



(51) Internationale Patentklassifikation ⁷ : C12N 15/867, 5/10, 7/01, A61K 48/00	A2	(11) Internationale Veröffentlichungsnummer: WO 00/53789 (43) Internationales Veröffentlichungsdatum: 14. September 2000 (14.09.00)
(21) Internationales Aktenzeichen: PCT/EP00/02064 (22) Internationales Anmeldedatum: 9. März 2000 (09.03.00) (30) Prioritätsdaten: 199 10 650.9 10. März 1999 (10.03.99) DE (71) Anmelder (für alle Bestimmungsstaaten ausser US): GSF-FORSCHUNGSZENTRUM FÜR UMWELT UND GESUNDHEIT GMBH [DE/DE]; Ingolstädter Landstrasse 1, D-85764 Neuherberg (DE). (72) Erfinder; und (75) Erfinder/Anmelder (nur für US): LEIB-MÖSCH, Christine [DE/DE]; Nadistrasse 23, D-80809 München (DE). SCHÖN, Ulrike [DE/DE]; Dachauer Strasse 73, D-80335 München (DE). BAUST, Corinna [DE/DE]; Schlossweg 30a, D-69190 Waldorf (DE). (74) Anwalt: BEHNISCH, Werner; Reinhard-Skuhra-Weise & Partner GbR, Friedrichstrasse 31, Postfach 44 01 51, D-80750 München (DE).	(81) Bestimmungsstaaten: JP, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i>	
(54) Title: RETROVIRAL EXPRESSION VECTORS ON THE BASIS OF HERV- LONG TERMINAL REPEAT SEQUENCES		
(54) Bezeichnung: RETROVIRALE EXPRESSIONSVEKTOREN AUF DER BASIS VON HERV-LTR-SEQUENZEN		
(57) Abstract		
<p>The invention relates to retroviral expression vectors with cell-specifically modulatable promoters. The vectors can be used, for example, for the cell-specific expression of therapeutically valuable genes in gene therapy. The invention specifically relates to retroviral expression vectors containing at least the following elements, in a functional configuration: a) DNA sequences for the packaging of the vector RNA and for the cell-specific expression of proteins or peptides which are coded by heterologous DNA nucleotide sequences; b) one or more DNA nucleotide sequences coding for a protein or a peptide and characterized in that for cell-specific expression said DNA sequences contain a cell-specifically modulatable promotor region of a human endogenous retroviral DNA nucleotide sequence (HERV).</p>		
(57) Zusammenfassung		
<p>Die vorliegende Erfindung betrifft retrovirale Expressionsvektoren mit zellspezifisch regulierbaren Promotoren. Die Vektoren sind beispielsweise zur zellspezifischen Expression therapeutisch wertvoller Gene im Rahmen einer Gentherapie einsetzbar. Die Erfindung beschreibt retrovirale Expressionsvektoren, enthaltend zumindest die nachfolgenden Elemente in funktioneller Anordnung: a) DNA-Sequenzen zur Verpackung der Vektor-RNA und zur zellspezifischen Expression von Proteinen oder Peptiden, die von heterologen DNA-Nukleotidsequenzen kodiert werden; b) eine oder mehrere für ein Protein oder Peptid kodierende DNA-Nukleotidsequenzen, dadurch gekennzeichnet, daß die DNA-Sequenzen zur zellspezifischen Expression eine zellspezifisch regulierbare Promotorregion aus einer humanen endogenen retroviralen DNA-Nukleotidsequenz (HERV) enthalten.</p>		

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidsschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Retrovirale Expressionsvektoren auf der Basis von HERV-LTR-Sequenzen

Die vorliegende Erfindung betrifft retrovirale Expressionsvektoren mit zellspezifisch regulierbaren Promotoren. Die Vektoren sind beispielsweise zur zellspezifischen Expression therapeutisch wertvoller Gene im Rahmen einer Gentherapie einsetzbar.

Retroviren sind RNA-Viren, bei denen die viralen Gene von einem einzelsträngigen RNA-Molekül kodiert werden. Nach Eindringen der Viren in die Zelle wird die virale RNA in ein doppelsträngiges DNA Molekül durch die reverse Transkription umgewandelt. Die DNA dringt in den Kern ein und integriert in das zelluläre Chromosom. Die integrierte virale DNA-Form, das sogenannte Provirus, bildet die Matrize zur Expression der viralen Gene.

Die Integration des viralen Genoms in das zelluläre Chromosom ist ein obligatorischer Teil der viralen Replikation und wird durch viral kodierte Enzyme vermittelt. Es scheint, daß, mit wenigen Ausnahmen, das Vorliegen des retroviralen Genoms in der Zelle, die Expression seiner Gene und die Bildung von Viruspartikeln die Lebensfähigkeit der infizierten Zelle nicht oder kaum beeinträchtigt.

Retroviraler Gentransfer wird dazu benutzt, funktionelle Gene, insbesondere therapeutisch wertvolle Gene, in die Zellen einzuführen, ohne die Proliferationsfähigkeit der Wirtszelle zu beeinflussen. Aufgrund ihres Replikationsmodus eignen sich Retroviren für eine derartigen Gentransfer. In der einfachsten Ausgestaltungsform wird zumindest ein Teil der viralen Gene durch ein Gen von Interesse ersetzt und, unter Zuhilfenahme des effizienten viralen Infektionsprozesses, wird dieses Gen von Interesse in die Zielzelle transferiert.

Retrovirale Vektoren eignen sich deshalb zur Gentherapie, weil die Infektion durch Retroviren hocheffizient verläuft und die retroviralen Vektoren so modifizierbar sind, daß sie heterologe DNA aufnehmen und stabil in das Wirtszellgenom integrieren können. Eine Vielzahl von retroviralen Vektoren wurde in den letzten Jahren entwickelt, und nur beispielsweise wird hier auf die Übersichtsartikel von Günzburg et al. (1996) und Robbins et al. (1998) hingewiesen.

Eine mögliche bevorzugte Ausgestaltungsform für retrovirale Vektoren sind sogenannte Pro-Con-Vektoren, die zum ersten Mal in der WO 96/07748 beschrieben wurden. Zur Offenbarung wird vollständig auf diese Druckschrift Bezug genommen.

ProCon-Vektoren tragen heterologe Promotor- und wahlweise weitere Regulationselemente in der 3'LTR, die, nach Infektion, dupliziert und an die 5'LTR in der Zielzelle transloziert werden und die befähigt sind, die Expression von Markergenen oder therapeutischen Genen zu regulieren. Diese heterologen Gene sind nicht direkt mit dem Promotor verbunden, sondern werden in das Innere des Vektors inseriert.

ProCon-Vektoren umfassen einen 5'LTR-Abschnitt der Struktur U3, R, U5 sowie zumindest eine kodierende und/oder nicht-kodierende Sequenz sowie einen 3'LTR-Bereich, der einen vollständig oder teilweise deletierten U3-Abschnitt umfaßt, wobei der deletierte U3-Abschnitt durch eine Polylinker-Sequenz ersetzt wurde, worauf sich der R- und U5-Abschnitt anschließt.

Die Vermehrung dieser Vektoren erfolgt mit Hilfe einer Helferzelllinie, die große Mengen der viralen Proteine produziert, die vom Expressionsvektor selbst nicht mehr synthetisiert werden. Die Helferzelllinie ist jedoch nicht mehr in der Lage, ein replikationskompetentes Virus zu produzieren. Diese Zelllinie wird auch als Verpackungszelllinie bezeichnet und umfaßt eine mit mindestens einem zweiten Plasmid transfizierte Zelllinie, das diejenigen Gene trägt, die eine Verpackung des modifizierten retroviralen Vektors ermöglichen. Wir verweisen hier z.B. auf die W092/10564, auf die hier vollinhaltlich Bezug genommen wird.

Die für das modifizierte Retrovirus (Expressionsvektor) kodierende DNA wird in die Verpackungszelllinie transfiziert. Unter diesen Bedingungen wird das modifizierte retrovirale Genom einschließlich der inserierten therapeutischen Gene bzw. Markergene transkribiert und in die retrovirale Partikel verpackt (rekombinate virale Partikel). Dieses rekombinante Virus wird dann zur Infektion der Zielzellen eingesetzt; das Genom des modifizierten Retrovirus, d.h. des Expressionsvektors, wird in das Genom der Zielzellen integriert, und zwar zusammen mit den Markergenen bzw. den therapeutischen Genen. Eine mit den so erzeugten rekombinanten viralen Teilchen infizierte Zelle kann ein neues Vektorvirus nicht mehr produzieren, da in diesen Zellen keine weiteren viralen Proteine mehr vorliegen. Die in die Wirts -

zelle integrierte DNA des Expressionsvektors mit den therapeutisch wertvollen Genen bzw. den Markergenen liegt in der zellulären DNA integriert vor und kann nunmehr in der Zelle exprimiert werden.

Therapeutisch wertvolle Gene, die über derartige retrovirale Expressionsvektoren zur Expression gebracht werden, sollen bevorzugt in zell- und gewebespezifischer Form exprimiert werden. Hierzu werden zellspezifische Regulationssequenzen in der LTR-Sequenz des Expressionsvektors eingesetzt. Diese zellspezifischen Regulationssequenzen umfassen beispielsweise zellspezifisch regulierbare Promotorabschnitte, zellspezifische Enhancersequenzen sowie Transkriptionsfaktor-Bindungsstellen. Die Promotoren liegen im U3-Abschnitt der LTR.

Zur Zeit werden in retrovirale Vektoren zelluläre Promotorsequenzen oder Promotorsequenzen exogener Retroviren eingesetzt.

Zelluläre Promotoren sind häufig auf zusätzliche Signalstrukturen angewiesen, die in großer Distanz stromaufwärts oder stromabwärts des Promotors liegen können. Es hat sich deshalb immer wieder als schwierig erwiesen, starke, gewebespezifische zelluläre Promotorsequenzen zu isolieren und in retrovirale Vektoren zu klonieren. Promotoren exogener Retroviren haben den Vorteil, daß sie in der retroviralen LTR auf kleinem Raum alle benötigten regulativen Elemente enthalten und daher weitgehend unabhängig von den benachbarten DNA-Sequenzen des Integrationsortes transkribiert werden. Ein schwerwiegender Nachteil ist jedoch, daß sie zwar sehr stark, aber nicht gewebespezifisch sind und in der Regel in allen Zelltypen gleich stark exprimiert werden.

Es ist deshalb eine Aufgabe der vorliegenden Erfindung, neue retrovirale Expressionsvektoren bereitzustellen, die zwar die Vorteile der retroviralen Promotoren nutzen, alle für die Transkription erforderlichen Signalstrukturen auf engem Raum innerhalb der U3 und R-Region zu konzentrieren, die damit verbundenen Nachteile jedoch vermeiden.

Diese Aufgabe wird erfindungsgemäß dadurch gelöst, daß in retrovirale Expressionsvektoren ein zellspezifisch regulierbarer Promotorabschnitt aus einer humanen endogenen retroviralen DNA-Nukleotidsequenz (HERV) eingesetzt wird. Diese Promotorsequenzen humaner endogener retroviraler Viren liegen bereits in der Wirtszelle vor, und es wurde erfindungs-

gemäß gefunden, daß sie sich in hervorragender Weise zur Regulation der zellspezifischen Expression von Markergenen und therapeutisch wertvollen Genen eignen.

Endogene Retroviren (ERV) sind im Genom aller Zellen eines Organismus zu finden. Sie werden vertikal über Keimbahnzellen übertragen und können durch Umwelteinflüsse reaktiviert werden. Das menschliche Genom besteht zu etwa 2% aus endogenen Retroviren und retroviralen Sequenzen, wobei solitäre HERV-LTRs mit 20000 - 40000 Kopien pro Genom vertreten sind (Tab. 1) (Leib-Mösch et al., 1993; Wilkinson et al., 1994, Patience et al., 1997).

Da HERV-Sequenzen bereits vor 30 - 40 Millionen Jahren in das Genom von Primaten integriert sind, kann davon ausgegangen werden, daß im Verlauf der Evolution die meisten pathogenen Sequenzen durch Mutationen und Rearrangements aus dem Provirus eliminiert bzw. so verändert wurden, daß sie für den Organismus nicht mehr von Nachteil sind. Retrovirale Vektoren, die aus derartigen Sequenzen konstruiert werden, haben gegenüber Vektoren aus animalen Viren den Vorteil, daß keine neuen viralen Sequenzen in das Genom eingebracht werden müssen. Auch durch Rekombination mit bereits im Genom vorhandenen HERV-Sequenzen können keine neuartigen Retroviren entstehen, wie das z.B. der Fall wäre, wenn Retroviren anderer Spezies als Vektoren eingesetzt werden. Aus diesem Grund kann die Verwendung dieser Sequenzen zur Konstruktion retroviraler Vektoren das Sicherheitsrisiko minimieren. Desweiteren könnten im Genom enthaltene homologe Regionen für eine gewebespezifische Integration der retroviralen Vektoren in bestimmte Stellen eines Chromosoms genutzt werden.

HERV-Elemente haben während der Evolution eine Reihe zellulärer Funktionen übernommen. So werden Promotor- und Enhancer-Elemente von HERV-LTRs zur Steuerung der Transkription zellulärer Gene benutzt (Kato et al., Feuchter-Murthy et al., 1993; Di Christofano et al., 1995). Ein Beispiel für die Verwendung von LTR-Regulationselementen zur gewebespezifischen Expression eines zellulären Gens ist das Amylase-Gen des Menschen. Dieses Gen wird durch die LTR eines HERV-E Elements kontrolliert und dadurch spezifisch nur in den Speicheldrüsen exprimiert (Ting et al., 1992). Darüber hinaus haben Schulte und Mitarbeiter (1996) gezeigt, daß die Insertion eines endogenen Retrovirus in die 5' nicht translatierte Region des Pleiotrophin-Gens für dessen Trophoblasten-spezifische Aktivität verant-

wortlich ist (Schulte et al., 1996). In anderen Fällen können polyA-Signale von HERV-LTRs auch dazu dienen, zelluläre Transkripte zu polyadenylieren (Mager, 1989; Goodchild et al., 1992).

Der große Vorteil beim Einsatz retroviraler Promotoren liegt darin, daß, wie bereits oben ausgeführt, alle für die Transkription erforderlichen Signalstrukturen auf engem Raum innerhalb des U3-Bereichs und des R-Bereichs der LTR lokalisiert sind, da Retroviren möglichst unabhängig von der Umgebung ihres Integrationsortes transkriptionsaktiv bleiben müssen. Da diese HERV-Promotoren seit Millionen Jahren im Primatengenom persistieren, haben sie sich im Laufe der Evolution so angepaßt, daß sie wie zelluläre Promotoren zelltyp-spezifisch aktiv sind, und somit die Vorteile zellulärer und retroviraler Promotoren in sich vereinen.

Die Transkription der HERVs wird von einem klassischen RNA Polymerasell-Promotor aus gestartet (Wilkinson et al., 1994). Dieser Promotor ist innerhalb der LTR-Region lokalisiert. Das HERV-Transkript besteht daher nicht aus der vollständigen Kopie des Provirus. Um den Verlust der Transkriptionskontrollelemente zu kompensieren, haben diese Elemente den Mechanismus der reversen Transkription entwickelt, mit dem die verlorengegangenen Sequenzen an beiden Seiten der Elemente regeneriert werden, woraus wiederum die LTRs entstehen. Neben Promotorsequenzen enthalten HERV-LTRs noch eine Vielzahl verschiedener Bindungsstellen für Transkriptionsfaktoren (Seifarth et al., 1998), die für die Gewebespezifität der Expression verantwortlich sind.

Obwohl eine gewisse strukturelle Ähnlichkeit zwischen HERVs und exogenen animalen Retroviren wie MLV oder MMTV besteht, wurden HERV-Sequenzen und insbesondere Promotoren von HERVs nie als mögliche Kandidaten für die Entwicklung retroviraler Expressionsvektoren in Betracht gezogen. Sie wurden, ganz im Gegenteil, im Zusammenhang mit einer Gentherapie bisher ausschließlich als Störfaktoren betrachtet (Patience et al., 1997). Es wurde befürchtet, daß sie in der Zielzelle durch Rekombinationen aufgrund von Sequenzhomologien mit dem therapeutischen Vektor interferieren könnten. Diese Befürchtungen konnten bislang zwar nicht experimentell bestätigt werden, jedoch mit der Entwicklung sehr effizienter humaner Verpackungszelllinien tauchte das Problem des Mitverpackens und der ungewollten Übertragung möglicherweise infektiöser HERV-Sequenzen auf. Daher wurde die mögliche Verpackung exprimierter HERV-Sequenzen in Virionen, die auf MLV basieren,

ausführlich untersucht. Patience et al. (1998) identifizierten mRNA-Transkripte mehrerer verschiedener HERV-Familien wie HERV-K und HERV-H in humanen Verpackungszelllinien. Jedoch konnten selbst mit einem hochsensitiven RT-PCR Test keine dieser Sequenzen in den von den Zellen freigesetzten MLV-Vektor-Partikeln nachgewiesen werden.

Nach diesen Befunden schien eine Verpackung und Übertragung von HERV-Sequenzen und damit schließlich auch HERV-basierender Vektoren in MLV-Verpackungssystemen ausgeschlossen zu sein. Zwar weisen HERV-Gene eine Sequenzhomologie zu MLV-Genen von 50 - 65 % auf, jedoch besitzen gerade die für die Verpackung und Infektion essentiellen Bereiche, insbesondere das zwischen der 5' LTR und der gag-Region lokalisierte Verpackungssignal und die LTR selbst, keine erkennbare Sequenzhomologie zu den entsprechenden MLV-Sequenzen. Effiziente HERV-Verpackungssysteme hingegen sind zur Zeit ebenfalls nicht vorstellbar, da bislang keine Zelllinien bekannt sind, die in ausreichenden Mengen HERV-Partikel produzieren.

Die vorliegende Erfindung löst somit das Problem von retroviralen Expressionsvektoren, die eine zell- und gewebespezifische Expression von Fremdgenen (Gen von Interesse) steuern, dadurch, daß Expressionsvektoren bereitgestellt werden, die, in funktioneller Anordnung, zumindest die nachfolgenden Elemente enthalten:

- a) DNA-Sequenzen zur Verpackung der Vektor-RNA und zur zellspezifischen Expression von Proteinen oder Peptiden, die von heterologen DNA-Nukleotidsequenzen kodiert werden;
- b) ein oder mehrere für ein Protein oder Peptid kodierende heterologe DNA-Nukleotidsequenzen (Transkriptionseinheit);
die DNA-Sequenzen zur zellspezifischen Expression sind dadurch charakterisiert, daß sie eine zellspezifisch regulierbare Promotorregion umfassen, die aus einem humanen endogenen retroviralen Virus, insbesondere der LTR-Sequenz dieses Virus, stammen;

Der Promotorbereich aus einer HERV-Sequenz kann den gesamten LTR-Bereich des HERV umfassen. In weiteren Ausführungsformen der Erfindung umfaßt der Promotorbereich jedoch nur den U3-Bereich oder den R-U3-Bereich der HERV-LTR. In einer weiteren bevorzugten Ausführungsform der Erfindung wird neben diesen Bereichen zusätzlich der nicht translatierte Bereich zwischen der 5' LTR und den gag-Genen vom Promotorabschnitt mit umfaßt.

Erfindungsgemäß wurde nämlich festgestellt, daß auch in diesem Bereich Sequenzen lokalisiert sind, die eine zellspezifische Expression von Proteinen bzw. Peptiden steuern, d.h. für diese zellspezifische Expression zumindest mit verantwortlich sind.

Die Promotoren sind Teilabschnitte der DNA, die für den Start der Transkription der zugeordneten Strukturgene notwendig sind. Der Promotor beinhaltet den Startpunkt der Transkription, die Erkennungs- und Bindungsstelle für die RNA-Polymerase. Der Promotor kann auch weitere Sequenzen umfassen, an die regulatorische Proteine binden können und dadurch spezifisch die Initiation der Transkription regulieren. Beispiele für derartige Proteine sind Transkriptionsfaktoren und Repressoren. Beispiele für diese Regulationselemente der Transkriptionsaktivität sind die CAAT-Box, die GC-Box und die TATA-Box. Die Promotoren werden von einer Polymerase des Typs II erkannt.

Die Promotorabschnitte zur zellspezifischen Expression von Fremdproteinen aus HERVs können wahlweise mit weiteren Sequenzen kombiniert werden, die aus exogenen Retroviren stammen und die zellspezifische Expression fördern. Auch eine Kombination mit Regulationssequenzen aus zellulären Genen ist denkbar, um die zellspezifische Expression zu unterstützen.

Der erfindungsgemäße retrovirale Expressionsvektor enthält weiterhin zumindest DNA-Sequenzen zur Verpackung des Vektors durch eine Verpackungs-Helferzelllinie. Die DNA-Sequenzen zur Verpackung sind zwischen der 5' LTR und dem gag-Gen lokalisiert. Derartige Verpackungssignale sind in allen retroviralen Vektoren vorhanden und sind dem Fachmann deshalb bekannt. Beispiele für Verpackungssequenzen sind aufgeführt in Mann et al., 1985, sowie Rein, 1994, sowie der dort zitierten Literatur. Auf diese Literatur wird vollinhaltlich Bezug genommen.

Der retrovirale Expressionsvektor der vorliegenden Erfindung enthält ein oder mehrere Transkriptionseinheiten, die für eine Aminosäuresequenz kodieren. Die Aminosäuresequenz steht für ein Protein oder Peptid. Es können beliebige Sequenzen in den Expressionsvektor inseriert werden, die für ein beliebiges Protein oder Peptid von Interesse kodieren. Derartige Proteine oder Peptide können beispielsweise von Markergenen, therapeutisch wertvollen Genen, antiviral wirkenden Genen, Antitumorgenen und/oder Zytokingenen kodiert werden. Diese Liste ließe sich beliebig ergänzen. Die in den retroviralen Expressionsvektor einsetz-

baren Gene sind dem Fachmann bekannt. Die Art der eingesetzten Gene hängt vom Anwendungszweck des erfindungsgemäßen Vektors ab.

Die erfindungsgemäßen Vektoren können beispielsweise in der Gentherapie eingesetzt werden, um heterologe DNA in Targetzellen zu transferieren, um Krankheiten einer spezifischen Therapie zugänglich zu machen. Die Vektor-DNA wird in die ausgewählten Targetzellen derart eingeführt, daß die heterologe DNA in der Zielzelle exprimiert und das durch die DNA kodierte Produkt produziert wird. Hierunter fallen insbesondere solche Gene zur Expression von Proteinen, die in der Zielzelle nicht oder nicht mehr oder in nicht ausreichender Menge produziert werden, so daß sich ein Krankheitszustand einstellt. Die Erfindung umfaßt nicht nur solche Proteine bzw. Peptide, die natürlicherweise vorkommen, sondern auch solche, die in einer Weise abgeändert wurden, daß ein gewünschter Effekt erreicht wird, beispielsweise eine höhere Aktivität eines Enzyms, die Blockierung der Bindungsstelle von Viren, Zerstörung von Tumorzellen durch Suizidgene usw.

Bei den für ein Protein oder ein Peptid kodierenden DNA-Nukleotidsequenzen handelt es sich im allgemeinen um heterologe DNA, die für RNA und Proteine kodiert, die normalerweise in vivo von der Zelle, in der die Proteine bzw. Peptide exprimiert werden, nicht produziert werden. Sie kann auch als Fremd-DNA bezeichnet werden. Beliebige Proteine, beispielsweise Enzyme, Hormone und Antikörper, fallen hierunter. Die erfindungsgemäß bereitgestellten retroviralen Expressionsvektoren sind deshalb so ausgelegt, daß sie in menschlichen Zellen Proteine von Interesse exprimieren können.

Die erfindungsgemäß eingesetzten Promotorbereiche werden aus HERV-Sequenzen ausgewählt, die aus den bekannten HERV-Familien stammen. Beispiele hierfür sind HERV-K, HERV-H, HERV-E, HERV-L, HERV-T, HERV-R, HERV-I, HERV-P, ERV9, HERV-W.

Selbstverständlich können auch andere, zur Zeit noch nicht bekannte HERV-Familien gescreent werden, um noch unbekanntes, die zellspezifische Expression steuernde Promotorsequenzen aufzufinden.

Erfindungsgemäß bevorzugte LTR-Sequenzen aus HERVs, die zur gewebespezifischen Expression von Proteinen und Peptiden von Interesse einsetzbar sind, werden im Anhang

offenbart. Sie sind in retrovirale Expressionsvektoren einsetzbar, um die erfindungsgemäß gestellte Aufgabe zu lösen. Selbstverständlich können durch an sich bekannte Methoden auch nur Teile dieser LTRs ausgewählt werden, um die in den Vektor eingesetzten Sequenzen so klein wie möglich zu halten. Mit Hilfe verschiedener Deletionsmutanten können die geeigneten Fragmente ausgewählt werden. Auch weitere Abänderungen dieser LTR-Sequenzen sind möglich, z.B. Punktmutationen, Insertionen, Additionen, Austausch mehrerer Nucleotide etc., um die Effizienz der gewebespezifischen Expression zu steigern und an die gewünschte Funktion anzupassen.

In einer bevorzugten Ausführungsform der Erfindung werden die eingangs beschriebenen ProCon-Vektoren eingesetzt. Derartige ProCon-Vektoren umfassen einen 5' LTR-Abschnitt der Struktur U3-R-U5, ein oder mehrere Sequenzen, die für ein Protein oder Peptid kodieren, und wahlweise nicht kodierenden Sequenzen, und einen 3' LTR-Abschnitt, umfassend einen teilweise oder vollständig deletierten U3-Abschnitt, wobei der deletierte U3-Abschnitt zumindest die erfindungsgemäß eingesetzten HERV-LTR-Sequenzen umfaßt, gefolgt vom R-U5-Abschnitt. Nähere Einzelheiten sind beispielsweise in der WO96/07748 und der WO96/28564 beschrieben. Auf diese Schriften wird hier vollinhaltlich Bezug genommen.

Erfindungsgemäß wurde eine Strategie entwickelt, um zellspezifisch wirkende Promotorsequenzen aufzuspüren. Diese Strategie wird in der nachfolgenden Beschreibung näher erläutert. Selbstverständlich sind grundsätzlich auch andere Methoden zum Auffinden von zellspezifisch wirksamen HERV-LTR-Sequenzen denkbar und anwendbar. Die Erfindung ist somit nicht auf die nachfolgenden Ausführungsbeispiele beschränkt.

Die erfindungsgemäßen retroviralen Expressionsvektoren sind verpackungsdefizient, d.h. sie sind nicht in der Lage, ohne Hilfe einer Verpackungs-Helferzelllinie Viruspartikel zu produzieren. Die Erfindung umfaßt deshalb auch ein retrovirales Vektorsystem, das einen retroviralen Expressionsvektor, wie er in der vorliegenden Erfindung beschrieben wird, und eine Verpackungszelllinie mit zumindest einem retroviralen oder rekombinanten retroviralen Konstrukt enthält, das für die Verpackungsproteine des retroviralen Expressionsvektors kodiert. Derartige Verpackungszelllinien sind an sich bekannt und beschrieben. Beispielsweise wird hier auf die murine Verpackungszelllinie PA317 (Saller et al., 1998) hingewiesen.

Nachfolgend wird die Erfindung zunächst allgemein und dann anhand von Ausführungsbeispielen beschrieben.

Erfindungsgemäß wurde die Eignung humaner endogener Retroviren zur Entwicklung gewebespezifischer Vektoren für die Gentherapie untersucht. Dazu wurde zunächst die Gewebespezifität der HERV pol Transkription in verschiedenen Zelllinien wie T-Zellen, Keratinozyten und Brustkrebs-Zellen in einem "Reversen Dot Blot"-Verfahren überprüft. Die Expressionsmuster der verschiedenen HERV-Familien erwiesen sich dabei durchwegs als Zelltyp-abhängig. Zur Isolierung der transkriptionsaktiven HERV-LTRs aus verschiedenen Zelllinien und Geweben wurden Primer entwickelt, mit denen spezifisch die U3/R-Regionen aus mRNA-Präparationen amplifiziert werden können. Die isolierten LTR-Sequenzen sowie einzelne Vertreter bereits bekannter LTRs wurden in Expressionsvektoren eingebaut. Die Aktivität der LTR-Promotoren wurde nach transienter Transfektion der Reporterplasmide über Luziferaseaktivität bzw. über eGFP-Fluoreszenz in verschiedenen Zelllinien getestet. Es zeigte sich, daß die Promotoraktivitäten der einzelnen HERV-LTRs deutlich in Abhängigkeit von der getesteten Zelllinie variieren. Die Promotorregion einer HERV-H LTR, die aus Astrozyten und Leberzellen isoliert worden war und sich in mehreren Versuchen als besonders aktiv in Lungenfibroblastenzellen (LC5) erwiesen hatte, wurden in zwei retrovirale Promotorkonversionsvektoren (pLESN und PLX) eingebaut, in Verpackungszelllinien getestet, die Verpackungseffizienz untersucht und überprüft, ob nach Infektion der Zielzellen eine Promotorkonversion stattgefunden hat. Zum Nachweis der Transkriptionsaktivität in den Zielzellen wurden FACS-Analysen durchgeführt.

Es wurde somit eine Methode beschrieben, mit der HERV Promotorsequenzen (U3/R-Region), die eine gewebespezifische Expression vermitteln, identifiziert und isoliert werden können. Die Gewebespezifität und Promotoraktivität dieser Sequenzen wurde dann in einem transienten Transfektions-Assay in verschiedenen humanen Zelllinien getestet. Schließlich wurden geeignete Sequenzen ausgewählt, in einen Promotorkonversionsvektor (ProCon-Vektor) kloniert und dabei ihre Eignung zur Konstruktion gewebespezifischer Vektoren für die Gentherapie überprüft. Die Herstellung der erfindungsgemäßen retroviralen Expressionsvektoren erfolgt durch an sich bekannte rekombinante Techniken. Derartige Techniken sind beispielsweise beschrieben in Sambrook et al., 1989, und Perbal, 1984. Zur Konstruk-

tion der ProCon-Vektoren wird auf die bereits eingangs erwähnte WO 96/07748 und die damit zusammenhängende Literatur verwiesen.

3. Ergebnisse

3.1 Analyse der HERV-Transkription in unterschiedlichen Zelltypen

Zur Untersuchung der HERV Transkription in unterschiedlichen Zellen wurde im ersten Schritt eine Methode (Reverse Dot-Blot Hybridisierung) eingesetzt, die ursprünglich zum Nachweis der HERV-Expression in mononukleären Zellen des peripheren Bluts entwickelt worden war (Herrmann und Kalden, 1994). Dabei wurden klonierte und charakterisierte HERV *pol*-Genfragmente aus humaner genomischer DNA auf einer Membran fixiert und mit radioaktiven HERV-*pol* Sonden hybridisiert. Die Sonden wurden mittels RT-PCR aus mRNA verschiedener Zellen unter Verwendung degenerierter Oligonukleotide amplifiziert, die zu einem hochkonservierten Bereich retroviraler *pol*-Gene homolog sind (Shih *et al.*, 1989; Donehower *et al.*, 1990). Mit dieser Methode erhielten wir mit jeder bisher untersuchten Zelllinie ein charakteristisches Hybridisierungsmuster, was einen ersten Hinweis auf eine gewebespezifische Expression der HERV Elemente darstellte.

3.2 Isolierung von LTR U3-Regionen exprimierter HERVs

Die gewebespezifische Expression eines Retrovirus ist vor allem durch ihre U3-Region festgelegt. In diesem Bereich sind alle regulatorischen Sequenzen wie Promotor, Enhancer und die Bindungsstellen für verschiedene zelluläre Transkriptionsfaktoren lokalisiert. Aus diesem Grund wurden Primer entwickelt, mit denen diese HERV-Sequenzen durch eine RT-PCR aus mRNA unterschiedlicher Zelllinien gezielt isoliert werden konnten (Tab. 2; Abb.1). Auf diese Weise wurden ca. 30 verschiedene HERV-LTRs kloniert. Ein Teil dieser Sequenzen wurde in einem Reporterplasmid auf Promotoraktivität und Gewebespezifität getestet.

Im ersten Ansatz wurde für die PCR ein polydT-Primer mit einem Primer kombiniert, der komplementär zum Polypurin Trakt (PPT) der retroviralen RNA ist (Abb.1). Der PPT-Trakt ist eine konservierte Region im nicht translatierten Bereich zwischen dem *env* Gen und der U3-

Region der 3'-LTR. Die PPT-Region wird während der reversen Transkription des Retrovirus als Primerbindungsstelle für die Synthese des Plus-Stranges genutzt (Sorge und Hughes, 1982).

Durch Datenbankanalysen wurden PPT-Sequenzen unterschiedlicher HERV-Familien identifiziert und durch Vergleich ihrer Homologien in verschiedene Gruppen eingeteilt. Aus den Konsensussequenzen der einzelnen Gruppen wurden Oligonukleotide als Primer für die RT-PCR synthetisiert. Die mRNA wurde aus verschiedenen Zelllinien präpariert: Epithelzellen (HeLa, HaCaT), Fibroblastenzellen (LC5), T-Zellen (H9, HUT78), Lymphoblasten (CML), Gliomzellen (85HG66, U373), Pancreaszellen (MiaPaCa2, Panc1), Leberzellen (Chang Liver) und Brustkrebszelllinien (T47-D, MCF7). Darüber hinaus wurden zusätzlich cDNA Genbanken (Clontech) verschiedener menschlicher Gewebe (Gehirn, Herz, Leber, Niere, Lunge, Pankreas, Plazenta, Skelettmuskel) für die RT-PCR eingesetzt.

Die erhaltenen Fragmente wurden anschließend kloniert, sequenziert und mittels Datenbankvergleichen analysiert. Von den PCR-Fragmenten, die mit den PPT- und den polydT-Primern erhalten wurden, konnten zwei LTRs über Homologievergleiche der HERV-H und der HERV-K Familie zugeordnet werden. Durch die Verwendung von polydT-Primern in diesen PCR-Ansätzen wurden zahlreiche Sequenzen amplifiziert, die keine Homologien zu bekannten retroviralen LTRs und darüber hinaus auch keine Promotorstrukturelemente enthielten. Aus diesem Grund wurden weitere Sequenzen aus konservierten Bereichen der U3-Region und aus der R-Region von HERV-K und HERV-H Familien zur Primersynthese ausgesucht (Mold *et al.*, 1997), (Abb. 1, Tab. 1). Die resultierenden PCR-Produkte wurden nach Auftrennung im Agarosegel auf Nitrozellulosefilter transferiert und mit Sonden hybridisiert, die aus den LTR-Regionen verschiedener HERV-LTRs (HERV-K-pl167, HERV-H-H6, HERV-E, HERV-L) präpariert wurden. Die hybridisierenden Fragmente wurden anschließend in einen Vektor (pZERO, Invitrogen) kloniert und sequenziert. Mit dieser Methode konnten verschiedene HERV-LTRs isoliert werden, die in Tabelle 3 aufgelistet sind.

Die HERV-K-LTRs, die aus menschlichem Gehirn und Herzgewebe sowie aus T47-D Zellen isoliert wurden, zeigen sehr hohe Sequenzhomologien zu der 3'-LTR von HERV-K10. Die HERV-H LTRs zeigen dagegen sehr viel höhere Sequenzvariabilitäten. HERV-H31, HERV-H3, HERV-HCM1, HERV-HCM4, HERV-HMP23 sind homolog zu der von Mager *et al.* isolierten HERV-H-H6 LTR, die anderen HERV-H Sequenzen zeigen Homologien zu den von

Anderssen *et al.* (1997) isolierten HERV-H LTRs aus Meerkatze, Krallenneffe und Mensch. Die aus T47-D Zellen isolierten HERV-W LTRs sind mit der LTR des Klons CL6 (Komurian-Pradel, 1999) verwandt.

3.3 Analyse der Expression von HERV-Promotoren in einem transienten Luziferase-Assay

Zur Analyse der Promotoraktivität und der Gewebespezifität der isolierten HERV-LTRs wurden diese zunächst in ein Luziferasereporterplasmid (pBL, Butz, K., DKFZ, Heidelberg) kloniert. Dieser Vektor enthält das *Photinus pyralis* Luziferasegen, fusioniert an das SV40 polyA Signal von pBLCAT2 (Hoppe-Seyler *et al.*, 1991).

Die einzelnen Vektorkonstrukte wurden transient in verschiedene Zelllinien transfiziert. Nach 48 h wurde die Luziferaseaktivität aus dem Zellysat mit dem Luziferase-Assay Kit von Promega gemessen und nach Abgleichen der β -Galaktosidaseaktivität bzw. der Renilla-Luziferaseaktivität als relative Luziferaseaktivität bestimmt. Die Promotoraktivitäten der LTRs wurden in Epithelzellen (HeLa, HaCaT), Fibroblastenzellen (LC5), T-Zellen (H9, HUT78), Gliomzellen (85HG66, U373), Leberzellen (Chang Liver), Pankreaszellen (MiaPaCa2, Panc1) und Brustkrebszelllinien (T47-D, MCF7) bestimmt.

Die Ergebnisse sind in den Abbildungen 2a - 2f dargestellt. Die HERV-H-H6 LTR besitzt demnach von allen untersuchten endogenen LTRs den stärksten Promotor. Die HERV-K LTR aus Plazenta ist in HeLa-Zellen besonders aktiv. In allen anderen Zelllinien ist diese LTR nur sehr schwach aktiv. Ebenfalls eine starke Aktivität in HeLa Zellen zeigte HERV-K-T47-D, diese LTR war auch in HaCat-Zellen und Pankreaszellen aktiv. Die HERV-L LTR besitzt in Leberzellen eine starke Promotoraktivität und ist in T-Zellen und Pankreaszellen schwach aktiv. Die HERV-T-S71A und HERV-E LTRs waren in keiner getesteten Zelllinie aktiv. Ebenso konnte bisher keinerlei Aktivität einer HERV-LTR in CML-Zellen beobachtet werden.

Die klonierten HERV-H LTRs (HERV-H1, HERV-H8, HERV-H13, HERV-H19, HERV-H-H6, Tab. 3) waren fast alle in 85HG66 Zellen aktiv, wobei HERV-H1 und HERV-H8 in dieser Zelllinie die höchste Aktivität zeigten (ohne Abb). HERV-H19 war in HeLa-Zellen sehr aktiv. Die

HERV-HCM1 LTR zeigte in allen Zelllinien die höchste Promotoraktivität und war besonders in Lungenfibroblasten (LC5) aktiv (Abb. 3).

3.4 Konstruktion von HERV-Hybrid-Vektoren und Überprüfung der Aktivitäten von HERV- Promotoren in diesen Vektoren

Die Funktionsfähigkeit menschlicher endogener retroviraler LTR Sequenzen in retroviralen Vektoren wurde in zwei unterschiedlichen Promotorkonversionsvektoren (ProCon) ausgetestet. Dazu wurden hybride HERV/MLV Vektoren unter Verwendung zweier auf MLV basierender Vektoren pLESN-MMTV (Abb. 7) und pLX-MMTV (Abb. 8) konstruiert. Diese Vektoren enthalten das EGFP Gen als Reportergen, das von der 5' LTR (unterschiedlich, je nachdem ob vor oder nach der Promotorkonversion) aus exprimiert wird, sowie ein Neomycingen, das von einem SV40 Promotor aus exprimiert wird. Darüberhinaus enthält der Vektor pLX-MMTV einen prokaryotischen Replikationsorigin, der die Reklonierung des Provirus für weitere molekulare Charakterisierungen erlaubt.

Zur Konstruktion der HERV-Hybrid Vektoren wurde jeweils die MMTV-LTR durch die HERV-HCM1 LTR ersetzt (Abb. 7). Dazu wurde die LTR zunächst mittels PCR aus dem Vektor pBL-HERV-H mit spezifischen Primern amplifiziert, die zusätzliche Sequenzen für die Restriktionsenzyme MluI und SacII enthielten. Diese Fragmente wurden anschließend in die 3' U3 deletierten Vektoren eingebaut. Das Reportergen EGFP wird nach der Transfektion in die Verpackungszelllinie zunächst vom MLV Promotor aus exprimiert. (Abb 9a) Nach Infektion der Zielzellen und erfolgreicher Promotorkonversion durch die reverse Transkription in den Zielzellen liegt das Reportergen unter der Transkriptionskontrolle der HERV LTR vor.

Die HERV-Hybrid Vektorkonstrukte pLESN-HERV-H (Abb. 7) und pLX-HERV-H (Abb.8), sowie die Ursprungsvektoren pLESN-MMTV und pLX-MMTV wurden in die amphotrophe Verpackungszelllinie PA317 transfiziert. Die resultierenden retroviralen Vektorpartikel wurden anschließend benutzt, um die Zelllinien CrfK und LC5 zu infizieren.

Die infizierten Zelllinien wurden kloniert und die selektionierten Zellklone daraufhin untersucht, ob sie die Vektorkonstrukte enthielten und ob die Promotorkonversion stattgefunden hatte. Dazu wurde aus infizierten und nicht infizierten Zellen chromosomale DNA präpariert

und mittels PCR analysiert. Die Primer wurden aus der MLV U3 (P5) und R (P2) Region sowie aus der HERV-H Region (P1) ausgesucht und in Kombination mit einem Primer aus der EGFP Region für eine PCR eingesetzt (Abb. 9a). Die PCR-Produkte wurden mit HERV-H spezifischen Sonden hybridisiert (Abb. 9b). Dabei ergab die DNA, die mit den pLX HERV-H Partikeln infiziert wurden, nach Amplifikation mit den Primern P1 und P3 ein PCR Produkt von 1,1 kb, das mit der HERV-H-Sonde hybridisierte. Die Amplifikation mit den MLV U3 spezifischen Primern (P2/P3) ergab mit DNA von Zellen, die mit pLX und mit pLX HERV-H infiziert waren, PCR Produkte mit einer Größe von etwa 900bp, die nicht mit der HERV-H Sonde hybridisierten. Aus der Amplifikation mit den MLV R Primern (P5/P3) wurde kein PCR-Produkt erhalten, das mit der HERV-H Sonde hybridisierte. Diese Ergebnisse zeigen, daß eine Promotorkonversion stattgefunden hat und der MLV Promotor der 5'LTR durch den HERV-Promotor ersetzt worden war.

Die Promotoraktivität der HERV LTR in den retroviralen Vektoren wurde nach Integration in die DNA der Zielzellen mittels FACS-Analysen über die Messung der EGFP Fluoreszenz bestimmt (Abb. 10). Die Aktivität des Ausgangsvektors pLX-MMTV wurde dabei mit dem HERV-Vektor pLX-HERV-H (H6) vor und nach Induktion mit Dexamethason verglichen. Der Vektor mit der MMTV LTR ist durch Dexamethason aktivierbar. Der Vektor mit der HERV LTR wird durch Dexamethason nicht aktiviert, er ist jedoch um etwa den Faktor 10 aktiver als der Dexamethason-stimmulierte MMTV-Hybrid-Vektor.

3.5 Einfluß von Regulationselementen in R- und U5-Region auf die Promotoraktivität von HERV-Sequenzen.

Um auszutesten, welcher Sequenzbereich für einen funktionsfähigen HERV-Promotor erforderlich ist, wurde an einigen Beispielen der Einfluß von zusätzlichen LTR-Sequenzen, die außerhalb der U3-Region in der LTR lokalisiert sind, untersucht. Dazu wurde von 7 HERV-K LTRs (HERV-K-T47D, L5, L50, L8, L9, L48 und L20/49) die Aktivität der U3-Region mit der Aktivität der entsprechenden U3-R-Fragmente im Luziferase-Assay verglichen. Überraschenderweise zeigte sich, daß die verschiedenen R-Regionen den Promotor in der U3-Region sehr unterschiedlich beeinflussen können. Bei LTRs der Gruppe 1 (L5, L50, L8, L9) führt die Anwesenheit von R-Sequenzen in allen getesteten Zelllinien zu einem deutlichen Anstieg der Pro-

motoraktivität (Abb. 4a). Bei LTRs der Gruppe 2 (L20/L49) hingegen wird die Aktivität des HERV-Promotors durch die R-Region reduziert (Abb. 4b). Der HERV-K-T47D-Promotor (Abb. 5) und der L48-Promotor (ohne Abbildung) werden durch die entsprechenden R-Sequenzen nicht wesentlich beeinflusst. Interessanterweise haben jedoch im Falle der HERV-K-T47D-LTR Sequenzbereiche, die downstream der U3-R-Region lokalisiert sind und die U5-Region, sowie den 3' nicht translatierten Bereich und den Beginn des *gag*-Gens umfassen, einen deutlich aktivierenden Effekt (Abb. 5).

Die Sequenzanalyse der verschiedenen getesteten R-Regionen zeigte, daß LTRs der Gruppe 1 in der R-Region eine Bindungsstelle für den Transkriptionsfaktor SP1 aufweisen, die in der R-Region der Gruppe 2 LTR fehlt (Abb. 6). Dagegen enthält die Gruppe 2 R-Region eine potentielle Bindungsstelle für den Faktor TFS3, der als Transkriptionsrepressor wirkt. Dies zeigt, daß die Aktivität von HERV-Promotoren durch den Einbau zusätzlicher Regulations-elemente wie Transkriptionsfaktorbindungsstellen, Enhancer-Sequenzen oder negativ regulierende Elemente modifiziert werden kann.

Literatur

- Anderssen, S., Sjøttem, E., Svineng, G., Johansen, T. (1997). Comparative analyses of LTRs of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset, african green monkey, and man. *Virology* **234**. 14-30
- Di Christofano, A., Strazzullo, M., Longo, L., La Mantia, G. (1995). 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. *Nucl. Acid. Res.* **23**. 2823-2830
- Donehower L. A., Bohannon, R. C., Ford, R. A. (1990). The use of primers from highly conserved *pol* regions to identify uncharacterized retroviruses by the polymerase chain reaction. *J. Virol. Methods* **28**. 33-46
- Emmerman, M., Temin, M. H. (1986). Comparison of promotor suppression in avian and murine retrovirus vectors. *Nucl. Acid. Res.* **14**. 9381-9396
- Feuchter, A., Mager, D. (1990). Functional heterogeneity of a large family of human LTR-like promoters and enhancers. *Nucl. Acid. Res.* **18**. 1261-1270
- Feuchter-Murthy, A. E., Freeman, J. D., Mager, D. L. (1993). Splicing of a human endogenous retrovirus to a novel phospholipase A2 related gene. *Nucl. Acid. Res.* **21**. 135-143
- Goodchild N. L., Wilkinson, D. A., Mager, D. L. (1992). A human endogenous long terminal repeat provides a polyadenylation signal to a novel, alternatively spliced transcript in normal placenta. *Gene*, **121**. 287-294
- Herrmann, M., Kalden, J. R. (1994). PCR and reverse dot hybridisation for the detection of endogenous retroviral transcripts. *J. Virol. Methods* **46**. 333-348
- Hoppe-Seyler, F., Butz, K., zur Hausen, H. (1991). Repression of the Human Papillomavirus Type 18 Enhancer by the Cellular Transcription Factor Oct-1. *J. Virol.* **65**. 5613-5618
- Kato-N., Shimotohno, K., VanLeeuwen, D., Cohen, M. (1990). Human proviral mRNAs down-regulated in choriocarcinoma encodes zinc finger protein related to Kruppel. *Mol. Cell. Biol.* **10**. 4401-4405
- Komurian-Pradel, F., G. Paranhos-Baccala, F. Bedin, A. Ounanian-Paraz, M. Sodoyer, C. Ott, A. Rajoharison, E. Garcia, F. Mallet, B. Mandrand, und H. Perron, 1999. Molecular cloning and characterization of MSRNV-related sequences associated with retrovirus-like particles. *Virology*. 260:1-9.

- Leib-Mösch, C., Haltmeier, M., Werner, T., Geigl, E.-M., Brack-Werner, R., Francke, U., Erfle, V., Hehlmann, R. (1993). Genomic distribution and transcription of solitary HERV-K LTRs. *Genomics* **18**. 261-269
- Mager, D. L. (1989). Polyadenylation Function and Sequence Variability of the Long Terminal Repeats of Human Endogenous Retrovirus-like Family RTVL-H. *Virology* **173**. 591-599.
- Mann, R. und D. Baltimore. 1985 Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNAs *J. Virol.* 54:401-407.
- Mold, D. E., Wu, T., Askin, F., Huang, R. C. (1997). Four classes of HERV-K long terminal repeats and their relative promoter strengths for transcription. *Biomed. Science* **4**. 78-82.
- Perbal, B. 1984. *A Practical Guide to Molecular Cloning*. John Wiley & Sons.
- Price, J., Turner, D., Cepko, C. (1987). Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer. *Proc. Natl. Acad. Sci.* **89**. 9237-9241
- Rein A, 1994 Retroviral RNA packaging: a review. *Arch Virol Suppl* 9:513-22.
- Saller, R. M., Öztürk, F., Salmons, B., Günzburg, W. (1998). Construction and characterization of a hybrid mouse mammary tumor virus/murine leukemia virus-based retroviral vector. *J. Virol.* **2**. 1699-1703
- Salmons, B., Günzburg, W. H. (1993). Targeting of retroviral vectors for gene therapy. *Human Gene Therapy* **4**. 129-141
- Sambrook, J., E.F. Fritsch und T. Maniatis. 1989. *Molecular Cloning*. Cold Spring Harbor Laboratory Press, New York, USA.
- Schulte, A. M., Lai, S., Kurtz, A., Czubayko, F., Riegel, A. T. (1996). Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Proc. Natl. Acad. Sci.* **93**. 14759-14764
- Seifarth, W., Baust, C., Murr, A., Skladny, H., Krieg-Schneider, F., Blusch, J., Werner, T., Hehlmann, R., Leib-Mösch, C., (1998). Proviral Structure, chromosomal location and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47-D particles. *J. Virol.* **72**. *In press*
- Shih, A., Misra, R., Rush, M. G. (1989). Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: relation to primate retroviruses. *J. Virol.* **63**. 64-75

- Sorge, J., Hughes, S. H. (1982). Polypurine tract adjacent to the U₃ region of the Rous sarcoma virus genome provides a *cis*-acting function. *J. Virol.* **43**. 482-488
- Ting, C.-N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C., Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* **6**. 1457-1465.
- Wilkinson, D. A., Mager, D. L., Leong, J.-A. (1994). Endogenous human retroviruses. In "The retroviridae" Vol. **3** (J. Levy, ed.), pp. 465-535, Plenum Press, NY.
- Patience, C., Wilkinson, D. A., Weiss, R. A. (1997). Our retroviral heritage. *Trends Genet.* **13**, 116-120.
- Patience, C., Takeuchi, Y., Cosset, F.-L., Weiss, R. A. (1998). Packaging of endogenous retroviral sequences in retroviral vectors produced by murine and human packaging cells. *J. Virol.* **72**, 2671-2676.

Patentansprüche

1. Retroviraler Expressionsvektor, enthaltend zumindest die nachfolgenden Elemente in funktioneller Anordnung:
 - a) DNA-Sequenzen zur Verpackung der Vektor-RNA und zur zellspezifischen Expression von Proteinen oder Peptiden, die von heterologen DNA-Nukleotidsequenzen kodiert werden;
 - b) ein oder mehrere für ein Protein oder Peptid kodierende DNA-Nukleotidsequenzen,

dadurch gekennzeichnet,
daß die DNA-Sequenzen zur zellspezifischen Expression eine zellspezifisch regulierbare Promotorregion aus einer humanen endogenen retroviralen DNA-Nukleotidsequenz (HERV) enthalten.
2. Expressionsvektor nach Anspruch 1,
dadurch gekennzeichnet,
daß die DNA-Sequenzen zur zellspezifischen Expression aus dem LTR-Bereich und wahlweise dem nicht translatierten Bereich zwischen der 5' LTR und dem gag-Bereich von HERVs stammen.
2. Vektor nach Anspruch 1 oder Anspruch 2,
dadurch gekennzeichnet,
daß der gesamte LTR-Bereich, der U3-Bereich oder der R- und U3-Bereich aus einer humanen endogenen retroviralen Nukleotidsequenz stammt.
3. Vektor nach einem oder mehreren der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
daß die für ein oder mehrere Proteine oder Peptide kodierende Nukleotidsequenzen ausgewählt werden aus einem oder mehreren Elementen der Gruppe, bestehend aus Markergenen, therapeutischen Genen, antiviralen Genen, Antitumorgenen und Zytokinen.

4. Vektor nach einem oder mehreren der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die zellspezifisch regulierbare Promotorregion aus dem LTR-Bereich einer zellspezifisch exprimierten endogenen humanen retroviralen Nukleotidsequenz stammt.
5. Vektor nach einem oder mehreren der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die humanen endogenen retroviralen zellspezifisch regulierbaren Promotorsequenzen ausgewählt werden aus einem oder mehreren Promotorsequenzen von HERV-Familien der Gruppe, bestehend aus HERV-K, HERV-H, HERV-E, HERV-L, HERV-T, HERV-R, HERV-I, HERV-P, ERV9. HERV-W.
6. Vektor nach einem oder mehreren der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß die Promotorregion neben der TATA-Box weitere Erkennungs- und Bindungsstellen für regulatorische Proteine umfaßt.
7. Vektor nach Anspruch 7, dadurch gekennzeichnet, daß die Erkennungs- und Bindungsstellen für regulatorische Proteine die GC-Box, die CAAT-Box, Enhancersequenzen und Repressorsequenzen sowie Hormon-responsive Sequenzmotive umfassen und daß wahlweise zusätzlich Erkennungs- und Bindungsstellen für regulatorische Proteine aus der LTR-Region exogener Retroviren und/oder aus zellulären Genen enthalten sind.
8. Vektor nach einem oder mehreren der vorhergehenden Ansprüche, dadurch gekennzeichnet, daß der Vektor ein Promotor-Konversionsvektor ist, umfassend einen 5' LTR-Abschnitt der Struktur U3-R-U5, ein oder mehrere Sequenzen, ausgewählt aus kodierenden und nicht kodierenden Sequenzen, und einen 3' LTR-Abschnitt, umfassend einen teilweise oder vollständig deletierten U3-Abschnitt, wobei der deletierte U3-Abschnitt durch eine zellspezifisch regulierbare Promotorregion aus einer HERV-LTR-Sequenz ersetzt ist, gefolgt vom R-U5-Abschnitt.

9. mRNA oder RNA eines retroviralen Expressionsvektors nach einem oder mehreren der vorhergehenden Ansprüche.
10. Prokaryontenzelle oder Eukaryontenzelle, enthaltend einen retroviralen Expressionsvektor nach einem oder mehreren der vorhergehenden Ansprüche.
11. Eukaryontenzelle, enthaltend einen retroviralen Expressionsvektor nach einem oder mehreren der vorhergehenden Ansprüche in integrierter Form.
12. Verwendung eines zellspezifisch regulierbaren Promotorabschnitts aus einer humanen endogenen retroviralen DNA-Nukleotidsequenz zur Steuerung der Expression von Fremdgenen in retroviralen Expressionsvektoren, bevorzugt ProCon-Vektoren.
13. Verwendung eines Expressionsvektors nach einem oder mehreren der vorhergehenden Ansprüche zur Expression von Fremdgenen in der Gentherapie.
14. Virion, enthaltend eine retrovirale Expressionsvektor-RNA, abgeleitet von einer Expressionsvektor-DNA nach einem oder mehreren der vorhergehenden Ansprüche.
15. Verfahren zur Herstellung eines Virions nach Anspruch 15 zum Einführen einer oder mehreren für ein Protein oder Peptid kodierender Nukleotidsequenzen,
dadurch gekennzeichnet,
daß der retrovirale Expressionsvektor nach einem oder mehreren der vorhergehenden Ansprüche in eine geeignete Verpackungszelllinie unter solchen Bedingungen eingebracht wird, daß das Virion ausgebildet wird und von der Verpackungszelllinie freigesetzt wird.
16. Verfahren zum Einführen von Nukleotidsequenzen, die für ein oder mehrere Proteine oder Peptide kodieren, in eine Eukaryontenzelle,
dadurch gekennzeichnet,
daß die Zelle mit einem Virion nach Anspruch 15 unter solchen Bedingungen infiziert wird, daß die für das Protein oder Peptid kodierende Nukleotidsequenzen in die chromosomale DNA der Eukaryontenzelle inseriert wird.

17. Verfahren nach Anspruch 17,
dadurch gekennzeichnet,
daß die Eukaryontenzelle eine Säugerzelle ist.
18. Verfahren nach Anspruch 18,
dadurch gekennzeichnet,
daß die Säugerzelle eine menschliche Zelle ist.
19. Retrovirales Vektorsystem, umfassend einen retroviralen Expressionsvektor nach einem oder mehreren der vorhergehenden Ansprüche und eine Verpackungszelllinie mit zumindest einem retroviralen oder rekombinanten retroviralen Konstrukt, das für die Verpackungsproteine des retroviralen Expressionsvektors kodiert.
20. Retrovirales Vektorsystem nach Anspruch 20,
dadurch gekennzeichnet,
daß die Verpackungszelllinie retrovirale oder rekombinante retrovirale Konstrukte enthält, die für solche retroviralen Proteine kodieren, die vom retroviralen Expressionsvektor nicht kodiert werden.

Tab. 1: Human endogenous retroviral elements

	HERV family	Copy number	% of genome	
Class I HERVs (type C- related HERVs)	HERV-ERI HERV-E (4-1, ERVA, NP-2) HERV-E LTR 51-1 ERV1 HERV-R (ERV3) RRHERV-I	35 - 50 500 - 600 35 - 50 10 - 15 10 20	0.07 %	
	HERV-T (S71, CRTK1, CRTK6) HERV-T LTR	50 - 60 150 - 200		
	ERV-FRD	8		
	HERV-RW HERV-W (MSRV) HERV-R (ERV9) ERV9 LTR	25 - 50 30 - 40 3000 - 4000	0.2%	
	HERV-P (HuERS-P, HuRRS-P)	50 - 90	0.01%	
	HERV-IP HERV-I (RTVL-I) HERV-IP-T47D (ERV-FTD) HERV-IP LTR	25 - 50 35 1800 - 2000	0.01%	
	HERV-HF HERV-H (RTVL-H, RGH) HERV-F HERV-H-LTR	900 - 1000 16 1000	0.2%	
	Class II HERVs (type A-, B-, and D-related HERVs)	HERV-K <i>HERV-K(HML-1)</i> <i>HERV-K(HML-2)</i> HERV-K10 HERV-K-HTDV HERV-K-IDDM <i>HERV-K(HML3)</i> <i>HERV-K(HML-4)</i> HERV-K-T47D <i>HERV-K(HML-6)</i> HERV-K-HML-6p HERV-K-HML-6.17 <i>HERV-K(HML-7)</i> HERV-K-NMW7 <i>HERV-K(HML-8)</i> <i>HERV-K(HML-9)</i> HERV-K-NMW9 <i>HERV-K(HML-10)</i> HERV-KC4 HERV-K LTR	10 - 20 30 - 50 25 6 30 - 40 10 - 50 10 000 - 25 000	0.5%
		HERV-L	100 - 200	0.02%
		Foamy virus- related HERVs		

Tab. 2: Primers used for the amplification of different HERV-LTR-regions

Nr.	Primer	Sequence
34	HERV-K	ATGGCGGTTTTGTCGAA
35	HERV-K	GTTCCMTYAGTATTTATTGATC
36	HERV-K	ATGGAGCATACAATCGGG
3	HERV-K	AAGAAAAGGGGGAAATGTGGG
11	HERVKC4	AAAGGGAGGGGGGCATG
12	HERV-KT47-D	TAAAAGGGGGGAGATG
1	HERV-H	ATGTGAGCAACATGGCTGTTTATTC
2	HERV-H	TGTCAGGCCTCTGAGCCCAA
39	HERV-H	GCCATCTCGAGTGTGAGSCTCTGAGYCYARGC
37	HERV-H	TATCTTGAATTCGKGTGAGCAAYAARRCTTTA
31	polydT	TTTTTTTTTTTTTTTT
17	HERV-E	AAAGGGGGGAAATATG
18	HERV-L	AGGGGTGGGACTTGCGATG
19	HERV-W	TGTTGAGATGGGGGACTGAG
20	HERV-W	GCAGTTGCAAGATTTAATAGAG

Tab. 3: Analyzed HERV-LTRs

3/29

A: HERV-LTRs from different cell lines and tissues

	Primer	Herkunft	Homology
HERV-K2	34/36	T47-D	HERV-K10, M12854, 97,9 %
HERV-K3	34/36	T47-D	HERV-K10, M12854, 98,4 %
HERV-K22-K32-K27-K45	34/36	brain	HERV-K10, M12854, 98,6 % *
HERV-K30	3/31	heart	HERV-K10, M12854, 97,6 %
HERV-K-T47D-L5		T47-D	MRSV, AF127229
HERV-K-T47D-L50		T47-D	MRSV, AF127229
HERV-K-T47D-L8		T47-D	MRSV, AF127229
HERV-K-T47D-L9		T47-D	MRSV, AF127229
HERV-K-T47D-L48		T47-D	MRSV, AF127229
HERV-K-T47D-L20		T47-D	MRSV, AF127229
HERV-IP-T47D		T47-D	MRSV, AF127229
HERV-T47D-W2	19/20	T47-D	MRSV, AF127229
HERV-T47D-W4	19/20	T47-D	MRSV, AF127229
HERV-T47D-W5	19/20	T47-D	MRSV, AF127229
HERV-W1	19/20	T47-D	MRSV, AF127229
HERV-W10	19/20	T47-D	MRSV, AF127229
HERV-W11	19/20	T47-D	MRSV, AF127229
HERV-W18	19/20	T47-D	MRSV, AF127229
HERV-W2	19/20	T47-D	MRSV, AF127229
HERV-W22	19/20	T47-D	MRSV, AF127229
HERV-W23	19/20	T47-D	MRSV, AF127229
HERV-W4	19/20	T47-D	MRSV, AF127229
HERV-W5	19/20	T47-D	MRSV, AF127229
HERV-W6	19/20	T47-D	MRSV, AF127229
HERV-W8	19/20	T47-D	MRSV, AF127229
HERV-H1	1/2	H9	Cercopithecus aethiops ERV-H; U96012, 87,1%
HERV-H8	1/2	HUT	HERV-H LTR18106, 84,8%
HERV-H13	1/2	HUT	HERV-H LTR18106, 91,8%
HERV-H19	1/2	liver	Callithrix jacchus ERV-H, 5'LTR: U96052, 92,1%
HERV-H31	1/2	liver	HERV-H(H6) x12717, 99,8%
HERV-H3	1/31	85HG66	HERV-H(H6) x12717, 100 %
HERV-H CL1	1/2	Chang Liver	HERV-H(H6) x12717, 100 %
HERV-H CL2	1/2	Chang Liver	HERV-H LTR18106, 84 %
HERV-H CL3	1/2	Chang Liver	Callithrix jacchus ERV-H, 5'LTR Silva 5, U96057, 84,2 %
HERV-H CL4	1/2	Chang Liver	HERV-H(H6) x12717, 100 %
HERV-H PA7	1/2	Panc1	Callithrix jacchus ERV-H, 5'LTR Silva 4, U96062, 85,7 %
HERV-H PA8	1/2	Panc1	Cercopithecus aethiops ERV-H, Vero 22, U96012, 87,1%
HERV-H PA9	1/2	Panc1	HERV-H LTR18106, 85 %
HERV-H PA10	1/2	Panc1	Callithrix jacchus ERV-H, 5'LTR Silva 4, U96062, 85,6 %
HERV-H MC14	1/2	MCF7	Cercopithecus aethiops ERV-H, Vero 22, U96012, 86,6%
HERV-H MC15	1/2	MCF7	Cercopithecus aethiops ERV-H, U96012, 86,6 %
HERV-H MC16	1/2	MCF7	Callithrix jacchus ERV-H, 5'LTR Silva 4, U96062, 87,4 %
HERV-H MC17	1/2	MCF7	Cercopithecus aethiops ERV-H, Vero 22, U96012, 86,6%
HERV-H MP20	1/2	MiaPaca	Human beta globin retrovirus-like repetitive element, k01891, 92,8 %
HERV-H MP21	1/2	MiaPaca	HERV-H LTR18106, 89,2 %
HERV-H MP23	1/2	MiaPaca	HERV-H(H6) x12717, 99,5 %

B: HERV-LTRs published in the literature

	(bp)	reference
HERV-K-pI167	970	Leib-Mösch <i>et al.</i> , 1993
HERV-K-T47-D	1200	Seifarth <i>et al.</i> , 1998
HERV-H-H6	393	Feuchter und Mager, 1990
HERV-T-S71A	625	Murr, Dissertation, 1998
HERV-E	391	Steele <i>et al.</i> , 1984
HERV-L	462	Cordonnier <i>et al.</i> , 1995

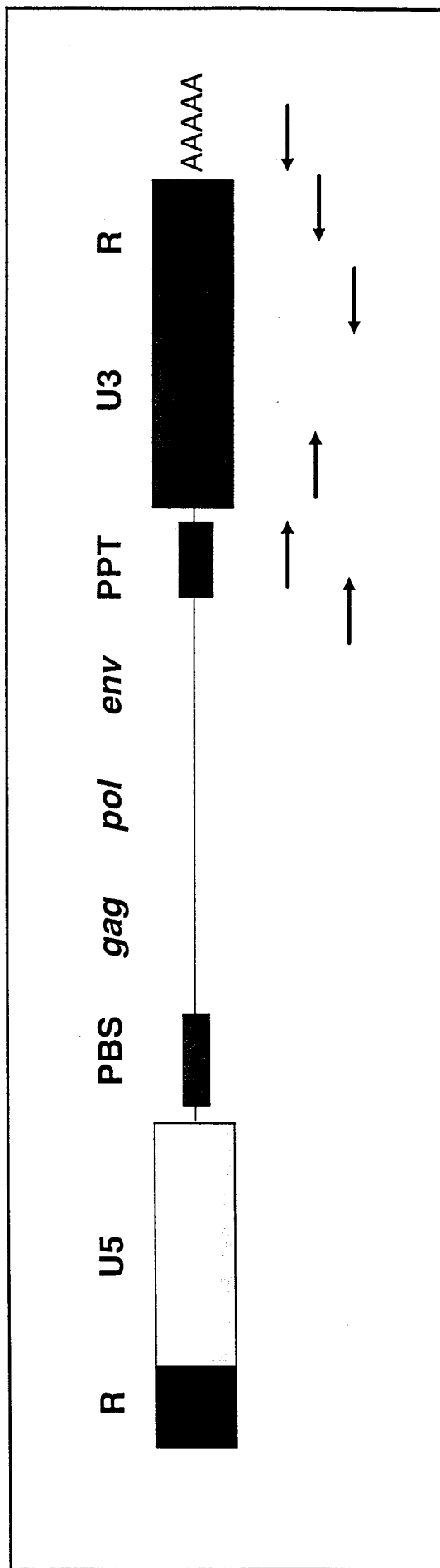
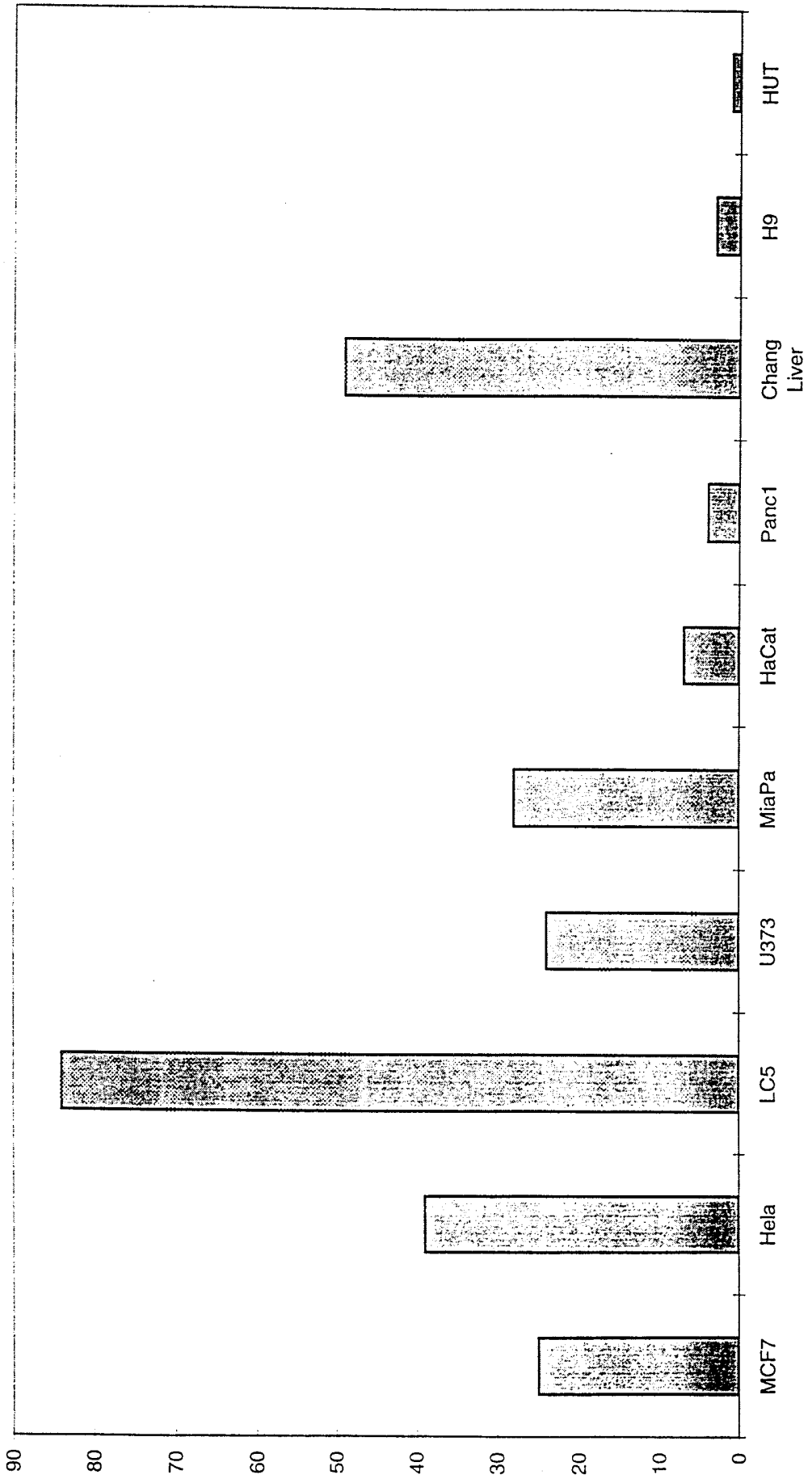


Fig.1: RT-PCR strategy to isolate U3/R-regions of transcribed HERVs

6/29

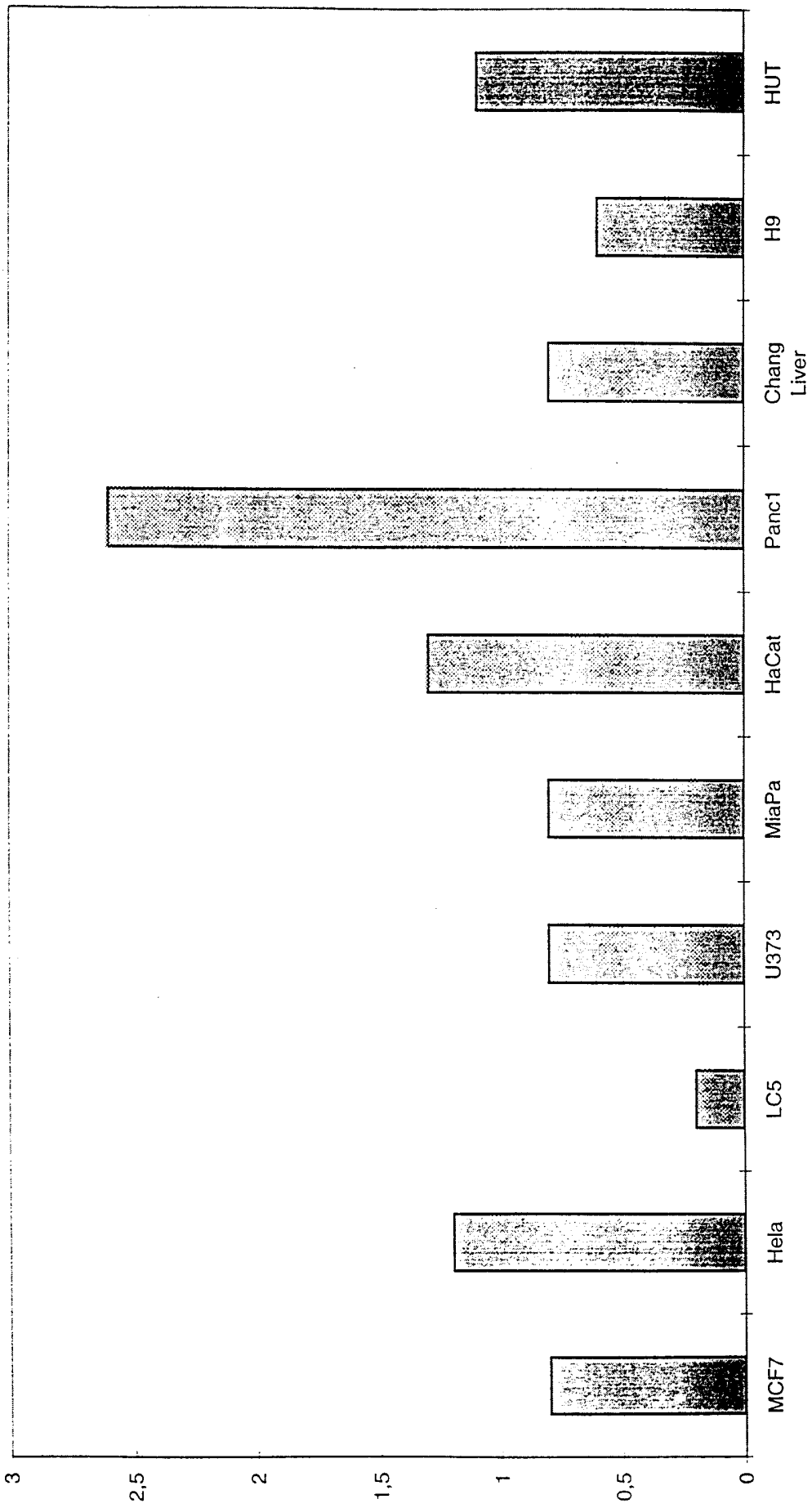
HERV-H-H6

Abb.2a)



HERV-E

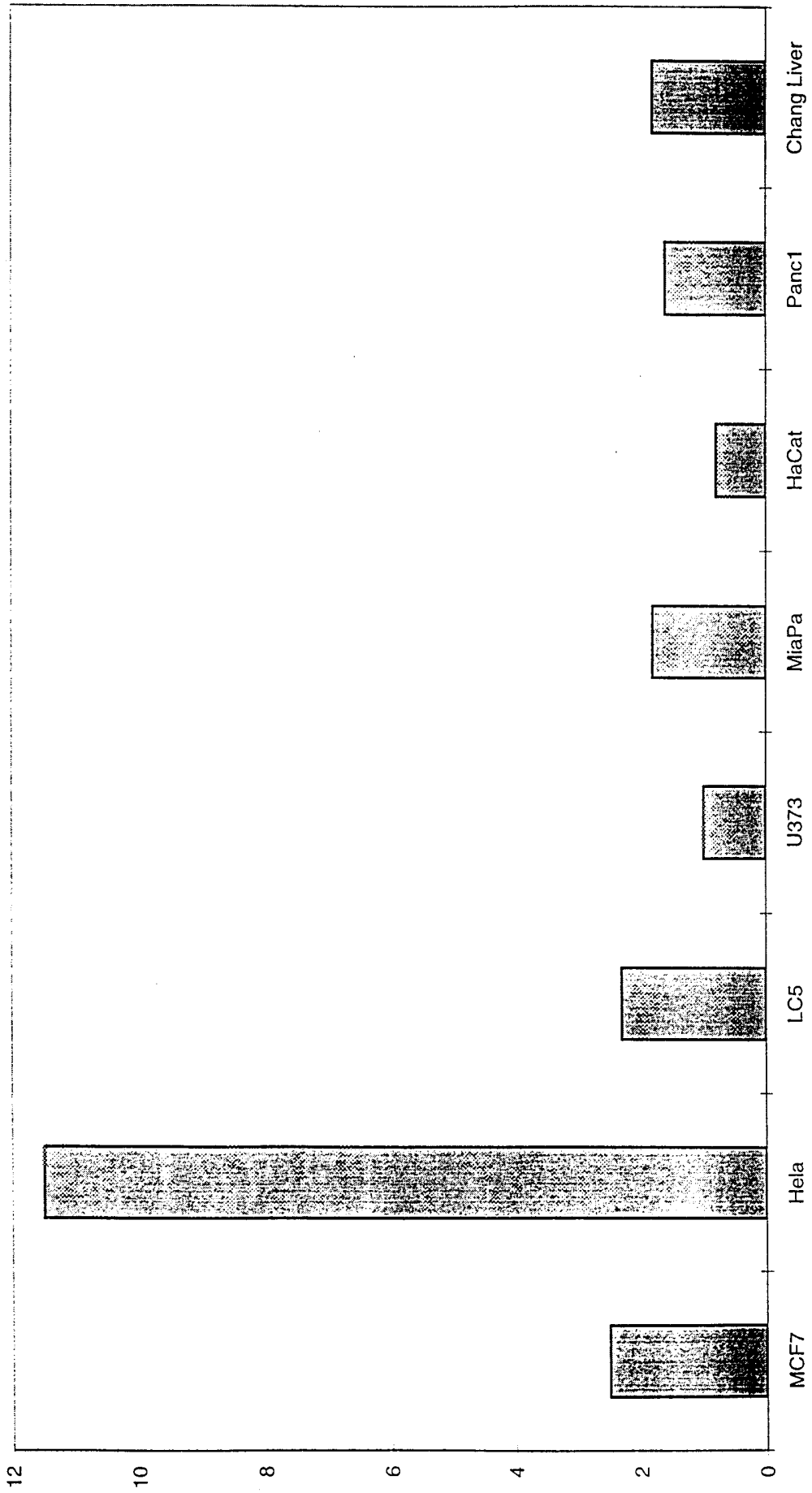
Abb.2b)



8/29

HERV-Kp1167

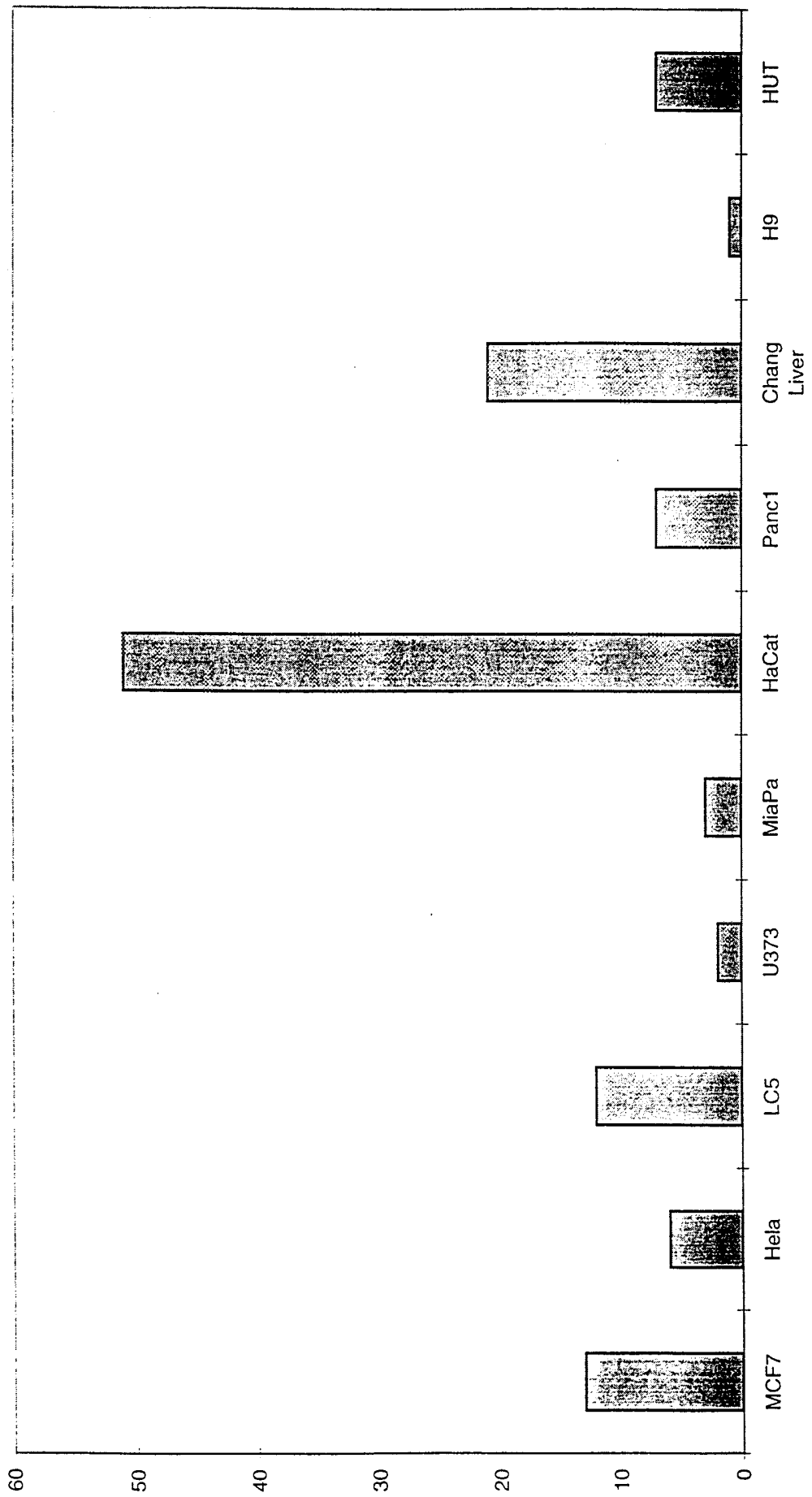
Abb.2c)



9/29

HERV-L

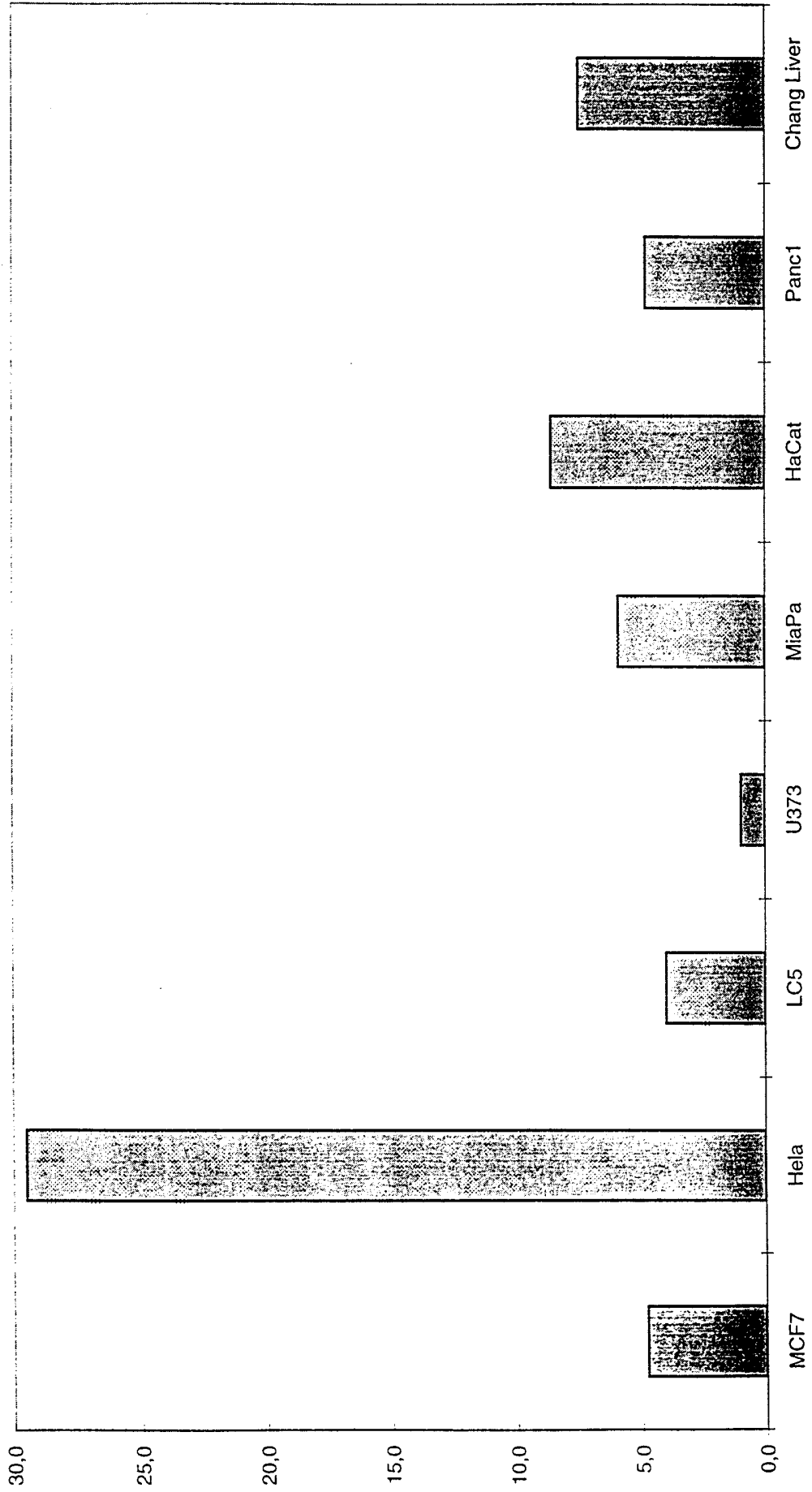
Abb.2d)



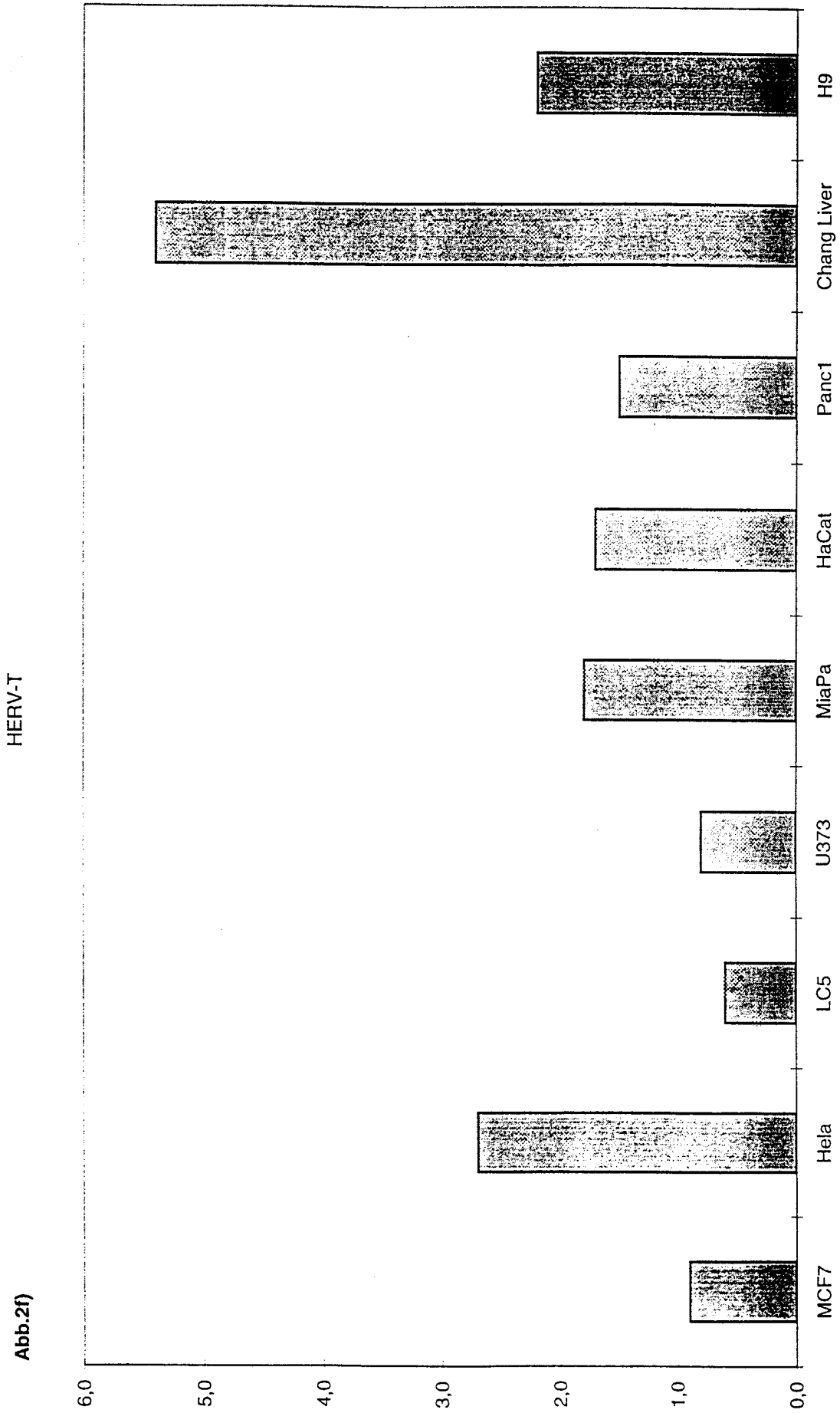
10/29

HERV-K-T47D

Abb.2e)



11/29



12/29

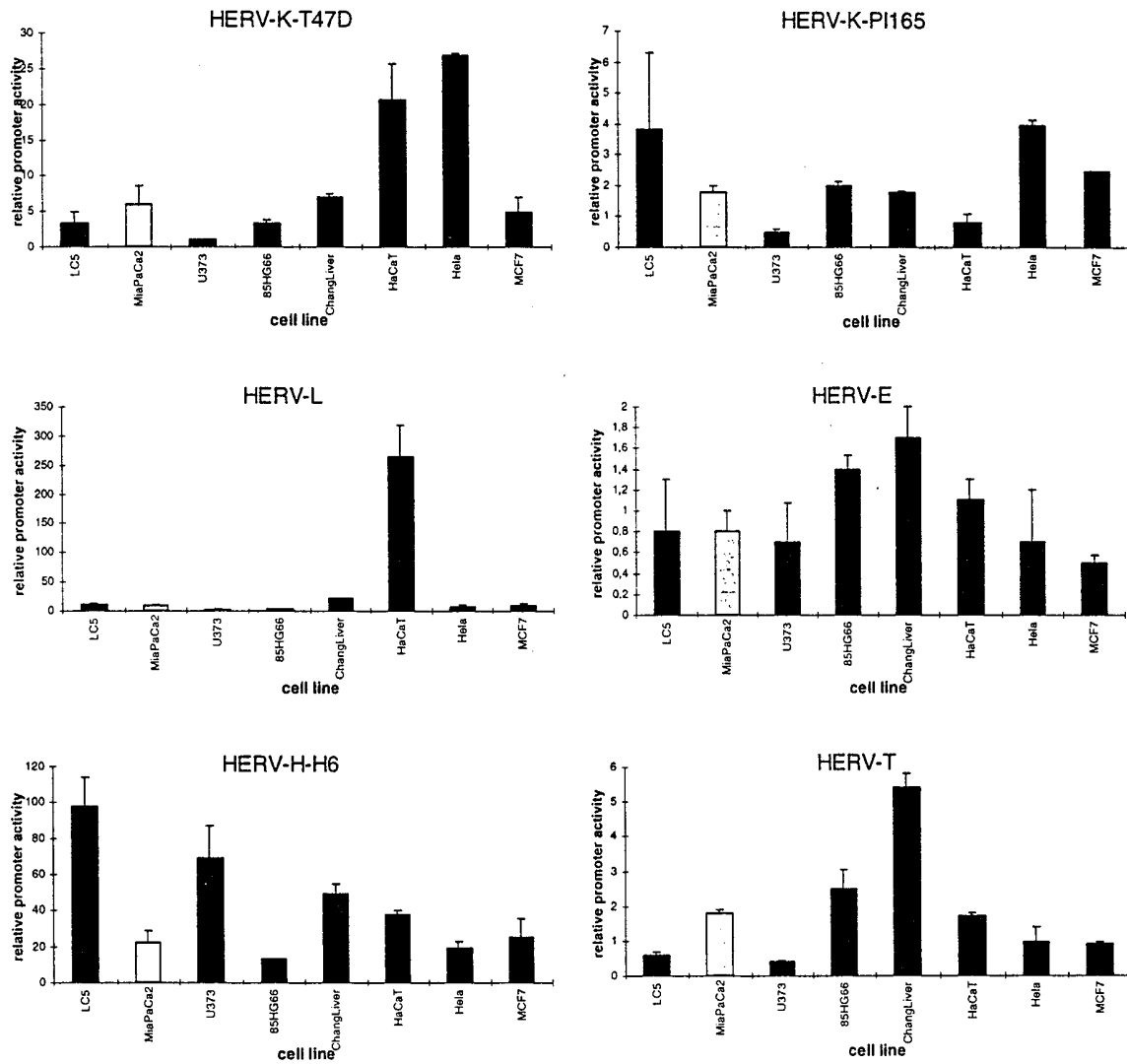


Abb. 2g: relative promoter activity of different HERV-LTRs in different cell lines

13/29

Fig. 3a

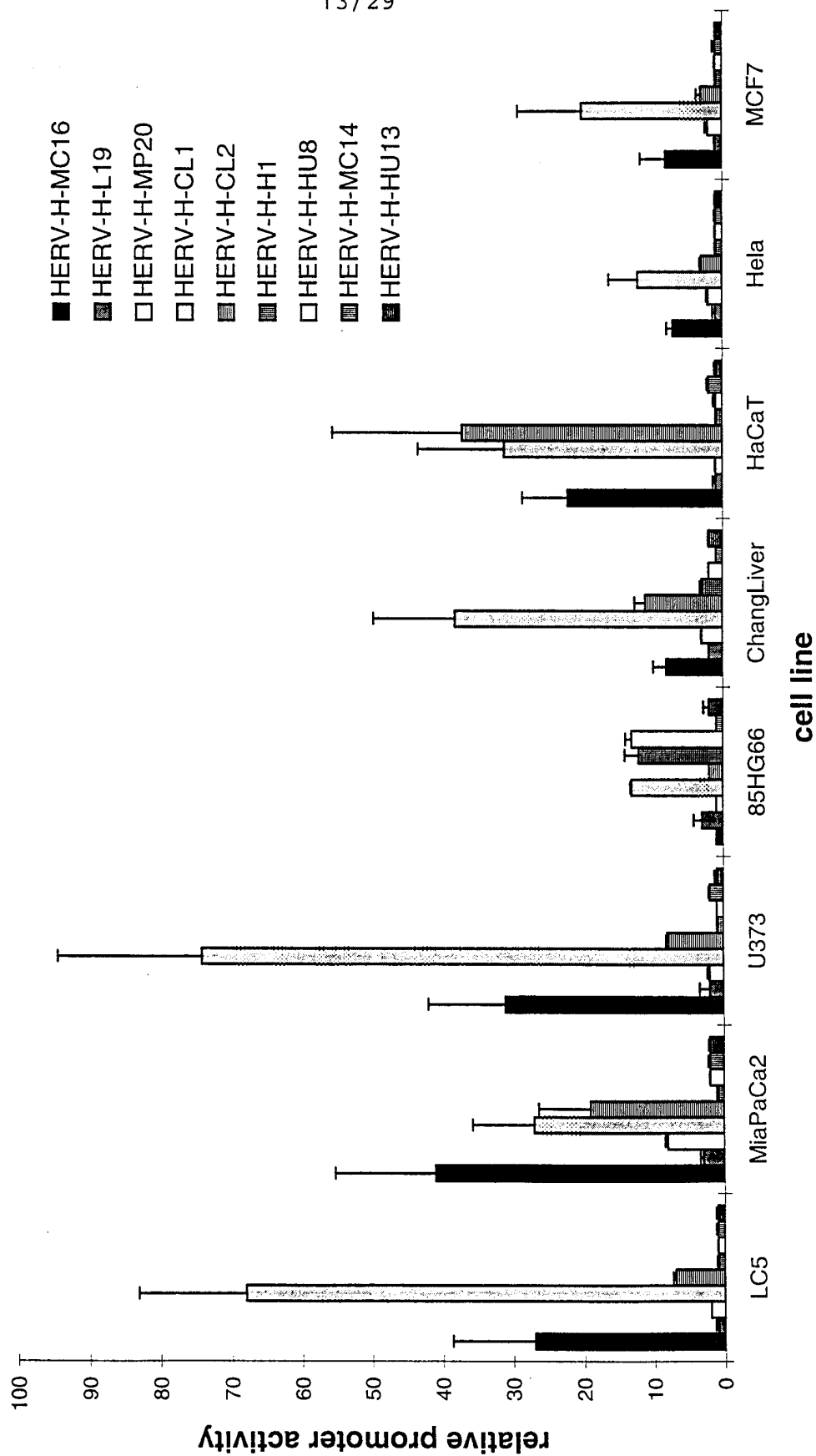
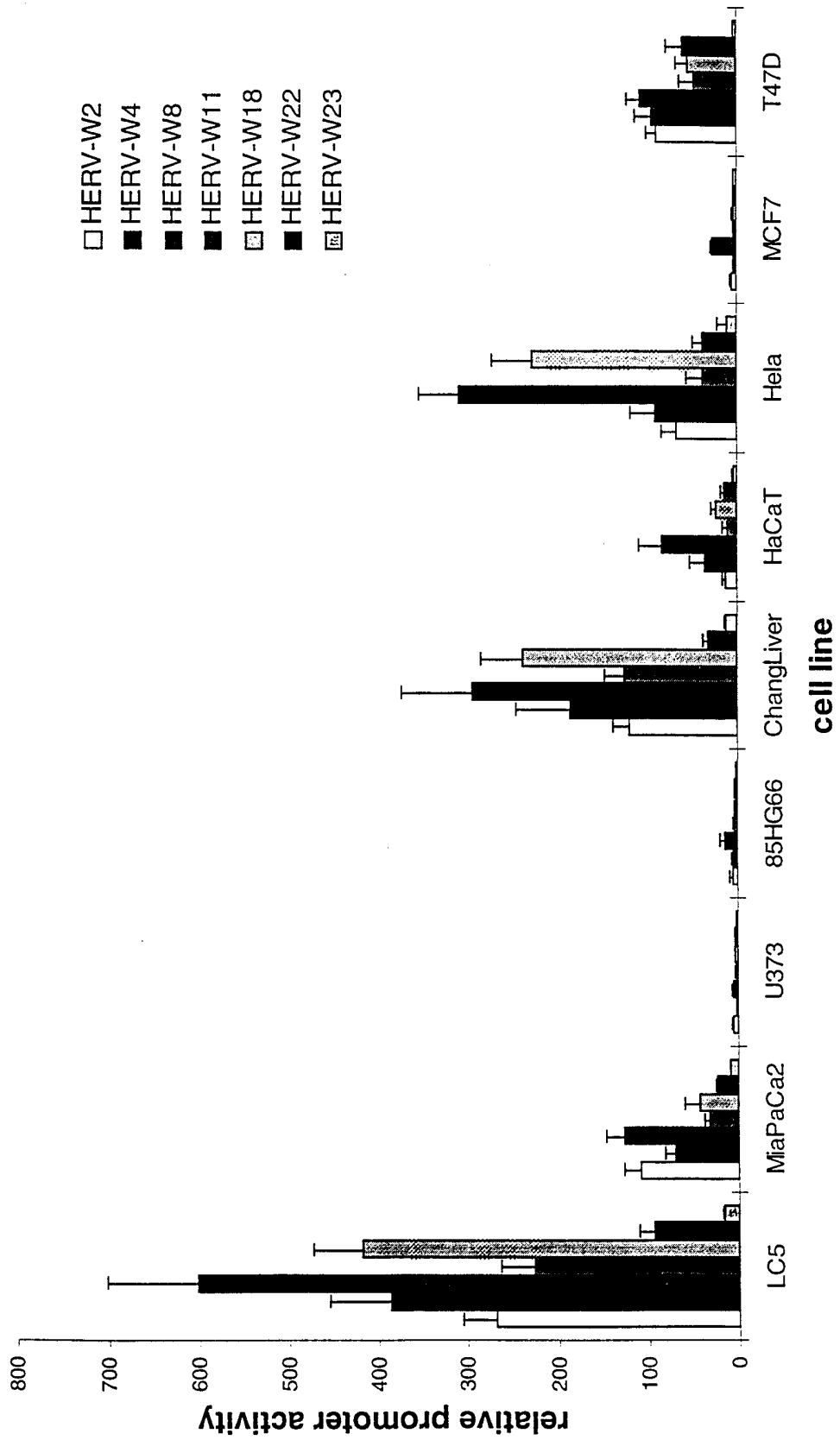


Fig. 3b



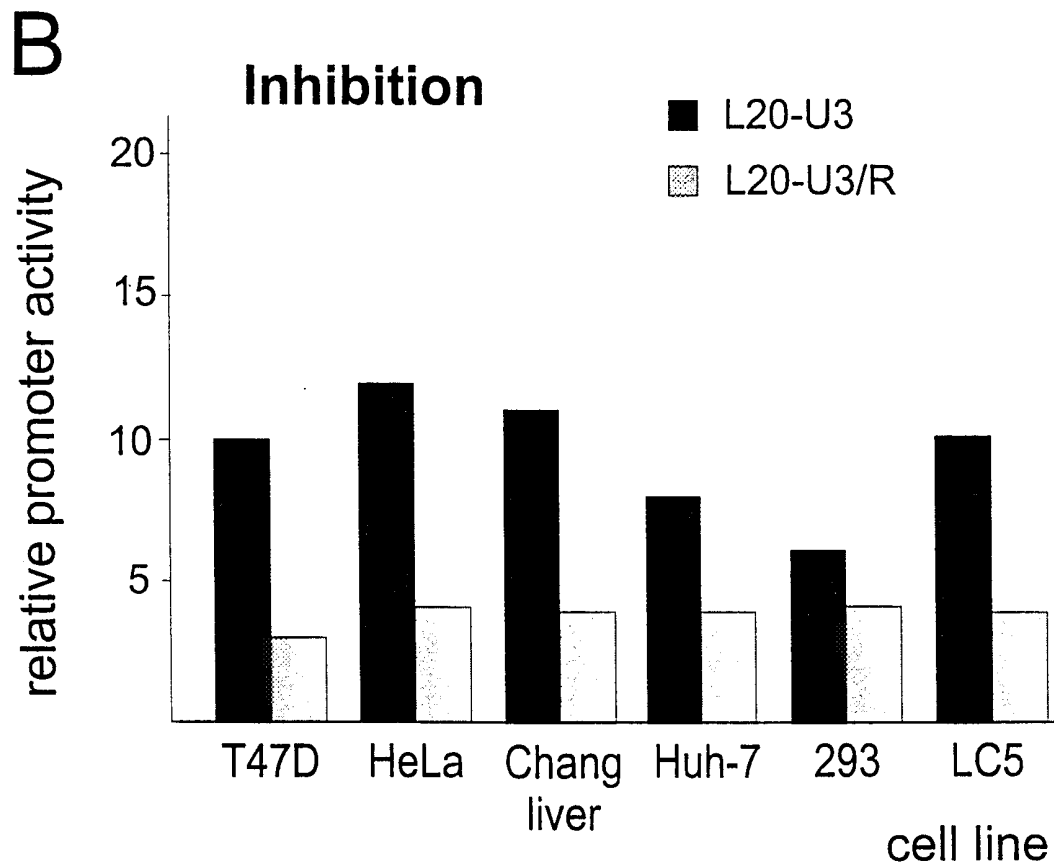
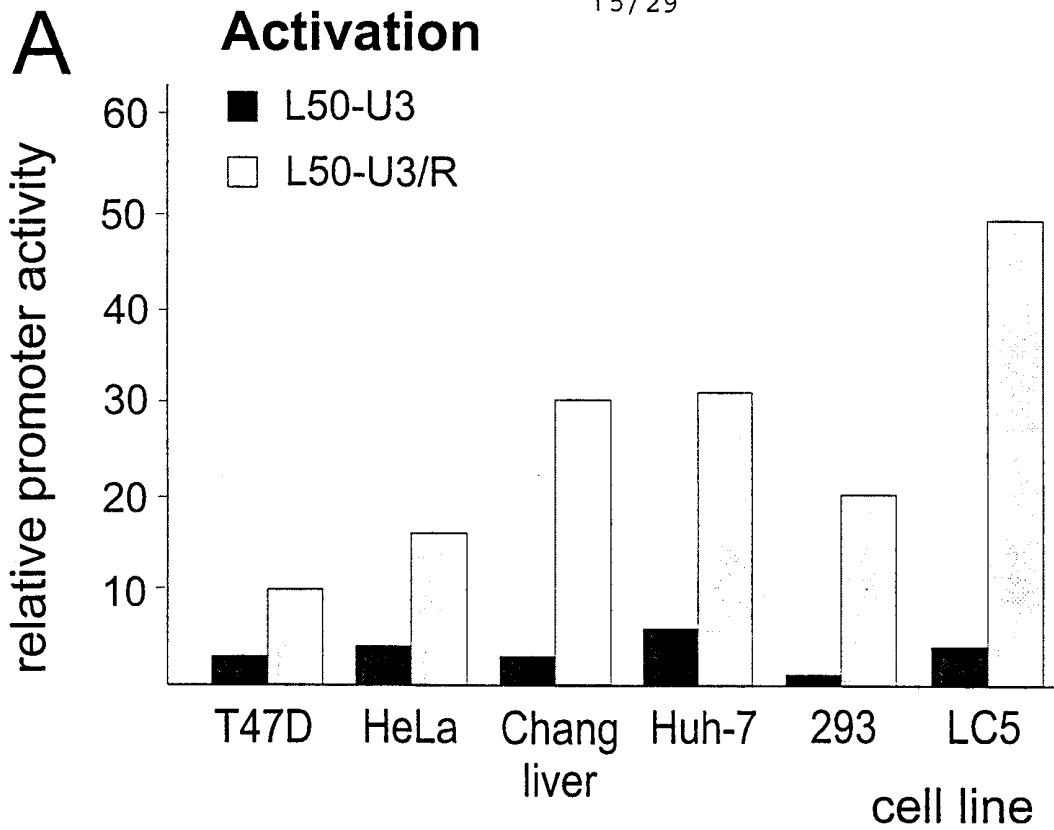


Fig. 4: LTR-R region modulates promoter activity of HERV-K-T47D related LTRs

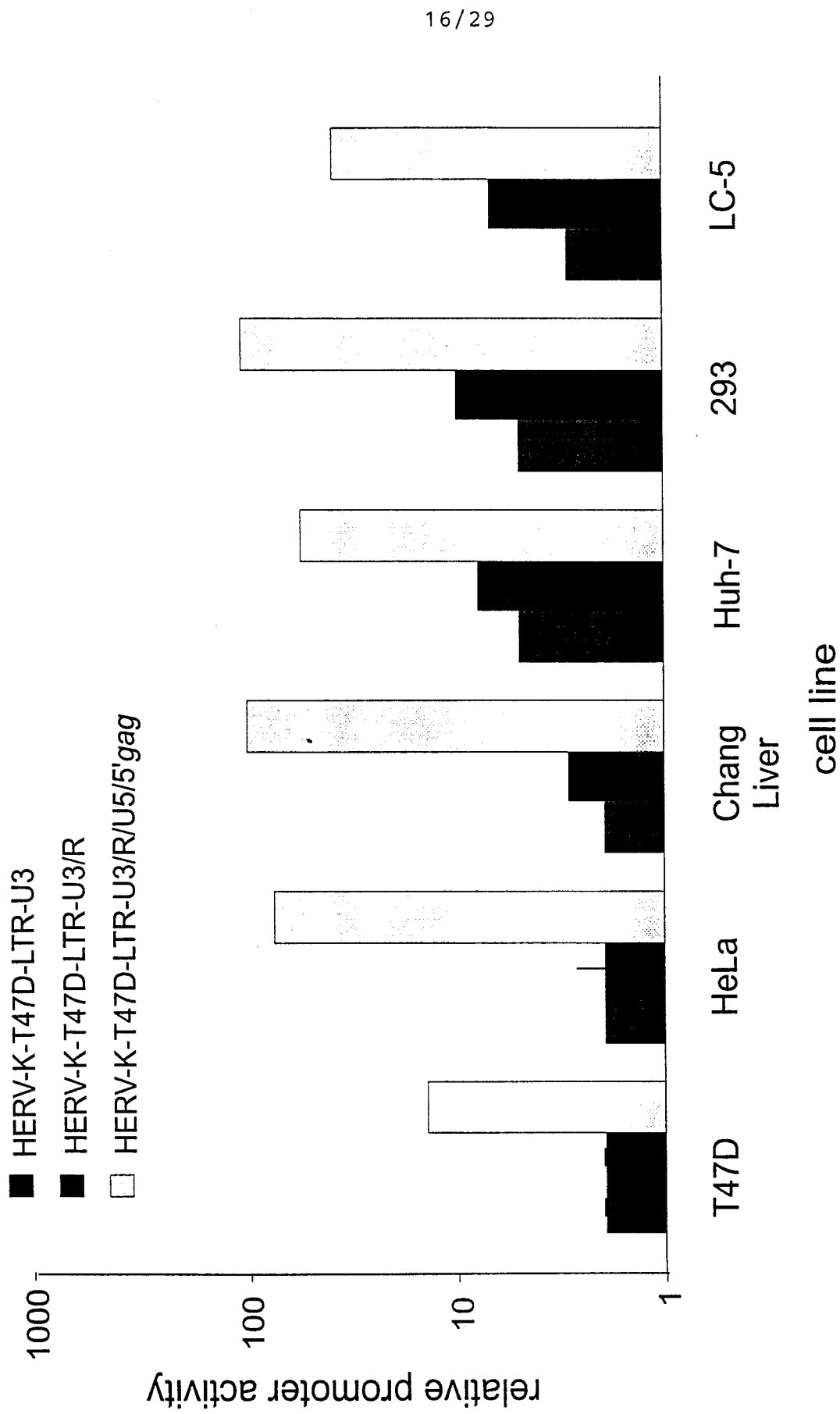


Fig. 5: Sequences downstream of LTR-R modulate promoter activity of HERV-K-T47D related LTRs

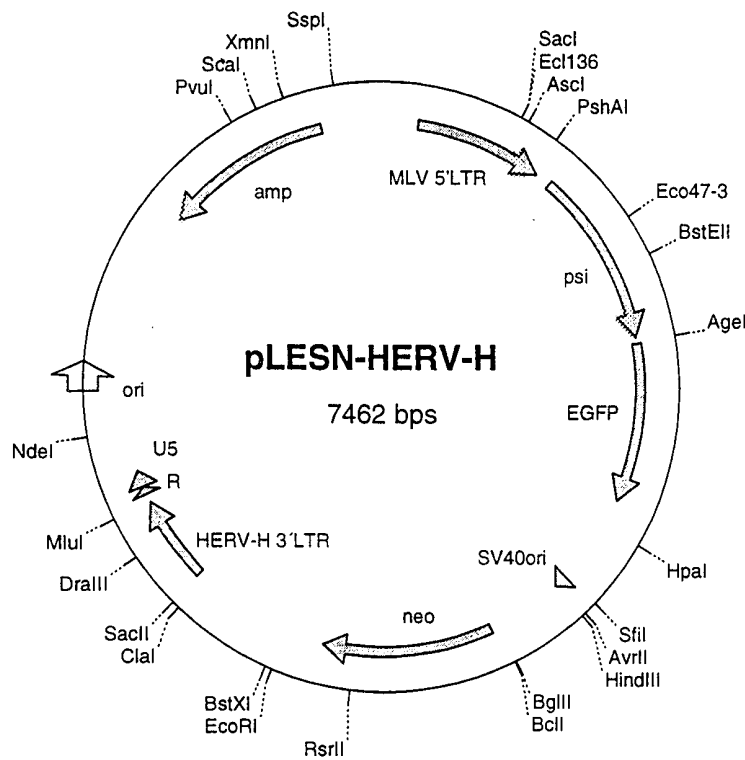
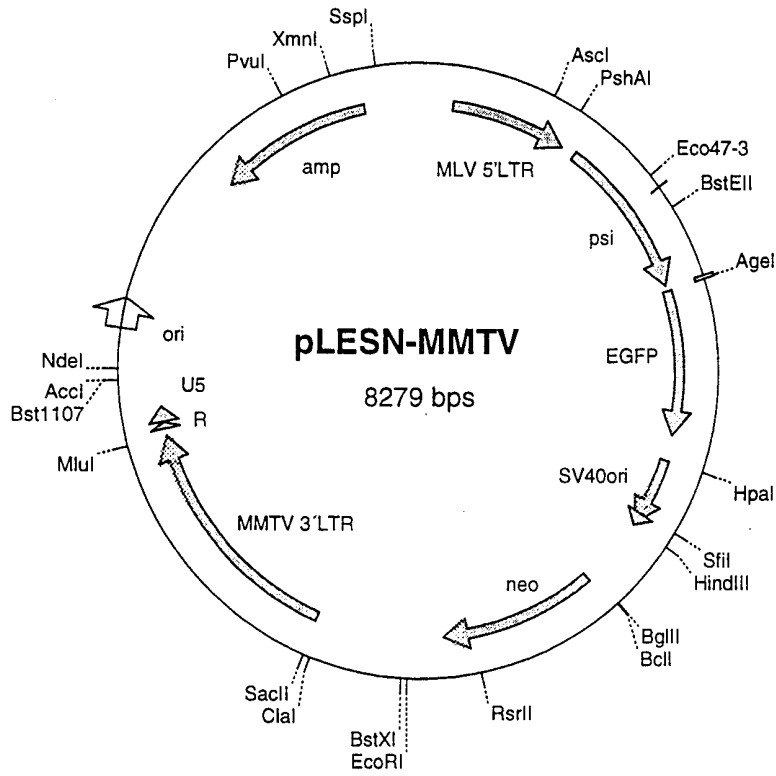


Fig.7: Retroviral ProCon vectors pLESN-MMTV and pLESN-HERV-H

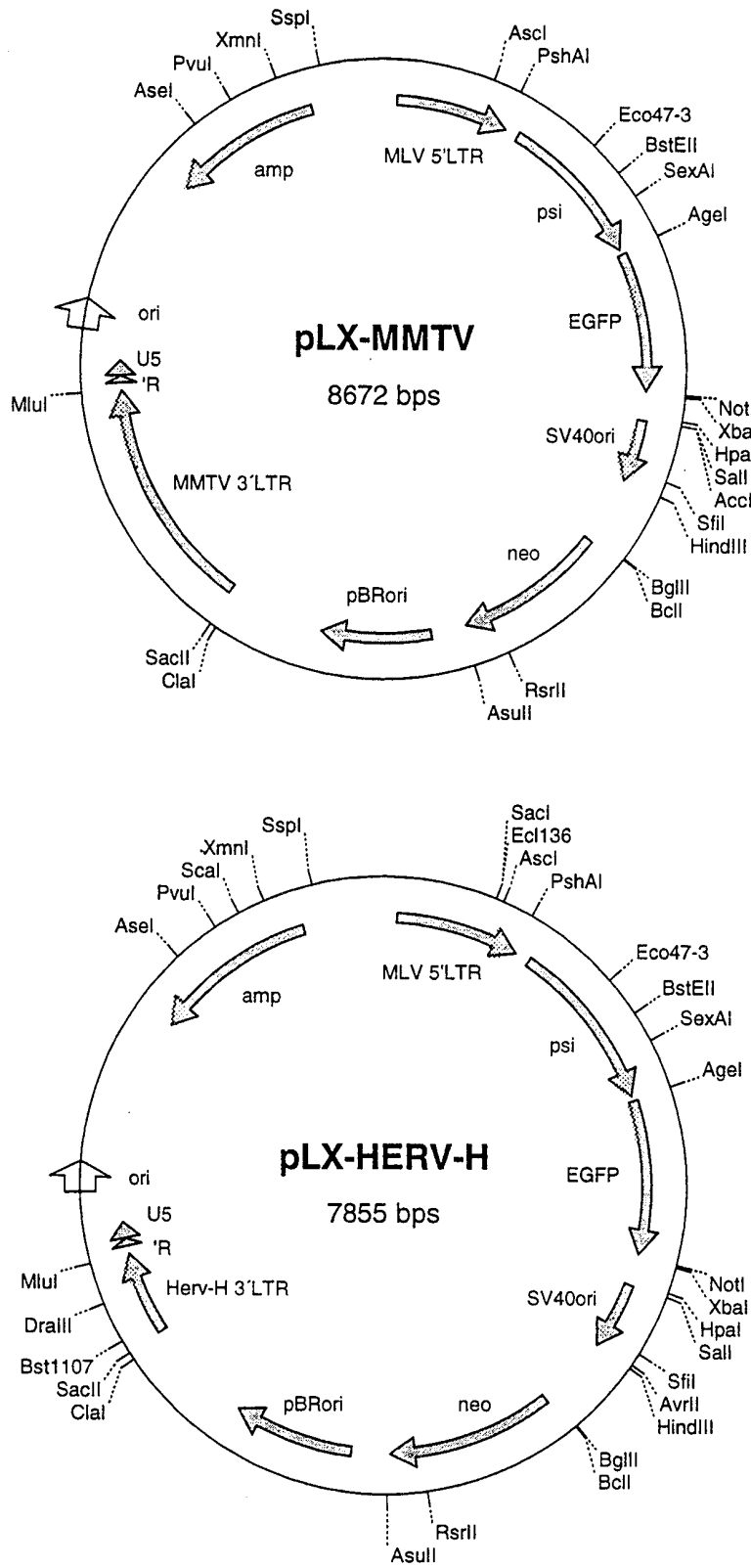
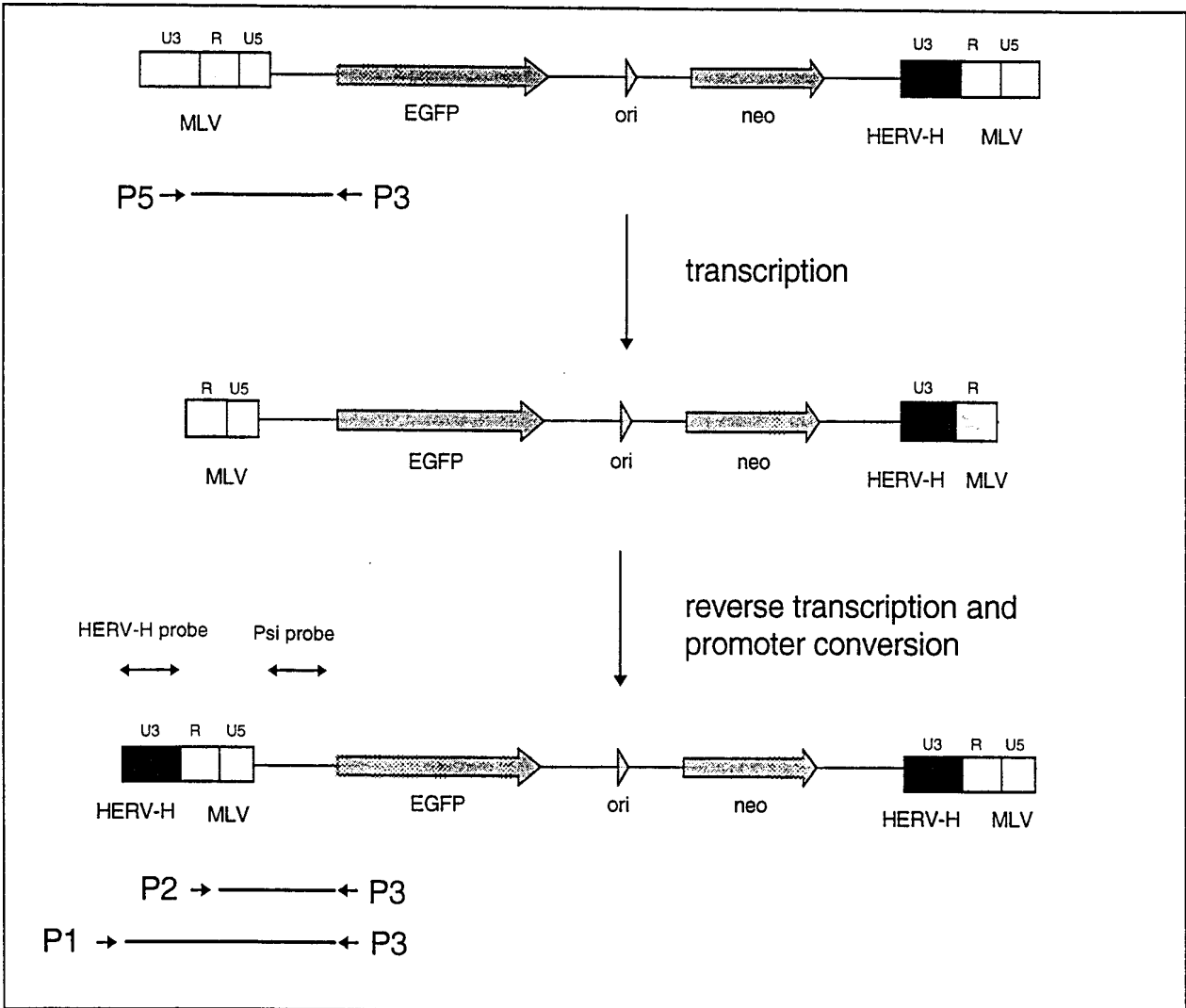


Fig.8: Retroviral ProCon vectors pLX-MMTV and pLX-HERV-H

a)

20/29



b)

HERV-H probe

Psi probe

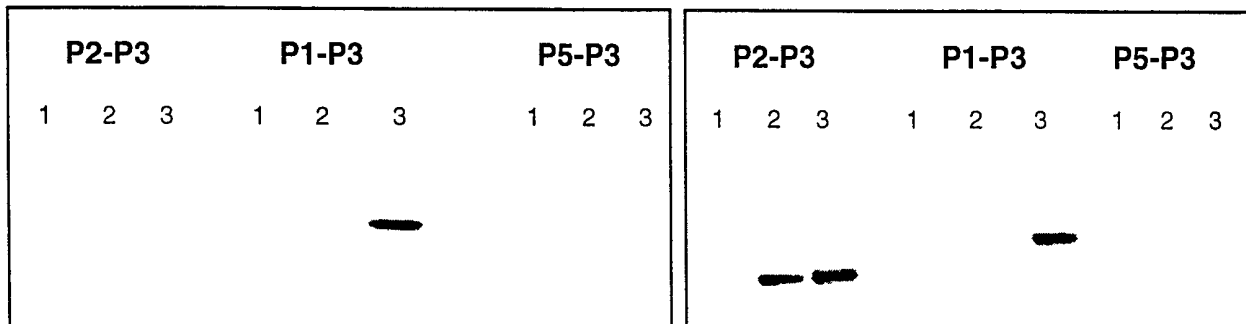
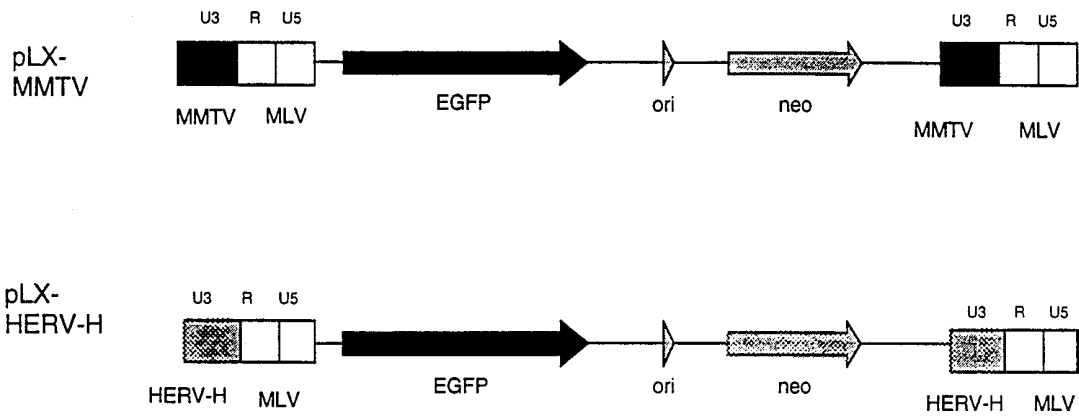


Fig. 9: a) Promoter conversion of the hybrid ProCon vectors

b) Demonstration of the correct promoter conversion with PCR and hybridization with a HERV-H and a psi probe (1:CK; 2:CK-pLX-MMTV; 3:CK-pLX-HERV-H)

a)



b)

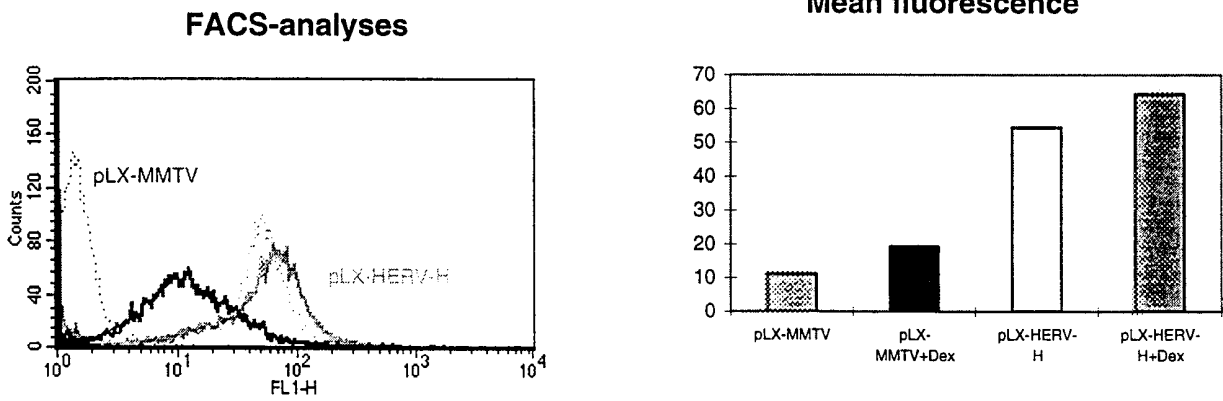


Fig.10: a) organization of the two ProCon vectors pLX-MMTV and pLX-HERV-H
 b) promoter activity of the HERV-H LTR in comparison to the MMTV-LTR by infection of CrfK cells

Appendix

A. HERV-H LTR sequences

	1				50
HERV-H L31	TGTCAGGCCT	CTGAGCCCAA	GCTAAGCCAT	CATATCCCCT	GTGACCTGCA
HERV-H HCM2	TGTCAGGCCT	CTGAGCCCAA	GCCAGGCCAT	CGCATCCCCT	GTGACTTGCA
HERV-H 19	TGTCAGGCCT	CTGAGCCCAA	GCTAAGCCAT	CATATCCCCT	GCGACCTGCA
HERV-H MP20	TGTCAGGCCT	CTGAGCCCAA	GCTAAGCCAT	CATATCCCCT	GTGACCTGCA
HERV-H CM3	TGTCAGGCCT	CTGAGCCCAA	GCCAAGCCAT	CGCATCCCCT	GTGACTTGCA
HERV-H MC16	TGTCAGGCCT	CTGAGCCCAA	GCC.	TGCA
HERV-H CM1	TGTCAGGCCT	CTGAGCCCAA	GCCAAGCCAT	CGCATCCCCT	GTGACTTGCA
HERV-H MP23	TGTCAGGCCT	CTGAGCCCAA	GCCAAGCCAT	CGCATCCCCT	GTGACTTGCA
HERV-H H13	TGTCAGGCCT	CTGAGCCCAA	GCTAAGCCAT	CATATCCCCA	GGGACCTGCA
HERV-H H1	TGTCAGGCCT	CTGAGCCCAA	GCTAAGCCAT	CAAATCCCCT	GTGACCTGCA
HERV-H HU8	TGTCAGGCCT	CTGAGCCCAA	GCTAAGCCAT	CATATCCC..	GTGACCTGCA
HERV-H PA7	TGTCAGGCCT	CTGAGCCCAA	GCTAAGCCAT	CAAATCCCCT	GTGACCTACA
	51				100
HERV-H L31	CGTATACATC	CAGATAGCC.TGAAG	CAACTG....
HERV-H HCM2	CGTATACATC	CAGATGGCC.TAAAG	TAAGTGAAGATCCA
HERV-H 19	CATATACATC	CAGATGGCC.TGAAG	TAAGTGAAGAATCA
HERV-H MP20	CGTACACATC	CAGATGGCCG	GTTCCCTGCCT	TAAGTGAAGA	CATTCCACCA
HERV-H CM3	CGTGTATGCC	CAGATGGCC.TGAAG	TAAGTGAAGAATCA
HERV-H MC16	CGTATACATC	CAGATG....AAG	CAAGTGAAGAATCA
HERV-H CM1	CGTATACGCC	CAGATGGCC.TGAAG	TAAGTGAAGAATCA
HERV-H MP23	CGTATACGCC	CAGATGGCC.TGAAG	TAAGTGAAGAATCA
HERV-H H13	CGTATACATC	CAGATGGCC.TGAAG	CAAGTGAAGATCCA
HERV-H H1	CGTGTACATC	CAGATGACC.TGAAG	CAAGTGAAGATCCA
HERV-H HU8	..TATACATC	CAGATGGCC.TGAAG	CAAGTGAAGATCCA
HERV-H PA7	CGTGTACATC	CAGATGACC.TGAAG	CAAGTGAAGATCCA
	101				150
HERV-H L31T	AAAAATATCC	TAACTGATG	ACA.....	..TTCCAATA
HERV-H HCM2	CAAAGAAGT	AAAAACAGCC	TAACTGATG	ACA.....	..TTCCAACA
HERV-H 19	CAAAGAAGT	GAAAATGGCC	TGTTCC....
HERV-H MP20	CGAAAGAAGT	GAAAATGACC	TGTTCC....
HERV-H CM3	CAAAGAAGT	GAAAAGGCC	TGCCCC....
HERV-H MC16	CAAAGAAGT	GAAAATGGCC	GGTTCC....
HERV-H CM1	CAAAGAAGT	GAAAAGGCC	TGCCCCGCCT	TAACTGATGA	CATTCCACCA
HERV-H MP23	CAAAGAAGT	GAAAAGGCC	TGCCCCGCCT	TAACTGATGA	CATTCCACCA
HERV-H H13	CAAAGGAAGT	GAAAATAGCC	TAACTGATG	ACA.....	..TTCCAACA
HERV-H H1	CAAAGAAGT	GAAAGTAGCC	TAACTGATG	ACA.....	..TTCCAACA
HERV-H HU8	CAAAGGAAGT	GAAAATAGCC	TAACTGATG	ACA.....	..TTCCAACA
HERV-H PA7	CAAAGAAGT	GAAAGTAGCC	TAACTGATG	ACA.....	..TTCCAACA
	151				200
HERV-H L31	TTGTGATTTG	TTTCTGCCCT	ACCCTGACTG	ATCAATGTGC	TTTGTAAATCT
HERV-H HCM2	TTGTGATTTG	TTCCTGCCCC	ACCCTAACTG	ATAAATGTAC	TTTGTAAATCT
HERV-H 19T	GCCTTAACTG	ATGACATTAC	CTTGTGAAAT
HERV-H MP20T	GCCTTAACTG	ATGACATTGT	CTTGTGAAAT
HERV-H CM3	ACCTTAACTG	AGTGATTAAC	CCCATGAATT
HERV-H MC16T	GCCTTAACTG	ATGACATTAC	CTTGTGAAAT
HERV-H CM1	TGGTGATTTG	TTCTTGCCCC	ACCTTAACTG	AGTGATTAAC	CCTGTGAATT
HERV-H MP23	TGGTGATTTG	TTCTTGCCCC	ACCTTAACTG	AGTGATTAAC	CCTGTGAATT
HERV-H H13	TTGTGATTTG	TTTCTGCCCT	ATCCTAACTG	ATCAATGTAC	TTTGTAAATCT
HERV-H H1	TTGTGATTTG	TTCCTGCCCC	ACGCTAACTG	AT.....AC	CATATATTCT
HERV-H HU8	TTGTGATTTG	TTCCTGCCCC	ACGCTAACTG	AT.....AC	CATATATTCT
HERV-H PA7	TTGTGATTTG	TTCCTGCCCT	ACGCTAGCTG	AT.....AC	CATATATTCT
	201				250
HERV-H L31	CCCCCACCCT	TCAGAAGGCT	CTTTGTAATC	CTCCCCACCC	TTGAGAATGG
HERV-H HCM2	CCCCCACCCT	TAAGAAGGTC	CTTTGTAATT	CTCCCCACCC	TTGAGAGTGT
HERV-H 19	TCCTTCTCCT	GGCTCATCCT	GGCTCAAAAG	CTC..CCGCA	CTGAG....C
HERV-H MP20	TCCTTCTCCT	GGCTCATCCT	GGCTCAAAAG	CTC..CCGCA	CTGAG....T
HERV-H CM3	TCCTTCCCCT	GGCTCAG...AAG	CTC..CCCCA	CTGAG....C
HERV-H MC16	TCCTTCTCCT	GGCTCAG...AAG	CTC..CCCCA	CTGAG....C

HERV-H CM1	TGCTTCTCCT	GGCTCAG...AAG	CTC..CCCCA	CTGAG....C
HERV-H MP23	TGCTTCTCCT	GGCTCAG...AAG	CTC..CCCCA	CTGAG....C
HERV-H H13	CTCCCACCCT	TAAGAAGGTT	CTTTGTAATT	CTCCCCACCC	TTGAGAGTGT
HERV-H H1	TCCCC.....CGCCC	TTGAGAATGT
HERV-H HU8	TCCCC.....CGCCC	TTGAGAATGT
HERV-H PA7	TCCCC.....CGCCC	TTGAGAATGT

	251				300
HERV-H L31	ACTTGGTGAG	ATCCACCCCC	TGCCTGCAAA	GCATTGCCCC	TAACTCCACC
HERV-H HCM2	ACTTTGTGAG	ATCCACACCT	GCCCACCAGA	GAACAAACCC	CCTTTGACTG
HERV-H 19	ACCTTGTGAC	CCCTGCCTCT	GCCCGCCAGA	GAGCAACCCC	CTCTTGACTG
HERV-H MP20	ACATTGTGAC	CCCCACTCCT	GCCCGCCAGA	GAACAGCCCC	CT.TTGACTG
HERV-H CM3	ACCTTGTGAC	CCCTGCCCCCT	GCCCACCAGA	GAACAAACCC	CT.TTGACTG
HERV-H MC16	ACCTTGTGAC	CCCCACTCCT	GCCCGCCAGA	GAACAAACCC	CT.TTGACTG
HERV-H CM1	ACCTTGTGAC	CCCCGCCCCCT	GCCCACCAGA	GAACAAACCC	CT.TTGACTG
HERV-H MP23	ACCTTGTGAC	CCCCGCCCCCT	GCCCACCAGA	GAACAGACCC	CT.TTGACTG
HERV-H H13	ACTTTGTGAG	ATCCACCCCC	TGCCGGCAAA	ACATTGCTCC	TAACCCAACC
HERV-H H1	ACTTTGTA..C
HERV-H HU8	ACTTTGTA..C
HERV-H PA7	ACTTTGTA..C

	301				350
HERV-H L31	GCCTGTCCCA	AAACCTATGA	GAA.CTAATG	ATA.....	ATCCC.ACCA
HERV-H HCM2	TAATTTTCCA	TTACCTTCCC	TAATCCTATA	AAACGGCCCC	ACCCC.ATCT
HERV-H 19	TAATTTTCCCT	TTACCTACCT	AAATCCTATA	AAATGGCCCC	ACTCCTATCT
HERV-H MP20	TAATTTTCCCT	TTATCTACCC	AAATCCTATA	AAACAGCCCC	ACCTTTATCT
HERV-H CM3	TAATTTTCCA	TTACTTTCCC	AAATCCTATA	AAACGGCCCC	ACCCCTATCT
HERV-H MC16	TAATTTTCCA	CTGCCCCGCC	AAACCCATA	AAACGGTCCC	ACCCC.ATCT
HERV-H CM1	TAATTTTCCA	TTACCTTCCC	AAATCCTATA	AAACGGCCCC	ACCCCTATCT
HERV-H MP23	TAATTTTCCA	TTACCTTCCC	AAATCCTATA	AAACGGCCCC	ACCCCTATCT
HERV-H H13	GCCTA.CCCC	AAACCTGTAA	GAA.CTAATG	ATA.....	ATCC..ACCA
HERV-H H1	ACCTATCCC.	AAACCTATAA	GAA.CTAATG	ATA.....	ATCCT.ACCA
HERV-H HU8	ACCTATCCC.	AAACCTATAA	GAA.CTAATG	ATA.....	ATCC..ACCA
HERV-H PA7	ACCTATCCC.	AAACCTATAA	GAA.CTAATG	ATA.....	ATCCT.ACCA

	351				400
HERV-H L31	CACCTTGCTG	ACTCTCTTTC	C...AGACTC	AGCCCGGCTG	CACCCAGGTG
HERV-H HCM2	CCCTTTGCTG	ACTCTCTTTT	C...GGACTC	AGCCCGCCTG	CACCCAGGTG
HERV-H 19	CCCTTCGCTG	ACTCTCTTTT	C...GGACTC	AGCCCGCCTG	TACCCAGGTG
HERV-H MP20	CCCTTGGCTG	ACTCTCTTTT	C...GGACTC	AGCCCGCCTG	CACCCAGGTG
HERV-H CM3	CCCTTCGCTG	ACTCTCTTTT	C...GGACTC	AGCCCGCCTG	CACCCAGGTG
HERV-H MC16	CCCTTCCCTG	ACTCTCTTTT	CTTCGGACTC	AGCCCGCCTG	CACCCAGGTG
HERV-H CM1	CCCTTCGCTG	ACTCTCTTTT	C...GGACTC	AGCCCGCCTG	CCCCCAGGTG
HERV-H MP23	CCCTTCGCTG	ACTCTCTTTT	C...GGACTC	AGCCCGCCTG	CCCCCAGGTG
HERV-H H13	CCCTTTGCTG	ACTC..TTTT	C...AGAATC	AGCCCGCCTG	CACCCAGGTG
HERV-H H1	CCCTTTGCTG	ACTCTCTTTT	T...GGACTC	AGCCCGCCTG	CACCCAGGTG
HERV-H HU8	CCCTTTGCTG	ACTCTCTTTT	T...GGACTC	AGCCCGCCTG	CACCCAGGTG
HERV-H PA7	CCCTTTGCTG	ACTCTCTTTT	T...GGACTC	AGCCCGCCTG	CACCCAGGTG

	401		425	
HERV-H L31	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H HCM2	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H 19	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H MP20	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H CM3	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H MC16	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H CM1	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H MP23	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H H13	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H H1	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H HU8	AAATAAACAG	CCATGTTGCT	CACAT	
HERV-H PA7	AAATAAACAG	CCATGTTGCT	CACAT	

B. HERV-W LTR sequences

	1				50
HERV-T47D-W2	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGCTGGATTT	CCTAGGCCGA
HERV-T47D-W4	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGCTGGATTT	CCTAGGCCGA
HERV-T47D-W5	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	ACCTGGATTT	CCTAGGCCGA
HERV-W1	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGCTGGATTT	CCTAGGCCAA
HERV-W10	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGCTGGATTC	CCTAGGCCGA
HERV-W11	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGTTGGATTT	CCTAGGCTGG
HERV-W18	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGTTGGATTT	CCTAGGCCGG
HERV-W2	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGCTGGATTT	CCTAGGCCAA
HERV-W22	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGCTGGATTT	CCTAGGCTGA
HERV-W23	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGCTGGATTT	CCTAGGCTGA
HERV-W4	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGTTGGATTT	CCTAGGCTGG
HERV-W5	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	ACCTGGATTT	CCTAGGCCAA
HERV-W6	TGTTGAGATG	GGGGACTGAG	AGACAGGACT	AGCTGGATTT	CCTAGGCCAA
HERV-W8	TGTTGAGATG	GGGGACTGAG	AAACAGGACT	AGCAGGATTT	CCTAGGCCGA

	51				100
HERV-T47D-W2	CTAAGAATCC	CTAAGCCTAG	CTGGGAAGGT	GACCGCATCC	ACCTTTAAAC
HERV-T47D-W4	CTAAGAATTC	CTAAGCCTAG	CTGGGAAGGT	GACCGCATCC	ATCTTTAAAC
HERV-T47D-W5	CTAAGAATCC	CTAAGCCTAG	CTGGGAAGGT	GACCACATCC	ACCTTTAAAC
HERV-W1	CTAAGAATCC	CTAAGCCTAG	CTGGGAAGGT	GACTACACCC	ACCTTTAAAC
HERV-W10	CTAAGAATCC	CTAAGCCTAG	CTGGGAAGGT	GACCACATCC	ACCTTTAAAC
HERV-W11	CTAAGAATCC	CTAAGCCTAG	CTGGGAAATT	GACCACGTCC	ACCTTTAAAC
HERV-W18	CTAAGAATCC	CTAAGCCTAG	CTGGGAAATT	GACCACGTCC	ACCTTTAAAC
HERV-W2	CTAAGAATCC	CTAAGCCTAG	CTGGGAAGGT	GACTACACCC	ACCTTTAAAC
HERV-W22	CTAAGAATCC	CTAAGCCTAG	CTGGGAAGGT	GACCGCATCC	ATCTTTAAAC
HERV-W23	CTAAGAATCC	CTAAGCCTAG	CTGGGAAGGT	GACTACACCC	ACCTTTAAAC
HERV-W4	CTAAGAATCC	CTAAGCCTAG	CTGGGAAATT	GACCACGTCC	ACCTTTAAAC
HERV-W5	CTAAGAATCT	CTAAGCCTAG	CTGGGAAGGT	GACCACATCC	ACCTTTAAAC
HERV-W6	CTAAGAATCC	CTAAGCCTAG	CTGGGAAGGT	GACTACACCC	ACCTTTAAAC
HERV-W8	TTAAGAATCC	CTAAGCCTAG	ATGGGAAGTT	GACCACATCC	ACCTTTAAAC

	101				150
HERV-T47D-W2	ACGGGGCTTG	CAACTTAGCT	CACACCCAAC	CAATCAGGTA	GTAAAGAGGG
HERV-T47D-W4	ATGGGGCTTG	CAACTTAACT	CATATCTGAC	CAATCAGGTA	GTAAAGAGAG
HERV-T47D-W5	ACAGGGCTTG	CAACTTAGCT	CACACTTGAC	CAGTCAGGTA	GTAAAGAGAG
HERV-W1	ATGGGGCTTG	CAACTTAGCT	CACACCCAAC	CAATCAGGTA	GTAAAGAGAG
HERV-W10	ACGGGGCTTG	CAACTTAGCT	CATACCCAAC	AAATCAGGTA	GTAAAGAGAG
HERV-W11	ACGGGGCTTG	CAATTTAGCT	CACACCCGAC	CAATCAGGTA	GTAAAGGGAG
HERV-W18	ACGGGGCTTG	CAATTTAGCT	CACACCCGAC	CAATCAGGTA	GTAAAGGGAG
HERV-W2	ACTAGGCTTG	CAACTTAGCT	CACACCCGAC	CAATCAGGTA	GTAAAGAGAG
HERV-W22	ATGGGGCTTG	CAACTTAACT	CATATCTGAC	CAATCAGGTA	GTAAAGAGAG
HERV-W23	ACTAGGCTTG	CAACTTAGCT	CACACCCGAC	CAATCAGGTA	GTAAAGAGAG
HERV-W4	ACGGGGCTTG	CAATTTAGCT	CACACCCGAC	CAATCAGGTA	GTAAAGGGAG
HERV-W5	ACAGGGCTTG	CAACTTAGCT	CACACCCGAC	CCATCAGGTA	AGAAAGAGAG
HERV-W6	ACTAGGCTTG	CAACTTAGCT	CACACCCGAC	CAATCAGGTA	GTAAAGAGAG
HERV-W8	ACGGGGCTTG	CAACTCAGCT	CACACCCGAC	CCATCAGGTA	AGAAAGAGAG

	151				200
HERV-T47D-W2	CTCACTAAAA	TGCTAATTAG	GCAAAGACAG	GAGGTAAAGA	AATAGCCAAT
HERV-T47D-W4	CTCACTAAAA	TGCTAATTAG	GCTAAAACAG	GAGGCAAAGA	AGTAGCCAAT
HERV-T47D-W5	CTCACTAAAA	TGCTAATTAG	GCTAAAACAG	GAGGTAAAGA	AATAGACAAT
HERV-W1	CTTGCTAAAA	TGCTAATTAG	GCAAAAACAG	GAGGTAAAGA	AATAGCCAGT
HERV-W10	CTCACTAAAA	TACTGATTAG	GCGAAAACAG	GAGGTAAGGA	AACAGCCAGT
HERV-W11	CTCACTAAAA	TGCTAATTAG	GGAAAACAG	GAGGTAAAGA	AGTAGCCAAT
HERV-W18	CTCACTAAAA	TGCTAATTAG	GGAAAACAG	GAGGTAAAGA	AGTAGCCAAT
HERV-W2	CTTGCTAAAA	TGCTAATTAG	GCAAAAACAG	GAGGTAGAGA	AATAGCCAAT
HERV-W22	CTTGCTAAAA	TGCTAATTAG	GCAAAAACAG	GAGGTAAAGA	AATAGCCAGT
HERV-W23	CTTGCTAAAA	TGCTAATTAG	GCAAAAACAG	GAGGTAAAGA	AATAGCCAGT
HERV-W4	CTCACTAAAA	TGCTAATTAG	GGAAAACAG	GAGGTAAAGA	AGTAGCCAAT
HERV-W5	CCCCTAAAA	TGCTAATTAG	GCAAAAACAG	GAGGTAAAGA	AATAGTCAAT
HERV-W6	CTTGCTAAAA	TGCTAATTAG	GCAAAAACAG	GAGGTAAAGA	AATAGCCAGT
HERV-W8	CCCCTAAAA	TGCTAATTAG	GCAAAAACAG	GAGGTAAAGA	AATAGCCAAT

	201				250
HERV-T47D-W2	CATCTATTGC	CTGAGAGCAC	AGCAGGAGGG	ACAATGATCG	GGATATAAAC
HERV-T47D-W4	CATCTGTTGC	CTGACAGCAC	AGCAGGAGGG	ACAATGATCG	GGATATAAAC
HERV-T47D-W5	CATCTATCAC	CTGAGAGCAC	AGTGGGAGGG	ACAATGATCG	GCATATAAAC
HERV-W1	CATCTATCGC	CTGACAGCAC	AAGGGGCGGG	ACAATGATCA	GGATATAAAC
HERV-W10	CATCTATCGC	CTGACAGCAC	AAGGGGCGGG	ACAATGATCA	GGATATAAAC
HERV-W11	CATCTATCGC	CTGAGAGCAC	AACAGGAGGG	ACAATGATCA	GGATATAAAC
HERV-W18	CATCTATCGC	CTGAGAGCAC	AACAGGAGGG	ACAATGATCA	GGATATAAAC
HERV-W2	CATCTATCGC	CTGAGAGCAC	AGCAGGAGGG	ACAATGATCC	GGATATAAAC
HERV-W22	CATCTATCGC	CTGACAGCAC	AAGGGGCGGG	ACAATGATCA	GGATATAAAC
HERV-W23	CATCTATCGC	CTGACAGCAC	AAGGGGCGGG	ACAATGATCA	GGATATAAAC
HERV-W4	CATCTATCGC	CTGAGAGCAC	AACAGGAGGG	ACAATGATCA	GGATATAAAC
HERV-W5	CATCTATTGC	CTGAGAGCAC	AGCGGGAGGG	ACAATGATCA	GGATATAAAC
HERV-W6	CATCTATCGC	CTGACAGCAC	AAGGGGCGGG	ACAATGATCA	GGATATAAAC
HERV-W8	CATCTATTGC	CTGAGAGCAC	AGCGGGAGGG	ACAATGATCA	GGATATAAAC

	251				300
HERV-T47D-W2	CCAAGTCTTC	GAGCCGGCAA	TGGCTACCCT	CTTTGGGTCC	CCTCCCTTTG
HERV-T47D-W4	CCAGGCATTC	GAGCCAGCTA	CAGCTACCCT	CTTTGGGTCC	CCTCCCTTTG
HERV-T47D-W5	CCAGGCATTC	GAGCCAGCAA	CAGCAACCCC	CTTTGGG...
HERV-W1	TCAGGCATTC	AAGCCAGCAA	TGGCTACCCA	CTTTGGGTCC	CCTCCCATTT
HERV-W10	TCAGGCATTC	AAGCCAGCAA	TGGCTACCCA	CTTTGGGTCC	CCTCCCATTT
HERV-W11	CCAGGCATTC	AAGCCAGCGG	TGGCTACCCT	CTTTGGGTCC	CCTCCCTTTG
HERV-W18	CCAGGCATTC	AAGCCAGCGG	TGGCTACCCT	CTTTGGGTCC	CCTCCCTTTG
HERV-W2	CCAAGCATTC	GAGCCAGCAA	TGGCTACCCT	CTTTGGGTCC	CCTCCCTTTG
HERV-W22	TCAGGCATTC	AAGCCAGCAA	TGGCTACCCA	CTTTGGGTCC	CCTCCCATTT
HERV-W23	TCAGGCATTC	AAGCCAGCAA	TGGCTACCCA	CTTTGGGTCC	CCTCCCATTT
HERV-W4	CCAGGCATTC	AAGCCAGCGG	TGGCTACCCT	CTTTGGGTCC	CCTCCCTTTG
HERV-W5	CCAGGCATTC	GAGCCGGCAA	CGACTACCCT	CTTTGGGTCC	CCTCCCTTTG
HERV-W6	TCAGGCATTC	AAGCCAGCAA	TGGCTACCCA	CTTTGGGTCC	CCTCCCATTT
HERV-W8	CCAGGCATTC	GAGCCGGCAA	CGACTACCCT	CTTTGGGTCC	CCTCCCTTTG

	301				343
HERV-T47D-W2	TATGGGAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	C..
HERV-T47D-W4	TATGGGAGCT	CTGTCTTCAC	TCTATTAAAT	CTTGCAACTG	C..
HERV-T47D-W5AGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	C..
HERV-W1	TATGGGAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W10	TATGGGAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W11	TATGGGAGCC	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W18	TATGGAAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W2	TATGGGAGCT	CTATTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W22	TATGGGAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W23	TATGGGAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W4	TATGGAAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W5	TATGGGAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W6	TATGGGAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA
HERV-W8	TATGGGAGCT	CTGTTTTTCAC	TCTATTAAAT	CTTGCAACTG	CAA

C. HERV-K LTR sequences

```

1.....50
HERV-K45 GCGACCGGT: GGATC:CCGG GCCCGCGG:T ACCGTCGACT :GCAGAATTC
HERV-K27 GCGACCGGT: GGATC:CCGG GCCCGCGG:T ACCGTCGACT :GCAGAATTC
HERV-K2 GCGACCGGT: GGATC:CCGG GCCCGCGG:T ACCGTCGACT :GCAGAATTC
HERV-K1 GCGACCGGT: GGATC:CCGG GCCCGCGG:T ACCGTCGACT :GCAGAATTC
HERV-K30 GTC CCACCTCCAG CCCTAAGGCG GTTTTTCCCT ATCTCAGTAG
HERV-K10 AGTAG

51.....100
HERV-K45 ATGGAGCATA CAATCGGGTT TTATACCGAG ACATTCCATT GCCCAGGGAC
HERV-K27 ATGGAGCATA CAATCGGGTT TTATACCGAG ACATTCCATT GCCCAGGGAC
HERV-K2 ATGGAGCATA CAATCGGGTT TTATACCGAG ACATTCCATT GCCCAGGGAC
HERV-K1 ATGGAGCATA CAATCGGGTT TTATACCGAG ACATTCCATT GCCCAGGGAC
HERV-K30 ATGGAGCATA CAATCGGGTT TTATACCGAG ACATTCCATT GCCCAGGGAC
HERV-K10 ATGGAGCATA CAATCGGGTT TTATACCGAG ACATTCCATT GCCCAGGGAC

101.....150
HERV-K45 AGGCAGGAGA CAGATGCCTT CCTCTTGTCT CAACTGCAAG AGGCATTCCCT
HERV-K27 AGGCAGGAGA CAGATGCCTT CCTCTTGTCT CAACTGCAAG AGGCATTCCCT
HERV-K2 AGGCAGGAGA CAGATGCCTT CCTCTTGTCT CAACTGCAAG AGGCATTCCCT
HERV-K1 AGGCAGGAGA CAGATGCCTT CCTCTTGTCT CAACTGCAAG AGGCATTCCCT
HERV-K30 AGGCAGGAGA CAGATGCCTT CCTCTTGTCT CAACTGCAAG AGGCATTCCCT
HERV-K10 AGGCAGGAGA CAGATGCCTT CCTCTTGTCT CAACTGCAAG AGGCATTCCCT

151.....200
HERV-K45 TCCTCTTATA CTAATCCTCC TCAGCACAGA CCCTTTACGG GTGTCGGGCT
HERV-K27 TCCTCTTATA CTAATCCTCC TCAGCACAGA CCCTTTACGG GTGTCGGGCT
HERV-K2 TCCTCTTATA CTAATCCTCC TCAGCACAGA CCCTTTACGG GTGTCGGGCT
HERV-K1 TCCTCTTATA CTAATCCTCC TCAGCACAGA CCCTTTACGG GTGTCGGGCT
HERV-K30 TCCTCTTATA CTAATCCTCC TCAGCACAGA CCCTTTACGG GTGTCGGGCT
HERV-K10 TCCTCTTTTA CTAATCCTCC TCAGCACAGA CCCTTTACAG GTGTCGGGCT

201.....250
HERV-K45 GGGGGACGGT CAGGTCTTTC CCTTCCCACG AGGCCATATT TCAGACTATC
HERV-K27 GGGGGACGGT CAGGTCTTTC CCTTCCCACG AGGCCATATT TCAGACTATC
HERV-K2 GGGGGATGGT CAGGTCTTTC CCTTCCCACG AGGCCATATT TCAGACTATC
HERV-K1 GGGGGACGGT CAGGTCTTTC CCTTCCCACG AGGCCATATT TCAGACTATC
HERV-K30 GGGGGACGGT CAGGTCTTTC CCTTCCCACG AGGCCATATT TCAGACTATC
HERV-K10 GGGGGACGGT CAGGTCTTTC CCTTCCCACG AGGCCATATT TCAGACTATC

251.....300
HERV-K45 ACATGGGGAG AAACCTTGGA CAATACCTGG CTTTCCTAGG CAGAGGTCCC
HERV-K27 ACATGGGGAG AAACCTTGGA CAATACCTGG CTTTCCTAGG CAGAGGTCCC
HERV-K2 ACATGGGAAG AAACCTTGGA CAATACCTGG CTTTCCTAGG CAGAGGTCCC
HERV-K1 ACATGGGGAG AAACCTTGGA CAATACCTGG CTTTCCTAGG CAGAGGTCCC
HERV-K30 ACATGGGGAG AAACCTTGGA CAATACCTGG CTTTCCTAGG CAGAGGTCCC
HERV-K10 ACATGGGGAG AAACCTTGGA CAATACCTGG CTTTCCTAGG CAGAGGTCCC

301.....350
HERV-K45 TGCGGCCTTC CGCAGTTTTT GTGT:CCTGG GACTTTGAGA TTAGGGAGTG
HERV-K27 TGCGGCCTTC CGCAGTTTTT GTGT:CCTGG GACTTTGAGA TTAGGGAGTG
HERV-K2 TGCGGCCTTC CGCAGTTTTT GTGT:CCTGG GACTTTGAGA TTAGGGAGTG
HERV-K1 TGCGGCCTTC CGCAGTTTTT GTGT:CCTGG GACTTTGAGA TTAGGGAGTG
HERV-K30 TGCGGCCTTC CGCAGTTTTT GTGTCC:TGG GACTTTGAGA TTAGGGAGTG
HERV-K10 TGCGGCCTTC TGCACTTTTT GTGTCCCTGG GACTTTGAGA TTAGGGAGTG

351.....400
HERV-K45 GTGATGACTC TTAAGGAGCA TGCTGCCTTC AAGCATCTGT TTAACAAAGC
HERV-K27 GTGATGACTC TTAAGGAGCA TGCTGCCTTC AAGCATCTGT TTAACAAAGC
HERV-K2 GTGATGACTC TTAAGGAGCA TGCTGCCTTC AAGCATCTGT TTAACAAAGC
HERV-K1 GTGATGACTC TTAAGGAGCA TGCTGCCTTC AAGCATCTGT TTAACAAAGC
HERV-K30 GTGATGACTC TTAAGGAGCA TGCTGCCTTC AAGCATCTGT TTAACAAAGC
HERV-K10 GTGATGACTC TTAAGGAGCA TGCTGCCTTC AAGCATCTGT TTAACAAAGC
    
```

401.....450
HERV-K45 ACATCCTGCA CCGCCCTTAA TCCATTCAAC CCTGAGTTGA CACAGCACAC
HERV-K27 ACATCCTGCA CCGCCCTTAA TCCATTCAAC CCTGAGTTGA CACAGCACAC
HERV-K2 ACATCCTGCA CTGCCCTTAA TCCATTCAAC CCTGAGTTGA CACAGCGCAC
HERV-K1 ACATCCTGCA CCGCCCTTAA TCCATTCAAC CCTGAGTTGA CACAGCACAC
HERV-K30 ACATCCTGCA CCGCCCTTAA TCCATTCAAC CCTGAGTTGA CACAGCACAC
HERV-K10 ACATCCTGCA CCGCCCTTAA TCCATTCAAC CCTGAGTTGA CACAGCACAT

451.....550
HERV-K45 GTTTCAGAGA GCACGGGGTT GGGGGTAAGG TCATAGATTA ACAGAATCTC
HERV-K27 GTTTCAGAGA GCACGGGGTT GGGGGTAAGG TCATAGATTA ACAGAATCTC
HERV-K2 GTTTCAGAGA GCACGGGGTT GGGGGTAAGG TCATAGATTA ACAGAATCTC
HERV-K1 GTTTCAGAGA GCACGGGGTT GGGGGTAAGG TCATAGATTA ACAGAATCTC
HERV-K30 GTTTCAGAGA GCACGGGGTT GGGGGTAAGG TCATAGATTA ACAGAATCTC
HERV-K10 GTTTCAGAGA GCACGGGGTT GGGGGTAAGG TCATAGATTA ACAGAATCTC

501.....550
HERV-K45 AAGGCAGAAG AATTTTTCTT AACACATAAC AAAATGGAGT CTCCCATGTC
HERV-K27 AAGGCAGAAG AATTTTTCTT AACACATAAC AAAATGGAGT CTCCCATGTC
HERV-K2 AAGGCAGAAG AATTTTTCTT AACACATAAC AAAATGGAGT CTCCCATGTC
HERV-K1 AAGGCAGAAG AATTTTTCTT AACACATAAC AAAATGGAGT CTCCCATGTC
HERV-K30 AAGGCAGAAG AATTTTTCTT AACACATAAC AAAATGGAGT CTCCCATGTC
HERV-K10 AAGGCAGAAG AATTTTTCTT AGCACATAAC AAAATGGAGT CTCCATGTC

551.....600
HERV-K45 TACTTCTTTC TACACAGACA CAGTAACAAT CTGATCTCTC TTGCTTTTCC
HERV-K27 TACTTCTTTC TACACAGACA CAGTAACAAT CTGATCTCTC TTGCTTTTCC
HERV-K2 TACTTCTTTC TACACAGACA CAGTAACAAT CTGATCTCTC TTGCTTTTCC
HERV-K1 TACTTCTTTC TACACAGACA CAGTAACAAT CTGATCTCTC TTGCTTTTCC
HERV-K30 TACTTCTTTC TACACAGACA CAGTAACAAT CTGATCCCTC TTGCTTTTCC
HERV-K10 TACTTCTTTC TACACAGACA CAGTAACAAT TTGATCTCTC TTGCTTTTCC

601.....650
HERV-K45 CCACATTTCC CCCTTTTCTT TTCG
HERV-K27 CCACATTTCC CCQTTTTCTT TTCGA
HERV-K2 CCACATTTCC CCCTTTTCTT TTCGACAAA
HERV-K1 CCACATTTCC CCCTTTTCTT TTCGACAAA CCGCCAT:CT CGAGATC:TG
HERV-K30 CCACATTTCC CCCTTTTCTT ATCCATCACA CTGGCGGCCG CTCGAGCATG
HERV-K10 CCACATTTCC CCCTTTTCTT TTCGACAAA CCGCCATC

651.....
HERV-K1 AGT
HERV-K30 CATCTAGAGG GCCCAATTCG CCCTATAGTG

HERV-K-T47D-5'LTR

TGTGGGCGAAGGATTACCCAGGTGCCGAGGCAAGAGACTGAAGGCACAACTGTTTCAGTATAATATAGAAAATAGCTAG
AATAAGAAATAGTTATAATAAAAAATTAGATATACACATGATCATGGACATTACCAATCATTACTACAAACATTGTTAATCA
TTAGCTTTTAAATATTACTCTTTGTTTATTACTAATATAACCAAGGAATAACCGGTAGCATACGGTCAGGTGCTGAAGGG
ACATTGTGAGAAGTGACCTAGAAGGCAAGAGGTGAGCCTTCTGTCACGCCGCATAAGGACAGCTTGAGGGCTCCTTGGT
CAAGCTGTAACACCAGTGCCTGGGAAGGCACCGTTACTTAGCAGACCATGAAAGGGAGTCTCCATTCCCTGGAGGAGTCA
GGAAACACTATGCTCCACCAGCTTCTTGTGTATCCAGCCCTGCCACAGTCATCCAGAGGCATAAACCCCTCCCTGTGG
TGCTGTGCTTCAATGGCCATGCTTCTTGTCCACTTTCATGTTCCCTCCTGTACTCCTGGTTCCTCTTTGAAGTTCGTAGAA
GATAATGGTAGAAGAAATAGTAAAAGTCTTTGATCTTTCTTATAAGTGCATAGAAGAAAACACTGATGTATGCCTGCCTT
CCCTCTCTGCTCAGCTACCTAAAAGGAAAGGCCCCCTTTCCCATGATCACATGACTTGCTGACCTTATCAATCACTTG
GAGGACTCACCCCTCCTTACCCTGTCCCTTTGTCTTGTATGCAATAAATATCAGCACGCCAGCCATTCCGGGGCCACTACT
GGTCTCCGCAACTTGGTGGTAGTGGTACCCTGGGCCAGCTGTTTTCTCTTTATCTCTTTTGTCTTGTGCTTTATTTCT
TACAATCTCTCATCTCTGCACATGGGAGAACACCGGCAAAGCCCGTAGGGCTGGACCTTACA

L48-LTR (U3-R)

TGTGGGCGGAAGAGTACCTAGGTGCCGAGGCAAGAGACTGAAGGCACAACTGTTTCAGTATAATAAAGAAAATAGAATA
AGAATAGTCATAATACAAATTAGATACAGCGATGATCATGAACAATTATCCATCATTATTATAAACATTATTAATCATT
GCTTTTAAATATTACTCTGTTGCATTAAATAATATAACCTAGGAATAACCGGCAGGTATAGGGTCAGGTGCTGAAGGGACAT
TGTGAGAAGTGAATAGAAGGCAAGAGGGGAGCCTTCTGTCATGCCCGCATAAGGGCCGCTTGAGGGCCCTTGGTCAAGC
GGTAACGCCAGTGTCTGGGAAGGCACCGTTACTGAGCAGACCCGGAAAGGGAGTCTCCTTTCCCTGGAGGAGTCAAGGA
ACGCTCTGCTCCACCAGCTTCTTGTGGGAGGCTGGATGTTACCCAGGCTGCCTGCAGTCATCCGGAGGCCGTAACCCCT
CCCTGTGGTGCCTCAATGGTACGTTCCCTTGTCCACTTTCATGCTCCTTCCGTACTCCTGGTTCCTCTTTGAAGTTCGTA
GTAGATAGCGGTAGAAGAAATAGTAAAAGTCTTAAAGTCTTTGATCTTATAAGTTCATAGAAGAAAACGCTGATGCCTGC
CGCCTTCTCTCTCTGCTTACAGCTACCTAAGAGGGAAGGGCCCGCTGTCTGTGATCAGGTGACTTGCTTACCTTGTCAA
TCACTTAGAAGACTGACCCCTCCTTATCCTGCCCCCTTGTCTTGTATGCAATAAATATCAGCGAGCCCAGCCGTTACAGGGC
CACTACCGGTCTCCGTGTCTTTGTGGTAGTGGTCCCCGGGCCAGCTGTTTTCTCTTT

L5-LTR (U3-R)

TGTGGGTGGAGGATTACCCAGGTGCCAAGGCAAGAGACTGAAGGCACAACTGTTTCAGTATAATAAAAAAATAGAATA
AGAATAGTCATAATACAAATTAGATATAGAGATGATCATGGACAATTAGCAATCACTATTAATCTTTAGCTTTTAAATATT
ACTCTTTGTTGCATTACTAATAAACCCTAGGAATAACCGGTGGGTATAGGGTCAGGTGCTGAAGGGACATTGTGTGAAGT
GACCTGGAAGGCAAGAGGTGAGCCCTCTGTCACGCCACATAAGGGCCGCTTGAGGGCTCCTTGGTCAAGTGGTAACGCC
AGTGTCTGGGAATGCACCCGTTAATTAGCAGACCCGAAAGGGAGTCTCCTTTCCCTGGAAGAGTTGGGGAACACTCTGC
TCCACCAGCTTCTTTGGAAGGCTGGATATATCCAGGCCCTGCCGCGAGTCATCCGGAGGCTTAAACCCCTCCCTGTGGT
GCTGTGCTTCAATGGTCCCCTCCTTGTCCACTTTCATGCTCCTCCCGTACTCCTGGTTCCTCTTTGAAGAGCGCAGTAG
ATAGCGGTAGAAGAAATAGTAAAAGTCTTAAAGTCTTTCGATCTTTCTTACAAGTGCAGAGAAGAAAACGCTGACATATGC
TGCCCTCCCTCTCTGCTTCGGCTACCTAAAAGGGAAGGGCCCGCTATCCTGTAATCACATGACTTGCTTACCTTGTCAA
TCACTTAGAAGATTCACTCTCCTTACCCTGCCCCCTTGTCTTGTATGCAATAAATATCAGTGACCCAGCCGTTACAGGGC
CACTACTGGTCTCCGCTCTTGTAGTGGTACCCCGGCC

L50-LTR (U3-R)

TGTGGGTGGAGGATTACCCAGGTGCCAAGGCAAGAGACTGAAGGCACAACTGTTTCAGTATAATAAAAAAATAGAATA
AGAATAGTCATAATACAAATTAGATATAGAGATGATCATGGACAATTAGCAATCACTATTAATCTTTAGCTTTTAAATATT
ACTCTTTGTTGCATTACTAATAAACCCTAGGAATAACCGGTGGGTATAGGGTCAGGTGCTGAAGGGACATTGTGAGAAGT
GACCTGGAAGGCAAGAGGTGAGCCCTCTGTCACGCCACATAAGGGCCGCTTGAGGGCTCCTTGGTCAAGTGGTAACGCC
AGTGTCTGGGAATGCACCCGTTAATTAGCAGACCCGAAAGGGAGTCTCCTTTCCCTGGAAGAGTTGGGGAACACTCTGC
TCCACCAGCTTCTTGTGGAAGGCTGGATATATCCAGGCCCTGCCGCGAGTCATCCGGAGGCTTAAACCCCTCCCTGTGGT
GCTGTGCTTCAATGGTCCCCTCCTTGTCCACTTTCATGCTCCTCCCGTACTCCTGGTTCCTCTTTGAAGAGCGCAGTAG
ATAGCGGTAGAAGAAATAGTAAAAGTCTTAAAGTCTTTCGATCTTTCTTACAAGTGCAGAGAAGAAAACGCTGACATATGC
TGCCCTCCCTCTCTGCTTCGGCTACCTAAAAGGGAAGGGCCCGCTATCCTGTAATCACATGACTTGCTTACCTTGTCAA
TCACTTAGAAGATTCACTCTCCTTACCCTGCCCCCTTGTCTTGTATGCAATAAATATCAGTGACCCAGCCGTTACAGGGC
CACTACTGGTCTCCGCTCTTGTAGTGGTACCCCGGCCAGGTGTTTTCTTT

L9-LTR (966 nt)

TGTGGGTGGAGGATTACCCAGGTGCCGAGGCAAGAGACTGAAGGCACAAACTGTTTCAGTATAATAAAGAAAAATGGTTAG
 AATAAGAATAGTCATAATACAAATTAGATATAGAGATGATCATGGACAATTATCAATCATTATTATAAACATTATTAATC
 ATTAGCTTTTAATATTACTCTTTGTTGCATTACTAATATAACCTAGGAATAACCGGTGGGTATAGGGTCAGGTGCTGAAA
 GGACATTGGGAGAAGTGACCTAGAAGGCAAGAGGTGAGTCTTCTGTCACGCCCGCATAAGGGTTGCTTGAGGGCTCCTTG
 GTCAAGTGGTAACGCCGGTGTCTGGGAAGGCACCTGTTACTTAGCCGACCACGAAAGGGAGTCTCCTTTCTTGAGGAG
 TCAGGGCGCACTCTGCTCCACCAGCTTCTTGTGGAAGGCTGGATATTATCCAGGCCCTGCCCGCAGTCATCCGGAGGCCTA
 AACCCCTCCCTGTGGTGTCTGTGCTTCAATGGGCACACTCCTCGTCCACTTTCATGTTCTCCCATACTCCTGGTTCTCT
 TTGAAGTTCGTAGTAGATAGTGGTAGAAGGAATAGGGAAAAATCTTAAAGTGTGTTGATCTTTCTTATAAGTGCATAGAAGA
 AAACGCTGACATATGCTGCCTTCTCTGTCTGCTTACCTAACAGAGGGAAGGGCCCCCTGTCCAGTGATCACGTGACT
 TGCTTACCTTGTCAATCACTTAGAAGATTCACCTCCTTACCCTGCCCCCTTGTCTTGTATGCAATAAATATCAGTGCA
 CCCAGCCTTTTCGGGGCCACTTACCAGTCTCCACGTCTTGGTGGTAGTGGTCCCCCGGGCCAGCTGTTTTCTCTTTATCT
 CTTTGTCTTGTGTCTTATTTATTACAATCTCTCGTCTCCGCACACAGGGAGAACACCCCGCTAAGCTCCGTAGGGCTGGAC
 CCTACA

L8-LTR (938 nt)

TGTGGGTGGAGGATTACCCAGGTGCCGAGGCAAGAGACTGAAGGCACAAACTGTTTCAGTATAATAAAGAAAAATGGTTAG
 AATAAGAATAGTCATAATACAAATTAGATATAGAGATGATCATGGACAATTATCAATCATTATTATAAACATTATTAATC
 ATTAGCTTTTAATATTACTCTTTGTTGCATTACTAATATAACCTAGGAATAACCGGTGGGTATAGGGTCAGGTGCTGAA
 GGACATTGGGAGAAGTGACCTAGAAGGCAAGAGGTGAGTCTTCTGTCACGCCCGCATAAGGGTTGCTTGAGGGCTCCTTG
 GTCAAGTGGTAACGCCGGTGTCTGGGAAGGCACCTGTTACTTAGCCGACCACGAAAGGGAGTCTCCTTTCTTGAGGAG
 TCAGGGCACACTCTGCTCCACCAGCTTCTTGTGGAAGGCTGGATATTATCCAGGCCCTGCCCGCAGTCATCCGGAGGCCTA
 AACCCCTCCCTGTGGTGTCTGTGCTTCAATGGGCACACTCCTCGTCCACTTTCATGTTCTCCCATACTCCTGGTTCTCT
 TTGAAGTTCGTAGTAGATAGTGGTAGAAGGAATAGGGAAAAATCTTAAAGTGTGTTGATCTTTCTTATAAGTGCATAGAAGA
 AAACGCTGACATATGCTGCCTTCTCTGTCTGCTTACCTAACAGAGGGAAGGGCCCCCTGTCCAGTGATCACGTGACT
 TGCTTACCTTGTCAATCACTTAGAAGATTCACCTCCTTACCCTGCCCCCTTGTCTTGTATGCAATAAATATCAGTGCA
 CCCAGCCTTTTCGGKKCACTTACCAGTCTCCACGTCTTGGTGGTAGTGGTCCCCCGGGCCAGCTGTTTTCTCTTTATCTCT
 TTGTCTTGTGTCTTATTTATTACAATCTCTCGTCTCCGCACACAGGGAGAACACCCCG (Abbruch 26 nt vor
 Ende der LTR)

L49-LTR = L20-LTR (963 nt)

TGTGGGCGAAAGATTACCTAGGTGCCGAGGCAAGAGACTGAAGGCACAAACTGTTTCAGTATAATAAAGAAAAATAGTTAA
 AATAAGAATAGTTATAATACAAATTAGATATAGAGATGATCATGGACAATTATCAATCATTATTATAAACATTAAATCATT
 AGCTTTTAATATTACTCTTTGTTGCTTTACTAATATAACCTAGGAATAACCGGTGGGTATAGGGTCAGGTGTTGACGGGA
 TATTGTGAGAAGTGACCTAGAAGGCAAGAGGTGAGCCTTCTGTCACGCCACATAAGGGCCGCTTGAGGGCTCTTTGGTC
 AAGTGGTAACGCCAGTGTCTGTGAAGGCACCTGTTACTTAGCAGACCAGGAAAGGGAGTCTCCTTTCTTGAGGAGTCA
 GGGAACACTCTGCTCCACCAGCTTCTTGTGGAAGGCTGGATATTATCTAGGCCCTGCCCGCAGTCATCTGGAGGCCTAAAC
 CCTCCCTGTGGTGTCTGTCTCAGTGGTCACTCTCCTTGTCCACTTTCATGTTCTCCGTACTIONCTGGTTCTCTTTG
 AAGTTCGTAGTAGATAGCAGTAGAAGAAATAGTAAAGTCTTAAAGTATTTGATCTTTCTTATAAGTGCATAGAAGAAAA
 CGCTGACATATGCTGCCTTCTATCTCTGCGGTGGCTACCTAAAAGGGAAGGGCCCCCTGTCCCATGATCATGTGACTT
 GCTTACCTTATCACTTAGAAGATTATCCTCCTTACCCTGCGCCCCCTCGTCTTGTATGCAATAAATATCAGCAGGCC
 AGTCGTTTGGAGCCACTGCCGGTCTCCCGCTCTTGGTGGTAGTGGTCCCCCGGGCCAGCTATTGTCTCTTTATCTCTTT
 GTCTTGTGTCTTTATTTATTACAATCTCTTGTCTCTGCACACAGGGAGAACACCTGCTAAGCCCCGTAGGACTGGACCT
 ACA

HERV-IP-T47D

TGTTCAATCTTTGCCTTCTACTTTTAACTTAACTTCCCTATAAAGCAACCTTTTTCAATCACCTGCTCCACTCTGACT
 CATTCTGATCACCTGCTCCACCTGACTCATTCCGATCACCTGATCCACTGTGACTCATTCCGATTACCCGCTCCACCT
 GACTCAATCTGATTTGATTTCTGCTCTGCCATAACCATTTTTCCCGCCAAACCACTCACCTGTCACTCTCTTTAAAT
 TAGCCAATTGGAATTAGTTTAGCCTGTGCGGTCAACCTAGCCAATAGGGGACTGACACAGCAGCAGGGGCCACATGTG
 TCAGGAATAAGACCCCTTCCCCTCCCTGCTCCAGATGTGTGCTCACCATTGCTCCATCTGTGAGGGCACACCCCTTATA
 GAAGTAAATTGCCTTGCTGAGAAGAAAAAAGAACATTTTATATTCAAGTCCATTTTCTTTTGTGTCACCGAACTTTA
 TTTATAACA

SEQUENZPROTOKOLL

<110> GSF - Forschungszentrum für Umwelt und Gesundheit

<120> Retrovirale Expressionsvektoren auf der Basis von
HERV-LTR-Sequenzen.

<130> P11045

<140> DE 199 10 650.9

<141> 1999-03-10

<160> 47

<170> PatentIn Ver. 2.1

<210> 1

<211> 375

<212> DNA

<213> Human endogenous retrovirus

<400> 1

```

tgtcaggcct ctgagcccaa gctaagccat catatcccct gtgacctgca cgtatacatc 60
cagatagcct gaagcaactg taaaaatata ctttaactgat gacattccaa tattgtgatt 120
tgtttctgcc ctaccctgac tgatcaatgt gctttgtaat ctccccacc cttcagaagg 180
ctctttgtaa tcctccccac ccttgagaat ggacttgggtg agatccacc cctgcctgca 240
aagcattgcc cctaactcca ccgcctgtcc caaacctat gagaactaat gataatccca 300
ccacactttg ctgactctct ttccagactc agccccggctg caccaggtg aaataaacag 360
ccatgttgct cacat 375

```

<210> 2

<211> 400

<212> DNA

<213> Human endogenous retrovirus

<400> 2

```

tgtcaggcct ctgagcccaa gccaggccat cgcatacccct gtgacttgca cgtatacatc 60
cagatggcct aaagtaactg aagatccaca aaagaagtaa aaacagcctt aactgatgac 120
attccaacat tgtgatttgt tcctgccccca ccctaactga taaatgtact ttgtaatctc 180
ccccaccctt aagaaggtcc tttgtaattc tccccaccct tgagagtgta ctttgtgaga 240
tccacacctg cccaccagag aacaaaacccc ctttgactgt aattttccat taccttccct 300
aatcctataa aacggcccca ccccatctcc ctttgctgac tctcttttcg gactcagccc 360
gcctgcaccc aggtgaaata aacagccatg ttgctcacat 400

```

<210> 3

<211> 361

<212> DNA

<213> Human endogenous retrovirus

<400> 3

```

tgtcaggcct ctgagcccaa gctaagccat catatcccct gcgacctgca catatacatc 60
cagatggcct gaagtaactg aagaatcaca aaagaagtga aaatggcctg ttctgcctt 120
aactgatgac attaccttgt gaaattcctt ctctggctc atcctggctc aaaagctccc 180
gcactgagca ctttgtgacc cctgcctctg ccgcagag agcaaccccc tcttgactgt 240
aattttcctt tacctaccta aatcctataa aatggcccca ctctatctc ctttgcctga 300
ctctcttttc ggactcagcc cgctgttacc caggtgaaat aaacagccat gttgctcaca 360
t 361

```

<210> 4

<211> 372

2

<212> DNA

<213> Human endogenous retrovirus

<400> 4

```

tgtcaggcct ctgagcccaa gctaagccat catatcccct gtgacctgca cgtacacatc 60
cagatggccg gttcctgcct taactgatga cattccacca cgaaagaagt gaaaatgacc 120
tgttcgtgcc ttaactgatg acattgtctt gtgaaattcc tcctcctggc tcatcctggc 180
tcaaaagctc cccgactgag tacattgtga cccccactcc tgcccgccag agaacagccc 240
cctttgactg taattttcct ttatctaccc aaatcctata aaacagcccc acctttatct 300
cccttgctg actctctttt cggactcagc ccgctgcac ccaggtgaaa taaacagcca 360
tgttgctcac at 372

```

<210> 5

<211> 349

<212> DNA

<213> Human endogenous retrovirus

<400> 5

```

tgtcaggcct ctgagcccaa gccaaagccat cgcaccccct gtgacttgca cgtgtatgcc 60
cagatggcct gaagtaactg aagaatcaca aaagaagtga aaaggccctg ccccaccta 120
actgagtgat taaccccatg aatttccttc ccctggctca gaagctcccc cactgagcac 180
cttgtgaccc ctgcccctgc ccaccagaga acaaccccct ttgactgtaa tttccatta 240
ctttccaaa tctataaaa cggccccacc cctatctccc ttcgctgact ctcttttcgg 300
actcagcccc cctgcaccca ggtgaaataa acagccatgt tgctcacat 349

```

<210> 6

<211> 324

<212> DNA

<213> Human endogenous retrovirus

<400> 6

```

tgtcaggcct ctgagcccaa gcctgcacgt atacatccag atgaagcaag tgaagaatca 60
caaaagaagt gaaaatggcc ggttctctgcc ttaactgatg acattacctt gtgaaattcc 120
ttctcctggc tcagaagctc cccccactgag caccttgtga cccccactcc tccccgccac 180
agaacaacc cctttgactg taattttcca ctgcccgcc aaaccctata aaacggtccc 240
acccatctc ccttccctga ctctctttt ttcggactca gcccgcctgc acccaggtga 300
aataaacagc catgttgctc acat 324

```

<210> 7

<211> 393

<212> DNA

<213> Human endogenous retrovirus

<400> 7

```

tgtcaggcct ctgagcccaa gccaaagccat cgcaccccct gtgacttgca cgtatacgcc 60
cagatggcct gaagtaactg aagaatcaca aaagaactga aaaggccctg ccccgccta 120
actgatgaca ttccaccatg gtgatttgtt cttgcccac ctttaactgag tgattaacc 180
tgatgaattg cttctcctgg ctcaagaagct cccccactga gcaccttggt acccccgcc 240
ctgcccacca gagaacaacc ccctttgact gtaattttcc attaccttc caaatcctat 300
aaaacggccc caccctatc tcccttctgt gactctctt tcggactcag cccgcctgcc 360
cccagtgaa ataacagcc atgttgctca cat 393

```

<210> 8

<211> 393

<212> DNA

<213> Human endogenous retrovirus

<400> 8

```

tgtcaggcct ctgagcccaa gccaaagccat cgcaccccct gtgacttgca cgtatacgcc 60

```

3

```

cagatggcct gaagtaactg aagaatcaca aaagaactga aaaggcctg ccccgctta 120
actgatgaca ttccaccatg gtgatttggt cttgccccac cttactgag tgattaaccc 180
tgtgaatttg cttctcctgg ctccagaagct cccccactga gcacctgtg acccccgcc 240
ctgcccacca gagaacagac ccccttgact gtaattttcc attacctcc caaatcctat 300
aaaacggccc caccctatc tccctcgct gactctcttt tcggactcag cccgcctgcc 360
cccaggtgaa ataacagcc atgttgctca cat 393

```

```

<210> 9
<211> 388
<212> DNA
<213> Human endogenous retrovirus

```

```

<400> 9
tgtcaggcct ctgagcccaa gctaagccat catatcccca gggacctgca cgtatacatc 60
cagatggcct gaagcaactg aagatccaca aaggaagtga aaatagcctt aactgatgac 120
attccaccat tgtgatttgg ttctgcccc tccctactga tcaatgtact ttgtaatctc 180
tcccaccctt aagaaggttc tttgtaattc tccccaccct tgagagtgtg ctttgtgaga 240
tccaccctt gccggcaaaa cattgctcct aaccacaaccg cctaccccaa acctgtaaga 300
actaatgata atccaccacc ctttgtgac tcttttcaga atcagcccgc ctgacccag 360
gtgaaataaa cagccatggt gctcacat 388

```

```

<210> 10
<211> 314
<212> DNA
<213> Human endogenous retrovirus

```

```

<400> 10
tgtcaggcct ctgagcccaa gctaagccat caaatcccct gtgacctgca cgtgtacatc 60
cagatgacct gaagcaactg aagatccaca aaagaagtga aagtagcctt aactgatgac 120
attccaccat tgtgatttgg tctgccccg cgctaactga taccatata tcttcccccg 180
cccttgagaa tgtactttgt acacctatcc caaacctata agaactaatg ataactctac 240
cacccttgc tgactctctt tttggactca gcccgctgc acccaggtga aataaacagc 300
catgttgctc acat 314

```

```

<210> 11
<211> 309
<212> DNA
<213> Human endogenous retrovirus

```

```

<400> 11
tgtcaggcct ctgagcccaa gctaagccat catatcccgt gacctgcata tacatccaga 60
tggcctgaag caactgaaga tccacaaagg aagtgaatag agccttaact gatgacattc 120
caccattgtg atttgttctt gccccacgct aactgatacc atatattctt cccccgccct 180
tgagaatgta ctttgtacac ctatcccaaa cctataagaa ctaatgataa tccaccacc 240
tttgctgact ctcttttgg actcagcccg cctgcaccca ggtgaaataa acagccatgt 300
tgctcacat 309

```

```

<210> 12
<211> 314
<212> DNA
<213> Human endogenous retrovirus

```

```

<400> 12
tgtcaggcct ctgagcccaa gctaagccat caaatcccct gtgacctaca cgtgtacatc 60
cagatgacct gaagcaactg aagatccaca aaagaagtga aagtagcctt aactgatgac 120
attccaccat tgtgatttgg tctgccccta cgctagctga taccatata tcttcccccg 180
cccttgagaa tgtactttgt acacctatcc caaacctata agaactaatg ataactctac 240
cacccttgc tgactctctt tttggactca gcccgctgc acccaggtga aataaacagc 300
catgttgctc acat 314

```

<210> 13
<211> 341
<212> DNA
<213> Human endogenous retrovirus

<400> 13
tgttgagatg ggggactgag agacaggact agctggattt cctaggccga ctaagaatcc 60
ctaagcctag ctgggaaggt gaccgcatcc acctttaaac acggggctcg caacttagct 120
cacaccaac caatcaggta gtaaagaggg ctactaaaa tgctaattag gcaaagacag 180
gaggtaaaga aatagccaat catctattgc ctgagagcac agcaggaggg acaatgatcg 240
ggatataaac ccaagtcttc gagccggcaa tggctacctt ctttgggtcc cctccctttg 300
tatgggagct ctgttttcac tctattaat cttgcaactg c 341

<210> 14
<211> 341
<212> DNA
<213> Human endogenous retrovirus

<400> 14
tgttgagatg ggggactgag agacaggact agctggattt cctaggccga ctaagaattc 60
ctaagcctag ctgggaaggt gaccgcatcc atctttaaac atggggcttg caacttaact 120
catatctgac caatcaggta gtaaagagag ctactaaaa tgctaattag gctaaaacag 180
gaggcaaaga agtagccaat catctgttgc ctgacagcac agcaggaggg acaatgatcg 240
ggatataaac ccaggcattc gagccagcta cagctaccct ctttgggtcc cctccctttg 300
tatgggagct ctgtcttcac tctattaat cttgcaactg c 341

<210> 15
<211> 322
<212> DNA
<213> Human endogenous retrovirus

<400> 15
tgttgagatg ggggactgag agacaggact acctggattt cctaggccga ctaagaatcc 60
ctaagcctag ctgggaaggt gaccacatcc acctttaaac acagggcttg caacttagct 120
cacacttgac cagtcaggta gtaaagagag ctactaaaa tgctaattag gctaaaacag 180
gaggtaaaga aatagacaat catctatcac ctgagagcac agtgggaggg acaatgatcg 240
gcatataaac ccaggcattc gagccagcaa cagcaacccc ctttgggagc tctgttttca 300
ctctatataa tcttgcaact gc 322

<210> 16
<211> 343
<212> DNA
<213> Human endogenous retrovirus

<400> 16
tgttgagatg ggggactgag agacaggact agctggattt cctaggccaa ctaagaatcc 60
ctaagcctag ctgggaaggt gactacaccc acctttaaac atggggcttg caacttagct 120
cacaccaac caatcaggta gtaaagagag cttgctaaaa tgctaattag gcaaaaacag 180
gaggtaaaga aatagccagt catctatcgc ctgacagcac aaggggcggg acaatgatca 240
ggatataaac tcaggcattc aagccagcaa tggctacca ctttgggtcc cctcccattt 300
tatgggagct ctgttttcac tctattaat cttgcaactg caa 343

<210> 17
<211> 343
<212> DNA
<213> Human endogenous retrovirus

5

<400> 17

```

tgttgagatg ggggactgag agacaggact agctggattc cctaggccga ctaagaatcc 60
ctaagcctag ctgggaaggt gaccacatcc acctttaaac acggggcttg caacttagct 120
catacccaac aatcaggta gtaaagagag ctactaaaa tactgattag gcgaaaacag 180
gaggtaagga aacagccagt catctatcgc ctgacagcac aaggggaggg acaatgatca 240
ggatataaac tcaggcattc aagccagcaa tggctaccca ctttgggtcc cctcccattt 300
tatgggagct ctgttttcac tctattaat cttgcaactg caa 343

```

<210> 18

<211> 343

<212> DNA

<213> Human endogenous retrovirus

<400> 18

```

tgttgagatg ggggactgag agacaggact agttggattt cctaggctgg ctaagaatcc 60
ctaagcctag ctgggaaatt gaccacgtcc acctttaaac acggggcttg caatttagct 120
cacacccgac caatcaggta gtaaaggag ctactaaaa tgctaattag ggaaaaacag 180
gaggtaaaga agtagccaat catctatcgc ctgagagcac aacaggaggg acaatgatca 240
ggatataaac ccaggcattc aagccagcgg tggctaccct ctttgggtcc cctccccttg 300
tatgggagcc ctgttttcac tctattaat cttgcaactg caa 343

```

<210> 19

<211> 343

<212> DNA

<213> Human endogenous retrovirus

<400> 19

```

tgttgagatg ggggactgag agacaggact agttggattt cctaggccgg ctaagaatcc 60
ctaagcctag ctgggaaatt gaccacgtcc acctttaaac acggggcttg caatttagct 120
cacacccgac caatcaggta gtaaaggag ctactaaaa tgctaattag ggaaaaacag 180
gaggtaaaga agtagccaat catctatcgc ctgagagcac aacaggaggg acaatgatca 240
ggatataaac ccaggcattc aagccagcgg tggctaccct ctttgggtcc cctccccttg 300
tatggaagct ctgttttcac tctattaat cttgcaactg caa 343

```

<210> 20

<211> 343

<212> DNA

<213> Human endogenous retrovirus

<400> 20

```

tgttgagatg ggggactgag agacaggact agctggattt cctaggccaa ctaagaatcc 60
ctaagcctag ctgggaaggt gactacaccc acctttaaac actaggcttg caacttagct 120
cacacccgac caatcaggta gtaaagagag cttgctaaaa tgctaattag gcaaaaacag 180
gaggtagaga aatagccaat catctatcgc ctgagagcac agcaggaggg acaatgatcc 240
ggatataaac ccaagcattc gagccagcaa tggctaccct ctttgtgtcc cctccccttg 300
tatgggagct ctattttcac tctattaat cttgcaactg caa 343

```

<210> 21

<211> 343

<212> DNA

<213> Human endogenous retrovirus

<400> 21

```

tgttgagatg ggggactgag agacaggact agctggattt cctaggctga ctaagaatcc 60
ctaagcctag ctgggaaggt gaccgcatcc atctttaaac atggggcttg caacttaact 120
catatctgac caatcaggta gtaaagagag cttgctaaaa tgctaattag gcaaaaacag 180
gaggtaaaga aatagccagt catctatcgc ctgacagcac aaggggaggg acaatgatca 240
ggatataaac tcaggcattc aagccagcaa tggctaccca ctttgggtcc cctcccattt 300
tatgggagct ctgttttcac tctattaat cttgcaactg caa 343

```

<210> 22
 <211> 343
 <212> DNA
 <213> Human endogenous retrovirus

<400> 22
 tgttgagatg ggggactgag agacaggact agctggattt cctaggctga ctaagaatcc 60
 ctaagcctag ctgggaaggt gactacaccc acctttaacc actaggcttg caacttagct 120
 cacacccgac caatcaggta gtaaagagag cttgctaaaa tgctaattag gcaaaaacag 180
 gaggtaaaga aatagccagt catctatcgc ctgacagcac aaggggcggg acaatgatca 240
 ggatataaac tcaggcattc aagccagcaa tggctaccca ctttgggtcc cctcccattt 300
 tatgggagct ctgttttcac tctattaat cttgcaactg caa 343

<210> 23
 <211> 343
 <212> DNA
 <213> Human endogenous retrovirus

<400> 23
 tgttgagatg ggggactgag agacaggact agttggattt cctaggctgg ctaagaatcc 60
 ctaagcctag ctgggaaatt gaccacgtcc acctttaaac acggggcttg caatttagct 120
 cacacccgac caatcaggta gtaaaggag ctaactaaaa tgctaattag ggaaaaacag 180
 gaggtaaaga agtagccaat catctatcgc ctgagagcac aacaggaggg acaatgatca 240
 ggatataaac ccaggcattc aagccagcgg tggctaccct ctttgggtcc cctccccttg 300
 tatggaagct ctgttttcac tctattaat cttgcaactg caa 343

<210> 24
 <211> 343
 <212> DNA
 <213> Human endogenous retrovirus

<400> 24
 tgttgagatg ggggactgag agacaggact acctggattt cctaggccaa ctaagaatct 60
 ctaagcctag ctgggaaggt gaccacatcc acctttaaac acagggcttg caacttagct 120
 cacacccgac ccatcaggta agaaagagag cccgctaaaa tgctaattag gcaaaaacag 180
 gaggtaaaga aatagtcaat catctattgc ctgagagcac agcgggaggg acaatgatca 240
 ggatataaac ccaggcattc gagccggcaa cgactaccct ctttgggtcc cctccccttg 300
 tatgggagct ctgttttcac tctattaat cttgcaactg caa 343

<210> 25
 <211> 343
 <212> DNA
 <213> Human endogenous retrovirus

<400> 25
 tgttgagatg ggggactgag agacaggact agctggattt cctaggccaa ctaagaatcc 60
 ctaagcctag ctgggaaggt gactacaccc acctttaacc actaggcttg caacttagct 120
 cacacccgac caatcaggta gtaaagagag cttgctaaaa tgctaattag gcaaaaacag 180
 gaggtaaaga aatagccagt catctatcgc ctgacagcac aaggggcggg acaatgatca 240
 ggatataaac tcaggcattc aagccagcaa tggctaccca ctttgggtcc cctcccattt 300
 tatgggagct ctgttttcac tctattaat cttgcaactg caa 343

<210> 26
 <211> 343
 <212> DNA
 <213> Human endogenous retrovirus

7

<400> 26

```

tgttgagatg ggggactgag aacaggact agcaggattt cctaggccga ttaagaatcc 60
ctaagcctag atgggaagtt gaccacatcc acctttaaac acggggcttg caactcagct 120
cacacccgac ccatcaggta agaaagagag cccgctaaaa tgctaattag gcaaaaacag 180
gaggtaaaga aatagccaat catctattgc ctgagagcac agcggggagg acaatgatca 240
ggatataaac ccaggcattc gagccggcaa cgactaccct ctttgggtcc cctccctttg 300
tatgggagct ctgttttcac tctattaat cttgcaactg caa 343

```

<210> 27

<211> 619

<212> DNA

<213> Human endogenous retrovirus

<400> 27

```

gcgaccggtg gatccccggc cgcggtacc gtcgactgca gaattcatgg agcatacaat 60
cgggttttat accgagacat tccattgccc agggacaggc aggagacaga tgccttcctc 120
ttgtctcaac tgcaagaggc attccttcct cttatactaa tcctcctcag cacagaccct 180
ttacgggtgt cgggctgggg gacggtcagg tctttccctt cccacgaggc catatttcag 240
actatcacat ggggagaaac cttggacaat acctggcttt cctaggcaga ggtccctgcg 300
gccttccgca gtttttgtgt cctgggtact tgagattagg gagtgggtgat gactcttaag 360
gagcatgctg ctttcaagca tctgtttaac aaagcacatc ctgcaccgcc cttaatccat 420
tcaaccctga gttgacacag cacacgtttc agagagcacg gggttggggg taaggtcata 480
gattaacaga atctcaaggc agaagaattt ttcttaacac ataacaaaat ggagtctccc 540
atgtctactt ctttctacac agacacagta acaatctgat ctctcttgct tttccccaca 600
tttccccctt ttcttttcg 619

```

<210> 28

<211> 620

<212> DNA

<213> Human endogenous retrovirus

<400> 28

```

gcgaccggtg gatccccggc cgcggtacc gtcgactgca gaattcatgg agcatacaat 60
cgggttttat accgagacat tccattgccc agggacaggc aggagacaga tgccttcctc 120
ttgtctcaac tgcaagaggc attccttcct cttatactaa tcctcctcag cacagaccct 180
ttacgggtgt cgggctgggg gacggtcagg tctttccctt cccacgaggc catatttcag 240
actatcacat ggggagaaac cttggacaat acctggcttt cctaggcaga ggtccctgcg 300
gccttccgca gtttttgtgt cctgggtact tgagattagg gagtgggtgat gactcttaag 360
gagcatgctg ctttcaagca tctgtttaac aaagcacatc ctgcaccgcc cttaatccat 420
tcaaccctga gttgacacag cacacgtttc agagagcacg gggttggggg taaggtcata 480
gattaacaga atctcaaggc agaagaattt ttcttaacac ataacaaaat ggagtctccc 540
atgtctactt ctttctacac agacacagta acaatctgat ctctcttgct tttccccaca 600
tttccccctt ttcttttcg 620

```

<210> 29

<211> 624

<212> DNA

<213> Human endogenous retrovirus

<400> 29

```

gcgaccggtg gatccccggc cgcggtacc gtcgactgca gaattcatgg agcatacaat 60
cgggttttat accgagacat tccattgccc agggacaggc aggagacaga tgccttcctc 120
ttgtctcaac tgcaagaggc attccttcct cttatactaa tcctcctcag cacagaccct 180
ttacgggtgt cgggctgggg gatggtcagg tctttccctt cccacgaggc catatttcag 240
actatcacat ggggagaaac cttggacaat acctggcttt cctaggcaga ggtccctgcg 300
gccttccgca gtttttgtgt cctgggtact tgagattagg gagtgggtgat gactcttaag 360
gagcatgctg ctttcaagca tctgtttaac aaagcacatc ctgcaactgcc cttaatccat 420
tcaaccctga gttgacacag cgacacgtttc agagagcacg gggttggggg taaggtcata 480
gattaacaga atctcaaggc agaagaattt ttcttaacac ataacaaaat ggagtctccc 540
atgtctactt ctttctacac agacacagta acaatctgat ctctcttgct tttccccaca 600

```

tttccccctt ttcttttcga caaa

624

<210> 30
 <211> 646
 <212> DNA
 <213> Human endogenous retrovirus

<400> 30
 gcgaccggtg gatccccggc cgcggttacc gtcgactgca gaattcatgg agcatacaat 60
 cgggtttttat accgagacat tccattgccc agggacaggc aggagacaga tgccttcctc 120
 ttgtctcaac tgcaagaggc attccttctt cttatactaa tcctcctcag cacagaccct 180
 ttacgggtgt cgggctgggg gacggtcagg tctttccctt cccacgaggc catatttcag 240
 actatcacat ggggagaaac cttggacaat acctggcttt ctaggcaga ggtccctgcg 300
 gccttcgca gttttgtgt cctgggtact tgagattagg gagggtgat gactcttaag 360
 gagcatgctg cttcaagca tctgtttaac aaagcacatc ctgcaccgcc cttaatccat 420
 tcaaccctga gttgacacg cacacgttcc agagagcacg ggggtggggg taaggtcata 480
 gattaacaga atctcaaggc agaagaattt ttcttaacac ataacaaaat ggagtctccc 540
 atgtctactt ctttctacac agacacagta acaatctgat ctctcttctt tttcccaca 600
 tttccccctt ttcttttcga caaaaccgcc atctcgagat ctgagt 646

<210> 31
 <211> 672
 <212> DNA
 <213> Human endogenous retrovirus

<400> 31
 gtccacctc cagccctaag gcggttttcc cctatctcag tagatggagc atacaatcgg 60
 gttttatacc gagacattcc attgccaggc gacaggcagg agacagatgc cttcctcttg 120
 tctcaactgc aagaggcatt ccttcctctt ataactaatc tcctcagcac agacccttta 180
 cgggtgtcgg gctgggggac ggtcaggctt tcccttccc acgaggccat atttcagact 240
 atcacatggg gagaaacctt ggacaatacc tggctttcct aggcagaggt cctgcgcc 300
 ttccgcagtt tttgtgtcct gggtacttga gattagggag tggatgatgac tcttaaggag 360
 catgctgcct tcaagcatct gtttaacaag gcacatcctg caccgccctt aatccattca 420
 accctgagtt gacacagcac acgtttcaga gagcacgggg ttgggggtaa ggtcatagat 480
 taacagaatc tcaaggcaga agaatttttc ttaacacata acaaaatgga gtctcccattg 540
 tctacttctt tctacacaga cacagtaaca atctgatccc tcttgctttt cccacattt 600
 ccccttttcc ttatccatca cactggcggc cgctcgagca tgcattctaga gggcccaatt 660
 cgccctatag tg 672

<210> 32
 <211> 593
 <212> DNA
 <213> Human endogenous retrovirus

<400> 32
 agtagatgga gcatacaatc gggttttata ccgagacatt ccattgcca gggacaggca 60
 ggagacagat gccttcctct tgtctcaact gcaagaggca ttccttcctc ttttactaat 120
 cctcctcagc acagaccctt tacagggtgtc gggctggggg acggtcaggc ctttcccttc 180
 ccacgaggcc atatttcaga ctatcacatg gggagaaacc ttggacaata cctggctttc 240
 ctaggcagag gtccctgagg cttctctgag tttttgtgtc cctgggtact tgagattagg 300
 gagggtgat gactcttaag gagcatgctg cttcaagca tctgtttaac aaagcacatc 360
 ctgcaccgcc cttaatccat tcaaccctga gttgacacag cacatgtttc agagagcacg 420
 ggggtggggg taaggtcata gattaacaga atctcaaggc agaagaattt ttcttagcac 480
 ataacaaaat ggagtctctt atgtctactt ctttctacac agacacagta acaatttgat 540
 ctctcttctt tttcccaca tttccccctt ttcttttcga caaaaccgcc atc 593

<210> 33
 <211> 943
 <212> DNA

<213> Human endogenous retrovirus

<400> 33

```
tgtgggcgaa ggattaccca ggtgccgagg caagagactg aaggcacaaa ctgtttcagt 60
ataatataga aaatagctag aataagaata gttataataa aaattagata tacacatgat 120
catggacatt accaatcatt actacaaaca ttgttaatca ttagctttta atattactct 180
ttgttttatt actaatataa ccaaggaata accggtagca tacggtcagg tgctgaaggg 240
acattgtgag aagtgaccta gaaggcaaga ggtgagcctt ctgtcacgcc tgcataagga 300
cagcttgagg gctccttggt caagctgtaa caccagtgcc tgggaaggca ccgttactta 360
gcagaccatg aaagggagtc tccattcctt ggaggagtca gggaaacact atgctccacc 420
agcttcttgt gtatccagcc ctgccacag tcatccagag gcataaaccc ctccctgtgg 480
tgctgtgctt caatggccat gcttcttgtc cactttcatg ttctctctgt actcctgggt 540
cctctttgaa gttcgtagaa gataatggta gaagaaatag tgaaagtctt tgatctttct 600
tataagtgca tagaagaaaa cactgatgta tgctgcctt ccctctctgc ttcagctacc 660
taaaaggaaa ggcccccttt cccatgatca catgacttgc ctgaccttat caatcacttg 720
gaggactcac cctccttacc ctgtcccttt gtcttgtatg caataaatat cagcacgccc 780
agccattcgg ggccactact ggtctccgca acttgggtgg agtggtagcc tgggcccagc 840
tgttttctct ttatctcttt tgtcttgggt ctttatttct tacaatctct catctctgca 900
catggggaga acaccggcaa agcccgtagg gctggacctt aca 943
```

<210> 34

<211> 389

<212> DNA

<213> Human endogenous retrovirus

<400> 34

```
aaaccctcc ctgtgggtgct gtgcttcaat ggccatgctt cttgtccact ttcatgttcc 60
tcctgtactc ctggttcctc tttgaagttc gtagaagata atggtagaag aaatagtgaa 120
agtctttgat ctttcttata agtgcataga agaaaacact gatgtatgcc tgccttccct 180
ctctgcttca gctacctaaa aggaaaggcc ccttttccca tgatcacatg acttgctga 240
ccttatcaat cacttgagg actcacctc cttaccctgt ccctttgtct tgtatgcaat 300
aaatatcagc acgcccagcc attcggggcc actactggtc tccgcaactt ggtggtagtg 360
gtaccctggg cccagctggt ttctcttta 389
```

<210> 35

<211> 858

<212> DNA

<213> Human endogenous retrovirus

<400> 35

```
tgtgggcgga agagtaccta ggtgccgagg caagagactg aaggcacaaa ctgtttcagt 60
ataataaaga aaatagaata agaatagtca taatacaaat tagatacagc gatgatcatg 120
aacaattatc catcattatt ataaaacatta ttaatcatta gcttttaata ttactctggt 180
gcattaataa tataacctag gaataaccgg caggtatagg gtcagggtgct gaaggacat 240
tgtgagaagt gaatagaagg caagaggga gccttctgtc atgcccgcac aagggccgct 300
tgagggcccc ttggtcaagc ggtaacgcca gtgtctggga aggcaccctg tactgagcag 360
accgggaaaag ggagtctcct ttccttggag gagtcaggga acgctctgct ccaccagctt 420
cttgtgggag gctggatggt acccaggcct gcctgcagtc atccggaggc ctgaaccct 480
ccctgtgggt cttcaatggt cacgttcctt gtccactttc atgctccttc cgtactcctg 540
gttccctctt gaagttcgta gtagatagcg gtagaagaaa tagtgaaagt cttaaagtct 600
ttgatcttat aagttcatag aagaaaacgc tgatgcctgc cgccttctct ctctgcttca 660
gctacctaaag aggggaaggc ccgctgtcct gtgatcaggt gacttgcttc accttgtaa 720
tacttagaa gactgacct ccttatcctg ccccttgtc ttgtatgcaa taaatatcag 780
cgagcccagc cgttcagggc cactaccggt ctccgtgtct ttgtggtagt ggtccccggg 840
cccagctggt ttctcttt 858
```

<210> 36

<211> 386

<212> DNA

<213> Human endogenous retrovirus

<400> 36
 gaacccctcc ctgtgggtgct tcaatggtea cgttccttgt ccactttcat gtccttccg 60
 tactcctggg tcctctttga agttcgtagt agatagcggg agaagaaata gtgaaagtct 120
 taaagtcttt gatcttataa gttcatagaa gaaaacgctg atgcctgccg ccttctctct 180
 ctgcttcagc tacctaagag ggaagggccc gctgtcctgt gatcagggtga cttgcttcac 240
 cttgtcaatc acttagaaga ctgacccctc ttatcctgcc cccttgtctt gtatgcaata 300
 aatatcagcg agcccagccg ttcagggcca ctaccggctc ccgtgtcttt gtggtagtgg 360
 tccccgggcc cagctgtttt ctcttt 386

<210> 37
 <211> 844
 <212> DNA
 <213> Human endogenous retrovirus

<400> 37
 tgtgggtgga ggattaccca ggtgccaagg caagagactg aaggcacaaa ctgtttcagt 60
 ataataaaaa aaatagaata agaatagtca taatacaaat tagatataga gatgatcatg 120
 gacaattagc aatcactatt aatcttttagc ttttaaatatt actctttggt gcattactaa 180
 tataacctag gaataaccgg tgggtatagg gtcagggtgct gaagggacat tgtgtgaagt 240
 gacctggaag gcaagagggt agccctctgt cacgcccaca taagggccgc ttgagggctc 300
 cttgggtcaag tggtaacgcc agtgtctggg aatgcacccg ttaattagca gaccgcaaaa 360
 gggagtctcc tttccttggg agagttgggg aacactctgc tccaccagct tcttgtggaa 420
 ggctggatat tatccaggcc tgcgcgcagt catccggagg cttaaaccct tccctgtggg 480
 gctgtgcttc aatggtccca ctcttgtcc actttcatgc tccctccgta ctctgggtc 540
 ctctttgaag agcgcagtag atagcggtag aagaaatagt gaaagtctta aagtcttcga 600
 tctttcttac aagtgcagag aagaaaacgc tgacatatgc tgccttcctt ctctgcttcg 660
 gctacctaaa agggaagggc cgcctatcct gtaatcacat gacttgcttc acctgtgcaa 720
 tcacttagaa gattcactct ccttaccctg ccccttgtc ttgtatgcaa taaatatcag 780
 tgacccagc cgttcagggc cactactggt ctccgcgtct tgatggtagt ggtcaccccg 840
 gcc 844

<210> 38
 <211> 381
 <212> DNA
 <213> Human endogenous retrovirus

<400> 38
 aaacccttcc ctgtgggtgct gtgcttcaat ggtcccactc cttgtccact ttcagtctcc 60
 tcccgctactc ctggttcctc tttgaagagc gcagtagata gcggtagaag aaatagttaa 120
 agttctaaag tcttcgatct ttcttacaag tgcagagaag aaaacgctga catatgctgc 180
 cttccctctc tgcttcgggt acctaaaagg gaagggccgc ctatcctgta atcacatgac 240
 ttgcttcacc ttgtcaatca cttagaagat tcaactctct taccctgccc ccttgtcttg 300
 tatgcaataa atatcagtga cccagccgt tcagggccac tactggtctc cgcgtcttga 360
 tggtagtggg cccccggcc c 381

<210> 39
 <211> 859
 <212> DNA
 <213> Human endogenous retrovirus

<400> 39
 tgtgggtgga ggattaccca ggtgccaagg caagagactg aaggcacaaa ctgtttcagt 60
 ataataaaaa aaatagaata agaatagtca taatacaaat tagatataga gatgatcatg 120
 gacaattagc aatcactatt aatcttttagc ttttaaatatt actctttggt gcattactaa 180
 tataacctag gaataaccgg tgggtatagg gtcagggtgct gaagggacat tgtgagaagt 240
 gacctggaag gcaagagggt agccctctgt cacgcccaca taagggccgc ttgagggctc 300
 cttgggtcaag tggtaacgcc agtgtctggg aatgcacccg ttaattagca gaccgcaaaa 360
 gggagtctcc tttccttggg agagttgggg aacactctgc tccaccagct tcttgtggaa 420
 ggctggatat tatccaggcc tgcgcgcagt catccggagg cttaaaccct tccctgtggg 480

```

gctgtgcttc aatggtecca ctcttgttc actttcatgc tctcccgtta ctctgggttc 540
ctctttgaag agcgcagtag atagcggtag aagaaatagt gaaagtctta aagtcttcga 600
tctttcttac aagtgcagag aagaaaacgc tgacatatgc tgccttcctt ctctgcttcg 660
gctacctaaa agggaaggcc gcctatctct gtaatcacat gacttgcttc accttgctca 720
tcacttagaa gattcaccct ccttaccctg ccccctgttc ttgtatgcaa taaatatcag 780
tgaccccgag cgttcagggc cactactggt ctccgcgtct tgatggtagt ggtcaccctg 840
gcccagggtg ttttctttt
    
```

```

<210> 40
<211> 396
<212> DNA
<213> Human endogenous retrovirus
    
```

```

<400> 40
aaacccttcc ctgtggtgct gtgcttcaat ggtcccactc cttgtccact ttcattgctcc 60
tcccgtatcc ctggttcctc tttgaagagc gcagtagata gcggtagaag aaatagttaa 120
agtcttaaag tcttcgatct ttcttacaag tgcagagaag aaaacgctga catatgctgc 180
cttccctctc tgcttcggct acctaaaagg gaagggccgc ctatctctga atcacatgac 240
ttgcttcacc ttgtcaatca cttagaagat tcaccctcct taccctgccc ccttgtcttg 300
tatgcaataa atatcagtag cccagcctgt tcagggccac tactggtctc cgcgtcttga 360
tggtagtggg caccgctggc caggtgtttt tctcttt
    
```

```

<210> 41
<211> 966
<212> DNA
<213> Human endogenous retrovirus
    
```

```

<400> 41
tgtgggtgga ggattaccca ggtgccgagg caagagactg aaggcacaaa ctgtttcagt 60
ataataaaga aaatggttag aataagaata gtcataatac aaattagata tagagatgat 120
catggacaat tatcaatcat tattataaac attattaatc attagctttt aatattactc 180
tttgttgcac tactaatata acctaggaat aaccgggtgg tataggggtca ggtgctgaaa 240
ggacattggg agaagtgacc tagaaggcaa gaggtgagtc ttctgtcacg cccgcataag 300
ggttgcttga gggctccttg gtcaagtggg aacgcccgtg tctgggaagg cacctgtttc 360
ttagccgacc acgaaaggga gtctcctttc cttggaggag tcagggcgca ctctgctcca 420
ccagcttctt gtggaaggct ggatattatc caggcctgcc cgcagtcac cggaggccta 480
aaccctccc tgtggtgctg tgcttcaatg ggcacactcc tcgtccactt tcatgttctt 540
cccatactcc tggtttctct ttgaagtctg tagtagatag tggtagaagg aaatagggaaa 600
atcttaaagt gtttgatctt tcttataagt gcatagaaga aaacgctgac atatgctgcc 660
ttctctgtct gcttcagcta cctaagaggg aagggccccc tgtccagtga tcacgtgact 720
tgcttcacct tgtcaatcac ttagaagatt caccctcctt accctgcccc cttgtcttgg 780
atgcaataaa tatcagtgca cccagccttt cggggccact taccggtctc cacgtcttgg 840
tggtagtggg cccccggg cagctgtttt ctctttatct ctttgtcttg tgtcttattt 900
attacaatct ctctctctcc cacacagga gaacaccgc taagctccgt agggctggac 960
cttaca
    
```

```

<210> 42
<211> 398
<212> DNA
<213> Human endogenous retrovirus
    
```

```

<400> 42
aaacccttcc ctgtggtgct gtgcttcaat gggcacactc ctctccact ttcattgctcc 60
tcccatactc ctggtttctc tttgaagtgc gtagtagata gtggtagaag gaatagggaa 120
aatcttaaag tgtttgatct ttcttataag tgcatagaag aaaacgctga catatgctgc 180
cttctctgtc tgcttcagct acctaagagg gaagggcccc ctgtccagtg atcaogtgac 240
ttgcttcacc ttgtcaatca cttagaagat tcaccctcct taccctgccc ccttgtcttg 300
tatgcaataa atatcagtag acccagcctt tcggggccac ttaccggtct ccacgtcttg 360
tggtagtggg tccccggg ccagctgttt tctcttta
    
```

<210> 43
 <211> 938
 <212> DNA
 <213> Human endogenous retrovirus

<400> 43
 tgtgggtgga ggattaccca ggtgccgagg caagagactg aaggcacaaa ctgtttcagt 60
 ataataaaga aaatggtttag aataagaata gtcataatac aaattagata tagagatgat 120
 catggacaat tatcaatcat tattataaac attattaatc attagctttt aatattactc 180
 tttgttgcat tactaatata acctaggaat aaccgggtggg tatagggtca ggtgctgaag 240
 ggacattggg agaagtgacc tagaaggcaa gaggtgagtc ttctgtcacg cccgcataag 300
 ggttgcttga gggctccttg gtcaagtggg aacgccgggtg tctgggaagg cacctgttac 360
 ttagccgacc acgaaagga gtctcctttc cttggaggag tcagggcaca ctctgctcca 420
 ccagcttctt gtggaaggct ggatattatc caggcctgcc cgcagtcac cggaggccta 480
 aaccctccc tgtggtgctg tgcttcaatg ggcacactcc tcgtccactt tcatgttctc 540
 cccatactcc tggttcctct ttgaagtctg tagtagatag tggtagaagg aatagggaaa 600
 atcttaaagt gtttgatctt tcttataagt gcatagaaga aaacgctgac atatgctgcc 660
 ttctctgtct gcttcagcta cctaagaggg aagggccccc tgtccagtga tcacgtgact 720
 tgcttcacct tgtcaatcac ttagaagatt caccctcctt accctgcccc cttgtcttgg 780
 atgcaataaa tatcagtgca ccagccttt cggkkcactt accggctctc acgtcttggg 840
 ggtagtggtc ccccggccca gctgttttct ctttatctct ttgtcttggg tcttatttat 900
 tacaatctct cgtctccgca cacagggaga acaccgcg 938

<210> 44
 <211> 396
 <212> DNA
 <213> Human endogenous retrovirus

<400> 44
 aaaccctccc ctgtggtgct gtgcttcaat gggcacactc ctggtccact ttcattgttcc 60
 tcccatactc ctggttcctc tttgaagttc gtagtagata gtggtagaag gaatagggaa 120
 aatcttaaag tgtttgatct ttcttataag tgcatagaag aaaacgctga catatgctgc 180
 cttctctgtc tgcttcagct acctaagagg gaagggccccc ctgtccagtg atcacgtgac 240
 ttgcttcacc ttgtcaatca cttagaagat tcaccctcct taccctgccc ccttgtcttg 300
 tatgcaataa atatcagtgac acccagcctt tcggkkcact taccggtctc cacgtcttgg 360
 tggtagtggt ccccggcccc agctgttttc tcttta 396

<210> 45
 <211> 963
 <212> DNA
 <213> Human endogenous retrovirus

<400> 45
 tgtgggcgaa agattaccta ggtgccgagg caagagactg aaggcacaaa ctgtttcagt 60
 ataataaaga aaatagttaa aataagaata gttataatac aaattagata tagagatgat 120
 catggacaat tatcaatcat tattataaac attaatcatt agcttttaatt attactcttt 180
 gttgctttac taatataacc taggaataac cgggtgggtat agggtcagggt gttgacggga 240
 tattgtgaga agtgacctag aaggcaagag gtgagccttc tgtcacgccc acataagggc 300
 cgcttgaggg ctctttgggtc aagtggtaac gccagtgtct gtgaaggcac ctgttactta 360
 gcagaccgag aaagggagtc tcctttcctt ggaggagtca gggaaactc tgctccacca 420
 gcttcttggg gaaggctgga tattatctag gcctgccccg agtcatctgg aggctaaaac 480
 ccctccctgt ggtgctgtgc ttcagtggtc actctccttg tccactttca tgttcctccc 540
 gtactcctgg ttctcttttg aagttcgtag tagatagcag tagaagaaat agtgaagtc 600
 ttaaagtatt tgatctttct tataagtga tagaagaaaa cgctgacata tgctgccttc 660
 tctatctctg cggtagctac ctaaaaggga agggccccct gtcccatgat catgtgactt 720
 gcttcacctt atcagctaga agattcatcc tccttaccct gcgccccctc gtcttgtatg 780
 caataaatat cagcacgccc agtcgtttga ggccactgcc ggtctccgag tcttgggtgg 840
 agtgggtccc cgggcccagc tattgtctct ttatctcttt gtcttgtgtc tttattttatt 900
 acaatctctt gtctctgcac acagggagaa cacctgctaa gccccgtagg actggaccct 960
 aca 963

<210> 46
 <211> 397
 <212> DNA
 <213> Human endogenous retrovirus

<400> 46
 aaaccctcc ctgtggtgct gtgcttcagt ggctactctc cttgtccact ttcattgttc 60
 tcccgtactc ctggttcctc tttgaagttc gtagtagata gcagtagaag aaatagttaa 120
 agtcttaaaag tatttgatct ttcttataag tgcatagaag aaaacgctga catatgctgc 180
 cttctctatc tctgcggtgg ctacctaaaa ggggaagggcc ccctgtcca tgatcatgtg 240
 acttgcttca ccttatcact tagaagattc atcctcctta ccctgcgcc cctcgtcttg 300
 tatgcaataa atatcagcac gccagtcgt ttgaggccac tgccggtctc cgcgtcttgg 360
 tggtagtggt cccccgggc cagctattgt ctcttta 397

<210> 47
 <211> 489
 <212> DNA
 <213> Human endogenous retrovirus

<400> 47
 tgttcaattc tttgccttct acttttaaac ttaacttctc cataaagcaa cctttttcaa 60
 tcacctgctc cactctgact cattctgatc acctgctcca ccctgactca ttccgatcac 120
 ctgatccact gtgactcatt ccgattaccg gctccaccct gactcattct gattctgatt 180
 tctgtctctg ccataacat ttttccgcc aaaccactca ccctgtcact ctctttaaatt 240
 tagccaattg gaattagttt agcctgtgcg gtctaaccct agccaatagg ggactgacac 300
 agcagcaggg gccacatgtg tcaggaataa gacccccctc ccctccctgt ccagatgtgt 360
 gctcaccatt gctccatctg tgagggcaca cccttctata gaagtaaatt gccttgctga 420
 gaagaaaaaa aagaacattt tatattcaag tcctatttct tttgctgcac cgaaacttta 480
 tttataaca 489