-3′
RLN1/2	Upper primer	5′-ACC TAG ACC AGT GGC AGA AAT TGT-3′
	Lower primer	5′-TGC TGT CTG CGG CTT CAC TT-3′
Leptin	Upper primer	5′-ACA TTT CAC ACA CGC AGT CAG T-3′
	Lower primer	5′-CCA TCT TGG ATA AGG TCA GGA T-3′

(Goeldi's marmoset) tissues were kind gifts from Dr. Jean-Luc Fausser (Institut d'Embryologie, EA3428, Strasbourg, France) and Dr. Eric Denamur (Laboratoire d'études de génétique bactérienne dans les infections de l'enfant, EA3105, Hôpital Robert Debré, Paris, France), respectively.

Isolation of INSL4 *Genomic Clones*

The human P1-derived artificial chromosome (PAC) DNA library (UK Human Genome Mapping Project [HGMP] Resource Center, Hinxton, Cambridge, U.K.) was screened at high stringency using a nick-translated (α-^{32}P)dCTP-labeled *INSL4* cDNA probe.

Genomic DNA and RNA Extraction

Total genomic DNA from a human P1-derived artificial chromosome and from *C. aethiops*, *Macaca mulatta* (rhesus monkey), *E. macaco*, and *C. goeldii* tissues was prepared according to the Protein K-phenol method [10].

Total RNA was extracted from cultured placental cells according to the acid-phenol guanidinium method [11]. The quality of the RNA samples was determined by electrophoresis through denaturing agarose gels and staining with ethidium bromide, with visualization of the 18S and 28S RNA bands under ultraviolet light.

Sequencing of PCR Products and Genomic Clones

PCR products and genomic clones were sequenced using the Applied Biosystems (ABI) PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin Elmer Applied Biosystems, Foster City, CA) and the Applied Biosystems model 377 DNA sequencer (Perkin Elmer).

Real-Time Reverse Transcriptase-Polymerase Chain Reaction

The theoretical and practical aspects of real-time quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) on the ABI Prism 7700 Sequence Detection System (Perkin-Elmer) have been described in detail elsewhere [12]. Briefly, total RNA is reverse-transcribed before real-time PCR amplification. Quantitative values are obtained from the threshold cycle (Ct) number at which the increase in the signal associated with exponential growth of PCR products begins to be detected using PE Biosystems analysis software, according to the manufacturer's manuals. The precise amount of total RNA added to each RT reaction mix (based on optical density) and its quality (i.e., lack of extensive degradation) are both difficult to assess. We therefore also quantified transcripts of the *RPLP0* gene (also known as 36B4, encoding human acidic ribosomal phosphoprotein P0) as the endogenous RNA control, and each sample was normalized on the basis of its *RPLP0* content. The relative target gene expression level was also normalized to a calibrator, or 1× sample. Calibrators were the normal human adult tissues that contained the smallest quantifiable amount of target gene mRNA (for the expression pattern of target genes in normal human adult tissues) and the matched cytotrophoblasts from eight placentas (for the expression of target genes during in vitro human cytotrophoblast differentiation).

Final results, expressed as N-fold differences in target gene expression relative to the *RPLP0* gene and the calibrator, termed ''Ntarget'', were determined as follows:

$$\text{Ntarget} = 2^{(\Delta Ct_{sample} - \Delta Ct_{calibrator})}$$

where ΔCt values of the sample and calibrator are determined by subtracting the average Ct value of the target gene from the average Ct value of the *RPLP0* gene. The nucleotide sequences of the oligonucleotide primers are shown in Table 1.

PCR was performed using the SYBR Green PCR Core Reagents kit (Perkin-Elmer). The thermal cycling conditions comprised an initial denaturation step at 95°C for 10 min, and 50 cycles at 95°C for 15 sec and 65°C for 1 min. Experiments were performed with duplicates for each data point.

Construction of INSL4 *Promoter/Luciferase Reporter Gene Plasmids*

Different lengths of the *INSL4* gene promoter were obtained by PCR amplification of human genomic DNA (Roche Diagnostics, Mannheim, Germany) using specific primers containing an artificial *Xma*I site. PCR products were purified and inserted into the multiple cloning site of the promoterless luciferase vector pGL3-basic (Promega, Madison, WI). Cloning details are available from the authors. Deletion mutant del(−719) (−719 to +95) was generated by *Alf*II/*Nhe*I digestion of del(−1707) (−1707 to +95). The identity of isolated clones was confirmed by nucleic acid sequencing.

Cell Culture and Luciferase Assay

The human choriocarcinoma cell line JEG-3 was purchased from American Type Culture Collection and cultured in Minimum Essential Medium (Invitrogen, Cergy Pontoise, France) with 10% fetal calf serum, 2 mM L-glutamine, 50 IU/ml penicillin, and 50 μg/ml streptomycin.

Cells were cotransfected with 10 μg of the appropriate *INSL4* promoter plasmid and 10 μg of pSV-β galactosidase as an internal control plasmid. The cells were harvested 48 h after transfection, and cell extracts were prepared using a cell culture lysis reagent (Promega).

After centrifugation, 20 μl of the supernatant was assayed for luciferase activity by using the Promega assay system on a Lumat LB 9501 luminometer (EG Instruments, Evry, France). The luciferase activity of each construct was normalized for differences in transfection efficiency, based on the results of the β-galactosidase assay. Each *INSL4* promoter construct was tested in duplicate in a minimum of three independent transfections.

Cytotrophoblast Cell Culture

Cytotrophoblast cells were isolated and purified from term chorionic villi after trypsin-DNase digestion and discontinuous Percoll gradient fractionation, as previously described [13]. The cells were plated in triplicate in 60-mm culture dishes (3×10^6 cells/dish) in 3 ml of Dulbecco modified Eagle medium, 2 mM glutamine, 10% heat-inactivated fetal calf serum, and antibiotics (100 IU/ml penicillin and 100 μg/ml streptomycin). The dishes were incubated at 37°C in a humid atmosphere containing 5% CO_2/95% air and were allowed to aggregate, fuse, and form syncytia. The purity of the cytotrophoblast cell population was checked by cytokeratin 7 immunostaining. Syncytium formation was followed by the distribution of desmoplakin and nuclei in cells after fixation and immunostaining [13]. The staining of desmoplakin present at the intercellular boundaries in aggregated cells progressively disappears with syncytium formation.

RESULTS

Genomic Organization of the INSL6, INSL4, RLN1, *and* RLN2 *Genes at Chromosome Region 9p23.3–p24.1*

A continuous genomic sequence determined by the Human Genome Project made it possible to examine the ge-

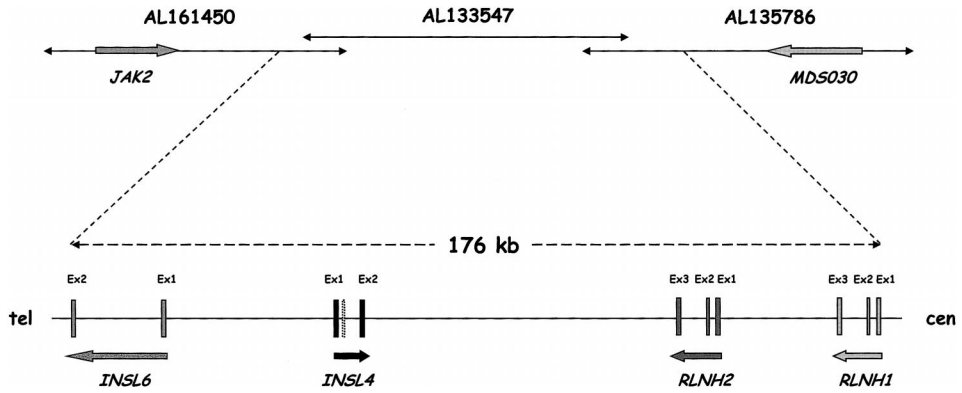

FIG. 1. Genomic organization of the insulin growth factor genes in chromosome region 9p23.3–p24.1.

nomic structure of the *INSL6, INSL4, RLN1,* and *RLN2* genes located on chromosome band 9p23.3–p24.1. We used the *INSL6* (GenBank accession number NM_007179), *INSL4* (NM_002195), *RLN1* (NM_006911), and *RLN2* (NM_134441) cDNA sequences to run a search in the nr (all non-redundant GenBank + EMBL + DDBJ + PDB sequences) and htgs (unfinished high throughput genomic sequences) database with the Gapped basic local alignment search tool (BLAST) program [14]. The four genes, identified on three contiguous overlapping human bacterial artificial chromosomes (BACs) (AL161450, AL133547, and AL135786), were tightly clustered within a 176-kilobase (kb) genomic segment between the *JAK2* gene (NM_004972, coding for a protein tyrosine kinase belonging to the Janus kinase family) and the *MDS030* gene (NM_018465, coding for an uncharacterized protein expressed in hematopoietic stem/progenitor cells).

Figure 1 depicts the order of the genes (*JAK2, INSL6, INSL4, RLN2, RLN1,* and *MDS030*) from telomere to centromere. All the insulin-like growth factor genes, apart from *INSL4,* are transcribed from their centromeric end to their telomeric end.

It is interesting that *INSL6* was physically linked to *JAK2* in chromosome region 9p23.3–p24.1, as are *INSL5* (NM_005478) and *JAK1* (NM_002227) in chromosome region 1p31.3, *INSL3* (NM_005543) and *JAK3* (NM_000215) in chromosome region 19p13.1, and *RLN3* (NM_080864) and *TYR2* (NM_003331) in chromosome region 19p13.2, which are additional members of the insulin gene superfamily (*INSL6, INSL5, INSL3,* and *RLN3*) and the *JAK* gene family (*JAK2, JAK1, JAK3,* and *TYR2*). These findings point to *trans* regional duplication during evolution of the vertebrate genome.

Identification of Two Novel Alternatively Spliced INSL4 *mRNA Species*

Gunnersen et al. [7] detected an alternatively spliced mRNA species incorporating an extra exon in the human *RLN1* and *RLN2* genes, which encodes novel relaxin-like peptides (named RLN1-like and RLN2-like, respectively). We therefore postulated that the human *INSL4* gene, like the RLN genes, was also composed of three exons and two introns.

We used the 1933-base pair (bp) genomic sequence of the *INSL4* intron to run a search in the dbEST database with the Gapped BLAST program. We detected two human expressed sequence tags (ESTs) (GenBank accession numbers H60640 and H02449), which are composite sequences containing either exon 1 (H60640) or exon 2 (H02449) plus intronic sequences of *INSL4.* Analysis of these ESTs and PCR products obtained with EST-specific primers identified two alternative forms of *INSL4* mRNA, *INSL4-like1* and *INSL4-like2,* generated by alternative splicing (Fig. 2A).

The *INSL4-like2* mRNA isoform is generated by the inclusion of an extra 149-bp exon, with exon/intron junctions corresponding to the consensus sequences of donor/acceptor splice sites. Splicing-in of the 149-bp exon introduces a reading-frame shift, leading to a novel predicted peptide (INSL4-like2) that would be identical to the INSL4 peptide in the amino-terminal domain but differ in the carboxy-terminal domain, in the same way as relaxin-like peptides [7].

The *INSL4-like1* mRNA isoform is generated by splicing with an alternative splice acceptor site located 27 bp downstream of the new exon. The use of this alternative splice acceptor site results in the insertion of nine in-frame amino acids relative to the INSL4-like2 peptide.

Comparison of the carboxy-terminal domains of the putative INSL4-like1, INSL4-like2, RLN1-like, and RLN2-like molecules revealed considerable homology, with the same carboxy-terminal end (Fig. 2B). Database searches using PSI-BLAST [14] revealed no significant homology of these four domains with any other human amino acid sequence.

The expression patterns of the *INSL4-like1* and *INSL4-like2* mRNA isoforms were determined by real-time quantitative RT-PCR at the mRNA level in a series of normal adult tissues including brain, thymus, placenta, liver, testis, breast, prostate, stomach, colon, skin, leukocytes, adrenal gland, pancreas, and uterus. Strong expression of the *INSL4-like1* and *INSL4-like2* mRNA isoforms was detected only in placental total RNA. The levels of *INSL4-like1* and *INSL4-like2* variants in placenta were similar to each other, but were markedly lower (~10 times) than the level of the standard *INSL4* transcript (data not shown).

Expression of the INSL6, INSL4, *and* RLN1/2 *Genes in Normal Human Adult Tissues*

Previous studies of *INSL6, INSL4,* and *RLN1/2* gene expression in human tissues have given somewhat inconsistent results [3, 7–9]. To better understand the regulation of *INSL4* gene expression in the insulin-like growth factor gene cluster at 9p23.3–p24.1, we developed real-time quantitative RT-PCR assays to analyze *INSL6* and *INSL4* alone and *RLN1* and *RLN2* together in a variety of normal human adult tissues (Table 2).

INSL4 was expressed strongly only in placenta. Much lower expression (~10000 times lower than in placenta) was detected in thymus, testis, breast, stomach, and uterus.

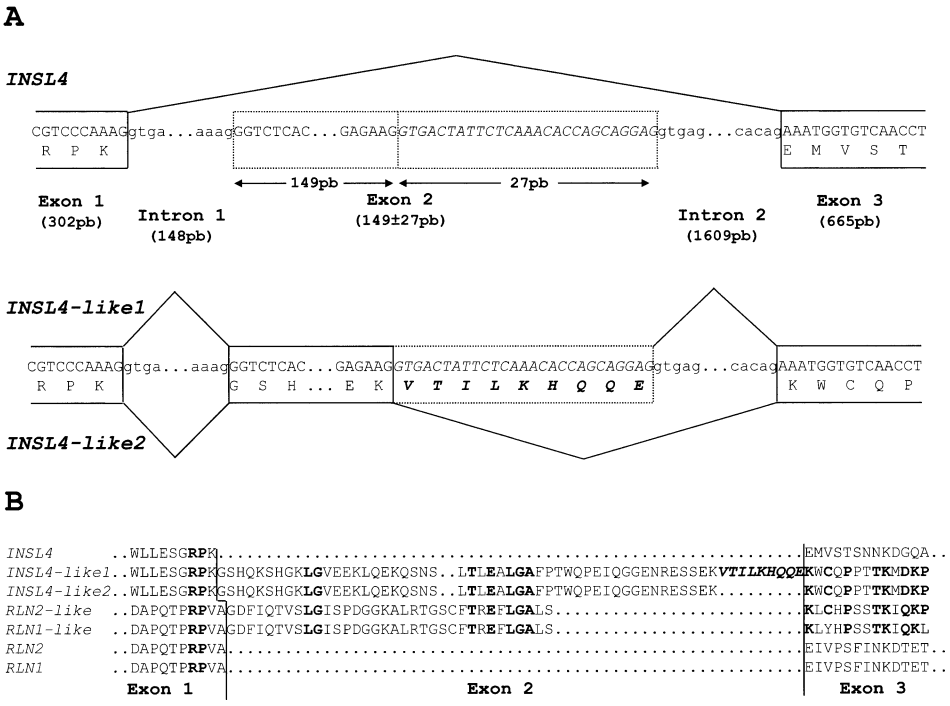

FIG. 2. **A**) Schematic representation of the *INSL4* splice sites. This gene is composed of three exons and two introns. The classical *INSL4* mRNA isoform is generated by alternative splicing of exon 2. The two additional mRNA isoforms (*INSL4-like1* and *INSL4-like2*) are generated by the inclusion of an additional exon, exon 2 (149 ± 27 bp). **B**) Carboxy-terminal domain alignment of the two putative INSL4-like peptides (INSL4-like1 and INSL4-like2) and the two relaxin-like peptides (RLN1-like and RLN2-like) [7]. The amino acid sequence is shown in the single-letter code. Positions of identity between the INSL4-like and RLN-like peptides are in boldface type.

The other tissues contained very little or no detectable *INSL4* mRNA.

INSL6 was expressed strongly only in testis. Much lower expression was found in placenta (~100 times lower) and in thymus and prostate (~1000 times lower). The other tissues contained very little or no detectable *INSL6* mRNA.

RLN1 and *RLN2* were expressed strongly only in prostate. Much lower expression (~1000 times lower) was detected in thymus, placenta, and testis. The other tissues tested contained very little or no detectable *RLN1/2* mRNA.

Identification of an HERV-K Element in the 5′ Flanking Region of the Human INSL4 *Gene*

To better understand the placenta-specific expression of the *INSL4* gene, we sequenced 4 kb of the 5′ flanking region of the *INSL4* gene from a P1-derived artificial chromosome (PAC number 75A3) obtained from a human PAC DNA library (UK Human Genome Mapping Project, HGMP). Using RepeatMasker version 2 (available at http://ftp.genome.washington.edu/cgi-bin/RM2), we found an unexpected HERV element inserted very close to the *INSL4* gene at position −381, with a sequence similar to the *env* gene, flanked by two LTRs. Including the two LTRs, the full-length element is 2434 bp in length.

TABLE 2. *INSL6, INSL4,* and *RLN1/2* gene expression in a series of normal adult tissues.*

Tissues	*INSL6*	*INSL4*	*RLN1/2*
Brain	0	0	0
Thymus	+	+	+
Placenta	++	+++++	+
Liver	0	0	0
Testis	++++	+	+
Breast	0	+	0
Prostate	+	0	++++
Stomach	0	+	0
Colon	0	0	0
Skin	0	0	0
Leukocytes	0	0	0
Adrenal gland	0	0	0
Pancreas	0	0	0
Uterus	0	+	0

* 0 indicates no signal detected after 35 cycles of PCR; + indicates low amount to target mRNA; ++, ++++, +++++ indicates 10-; 1000-; 10 000-fold differences in target gene expression relative to the samples scored +.

Compared with the full-length 3′ LTR (489 bp), the 5′ LTR (275 bp) shows a major deletion at its 3′ end, including the tRNA binding site. The shared portion of the two LTRs differs by a few deletions, insertions, and nucleotide substitutions, resulting in a homology of 89%.

The presence of multiple stop codons in the *env* reading frame, and the absence of matches with ESTs in the dbEST database, suggests that the *env* gene of the HERV is no longer transcriptionally active.

A search of the nr and htgs databases using the HERV genomic DNA as the query identified multiple sequences, including HERV-K(C4) (GenBank accession number X80240) located in the complement C4 gene cluster at 6p21.3 [15]. The most significant sequence hit (GenBank accession number AC010203, map position chromosome 12q22) contained a HERV element (6091 bp long) with sequences similar to *gag* and *env* genes, flanked by two LTRs. It is interesting that this HERV element showed identities to multiple ESTs, suggesting that it might be the ''master'' gene of the *INSL4*-HERV family (i.e., the HERV family member that remained transcriptionally active).

Finally, we tested three contiguous overlapping human 9p23.3–p24.1 BAC clones (AL161450, AL133547, and AL135786), including the *INSL6*, *INSL4*, *RLN1*, and *RLN2* genes, for additional interspersed repeat sequences. No additional HERV elements were detected.

Influence of the HERV 3′ LTR on Human INSL4 *Promoter Activity*

To examine the possible promoter activity of the HERV 3′ LTR, we analyzed the capacity of promoter fragments;

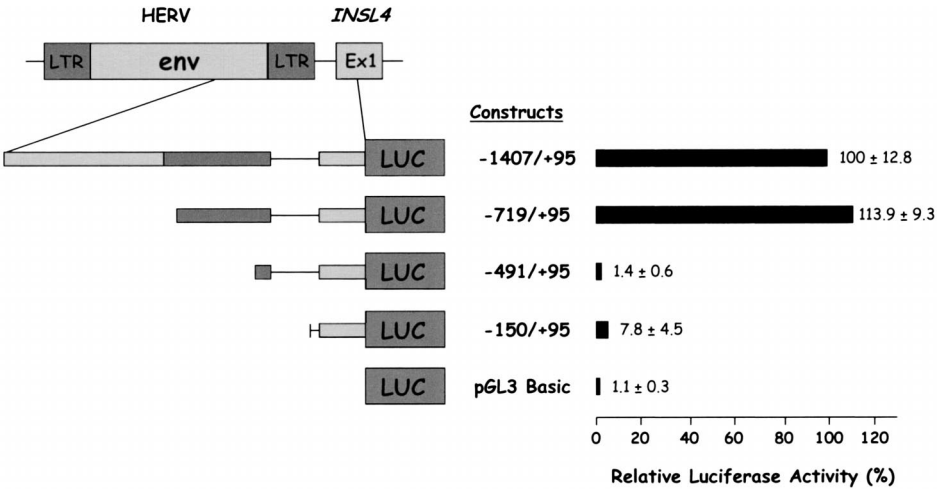

FIG. 3. Deletion analysis of the human INSL4 promoter. JEG-3 cells were cotransfected with the appropriate INSL4 promoter luciferase construct and an RSV β-galactosidase plasmid as described in ''Materials and Methods.'' Luciferase activity data, normalized to β-galactosidase activity, are shown to the right as a percentage of the activity of the construct containing the full-sized 3′ LTR (fragment −1407/+95). Final percentage results are displayed as the mean (± SD) of three independent experiment results.

namely, a 1502 bp fragment including the entire 3′ LTR element (fragment −1407/+95) and shorter fragments with 5′ to 3′ deletions, to drive the expression of the luciferase gene after cloning into the promoterless pGL3-basic vector. The resulting plasmids were transiently transfected into JEG-3 choriocarcinoma cells, and the promoter activity of the constructs was assessed by measuring luciferase activity relative to the transcriptional activity of the construct containing the full-sized 3′ LTR (fragment −1407/+95) (Fig. 3). The results clearly indicated that the 3′ LTR was important for *INSL4* expression in JEG-3 cells. Deletion of the retrovirus-derived promoter portion abolished its activity.

Expression of the INSL4, INSL6, RLN1/2 *Genes During Human Cytotrophoblast Differentiation In Vitro*

Recent studies suggest that HERV elements expressed in the syncytiotrophoblast may be involved in placental morphogenesis [16, 17]. Using an in vitro model of human villous cytotrophoblast differentiation into syncytiotrophoblast [13], we examined the possible role of the *INSL4* gene, via its HERV, in human placental morphogenesis. We quantified *INSL4*, *INSL6*, and *RLN1/2* expression at the mRNA level in matched cytotrophoblasts and syncytiotroblast from eight placentas. Isolated cytotrophoblasts aggregated and fused together in vitro to form a syncytiotrophoblast within 72 h.

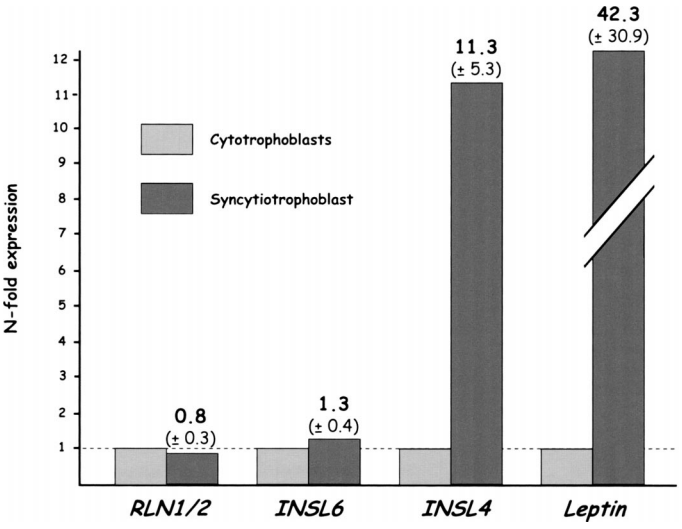

FIG. 4. *RLN1/2, INSL6, INSL4,* and *Leptin* mRNA levels during cytotrophoblast differentiation in vitro. Total RNA was isolated from paired cytotrophoblast and syncytiotroblast samples from eight placentas. Results are expressed as N-fold differences in target gene expression in the syncytiotrophoblast relative to matched cytotrophoblasts (or 1× samples). Final results are expressed as the mean (± SD) of the eight syncytiotrophoblasts results.

We also analyzed the *Leptin* gene, because its placenta-specific expression may be due the presence of an LTR element in its promoter [18].

As shown in Figure 4, *INSL4* expression was up-regulated more than 10-fold in syncytiotrophoblast relative to the corresponding cytotrophoblasts. Similar results were obtained with the *Leptin* gene (more than 40 times higher). In contrast, *INSL6* and *RLN1/2* mRNA levels did not change during cytotrophoblast differentiation.

Evolutionary Conservation of the INSL4 *Gene*

The human INSL4 peptide revealed 44% identity to the human RLN1 peptide, 43% identity to the human RLN2 peptide, and 15% identity to the human INSL6 peptide, suggesting that the *INSL4* gene likely arose from the common precursor *RLN* gene (not from *INSL6*) by a process of gene duplication.

By genomic sequence analysis, we confirmed that Old World monkeys, including the rhesus monkey (*M. mulatta*) and the Africa green monkey (*C. aethiops*), as well as New World monkeys, including Goeldi's marmoset (*C. goeldii*) and prosimians including the black lemur (*E. macaco*) have only one *RLN* gene.

The *INSL4* gene is probably the only insulin-like growth factor gene to be primate-specific. Indeed, 1) a search of the dbEST database using *INSL4* cDNA as the query identified only multiple human expressed sequence tags, whereas the other insulin-like growth factor cDNAs showed multiple murine ESTs; 2) hybridizing fragments using *INSL4* cDNA were observed in human and chimpanzee genomic DNA but not in horse or mouse genomic DNA (data not shown); 3) PCR using *INSL4* degenerated primers and sequence analysis identified the *INSL4* gene in the Old World monkeys (rhesus monkey and Africa green monkey) but not in mouse, horse, prosimians (black lemur), or New World monkeys (marmoset); 4) finally, a search of the nr and htgs databases using the murine *RLN* cDNA (GenBank accession number Z27088) as the query identified a 215-kb mouse genomic clone (AC093339; 10 unordered con-

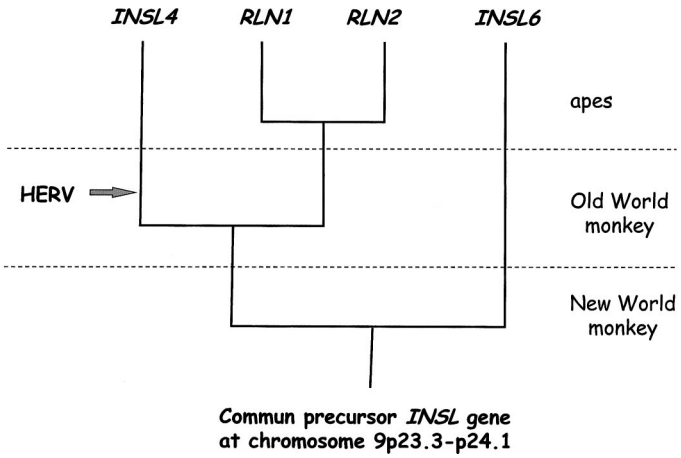

FIG. 5. Evolutionary tree of the insulin growth factor genes in chromosome region 9p23.3–p24.1, and approximate time of HERV element integration into the *INSL4* promoter.

tigs) and a 145-kb rat genomic clone (AC096324; 30 unordered contigs) in the ''working draft'' of the complete murine and rat genomes, which respectively include the mouse and rat genomic DNAs of the unique common precursor RLN gene, but no other significant homology (at the protein or nucleic acid levels) with other known insulin-family nucleic acid or protein sequences. Confirmation of the absence of murine *INSL4* must await full sequencing of the genomic clone AC093339.

Finally, by genomic sequence analysis, we identified the HERV element in the rhesus monkey (Old World monkeys), suggesting that HERV insertion into the *INSL4* promoter occurs relatively soon after the emergence of this gene (Fig. 5).

DISCUSSION

Acquisition of new genetic material during evolution of the vertebrate genome can occur through two duplication mechanisms: 1) *trans* regional duplication of an ancestor chromosome region involving several chromosomes; and 2) *cis* local duplication involving a single chromosome region and leading to the generation of gene clusters from a common precursor gene [6]. Both mechanisms give rise to groups of paralogous genes that constitute many gene families and superfamilies. Results of this study support the conclusion that duplication of an ancestral insulin gene by these two mechanisms generated 10 members of the insulin gene superfamily (*Insulin*, *IGF-I*, *IGF-II*, *RLN1*, *RLN2*, *RLN3*, *INSL3*, *INSL4*, *INSL5*, and *INSL6*) during evolution of the vertebrate genome.

Initially, the common precursor insulin-like growth factor gene at 9p23.3–p24.1 arose by *trans* duplication of an ancestor chromosome region, including one member of the JAK family and one member of the insulin gene family. Indeed, like *RLN1*, *RLN2*, *INSL4*, *INSL6*, and *JAK2* in chromosome region 9p23.3–p24.1, *INSL5* and *JAK1* are physically linked at 1p31.3, *INSL3* and *JAK3* at 19p13.1, and *RLN3* and *TYR2* (a additional member of the *JAK* gene family) at 19p13.2.

Subsequently, *cis* duplication at 9p23.3–p24.1, probably mediated by a series of local rearrangements, generated four insulin-like growth factor genes in chromosome region 9p23.3–p24.1. From telomere to centromere, *INSL6*, *INSL4*, *RLN2*, and *RLN1* are tightly clustered within a 176-kb genomic segment between the *JAK2* gene and the *MDS030* gene. We failed to detect other insulin-like growth factor genes within this cluster, or within the 100- to 200-kb sequences at the telomeric and centromeric ends of this cluster, despite systematic searches.

Contrary to the other insulin-like growth factor genes that have homologous murine genes, *INSL4* appears to be primate-specific. Indeed, we failed to identify any *INSL4* nucleic acid sequences in mouse, horse, and lemur samples despite careful searches, including cross-hybridizing bands, PCR analysis using degenerate *INSL4* primers, murine ESTs in the dbEST database, and INSL4 nucleic acid or protein sequences in two genomic clones (AC093339, AC096324) recently revealed by the working draft of the complete murine and rat genomes, and which contain the mouse and rat common precursor *RLN* genes, respectively.

However, we found *INSL4* homologue genes in the rhesus monkey and African green monkey genomes, suggesting that *INSL4* emerged after the divergence of New and Old World monkeys (~45 million years BP). *INSL6* and the common precursor *RLN* gene are present in the mouse genome [19, 20]. We confirmed the recent duplication event leading to the generation of two *RLN* genes, after the divergence of Old World monkeys and great apes (~25 million years BP). Evans et al. [21] showed that all great apes (chimpanzees, gorillas, and orangutans) have two *RLN* genes, whereas Old World monkeys (rhesus monkey) have only one *RLN* gene, and suggested that *RLN1* is probably nonfunctional in the orangutan and gorilla.

INSL4 probably arose through duplication of the common precursor *RLN* gene rather than *INSL6.* Indeed, the INSL4 peptide sequence is closer to that of RLN1/2 than INSL6. Moreover, we identified extra exons in the *INSL4* gene, as occur in *RLN* genes [7]. We failed to identify such exons in *INSL6* despite careful searches, including human intronic ESTs in the dbEST database and intronic homologies with the extra exons of *INSL4* and *RLN1/2* both at the protein and nucleic acid levels. The conservation of the open reading frames (ORFs) and the homologies between these INSL4-like peptides and relaxin-like peptides suggests that these conserved peptides may have a significant biological role. It is noteworthy that these INSL4-like and relaxin-like molecules lack the A-chain and, thus, conserved cysteines involved in disulfide bonds, which are characteristic of other members of the insulin-related protein family [22].

Surprisingly, we identified a HERV element inserted in the *INSL4* promoter. The 3′ LTR of this HERV appears to mediate the placenta-specific expression of the *INSL4* gene. The HERV element is inserted very close to the *INSL4* gene, at position −381, generating a promoter with new functional features. This HERV element is also present in the promoter of the rhesus monkey (*M. mulatta*) *INSL4* homologue gene, suggesting that genomic insertion occurred throughout Old World monkey evolution (between 25 and 45 million years BP), after duplication of the common precursor RLN gene gave rise to the *INSL4* gene (Fig. 5). Indeed, we failed to identify a HERV element in the 5′ flanking regions of the human *RLN1* and *RLN2* genes. This is in keeping with results reported by Medstrand and Mager [23], which showed that the oldest HERV, in evolutionary terms, is found in Old World monkey lineages.

This 2434-bp HERV element comprises a classical retroviral structure, except for the truncation of the 5′ LTR (including the tRNA binding site) and the absence of the *gag* and *pol* genes. This proviral genome could not produce a functional protein, as the *env* ORF is interrupted by frame

shifts and stop codons. Sequence homology with other HERV sequences suggests that this HERV should be included among the HERV-K elements [24]. We also identified a very closely related HERV sequence on chromosome region 12q22 (GenBank accession number AC010203); it possesses sequences similar to those of the *gag* and *env* genes, is flanked by two LTRs, and shows similarities to multiple ESTs. This HERV element is transcriptionally active and could thus be the ''master'' gene of the *INSL4*-HERV element.

We postulate that the placenta-specific expression of the *INSL4* gene is mediated by the 3′ LTR of the HERV element. Several published findings support this hypothesis. First, several HERV families, especially HERV-K, HERV-W, HERV-F, HERV-R (ERV-3), and HERV-H (RTLV-H), are mainly expressed in the placenta [16, 25–29]. In addition, Mi et al. [16] and Blond et al. [17] identified a HERV-W coding for an envelope protein (syncytin) that has fusogenic activity and therefore could be involved in syncytiotrophoblast formation. Second, alternative chimeric transcripts with placenta-specific expression can be created by HERV insertion into various gene promoters (the pleiotrophin and endothelin B receptor genes) [30, 31]. Third, isolated HERV LTR possess bidirectional promoter activity as well as both young LTR (emergence in the human genome 5 million years BP) and old LTR (emergence in the human genome 45 million years BP) show similar promoter activity [32]. Fourth, HERV LTRs contain elements that can regulate the transcription of neighboring genes. [18, 33]. For example, Bi et al. [18] identified a MER11 repetitive element in the *Leptin* promoter that mediates its placental expression. Because the MER11 elements are not present in the murine genome, this might explain why the human but not the murine placenta expresses leptin. We found that *INSL4* is the only member of the insulin-like growth factor gene cluster at 9p23.3–p24.1 to be specifically expressed in the placenta, and the only gene to have a HERV element inserted in its promoter.

Our results strongly support the conclusion that the 3′ LTR promotes human *INSL4* expression in JEG-3 cells. Indeed, transfection experiments demonstrated that the LTR was capable of directing luciferase expression in JEG-3 cells, and deletion of the retrovirus-derived promoter portion abolished its activity. HERVs (or isolated LTRs) may be present in the 5′ flanking region of other human genes with placenta-specific expression.

Several reports suggest a role for HERV in the creation of the syncytiotrophoblast layer [16, 26, 27]. HERV could mediate neighboring gene function. We studied *INSL4* expression in an in vitro model of human villous cytotrophoblast differentiation into the syncytiotrophoblast and found that expression of *INSL4* (but not that of *INSL6, RLN1,* or *RLN2*) was up regulated more than 10-fold in the syncytiotrophoblast relative to the corresponding cytotrophoblasts; in the same way, the *Leptin* gene, the placental expression of which is under control of a MER11 repeat element [18], was up-regulated more than 40-fold.

In conclusion, our results confirm that ancient retroviral infection may have been a major event in primate evolution, especially the functional evolution of the human placenta. The human hemochorial placenta is considered more invasive than the mouse counterpart in which trophoblast and maternal cells are rather demarcated from each other [34]. Some genes (such as *INSL4, Leptin, Pleiotrophin, Endothelin B receptor,* and *Syncytin*) with new placenta-specific expression in primate species via HERV element integration could contribute to the aggressive growth phenotype of the normal primate placenta.

ACKNOWLEDGMENT

We are indebted to Dr. Eric Denamur (Laboratoire d'études de génétique bactérienne dans les infections de l'enfant, EA3105, Hôpital Robert Debré, Paris, France) for kindly providing *C. goeldii* tissue specimen.

REFERENCES

1. Le Roith D. Insulin-like growth factor. N Engl J Med 1997; 336:633–640.
2. Conklin D, Lofton-Day CE, Haldeman BA, Ching A, Whitmore TE, Lok S, Jaspers S. Identification of INSL5, a new member of the insulin superfamily. Genomics 1999; 60:50–56.
3. Lok S, Johnston DS, Conklin D, Lofton-Day CE, Adams RL, Jelmberg AC, Whitmore TE, Schrader S, Griswold MD, Jaspers SR. Identification of INSL6, a new member of the insulin family that is expressed in the testis of the human and rat. Biol Reprod 2000; 62: 1593–1599.
4. Bathgate RA, Samuel CS, Burazin TCD, Layfield S, Claasz AA, Reytomas IGT, Dawson NF, Zhao C, Bond C, Summers RJ, Parry LJ, Wade JD, Tregear GW. Human relaxin gene 3 (H3) and the equivalent mouse relaxin (M3) gene. J Biol Chem 2002; 277:1148–1157.
5. Veitia R, Laurent A, Quintana-Murci L, Ottolenghi C, Fellous M, Vidaud M, McElreavey K. The INSL4 gene maps close to WI-5527 at 9p24.1–p23.3 clustered with two relaxin genes and outside the critical region for the monosomy 9p syndrome. Cytogenet Cell Genet 1998; 81:275–277.
6. Sankoff D. Gene and genome duplication. Curr Opin Genet Dev 2001; 11:681–684.
7. Gunnersen JM, Fu P, Roche PJ, Tregear GW. Expression of human relaxin genes: characterization of a novel alternatively-spliced human relaxin mRNA species. Mol Cell Endocrinol 1996; 118:85–94.
8. Chassin D, Laurent A, Janneau J-L, Berger R, Bellet D. Cloning of a new member of the insulin gene superfamily (INSL4) expressed in human placenta. Genomics 1995; 29:465–470.
9. Laurent A, Rouillac C, Delezoide A-L, Giovangrandi Y, Vekemans M, Bellet D, Abitbol M, Vidaud M. Insulin-like 4 (INSL4) gene expression in human embryonic and trophoblastic tissues. Mol Reprod Dev 1998; 51:123–129.
10. Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
11. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 1987; 162:156–159.
12. Bièche I, Parfait B, Le Doussal V, Olivi M, Rio M-C, Lidereau R, Vidaud M. Identification of CGA as a novel estrogen receptor-responsive gene in breast cancer: an outstanding candidate marker to predict the response to endocrine therapy. Cancer Res 2001; 61:1652–1658.
13. Frendo JL, Therond P, Bird T, Massin N, Muller F, Guibourdenche J, Luton D, Vidaud M, Anderson WB, Evain-Brion D. Overexpression of copper zinc superoxide dismutase impairs human trophoblast cell fusion and differentiation. Endocrinology 2001; 142:3638–3648.
14. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389–3402.
15. Tassabehji M, Strachan T, Anderson M, Campbell RD, Collier S, Lako M. Identification of a novel family of human endogenous retrovirus and characterization of one family member, HERV-K(C4), located in the complement C4 gene cluster. Nucleic Acids Res 1994; 22:5211–5217.
16. Mi S, Lee X, Li X-P, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang X-Y, Edouard P, Howes S, Keith JC, McCoy JM. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 2000; 403:785–789.
17. Blond J-L, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, Mandrand B, Mallet F, Cosset FL. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placental and fuses cells expressing the type D mammalian retrovirus receptor. J Virol 2000; 74:3321–3329.
18. Bi S, Gavrilova O, Gong D-W, Mason MM, Reitman M. Identification

of a placental enhancer for the human leptin gene. J Biol Chem 1997; 272:30583–30588.
19. Evans BA, John M, Fowler KJ, Summers RJ, Cronk M, Shine J, Tregear G. The mouse relaxin gene: nucleotide sequence and expression. J Mol Endocrinol 1993; 10:15–23.
20. Kasik J, Muglia L, Stephan DA, Menon RK. Identification, chromosomal mapping, and partial characterization of mouse INSL6: a new member of the insulin family. Endocrinology 2000; 141:458–461.
21. Evans BA, Fu P, Tregear GW. Characterization of two relaxin genes in the chimpanzee. J Endocrinol 1994; 140:385–392.
22. Duret L, Guex N, Peitsch MC, Bairoch A. New insulin-like proteins with atypical disulfide bond pattern characterized in *Caenorhabditis elegans* by comparative sequence analysis and homology modeling. Genome Res 1998; 8:348–353.
23. Medstrand P, Mager DL. Human-specific integrations of the HERV-K endogenous retrovirus family. J Virol 1998; 72:9782–9787.
24. Tönjes RR, Czauderna F, Kurth R. Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K. J Virol 1999; 73:9187–9195.
25. Johansen T, Holm T, Bjorklid E. Members of the RTVL-H family of human endogenous retrovirus-elements are expressed in placenta. Gene 1989; 79:259–267.
26. Boyd MT, Bax CMR, Bax BE, Bloxam DL, Weiss RA. The human endogenous retrovirus ERV-3 is upregulated in differentiating placental trophoblast cells. Virology 1993; 196:905–909.
27. Venables PJ, Brookes SM, Griffiths D, Weiss RA, Boyd MT. Abundance of an endogenous retroviral envelope protein in placenta trophoblasts suggests a biological function. Virology 1995; 211:589–592.
28. Blond J-L, Besème F, Duret L, Bouton O, Bedin F, Perron H, Mandrand B, Mallet F. Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. J Virol 1999; 73:1175–1185.
29. Kjellman C, Sjogren HO, Salford LG, Widegren B. HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. Gene 1999; 239:99–107.
30. Schulte AM, Lai S, Kurtz A, Czubayko F, Riegel T, Wellstein A. Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. Proc Natl Acad Sci U S A 1996; 93:14759–14764.
31. Medstrand P, Landry J-R, Mager DL. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. J Biol Chem 2001; 276:1896–1903.
32. Domansky AN, Kopantzev EP, Snezhkov EV, Lebedev YB, Leib-Mosch C, Sverdlov ED. Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. FEBS Lett 2000; 472:191–195.
33. Di Cristofano A, Strazullo M, Longo L, La Mantia G. Characterization and genomic mapping of the ZNF80 locus: expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. Nucleic Acids Res 1995; 23:2823–2830.
34. Graham CH, Lala PK. Mechanisms of placental invasion of the uterus and their control. Biochem Cell Biol 1992; 70:867–874.