(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 November 2012 (22.11.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/156839 A2

(51) **International Patent Classification**:
C12N 15/86 (2006.01) *A61K 48/00* (2006.01)

(21) **International Application Number**:
PCT/IB2012/052041

(22) **International Filing Date**:
23 April 2012 (23.04.2012)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/487,891 19 May 2011 (19.05.2011) US

(71) **Applicants** *(for all designated States except US)*: **OSPEDALE SAN RAFFAELE S.R.L.** [IT/IT]; Via Olgettina, 60, I-20132 Milano (IT). **Fondazione Telethon** [IT/IT]; Via Carlo Spinola, 16, I-00154 Rome (IT).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **MONTINI, Eugenio** [IT/IT]; San Raffaele Telethon Institute for Gene Therapy, Via Olgettina, 58, I-20132 Milan (IT). **NALDINI, Luigi** [IT/IT]; San Raffaele Telethon Institute for Gene Therapy, Via Olgettina, 58, I-20132 Milan (IT). **FERRARI, Giuliana** [IT/IT]; San Raffaele Telethon Institute for Gene Therapy, Via Olgettina, 58, I-20132 Milan (IT). **MAVILIO, Fulvio** [IT/IT]; Gene Expression Unit, San Raffaele Scientific Institute, Via Olgettina, 58, I-20132 Milan (IT).

(74) **Agent**: **O'Brien, Simon**; D Young & Co LLP, 120 Holborn, London EC1N 2DY (GB).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2012/156839 A2

(54) **Title**: NEW GENERATION OF SPLICE-LESS LENTIVIRAL VECTORS FOR SAFER GENE THERAPY APPLICATIONS

(57) **Abstract**: Use of a lentiviral vector containing a lentiviral backbone in which at least two of the splice sites have been eliminated to improve the safety profile of the lentiviral vector.

New Generation of Splice-less Lentiviral Vectors for Safer Gene Therapy Applications

Field of the Invention

The present invention generally relates to lentiviral vectors for use in gene transfer and therapy applications, and to methods of producing them, and uses thereof.

Background to the Invention

Lentiviral vectors (LVs) and other viral vectors are an attractive tool for gene therapy. LVs can transduce a broad range of tissues, including non-dividing cells such as hepatocytes, neurons and hematopoietic stem cells. Moreover, LVs integrate into target cell genomes and provide long-term transgene expression.

In order to overcome the risk of insertional mutagenesis of MoMLV-based retroviral vectors in gene therapy, the usage of self-inactivated lentiviral vectors (LVs) has been taken into account as a safer tool for the treatment of genetic disorders.

However, it has now been found that LVs integrate preferentially inside active genes and possess strong splicing and polyadenylation signals that could lead to the formation of aberrant and possibly truncated transcripts.

Retroviruses, transposons and gene therapy vectors integrate into the genome of host cells and are able to trigger oncogenesis by a process known as insertional mutagenesis, which consists in the deregulation of proto-oncogenes found at or nearby the insertion site via different molecular mechanisms (Uren et al. Oncogene 24:7656-7672; Baum, C. 2007. Et al. Curr Opin Hematol 14:337-342). As demonstrated in several different mouse models of oncogene tagging and in gene therapy clinical trials enhancer mediated activation is the most prominent mechanism involved in oncogene activation. Such enhancer mediated activation involves short and long-range interaction of viral enhancer sequences with cellular promoters to increase the mRNA levels of a proto-oncogene (Coffin et al. H. 1997. *Retroviruses*. Plainview, N.Y.: Cold Spring Harbor Laboratory Press. xv, 843 p. pp.)

However, additional mechanisms of proto-oncogene activation may involve the generation of chimeric transcripts originating from the interaction of promoter

elements or splice sites contained in the genome of the insertional mutagen with the cellular transcriptional unit targeted by integration (Gabriel et al. 2009. Nat Med 15:1431-1436; Bokhoven, et al. J Virol 83:283-29). Chimeric fusion transcripts comprising vector sequences and cellular mRNAs can be generated either by read-through transcription starting from vector sequences and proceeding into the flanking cellular genes, or *vice versa*. *In vitro* genotoxicity assays and mouse studies show that when retroviruses, transposons, γRV or LVs with active LTRs integrate downstream the promoters of cellular genes in the same transcriptional orientation, gene transcription is put under the control of the viral promoter present in the 5' or 3' LTR (Kool et al. Nat Rev Cancer 9:389-399; Bokhoven et al. 2009 *J Virol* 83:283-294). In a previous study, using a tumor prone mouse model for LV genotoxicity testing, a tumor was found harboring an integration of an LV with active LTRs within the *Braf* transcription unit (Montini, et al. 2009; *J Clin Invest* 119:964-975). This integration led to the formation of an aberrant transcript encoding for a truncated Braf protein lacking the regulatory domain and endowed with oncogenic activity. Specifically, the canonical LV splice donor sequence placed downstream the active LTR proficiently interacted with the splice acceptor of the 13th exon of *Braf* to form this aberrant transcript. The same mechanism of vector LTR-driven read-through and splicing capture was also responsible for several independent gene activation events in an *in vitro* genotoxicity assay (Bokhoven et al. 2009 *J Virol* 83:283-294; Knight et al. *Journal of virology* 84:4856-4859.). Aberrant transcript formation can even be caused by vectors with self-inactivating (SIN) LTRs, which are devoid of strong enhancer-promoter sequences.

In a recent LV-based gene therapy trial for the treatment of β-thalassemia, a transplanted patient displayed a dominant myeloid cell clone harboring an integrated vector copy within *HMGA2*. Vector integration triggered the fusion of the splice donor sequence of the third exon of *HMGA2* with a cryptic splice acceptor sequence present within an insulator element inserted in the vector LTR. Interestingly, this new splicing event caused activation of the viral polyadenylation signal in the LV LTR and thus induced premature *HMGA2* transcript termination. This aberrant mRNA, lacking let7 miRNA binding sites, displayed a higher stability that in turn leads to increased protein levels. Although not fully proven, the activation of *HMGA2* has been suggested to be causative of the clonal dominance (Cavazzana-Calvo et al. 2010. *Nature* 467:318-322). Thus, there is emerging evidence that the potential of inducing aberrant transcripts might constitute a previously unappreciated genotoxicity factor for gene therapy vectors. How to reduce these splicing capture events and aberrant

transcript formation triggered by vector integration is still unclear. The present invention addresses this problem.

Summary of the Invention

The invention relates to design of safer integrative lentiviral vectors (LV) to avoid generation of aberrant transcripts (aberrantly spliced mRNAs that contain lentiviral vector sequences fused with cellular transcripts) for reducing their potential post-transcriptional genotoxicity. New LV backbones in which the splice sites have been recoded and eliminated are safer.

In more detail, HIV-derived, self-inactivating (SIN) lentiviral vectors (LVs), depleted of LTR enhancer and promoter regions, provide an efficient, versatile and relatively safe gene delivery system. Nevertheless, since LVs integrate preferentially into active genes, they have the potential to de-regulate gene expression at the post-transcriptional level, by interfering with the normal splicing and polyadenylation of primary transcripts. To test this hypothesis we developed a new PCR technique named Linear Amplification-Mediated PCR on cDNA (cLAM-PCR) which allows retrieving aberrantly spliced mRNAs that contain LV sequences fused with cellular transcripts. We applied cLAM-PCR on lentiviral vector (LV)-transduced cell lines and primary human HSCs and identified several splice sites within the LV backbone that participate in the aberrant splicing process with variable efficiency.

Furthermore, splice sites were identified by transduction of human T cells, myeloid cells and keratinocytes with a specifically designed, "splice trap" LV, and the analysis of the expression of a promoter-less GFP gene placed downstream of the constitutive HIV GAG splice acceptor site. Cells were transduced also with SIN-LVs carrying internal GFP expression cassettes or a full human β-globin gene driven by a β-globin promoter and a reduced-size LCR. Cells were randomly cloned and integration sites mapped by LM-PCR in individual clones. Chimeric transcripts were identified by RACE PCR and exon-specific RT-PCR. Abnormal, chimeric transcripts were identified in >50% of the LV target genes in all cell types. Semi-quantitative RT-PCR revealed that fusion transcripts were mostly represented at low level compared to constitutively spliced, wild-type transcripts. Fusion transcripts were also generated through aberrant splicing caused by the usage of both constitutive and cryptic splice

sites located in the viral intron and the U5 portion of the 5' LTR and in the β-globin transcriptional cassette.

In summary, we have identified a set of splice sites that are responsible for most of the aberrantly spliced transcripts, providing a platform to recode vector backbones in order to reduce their potential post-transcriptional genotoxicity.

Statements of the Invention

The invention refers to novel lentiviral backbone constructs with reduced capability of interaction with the cellular splicing machinery and consequent reduction of chimerical LV/cellular transcript formation.

The aim of the invention is to obtain a novel lentiviral vector (LV) backbone devoid of sequences that have been demonstrated to be or are potentially involved in aberrant splicing formation.

According to one aspect of the present invention there is provided use of a lentiviral vector containing a lentiviral backbone, i.e. polynucleotide sequence, in which at least two, three four, five, six seven, eight, nine, ten, eleven or all etc of the splice sites have been eliminated to improve the safety profile of the lentiviral vector.

The safety profile is reduced compared to an equivalent lentiviral vector in which the corresponding splice sites have not been eliminated. By improved safety profile we include that the ability of the lentiviral vector to generate a lentiviral sequence fused to a cellular transcript is reduced. This can be measured using the techniques described herein.

According to another aspect of the present inventoin there is provided a polynucleotide sequence comprising a lentiviral nucleotide sequence wherein at least one of the following splice sites is inactivated (i.e., i.e at least one of the nucleotides corresponding to the following splice sites is inactivated):

SPLICE ACCEPTOR GROUP 1

SA1 - corresponding to nucleotides 3127-3128 of SEQ ID NO:1 or nucleotides 3130-3131 of SEQ ID NO:3

SA2 – corresponding to nucleotides 4341-4342 of SEQ ID NO:1 or nucleotides 4344-4345 of SEQ ID NO:3.

SA3 – corresponding to nucleotides 3071-3072 of SEQ ID NO:1 or nucleotides 3071-3072 of SEQ ID NO:3.

SA4 – corresponding to nucleotides 3068-3069 of SEQ ID NO:1 or nucleotides 3068-3069 of SEQ ID NO:3.

SA5 - corresponding to nucleotides 4069-4070 of SEQ ID NO:1 or nucleotides 4072-4073 of SEQ ID NO:3.

SA6 - corresponding to nucleotides 3947-3948 of SEQ ID NO:1 or nucleotides 3950-3951 of SEQ ID NO:3.

SA7 - corresponding to nucleotides 3597-3598 (complement) of SEQ ID NO:1 or nucleotides 3600-3601 (complement) of SEQ ID NO:3.

SA9 - corresponding to nucleotides 3431-3432 of SEQ ID NO:1 or nucleotides 3434-3435 of SEQ ID NO:3.

SA10 - corresponding to nucleotides 4361-4362 of SEQ ID NO:1 or nucleotides 4364-4365 of SEQ ID NO:3.

SA11 - corresponding to nucleotides 4373-4374 of SEQ ID NO:1 or nucleotides 4376-4377 of SEQ ID NO:3.

SA20 - corresponding to nucleotides 3933-3934 (complement) of SEQ ID NO:1 or nucleotides 3936-3937 (complement) of SEQ ID NO:3.

SA21 - corresponding to nucleotides 3929-3930 (complement) of SEQ ID NO:1 or nucleotides 3932-3933 (complement) of SEQ ID NO:3.

SPLICE DONOR GROUP 1

SD1 - corresponding to nucleotides 3178-3179 of SEQ ID NO:1 or nucleotides 3181-3182 of SEQ ID NO:3.

SD2 - corresponding to nucleotides 3557-3558 of SEQ ID NO:1 or nucleotides 3560-3561 of SEQ ID NO:3.

SD3 - corresponding to nucleotides 3920-3921 of SEQ ID NO:1 or nucleotides 3923-3924 of SEQ ID NO:3.

SD4 - corresponding to nucleotides 4450-4451 of SEQ ID NO:1 or nucleotides 4453-4454 of SEQ ID NO:3.

SD5- corresponding to nucleotides 2974-2975 (complement) of SEQ ID NO:1 or nucleotides 2974-2975 (complement) of SEQ ID NO:3.

SD6- corresponding to nucleotide 4347-4348 (complement) of SEQ ID NO:1 or nucleotides 4350-4351 (complement) of SEQ ID NO:3.

SD14- corresponding to nucleotides 6500-6501(complement) of SEQ ID NO:1 or nucleotides 6503-6504 (complement) of SEQ ID NO:3.

SD15- corresponding to nucleotides 6520-6521(complement) of SEQ ID NO:1 or nucleotides 6523-6524 (complement) of SEQ ID NO:3.

SPLICE ACCEPTOR GROUP 2 (referred to as cryptic splice acceptor sites in SEQ ID NO:1)

SA1 - corresponding to nucleotides 3040-3041 of SEQ ID NO:1 or nucleotides 3040-3041 of SEQ ID NO:3.

SA4 - corresponding to nucleotides 3077-3078 of SEQ ID NO:1 or nucleotides 3077-3078 of SEQ ID NO:3.

SA5 - corresponding to nucleotides 3089-3090 of SEQ ID NO:1 or nucleotides 3089-3090 of SEQ ID NO:3.

SA6 - corresponding to nucleotides 3108-3109 of SEQ ID NO:1 or nucleotides 3108-3109 of SEQ ID NO:3.

SA8 - corresponding to nucleotides 3130-3131 of SEQ ID NO:1 or nucleotides 3133-3134 of SEQ ID NO:3.

In various embodiments at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 of the splice sites shown above are inactivated.

It will be appreciated that the corresponding splice site sequences in SEQ ID NO1 and SEQ ID NO:3 are the same. However, the positon of the splice sites within these sequences differs in some cases by three nucleotides. This is becasue SEQ ID NO:3 is three nucloetides longer than SEQ ID NO:1. Table A below summarises the relative positions of the corresponding splice sites in SEQ ID NOs 1 and 3:

Name[a]	Dinucleotide number In SEQ ID NO:1[b]	Dinucleotide number In SEQ ID NO:3[c]	Strand[d]	Recoded sites[e]
SPLICE ACCEPTOR GROUP 1				
SA1	3127-3128	3130-3131	sense	
SA2	4341-**4342**	4344-**4345**	sense	Mut9
SA3	3071-3072	3071-3072	sense	
SA4	3068-3069	3068-3069	sense	
SA5	**4069**-4070	**4072**-4073	Antisense	Mut8
SA6	**3947**-3948	**3950**-3951	Antisense	Mu7
SA7	**3597**-3598	**3600**-3601	Antisense	Mut3
SA9	3431-**3432**	3434-**3435**	sense	Mut1
SA10	4361-**4362**	4364-**4365**	sense	Mut11
SA11	4373-**4374**	4376-**4377**	sense	Mut12
SA20	**3933**-3934	**3936**-3937	Antisense	Mut6
SA21	**3929**-3930	**3932**-3933	Antisense	Mut5
SPLICE DONOR GROUP 1				
SD1	**3178**-3179	**3181**-3182	sense	MutSD
SD2	**3557**-3558	**3560**-3561	sense	Mut2
SD3	**3920**-3921	**3923**-3924	sense	Mut4
SD4	**4450**-4451	**4453**-4454	sense	Mut13
SD5	2974-2975	2974-2975	Antisense	
SD6	4347-**4348**	4350-**4351**	Antisense	Mut10
SD14	6500-**6501**	6503-**6504**	Antisense	Mut14
SD15	6520-**6521**	6523-**6524**	Antisense	Mut15
SPLICE ACCEPTOR GROUP 2				
SA1	3040-3041	3040-3041	sense	
SA2	3068-3069	3068-3069	sense	
SA3	3071-3072	3071-3072	sense	
SA4	3077-3078	3077-3078	sense	
SA5	3089-3090	3089-3090	sense	
SA6	3108-3109	3108-3109	sense	
SA7	3127-3128	3130-3131	sense	
SA8	3130-3131	3133-3134	sense	

Table A

Table Legend

a) Name: name of the splice site
b) Dinucleotide number: position of the dinucleotide required for the splicing process and recognize from the spliceosome, according toSEQ ID NO:1 (the vectorNTI map plasmid #277).
c) Dinucleotide number: position of the dinucleotide required for the splicing process and recognize from the spliceosome, according toSEQ ID NO:3 (the vectorNTI map plasmid #743).
d) Strand: strand of the vector sequence(SEQ ID NO:1 and SEQ ID NO:3) where the splice site has been identified
e) Recoded site: name of the recoded site

According to another aspect of the present invention there is provided a polynucleotide sequence comprising a lentiviral nucleotide sequence wherein at least one of the splice sites shown in Table 1 (identified with respect to the dinucleotides numbers of SEQ ID NO:1) is/are inactivated. In various embodiments at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 of the splice sites shown in Table 1 are inactivated. In a preferred embodiment, the splice sites are inactivated using the recoded nucleotides shown in Table 1.

Name[a]	original sequence (dinucleotide underlined, Recoded nucleotide in bold)[b]	Dinucleotide number In SEQ ID NO:1[c]	Strand[d]	Recoded sites[e]	Recoded nucleotide number[f]	Recoded nucleotide[g]
SD1	GCGGCGACTG^GTGAGTACGC	3178-3179	sense	MutSD	3178	A
SA1	CTCGACGCAG^GACTCGGCTT	3127-3128	sense			
SA9	ATCCCTTCAG^ACAGGATCAG	3431-3432	sense	Mut1	3432	A
SD2	CAAAACAAAA^GTAAGACCAC	3557-3558	sense	Mut2	3557	A
SA7	CCTCCTCCAG^GTCTGAAGAT	3597-3598	Antisense	Mut3	3597	A
SD3	GCAGCTCCAG^GCAAGAATCC	3920-3921	sense	Mut4	3920	A
SA21	CCACAGCCAG^GATTCTTGCC	3929-3930	Antisense	Mut5	3929	A
SA20	CTTTCCACAG^CCAGGATTCT	3933-3934	Antisense	Mut6	3933	A
SA6	GATCCTTTAG^GTATCTTTCC	3947-3948	Antisense	Mu7	3947	A
SA5	CTCCATCCAG^GTCGTGTGAT	4069-4070	Antisense	Mut8	4069	A
SA2	ATCGTTTCAG^ACCCACCTCC	4341-4342	sense	Mut9	4342	A
SD6	GGGTTGGGAG^GTGGGTCTGA	4347-4348	Antisense	Mut10	4348	A
SA10	CAACCCCGAG^GGGACCCGAC	4361-4362	sense	Mut11	4362	A
SA11	GACCCGACAG^GCCCGAAGGA	4373-4374	sense	Mut12	4374	A
SD4	GATCTCGACG^GTATCGGTTA	4450-4451	sense	Mut13	4450	A
SD14	GGTCTTAAAG^GTACCGAGCT	6500-6501	Antisense	Mut14	6501	A
SD15	CAGCTGCCTT^GTAAGTCATT	6520-6521	Antisense	Mut15	6521	A
SA3	TCTCTAGCAG^TGGCGCCCGA	3071-3072	sense			
SA4	AAATCTCTAG^CAGTGGCGCC	3068-3069	sense			
SD5	AGCACTCAAG^GCAAGCTTTA	2974-2975	Antisense			

Table 1

Table Legend

a) Name: name of the splice site
b) Original sequence: nucleotide sequence involved in splicing. Underlined are indicated the dinucleotide required for splicing (position indicated in b) and ^ indicate the splice site. In bold are indicated the recoded nucleotide.
c) Dinucleotide number: position of the dinucleotide required for the splicing process and recognize from the spliceosome, according toSEQ ID NO:1 (the vectorNTI map plasmid #277).
d) Strand: strand of the vector sequence(SEQ ID NO:1)where the splice site has been identified
e) Recoded site: name of the recoded site
f) Recoded nucleotide number: nucleotide position of the recoded nucleotide
g) Recoded nucleotide: each G of the dinucleotide has been recoded in A

In one embodiment at least one of the nucleotides corresponding to the splice site is replaced by another nucleotide.

In another embodiment, the splice site G is changed to A.

In one aspect the polynucleotide of the present invention is employed in the aforementioned uses. In one embodiment, the polynucloetide is used in therapy.

According to another aspect of the present invention there is provided a polynucleotide sequence comprising a HS3 region of b-globin locus control region nucleotide sequence wherein at least one of the following splice sites (forward and/or reverse) is inactivated (i.e., at least one of the nucleotides corresponding to the following splice sites is inactivated):

Splice acceptor	**corresponds to site shown in**
SA A	nucleotides 8106-8107 of SEQ ID NO: 2
SA B	nucleotides 8067-8068 of SEQ ID NO: 2
SA C	nucleotides 7474-7475 of SEQ ID NO: 2
SA D	nucleotides 5423-5424 of SEQ ID NO: 2
Splice donor	**corresponds to site shown in**
SD A	nucleotides 7912-7913 of SEQ ID NO: 2
SD B	nucleotides 7837-783 of SEQ ID NO: 2
SD C	nucleotides 7821-7822 of SEQ ID NO: 2
SD D	nucleotides 7797-7798 of SEQ ID NO: 2
SD E	nucleotides 7367-7668 of SEQ ID NO: 2
SD F	nucleotides 7363-73642 of SEQ ID NO: 2

In various embodiments at least 2, 3, 4, 5, 6, 7, 8 or 9 of the splice sites shown in the HS3 region of b-globin locus control region are inactivated.

The present invention also provides a viral vector comprising the polynucleotide sequence of the present invention, preferably in the form of a lentiviral vector particle, even more preferably derived from HIV or EIAV.

The present invention further provides a packaging, producer or host cell comprising the polynucleotide sequence or vector of the present invention.

The present invention additional provides a pharmaceutical composition comprising a polynucleotide sequence, vector or cell according to the present invention.

The gene vector or gene transfer vector of the present invention may be used to deliver a transgene to a site or cell of interest. The polynucleotide of the present invention may be delivered to a target site by a viral or non-viral vector, but is preferably delivered in a viral, more preferably, lenitviral vector in which the polynucleotide forms back of the viral backbone.

A vector is a tool that allows or facilitates the transfer of an entity from one environment to another. By way of example, some vectors used in recombinant DNA techniques allow entities, such as a segment of DNA (such as a heterologous DNA segment, such as a heterologous cDNA segment), to be transferred into a target cell. Optionally, once within the target cell, the vector may then serve to maintain the heterologous DNA within the cell or may act as a unit of DNA replication.

The term "vector particle" refers to the packaged retroviral vector, that is preferably capable of binding to and entering target cells. The components of the particle, as already discussed for the vector, may be modified with respect to the wild type retrovirus. For example, the Env proteins in the proteinaceous coat of the particle may be genetically modified in order to alter their targeting specificity or achieve some other desired function.

Preferably, the viral vector preferentially transduces a certain cell type or cell types.

More preferably, the viral vector is a targeted vector, that is it has a tissue tropism which is altered compared to the native virus, so that the vector is targeted to particular cells.

According to another aspect of the present invention there is provided a set of DNA constructs for producing the viral vector particle comprising a DNA construct encoding a packagable vector genome comprising a polynucleotide of the present invention, and optionally a transgene. By packagable vector genome we mean that the vector genome is in an environment where it can be packaged into a viral vector particle. This generally requires the present of Gag-Pol and Env.

According to another aspect of the present invention there is provided a process for preparing a viral vector particle comprising introducing the set of DNA constructs of claim into a host cell, and obtaining the viral vector particle.

According to another aspect of the present invention there is provided a viral vector particle produced by the process of the present invention.

According to another aspect of the present invention there is provided a pharmaceutical composition comprising the gene vector or vector particle according to the present invention together with a pharmaceutically acceptable diluent, excipient or carrier.

According to a further aspect of the present invention there is provided a cell infected or transduced with the vector particle of the present invention. The cell may be transduced or infected in an in vivo or in vitro scenario. The cell may be derived from or form part of an animal, preferably a mammal, such as a human or mouse. Thus it will be appreciated that the present invention is useful in providing transgenic animals e.g., for use as disease models. In one embodiment, the mammal is a non-human mammal.

Key features provided by the present invention include:

1) Novel vector designs with reduced impact on cellular splicing are likely to be the next generation of lentiviral vectors to be used in clinical applications (safer vectors).
2) Identified the relevant portions of the LV backbone involved in aberrant splicing.
3) Established and validated novel PCR techniques and quantitative assays to evaluate the types and amount of aberrant splicing for any integrating vector.
4) Integration of lentiviral vectors within transcribed regions causes abnormal splicing in a variable but significant percentage of targeted genes in all tested cell types.
5) Abnormal splicing is due to the usage of constitutive and cryptic splice signals located on both strands of the integrated provirus.
6) The proportion of aberrant, alternatively spliced transcripts is on average low compared to constitutively spliced transcripts.
7) The relative usage of cryptic splice sites is proportional to their homology to the mammalian splice consensus sequences.

8) The strength of constitutive splice signals in the targeted gene does not predict the extent of vector-induced alternative splicing.
9) Systematic analysis of alternatively spliced transcripts can be used to recode vector backbones and reduce their potential genotoxicity

According to another aspect of the invention there is provided a vector comprising an miRNA target sequence wherien said miRNA target sequence is positioned upstream of a splice donor site or downstream of a splice acceptor site, wherein said splice donor or splice acceptor site is responsible for splicing events that generate unwanted fusion transcripts comprisng vector sequences and cellular mRNAs, wherein said miRNA target sequence casues degradation of said unwanted fusion transcripst in a cell comprising a corresponding enogenous miRNA.

Preferably the miRNA target sequence is recognised by enodenous miRNA expressed in hematopoietic or hepatic cells such that the fusion transcript is selectively degraded in said cells.

Preferably the miRNA target sequence is one targeted by hsa-mir-142as (also called hsa-mir-142-3p) miRNA, let-7a, mir-15a, mir-16, mir-17-5p, mir-19, mir-142-5p, mir-145 and/or mir-218 miRNA.

Preferably the vector is a lentiviral vector.

Preferably the splice acceptor or splice donor site is a splice site recited herein.

Detailed description

Various preferred features and embodiments of the present invention will now be described by way of non-limiting examples.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA and immunology, which are within the capabilities of a person of ordinary skill in the art. Such techniques are explained in the literature. See, for example, J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel,

F. M. et al. (1995 and periodic supplements; *Current Protocols in Molecular Biology*, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.); B. Roe, J. Crabtree, and A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; J. M. Polak and James O'D. McGee, 1990, *In Situ Hybridization: Principles and Practice*; Oxford University Press; M. J. Gait (Editor), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; D. M. J. Lilley and J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA* Methods in Enzymology, Academic Press; and E. M. Shevach and W. Strober, 1992 and periodic supplements, *Current Protocols in Immunology*, John Wiley & Sons, New York, NY. Each of these general texts is herein incorporated by reference.

Oncogenesis induced by insertional mutagenesis with gene therapy vectors occurs mainly by deregulation of protooncogenes found at or nearby the insertion site. Proto-oncogene activation occurs by an enhancer-mediated mechanism or by a process of splicing capture which generates chimeric transcripts comprising portions of vector and cellular mRNAs. Although the activation of oncogenes may be reduced by the use of self-inactivating design and moderate cellular promoters, how to reduce genotoxic splicing capture events and aberrant transcript formation triggered by vector integration is still unclear. In this perspective, we developed a new PCR technique named Linear Amplification-Mediated PCR on cDNA (cLAM-PCR) which allows retrieving aberrantly spliced mRNAs that contain LV sequences fused with cellular transcripts.

We applied cLAM-PCR on lentiviral vector (LV)-transduced cell lines and primary human HSCs and identified several established and previously unknown splice sites within the LV backbone that participate in the aberrant splicing process with variable efficiency. Preliminary results with different LV designs show that the integrated LV can perturb the processing of cellular transcripts by interacting with the cellular splicing machinery and fusing with its own splice sites to cellular splice sites both upstream and downstream the integration site. Moreover, qPCR on different LV portions allowed us to identify different splice sites as major or minor contributors to the aberrant splicing process. This strategy will allow characterizing the mechanism and genetic features that modulate vector-induced aberrant splicing. In a biosafety perspective, elimination of splice sites within the lentiviral backbone will be instrumental in enhancing the safety of LVs.

In the present invention the splice sites may be inactivated by altering the sequence from the wild type sequence, including replacing a nucleotide by another nucleotide.

As used herein, the term "replaced by another nucleotide" means replaced by a nucleotide that differs from the wild type sequence. The replacements are made such that the relevant splice donor site or splice acceptor sites are removed.

The splice site may be readily isolated and mutated as described below, in order to construct nucleic acid molecules comprising a splice site comprising one or more mutations which substantially reduce the splicing of the sequence, as compared to unmutated sequence. As utilised herein, it should be understood that "unmutated sequence" refers to native or wild-type splice site.

It should be noted that in this application nucleotide positions are referred to by reference to a position in the Figures, Table 1 and sequence listings. However, when such references are made, it will be understood that the invention is not to be limited to the exact sequence as set out in the Figures, Table 1 and sequence listings but includes variants and derivatives thereof. Thus, identification of nucleotide locations in other sequences are contemplated (i.e., nucleotides at positions which the skilled person would consider correspond to the positions identified in SEQ ID NO1, 2 or 3). The person skilled in the art can readily align similar sequences and locate the same nucleotide locations.

Construction of Splice site mutants

Splice site mutants of the present invention may be constructed using a variety of techniques. For example, mutations may be introduced at particular loci by synthesising oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence comprises a derivative having the desired nucleotide insertion, substitution, or deletion.

Alternatively, oligonucleotide-directed site-specific (or segment specific) mutagenesis procedures may be employed to provide an altered sequence having particular codon altered according to the substitution, deletion, or insertion required. Deletion or truncation derivatives of splice site mutants may also be constructed by utilising convenient restriction endonuclease sites adjacent to the desired deletion.

Subsequent to restriction, overhangs may be filled in, and the DNA religated. Exemplary methods of making the alterations set forth above are disclosed by Sambrook et al. (*Molecular cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, 1989).

Splice site mutants may also be constructed utilising techniques of PCR mutagenesis, chemical mutagenesis, chemical mutagenesis (Drinkwater and Klinedinst, 1986) by forced nucleotide misincorporation (*e.g.,* Liao and Wise, 1990), or by use of randomly mutagenised oligonucleotides (Horwitz et al., 1989).

In a preferred embodiment of the present invention, the nucleotides are modified taking note of the genetic code such that a codon is changed to a degenerate codon which codes for the same amino acid residue. In this way, it is possible to make coding regions of the protein of interest which encode wild type protein but which do not contain a functional splice site.

In a particular aspect of the present invention elimination of unwanted fusion transcripts may be achieved by adding tags recognized by specific microRNAs that will trigger their selective degradation. For example, if detrimental mutations of splice sites cannot be re-recoded without reducing the vector titer then alternative micro-RNA-based strategies can be used. Indeed we identified exonic LV sequences most frequently present in chimeric transcripts. Therefore these LV exon sequences could be tagged by sequences complementary to microRNAs highly active in, for example, hematopoietic or hepatic cells (but not in vector-producer cells). Aberrant transcripts will be thus recognized by the endogenous microRNA and selectively degraded by the miRNA pathway.

Splice sites

The proportion of RNA which is removed (or "spliced out") during splicing is typically called an intron, and the two pieces of RNA either side of the intron that are joined by splicing are typically called exons.

A splice donor site is a site in RNA which lies at the 5' side of the RNA which is removed during the splicing process and which contains the site which is cut and rejoined to a nucleotide residue within a splice acceptor site. Thus, a splice donor site

is the junction between the end of an exon and the start of the intron, typically terminating in the dinucleotide GU. In a preferred embodiment of the present invention, one or both of the terminal GU dinucleotides (or GT dinucleotides in the corresponding DNA sequence) of the splice donor site is/are altered to remove the splice site.

A splice acceptor site is a site in RNA which lies at the 3' side of the RNA which is removed during the splicing process and which contains the site which is cut and rejoined to a nucleotide residue within a splice donor site. Thus, a splice acceptor site is the junction between the end of an intron (typically terminating with the dinucleotide AG) and the start of the downstream exon. In a preferred embodiment of the present invention, one or both of the terminal AG dinucleotides of the splice acceptor site is/are altered to remove the splice site.

Polynucleotides

Polynucleotides used in the invention may comprise DNA or RNA. They may be single-stranded or double-stranded. It will be understood by a skilled person that numerous different polynucleotides can encode the same polypeptide as a result of the degeneracy of the genetic code. In addition, it is to be understood that skilled persons may, using routine techniques, make nucleotide substitutions that do not affect any polypeptide sequence encoded by the polynucleotides used in the invention to reflect the codon usage of any particular host organism in which the polypeptides are to be expressed. The polynucleotides may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or life span of polynucleotides of the invention.

Retroviruses

During the past decade, gene therapy has been applied to the treatment of disease in hundreds of clinical trials. Various tools have been developed to deliver genes into human cells; among them, genetically engineered retroviruses, including lentiviruses, are currently amongst the most popular tool for gene delivery. Most of the systems contain vectors that are capable of accommodating genes of interest and helper cells that can provide the viral structural proteins and enzymes to allow for the generation of vector-containing infectious viral particles. Retroviridae is a family of retroviruses

that differs in nucleotide and amino acid sequence, genome structure, pathogenicity, and host range. This diversity provides opportunities to use viruses with different biological characteristics to develop different therapeutic applications. As with any delivery tool, the efficiency, the ability to target certain tissue or cell type, the expression of the gene of interest, and the safety of retroviral-based systems are important for successful application of gene therapy. Significant efforts have been dedicated to these areas of research in recent years. Various modifications have been made to retroviral-based vectors and helper cells to alter gene expression, target delivery, improve viral titers, and increase safety. The present invention represents an improvement in this design process in that it acts to efficiently deliver genes of interest into such viral vectors.

Viruses are logical tools for gene delivery. They replicate inside cells and therefore have evolved mechanisms to enter the cells and use the cellular machinery to express their genes. The concept of virus-based gene delivery is to engineer the virus so that it can express the gene of interest. Depending on the specific application and the type of virus, most viral vectors contain mutations that hamper their ability to replicate freely as wild-type viruses in the host.

Viruses from several different families have been modified to generate viral vectors for gene delivery. These viruses include retroviruses, lentivirus, adenoviruses, adeno-associated viruses, herpes simplex viruses, picornaviruses, and alphaviruses. The present invention preferably employs retroviruses, including lentiviruses.

An ideal retroviral vector for gene delivery must be efficient, cell-specific, regulated, and safe. The efficiency of delivery is important because it can determine the efficacy of the therapy. Current efforts are aimed at achieving cell-type-specific infection and gene expression with retroviral vectors. In addition, retroviral vectors are being developed to regulate the expression of the gene of interest, since the therapy may require long-lasting or regulated expression. Safety is a major issue for viral gene delivery because most viruses are either pathogens or have a pathogenic potential. It is important that during gene delivery, the patient does not also inadvertently receive a pathogenic virus that has full replication potential.

Retroviruses are RNA viruses that replicate through an integrated DNA intermediate. Retroviral particles encapsidate two copies of the full-length viral RNA, each copy containing the complete genetic information needed for virus replication. Retroviruses

possess a lipid envelope and use interactions between the virally encoded envelope protein that is embedded in the membrane and a cellular receptor to enter the host cells. Using the virally encoded enzyme reverse transcriptase, which is present in the virion, viral RNA is reverse transcribed into a DNA copy. This DNA copy is integrated into the host genome by integrase, another virally encoded enzyme. The integrated viral DNA is referred to as a provirus and becomes a permanent part of the host genome. The cellular transcriptional and translational machinery carries out expression of the viral genes. The host RNA polymerase II transcribes the provirus to generate RNA, and other cellular processes modify and transport the RNA out of the nucleus. A fraction of viral RNAs are spliced to allow expression of some genes whereas other viral RNAs remain full-length. The host translational machinery synthesizes and modifies the viral proteins. The newly synthesized viral proteins and the newly synthesized full-length viral RNAs are assembled together to form new viruses that bud out of the host cells.

Based on their genome structures, retroviruses can be classified into simple and complex retroviruses. Simple and complex retroviruses encode gag (group-specific antigen), pro (protease), pol (polymerase), and env (envelope) genes. In addition to these genes, complex retroviruses also encode several accessory genes.

Retroviruses can also be classified into oncoviruses, lentiviruses, and spumaviruses. Most oncoviruses are simple retroviruses. Lentiviruses, spumaviruses, and some oncoviruses are complex retroviruses.

When a replication-competent retrovirus infects a natural host cell, it can form a provirus in the host genome, express viral genes, and release new infectious particles to infect other hosts. In most gene therapy applications, it is not desirable to deliver a replication-competent virus into a patient because the virus may spread beyond the targeted tissue and cause adverse pathogenic effects. Therefore, in most retroviral systems designed for gene delivery, the viral components are divided into a vector and a helper construct to limit the ability of the virus to replicate freely.

The term vector generally refers to a modified virus that contains the gene(s) of interest (or transgene) and cis-acting elements needed for gene expression and replication. Most vectors contain a deletion(s) of some or all of the viral protein coding sequences so that they are not replication-competent. Helper constructs are designed to express viral genes lacking in the vectors and to support replication of

the vectors. The helper function is most often provided in a helper cell format although it can also be provided as a helper virus or as cotransfected plasmids.

Helper cells are engineered culture cells expressing viral proteins needed to propagate retroviral vectors; this is generally achieved by transfecting plasmids expressing viral proteins into culture cells. Most helper cell lines are derived from cell clones to ensure uniformity in supporting retroviral vector replication. Helper viruses are not used often because of the likelihood that a replication-competent virus could be generated through high frequency recombination. Helper functions can also be provided by transient transfection of helper constructs to achieve rapid propagation of the retroviral vectors.

Most retroviral vectors are maintained as bacterial plasmids to facilitate the manipulation and propagation of the vector DNA. These double-stranded DNA vectors can be introduced into helper cells by conventional methods such as DNA transfection, lipofection, or electroporation. The helper cell shown expresses all of the viral proteins (Gag, Gag-Pol, and Env) but lacks RNA containing the packaging signal. Viral RNA is necessary for the formation and release of infectious viral particles, but it is not necessary for the formation of "empty" noninfectious viral particles. When the vector DNA is introduced into the helper cells, vector RNA containing a packaging signal is transcribed and efficiently packaged into viral particles. The viral particles contain viral proteins expressed from helper constructs and RNA transcribed from the vector. These viral particles can infect target cells, reverse transcribe the vector RNA to form a double-stranded DNA copy, and integrate the DNA copy into the host genome to form a provirus. This provirus encodes the gene(s) of interest and is expressed by the host cell machinery. However, because the vector does not express any viral proteins, it cannot generate infectious viral particles that can spread to other target cells.

Helper cells are designed to support the propagation of retroviral vectors. The viral proteins in the helper cells are expressed from helper constructs that are transfected into mammalian cells. Helper constructs vary in their mode of expression and in the genes they encode.

One-Genome Helper Constructs

In helper cell lines that were initially developed, all of the viral genes were expressed from one helper construct. Examples of these helper cells are C3A2 and -2. The helper constructs for these cell lines were cloned proviral DNAs that lacked the packaging signals. These helper cells can support efficient propagation of retroviral vectors. However, a major problem with these helper cells is that replication-competent viruses can be frequently generated during the propagation of the viral vector. The helper construct contains most of the viral genome and thus shares significant sequence homology with the retroviral vector. The sequence homology can facilitate recombination between the helper construct and the retroviral vector to generate replication-competent viruses. Although the helper RNA lacks the packaging signal, it can still be packaged into a virion with a low efficiency (approximately 100- to 1,000-fold less than RNAs containing). Retroviral recombination occurs frequently between the two copackaged viral RNAs to generate a DNA copy that contains genetic information from both parents. If the helper RNA and the vector RNA are packaged into the same virion, the large regions of sequence homology between the two RNAs can facilitate homologous recombination during reverse transcription to generate a replication-competent virus. A similar recombination event can also occur between the helper RNA and RNA derived from an endogenous virus at a lower efficiency to generate replication-competent viruses.

Split-Genome Helper Constructs

The safety concern associated with the generation of replication-competent viruses has provoked the design of many helper cell lines using "split genomes", including CRIP, GP+envAm12, and DSN. In these helper cells, the viral Gag/Gag-Pol polyproteins are expressed from one plasmid and the Env proteins are expressed from another plasmid. Furthermore, the two helper constructs also contain deletions of viral cis-acting elements to reduce or eliminate sequence homology with the retroviral vector. In these helper cells, genes encoding viral proteins are separated into two different constructs and the viral cis-acting elements are located in the vector. Therefore, several recombination events have to occur to reconstitute the viral genome. In addition, reducing the regions of homology decreases the probability that these recombination events will occur. Therefore, helper cells containing split-genome helper constructs are considered safer than helper cells containing one-genome helper constructs.

Inducible Helper Constructs

In contrast to the helper cell lines described above that express viral proteins constitutively, some helper cell lines have been designed to express the viral proteins in an inducible manner. One rationale for the generation of an inducible helper cell line is that some viral proteins are cytotoxic and cannot be easily expressed at high levels. By using an inducible system, expression of the cytotoxic proteins can be limited to the stage in which virus is propagated. By controlling the expression of the cytotoxic proteins, high viral titers can be achieved. Examples of the inducible helper cells include the 293GPG cells and HIV-1 helper cell lines.

Transient Transfection Systems

With the development of efficient transfection methods, transient transfection systems have also been developed for propagation of retroviral vectors. In these systems, helper functions are generally expressed from two different constructs, one expressing gag-pol and another expressing env. These two constructs generally share little sequence homology. The retroviral vector and the helper constructs are transfected into cells, and viruses are harvested a few days after transfection

Systems That Generate Pseudotyped Viruses

Pseudotyping refers to viral particles containing a viral genome from one virus and part (or all) of the viral proteins from a different virus. The most common form of pseudotyping involves one virus using the envelope protein of another virus. Some of the helper cell lines contain helper constructs that express gag-pol from one virus and env from another virus. Since the Gag polyproteins select the viral RNA, the viral vector to be propagated contains an RNA that is recognized by the Gag polyprotein expressed in these cells. However, the viral particles produced contain the Env protein derived from another virus. Therefore, these viral particles can only infect cells that express a receptor that can interact with the heterologous envelope protein. For example, the helper cell line PG13 expresses gag-pol from MLV and env from gibbon ape leukemia virus (GaLV). Because the PG13 cell line expresses MLV Gag polyprotein, it can efficiently package MLV-based retroviral vectors. It has also been shown that some envelopes derived from viruses of a different family can also pseudotype retroviruses and generate infectious viral particles. For example, the G protein of vesicular stomatitis virus (VSV), a rhabdovirus, can be used to generate

pseudotyped retroviral vectors. These VSV G pseudotyped viruses exhibit a very broad host range and can infect a variety of cells that cannot normally be infected with retroviruses. Other envelopes that can be used for vector pseudotyping are those of the following viruses: the RD114 endogenous feline retrovirus, which effectively targets hematopoietic cells, the Lymphocytic ChorioMeningitis Virus (LCMV), the Rabies virus, the Ebola and Mokola viruses, the Ross River and Semliki Forest virus, and the baculovirus gp64 envelope.

Pseudotyping may involve for example a retroviral genome based on a lentivirus such as an HIV or equine infectious anaemia virus (EIAV) and the envelope protein may for example be the amphotropic envelope protein designated 4070A. Alternatively, envelope protein may be a protein from another virus such as an Influenza haemagglutinin. In another alternative, the envelope protein may be a modified envelope protein such as a mutant, truncated or engineered envelope protein (such as the engineered RD114 envelope). Modifications may be made or selected to introduce targeting ability or to reduce toxicity or for another purpose.

Systems Containing Genetically Modified env for Cell or Tissue Targeting
Interactions between the viral envelope proteins and the cellular receptors determine the host range of the virus. Strategies have been developed to target virus delivery into certain cell types by modifying the viral Env. After translation and modification, the SU portion of Env interacts with a cellular receptor. The modification of the SU portion of Env is often achieved by deletion of a part of the coding region for SU and replacing it with regions of other proteins. Proteins that have been used to modify the SU portion of Env include erythropoietin, heregulin, insulin-like growth factor I, and single-chain variable fragment antibodies against various proteins.

Vectors Derived from Lentiviruses

Lentiviruses have been shown to infect nondividing, quiescent cells. Lentiviruses are complex retroviruses that may need to express accessory proteins for regulation of their replication cycle. Some of these accessory proteins bind to regions of the viral genome to regulate gene expression. Therefore, lentivirus-based vectors need to incorporate additional cis-acting elements so that efficient viral replication and gene expression can occur. As examples of lentivirus-based vectors, HIV-1- and HIV-2-based vectors are described below.

The HIV-1 vector contains cis-acting elements that are also found in simple retroviruses. It has been shown that sequences that extend into the gag open reading frame are important for packaging of HIV-1. Therefore, HIV-1 vectors often contain the relevant portion of gag in which the translational initiation codon has been mutated. In addition, most HIV-1 vectors also contain a portion of the env gene that includes the RRE. Rev binds to RRE, which permits the transport of full-length or singly spliced mRNAs from the nucleus to the cytoplasm. In the absence of Rev and/or RRE, full-length HIV-1 RNAs accumulate in the nucleus. Alternatively, a constitutive transport element from certain simple retroviruses such as Mason-Pfizer monkey virus can be used to relieve the requirement for Rev and RRE. Efficient transcription from the HIV-1 LTR promoter requires the viral protein Tat. Therefore, it is important that Tat is expressed in target cells if efficient transcription from the HIV-1 LTR is needed. The need for Tat expression can be met by expressing the Tat gene from the retroviral vector. Alternatively, expressing the gene of interest from a heterologous internal promoter can circumvent the need for Tat expression.

Most HIV-2-based vectors are structurally very similar to HIV-1 vectors. Similar to HIV-1-based vectors, HIV-2 vectors also require RRE for efficient transport of the full-length or singly spliced viral RNAs.

It has also been demonstrated that the HIV-1 vector can be propagated to high viral titers using viral proteins from simian immunodeficiency virus. In one system, the vector and helper constructs are from two different viruses, and the reduced nucleotide homology may decrease the probability of recombination. In addition to vectors based on the primate lentiviruses, vectors based on feline immunodeficiency virus have also been developed as an alternative to vectors derived from the pathogenic HIV-1 genome. The structures of these vectors are also similar to the HIV-1 based vectors.

Design of Retroviral Vectors

Retroviral vectors may contain many different modifications that serve various purposes for the gene therapist. These modifications may be introduced to permit the expression of more than one gene, regulate gene expression, activate or inactivate the viral vectors, and eliminate viral sequences to avoid generation of a replication-competent virus. Some examples of these modifications are described below.

A. Standard Vectors

1. U3 Promoter-Driven Gene Expression. Full-length viral RNA is expressed from the retroviral promoter located in the U3 region of the 5' LTR. The viral RNA contains the R, U5, 5' untranslated region, a gene of interest, 3' untranslated region, U3, and R. The gene inserted between the 5' and 3' untranslated regions can be translated from the full-length RNA that is transcribed from the U3 promoter.

During the propagation of viral stocks, it is often desirable to express a selectable marker gene in the vector so that helper cells transfected or infected by the viral vectors can be selected. Therefore, it is often necessary to design retroviral vectors that express a selectable marker gene as well as a gene of interest. Drug resistance genes are frequently used as selectable markers, but other marker genes, such as the green fluorescent protein gene, can also be used to select for transfected or infected cells. The expression of two genes in a retroviral vector can be achieved by expressing the 3' gene by using an internal promoter, RNA splicing, or an internal ribosomal entry site (IRES).

2. Vectors That Use an Internal Promoter to Express Additional Genes. An example of gene expression from a retroviral vector containing an internal promoter where, e.g., the full-length RNA that is expressed from the viral U3 promoter is used to translate a first gene of interest(s). The subgenomic RNA that is expressed from the internal promoter is used to translate a second gene of interest(s).

3. Vectors That Use Splicing to Express Additional Genes. Retroviruses express env by regulated splicing. The splice donor site that is used to express env is located in the 5' untranslated region of retroviruses. During replication, some full-length viral RNAs are spliced to produce subgenomic viral RNAs that are used to express the Env proteins. Splicing vectors were developed by using the same principle to express two different genes by using the viral splice donor and splice acceptor sites. The advantage of splicing vectors is that only one promoter is necessary, and any potential for promoter interference is eliminated.

4. Vectors That Use Translational Control Signals to Express Additional Genes. It was first demonstrated in picornaviruses that sequences in the mRNA can serve as signals that allow the ribosome to bind to the middle of an mRNA and translate a

gene far from the 5' end of the mRNA. These sequences (named IRES), are now commonly used in retroviral vectors. In addition to the IRES sequences identified in picornaviruses, IRES sequences have also been identified in the 5' untranslated regions of some retroviruses such as MLV, SNV, and an endogenous virus like particle (VL30). Therefore, it is also possible to use these retroviral IRES sequences to express a second gene. Other sequences allowing expression of multiple proteins from a single transcript are self-cleaving 2A-like peptides (also called CHYSEL, cis-acting hydrolase elements) derived from the Foot-and-Mouth disease virus and other picoRNA viruses. Alternatively bidirectional promoters can be used to express two genes from the same promoter.

B. Double-Copy Vectors

The fact that the LTR sequences are duplicated in retroviral vectors has been exploited to construct vectors containing two copies of the gene of interest. For example, the first set of double-copy vectors contains the gene of interest in the U3 region upstream of the viral. These genes are expressed using either an RNA polymerase II promoter or an RNA polymerase III promoter. This strategy has been shown to successfully increase the level of gene expression. In another example of a double-copy vector the vector contains the gene of interest in the middle of the R region.

C. Self-Inactivating Vectors

One safety concern associated with using retroviral vectors for gene therapy is that a replication-competent virus can be generated during propagation of the vectors, which can lead to inadvertent spread of the therapeutic vector to nontarget tissues. To address this concern, a class of vectors was designed to undergo self-inactivation. The principle is that after gene delivery, the vector will delete some of the cis-acting elements needed to complete another round of replication. Therefore, even in the presence of a replication-competent virus, these vectors cannot be transferred to other target cells efficiently. The generation of a replication-competent virus sometimes involves recombination between the defective helper plasmid and the vector encoding the gene of interest. Therefore, another possible benefit of the self-inactivating vector is that it may decrease the probability of generating a replication-competent virus.

1. U3 Minus Vectors. U3 minus vectors were the first self-inactivating retroviral vectors to be developed. These vectors are designed to delete the viral U3 promoter during reverse transcription so that the provirus in the target cell lacks a viral

promoter. In these vectors, the U3 of the 5' LTR is intact, whereas the U3 of the 3' LTR is inactivated by a large deletion. The RNA generated from this vector contains R, U5, 5' untranslated region, gene(s) of interest, 3' untranslated region, a deleted U3, and R. During reverse transcription, the U3 at the 3' end of the viral RNA is normally used as a template to generate the LTR. Therefore, the viral DNA that is synthesized from the U3 minus vector through reverse transcription contains deleted U3 sequences in both LTRs. Since the viral promoter is deleted during reverse transcription, the gene of interest is under the control of an internal promoter. The advantage of the U3 minus vector is that it is potentially safer, since the probability of generation of a replication-competent virus is reduced. However, at a low frequency, recombination during DNA transfection can occur to regenerate the U3 at the 3' LTR. If this occurs, the resulting vector will still contain the promoter in the U3 and thus retain two complete LTRs. Additional modifications have been made in some U3 minus vectors to decrease the homology between the 5' and 3' LTRs, which reduces the probability of recombination and regeneration of an intact LTR during DNA transfection.

2. Cre/loxP Vectors. The Cre recombinase, a naturally occurring site-specific recombinase of bacteriophage P1, recognizes a 32-bp sequence named loxP. Cre can efficiently mediate site-specific recombination using two loxP sites separated by sequences of variable lengths. The recombination events include deletion, insertion, and inversion of the sequences between the loxP sites. This system has been exploited to develop self-inactivating retroviral vectors (Choulika et al., 1996; Russ et al., 1996). An example of such a vector contains an intact 5' LTR and all of the cis-acting elements needed for retroviral replication. The vector contains the cre recombinase gene that is expressed using an internal promoter. The 3' LTR has been modified by insertion of several sequences in the U3, including a loxP site, a promoter, and a gene of interest; in addition, the 3' U3 often contains a deletion to reduce the promoter activity. The full-length viral RNA is packaged into virion, and upon infection of target cells, the viral RNA is reverse-transcribed. The 3' U3 sequence is used as a template to synthesize both LTRs; consequently, the sequences in both LTRs contain a copy of the loxP site, a promoter, and a gene of interest. The cre gene is expressed, and the Cre recombinase is synthesized in the infected target cells. The Cre recombinase then mediates the deletion of sequences between the two loxP sites in the viral DNA, which results in deletion of the 5' LTR, the 5' untranslated region, the internal promoter, and cre. As a result, the provirus in the target cells contains only one LTR that expresses the gene of interest.

Using the same principle, the Cre/loxP system can be used to delete different sequences in the retroviral vector as well as delete portions of the helper construct in the packaging cells. Another application of the Cre/loxP system is that it can be used to delete the selectable marker from a retroviral vector after the viral DNA is integrated into the chromosome of the target cells. The selectable marker is included in the vector so that helper cells transfected with the vector DNA can be selected. Deletion of the selectable marker is desirable because the presence of the selectable marker can lead to promoter interference or an immune response against the transduced cells. Deletion of the selectable marker is accomplished by insertion of two loxP sites that flank the selectable marker gene. After the vector is introduced into target cells by infection, the target cells are infected with another vector that expresses the Cre recombinase. The Cre recombinase then deletes sequences between the two loxP sites, which include the selectable marker. As a result, the final provirus expresses only the gene of interest.

D. Self-Inactivating and Self-Activating Vectors

Depending on the properties and effects of the gene products, it may be desirable to have an inactivated gene of interest in the helper cells and activate this gene after it is delivered to target cells. For example, if the product from the gene of interest is cytotoxic, then expressing the gene in helper cells would result in toxicity and most likely reduce or eliminate viral production. A series of vectors have been generated to simultaneously activate a gene and inactivate the vector during gene delivery. This is accomplished by the frequent deletion of directly repeated sequences during reverse transcription. If directly repeated sequences are present in a virus, one copy of the direct repeat and all of the sequences between the two repeats can be deleted at high frequencies during reverse transcription. This property of reverse transcriptases has been exploited to generate the self-activating and self-inactivating retroviral vectors.

E. Vectors Targeted to Specific Cells

An important goal for gene therapists is to develop a means to target gene delivery to specific cell types or tissues. At least two strategies have been used in an effort to target gene delivery using retroviral vectors. One strategy is designed to control gene delivery at the point of virus entry into the host cell by using natural or genetically engineered envelope proteins that interact with cell-type-specific receptors. Another

strategy is designed to control expression of the therapeutic gene in specific cell types by using tissue-specific promoters.

F. Vectors That Utilize Cell-Type-Specific Promoters

Promoters that are active in certain tissues or respond to certain reagents can be used to regulate the expression of a gene of interest. These promoters can be inserted between the LTRs of a retroviral vector. Alternatively, the regulated promoter can be used to replace the viral promoter in the U3 region. The design of a retroviral vector with an internal tissue-specific promoter is similar to that of other retroviral vectors containing internal promoters.

Virus Host Range

1. Considerations for Envelope Selection and Virus Host Range. The nature of the viral envelope protein determines whether a certain virus can enter a target cell. Therefore, it is important to consider whether the target cells have the correct cell surface receptor before the selection of an envelope protein that will be used for virus production (as discussed above).

The retroviral vector particle according to the invention will also be capable of transducing cells which are slowly-dividing, and which non-lentiviruses such as MLV would not be able to efficiently transduce. Slowly-dividing cells divide once in about every three to four days including certain tumour cells. Although tumours contain rapidly dividing cells, some tumour cells especially those in the centre of the tumour, divide infrequently. Alternatively the target cell may be a growth-arrested cell capable of undergoing cell division such as a cell in a central portion of a tumour mass or a stem cell such as a haematopoietic stem cell or a CD34-positive cell. As a further alternative, the target cell may be a precursor of a differentiated cell such as a monocyte precursor, a CD33-positive cell, or a myeloid precursor. As a further alternative, the target cell may be a differentiated cell such as a neuron, astrocyte, glial cell, microglial cell, macrophage, monocyte, epithelial cell, endothelial cell or hepatocyte. Target cells may be transduced either *in vitro* after isolation from a human individual or may be transduced directly *in vivo*.

Administration

The delivery vehicles of the present invention may be administered to a patient or used to produce a transgenic plant or non-human animal. A skilled worker would be able to

determined appropriate dosage rates. The term "administered" includes delivery by viral or non-viral techniques. Viral delivery mechanisms include but are not limited to adenoviral vectors, adeno-associated viral (AAV) vectors, herpes viral vectors, retroviral vectors, lentiviral vectors, and baculoviral vectors etc as described above. Non-viral delivery mechanisms include lipid mediated transfection, liposomes, immunoliposomes, lipofectin, cationic facial amphiphiles (CFAs) and combinations thereof.

Diseases

The delivery of one or more therapeutic genes by a vector system according to the present invention may be used alone or in combination with other treatments or components of the treatment.

For example, the vector of the present invention may be used to deliver one or more transgene(s) useful in the treatment of the disorders listed in WO-A-98/05635. For ease of reference, part of that list is now provided: cancer, inflammation or inflammatory disease, dermatological disorders, fever, cardiovascular effects, haemorrhage, coagulation and acute phase response, cachexia, anorexia, acute infection, HIV infection, shock states, graft-versus-host reactions, autoimmune disease, reperfusion injury, meningitis, migraine and aspirin-dependent anti-thrombosis; tumour growth, invasion and spread, angiogenesis, metastases, malignant, ascites and malignant pleural effusion; cerebral ischaemia, ischaemic heart disease, osteoarthritis, rheumatoid arthritis, osteoporosis, asthma, multiple sclerosis, neurodegeneration, Alzheimer's disease, atherosclerosis, stroke, vasculitis, Crohn's disease and ulcerative colitis; periodontitis, gingivitis; psoriasis, atopic dermatitis, chronic ulcers, epidermolysis bullosa; corneal ulceration, retinopathy and surgical wound healing; rhinitis, allergic conjunctivitis, eczema, anaphylaxis; restenosis, congestive heart failure, endometriosis, atherosclerosis or endosclerosis.

In addition, or in the alternative, the vector of the present invention may be used to deliver one or more transgene(s) useful in the treatment of disorders listed in WO-A-98/07859. For ease of reference, part of that list is now provided: cytokine and cell proliferation/differentiation activity; immunosuppressant or immunostimulant activity (e.g. for treating immune deficiency, including infection with human immune deficiency virus; regulation of lymphocyte growth; treating cancer and many autoimmune diseases, and to prevent transplant rejection or induce tumour

immunity); regulation of haematopoiesis, e.g. treatment of myeloid or lymphoid diseases; promoting growth of bone, cartilage, tendon, ligament and nerve tissue, e.g. for healing wounds, treatment of burns, ulcers and periodontal disease and neurodegeneration; inhibition or activation of follicle-stimulating hormone (modulation of fertility); chemotactic/chemokinetic activity (e.g. for mobilising specific cell types to sites of injury or infection); haemostatic and thrombolytic activity (e.g. for treating haemophilia and stroke); antiinflammatory activity (for treating e.g. septic shock or Crohn's disease); as antimicrobials; modulators of e.g. metabolism or behaviour; as analgesics; treating specific deficiency disorders; in treatment of e.g. psoriasis, in human or veterinary medicine.

In addition, or in the alternative, the retroviral vector of the present invention may be used to deliver one or more transgenes(s) useful in the treatment of disorders listed in WO-A-98/09985. For ease of reference, part of that list is now provided: macrophage inhibitory and/or T cell inhibitory activity and thus, anti-inflammatory activity; anti-immune activity, i.e. inhibitory effects against a cellular and/or humoral immune response, including a response not associated with inflammation; inhibit the ability of macrophages and T cells to adhere to extracellular matrix components and fibronectin, as well as up-regulated fas receptor expression in T cells; inhibit unwanted immune reaction and inflammation including arthritis, including rheumatoid arthritis, inflammation associated with hypersensitivity, allergic reactions, asthma, systemic lupus erythematosus, collagen diseases and other autoimmune diseases, inflammation associated with atherosclerosis, arteriosclerosis, atherosclerotic heart disease, reperfusion injury, cardiac arrest, myocardial infarction, vascular inflammatory disorders, respiratory distress syndrome or other cardiopulmonary diseases, inflammation associated with peptic ulcer, ulcerative colitis and other diseases of the gastrointestinal tract, hepatic fibrosis, liver cirrhosis or other hepatic diseases, thyroiditis or other glandular diseases, glomerulonephritis or other renal and urologic diseases, otitis or other oto-rhino-laryngological diseases, dermatitis or other dermal diseases, periodontal diseases or other dental diseases, orchitis or epididimo-orchitis, infertility, orchidal trauma or other immune-related testicular diseases, placental dysfunction, placental insufficiency, habitual abortion, eclampsia, pre-eclampsia and other immune and/or inflammatory-related gynaecological diseases, posterior uveitis, intermediate uveitis, anterior uveitis, conjunctivitis, chorioretinitis, uveoretinitis, optic neuritis, intraocular inflammation, e.g. retinitis or cystoid macular oedema, sympathetic ophthalmia, scleritis, retinitis pigmentosa, immune and inflammatory components of degenerative fondus disease, inflammatory

components of ocular trauma, ocular inflammation caused by infection, proliferative vitreo-retinopathies, acute ischaemic optic neuropathy, excessive scarring, e.g. following glaucoma filtration operation, immune and/or inflammation reaction against ocular implants and other immune and inflammatory-related ophthalmic diseases, inflammation associated with autoimmune diseases or conditions or disorders where, both in the central nervous system (CNS) or in any other organ, immune and/or inflammation suppression would be beneficial, Parkinson's disease, complication and/or side effects from treatment of Parkinson's disease, AIDS-related dementia complex HIV-related encephalopathy, Devic's disease, Sydenham chorea, Alzheimer's disease and other degenerative diseases, conditions or disorders of the CNS, inflammatory components of stokes, post-polio syndrome, immune and inflammatory components of psychiatric disorders, myelitis, encephalitis, subacute sclerosing pan-encephalitis, encephalomyelitis, acute neuropathy, subacute neuropathy, chronic neuropathy, Guillaim-Barre syndrome, Sydenham chora, myasthenia gravis, pseudo-tumour cerebri, Down's Syndrome, Huntington's disease, amyotrophic lateral sclerosis, inflammatory components of CNS compression or CNS trauma or infections of the CNS, inflammatory components of muscular atrophies and dystrophies, and immune and inflammatory related diseases, conditions or disorders of the central and peripheral nervous systems, post-traumatic inflammation, septic shock, infectious diseases, inflammatory complications or side effects of surgery, bone marrow transplantation or other transplantation complications and/or side effects, inflammatory and/or immune complications and side effects of gene therapy, e.g. due to infection with a viral carrier, or inflammation associated with AIDS, to suppress or inhibit a humoral and/or cellular immune response, to treat or ameliorate monocyte or leukocyte proliferative diseases, e.g. leukaemia, by reducing the amount of monocytes or lymphocytes, for the prevention and/or treatment of graft rejection in cases of transplantation of natural or artificial cells, tissue and organs such as cornea, bone marrow, organs, lenses, pacemakers, natural or artificial skin tissue.

The present invention also provides a pharmaceutical composition for treating an individual by gene therapy, wherein the composition comprises a therapeutically effective amount of the vector of the present invention comprising one or more deliverable therapeutic and/or diagnostic transgenes(s) or a viral particle produced by or obtained from same. The pharmaceutical composition may be for human or animal usage. Typically, a physician will determine the actual dosage which will be most suitable for an individual subject and it will vary with the age, weight and response of the particular individual.

The composition may optionally comprise a pharmaceutically acceptable carrier, diluent, excipient or adjuvant. The choice of pharmaceutical carrier, excipient or diluent can be selected with regard to the intended route of administration and standard pharmaceutical practice. The pharmaceutical compositions may comprise as - or in addition to - the carrier, excipient or diluent any suitable binder(s), lubricant(s), suspending agent(s), coating agent(s), solubilising agent(s), and other carrier agents that may aid or increase the viral entry into the target site (such as for example a lipid delivery system).

Where appropriate, the pharmaceutical compositions can be administered by any one or more of: inhalation, in the form of a suppository or pessary, topically in the form of a lotion, solution, cream, ointment or dusting powder, by use of a skin patch, orally in the form of tablets containing excipients such as starch or lactose, or in capsules or ovules either alone or in admixture with excipients, or in the form of elixirs, solutions or suspensions containing flavouring or colouring agents, or they can be injected parenterally, for example intracavernosally, intravenously, intramuscularly or subcutaneously. For parenteral administration, the compositions may be best used in the form of a sterile aqueous solution which may contain other substances, for example enough salts or monosaccharides to make the solution isotonic with blood. For buccal or sublingual administration the compositions may be administered in the form of tablets or lozenges which can be formulated in a conventional manner.

The delivery of one or more therapeutic genes by a vector system according to the invention may be used alone or in combination with other treatments or components of the treatment. Diseases which may be treated include, but are not limited to: cancer, neurological diseases, inherited diseases, heart disease, stroke, arthritis, viral infections and diseases of the immune system. Suitable therapeutic genes include those coding for tumour suppressor proteins, enzymes, pro-drug activating enzymes, immunomodulatory molecules, antibodies, engineered immunoglobulin-like molecules, fusion proteins, hormones, membrane proteins, vasoactive proteins or peptides, cytokines, chemokines, anti-viral proteins, antisense RNA and ribozymes.

MicroRNAs (miRNAs)

miRNAs are small, RNA molecules encoded in the genomes of plants and animals.

These highly conserved, ~ 21-mer RNAs regulate the expression of genes by binding to specific mRNAs (He and Harmon, 2004).

The founding members of the miRNA family, lin-4 and let-7, were identified through genetic screens for defects in the temporal regulation of Caenorhabditis elegans larval development. - Owing to genome- wide cloning efforts, hundreds of miRNAs have now been identified in almost all metazoans, including flies, plants and mammals.

MiRNAs exhibit temporally and spatially regulated expression patterns during diverse developmental and physiological processes.

The majority of the animal miRNAs that have been characterized so far affect protein synthesis from their target mRNAs. On the other hand, most of the plant miRNAs studied so far direct the cleavage of their targets.

The degree of complementarity between a miRNA and its target, at least in part, determines the regulatory mechanism.

In animals, primary transcripts of miRNAs are processed sequentially by two RNase-III enzymes, Drosha and Dicer, into a small, imperfect dsRNA duplex (miRNArmiRNA) that contains both the mature miRNA strand and its complementary strand (miRNA<*>). Relative instability at the 5' end of the mature miRNA leads to the asymmetric assembly of the mature miRNA into the effector complex, the RNA-induced silencing complex (RISC). - Ago proteins are a key component of the RISC. Multiple Ago homologues in various metazoan genomes indicate the existence of multiple RISCs that carry out related but specific biological functions.

Bioinformatic prediction of miRNA targets has provided an important tool to explore the functions of miRNAs.

Several hundred miRNAs have been cloned and sequenced from mouse, human, Drosphila, C, elegans and Arabidopsis. Examples of such sequences may be found on www.sanger.ac.uk (Griffiths-Jones et al., 2006). Further miRNA target sequences may be searched at www.miRNA. or g.

Like niRNAs, miRNA expression profiles appear to vary from tissue to tissue but a

similar for identical tissues in different individuals(Baskerville and Bartel, 2005). Determining an miRNA with the desired expression profile may be achieved using techniques known to those skilled in the art. Once, the miRNA has been identified the corresponding target sequence can readily be determined using, for example, the databases indicated above.

For example, the mzVVana (TM) miTNA Probe Set and mJrVana (TM) miTNA Labelling Kit available from Ambion, Inc. may be used to compare the miRNA expression profiles in human tissues according to the manufacturer's instructions.

Another common way of identifying tissue-specifc rm<'>RNAs is using Northern Blot. An example of such a technique is described in Lagos-Quintana M et al, Current Biol (2002) 12:735-739 in which they identify 34 novel miRNAs by tissue-specific cloning of approximately 21-nucleotide RNAs from mouse(Lagos-Quintana et al., 2002).

Similarly, Michael M et al, Mol Can Res (2003) 1 :882-891 describes the identification of 28 different miRNA sequences in colonic adenocarcinomas and normal mucosa.

Chen C-Z et al, Science (2004) 303:83-86 describes three miRNAs, miR-181, miR-142 and miR-223 which are specifically expressed in hematopoietic cells(Chen et al., 2004).

Sempere L et al, Genome Biology (2004) 5:R13 discloses a total of 17 miRNAs detected exclusively in a particular mouse organ; these included: seven brain-specific miRNAs (miR-9, -124a, -124b, -135, -153, -183, -219), six lung-specific miRNAs (miR-18, -19a, -24, -32, -130, -213), two spleen-specific miRNAs (miR-189, -212), one liver-specific miRNA (miR-122a), and one heart-specific miRNA (miR-208). All of the indicated mouse brain-, liver- and heart-specific miRNAs were also detected in the human counterpart organs (miRNA expression was not examined in human kidney, lung or spleen), with the exception of miR-183 in the human brain. Among the 75 miRNAs that were detected in two or more mouse organs, the levels of 14 of these were detected in a particular mouse organ at levels at least two-fold higher than in any other organ; these included: seven brain-enriched miRNAs (miR-9*, -125a, -125b, - 128, -132, -137, -139), three skeletal muscle-enriched miRNAs (miR-ld, -133, -206), two kidney-enriched miRNAs (miR-30b, -30c), and one spleen-enriched miRNA (miR- 99a). All brain-enriched and skeletal muscle-enriched miRNAs had similar elevated levels in the human counterpart organs. The high

conservation of expression of these organ-specific and organ-enriched miRNAs between mouse and human suggests that they may play a conserved role in the establishment and/or maintenance of a cell or tissue type of that particular organ(Sempere et al., 2004).

Baskerville & Bartel, RNA (2005) 11:241-247 discloses a microarray profiling survey and the expression patterns of 175 human miRNAs across 24 different human organs. The results show that proximal pairs of miRNAs are generally coexpressed (Baskerville and Bartel, 2005). In addition, an abrupt transition in the correlation between pairs of expressed miRNAs occurs at a distance of 50 kb, implying that miRNAs separated by <50 kb typically derive from a common transcript. Some miRNAs are within the introns of host genes. Intronic miRNAs are usually coordinately expressed with their host gene mRNA, implying that they also generally derive from a common transcript, and that in situ analyses of host gene expression can be used to probe the spatial and temporal localization of intronic miRNAs.

Barad et al, Genome Research (2004) 14:2486-2494 establishes a miRNA-specific oligonucleotide microarray system that enables efficient analysis of the expression of the human miRNAs identified so far. It shows that the 60-mer oligonucleotide probes on the microarrays hybridize with labeled cRNA of miRNAs, but not with their precursor hairpin RNAs, derived from amplified, size-fractionated, total RNA of human origin. Signal intensity is related to the location of the miRNA sequences within the 60- mer probes, with location at the 5' region giving the highest signals, and at the 3' end, giving the lowest signals. Accordingly, 60-mer probes harboring one miRNA copy at the 5' end gave signals of similar intensity to probes containing two or three miRNA copies. Mismatch analysis shows that mutations within the miRNA sequence significantly reduce or eliminate the signal, suggesting that the observed signals faithfully reflect the abundance of matching miRNAs in the labeled cRNA. Expression profiling of 150 miRNAs in five human tissues and in HeLa cells revealed a good overall concordance with previously published results, but also with some differences. They present data on miRNA expression in thymus, testes, and placenta, and have identified miRNAs highly enriched in these tissues. Taken together, these results highlight the increased sensitivity of the DNA microarray over other methods for the detection and study of miRNAs, and the immense potential in applying such microarrays for the study of miRNAs in health and disease(Barad et al., 2004).

Kasashima K et al, Biochem Biophys Res Commun (2004) 322(2):403-10 describes the identification of three novel and 38 known miRNAs expressed in human leukemia cells (HL-60)(Kasashima et al., 2004).

Mansfield J et al, Nature Genetics (2004) 36:1079-1083 discloses the tissue-specific expression of several miRNAs during embryogenesis, including miR-10a and miR-196a(Mansfield et al., 2004).

Chen C-Z and Lodish H, Seminars in Immunology (2005) 17(2):155-165 discloses miR-181, a miRNA specifically expressed in B cells within mouse bone marrow(Chen and Lodish, 2005). It also discloses that some human miRNAs are linked to leukemias; the miR-15a/miR-16 locus is frequently deleted or down-regulated in patients with B cell chronic lymphocytic leukemia and miR-142 is at a translocation site found in a case of aggressive B cell leukemia. It is stated that these results indicate that miRNAs may be important regulators of mammalian hematopoiesis.

Methods of identifying new miRNAs and their target sequences using a computation approach are disclosed in WO2004/066183 and Brennecke J et al, PLoS Biology (2005) 3(3):0404-0418 (Brennecke et al., 2005).

Description of the Figures

Figure 1. Schematic maps depicting the model vectors used for the identification of aberrant splicing events

SIN.LV.PGK.GFPwPRE: representative LV with self inactivating Long Terminal Repeats (indicated 5' and 3' SIN LTR). The internal human phosphoglycerate kinase (PGK) promoter drives the expression of the GFP marker transgene. In other LV constructs with similar design the GFP was under the control of different promoters (K14 or CMV) or the IRES sequence. **LV.SF.LTR:** the strong enhancer/promoter form the Spleen Focus Forming Virus (SF) is placed within the LV LTR and drives the expression of GFP. **GLOBE:** SIN LV with the expression cassette cloned in opposite orientation with respect the vector genome. B-glob: β-globin transgene; SH2-3: β-globin promoter and hypersensitive sites 2 and 3 from the LCR. SD: Splice Donor; SA: Splice Acceptor; wPRE: RNA stabilization element.

Figure 2. Experimental strategy for the detection of splicing sites in LV backbone.

A) cLAM PCR strategy for genome-wide identification of aberrantly LV spiced trascripts:

Cellular PolyA+ mRNA is retrotranscribed into double-stranded cDNA using an oligo-dT primer for the first strand synthesis. A linear PCR is performed using a biotinylated primer located upstream a known LV splice site, allowing to extend into the unknown cellular portion of a chimeric transcript. The single stranded product is purified by streptavidin-coupled magnetic beads and subsequently made double stranded using Klenow enzyme and cut using an appropriate restriction enzyme (RE). A linker cassette compatible with the RE cut is ligated and two sequential nested PCRs are performed with primers located either upstream the LV splice site and on the linker cassette to enrich for chimeric transcripts. The final PCR products are then sequenced either by shotgun cloning and Sanger sequencing or by 454 Pyrosequencing.

B) Identification of aberrant splicing products on isolated cell clones

PolyA+ mRNA isolated form cell clones was subjected to different procedures (indicated) to identify aberrant splicing products. GSP: Gene Specific Primer; Primer E: exonic primer designed after identification of the LV genomic integration site in cell clones.

Figure 3. Representation of splice donor and acceptor sites identified within the LV backbone.

A) Schematic representation of splice sites identified by cLAM PCR in CD34+HSCPs and JY cells transduced with SIN.LV.PGK.GFP.wPRE. Chimeras involving interactions between LV splice donors and downstream cellular splice acceptors are indicated above and marked "3' events". Chimeras involving interactions between LV splice acceptors and upstream cellular splice donors are indicated below and marked "5' events". SD1: canonical LV splice donor; SA2: canonical LV splice acceptor; SA1: cryptic LV splice acceptor; SD4: cryptic LV splice donor; SA7: cryptic LV splice acceptor; cryptic splice sites are between parentheses. cLAM-PCR primer sets for each splice site are indicated: UPLVSD (SD1), DWLVSA (SA2), UPcrypSD (SD4), DWcrypSA (SA1). LV.exon_1 and LV.exon_2, as defined by their boundary splice sites, are indicated.

B) Transfer plasmid map of a SIN LV backbone (plasmid #277 in this example) showing all splice sites identified by the different experimental approaches and by bioinformatic predictions. Cryptic SA were found in isolated cell clones as explained in Figure 2B.

C) Splicing sites of the GLOBE LV are indicated The HS3 region was exhaustively analyzed and contain several cryptic splice sites. Annotated Gene Bank files with the of the plasmids depicted in B and C are available ad single files and embedded below.

Figure 4. qPCR for different vector portions performed on RNA extracted from JY cells transduced with LV.SIN.PGK or LV.SF.LTR at two different Multiplicity of Infection (MOI).

a. schematic representation of the position of the four TaqMan primer sets on each of the two vectors. U3RU5 recognizes the portion from the LV LTR to the SD1, encompassing the cryptic splice acceptor SA1. LV.FUSION recognizes the internally spliced transcript (SD1 to SA2). SA-PPT recognizes the sequence downstream the canonical splice acceptor SA2, encompassing the cryptic donor SD4. GFP recognizes the GFP transgene sequence.

b. Dct values obtained using β2 microglobulin as normalizer. Mean values of three biological replicates and SD are indicated. Experiments have been performed at two different MOIs to confirm that relative transcript abundance is not influenced by integrated vector load. All differences are statistically significant unless otherwise indicated (n.s.).

Figure 5. Scheme to eliminate putative unwanted fusion transcripts by adding tags recognized by specific microRNAs that will trigger their selective degradation.

On Top is represented a schematic representation of the LV backbone with the SIN LTRs depicted as boxes and the 4 repetitions of micro RNA tags complementary to the microRNAs 126 and 142.3p. In the middle, schematic representation of the LV integrated in within a coding gene. The transcription of the cellular gene (orientation indicated by the arrow) drives the expression of an mRNA that is aberrantly spliced (splicing events indicated by dashed lines) produced by the fusion between gene exons (boxes) and LV exons (lines). The chimerical transcript (depicted at the bottom) contains the sequences complementary to cellular microRNAs 126 and 143.3p (boxes). Recognition by the perfect complementary microRNAs will promote the degradation of the mRNAs containing LV sequences.

Figure 6. cLAM-PCR procedure for the retrieval of LV cellular fusion transcripts.

A) Scheme of the experimental procedure for cLAM-PCR. Total mRNA is retrotranscribed into double-stranded cDNA using oligo-dT primers. Linear PCR uses biotinylated primer located upstream/downstream a known LV splice site, allowing extension into vector or unknown cellular portion of a chimeric transcript. Single stranded product is purified by streptavidin-coupled magnetic beads, double-stranded using Klenow enzyme and cut using restriction enzymes (RE). A linker cassette compatible with the RE cut is ligated and two sequential nested PCRs are performed. The final PCR products are then sequenced. B) FACS plots showing % of GFP+ in JY cells and CD34+ HSPCs after SIN.LV.PGK transduction. The vector copy number and the MOI are indicated. C) Representative band pattern of cLAM-PCR performed on mRNA from SIN.LV.PGK-transduced cells. Retrotranscribed mRNA (RT+) and negative controls (RT-) were used. By sequencing, bands in RT+ samples corresponded to aberrant transcripts or unspliced internal control sequences. Rare faint bands in RT- controls corresponded to oligonucleotide dimers or concatamers. M: marker; H2O, from the linear amplification reaction to the second exponential PCR. D) Cryptic splice sites identified by cLAM in the LV backbone are in parentheses: SA1; SD4, SA7, SA3,SA4 and SD5. cLAM-PCR primer sets: UPLVSD and DWLVSA, UPcrypSD and DWcrypSA are indicated. LV.exon_1, LV.exon_1a, LV.exon_1b, LV.exon_2 and LV.exon_3 as defined by their boundary splice sites, are indicated.

Figure 7. Examples of chimeric LV/cellular gene/genome transcripts. Chimeric sequences are aligned on the human genome sequence using BLAT and shown on the UCSC genome browser. Sequences aligned to exonic sequences (black boxes) of know transcripts (chromosomal coordinates and size interval are shown above each panel). Orientation of vector and genes with respect to genome are indicated by orientation of triangles and arrows respectively. Vector position and size are arbitrary. The ten bases surrounding the vector/genomic junction are indicated: black text on white background indicates vector sequence, white text on black background indicates genomic sequence. In the 3 upper panels, LV integrations in the same gene transcriptional orientation involved: the canonical vector splice donor site SD1 sequence fused downstream the SA site of a gene exon (i.e. *RPL22*, first panel above); the vector splice acceptor sequence SA1 fuses to a cellular exons upstream (i.e. *BLNK*, second panel); in some cases junctions with a splice site in an unannotated exon within gene introns were found (i.e. *USP49*, third

panel); In some cases fusion transcripts aligned discontinuously to genomic portions without annotated transcripts were identified (bottom panel).

Figure 8. Representation of aberrant splicing events within the LV backbone and quantification of transcription levels of LV backbone portions.
A) Schematic representation of the position of the recoded splice sites within the LV backbone. The different mutations were distributed in 3 different vector constructs (indicated as MutSD, Mut 1_13 and Mut 14_15). B) Titers of the 3 different recoded vectors. The titer is defined as number of transducing units per milliliter (TU/ml) of vector preparation. C) Representation of the position of the 4 TaqMan primer sets on SIN.LV.PGK and LV.SF.LTR vector. U3RU5: recognizes the portion from the LV.LTR to the SD1, encompassing the cryptic splice acceptor SA1; LV.FUSION: recognizes the internally spliced transcript (SD1 to SA2); SAPPT: recognizes the sequence downstream the canonical splice acceptor SA2, encompassing the cryptic donor SD4; GFP recognizes the GFP transgene sequence. D-G) RT-qPCR results on transcription levels of different LV backbone portions performed on JY cells transduced with SIN.LV.PGK or LV.SF.LTR at MOI 0.1 (white bars), or MOI 10, (black bars). ΔCt values obtained using B2 microglobulin (Vs B2M) as normalizer measure the relative expression levels with respect to a housekeeping cellular gene. ΔCt values obtained using GFP as normalizer to measure the relative expression levels with respect to transgene expression. H) ΔCt values obtained using GFP as normalizer from JY cells transduced with SIN.LV.PGK and the recoded vectors are indicated. Probe set used are indicated. Statistical evaluation was performed by One Way Anova with Bonferroni's correction (*** $p<0.001$; **** $p<0.0001$).

Figure 9. Analysis of GLOBE harbouring reverse-orientated proviruses in which the B-globin transgene is in the same transcriptional orientation as the target gene.

EXAMPLES

Example 1 –Initial Experimental Findings

Materials and Methods

Lentiviral Vector production

All VSV-pseudotyped LV concentrated stocks were produced and titered as described in Follenzi A, Naldini L. HIV-based vectors. Preparation and use. Methods Mol Med. 2002;69:259-274.

Isolation and transduction of human HSPC

Human HSC were obtained by positive selection of CD34-expressing cells (CD34 progenitor cell isolation kit, MACS; Miltenyi Biotec, Bergisch Gladbach, Germany) from bone marrow (BM) aspirates or from mobilized peripheral blood (MPB) from healthy donors upon informed consent collection (in the context of the TIGET01 protocol that was approved by San Raffaele Scientific Institute Ethical Committee). Alternatively, purified CD34+ cells from healthy donors' BM were provided by Lonza (Human Bone Marrow CD34+ Progenitors 2M-101, Lonza). Soon after purification or thawing, cells were placed in culture on retronectin-coated wells (T100A Takara) in CellGro SCGM medium (2001 CellGenix) at a concentration of 1-1.5x10^6 cells/ml in the presence of cytokines (interleukin-3 [IL-3, 60 ng/μl], thrombopoietin [TPO, 100 ng/μl], stem cell factor [SCF, 300 ng/μl], and Flt3 ligand [Flt3-L, 300 ng/μl]; PeproTech, Rocky Hill, NJ) for 24-48 hours of pre-stimulation. Cells were then transduced with the differnt LVs at a multiplicity of infection [MOI] as indicated for 12 hours. Cells were plated in Iscove's modified Dulbecco's medium (IMDM) –10% fetal bovine serum (FBS) with cytokines (IL-3, 60 ng/μl; IL-6, 60 ng/μl; SCF, 300 ng/μl) and cultured for a total of 14 days. Thereafter, cells were collected for molecular, biochemical and flow cytometry studies.

Flow cytometry

Before pre-stimulation and at the end of transduction, 5x10^4 cells were stained with one μl of PE-conjugated anti-CD34 and FITC-conjugated anti-CD45 antibodies or IgG isotype controls (Dako, Glostrup, Denmark). After 20 min on ice, cells were washed, re-suspended in PBS with 2% FBS and 1% paraformaldehyde (PFA) and analyzed by flow cytometry (FACSCalibur; BD Biosciences Immunocytometry Systems, SanJose, CA). The percentage of CD34+ cells was calculated on the gated CD45+ population.

At the end of the 14-day culture period, 1x10^5 GFP-transduced cells were collected and GFP fluorescence (measured as percentage of positive cells and mean fluorescence intensity – MFI) was detected with detector channel FL1 calibrated to the fluorescein isothiocyanate (FITC) emission profile.

Quantitative PCR

Genomic DNA was extracted from CD34+ liquid culture samples with QIAamp DNA Blood Mini Kit-Quiagen, and from murine tissues with the Blood & Cells DNA Midi Kit-Quiagen after o/n digestion with proteinase K (PK)(Roche). Single Colonies were digested 4 h 37°C in Monini's Buffer (500ul Lauryl Ether 10%, 500ul Tris Hcl 1M, 100ug/ml PK). After PK inactivation (10' 95°C), lysates were centrifuged at 8000 rpm 10'. 10 µl of supernatants were used for quantitative PCR. LV sequences were detected by quantitative PCR on 50 ng of total genomic DNA.

cLAM-PCR and genomic integration site analysis

All procedures for cLAM are similar to those described in: Montini E, Cesana D, Schmidt M, et al. The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J Clin Invest. 2009;119:964-975, and Schmidt M, Schwarzwaelder K, Bartholomae C, et al. High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat Methods. 2007;4:1051-1057.

Briefly, 500 ng of poly A+ mRNA was converted into cDNA using the Invitrogen SuperScript double stranded cDNAsynthesis kit following manufacturer instructions. Initally, 50-cycle linear PCR was performed, double trand systhesis and restriction digest using Tsp509I or HpyCH4IV and ligation of a restriction site–complementary linker cassette. The first exponential biotinylated PCR product was captured via magnetic beads and reamplified by a nested PCR. LAM-PCR products were separated by Spreadex gel electrophoresis (Elchrom Scientific) to verify the presence and number of bands. Each LAM PCR was shotgun cloned into the TOPO TA vector (Invitrogen) and sequenced by Sanger sequencing (GATC Biotech). Sequences were validated and classified using specific PERL scripts and aligned to the human genome (freeze March 2006, UCSC).

Results

Frequency of aberrant splicing events was determined by 5'RACE and RT-PCR. cDNA was prepared with a primer annealing to an internal viral region (GSP5). PCR primers were designed upstream the LV integration (E FOR) and downstream HIV splicing signals inside the vector (GSP7). Aberrant splicing events were detected for more than 60% of the integration sites in all cell types.

Oncogenesis induced by insertional mutagenesis with gene therapy vectors occurs mainly by activation of proto-oncogenes found at or nearby the insertion site. This activation often occurs by an enhancer-mediated mechanism or by a process of splicing capture which generates chimeric transcripts comprising portions of vector and cellular mRNAs. Although the activation of oncogenes may be reduced by the use of self-inactivating (SIN) design and moderate cellular promoters, how to reduce genotoxic splicing capture events and aberrant transcript formation triggered by vector integration is still unclear. We developed a modified Linear Amplification-Mediated (LAM) PCR technique, named cDNA LAM PCR (cLAM-PCR), aimed at retrieving, from the whole transcriptome of LV-transduced cells aberrantly spliced mRNAs that contain lentiviral vector (LV) sequences fused with cellular transcripts in a high-throughput fashion. The sequences of cLAM-PCR products were obtained by 454 pyrosequencing and analyzed by a purposely build high-throughput computational pipeline. Our pipeline is based on a map-reduce parallelization model, running in a private computer cluster and use a dynamic analysis process composed by different steps implemented as map-reduce applications. Thus, chimeric LV-genome sequences are recognized, the nucleotide position of the fused sequence is identified (the splice site), and the remaining portion mapped on the appropriate genome assembly by BLAST. Results obtained with different LV constructs show that integrated LVs can perturb the processing of cellular transcripts by interacting with the cellular splicing machinery and fusing with its own splice sites to cellular splice sites both upstream and downstream the integration site. So far, 70 different fusion transcripts could be identified in total, 84% of which were fused to known splice sites of gene exons, 6% were fused to uncharacterized cryptic splice sites located in introns and the remaining 10% were fused to genomic sequences not corresponding to any annotated gene. We identified several established and previously unknown splice sites within the LV backbone that participate in the aberrant splicing process with variable efficiency. Quantitative PCR on different portions of the LV backbone allows measuring the relative contribution to the aberrant splicing process of each LV splice site identified. The amount of transcription occurring in regions outside the expression cassette reaches up to the 3% of the entire transgene expression. The cLAM-PCR technique, coupled to high-throughput sequencing and the computational power of our specialized data analysis pipeline allows gaining insights into the biology of vector-mediated splicing alteration.

Splice sites within the lentiviral vector backbone were identified from different human cell sources such as: Human primary cord blood derived CD34+ hematopoietic stem

progenitor cells, the lymphoblastoid B-cell line JY, primary T cells, myeloid cells and keratinocytes.

Cells were transduced with a battery of lentiviral vectors such as: SIN.LV.PGK.GFP.wPRE, and SIN LV SIN-LVs with identical design but for the internl promoter used upstream the GFP transgene LV.SF.LTR.GFP.wPRE vector (Figure 1), cells were transduced also with SIN-LVs carrying internal GFP expression cassettes or a full human β-globin gene driven by a β-globin promoter and a reduced-size LCR (specific for the therapy of β-thalassemia).

By the use of an in-house developed PCR technique named **c**omplementary **L**inear **A**mplification **M**ediated **(cLAM)-PCR,** we have been able to retrieve fusion transcripts between LV sequences and cellular mRNAs in human primary cord blood derived CD34+ hematopoietic stem progenitor cells and JY cells (Figure 2A). Moreover SIN LV-marked cells were randomly cloned and integration sites mapped by LM-PCR in individual clones. Chimeric transcripts were identified by RACE PCR and exon-specific/LV RT-PCR. Abnormal, chimeric transcripts were identified in >50% of the LV target genes in all cell types (Figure 2B).

With these approaches we demonstrated that known and previously uncharacterized LV splice sites interact with the cellular splicing machinery triggering the formation of chimerical fusion mRNAs with cellular genes surrounding the integration site.

Several genes targeted by the aberrant splicing events are oncogenes or tumor suppressors (for example, PTEN in two different T-cell clones). Splicing events occurring in these genes may represent a potential risk of post-transcriptional genotoxicity of LVs.

Coupling our experimental data and bioinformatic splice site prediction analysis (NetGene2 server splice site prediction software), we identified a set of splicing consensus sequences in the LV backbone (splice donors and splice acceptors, both in forward and in reverse orientation with respect the LV genome) that are most prone to the induction of chimerical aberrant transcripts (Figure 3).

Furthermore, we set up qPCR assays to quantitatively assess the strength of different splice sites and to identify such sites as major o minor contributors to the

aberrant splicing process. With this approach we found that at the genome-wide level LV sequences outside the expression cassette account for up to 5% of the overall transcription produced by the internal expression cassette, which is remarkably high. Moreover the transcription of LV backbone s dependent on the presence/position of the enhancer promoter sequences within the vector. Indeed, the SF sequence within the LTR of the LV.SF.LTR stabilizes the usage of 3 spliced (exons) within the LV backbone. On the other hand the SIN.LV.PGK.GFP.wPRE shows a decreasing transcription (splicing) rate from 5' to 3' along the vector backbone (Figure 4).

RT-PCR analyses showed that aberrant splicing events generating chimeric transcripts occur at >60% of the integration sites in both keratinocytes and T cells. Semi-quantitative RT-PCR revealed that fusion transcripts were mostly represented at low level compared to constitutively spliced, wild-type transcripts. The incidence of aberrant splicing is similarly high for vectors lacking an internal promoter ("read-through trap" strategy) and LV harboring CMV or K14 promoters.

Fusion transcripts were generated through aberrant splicing caused by the usage of both constitutive and cryptic splice sites located in the viral intron and the U5 portion of the 5' LTR and in the β-globin transcriptional cassette. A high relative abundance of aberrant transcripts compared to WT mRNA was detected by semiquantitative PCR. Preliminary data indicate that the aberrant transcripts terminate at the cellular polyA signal.

Elimination of these splice from the LV backbone is instrumental in reducing its aberrant splicing potential. We are now generating new LV constructs to test the production and transduction efficiency of vectors in which the splice sites have been recoded and eliminated. In particular, two recoded splice sites are being evaluated individually in two different constructs, since they lie in regions with a highly conserved secondary structure, whereas other 15 recoded splice sites located in different LV regions are being tested on a third construct. Finally, an extensively recoded LV carrying 17 single nucleotide mutations can be generated and validated with respect to production and transduction efficiency, and with respect to aberrant splicing potential using cLAM-PCR and qPCR.

position (plasm#277)	phase	strand	splice site score (0 to 1)	original sequence	name	new recoded sequence
3178	2	1	0.83	GCGGCGACTG^GTGAGTACGC	SD1	GCGGCGACTG^ATGAGTACGC
3128	1	1	0.33	CTCGACGCAG^GACTCGGCTT	SA1	CTCGACGCAA^GACTCGGCTT
3432	2	1	0.16	ATCCCTTCAG^ACAGGATCAG	SA9	ATCCCTTCAA^ACAGGATCAG
3557	1	1	0.51	CAAAACAAAA^GTAAGACCAC	SD2	CAAAACAAAA^ATAAGACCAC
3597	1	2	0.43	CCTCCTCCAG^GTCTGAAGAT	SA7	CCTCCTCCAA^GTCTGAAGAT
3920	0	1	0.30	GCAGCTCCAG^GCAAGAATCC	SD3	GCAGCTCCAG^ACAAGAATCC
3929	0	2	0.14	CCACAGCCAG^GATTCTTGCC	SA21	CCACAGCCAA^GATTCTTGCC
3933	2	2	0.15	CTTTCCACAG^CCAGGATTCT	SA20	CTTTCCACAA^CCAGGATTCT
3947	0	2	0.33	GATCCTTTAG^GTATCTTTCC	SA6	GATCCTTTAA^GTATCTTTCC
4069	2	2	0.39	CTCCATCCAG^GTCGTGTGAT	SA5	CTCCATCCAA^GTCGTGTGAT
4342	2	1	0.28	ATCGTTTCAG^ACCCACCTCC	SA2	ATCGTTTCAA^ACCCACCTCC
4348	0	2	0.49	GGGTTGGGAG^GTGGGTCTGA	SD6	GGGTTGGGAG^ATGGGTCTGA
4362	1	1	0.19	CAACCCCGAG^GGGACCCGAC	SA10	CAACCCCGAA^GGGACCCGAC
4374	1	1	0.18	GACCCGACAG^GCCCGAAGGA	SA11	GACCCGACAA^GCCCGAAGGA
4450	2	1	0.55	GATCTCGACG^GTATCGGTTA	SD4	GATCTCGACG^ATATCGGTTA
6501	1	2	0.41	GGTCTTAAAG^GTACCGAGCT	SD2	GGTCTTAAAG^ATACCGAGCT
6521	0	2	0.34	CAGCTGCCTT^GTAAGTCATT	SD1	CAGCTGCCTT^ATAAGTCATT

Table 1: list of recoded splice sites.

Position, nucleotide number according to vectorNTI map #277 (Lab Naldini); Strand, 1=forward 2=reverse; Score, as predicted by NetGene2 server splice site prediction software; Name, SD# indicates splice donors and SA# indicates splice acceptors.

Example 2 – Up-dated experimental findings relating to identification of the following splice sites:

SPLICE ACCEPTOR GROUP 1

SA1 - corresponding to nucleotides 3127-3128 of SEQ ID NO:1 or nucleotides 3130-3131 of SEQ ID NO:3

SA2 – corresponding to nucleotides 4341-4342 of SEQ ID NO:1 or nucleotides 4344-4345 of SEQ ID NO:3.

SA3 – corresponding to nucleotides 3071-3072 of SEQ ID NO:1 or nucleotides 3071-3072 of SEQ ID NO:3.

SA4 – corresponding to nucleotides 3068-3069 of SEQ ID NO:1 or nucleotides 3068-3069 of SEQ ID NO:3.

SA5 - corresponding to nucleotides 4069-4070 of SEQ ID NO:1 or nucleotides 4072-4073 of SEQ ID NO:3.

SA6 - corresponding to nucleotides 3947-3948 of SEQ ID NO:1 or nucleotides 3950-3951 of SEQ ID NO:3.

SA7 - corresponding to nucleotides 3597-3598 (complement) of SEQ ID NO:1 or nucleotides 3600-3601 (complement) of SEQ ID NO:3.

SA9 - corresponding to nucleotides 3431-3432 of SEQ ID NO:1 or nucleotides 3434-3435 of SEQ ID NO:3.

SA10 - corresponding to nucleotides 4361-4362 of SEQ ID NO:1 or nucleotides 4364-4365 of SEQ ID NO:3.

SA11 - corresponding to nucleotides 4373-4374 of SEQ ID NO:1 or nucleotides 4376-4377 of SEQ ID NO:3.

SA20 - corresponding to nucleotides 3933-3934 (complement) of SEQ ID NO:1 or nucleotides 3936-3937 (complement) of SEQ ID NO:3.

SA21 - corresponding to nucleotides 3929-3930 (complement) of SEQ ID NO:1 or nucleotides 3932-3933 (complement) of SEQ ID NO:3.

SPLICE DONOR GROUP 1

SD1 - corresponding to nucleotides 3178-3179 of SEQ ID NO:1 or nucleotides 3181-3182 of SEQ ID NO:3.

SD2 - corresponding to nucleotides 3557-3558 of SEQ ID NO:1 or nucleotides 3560-3561 of SEQ ID NO:3.

SD3 - corresponding to nucleotides 3920-3921 of SEQ ID NO:1 or nucleotides 3923-3924 of SEQ ID NO:3.

SD4 - corresponding to nucleotides 4450-4451 of SEQ ID NO:1 or nucleotides 4453-4454 of SEQ ID NO:3.

SD5- corresponding to nucleotides 2974-2975 (complement) of SEQ ID NO:1 or nucleotides 2974-2975 (complement) of SEQ ID NO:3.

SD6- corresponding to nucleotide 4347-4348 (complement) of SEQ ID NO:1 or nucleotides 4350-4351 (complement) of SEQ ID NO:3.

SD14- corresponding to nucleotides 6500-6501(complement) of SEQ ID NO:1 or nucleotides 6503-6504 (complement) of SEQ ID NO:3.

SD15- corresponding to nucleotides 6520-6521(complement) of SEQ ID NO:1 or nucleotides 6523-6524 (complement) of SEQ ID NO:3.

Material and Methods

Vector production and titration

LV.SF.LTR and SINLV.PGK constructs were previously generated (Gabriel et al., 2009, *Nat Med* 15:1431-1436; Follenzi et al. 2000, *Nat Genet* 25:217-222). Concentrated lentiviral vector stocks,pseudotyped with the Vesicular Stomatitis Virus envelope were produced by transient 4 plasmid cotransfection of 293T cells and titered on HeLa cells as described (Follenzi et al. 2000;,*Nat Genet* 25:217-222). Recoded vectors were generated by DNA gene synthesis (GeneArt) and cloned in the SINLV.PGK transfer plasmid. A 728 bp DNA fragment harboring the mutation in the canonical HIV1 splice donor was cloned using NruI and NdeI restriction sites. A 1331 bp DNA fragment harboring the mutations from 1 to 13 was cloned using NruI and XhoI restriction sites. A 751 bp DNA fragment harboring the mutations from 14 and 15 was cloned using SacII and AvrII restriction sites.

Isolation and transduction of human HSPC and JY cells

Cord blood-derived cells were harvested, cultured and transduced as previously described. Human HSPC were obtained by positive selection of CD34-expressing cells (CD34 progenitor cell isolation kit, MACS; Miltenyi Biotec) from cord blood from healthy donors. Soon after purification or thawing, cells were placed in culture at a concentration of 1-1.5x106 cells/ml in the presence of cytokines (interleukin-3 [IL-3, 60 ng/µl], thrombopoietin [TPO, 100 ng/µl], stem cell factor [SCF, 300 ng/µl], and Flt3 ligand [Flt3-L, 300 ng/µl]; PeproTech) for 24-48 hours of pre-stimulation. Cells were then transduced with the different LVs at a multiplicity of infection [MOI] as indicated for 12 hours. Cells were plated in Iscove's modified Dulbecco's medium (Euroclone) –10% fetal bovine serum (Euroclone) with cytokines (IL-3, 60 ng/µl; IL-6, 60 ng/µl; SCF, 300 ng/µl) and cultured for a total of 14 days. Thereafter, cells were collected for molecular, biochemical and flow cytometry studies.

JY cells were grown in RPMI, 10% FBS supplemented with penicillin and streptomycin, and transduced at MOI 10, 1 or 0.1 in a single round of infection.

Flow cytometry

Before pre-stimulation and at the end of transduction, $5x10^4$ cells were stained with one µl of PE-conjugated anti-CD34 and FITC-conjugated anti-CD45 antibodies or

IgG isotype controls (Dako). After 20 min on ice, cells were washed, re-suspended in PBS with 2% FBS and 1% paraformaldehyde (PFA) and analyzed by flow cytometry (FACSCalibur; BD Biosciences Immunocytometry Systems). The percentage of CD34+ cells was calculated on the gated CD45+ population.

At the end of the 14-day culture period, 1x105 GFP-transduced cells were collected and GFP fluorescence (measured as percentage of positive cells and mean fluorescence intensity – MFI) was detected with detector channel FL1 calibrated to the fluorescein isothiocyanate (FITC) emission profile.

DNA and RNA isolation, cDNA synthesis and quantitative PCR

Genomic DNA was extracted from CD34+ liquid culture samples with QIAamp DNA Blood Mini Kit-Quiagen, and from murine tissues with the Blood & Cells DNA Midi Kit-Quiagen after o/n digestion with proteinase K (Roche). Total RNA from JY or CD34+ cells was isolated with the RNeasy Mini Kit following manufacturer instructions (QIAGEN). Double-stranded (ds) cDNA preparation was performed using SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen). cDNA was used as template for Custom Plus TaqMan Gene Expression Assays specific to each LV portion (Applied Biosystems). Amplification reactions were performed on a 7900HT Real-Time PCR Thermal Cycler (Applied Biosystems). The relative expression level of each gene was calculated by the ΔCt method normalizing to Beta-2 microglobulin (housekeeping gene control) or GFP expression.

cLAM-PCR amplification

We used 500 ng of ds cDNA as template for cLAM-PCR. cLAM-PCR was initiated with a 100-cycle linear PCR using a biotinylated primer (primer_1), second strand synthesis by Klenow fragment and random hexamers, restriction digest using Tsp509I or HpyCH4IV and ligation of a restriction site–complementary linker cassette. The biotinylated PCR product was captured via magnetic beads and reamplified by two nested PCRs using primers downstream primer_1 (primer_2, primer_3) and primers complementary to the linker cassette. Primer sequences for the 4 primer sets are: (UPLVSD_1 GAAAGCGAAAGGGAAACCAGA, UPLVSD_2 GACGCAGGACTCGGCTTG, UPLVSD_3 ACGGCAAGAGGCGAGG; DWLVSA_1 TCGAGATCCGTTCACTAATCG, DWLVSA_2 ATGGATCTGTCTCTGTCTCTCTCT, DWLVSA_3 CCACCTTCTTCTTCTATTCCTTC; UPcrypSD_1

GAGGGGACCCGACAGG, UPcrypSD_2 CCGAAGGAATAGAAGAAGAAGG, UPcrypSD_3 CAGAGACAGATCCATTCGATTAGTG; DWcrypSA_1 CCTCGCCTCTTGCCGTGC, DWcrypSA_2 CTTCAGCAAGCCGAGTCC). Linker cassette primers were previously described.

cLAM-PCR products were separated by Spreadex gel electrophoresis (Elchrom Scientific) to verify the presence and number of bands. cLAM-PCR was shotgun cloned into the TOPO TA vector (Invitrogen) and sequenced by Sanger sequencing (GATC Biotech) or directly sequenced by 454 pyrosequencing after a PCR reamplification with the use of oligonucleotides with specific 6-nucleotide sequence tags for sample identification. Sequences were validated and classified with specific scripts and aligned to the human genome (GRCh37/hg19) or with the use of the UCSC BLAT genome browser.

Gene ontology analysis

Analysis of overrepresentation of gene classes in integration data sets was performed with the DAVID-EASE software (http://david.niaid.nih.gov/ david/ease.htm) using the stringency setting "high".

Statistics overview.

For gene ontology analyses we considered significant only those classes represented by at least 3 genes, a fold increase >3 and a p value<0.05. The results were corrected for multiple testing errors within each data set/system combination with the Bonferroni's method. For gene expression analyses One-way ANOVA test with Bonferroni's multiple comparison post-test was used to assess statistical significance of differences among all samples ($p< 0.05$). In all the graphs the mean ± standard deviations are indicated.

Study Approval

Cord blood-derived Human CD34+ was collected upon informed consent in the context of the TIGET01 protocol that was approved by San Raffaele Scientific Institute Ethical Committee.

Results

cDNA Linear Amplification Mediated (cLAM)-PCR technique to study LV-induced aberrant splicing in primary Human Stem Progenitor Cells

cLAM PCR is aimed at retrieving in a high-throughput fashion aberrantly spliced mRNAs that contain LV sequences fused with cellular transcripts from the whole transcriptome of LV-transduced cells (schematics in Fig. 6A). Similarly to the previously published LAM PCR technique (Schmidt,et al.. 2007. *Nat Methods* 4:1051-1057), a biotinylated oligonucleotide is designed on a sequence complementary to the HIV backbone and used for linear amplification on single or double stranded cDNA from LV transduced cells. The resulting single stranded DNA molecules will contain expressed portions of the HIV backbone and, in chimeric transcripts, may also contain unknown cellular sequences. The linear amplification products are then purified with streptavidin-coupled paramagnetic beads and subsequently subjected to double strand synthesis, and digested with a restriction enzyme to ligate a linker cassette. The restriction enzymes used in this study were Tsp509I (AATT) and HpyCH4IV (AGCT) as their efficacy in LAM-PCR protocols has been previously confirmed.. The resulting products are then amplified by exponential PCR using nested oligonucleotides complementary to the HIV backbone and the linker cassette. The final cLAM-PCR products were sequenced by 454 pyrosequencing and analyzed by dedicated high throughput computational pipeline. This computational pipeline has been developed to recognize and annotate chimeric LV-genome transcripts that contain LV sequences fused to host cell sequences. LV sequences are recognized and the nucleotide position at the fusion point identified (splice site) on the LV genome. The remaining sequence portion, after removal of the LV and linker cassette sequences, is mapped on the appropriate genome by BLAST. With this technique, by designing the proper oligonucleotide sets in different portions of the LV backbone, it is possible to interrogate different LV sequences for their ability to generate aberrant splicing events. The most obvious choice was to design an oligonucleotide set upstream to the canonical LV splice donor site and in forward orientation with respect to the HIV genome (oligonucleotide set named UPLVSD). The second cLAM oligonucleotide set was designed downstream the canonical splice acceptor site sequence in reverse orientation with respect to HIV transcription (oligonucleotide set named DWLVSA). These two cLAM oligonucleotide sets encompass the 1165 bp HIV intron, which, based on our previous results of

LV.SF.LTR-induced Braf activation in *Cdkn2a-/-* tumors, likely play a relevant role in the splicing capture process.

We investigated the aberrant splicing induced by an LV with self-inactivating long terminal repeats (SIN LTRs), containing the human phosphoglycerate kinase (hPGK) promoter in internal position driving the expression of the GFP (SINLV.PGK). This vector was used to transduce a human B-lymphoblastoid cell line (JY cells) and the clinically relevant human primary cord blood-derived CD34+ Hematopoietic Stem Progenitor Cells (HSPCs). JY cells were transduced at different multiplicity of infection (MOI) 0.1 or 10, obtaining an average Vector Copy Number (VCN) of 0.18 and 15 and a percentage of vector-marked cells of 18% and 100%, respectively. Human CD34+ HSPCs were transduced at MOI 100 using an established clinical protocol(23), obtaining an average VCN of 4.7 and 78% of vector-marked cells (Fig.6B).

Spreadex gel electrophoresis of cLAM PCR products obtained from transduced JY and CD34+HSPCs showed several bands of variable molecular size, ranging from 100 to 600 bp. On the other hand, non-retrotranscribed RNA controls (RT- controls) yielded rare and faint bands corresponding to primer dimers or concatemers. The complexity of band patterns correlated with the marking levels of the samples tested. Cells transduced with high vector loads produced many bands of different molecular size while samples from low MOI showed smaller number of bands (Fig.6C).

cLAM PCR products were shotgun cloned into plasmids and sequenced by the Sanger method or tagged by PCR with adapter primers designed to include a sequence bar code tag (DNA barcoding) and subjected to 454-pyrosequencing. The information contained in the DNA bar codes allows the simultaneous sequencing of pooled amplicons from different samples. By this approach we identified a total of 8 splice sites within the LV backbone that participate in the aberrant splicing process with variable efficiency. Based on these preliminary data, two additional cLAM primer sets were designed to interrogate the activity of SA1 and SD4 (DWcrypSA and UPcrypSD sets respectively) (Fig. 6D).

Sequence analysis of aberrantly spliced transcripts and LV sequences participating in the aberrant splicing process

Overall, using the 4 cLAM PCR primer sets we obtained 39430 sequencing reads from SINLV.PGK-transduced JY cells and CD34+ HSCPs. A dedicated pipeline was used to eliminate the LV sequence complementary to the oligonucleotide used and the liker cassette. The remaining sequence was mapped on the LV and the human genomes to precisely identify the sequences involved in the fusion process. The majority of the sequencing reads (n=28216, 71.5%) were too short to be univocally mapped on the LV or human genome (less than 20 nucleotides) or lack the LV sequences required to validate the PCR products as genuine LV-originated transcripts. Although the process may appear to be relatively inefficient, we were still able to validate 11214 sequence reads sequencing reads (28.5%) as genuine transcripts containing LV backbone sequences. After exclusion of sequencing reads containing only LV genome (n= 8720) and pooling all the redundant sequencing reads (n=2494), we identified 317 unique LV fusion transcripts with cellular gene exons or genomic sequences. The fusion transcripts were generated using the LV canonical splice acceptor or donor sites (17%) or the other splice sites within the LV backbone (SA1: 14.8%, SA3: 10.7%, SA4: 37.2%, SD4: 3.5%, SD5: 16.7%). Overall, the retrieved transcripts were fusions between LV splice sites and: i) known gene exons (88.6%); ii) cryptic splice sites located in known gene introns (6.6%); iii) 3' UTRs (0.6%); iv) cryptic splice sites located in intergenic regions (4.1%) (Fig. 7). The latter cases are quite peculiar as it appears that LV genomic integrations are able to tag unknown human transcripts or induce the formation of novel transcripts. All the splice acceptor sites identified within the LV backbone have the typical AG dinucleotide. Two out of 3 LV splice donors have the GT dinucleotide, while the splice donor SD5 has a GC dinucleotide. We observed that 237 fusion transcripts (75%) show the expected GT/AG junction, a frequency lower than the 98-99% reported for the genomic splice junctions. On the other hand, 53 fusion transcripts (16.7%) were generated by using the LV splice donor SD5 thus generating GC/AG sequences at the putative splice junctions. The remaining 27 fusion transcripts (7.3%) contained non-canonical splice junctions (for example GC/AG, TC/AG and AC/AG). Interestingly, the latter class of transcripts was mainly (25 out of 27) the fusion of LV sequences with introns (n=6), intergenic (n=4) or within exonic sequences (n=15). To understand if the genes subjected to aberrant splicing were enriched for specific gene classes, we performed Gene Ontology analysis using the DAVID EASE online software (http://david.abcc.ncifcrf.gov). From this analysis we found that, LV chimeric transcripts were significantly overrepresented for gene classes such as ubiquitin-protein ligase activity (p=2.6x10-3, fold change=3.8), nuclear export (p=3.3x10-3, fold change=5.9), Lymphocyte Activation (p=2.0x10-3,

fold change=3.3), Lymphocyte differentiation (p=3.0x10-2, fold change=3.4), Positive Regulation of Growth (p=3.4x10-2, fold change=5.6), RNA splicing (p=4.4x 10-3, fold change=3.5), nuclear mRNA splicing (p=4.4x10-3, fold change=3.5) and ATP catabolic process (p=2.9 x 10-2, fold change=11) (Table 2). This bias towards these gene classes overlapped only partially to the typical LV integration bias reported in hematopoietic cells. To directly test the LV integration bias in our cells, we performed LAM PCR on the genomic DNA of the same SINLVPGK transduced JY and CD34+ cell preparation used for cLAM PCR. Overall, we mapped 1630 unique LV integration sites and addressed the integration bias into specific gene classes by Gene Ontology analysis. Similarly to the reported LV integration bias reported in other hematopoietic cells we observed the marked tendency to integrate into genes which are enriched for chromatin remodeling functions. Interestingly, several gene classes significantly overrepresented in LV-mediated aberrant splicing formation such as Positive Regulation of Growth, RNA Splicing, and lymphocyte activation and differentiation are different from those found in our and other previously reported genomic integration profiles on hematopoietic cells.

Impact of spice site recoding on vector infectivity and levels of read-through transcription on LV backbone

The identification of the 8 splice sites within the LV backbone by cLAM PCR provides important information on how to recode the sequences to reduce the aberrant splicing potential events. Moreover, we used the "NetGene2 server" splice site prediction software (http://www.cbs.dtu.dk/services/NetGene2/) to identify other potential splice sites within the LV backbone. The software identified 5 experimentally validated splice sites with levels of confidence ranging from 0.28 to 0.83 and 15 additional putative splice sites (levels of confidence > 0.14) (Table 3). Since some splice sites are located in regions with a highly conserved secondary structure, 3 sets of mutations were distributed into the parental SINLV.PGK vector : a construct containing only the recoding of the canonical splice donor site SD1 (SINLV.MutSD.PGK.GFP.mwPREpre, named Mut SD); a construct containing 13 recoded splice sites in a region comprised between SD1 and the cryptic SD4 (SA9, SD2, SA7, SD3, SA21, SA20, SA6, SA5, SA2, SD6, SA10, SA11, SD4) (SINLV.Mut1_13.PGK.GFP.mwPREpre, named Mut 1_13); and a construct harboring two recoded splice sites near the 3'LTR (SD14, SD15) (SINLV.Mut14_15.PGK.GFP.mwPRE, named Mut 14_15) (Fig. 8A). Each recoded LV construct was then tested by Fluorescence Activated Cell Sorting (FACS) to

evaluate vector titer and by RT qPCR to measure transgene expression and read-through transcription within the LV backbone as surrogate of aberrant splicing potential. The constructs harboring the mutated SD1 and the 13 mutations showed a significant reduction in infectivity (10 fold reduction), while the vector with the two mutations near the 3'LTR (Mut 14 15) had a comparable infectivity to the standard SINLV.PGK (Figure 8B). Median GFP fluorescence intensity (MFI) of single copy transduced JY cells was similar among all the different vectors (MFI range 7400-7900), indicating that the recoding does not affect transgene expression in any case. We set up RT-qPCR assays on cDNA from transduced JY cells to probe the transcription levels in different portions of the different LV backbones (Fig. 8C). The oligonucleotides and probes used for the RT-qPCR were designed to amplify different portions of the LV backbone encompassing the splice sites identified in this study: the U3RU5 RT-PCR assay, encompassing the SA1; the LV.FUSION, encompassing the HIV1 intron and measuring only spliced LV mRNAs; SA.PPT, encompassing the SD4; the GFP assay, complementary to the GFP transgene sequence. JY cells were initially transduced with two different LVs: the previously mentioned SINLV.PGK, and an LV harboring the strong enhancer sequence of the Spleen Focus Forming Virus (SF) promoter within the LTR (LV.SF.LTR) and driving GFP expression.. The latter vector was used as positive control of transcription within the LV backbone as the SF promoter transcription starts upstream the regions tested for expression and the transcript is extended through the GFP transgene and terminates at the polyadenylation site in the 3' LTR. The relative levels of read through transcription within the LV backbone were normalized to the endogenous housekeeping β2-microglobulin (*β2M*) gene. The expression measured by the U3RU5 probe in SINLV.PGK transduced cells at MOI 10 or 0.1 showed a ΔCt of 4.09 and 5.67, corresponding to 6.3% and 1.9% of the housekeeping gene expression level, respectively. Interestingly, the other probes (LV-FUSION and SA-PPT) showed a decrease in expression levels from 5' to the 3' of the LV backbone (Fig. 8D). As the amounts of read through transcription vary according to the number of vector integrations in expression-permissive genome, the relative expression level of each LV portion tested was also normalized to the GFP level, which depends directly on the integrated vector load. For both MOIs, U3RU5 probe showed a ΔCt of 6.4 ±0.1 with respect to GFP, indicating that read-through transcription was >1.2% of the overall GFP transcripts produced by the internal expression cassette (Fig.8E). Also with this normalization, both the LV-FUSION and SA-PPT showed a progressive reduction of the expression levels from 5' to the 3' of the LV backbone. Using the same probe sets, we measured the transcription levels of the LV.SF.LTR backbone

in transduced JY cells. The presence of the strong SF enhancer/promoter within the LTR resulted in a much higher level of expression compared to the backbone transcription measured for the SINLV.PGK, indicating a previously underappreciated advantage of the SIN LTR design (Fig.8 F,G). We then performed our RT-qPCR analysis on JY cells transduced with the recoded constructs (MutSD, Mut1_13 and Mut 14_15). We used the U3RU5, SA-PPT probe sets and an additional probe set encompassing the canonical HIV1 splice donor (HIV-SD). GFP normalized values show that splice site mutagenesis can further and significantly reduce the residual backbone transcription in MutSD and Mut1_13 vectors when compared with the expression of the parental vector (SINLV.PGK Vs. Mut SD with HIVSD, p=0.001 and SINLV.PGK Vs Mut1_13 with SA-PPT, p<0.0001) (Fig. 8H)

SYSTEM	GENE CLASSES	COUNT	p VALUE	FOLD CHANGE
MF	histone acetyl-lysine binding	3	1.50E-03	46
BP	lymphocyte activation	11	2.00E-03	3.3
MF	ubiquitin-protein ligase activity	9	2.60E-03	3.8
BP	nuclear export	6	3.30E-03	5.9
BP	RNA splicing, via transesterification reactions	9	4.40E-03	3.5
BP	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	9	4.40E-03	3.5
BP	nuclear mRNA splicing, via Spliceosome	9	4.40E-03	3.5
MF	small conjugating protein ligase activity	9	5.40E-03	3.4
BP	telomere organization	4	1.20E-02	8.1
MF	non-membrane spanning protein tyrosine kinase activity	4	2.80E-02	6
BP	ATP catabolic process	3	2.90E-02	11
BP	lymphocyte differentiation	6	3.00E-02	3.4
BP	positive regulation of cell Growth	4	3.40E-02	5.6
BP	ribonucleoside triphosphate catabolic process	3	4.00E-02	9.3
BP	purine ribonucleoside triphosphate catabolic process	3	4.00E-02	9.3
BP	positive regulation of cell size	4	4.70E-02	4.9
BP	endoplasmic reticulum unfolded protein response	3	4.80E-02	8.4

BP	cellular response to unfolded protein	3	4.80E-02	8.4
BP	purine nucleoside triphosphate catabolic process	3	4.80E-02	8.4

Table 2: Overrepresentation analysis of the gene types involved in the generation of LV/cellular gene fusion transcripts.

Cellular genes involved in aberrant splicing formation with LV sequences were clustered in large classes with similar biological process and functions (Column System; MF, Molecular Function; BP, Biological Process). Gene classes are indicated. Count: number of genes identified in the dataset belonging to each specific class. P value: P values < 0.05 are shown. Significant P values after Bonferroni correction are highlighted. Fold Change: Fold increase over the predicted random distribution.

Strand	Net Gene splice site Score	original sequence	Name	cLAM
1	0.83	GCGGCGACTG^GTGAGTACGC	SD1	x
1	0.33	CTCGACGCAG^GACTCGGCTT	SA1	x
1	0.16	ATCCCTTCAG^ACAGGATCAG	SA9	
1	0.51	CAAAACAAAA^GTAAGACCAC	SD2	
1	0.43	CCTCCTCCAG^GTCTGAAGAT	SA7	x
1	0.3	GCAGCTCCAG^GCAAGAATCC	SD3	
2	0.14	CCACAGCCAG^GATTCTTGCC	SA21	
2	0.15	CTTTCCACAG^CCAGGATTCT	SA20	
2	0.33	GATCCTTTAG^GTATCTTTCC	SA6	
2	0.39	CTCCATCCAG^GTCGTGTGAT	SA5	
1	0.28	ATCGTTTCAG^ACCCACCTCC	SA2	x
2	0.49	GGGTTGGGAG^GTGGGTCTGA	SD6	
1	0.19	CAACCCCGAG^GGGACCCGAC	SA10	
1	0.18	GACCCGACAG^GCCCGAAGGA	SA11	
1	0.55	GATCTCGACG^GTATCGGTTA	SD4	x
2	0.41	GGTCTTAAAG^GTACCGAGCT	SD14	
2	0.34	CAGCTGCCTT^GTAAGTCATT	SD15	
1	na	TCTCTAGCAG^TGGCGCCCGA	SA3	x
1	na	AAATCTCTAG^CAGTGGCGCC	SA4	x
2	na	TAAAGCTTGC^CTTGAGTGCT	SD5	x

Table 3: Internal splice sites identified by cLAM and Net2Gene splice site prediction

Strand: is the strand in which is located the splice site. 1 is the positive strand, 2 is the negative strand

Net Gene splice site score: confidence value provided by NetGene software that estimates the probability that a given sequence is a true splice site (1= maximum value)

Original Sequence: Lentiviral backbone sequence with the splice sites indicated by ^

Name: Splice site ID

cLAM: Splice sites identified by cLAM are indicated by X

Example 3 – Further splice sites defined herein were identified by splice trap experiments.

To test the consequences of LV integration on the expression of targeted genes, Jurkat and SupT1 human T cell lines were transduced at high MOI with a splice trap, HIV-derived SIN LV lacking an internal promoter and carrying the EGFP reporter gene downstream of an internal ribosomal entry site (referred to as the IRES-GFP) vector. Upon vector integration in a transcribed gene in the same orientation, a splicing event trapping the IRES-GFP cassette into a mature transcript may result in the expression of the reporter gene.

To characterize the chimeric transcripts generated by the IRES-GFP vector, nested 5′ rapid amplification of cDNA ends (RACE) PCR or RT-PCR were performed on poly(A)+ RNA obtained from 38 selected Jurkat and SupT1 clones, using forward primers annealing to the exons upstream of the LV integration sites and a reverse primer (Lenti-rev) annealing to the provirus downstream the HIV *gag* major splice acceptor (SA) site SA7. Fusion transcripts were detected for more than 60% of the mapped integration sites. In many cases, we detected amplicons of different molecular weight with the same primer pairs, indicating the existence of multiple gene-vector fusion transcripts from the same gene. Sequencing of the PCR products allowed the identification of the SA sites used in combination with splice donor (SD) sites in the upstream exon to generate the fusion transcripts.

Example 4 - A b-globin transgene provides alternative splicing signals when integrated into active genes.

A popular way to express an intron-containing transgene in a LV is to insert it in reverse transcriptional orientation with respect to the vector backbone. The paradigm antisense LVs were those expressing the human b-globin gene, such as GLOBE. To analyze the consequences of integrating a transgene containing natural intron-exon junctions on target gene expression, we analyzed HEL clones transduced with GLOBE and harboring reverse-oriented proviruses, in which the b-globin transgene is in the same transcriptional orientation as the target gene. RT-PCR analysis was performed by using a forward primer annealing to the exon upstream of the b-globin

transgene (E-for, Figure 9A) and a reverse primer specific for the b-globin third exon (Globin-rev, Figure 9A). We were able to detect chimeric transcripts in 55% (12 out of 22) of the analyzed proviruses in 13 HEL clones. Cloning and sequencing of the PCR products identified 4 species of transcripts: type-4 transcripts, splicing the upstream exon SD site to the constitutive SA site of the second intron of the b globin gene; type-5 transcripts, splicing the upstream exon SD to a cryptic SA site located in the first exon of the b-globin; type-6 transcripts, splicing the upstream exon SD to cryptic SA sites located in the LCR HS3 element and cryptic SD sites in HS3 to the constitutive SA site of the b-globin second intron; and type-7 transcripts, splicing cryptic SD sites in HS3 to the cryptic SA site in the b-globin first exon (Figure 9B). Constitutive splicing of the b-globin second and third exons occurred in all transcript types, while the first exon was retained in type-5 and type-7 transcripts (Figure 9B). In terms of relative frequency, aberrant type-4, -6, and -7 transcripts were all found in approximately 27% of the cases (9, 9, and 8 out of 29 sequenced transcripts, respectively), while type-5 transcripts were detected in only 3 cases. Sequencing of the splice junctions identified 6 different cryptic SD sites (sites A–F, Figure 9C) and 3 SA sites (sites J–L, Figure 9C) in the HS3 element and 1 cryptic SA site in the 5′ UTR of the b-globin first exon (site M, Figure 9C). The most frequently used SA sites were J and M, identified in 10 and 13 out of 27 sequenced transcripts, respectively. The most frequently used SD sites were B, C, and F, mapped in 10, 6, and 4 out of 24 transcripts, respectively, while the A, D, and E sites were found in only 1 or 2 cases.

Example 5 - Elimination of unwanted fusion transcripts by adding tags recognized by specific microRNAs that will trigger their selective degradation.

According to another aspect of the present invention there is provided a vector comprising an miRNA target sequence wherien said miRNA target sequence is positioned upstream of a splice donor site or downstream of a splice acceptor site, wherein said splice donor or splice acceptor site is responsible for splicing events that generate unwanted fusion transcripts comprisng vector sequences and cellular mRNAs, wherein said miRNA target sequence casues degradation of said unwanted fusion transcripst comprising a corresponding miRNA.

To effectively blunt the genotoxicity of aberrant splicing, sequences complementary of the hematopoietic microRNA142 is cloned in the exonic portion upstream the

splice donor sequence or downstream the splice acceptor sequence on the genotoxic LV.SF.LTR (LV.SF.LTR.mirT.SD or LV.SF.LTR.SA.mirT). Because the already identified LV exons map in positions that play an important role in the viral live cycle, insertion of these microRNAs tags could be detrimental to vector titer. Therefore, insertions of different numbers (1 to 8) and combination of targets in different sites within these vectors may be tested (fragments generated by outsourced DNA synthesis) for vector production efficiency by comparison to standard LV designs. The modified vector may be used to in vitro transduce hematopoietic cell lines that express high levels of the microRNA142 (for example JY or U937 cell lines). The presence of the spliced transcript may be then evaluated by RT QPCR and the expression levels compared to non modified parental vector. The strategy is depicted in Figure 5.

Summary

The invention is designed to generate a novel LV backbone to increase the efficacy and safety of gene transfer and therapy by reducing the probability of formation of potentially dangerous chimerical LV-cellular mRNAs.

Comprehensive efficacy and safety data have been provided for lentiviral vectors (LV) with self-inactivating (SIN) Long terminal Repeats (LTRs). On the other hand, an LV with active LTRs (LV.SF.LTR) was oncogenic in an in vivo genotoxicity assay based on transduction and transplantation of tumor prone hematopoietic stem cells (HSCs). LV.SF.LTR-mediated oncogenesis occurs by deregulating proto-oncogenes found at or nearby the insertion site, or by a process of splicing capture generating chimerical transcripts between LV and cellular mRNAs. Moreover, in a recent LV-based clinical trial for the cure of beta-thalassemia, a dominant cell clone harbored an integrated vector copy that caused transcriptional activation of HMGA2 by aberrant splicing. The activation of HMGA2 has been suggested to be causative of the observed clonal dominance. Thus, there is emerging evidence that the potential of inducing aberrant transcripts might constitute a previously unappreciated genotoxicity factor for gene therapy vectors. In this perspective, the generation of novel LV constructs in which splice sites are recoded and eliminated is of particular importance for future gene therapy applications.

Appendix – Sequence ID Nos: 1, 2 and 3

Sequence ID No: 1 - Annotated Gene Bank files of LV transfer plasmids with all splice sites identified. #277.pCCLsin.PPT contains all splice sites identified in the lentiviral backbone.

Sequence ID No: 2 - GLOBE contains splice sites contained in the HS3 region of the β-globin LCR. Note that Gene Bank files are also included as independent files that can be opened by the proper software (for example VectorNTi) for the graphical visualization of the maps and full report of the different features contained in the plasmids.

Sequence ID No: 3 - Annotated Gene Bank files of LV transfer plasmids with all splice sites identified. #743.pCCLsin.PPT contains all splice sites identified in the lentiviral backbone.

Appendix - SEQ ID NO1

```
LOCUS       #277.pCCLsin.PPT        7827 bp    DNA     circular     20-APR-2012
SOURCE
  ORGANISM
COMMENT     This file is created by Vector NTI
            http://www.invitrogen.com/
COMMENT     ORIGDB|GenBank
COMMENT     VNTDATE|623877990|
COMMENT     VNTDBDATE|623879420|
COMMENT     LSOWNER|
COMMENT     VNTNAME|#277.pCCLsin.PPT.hPGK.GFP.pre con splicing SIMPLIFIED|
COMMENT     VNTAUTHORNAME|Demo User|
COMMENT     VNTOAUTHORNAME|IRCC - ROSSO 3|
COMMENT     Vector_NTI_Display_Data_(Do_Not_Edit!)
COMMENT     (SXF
COMMENT      (CGexDoc "#277.pCCLsin.PPT.hPGK.GFP.pre con splicing SIMPLIFIED" 0 7827
COMMENT       (CDBMol 0 0 1 1 1 0 0 0 0 "" "" 0 0 0 0 (CObList) (CObList) (CObList)
COMMENT        (CObList) -1 "")
COMMENT       (CDocSetData 1 0 0 0 0 0 "MAIN" 1 1 1 1 0 0 1 1 0 1 10 10 4294967295 50 0
COMMENT        1 0 (CHomObj 0 0 0 3 75) (CWordArray) (CWordArray)
COMMENT        (CStringList "ApaLI" "AvaI" "BamHI" "ClaI" "EcoRI" "HindIII" "NcoI"
COMMENT         "PstI" "SmaI" "XmaI") (CStringList "atg" "gtg")
COMMENT        (CStringList "taa" "tga" "tag") (CObList) 1 "{(0,1),2}" 0 0 "" 0
COMMENT        4294967295 0 1 0 0 0 0 0 "MAIN" 0 0 30 0
COMMENT        (CProteinMotifSearchObject 70 20 1 1 1 1 0 0 1 0 0 0 0 0))
COMMENT       (CMolPar 0 0 0 0 4294967295 1 7827 0 0 0 0 0 0 0 0) (CStringList)
COMMENT       (CStringList) (CObList) (COAPar 25 250 50 0 6 4 3 7)
COMMENT       (COAPar 25 250 50 0 6 4 3 7) (COAPar 25 250 50 0 6 4 3 7) (CObList)
COMMENT       (CObList
COMMENT        #0=(CFSignal (CObList) "GFP" 4 0 0 5159 5878 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "5159..5878")
COMMENT        #1=(CFSignal (CObList) "Wpre" 85 0 0 5888 6484 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "5888..6484")
COMMENT        #2=(CFSignal (CObList) "hPGK" 29 0 0 4622 5137 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "4622..5137")
COMMENT        #3=(CFSignal (CObList) "dR3RU5" 19 0 0 6568 6801 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "6568..6801")
COMMENT        #4=(CFSignal (CObList) "RU5" 19 0 0 2891 3072 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "2891..3072")
COMMENT        #5=(CFSignal (CObList) "SA1" 38 0 0 3127 3128 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "3127..3128")
COMMENT        #6=(CFSignal (CObList) "SA9" 38 0 0 3431 3432 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "3431..3432")
COMMENT        #7=(CFSignal (CObList) "SD2" 38 0 0 3557 3558 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "3557..3558")
COMMENT        #8=(CFSignal (CObList) "SD1" 38 0 0 3178 3179 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "3178..3179")
COMMENT        #9=(CFSignal (CObList) "SA20" 38 0 1 3933 3934 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "complement(3933..3934)")
COMMENT        #10=(CFSignal (CObList) "SA5" 38 0 0 4069 4070 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "4069..4070")
COMMENT        #11=(CFSignal (CObList) "SA2" 38 0 0 4341 4342 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "4341..4342")
COMMENT        #12=(CFSignal (CObList) "SA10" 38 0 0 4361 4362 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "4361..4362")
COMMENT        #13=(CFSignal (CObList) "SA11" 38 0 0 4373 4374 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "4373..4374")
COMMENT        #14=(CFSignal (CObList) "SD4" 37 0 0 4450 4451 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "4450..4451")
COMMENT        #15=(CFSignal (CObList) "SD2" 38 0 1 6511 6512 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "complement(6511..6512)")
COMMENT        #16=(CFSignal (CObList) "SD14" 21 0 0 6500 6501 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "6500..6501")
COMMENT        #17=(CFSignal (CObList) "SA4" 21 0 0 3068 3069 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "3068..3069")
COMMENT        #18=(CFSignal (CObList) "SA3" 21 0 0 3071 3072 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "3071..3072")
COMMENT        #19=(CFSignal (CObList) "SD5" 21 0 0 2974 2975 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "2974..2975")
COMMENT        #20=(CFSignal (CObList) "SA7" 38 0 1 3597 3598 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "complement(3603..3604)")
COMMENT        #21=(CFSignal (CObList) "SD3" 38 0 0 3920 3921 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "3918..3919")
COMMENT        #22=(CFSignal (CObList) "SA21" 38 0 1 3929 3930 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "complement(3927..3928)")
COMMENT        #23=(CFSignal (CObList) "SA6" 38 0 0 3947 3948 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "3955..3956")
COMMENT        #24=(CFSignal (CObList) "SD6" 38 0 1 4347 4348 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "complement(4349..4350)")
COMMENT        #25=(CFSignal (CObList) "SD15" 21 0 0 6520 6521 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "")) (CObList) (CObList) (CObList) (CObList)
COMMENT       (CObList) (CObList)
COMMENT       (CTextView 0
COMMENT        #26=(CGroupPar (CParagraph 0 (0 0) 1 2 0 0 180)
```

Appendix - SEQ ID NO1

```
COMMENT              (CObjectList
COMMENT               #27=(CRefLinePar
COMMENT                    (CLinePar (CParagraph 0 (0 0) 0 2 0 0 233)
COMMENT                     "#277.pCCLsin.PPT.hPGK.GFP.pre con splicing SIMPLIFIED" 2) 5
COMMENT                    "" 0 4)
COMMENT               #28=(CFolderPar
COMMENT                    (CGroupPar (CParagraph 1 (0 0) 1 1 0 0 178)
COMMENT                     (CObjectList
COMMENT                      #29=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "DNA '#277.pCCLsin.PPT.hPGK.GFP.pre con splicing SIMPLIFIED'"
COMMENT                          1)
COMMENT                      #30=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Currently local object. Original author: IRCC - ROSSO 3"
COMMENT                          1)
COMMENT                      #31=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Created: 04/20/12 07:26PM" 1)
COMMENT                      #32=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Last Modified: 04/20/12 07:50PM" 1)
COMMENT                      #33=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "length: 7827 bp" 1)
COMMENT                      #34=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "storage type: Basic" 1)
COMMENT                      #35=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "form: Circular" 1))) "General Description")
COMMENT               #36=(CFolderPar
COMMENT                    (CGroupPar (CParagraph 2 (0 0) 1 1 0 0 178)
COMMENT                     (CObjectList
COMMENT                      #37=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Original Source Database: GenBank" 1)
COMMENT                      #38=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Modification Date in the Original DB: 20-APR-2012" 1)))
COMMENT                    "Standard Fields")
COMMENT               #39=(CFolderPar
COMMENT                    (CGroupPar (CParagraph 5 (0 0) 1 1 0 0 178)
COMMENT                     (CObjectList
COMMENT                      #40=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "IRCC - ROSSO 3" 1))) "Original Author")
COMMENT               #41=(CRefLinePar
COMMENT                    (CLinePar (CParagraph 0 (0 0) 0 2 0 0 233) "Comments" 2) 1 ""
COMMENT                    0 0)
COMMENT               #42=(CFolderPar
COMMENT                    (CGroupPar (CParagraph 8 (0 0) 1 2 0 0 178) (CObjectList))
COMMENT                    "Annotations")
COMMENT               #43=(CFolderPar
COMMENT                    (CGroupPar (CParagraph 12 (6 0) 1 1 0 0 178)
COMMENT                     (CObjectList
COMMENT                      #44=(CFolderPar
COMMENT                           (CGroupPar (CParagraph 4 (7 4 0) 1 1 1 0 178)
COMMENT                            (CObjectList
COMMENT                             #45=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 5159 (3 #0# 0) 1 2 2 0 327)
COMMENT                                   (CObjectList
COMMENT                                    #46=(CLinePar
COMMENT                                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                         "Start: 5159  End: 5878" 1)
COMMENT                                    #47=(CLinePar
COMMENT                                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                         "Original Location Description:" 1)
COMMENT                                    #48=(CLinePar
COMMENT                                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                         "   5159..5878" 1))) "GFP")))
COMMENT                           "CDS (1 total)")
COMMENT                      #49=(CFolderPar
COMMENT                           (CGroupPar (CParagraph 19 (7 19 0) 1 1 1 0 178)
COMMENT                            (CObjectList
COMMENT                             #50=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 2891 (3 #4# 0) 1 2 2 0 194)
COMMENT                                   (CObjectList
COMMENT                                    #51=(CLinePar
COMMENT                                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                         "Start: 2891  End: 3072" 1)
COMMENT                                    #52=(CLinePar
COMMENT                                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                         "Original Location Description:" 1)
COMMENT                                    #53=(CLinePar
COMMENT                                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                         "   2891..3072" 1))) "RU5")
COMMENT                             #54=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 6568 (3 #3# 0) 1 2 2 0 194)
COMMENT                                   (CObjectList
COMMENT                                    #55=(CLinePar
COMMENT                                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                         "Start: 6568  End: 6801" 1)
```

```
COMMENT                                  Appendix - SEQ ID NO1
COMMENT                                     #56=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Original Location Description:" 1)
COMMENT                                     #57=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "    6568..6801" 1))) "dR3RU5")))
COMMENT                               "LTR (2 total)")
COMMENT                          #58=(CFolderPar
COMMENT                               (CGroupPar (CParagraph 21 (7 21 0) 1 1 1 0 178)
COMMENT                                (CObjectList
COMMENT                                 #59=(CFolderPar
COMMENT                                      (CGroupPar
COMMENT                                       (CParagraph 2974 (3 #19# 0) 1 2 2 0 194)
COMMENT                                       (CObjectList
COMMENT                                        #60=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Start: 2974  End: 2975" 1)
COMMENT                                        #61=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Original Location Description:" 1)
COMMENT                                        #62=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "    2974..2975" 1))) "SD5")
COMMENT                                 #63=(CFolderPar
COMMENT                                      (CGroupPar
COMMENT                                       (CParagraph 3068 (3 #17# 0) 1 2 2 0 194)
COMMENT                                       (CObjectList
COMMENT                                        #64=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Start: 3068  End: 3069" 1)
COMMENT                                        #65=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Original Location Description:" 1)
COMMENT                                        #66=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "    3068..3069" 1))) "SA4")
COMMENT                                 #67=(CFolderPar
COMMENT                                      (CGroupPar
COMMENT                                       (CParagraph 3071 (3 #18# 0) 1 2 2 0 194)
COMMENT                                       (CObjectList
COMMENT                                        #68=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Start: 3071  End: 3072" 1)
COMMENT                                        #69=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Original Location Description:" 1)
COMMENT                                        #70=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "    3071..3072" 1))) "SA3")
COMMENT                                 #71=(CFolderPar
COMMENT                                      (CGroupPar
COMMENT                                       (CParagraph 6500 (3 #16# 0) 1 2 2 0 194)
COMMENT                                       (CObjectList
COMMENT                                        #72=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Start: 6500  End: 6501" 1)
COMMENT                                        #73=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Original Location Description:" 1)
COMMENT                                        #74=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "    6500..6501" 1))) "SD14")
COMMENT                                 #75=(CFolderPar
COMMENT                                      (CGroupPar
COMMENT                                       (CParagraph 6520 (3 #25# 0) 1 2 2 0 194)
COMMENT                                       (CObjectList
COMMENT                                        #76=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Start: 6520  End: 6521" 1))) "SD15")))
COMMENT                               "Misc. Feature (5 total)")
COMMENT                          #77=(CFolderPar
COMMENT                               (CGroupPar (CParagraph 29 (7 29 0) 1 1 1 0 178)
COMMENT                                (CObjectList
COMMENT                                 #78=(CFolderPar
COMMENT                                      (CGroupPar
COMMENT                                       (CParagraph 4622 (3 #2# 0) 1 2 2 0 194)
COMMENT                                       (CObjectList
COMMENT                                        #79=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Start: 4622  End: 5137" 1)
COMMENT                                        #80=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Original Location Description:" 1)
COMMENT                                        #81=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "    4622..5137" 1))) "hPGK")))
```

```
                    Appendix - SEQ ID NO1
COMMENT          "Promoter Eukaryotic (1 total)")
COMMENT     #82=(CFolderPar
COMMENT          (CGroupPar (CParagraph 37 (7 37 0) 1 1 1 0 178)
COMMENT           (CObjectList
COMMENT            #83=(CFolderPar
COMMENT                 (CGroupPar
COMMENT                  (CParagraph 4450 (3 #14# 0) 1 2 2 0 194)
COMMENT                  (CObjectList
COMMENT                   #84=(CLinePar
COMMENT                        (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                        "Start: 4450  End: 4451" 1)
COMMENT                   #85=(CLinePar
COMMENT                        (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                        "Original Location Description:" 1)
COMMENT                   #86=(CLinePar
COMMENT                        (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                        "   4450..4451" 1))) "SD4")))
COMMENT          "Silencer (1 total)")
COMMENT     #87=(CFolderPar
COMMENT          (CGroupPar (CParagraph 38 (7 38 0) 1 1 1 0 178)
COMMENT           (CObjectList
COMMENT            #88=(CFolderPar
COMMENT                 (CGroupPar
COMMENT                  (CParagraph 3127 (3 #5# 0) 1 2 2 0 194)
COMMENT                  (CObjectList
COMMENT                   #89=(CLinePar
COMMENT                        (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                        "Start: 3127  End: 3128" 1)
COMMENT                   #90=(CLinePar
COMMENT                        (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                        "Original Location Description:" 1)
COMMENT                   #91=(CLinePar
COMMENT                        (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                        "   3127..3128" 1))) "SA1")
COMMENT             #92=(CFolderPar
COMMENT                  (CGroupPar
COMMENT                   (CParagraph 3178 (3 #8# 0) 1 2 2 0 194)
COMMENT                   (CObjectList
COMMENT                    #93=(CLinePar
COMMENT                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                         "Start: 3178  End: 3179" 1)
COMMENT                    #94=(CLinePar
COMMENT                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                         "Original Location Description:" 1)
COMMENT                    #95=(CLinePar
COMMENT                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                         "   3178..3179" 1))) "SD1")
COMMENT             #96=(CFolderPar
COMMENT                  (CGroupPar
COMMENT                   (CParagraph 3431 (3 #6# 0) 1 2 2 0 194)
COMMENT                   (CObjectList
COMMENT                    #97=(CLinePar
COMMENT                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                         "Start: 3431  End: 3432" 1)
COMMENT                    #98=(CLinePar
COMMENT                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                         "Original Location Description:" 1)
COMMENT                    #99=(CLinePar
COMMENT                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                         "   3431..3432" 1))) "SA9")
COMMENT             #100=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3557 (3 #7# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #101=(CLinePar
COMMENT                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                           "Start: 3557  End: 3558" 1)
COMMENT                     #102=(CLinePar
COMMENT                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                           "Original Location Description:" 1)
COMMENT                     #103=(CLinePar
COMMENT                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                           "   3557..3558" 1))) "SD2")
COMMENT             #104=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3597 (3 #20# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #105=(CLinePar
COMMENT                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                           "Start: 3597  End: 3598 (Complementary)"
COMMENT                           1)
COMMENT                     #106=(CLinePar
COMMENT                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                           "Original Location Description:" 1)
COMMENT                     #107=(CLinePar
COMMENT                           (CParagraph 0 (0 0) 1 2 3 0 180)
```

```
COMMENT                                   Appendix - SEQ ID NO1
COMMENT                                          "    complement(3603..3604)" 1))) "SA7")
COMMENT                                 #108=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 3920 (3 #21# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #109=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 3920  End: 3921" 1)
COMMENT                                         #110=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #111=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "    3918..3919" 1))) "SD3")
COMMENT                                 #112=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 3929 (3 #22# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #113=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 3929  End: 3930 (Complementary)"
COMMENT                                               1)
COMMENT                                         #114=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #115=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "    complement(3927..3928)" 1)))
COMMENT                                       "SA21")
COMMENT                                 #116=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 3933 (3 #9# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #117=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 3933  End: 3934 (Complementary)"
COMMENT                                               1)
COMMENT                                         #118=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #119=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "    complement(3933..3934)" 1)))
COMMENT                                       "SA20")
COMMENT                                 #120=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 3947 (3 #23# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #121=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 3947  End: 3948" 1)
COMMENT                                         #122=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #123=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "    3955..3956" 1))) "SA6")
COMMENT                                 #124=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 4069 (3 #10# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #125=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 4069  End: 4070" 1)
COMMENT                                         #126=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #127=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "    4069..4070" 1))) "SA5")
COMMENT                                 #128=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 4341 (3 #11# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #129=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 4341  End: 4342" 1)
COMMENT                                         #130=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #131=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "    4341..4342" 1))) "SA2")
COMMENT                                 #132=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 4347 (3 #24# 0) 1 2 2 0 194)
```

```
COMMENT                                  Appendix - SEQ ID NO1
COMMENT                                      (CObjectList
COMMENT                                       #133=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Start: 4347  End: 4348 (Complementary)"
COMMENT                                             1)
COMMENT                                       #134=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "Original Location Description:" 1)
COMMENT                                       #135=(CLinePar
COMMENT                                             (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                             "   complement(4349..4350)" 1))) "SD6")
COMMENT                                 #136=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 4361 (3 #12# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #137=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 4361  End: 4362" 1)
COMMENT                                         #138=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #139=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "   4361..4362" 1))) "SA10")
COMMENT                                 #140=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 4373 (3 #13# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #141=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 4373  End: 4374" 1)
COMMENT                                         #142=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #143=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "   4373..4374" 1))) "SA11")
COMMENT                                 #144=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 6511 (3 #15# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #145=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 6511  End: 6512 (Complementary)"
COMMENT                                               1)
COMMENT                                         #146=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #147=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "   complement(6511..6512)" 1))) "SD2")))
COMMENT                                "Splicing Signal (15 total)")
COMMENT                           #148=(CFolderPar
COMMENT                                 (CGroupPar (CParagraph 85 (7 85 0) 1 1 1 0 178)
COMMENT                                  (CObjectList
COMMENT                                   #149=(CFolderPar
COMMENT                                         (CGroupPar
COMMENT                                          (CParagraph 5888 (3 #1# 0) 1 2 2 0 194)
COMMENT                                          (CObjectList
COMMENT                                           #150=(CLinePar
COMMENT                                                 (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                                 "Start: 5888  End: 6484" 1)
COMMENT                                           #151=(CLinePar
COMMENT                                                 (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                                 "Original Location Description:" 1)
COMMENT                                           #152=(CLinePar
COMMENT                                                 (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                                 "   5888..6484" 1))) "Wpre")))
COMMENT                                "Misc. Difference (1 total)"))) "Feature Map"))))
COMMENT                 (CGraphView
COMMENT                  (CWStyleSheet
COMMENT                   (CObjectList
COMMENT                    #153=(CWidgetStyle "RSite Label" 1 (LOGPEN 0 0 13408563) 1 0 1
COMMENT                          (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Georgia") 0.555556 0 1 5
COMMENT                          "@N (@S)" 0)
COMMENT                    #154=(CWidgetStyle "Signal Label" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                          (LOGFONT 0 0 0 0 700 0 0 0 0 3 2 1 34 "Arial") 0.666667 0 1 1
COMMENT                          "@N" 0)
COMMENT                    #155=(CWidgetStyle "Molecule Label 2" 0 0 1
COMMENT                          (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Georgia") 0.555556 0 1 16
COMMENT                          "@L bp" 0)
COMMENT                    #156=(CWidgetStyle "Molecule Label 1" 0 0 1
COMMENT                          (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Verdana") 0.833333 0 1 1
COMMENT                          "@N" 0)
COMMENT                    #157=(CWidgetStyle "Shape 3" 1 (LOGPEN 0 0 3355545) 1 1
COMMENT                          (LOGBRUSH 0 6724095 0) 0 0 1 (LOGSHAPE 9 1 0.8 1.8 0))
```

Appendix - SEQ ID NO1

```
COMMENT          #158=(CWidgetStyle "Shape 1" 1 (LOGPEN 0 0 6723840) 1 1
COMMENT                (LOGBRUSH 0 10079334 0) 0 0 1 (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT          #159=(CWidgetStyle "Axis" 1 (LOGPEN 0 0 10079436) 2 1
COMMENT                (LOGBRUSH 0 13434879 0) 0 0 1 (LOGSHAPE 10 1 0 0 0))
COMMENT          #160=(CWidgetStyle "Line 2" 1 (LOGPEN 0 0 6723840) 8 0 0 0 1
COMMENT                (LOGSHAPE 1 1.9 0 0 0))
COMMENT          #161=(CWidgetStyle "RSite" 1 (LOGPEN 0 0 10053171) 8 0 0 0 1
COMMENT                (LOGSHAPE 1 1.9 0 0 0))
COMMENT          #162=(CWidgetStyle "Short Signal" 1 (LOGPEN 0 0 13395507) 10 0 0 0 1
COMMENT                (LOGSHAPE 1 1.9 0 0 0))
COMMENT          #163=(CWidgetStyle "Uniq RSite Label" 1 (LOGPEN 0 0 153) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Georgia") 0.555556 128 1
COMMENT                5 "@N (@S)" 0)
COMMENT          #164=(CWidgetStyle "Vanilla" 1 (LOGPEN 0 0 0) 1 1
COMMENT                (LOGBRUSH 0 16777215 0) 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 7 48 2 18 "Times New Roman") 0.8 0
COMMENT                1 2 "?" 0)
COMMENT          #165=(CWidgetStyle "Mark 1" 0 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 2 7 48 2 2 "Windings") 0.7 0 1 2 "?"
COMMENT                0)
COMMENT          #166=(CWidgetStyle "Motif Label" 1 (LOGPEN 0 0 16744512) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.611111 8388608
COMMENT                1 65535 "@N (@H)" 0)
COMMENT          #167=(CWidgetStyle "Fragment Label 2" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 49 "Courier New") 1.05 0 1 48
COMMENT                "@F bp (molecule @L bp)" 0)
COMMENT          #168=(CWidgetStyle "Fragment Label 1" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.91 0 1 1
COMMENT                "Fragment of @N" 0)
COMMENT          #169=(CWidgetStyle "Shape 4" 1 (LOGPEN 0 0 0) 1 1
COMMENT                (LOGBRUSH 2 8388608 5) 0 0 0)
COMMENT          #170=(CWidgetStyle "Shape 2" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 0 128 0) 0
COMMENT                0 0)
COMMENT          #171=(CWidgetStyle "Shape 0" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 0 0 0) 0 0
COMMENT                0)
COMMENT          #172=(CWidgetStyle "ORF" 1 (LOGPEN 0 0 16384) 8 0 0 0 1
COMMENT                (LOGSHAPE 7 0.2 3.41182 2.86186 0.609808))
COMMENT          #173=(CWidgetStyle "Line 4" 1 (LOGPEN 0 0 32768) 8 0 0 0 0)
COMMENT          #174=(CWidgetStyle "Line 3" 1 (LOGPEN 0 0 16711680) 8 0 0 0 0)
COMMENT          #175=(CWidgetStyle "Line 1" 1 (LOGPEN 0 0 16711680) 1 0 0 0 0)
COMMENT          #176=(CWidgetStyle "Short Promoter" 1 (LOGPEN 0 0 128) 6 0 0 0 0)
COMMENT          #177=(CWidgetStyle "Motif" 1 (LOGPEN 0 0 0) 1 0 0 0 0)
COMMENT          #178=(CWidgetStyle "Line 0" 1 (LOGPEN 0 0 0) 8 0 0 0 0)
COMMENT          #179=(CWidgetStyle "Void" 0 0 0 0 0)
COMMENT          #180=(CWidgetStyle "General Label" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman") 0.91 0
COMMENT                1 3 "@T @N " 0)
COMMENT          #181=(CWidgetStyle "Position" 1 (LOGPEN 0 0 0) 1 0 0 0 0)
COMMENT          #182=(CWidgetStyle "Annotation" 0 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman") 0.91 0
COMMENT                0 0)
COMMENT          #183=(CWidgetStyle "Position Label" 1 (LOGPEN 0 0 8388608) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 1 0 0 3 2 1 34 "Arial") 0.63 8388608 1 1
COMMENT                "@N" 0)
COMMENT          #184=(CWidgetStyle "Range" 1 (LOGPEN 0 0 0) 1 1
COMMENT                (LOGBRUSH 0 16777215 0) 0 0 0)
COMMENT          #185=(CWidgetStyle "Range Label" 1 (LOGPEN 0 0 8388608) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 1 0 0 3 2 1 34 "Arial") 0.63 8388608 1 1
COMMENT                "@N" 0)
COMMENT          #186=(CWidgetStyle "ORF Label" 1 (LOGPEN 0 0 49216) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman")
COMMENT                0.611111 0 1 65535 "@N" 0)
COMMENT          #187=(CWidgetStyle "CDS Label" 1 (LOGPEN 0 0 4227264) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.555556 255 1 1
COMMENT                "@N" 0)
COMMENT          #188=(CWidgetStyle "Shape 5" 1 (LOGPEN 0 0 0) 3 1
COMMENT                (LOGBRUSH 0 16777113 0) 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 7 48 2 50 "Arial") 0.9 0 0 1
COMMENT                (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT          #189=(CWidgetStyle "CDS" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 2 39423 3) 0 0
COMMENT                1 (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT          #190=(CWidgetStyle "Label 2" 1 (LOGPEN 0 0 4227264) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.944444 8388608
COMMENT                1 1 "@N" 0)
COMMENT          #191=(CWidgetStyle "Label 3" 1 (LOGPEN 0 0 8421376) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 700 255 0 0 0 3 2 1 34 "Arial") 0.833333 255 1
COMMENT                5 "@N (@S)" 0)
COMMENT          #192=(CWidgetStyle "Label 4" 1 (LOGPEN 0 0 8437824) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.722222 0 1 5
COMMENT                "@N (@S)" 0)) 0.164644 1.74233 0.164644 2.53336
COMMENT         (70 (CShapeMapEntry 0 "Unique RSite" 1 "Uniq RSite Label") 67
COMMENT          (CShapeMapEntry 0 "ORF" 0 "ORF Label")) 40.0378 40.0378 39 39 0.1
COMMENT         -7827) 1 0 1 1 1
COMMENT        (mapper: 14.9597 -14.6005 87.75 87.75 0.01 10 14 7827 7827 1 0 0)
COMMENT        #193=(CGroupWidget (CWidget 0 (0 0) 1 2 0 0 Nil 1768117093 100)
COMMENT             (CObjectList
```

Appendix - SEQ ID NO1

```
COMMENT                 #194=(CGroupWidget (CWidget 1 (0 0) 1 2 0 0 Nil -399 100)
COMMENT                       (CObjectList
COMMENT                        #195=(CAxis
COMMENT                              (CWideLine
COMMENT                               (CWidget 0 (0 0) 1 2 0 0 #159# 59085940 0)
COMMENT                               (LOGPEN 0 2 10079436) 2 (LOGBRUSH 0 13434879 0) 1
COMMENT                               6.27499 6.27299 1 0.0214037) 0.080148)
COMMENT                        #196=(CLabel (CWidget 1001 (0 0) 1 2 0 0 #156# 0 100)
COMMENT                              (LOGPEN 0 0 0) 1
COMMENT                              (LOGFONT 92 35 0 0 400 0 0 0 0 3 2 1 34 "Verdana")
COMMENT                              2.53336 0.833333 0
COMMENT                              "#277.pCCLsin.PPT.hPGK.GFP.pre con splicing SIMPLIFIED"
COMMENT                              "@N" 1 0 0.5 0 -4.55956 23.2821 1.04843 Nil)
COMMENT                        #197=(CLabel
COMMENT                              (CWidget 1002 (0 0) 1 2 0 0 #155# 55927284 100)
COMMENT                              (LOGPEN 0 0 0) 1
COMMENT                              (LOGFONT 61 23 0 0 400 0 0 0 0 3 2 1 18 "Georgia")
COMMENT                              2.53336 0.555556 0 "7827 bp" "@L bp" 16 0 -0.8 0
COMMENT                              -5.82624 2.11966 0.683761 Nil)) (CObjectList))
COMMENT                 #198=(CGroupWidget (CWidget 10 (6 0) 1 2 0 0 Nil -399 100)
COMMENT                       (CObjectList
COMMENT                        #199=(CGroupWidget
COMMENT                              (CWidget 4 (7 4 0) 1 2 0 0 Nil -250 100)
COMMENT                              (CObjectList
COMMENT                               #200=(CWideArrow
COMMENT                                     (CWideLine
COMMENT                                      (CWidget 0 (3 #0# 0) 1 2 0 0 #157#
COMMENT                                       2082603970 100) (LOGPEN 0 0 3355545) 1
COMMENT                                      (LOGBRUSH 0 6724095 0) 1 1.56208 2.13915 1
COMMENT                                      0.082322) 0.8 1.8 0)
COMMENT                               #201=(CLabel
COMMENT                                     (CWidget 0 (0 0) 1 2 0 0 #154# 920 100)
COMMENT                                     (LOGPEN 0 0 0) 1
COMMENT                                     (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                      "Arial") 2.53336 0.666667 0 "GFP" "@N" 1 0 0
COMMENT                                     6.58999 1.97573 1.4359 0.843305 #200#))
COMMENT                              (CObjectList))
COMMENT                        #202=(CGroupWidget
COMMENT                              (CWidget 19 (7 19 0) 1 2 0 0 Nil -152 100)
COMMENT                              (CObjectList
COMMENT                               #203=(CWideArrow
COMMENT                                     (CWideLine
COMMENT                                      (CWidget 0 (3 #3# 0) 1 2 0 0 #169# 826617139
COMMENT                                       100) (LOGPEN 0 0 0) 1
COMMENT                                      (LOGBRUSH 2 8388608 5) 1 0.822319 1.00986 1
COMMENT                                      0.082322) 0.8 1.8 0)
COMMENT                               #204=(CLabel
COMMENT                                     (CWidget 0 (0 0) 1 2 0 0 #154# 56733512 100)
COMMENT                                     (LOGPEN 0 0 0) 1
COMMENT                                     (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                      "Arial") 2.53336 0.666667 0 "dR3RU5" "@N" 1
COMMENT                                     0 0 11.4769 1.97573 2.64387 0.843305 #203#)
COMMENT                               #205=(CWideArrow
COMMENT                                     (CWideLine
COMMENT                                      (CWidget 0 (3 #4# 0) 1 2 0 0 #169#
COMMENT                                       2082603826 100) (LOGPEN 0 0 0) 1
COMMENT                                      (LOGBRUSH 2 8388608 5) 1 3.81104 3.95691 1
COMMENT                                      0.082322) 0.8 1.8 0)
COMMENT                               #206=(CLabel
COMMENT                                     (CWidget 0 (0 0) 1 2 0 0 #154# 444 100)
COMMENT                                     (LOGPEN 0 0 0) 1
COMMENT                                     (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                      "Arial") 2.53336 0.666667 0 "RU5" "@N" 1 0 0
COMMENT                                     -4.74811 3.24241 1.39031 0.843305 #205#))
COMMENT                              (CObjectList))
COMMENT                        #207=(CGroupWidget
COMMENT                              (CWidget 21 (7 21 0) 1 2 0 0 Nil -448 100)
COMMENT                              (CObjectList
COMMENT                               #208=(CScratch
COMMENT                                     (CWidget 0 (3 #16# 0) 1 2 0 0 #162# 0 100)
COMMENT                                     (LOGPEN 0 16 13395507) 10 1 1.06437 1.9
COMMENT                                     0.082322 1)
COMMENT                               #209=(CLabel
COMMENT                                     (CWidget 0 (0 0) 1 2 0 0 #154# 268786312 100)
COMMENT                                     (LOGPEN 0 0 0) 1
COMMENT                                     (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                      "Arial") 2.53336 0.666667 0 "SD14" "@N" 1 0
COMMENT                                     0 10.3848 5.77577 1.7208 0.843305 #208#)
COMMENT                               #210=(CScratch
COMMENT                                     (CWidget 0 (3 #17# 0) 1 2 0 0 #162# 0 100)
COMMENT                                     (LOGPEN 0 16 13395507) 10 1 3.81505 1.9
COMMENT                                     0.082322 1)
COMMENT                               #211=(CLabel
COMMENT                                     (CWidget 0 (0 0) 1 2 0 0 #154# 55931316 100)
COMMENT                                     (LOGPEN 0 0 0) 1
COMMENT                                     (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
```

Appendix - SEQ ID NO1

```
COMMENT                    "Arial") 2.53336 0.666667 0 "SA4" "@N" 1 0 0
COMMENT                   -4.45072 4.50909 1.35613 0.843305 #210#)
COMMENT                 #212=(CScratch
COMMENT                    (CWidget 0 (3 #18# 0) 1 2 0 0 #162# 0 100)
COMMENT                    (LOGPEN 0 16 13395507) 10 1 3.81264 1.9
COMMENT                    0.082322 1)
COMMENT                 #213=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #154# 55929300 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SA3" "@N" 1 0 0
COMMENT                    -4.44035 5.77577 1.35613 0.843305 #212#)
COMMENT                 #214=(CScratch
COMMENT                    (CWidget 0 (3 #19# 0) 1 2 0 0 #162# 16848848
COMMENT                     100) (LOGPEN 0 16 13395507) 10 1 3.89038 1.9
COMMENT                    0.082322 1)
COMMENT                 #215=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #154# 1667721076 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SD5" "@N" 1 0 0
COMMENT                    -4.77577 1.97573 1.35613 0.843305 #214#)
COMMENT                 #216=(CScratch
COMMENT                    (CWidget 0 (3 #25# 0) 1 2 0 0 #162# 16933184
COMMENT                     100) (LOGPEN 0 16 13395507) 10 1 1.04834 1.9
COMMENT                    0.082322 1)
COMMENT                 #217=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #154# 1667327860 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SD15" "@N" 1 0
COMMENT                    0 10.4585 3.24241 1.7208 0.843305 #216#))
COMMENT              (CObjectList))
COMMENT           #218=(CGroupWidget
COMMENT              (CWidget 29 (7 29 0) 1 2 0 0 Nil -102 100)
COMMENT              (CObjectList
COMMENT               #219=(CWideArrow
COMMENT                    (CWideLine
COMMENT                     (CWidget 0 (3 #2# 0) 1 2 0 0 #158# 21 100)
COMMENT                     (LOGPEN 0 0 6723840) 1
COMMENT                     (LOGBRUSH 0 10079334 0) 1 2.15598 2.56955 1
COMMENT                     0.082322) 0.8 1.8 0)
COMMENT                    #220=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #154# 658328553 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "hPGK" "@N" 1 0
COMMENT                    0 4.46582 1.97573 1.91453 0.843305 #219#))
COMMENT              (CObjectList))
COMMENT           #221=(CGroupWidget
COMMENT              (CWidget 37 (7 37 0) 1 2 0 0 Nil -497 100)
COMMENT              (CObjectList
COMMENT               #222=(CScratch
COMMENT                    (CWidget 0 (3 #14# 0) 1 2 0 0 #162# 0 100)
COMMENT                    (LOGPEN 0 16 13395507) 10 1 2.7074 1.9
COMMENT                    0.082322 1)
COMMENT                    #223=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #154# 1668571491 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SD4" "@N" 1 0 0
COMMENT                    2.65035 1.97573 1.35613 0.843305 #222#))
COMMENT              (CObjectList))
COMMENT           #224=(CGroupWidget
COMMENT              (CWidget 38 (7 38 0) 1 2 0 0 Nil -497 100)
COMMENT              (CObjectList
COMMENT               #225=(CScratch
COMMENT                    (CWidget 0 (3 #5# 0) 1 2 0 0 #173# 0 100)
COMMENT                    (LOGPEN 0 14 32768) 8 1 3.76776 1.9 0.082322
COMMENT                    1)
COMMENT                    #226=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #154# 55928152 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SA1" "@N" 1 0 0
COMMENT                    -4.2467 7.04245 1.35613 0.843305 #225#)
COMMENT                    #227=(CScratch
COMMENT                    (CWidget 0 (3 #6# 0) 1 2 0 0 #173# 0 100)
COMMENT                    (LOGPEN 0 14 32768) 8 1 3.52411 1.9 0.082322
COMMENT                    1)
COMMENT                    #228=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #154# 658328013 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SA9" "@N" 1 0 0
COMMENT                    -3.19547 9.57581 1.35613 0.843305 #227#)
```

Appendix - SEQ ID NO1


```
COMMENT #229=(CScratch
COMMENT      (CWidget 0 (3 #7# 0) 1 2 0 0 #173# 669 100)
COMMENT      (LOGPEN 0 14 32768) 8 1 3.42312 1.9 0.082322
COMMENT      1)
COMMENT #230=(CLabel
COMMENT      (CWidget 0 (0 0) 1 2 0 0 #154# 268786312 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT       "Arial") 2.53336 0.666667 0 "SD2" "@N" 1 0 0
COMMENT      -2.75977 10.8425 1.35613 0.843305 #229#)
COMMENT #231=(CScratch
COMMENT      (CWidget 0 (3 #8# 0) 1 2 0 0 #173# 651165235
COMMENT       100) (LOGPEN 0 14 32768) 8 1 3.72688 1.9
COMMENT      0.082322 1)
COMMENT #232=(CLabel (CWidget 0 (0 0) 1 2 0 0 #154# 0 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT       "Arial") 2.53336 0.666667 0 "SD1" "@N" 1 0 0
COMMENT      -4.07034 8.30913 1.35613 0.843305 #231#)
COMMENT #233=(CScratch
COMMENT      (CWidget 0 (3 #9# 0) 1 2 0 0 #173# 56994272
COMMENT       100) (LOGPEN 0 14 32768) 8 1 3.12177 1.9
COMMENT      0.082322 1)
COMMENT #234=(CLabel (CWidget 0 (0 0) 1 2 0 0 #154# 0 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT       "Arial") 2.53336 0.666667 0 "SA20" "@N" 1 0
COMMENT      0 0.924503 10.8425 1.7208 0.843305 #233#)
COMMENT #235=(CScratch
COMMENT      (CWidget 0 (3 #10# 0) 1 2 0 0 #173# 0 100)
COMMENT      (LOGPEN 0 14 32768) 8 1 3.01276 1.9 0.082322
COMMENT      1)
COMMENT #236=(CLabel
COMMENT      (CWidget 0 (0 0) 1 2 0 0 #154# 55931092 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT       "Arial") 2.53336 0.666667 0 "SA5" "@N" 1 0 0
COMMENT      1.24622 8.30913 1.35613 0.843305 #235#)
COMMENT #237=(CScratch
COMMENT      (CWidget 0 (3 #11# 0) 1 2 0 0 #173# 0 100)
COMMENT      (LOGPEN 0 14 32768) 8 1 2.79476 1.9 0.082322
COMMENT      1)
COMMENT #238=(CLabel
COMMENT      (CWidget 0 (0 0) 1 2 0 0 #154# 1953789027 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT       "Arial") 2.53336 0.666667 0 "SA2" "@N" 1 0 0
COMMENT      2.24864 7.04245 1.35613 0.843305 #237#)
COMMENT #239=(CScratch
COMMENT      (CWidget 0 (3 #12# 0) 1 2 0 0 #173# 0 100)
COMMENT      (LOGPEN 0 14 32768) 8 1 2.77873 1.9 0.082322
COMMENT      1)
COMMENT #240=(CLabel
COMMENT      (CWidget 0 (0 0) 1 2 0 0 #154# 1667719540 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT       "Arial") 2.53336 0.666667 0 "SA10" "@N" 1 0
COMMENT      0 2.50184 4.50909 1.7208 0.843305 #239#)
COMMENT #241=(CScratch
COMMENT      (CWidget 0 (3 #13# 0) 1 2 0 0 #173# 3086 100)
COMMENT      (LOGPEN 0 14 32768) 8 1 2.76911 1.9 0.082322
COMMENT      1)
COMMENT #242=(CLabel
COMMENT      (CWidget 0 (0 0) 1 2 0 0 #154# 1667462243 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT       "Arial") 2.53336 0.666667 0 "SA11" "@N" 1 0
COMMENT      0 2.54606 3.24241 1.7208 0.843305 #241#)
COMMENT #243=(CScratch
COMMENT      (CWidget 0 (3 #15# 0) 1 2 0 0 #173# 651165415
COMMENT       100) (LOGPEN 0 14 32768) 8 1 1.05555 1.9
COMMENT      0.082322 1)
COMMENT #244=(CLabel
COMMENT      (CWidget 0 (0 0) 1 2 0 0 #154# 55927508 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT       "Arial") 2.53336 0.666667 0 "SD2" "@N" 1 0 0
COMMENT      10.2459 4.50909 1.35613 0.843305 #243#)
COMMENT #245=(CScratch
COMMENT      (CWidget 0 (3 #20# 0) 1 2 0 0 #173# 651165687
COMMENT       100) (LOGPEN 0 14 32768) 8 1 3.39106 1.9
COMMENT      0.082322 1)
COMMENT #246=(CLabel
COMMENT      (CWidget 0 (0 0) 1 2 0 0 #154# 55929524 100)
COMMENT      (LOGPEN 0 0 0) 1
COMMENT      (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
```

Appendix - SEQ ID NO1

```
COMMENT                                  "Arial") 2.53336 0.666667 0 "SA7" "@N" 1 0 0
COMMENT                                 -2.62145 12.1092 1.35613 0.843305 #245#)
COMMENT                            #247=(CScratch
COMMENT                                 (CWidget 0 (3 #21# 0) 1 2 0 0 #173# 651165311
COMMENT                                  100) (LOGPEN 0 14 32768) 8 1 3.13218 1.9
COMMENT                                 0.082322 1)
COMMENT                            #248=(CLabel (CWidget 0 (0 0) 1 2 0 0 #154# 0 100)
COMMENT                                 (LOGPEN 0 0 0) 1
COMMENT                                 (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                  "Arial") 2.53336 0.666667 0 "SD3" "@N" 1 0 0
COMMENT                                 0.697107 13.3758 1.35613 0.843305 #247#)
COMMENT                            #249=(CScratch
COMMENT                                 (CWidget 0 (3 #22# 0) 1 2 0 0 #173# 3146 100)
COMMENT                                 (LOGPEN 0 14 32768) 8 1 3.12497 1.9 0.082322
COMMENT                                 1)
COMMENT                            #250=(CLabel
COMMENT                                 (CWidget 0 (0 0) 1 2 0 0 #154# 1734435175 100)
COMMENT                                 (LOGPEN 0 0 0) 1
COMMENT                                 (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                  "Arial") 2.53336 0.666667 0 "SA21" "@N" 1 0
COMMENT                                 0 0.909762 12.1092 1.7208 0.843305 #249#)
COMMENT                            #251=(CScratch
COMMENT                                 (CWidget 0 (3 #23# 0) 1 2 0 0 #173# 0 100)
COMMENT                                 (LOGPEN 0 14 32768) 8 1 3.11054 1.9 0.082322
COMMENT                                 1)
COMMENT                            #252=(CLabel
COMMENT                                 (CWidget 0 (0 0) 1 2 0 0 #154# 1952671585 100)
COMMENT                                 (LOGPEN 0 0 0) 1
COMMENT                                 (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                  "Arial") 2.53336 0.666667 0 "SA6" "@N" 1 0 0
COMMENT                                 0.796611 9.57581 1.35613 0.843305 #251#)
COMMENT                            #253=(CScratch
COMMENT                                 (CWidget 0 (3 #24# 0) 1 2 0 0 #173# 3206 100)
COMMENT                                 (LOGPEN 0 14 32768) 8 1 2.78995 1.9 0.082322
COMMENT                                 1)
COMMENT                            #254=(CLabel
COMMENT                                 (CWidget 0 (0 0) 1 2 0 0 #154# 1668573027 100)
COMMENT                                 (LOGPEN 0 0 0) 1
COMMENT                                 (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                  "Arial") 2.53336 0.666667 0 "SD6" "@N" 1 0 0
COMMENT                                 2.27075 5.77577 1.35613 0.843305 #253#))
COMMENT                            (CObjectList))
COMMENT                       #255=(CGroupWidget
COMMENT                            (CWidget 85 (7 85 0) 1 2 0 0 Nil -152 100)
COMMENT                            (CObjectList
COMMENT                             #256=(CWideArrow
COMMENT                                   (CWideLine
COMMENT                                    (CWidget 0 (3 #1# 0) 1 2 0 0 #158#
COMMENT                                     1397443669 100) (LOGPEN 0 0 6723840) 1
COMMENT                                    (LOGBRUSH 0 10079334 0) 0.835356 1.07639
COMMENT                                    1.55487 1 0.082322) 0.8 1.8 0)
COMMENT                             #257=(CLabel (CWidget 0 (0 0) 1 2 0 0 #154# 0 100)
COMMENT                                   (LOGPEN 0 0 0) 1
COMMENT                                   (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                    "Arial") 2.53336 0.666667 0 "Wpre" "@N" 1 0
COMMENT                                   0 9.22945 7.04245 1.67521 0.843305 #256#))
COMMENT                            (CObjectList))) (CObjectList))
COMMENT                  #258=(CGroupWidget (CWidget 14 (16 0) 1 2 0 0 Nil -349 100)
COMMENT                       (CObjectList) (CObjectList))
COMMENT                  #259=(CGroupWidget (CWidget 11 (0 0) 1 2 0 0 Nil -300 100)
COMMENT                       (CObjectList) (CObjectList))
COMMENT                  #260=(CGroupWidget (CWidget 12 (0 0) 1 2 0 0 Nil -201 100)
COMMENT                       (CObjectList) (CObjectList))) (CObjectList)))
COMMENT             (CSeqView 10 10 (CObjectList)
COMMENT              (CObList #261=(AnlzFontColorItem 3894 3918 37173247)
COMMENT               #262=(AnlzFontColorItem 3595 3596 37173247)
COMMENT               #263=(AnlzFontColorItem 6521 6535 33554432)) 1
COMMENT              (CObList #264=(AnlzFontBackColorItem 3917 3918 33606707)
COMMENT               #265=(AnlzFontBackColorItem 3596 3605 33606707)
COMMENT               #266=(AnlzFontBackColorItem 3928 3929 40304639)
COMMENT               #267=(AnlzFontBackColorItem 3932 3933 40304639)
COMMENT               #268=(AnlzFontBackColorItem 3946 3947 40304639)
COMMENT               #269=(AnlzFontBackColorItem 4340 4341 40304639)
COMMENT               #270=(AnlzFontBackColorItem 4360 4361 40304639)
COMMENT               #271=(AnlzFontBackColorItem 4372 4373 40304639)
COMMENT               #272=(AnlzFontBackColorItem 4449 4450 40304639)
COMMENT               #273=(AnlzFontBackColorItem 6499 6500 40304639)
COMMENT               #274=(AnlzFontBackColorItem 6519 6520 40304639)
COMMENT               #275=(AnlzFontBackColorItem 3070 3071 40304639)
COMMENT               #276=(AnlzFontBackColorItem 3067 3068 40304639)
COMMENT               #277=(AnlzFontBackColorItem 2973 2974 40304639))) (CObList) 16777215
COMMENT             (CStringList) 56747520 622264320 (CObList)))
FEATURES             Location/Qualifiers
     CDS             5159..5878
                     /vntifkey="4"
                     /label=GFP
```

Appendix - SEQ ID NO1

misc_difference 5888..6484
/vntifkey="85"
/label=Wpre
promoter 4622..5137
/vntifkey="29"
/label=hPGK
LTR 6568..6801
/vntifkey="19"
/label=dR3RU5
LTR 2891..3072
/vntifkey="19"
/label=RU5
splicing_signal 3127..3128
/vntifkey="38"
/label=SA1
splicing_signal 3431..3432
/vntifkey="38"
/label=SA9
splicing_signal 3557..3558
/vntifkey="38"
/label=SD2
splicing_signal 3178..3179
/vntifkey="38"
/label=SD1
splicing_signal complement(3597..3598)
/vntifkey="38"
/label=SA7
splicing_signal 3920..3921
/vntifkey="38"
/label=SD3
splicing_signal complement(3929..3930)
/vntifkey="38"
/label=SA21
splicing_signal complement(3933..3934)
/vntifkey="38"
/label=SA20
splicing_signal 3947..3948
/vntifkey="38"
/label=SA6
splicing_signal 4069..4070
/vntifkey="38"
/label=SA5
splicing_signal 4341..4342
/vntifkey="38"
/label=SA2
splicing_signal complement(4347..4348)
/vntifkey="38"
/label=SD6
splicing_signal 4361..4362
/vntifkey="38"
/label=SA10
splicing_signal 4373..4374
/vntifkey="38"
/label=SA11
silencer 4450..4451
/vntifkey="37"
/label=SD4
splicing_signal complement(6511..6512)
/vntifkey="38"
/label=SD2
misc_feature complement (6500..6501)
/vntifkey="21"
/label=SD14
misc_feature 3068..3069
/vntifkey="21"
/label=SA4
misc_feature 3071..3072
/vntifkey="21"
/label=SA3
misc_feature complement (2974..2975)
/vntifkey="21"
/label=SD5
misc_feature complement (6520..6521)
/vntifkey="21"
/label=SD15
misc_feature 3040..3041
/vntifkey="21"
/label=Cryptic SA1
misc_feature 3068..3069
/vntifkey="21"
/label=Cryptic SA2
misc_feature 3071..3072
/vntifkey="21"
/label=Cryptic SA3
misc_feature 3077..3078
/vntifkey="21"

Appendix - SEQ ID NO1

```
                     /label=Cryptic SA4
     misc_feature    3089..3090
                     /vntifkey="21"
                     /label=Cryptic SA5
     misc_feature    3108..3109
                     /vntifkey="21"
                     /label=Cryptic SA6
     misc_feature    3127..3128
                     /vntifkey="21"
                     /label=Cryptic SA7
     misc_feature    3130..3131
                     /vntifkey="21"
                     /label=Cryptic SA8

BASE COUNT     1956 a      1974 c      2023 g      1874 t
ORIGIN
        1 caggtggcac ttttcgggga aatgtgcgcg gaacccctat ttgtttattt ttctaaatac
       61 attcaaatat gtatccgctc atgagacaat aaccctgata aatgcttcaa taatattgaa
      121 aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt tttgcggcat
      181 tttgccttcc tgtttttgct cacccagaaa cgctggtgaa agtaaaagat gctgaagatc
      241 agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag atccttgaga
      301 gttttcgccc cgaagaacgt tttccaatga tgagcacttt taaagttctg ctatgtggcg
      361 cggtattatc ccgtattgac gccgggcaag agcaactcgg tcgccgcata cactattctc
      421 agaatgactt ggttgagtac tcaccagtca cagaaaagca tcttacggat ggcatgacag
      481 taagagaatt atgcagtgct gccataacca tgagtgataa cactgcggcc aacttacttc
      541 tgacaacgat cggaggaccg aaggagctaa ccgcttttt gcacaacatg ggggatcatg
      601 taactcgcct tgatcgttgg gaaccggagc tgaatgaagc cataccaaac gacgagcgtg
      661 acaccacgat gcctgtagca atggcaacaa cgttgcgcaa actattaact ggcgaactac
      721 ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa gttgcaggac
      781 cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct ggagccggtg
      841 agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc tcccgtatcg
      901 tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga cagatcgctg
      961 agataggtgc ctcactgatt aagcattggt aactgtcaga ccaagtttac tcatatatac
     1021 tttagattga tttaaaactt catttttaat ttaaaaggat ctaggtgaag atccttttg
     1081 ataatctcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg tcagaccccg
     1141 tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc
     1201 aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag ctaccaactc
     1261 tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtc cttctagtgt
     1321 agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac ctcgctctgc
     1381 taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc gggttggact
     1441 caagacgata gttaccggat aaggcgcagc ggtcgggctg aacgggggt tcgtgcacac
     1501 agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt gagctatgag
     1561 aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc ggcagggtcg
     1621 gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg
     1681 tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca ggggggcgga
     1741 gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt tgctggcctt
     1801 ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt attaccgcct
     1861 ttgagtgagc tgataccgct cgccgcagcc gaacgaccga gcgcagcgag tcagtgagcg
     1921 aggaagcgga agagcgccca atacgcaaac cgcctctccc cgcgcgttgg ccgattcatt
     1981 aatgcagctg gcacgacagg tttcccgact ggaaagcggg cagtgagcgc aacgcaatta
     2041 atgtgagtta gctcactcat taggcacccc aggctttaca ctttatgctt ccggctcgta
     2101 tgttgtgtgg aattgtgagc ggataacaat ttcacacagg aaacagctat gaccatgatt
     2161 acgccaagcg cgcaattaac cctcactaaa gggaacaaaa gctggagctg caagcttggc
     2221 cattgcatac gttgtatcca tatcataata tgtacattta tattggctca tgtccaacat
     2281 taccgccatg ttgacattga ttattgacta gttattaata gtaatcaatt acggggtcat
     2341 tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaat ggcccgcctg
     2401 gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatgtt cccatagtaa
     2461 cgccaatagg gactttccat tgacgtcaat gggtggagta tttacggtaa actgcccact
     2521 tggcagtaca tcaagtgtat catatgccaa gtacgccccc tattgacgtc aatgacggta
     2581 aatggcccgc ctggcattat gcccagtaca tgaccttatg ggactttcct acttggcagt
     2641 acatctacgt attagtcatc gctattacca tggtgatgcg gttttggcag tacatcaatg
     2701 ggcgtggata gcggtttgac tcacggggat ttccaagtct ccaccccatt gacgtcaatg
     2761 ggagtttgtt ttggcaccaa aatcaacggg actttccaaa atgtcgtaac aactccgccc
     2821 cattgacgca aatgggcggt aggcgtgtac ggtgggaggt ctatataagc agagctcgtt
     2881 tagtgaaccg gggtctctct ggttagacca gatctgagcc tgggagctct ctggctaact
     2941 agggaaccca ctgcttaagc ctcaataaag cttgccttga gtgcttcaag tagtgtgtgc
     3001 ccgtctgttg tgtgactctg gtaactagag atccctcaga ccctttttagt cagtgtggaa
     3061 aatctctagc agtggcgccc gaacagggac ctgaaagcga aagggaaacc agagctctct
     3121 cgacgcagga ctcggcttgc tgaagcgcgc acggcaagag gcgaggggcg gcgactggtg
     3181 agtacgccaa aaattttgac tagcggaggc tagaaggaga gagatgggtg cgagagcgtc
     3241 agtattaagc gggggagaat tagatcgcga tgggaaaaaa ttcggttaag gccaggggga
     3301 aagaaaaaat ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca
     3361 gttaatcctg gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa
     3421 ccatcccttc agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc
     3481 tattgtgtgc atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag
     3541 gaagagcaaa acaaaagtaa gaccaccgca cagcaagcgg ccgctgatct tcagacctgg
     3601 aggaggagat atgagggaca attggagaag tgaattatat aaatataaag tagtaaaaat
     3661 tgaaccatta ggagtagcac ccaccaaggc aaagagaaga gtggtgcaga gagaaaaaag
     3721 agcagtggga ataggagctt tgttccttgg gttcttggga gcagcaggaa gcactatggg
     3781 cgcagcctca atgacgctga cggtacaggc cagacaatta ttgtctggta tagtgcagca
     3841 gcagaacaat ttgctgaggg ctattgaggc gcaacagcat ctgttgcaac tcacagtctg
     3901 gggcatcaag cagctccagg caagaatcct ggctgtggaa agatacctaa aggatcaaca
     3961 gctcctgggg atttggggtt gctctggaaa actcatttgc accactgctg tgccttggaa
     4021 tgctagttgg agtaataaat ctctggaaca gatttggaat cacacgacct ggatggagtg
     4081 ggacagagaa attaacaatt acacaagctt aatacactcc ttaattgaag aatcgcaaaa
     4141 ccagcaagaa aagaatgaac aagaattatt ggaattagat aaatgggcaa gtttgtggaa
```

Appendix - SEQ ID NO1

```
4201 ttggtttaac ataacaaatt ggctgtggta tataaaatta ttcataatga tagtaggagg
4261 cttggtaggt ttaagaatag tttttgctgt actttctata gtgaatagag ttaggcaggg
4321 atattcacca ttatcgtttc agacccacct cccaaccccg aggggacccg acaggcccga
4381 aggaatagaa gaagaaggtg gagagagaga cagagacaga tccattcgat tagtgaacgg
4441 atctcgacgg tatcggttaa ctttaaaag aaaaggggg attgggggt acagtgcagg
4501 ggaaagaata gtagacataa tagcaacaga catacaaact aaagaattac aaaaacaaat
4561 tacaaaaatt caaaatttta tcgatcacga gactagcctc gagaagcttg atatcgaatt
4621 cccacggggt tggggttgcg ccttttccaa ggcagccctg ggtttgcgca gggacgcggc
4681 tgctctgggc gtggttccgg gaaacgcagc ggcgccgacc ctgggtctcg cacattcttc
4741 acgtccgttc gcagcgtcac ccggatcttc gccgctaccc ttgtgggccc cccggcgacg
4801 cttcctgctc cgcccctaag tcgggaaggt tccttgcggt tcgcggcgtg ccggacgtga
4861 caaacggaag ccgcacgtct cactagtacc ctcgcagacg gacagcgcca gggagcaatg
4921 gcagcgcgcc gaccgcgatg ggctgtggcc aatagcggct gctcagcggg gcgcgccgag
4981 agcagcggcc gggaaggggc ggtgcgggag gcggggtgtg gggcggtagt gtgggccctg
5041 ttcctgcccg cgcggtgttc cgcattctgc aagcctccgg agcgcacgtc ggcagtcggc
5101 tccctcgttg accgaatcac cgacctctct ccccaggggg atccaccggt cgccaccatg
5161 gtgagcaagg gcgaggagct gttcaccggg gtggtgccca tcctggtcga gctggacggc
5221 gacgtaaacg gccacaagtt cagcgtgtcc ggcgagggcg agggcgatgc cacctacggc
5281 aagctgaccc tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg gcccacctc
5341 gtgaccaccc tgacctacgg cgtgcagtgc ttcagccgct accccgacca catgaagcag
5401 cacgacttct tcaagtccgc catgcccgaa ggctacgtcc aggagcgcac catcttcttc
5461 aaggacgacg gcaactacaa gacccgcgcc gaggtgaagt tcgagggcga caccctggtg
5521 aaccgcatcg agctgaaggg catcgacttc aaggaggacg gcaacatcct ggggcacaag
5581 ctggagtaca actacaacag ccacaacgtc tatatcatgg ccgacaagca gaagaacggc
5641 atcaaggtga acttcaagat ccgccacaac atcgaggacg gcagcgtgca gctcgccgac
5701 cactaccagc agaacacccc catcggcgac ggccccgtgc tgctgcccga caaccactac
5761 ctgagcaccc agtccgccct gagcaaagac cccaacgaga agcgcgatca catggtcctg
5821 ctggagttcg tgaccgccgc cgggatcact ctcggcatgg acgagctgta caagtaaagc
5881 ggccgcgtcg acaatcaacc tctggattac aaaatttgtg aaagattgac tggtattctt
5941 aactatgttg ctccttttac gctatgtgga tacgctgctt taatgccttt gtatcatgct
6001 attgcttccc gtatggcttt cattttctcc tccttgtata aatcctggtt gctgtctctt
6061 tatgaggagt tgtggcccgt tgtcaggcaa cgtggcgtgg tgtgcactgt gtttgctgac
6121 gcaaccccca ctggttgggg cattgccacc acctgtcagc tcctttccgg gactttcgct
6181 ttccccctcc ctattgccac ggcggaactc atcgccgcct gccttgcccg ctgctggaca
6241 ggggctcggc tgttgggcac tgacaattcc gtggtgttgt cggggaagct gacgtccttt
6301 ccatggctgc tcgcctgtgt tgccacctgg attctgcgcg ggacgtcctt ctgctacgtc
6361 ccttcggccc tcaatccagc ggaccttcct tcccgcggcc tgctgccggc tctgcggcct
6421 cttccgcgtc ttcgccttcg ccctcagacg agtcggatct ccctttgggc cgcctccccg
6481 cctggaattc gagctcggta cctttaagac caatgactta caaggcagct gtagatctta
6541 gccacttttt aaaagaaaag gggggactgg aagggctaat tcactcccaa cgaagacaag
6601 atctgctttt tgcttgtact gggtctctct ggttagacca gatctgagcc tgggagctct
6661 ctggctaact agggaaccca ctgcttaagc ctcaataaag cttgccttga gtgcttcaag
6721 tagtgtgtgc ccgtctgttg tgtgactctg gtaactagag atccctcaga cccttttagt
6781 cagtgtggaa aatctctagc agtagtagtt catgtcatct tattattcag tatttataac
6841 ttgcaaagaa atgaatatca gagagtgaga ggaacttgtt tattgcagct tataatggtt
6901 acaaataaag caatagcatc acaaatttca caaataaagc attttttca ctgcattcta
6961 gttgtggttt gtccaaactc atcaatgtat cttatcatgt ctggctctag ctatcccgcc
7021 cctaactccg cccagttccg cccattctcc gccccatggc tgactaattt ttttattta
7081 tgcagaggcc gaggccgcct cggcctctga gctattccag aagtagtgag gaggcttttt
7141 tggaggccta ggcttttgcg tcgagacgta cccaattcgc cctatagtga gtcgtattac
7201 gcgcgctcac tggccgtcgt tttacaacgt cgtgactggg aaaaccctgg cgttacccaa
7261 cttaatcgcc ttgcagcaca tccccctttc gccagctggc gtaatagcga agaggcccgc
7321 accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatggcgcga cgcgccctgt
7381 agcggcgcat taagcgcggc gggtgtggtg gttacgcgca gcgtgaccgc tacacttgcc
7441 agcgccctag cgcccgctcc tttcgctttc ttcccttcct ttctcgccac gttcgccggc
7501 tttccccgtc aagctctaaa tcgggggctc cctttagggt tccgatttag tgctttacgg
7561 cacctcgacc ccaaaaaact tgattagggt gatggttcac gtagtgggcc atcgccctga
7621 tagacggttt ttcgcccttt gacgttggag tccacgttct ttaatagtgg actcttgttc
7681 caaactggaa caacactcaa ccctatctcg gtctattctt ttgatttata agggattttg
7741 ccgatttcgg cctattggtt aaaaaatgag ctgatttaac aaaaatttaa cgcgaatttt
7801 aacaaaatat taacgtttac aatttcc
```

//

Appendix - SEQ ID No: 2

```
LOCUS       GLOBE\Splicing\s        9771 bp    DNA     circular     13-APR-2011
DEFINITION  Complementary copy of MA primm.
SOURCE
  ORGANISM
COMMENT     http://www.informaxinc.com/
COMMENT     This file is created by Vector NTI
            http://www.invitrogen.com/
COMMENT     ORIGDB|GenBank
COMMENT     VNTDATE|587327355|
COMMENT     VNTDBDATE|587332331|
COMMENT     LSOWNER|
COMMENT     VNTNAME|GLOBE Splicing sites|
COMMENT     VNTAUTHORNAME|Ferrari Giuliana|
COMMENT     Vector_NTI_Display_Data_(Do_Not_Edit!)
COMMENT     (SXF
COMMENT      (CGexDoc "GLOBE Splicing sites" 0 9771
COMMENT       (CDBMol 0 0 1 1 1 0 0 0 0 "" "" 0 0 0 0 (CObList) (CObList) (CObList)
COMMENT        (CObList) -1 "")
COMMENT       (CDocSetData 1 1 0 0 0 1 "MAIN" 1 1 1 1 0 0 1 1 0 1 10 10 4294967295 50 0
COMMENT        1 0 (CHomObj 0 0 0 3 75) (CWordArray) (CWordArray) (CStringList)
COMMENT        (CStringList "atg" "gtg") (CStringList "taa" "tga" "tag") (CObList) 1
COMMENT        "{(0,1),2}" 0 0 "" 0 4294967295 0 1 0 0 0 0 0 "MAIN" 0 0 30 0
COMMENT        (CProteinMotifSearchObject 70 20 1 1 1 1 0 0 1 0 0 0 0 0))
COMMENT       (CMolPar 0 0 0 0 0 1 9771 0 0 0 0 0 0 0 0 0) (CStringList) (CStringList)
COMMENT       (CObList) (COAPar 25 250 50 0 6 4 3 7) (COAPar 25 250 50 0 6 4 3 7)
COMMENT       (COAPar 25 250 50 0 6 4 3 7) (CObList)
COMMENT       (CObList
COMMENT        #0=(CFSignal (CObList) "RU5" 21 0 0 2447 2627 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "2447..2627")
COMMENT        #1=(CFSignal (CObList) "sin18 RU5" 21 0 0 8509 8743 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "8509..8743")
COMMENT        #2=(CFSignal (CObList) "GAG" 21 0 0 2783 3147 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "2783..3147")
COMMENT        #3=(CFSignal (CObList) "HS3" 22 8 0 7199 8400 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "7199..8400")
COMMENT        #4=(CFSignal (CObList) "HS2" 85 8 0 5758 7192 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "5758..7192")
COMMENT        #5=(CFSignal (CObList) "Exon 2" 61 0 1 4960 5182 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "complement(4960..5182)")
COMMENT        #6=(CFSignal (CObList) "Exon 3" 61 0 1 4574 4702 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "complement(4574..4702)")
COMMENT        #7=(CFSignal (CObList) "Exon 1" 61 0 1 5313 5408 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "complement(5313..5408)")
COMMENT        #8=(CFSignal (CObList) "HIV-SA" 21 0 0 3887 3898 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "3887..3898")
COMMENT        #9=(CFSignal (CObList) "RRE" 20 0 0 3293 3534 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "3293..3534")
COMMENT        #10=(CFSignal (CObList) "\247-globin polyA" 25 0 0 4461 4466 0
COMMENT             (CStringList) (CStringList) 1 1 1 1 "4461..4466")
COMMENT        #11=(CFSignal (CObList) "3' UTR" 50 0 0 4442 4573 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "4442..4573")
COMMENT        #12=(CFSignal (CObList) "RSV" 29 0 0 2216 2447 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "2216..2447")
COMMENT        #13=(CFSignal (CObList) "Amp r" 4 0 0 132 992 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "132..992")
COMMENT        #14=(CFSignal (CObList) "pUC19" 21 0 0 1 1999 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "1..1999")
COMMENT        #15=(CFSignal (CObList) "SA_A" 38 0 0 8106 8107 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "8106..8107")
COMMENT        #16=(CFSignal (CObList) "SA_B" 38 0 0 8067 8068 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "8067..8068")
COMMENT        #17=(CFSignal (CObList) "SA_C" 38 0 0 7474 7475 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "7474..7475")
COMMENT        #18=(CFSignal (CObList) "SA_D" 38 0 0 5423 5424 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "5423..5424")
COMMENT        #19=(CFSignal (CObList) "SD_A" 38 0 0 7912 7913 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "7912..7913")
COMMENT        #20=(CFSignal (CObList) "SD_B" 38 0 0 7837 7838 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "7837..7838")
COMMENT        #21=(CFSignal (CObList) "SD_C" 38 0 0 7821 7822 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "7821..7822")
COMMENT        #22=(CFSignal (CObList) "SD_D" 38 0 0 7797 7798 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "7797..7798")
COMMENT        #23=(CFSignal (CObList) "SD_E" 38 0 0 7367 7368 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "7367..7368")
COMMENT        #24=(CFSignal (CObList) "SD_F" 38 0 0 7363 7364 0 (CStringList)
COMMENT             (CStringList) 1 1 1 1 "7363..7364")) (CObList) (CObList) (CObList)
COMMENT       (CObList) (CObList) (CObList)
COMMENT       (CTextView 0
COMMENT        #25=(CGroupPar (CParagraph 0 (0 0) 1 2 0 0 180)
COMMENT             (CObjectList
COMMENT              #26=(CRefLinePar
COMMENT                   (CLinePar (CParagraph 0 (0 0) 0 2 0 1 233)
COMMENT                    "GLOBE Splicing sites" 2) 5 "" 0 4)
COMMENT              #27=(CFolderPar
```

```
COMMENT                    (CGroupPar (CParagraph 1 (0 0) 1 1 0 0 178)
COMMENT                     (CObjectList
COMMENT                      #28=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "DNA 'GLOBE Splicing sites'" 1)
COMMENT                      #29=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Complementary copy of MA primm." 1)
COMMENT                      #30=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Foreign object. Author: Ferrari Giuliana. Original author:
Ferrari Giuliana"
COMMENT                           1)
COMMENT                      #31=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Created: 04/13/11 06:29PM" 1)
COMMENT                      #32=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "Last Modified: 04/13/11 07:52PM" 1)
COMMENT                      #33=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "length: 9771 bp" 1)
COMMENT                      #34=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "storage type: Basic" 1)
COMMENT                      #35=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                           "form: Circular" 1))) "General Description")
COMMENT                #36=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 2 (0 0) 1 1 0 0 178)
COMMENT                      (CObjectList
COMMENT                       #37=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Original Source Database: GenBank" 1)
COMMENT                       #38=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Modification Date in the Original DB: 13-APR-2011" 1)))
COMMENT                     "Standard Fields")
COMMENT                #39=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 4 (0 0) 1 1 0 0 178)
COMMENT                      (CObjectList
COMMENT                       #40=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Ferrari Giuliana" 1))) "Author")
COMMENT                #41=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 5 (0 0) 1 1 0 0 178)
COMMENT                      (CObjectList
COMMENT                       #42=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Ferrari Giuliana" 1))) "Original Author")
COMMENT                #43=(CRefLinePar
COMMENT                     (CLinePar (CParagraph 0 (0 0) 0 2 0 0 233) "Comments" 2) 1 ""
COMMENT                     0 0)
COMMENT                #44=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 8 (0 0) 1 2 0 0 178) (CObjectList))
COMMENT                     "Annotations")
COMMENT                #45=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 12 (6 0) 1 1 0 0 178)
COMMENT                      (CObjectList
COMMENT                       #46=(CFolderPar
COMMENT                            (CGroupPar (CParagraph 4 (7 4 0) 1 1 1 0 178)
COMMENT                             (CObjectList
COMMENT                              #47=(CFolderPar
COMMENT                                   (CGroupPar
COMMENT                                    (CParagraph 132 (3 #13# 0) 1 2 2 0 327)
COMMENT                                    (CObjectList
COMMENT                                     #48=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Start: 132   End: 992 " 1)
COMMENT                                     #49=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Original Location Description:" 1)
COMMENT                                     #50=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "   132..992" 1))) "Amp r")))
COMMENT                            "CDS (1 total)")
COMMENT                       #51=(CFolderPar
COMMENT                            (CGroupPar (CParagraph 20 (7 20 0) 1 1 1 0 178)
COMMENT                             (CObjectList
COMMENT                              #52=(CFolderPar
COMMENT                                   (CGroupPar
COMMENT                                    (CParagraph 3293 (3 #9# 0) 1 2 2 0 194)
COMMENT                                    (CObjectList
COMMENT                                     #53=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Start: 3293  End: 3534" 1)
COMMENT                                     #54=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Original Location Description:" 1)
COMMENT                                     #55=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "   3293..3534" 1))) "RRE")))
COMMENT                            "Misc. Binding Site (1 total)")
COMMENT                       #56=(CFolderPar
COMMENT                            (CGroupPar (CParagraph 21 (7 21 0) 1 1 1 0 178)
COMMENT                             (CObjectList
COMMENT                              #57=(CFolderPar
COMMENT                                   (CGroupPar
COMMENT                                    (CParagraph 1 (3 #14# 0) 1 2 2 0 194)
```

```
COMMENT                         (CObjectList
COMMENT                          #58=(CLinePar
COMMENT                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                               "Start: 1     End: 1999" 1)
COMMENT                          #59=(CLinePar
COMMENT                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                               "Original Location Description:" 1)
COMMENT                          #60=(CLinePar
COMMENT                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                               "   1..1999" 1))) "pUC19")
COMMENT                    #61=(CFolderPar
COMMENT                         (CGroupPar
COMMENT                          (CParagraph 2447 (3 #0# 0) 1 2 2 0 194)
COMMENT                          (CObjectList
COMMENT                           #62=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "Start: 2447  End: 2627" 1)
COMMENT                           #63=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "Original Location Description:" 1)
COMMENT                           #64=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "   2447..2627" 1))) "RU5")
COMMENT                    #65=(CFolderPar
COMMENT                         (CGroupPar
COMMENT                          (CParagraph 2783 (3 #2# 0) 1 2 2 0 194)
COMMENT                          (CObjectList
COMMENT                           #66=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "Start: 2783  End: 3147" 1)
COMMENT                           #67=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "Original Location Description:" 1)
COMMENT                           #68=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "   2783..3147" 1))) "GAG")
COMMENT                    #69=(CFolderPar
COMMENT                         (CGroupPar
COMMENT                          (CParagraph 3887 (3 #8# 0) 1 2 2 0 194)
COMMENT                          (CObjectList
COMMENT                           #70=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "Start: 3887  End: 3898" 1)
COMMENT                           #71=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "Original Location Description:" 1)
COMMENT                           #72=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "   3887..3898" 1))) "HIV-SA")
COMMENT                    #73=(CFolderPar
COMMENT                         (CGroupPar
COMMENT                          (CParagraph 8509 (3 #1# 0) 1 2 2 0 194)
COMMENT                          (CObjectList
COMMENT                           #74=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "Start: 8509  End: 8743" 1)
COMMENT                           #75=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "Original Location Description:" 1)
COMMENT                           #76=(CLinePar
COMMENT                                (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                "   8509..8743" 1))) "sin18 RU5")))
COMMENT                    "Misc. Feature (5 total)")
COMMENT               #77=(CFolderPar
COMMENT                    (CGroupPar (CParagraph 22 (7 22 0) 1 1 1 0 178)
COMMENT                     (CObjectList
COMMENT                      #78=(CFolderPar
COMMENT                           (CGroupPar
COMMENT                            (CParagraph 7199 (3 #3# 0) 1 2 2 0 194)
COMMENT                            (CObjectList
COMMENT                             #79=(CLinePar
COMMENT                                  (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                  "Start: 7199  End: 8400" 1)
COMMENT                             #80=(CLinePar
COMMENT                                  (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                  "Original Location Description:" 1)
COMMENT                             #81=(CLinePar
COMMENT                                  (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                  "   7199..8400" 1))) "HS3")))
COMMENT                    "Misc. Marker (1 total)")
COMMENT               #82=(CFolderPar
COMMENT                    (CGroupPar (CParagraph 25 (7 25 0) 1 1 1 0 178)
COMMENT                     (CObjectList
COMMENT                      #83=(CFolderPar
COMMENT                           (CGroupPar
COMMENT                            (CParagraph 4461 (3 #10# 0) 1 2 2 0 194)
COMMENT                            (CObjectList
```

```
COMMENT                                  #84=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 4461  End: 4466" 1)
COMMENT                                  #85=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Original Location Description:" 1)
COMMENT                                  #86=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "   4461..4466" 1))) "\247-globin polyA")))
COMMENT                             "PolyA Signal (1 total)")
COMMENT                        #87=(CFolderPar
COMMENT                             (CGroupPar (CParagraph 29 (7 29 0) 1 1 1 0 178)
COMMENT                              (CObjectList
COMMENT                               #88=(CFolderPar
COMMENT                                    (CGroupPar
COMMENT                                     (CParagraph 2216 (3 #12# 0) 1 2 2 0 194)
COMMENT                                     (CObjectList
COMMENT                                      #89=(CLinePar
COMMENT                                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                           "Start: 2216  End: 2447" 1)
COMMENT                                      #90=(CLinePar
COMMENT                                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                           "Original Location Description:" 1)
COMMENT                                      #91=(CLinePar
COMMENT                                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                           "   2216..2447" 1))) "RSV")))
COMMENT                             "Promoter Eukaryotic (1 total)")
COMMENT                        #92=(CFolderPar
COMMENT                             (CGroupPar (CParagraph 38 (7 38 0) 1 1 1 0 178)
COMMENT                              (CObjectList
COMMENT                               #93=(CFolderPar
COMMENT                                    (CGroupPar
COMMENT                                     (CParagraph 5423 (3 #18# 0) 1 2 2 0 194)
COMMENT                                     (CObjectList
COMMENT                                      #94=(CLinePar
COMMENT                                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                           "Start: 5423  End: 5424" 1)
COMMENT                                      #95=(CLinePar
COMMENT                                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                           "Original Location Description:" 1)
COMMENT                                      #96=(CLinePar
COMMENT                                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                           "   5423..5424" 1))) "SA_D")
COMMENT                               #97=(CFolderPar
COMMENT                                    (CGroupPar
COMMENT                                     (CParagraph 7363 (3 #24# 0) 1 2 2 0 194)
COMMENT                                     (CObjectList
COMMENT                                      #98=(CLinePar
COMMENT                                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                           "Start: 7363  End: 7364" 1)
COMMENT                                      #99=(CLinePar
COMMENT                                           (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                           "Original Location Description:" 1)
COMMENT                                      #100=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "   7363..7364" 1))) "SD_F")
COMMENT                               #101=(CFolderPar
COMMENT                                    (CGroupPar
COMMENT                                     (CParagraph 7367 (3 #23# 0) 1 2 2 0 194)
COMMENT                                     (CObjectList
COMMENT                                      #102=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "Start: 7367  End: 7368" 1)
COMMENT                                      #103=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "Original Location Description:" 1)
COMMENT                                      #104=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "   7367..7368" 1))) "SD_E")
COMMENT                               #105=(CFolderPar
COMMENT                                    (CGroupPar
COMMENT                                     (CParagraph 7474 (3 #17# 0) 1 2 2 0 194)
COMMENT                                     (CObjectList
COMMENT                                      #106=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "Start: 7474  End: 7475" 1)
COMMENT                                      #107=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "Original Location Description:" 1)
COMMENT                                      #108=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "   7474..7475" 1))) "SA_C")
COMMENT                               #109=(CFolderPar
COMMENT                                    (CGroupPar
COMMENT                                     (CParagraph 7797 (3 #22# 0) 1 2 2 0 194)
COMMENT                                     (CObjectList
COMMENT                                      #110=(CLinePar
```

```
COMMENT                                    (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                    "Start: 7797  End: 7798" 1)
COMMENT                                  #111=(CLinePar
COMMENT                                    (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                    "Original Location Description:" 1)
COMMENT                                  #112=(CLinePar
COMMENT                                    (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                    "   7797..7798" 1))) "SD_D")
COMMENT                                #113=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 7821 (3 #21# 0) 1 2 2 0 194)
COMMENT                                   (CObjectList
COMMENT                                    #114=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Start: 7821  End: 7822" 1)
COMMENT                                    #115=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Original Location Description:" 1)
COMMENT                                    #116=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "   7821..7822" 1))) "SD_C")
COMMENT                                #117=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 7837 (3 #20# 0) 1 2 2 0 194)
COMMENT                                   (CObjectList
COMMENT                                    #118=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Start: 7837  End: 7838" 1)
COMMENT                                    #119=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Original Location Description:" 1)
COMMENT                                    #120=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "   7837..7838" 1))) "SD_B")
COMMENT                                #121=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 7912 (3 #19# 0) 1 2 2 0 194)
COMMENT                                   (CObjectList
COMMENT                                    #122=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Start: 7912  End: 7913" 1)
COMMENT                                    #123=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Original Location Description:" 1)
COMMENT                                    #124=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "   7912..7913" 1))) "SD_A")
COMMENT                                #125=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 8067 (3 #16# 0) 1 2 2 0 194)
COMMENT                                   (CObjectList
COMMENT                                    #126=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Start: 8067  End: 8068" 1)
COMMENT                                    #127=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Original Location Description:" 1)
COMMENT                                    #128=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "   8067..8068" 1))) "SA_B")
COMMENT                                #129=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 8106 (3 #15# 0) 1 2 2 0 194)
COMMENT                                   (CObjectList
COMMENT                                    #130=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Start: 8106  End: 8107" 1)
COMMENT                                    #131=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Original Location Description:" 1)
COMMENT                                    #132=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "   8106..8107" 1))) "SA_A")))
COMMENT                              "Splicing Signal (10 total)")
COMMENT                            #133=(CFolderPar
COMMENT                              (CGroupPar (CParagraph 50 (7 50 0) 1 1 1 0 178)
COMMENT                               (CObjectList
COMMENT                                #134=(CFolderPar
COMMENT                                  (CGroupPar
COMMENT                                   (CParagraph 4442 (3 #11# 0) 1 2 2 0 194)
COMMENT                                   (CObjectList
COMMENT                                    #135=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Start: 4442  End: 4573" 1)
COMMENT                                    #136=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Original Location Description:" 1)
```

```
COMMENT                                    #137=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "   4442..4573" 1))) "3' UTR")))
COMMENT                              "3' UTR (1 total)")
COMMENT                         #138=(CFolderPar
COMMENT                               (CGroupPar (CParagraph 61 (7 61 0) 1 1 1 0 178)
COMMENT                                (CObjectList
COMMENT                                 #139=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 4574 (3 #6# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #140=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 4574  End: 4702 (Complementary)"
COMMENT                                               1)
COMMENT                                         #141=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #142=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "   complement(4574..4702)" 1)))
COMMENT                                      "Exon 3")
COMMENT                                 #143=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 4960 (3 #5# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #144=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 4960  End: 5182 (Complementary)"
COMMENT                                               1)
COMMENT                                         #145=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #146=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "   complement(4960..5182)" 1)))
COMMENT                                      "Exon 2")
COMMENT                                 #147=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 5313 (3 #7# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #148=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 5313  End: 5408 (Complementary)"
COMMENT                                               1)
COMMENT                                         #149=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #150=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "   complement(5313..5408)" 1)))
COMMENT                                      "Exon 1"))) "Exon (3 total)")
COMMENT                         #151=(CFolderPar
COMMENT                               (CGroupPar (CParagraph 85 (7 85 0) 1 1 1 0 178)
COMMENT                                (CObjectList
COMMENT                                 #152=(CFolderPar
COMMENT                                       (CGroupPar
COMMENT                                        (CParagraph 5758 (3 #4# 0) 1 2 2 0 194)
COMMENT                                        (CObjectList
COMMENT                                         #153=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Start: 5758  End: 7192" 1)
COMMENT                                         #154=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "Original Location Description:" 1)
COMMENT                                         #155=(CLinePar
COMMENT                                               (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                               "   5758..7192" 1))) "HS2")))
COMMENT                              "Misc. Difference (1 total)"))) "Feature Map"))))
COMMENT                    (CGraphView
COMMENT                     (CWStyleSheet
COMMENT                      (CObjectList
COMMENT                       #156=(CWidgetStyle "RSite Label" 1 (LOGPEN 0 0 13408563) 1 0 1
COMMENT                             (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Georgia") 0.555556 0 1 5
COMMENT                             "@N (@S)" 0)
COMMENT                       #157=(CWidgetStyle "Signal Label" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                             (LOGFONT 0 0 0 0 700 0 0 0 0 3 2 1 34 "Arial") 0.666667 0 1 1
COMMENT                             "@N" 0)
COMMENT                       #158=(CWidgetStyle "Molecule Label 2" 0 0 1
COMMENT                             (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Georgia") 0.555556 0 1 16
COMMENT                             "@L bp" 0)
COMMENT                       #159=(CWidgetStyle "Molecule Label 1" 0 0 1
COMMENT                             (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Verdana") 0.833333 0 1 1
COMMENT                             "@N" 0)
COMMENT                       #160=(CWidgetStyle "Shape 3" 1 (LOGPEN 0 0 3355545) 1 1
COMMENT                             (LOGBRUSH 0 6724095 0) 0 0 1 (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT                       #161=(CWidgetStyle "Shape 1" 1 (LOGPEN 0 0 6723840) 1 1
```

```
COMMENT                 (LOGBRUSH 0 10079334 0) 0 0 1 (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT               #162=(CWidgetStyle "Axis" 1 (LOGPEN 0 0 10079436) 2 1
COMMENT                 (LOGBRUSH 0 13434879 0) 0 0 1 (LOGSHAPE 10 1 0 0 0))
COMMENT               #163=(CWidgetStyle "Line 2" 1 (LOGPEN 0 0 6723840) 8 0 0 0 1
COMMENT                 (LOGSHAPE 1 1.9 0 0 0))
COMMENT               #164=(CWidgetStyle "RSite" 1 (LOGPEN 0 0 10053171) 8 0 0 0 1
COMMENT                 (LOGSHAPE 1 1.9 0 0 0))
COMMENT               #165=(CWidgetStyle "Short Signal" 1 (LOGPEN 0 0 13395507) 10 0 0 0 1
COMMENT                 (LOGSHAPE 1 1.9 0 0 0))
COMMENT               #166=(CWidgetStyle "Uniq RSite Label" 1 (LOGPEN 0 0 153) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Georgia") 0.555556 128 1
COMMENT                 5 "@N (@S)" 0)
COMMENT               #167=(CWidgetStyle "Vanilla" 1 (LOGPEN 0 0 0) 1 1
COMMENT                 (LOGBRUSH 0 16777215 0) 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 7 48 2 18 "Times New Roman") 0.8 0
COMMENT                 1 2 "?" 0)
COMMENT               #168=(CWidgetStyle "Mark 1" 0 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 2 7 48 2 2 "Windings") 0.7 0 1 2 "?"
COMMENT                 0)
COMMENT               #169=(CWidgetStyle "Motif Label" 1 (LOGPEN 0 0 16744512) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.611111 8388608
COMMENT                 1 65535 "@N (@H)" 0)
COMMENT               #170=(CWidgetStyle "Fragment Label 2" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 49 "Courier New") 1.05 0 1 48
COMMENT                 "@F bp (molecule @L bp)" 0)
COMMENT               #171=(CWidgetStyle "Fragment Label 1" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.91 0 1 1
COMMENT                 "Fragment of @N" 0)
COMMENT               #172=(CWidgetStyle "Shape 4" 1 (LOGPEN 0 0 0) 1 1
COMMENT                 (LOGBRUSH 2 8388608 5) 0 0 0)
COMMENT               #173=(CWidgetStyle "Shape 2" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 0 128 0) 0
COMMENT                 0 0)
COMMENT               #174=(CWidgetStyle "Shape 0" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 0 0 0) 0 0
COMMENT                 0)
COMMENT               #175=(CWidgetStyle "ORF" 1 (LOGPEN 0 0 16384) 8 0 0 0 1
COMMENT                 (LOGSHAPE 7 0.2 3.41182 2.86186 0.609808))
COMMENT               #176=(CWidgetStyle "Line 4" 1 (LOGPEN 0 0 32768) 8 0 0 0 0)
COMMENT               #177=(CWidgetStyle "Line 3" 1 (LOGPEN 0 0 16711680) 8 0 0 0 0)
COMMENT               #178=(CWidgetStyle "Line 1" 1 (LOGPEN 0 0 16711680) 1 0 0 0 0)
COMMENT               #179=(CWidgetStyle "Short Promoter" 1 (LOGPEN 0 0 128) 6 0 0 0 0)
COMMENT               #180=(CWidgetStyle "Motif" 1 (LOGPEN 0 0 0) 1 0 0 0 0)
COMMENT               #181=(CWidgetStyle "Line 0" 1 (LOGPEN 0 0 0) 8 0 0 0 0)
COMMENT               #182=(CWidgetStyle "Void" 0 0 0 0 0)
COMMENT               #183=(CWidgetStyle "General Label" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman") 0.91 0
COMMENT                 1 3 "@T @N " 0)
COMMENT               #184=(CWidgetStyle "Position" 1 (LOGPEN 0 0 0) 1 0 0 0 0)
COMMENT               #185=(CWidgetStyle "Annotation" 0 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman") 0.91 0
COMMENT                 0 0)
COMMENT               #186=(CWidgetStyle "Position Label" 1 (LOGPEN 0 0 8388608) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 1 0 0 3 2 1 34 "Arial") 0.63 8388608 1 1
COMMENT                 "@N" 0)
COMMENT               #187=(CWidgetStyle "Range" 1 (LOGPEN 0 0 0) 1 1
COMMENT                 (LOGBRUSH 0 16777215 0) 0 0 0)
COMMENT               #188=(CWidgetStyle "Range Label" 1 (LOGPEN 0 0 8388608) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 1 0 0 3 2 1 34 "Arial") 0.63 8388608 1 1
COMMENT                 "@N" 0)
COMMENT               #189=(CWidgetStyle "ORF Label" 1 (LOGPEN 0 0 49216) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman")
COMMENT                 0.611111 0 1 65535 "@N" 0)
COMMENT               #190=(CWidgetStyle "CDS Label" 1 (LOGPEN 0 0 4227264) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.555556 255 1 1
COMMENT                 "@N" 0)
COMMENT               #191=(CWidgetStyle "Shape 5" 1 (LOGPEN 0 0 0) 3 1
COMMENT                 (LOGBRUSH 0 16777113 0) 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 7 48 2 50 "Arial") 0.9 0 0 1
COMMENT                 (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT               #192=(CWidgetStyle "CDS" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 2 39423 3) 0 0
COMMENT                 1 (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT               #193=(CWidgetStyle "Label 2" 1 (LOGPEN 0 0 4227264) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.944444 8388608
COMMENT                 1 1 "@N" 0)
COMMENT               #194=(CWidgetStyle "Label 3" 1 (LOGPEN 0 0 8421376) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 700 255 0 0 0 3 2 1 34 "Arial") 0.833333 255 1
COMMENT                 5 "@N (@S)" 0)
COMMENT               #195=(CWidgetStyle "Label 4" 1 (LOGPEN 0 0 8437824) 1 0 1
COMMENT                 (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.722222 0 1 5
COMMENT                 "@N (@S)" 0)) 0.164644 1.74233 0.164644 2.53336
COMMENT               (70 (CShapeMapEntry 0 "Unique RSite" 1 "Uniq RSite Label") 67
COMMENT                (CShapeMapEntry 0 "ORF" 0 "ORF Label")) 40.0378 40.0378 87.75 87.75
COMMENT               0.1 -9771) 1 1 1 1 1
COMMENT              (mapper: 15.4554 -9.71536 87.75 87.75 0.01 10 14 9771 9771 1 0 0)
COMMENT              #196=(CGroupWidget (CWidget 0 (0 0) 1 2 0 0 Nil 745434985 100)
COMMENT                   (CObjectList
COMMENT                    #197=(CGroupWidget (CWidget 1 (0 0) 1 2 0 0 Nil -1255 100)
COMMENT                         (CObjectList
```

```
COMMENT                    #198=(CAxis
COMMENT                          (CWideLine
COMMENT                           (CWidget 0 (0 0) 1 2 0 0 #162# 85831500 0)
COMMENT                           (LOGPEN 0 2 10079436) 2 (LOGBRUSH 0 13434879 0) 1
COMMENT                           6.27483 6.27283 1 0.0214037) 0.0642021)
COMMENT                    #199=(CLabel
COMMENT                          (CWidget 1001 (0 0) 1 2 0 0 #159# 61695508 100)
COMMENT                          (LOGPEN 0 0 0) 1
COMMENT                          (LOGFONT 92 35 0 0 400 0 0 0 0 3 2 1 34 "Verdana")
COMMENT                          2.53336 0.833333 0 "GLOBE Splicing sites" "@N" 1 0
COMMENT                          0.871165 -9.93732 8.38631 8.33048 1.04843 Nil)
COMMENT                    #200=(CLabel
COMMENT                          (CWidget 1002 (0 0) 1 2 0 0 #158# 83486200 100)
COMMENT                          (LOGPEN 0 0 0) 1
COMMENT                          (LOGFONT 61 23 0 0 400 0 0 0 0 3 2 1 18 "Georgia")
COMMENT                          2.53336 0.555556 0 "9771 bp" "@L bp" 16 0 -0.871165
COMMENT                          -3.63533 8.3504 2.07407 0.683761 Nil)) (CObjectList))
COMMENT               #201=(CGroupWidget (CWidget 10 (6 0) 1 2 0 0 Nil -1144 100)
COMMENT                     (CObjectList
COMMENT                      #202=(CGroupWidget
COMMENT                          (CWidget 4 (7 4 0) 1 2 0 0 Nil -922 100)
COMMENT                          (CObjectList
COMMENT                           #203=(CWideArrow
COMMENT                                (CWideLine
COMMENT                                 (CWidget 0 (3 #13# 0) 1 2 0 0 #160# 0 100)
COMMENT                                 (LOGPEN 0 0 3355545) 1
COMMENT                                 (LOGBRUSH 0 6724095 0) 1 5.6363 6.18908 1
COMMENT                                 0.082322) 0.8 1.8 0)
COMMENT                           #204=(CLabel
COMMENT                                (CWidget 0 (0 0) 1 2 0 0 #157# 61693716 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "Amp r" "@N" 1
COMMENT                                7.82381 10.454 -13.6904 -5.13538 1.91453
COMMENT                                0.843305 #203#)) (CObjectList))
COMMENT                      #205=(CGroupWidget
COMMENT                          (CWidget 20 (7 20 0) 1 2 0 0 Nil -255 100)
COMMENT                          (CObjectList
COMMENT                           #206=(CLine
COMMENT                                (CWidget 0 (3 #9# 0) 1 2 0 0 #163# 0 100)
COMMENT                                (LOGPEN 0 14 6723840) 8 0.835356 4.00428
COMMENT                                4.15965)
COMMENT                           #207=(CLabel (CWidget 0 (0 0) 1 2 0 0 #157# 50 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "RRE" "@N" 1
COMMENT                                11.0245 -6.96932 -5.93426 -8.44645 1.47009
COMMENT                                0.843305 #206#)) (CObjectList))
COMMENT                      #208=(CGroupWidget
COMMENT                          (CWidget 21 (7 21 0) 1 2 0 0 Nil -477 100)
COMMENT                          (CObjectList
COMMENT                           #209=(CWideArrow
COMMENT                                (CWideLine
COMMENT                                 (CWidget 0 (3 #0# 0) 1 2 0 0 #161# 0 100)
COMMENT                                 (LOGPEN 0 0 6723840) 1
COMMENT                                 (LOGBRUSH 0 10079334 0) 1 4.5866 4.7028 1
COMMENT                                 0.082322) 0.8 1.8 0)
COMMENT                           #210=(CLabel
COMMENT                                (CWidget 0 (0 0) 1 2 0 0 #157# 83486440 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "RU5" "@N" 1
COMMENT                                12.8882 -1.74233 -7.54735 -5.80603 1.39031
COMMENT                                0.843305 #209#)
COMMENT                           #211=(CWideArrow
COMMENT                                (CWideLine
COMMENT                                 (CWidget 0 (3 #1# 0) 1 2 0 0 #161# 0 100)
COMMENT                                 (LOGPEN 0 0 6723840) 1
COMMENT                                 (LOGBRUSH 0 10079334 0) 1 0.659997 0.810872
COMMENT                                 1 0.082322) 0.8 1.8 0)
COMMENT                           #212=(CLabel (CWidget 0 (0 0) 1 2 0 0 #157# 0 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "sin18 RU5" "@N"
COMMENT                                1 -11.0942 13.9386 12.1409 -3.32342 3.20228
COMMENT                                0.843305 #211#)
COMMENT                           #213=(CWideArrow
COMMENT                                (CWideLine
COMMENT                                 (CWidget 0 (3 #2# 0) 1 2 0 0 #161# 0 100)
COMMENT                                 (LOGPEN 0 0 6723840) 1
COMMENT                                 (LOGBRUSH 0 10079334 0) 1 4.25275 4.48708 1
COMMENT                                 0.082322) 0.8 1.8 0)
COMMENT                           #214=(CLabel
COMMENT                                (CWidget 0 (0 0) 1 2 0 0 #157# 425726633 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "GAG" "@N" 1
```

```
COMMENT                                    12.5916 -3.48466 -6.74925 -7.18322 1.59544
COMMENT                                    0.843305 #213#)
COMMENT                                 #215=(CScratch
COMMENT                                    (CWidget 0 (3 #8# 0) 1 2 0 0 #165# 268697852
COMMENT                                     100) (LOGPEN 0 16 13395507) 10 1 3.77829 1.9
COMMENT                                    0.082322 1)
COMMENT                                 #216=(CLabel
COMMENT                                    (CWidget 0 (0 0) 1 2 0 0 #157# 425726605 100)
COMMENT                                    (LOGPEN 0 0 0) 1
COMMENT                                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                     "Arial") 2.53336 0.666667 0 "HIV-SA" "@N" 1
COMMENT                                    9.25844 -10.454 -3.96879 2.6208 2.2792
COMMENT                                    0.843305 #215#)
COMMENT                                 #217=(CWideArrow
COMMENT                                    (CWideLine
COMMENT                                     (CWidget 0 (3 #14# 0) 1 2 0 0 #161# 1 100)
COMMENT                                     (LOGPEN 0 0 6723840) 1
COMMENT                                     (LOGBRUSH 0 10079334 0) 0.835356 4.98979
COMMENT                                     6.27319 1 0.082322) 0.8 1.8 0)
COMMENT                                 #218=(CLabel
COMMENT                                    (CWidget 0 (0 0) 1 2 0 0 #157# 61694136 100)
COMMENT                                    (LOGPEN 0 0 0) 1
COMMENT                                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                     "Arial") 2.53336 0.666667 0 "pUC19" "@N" 1
COMMENT                                    10.1479 8.71165 -10.5684 -7.24192 2.15385
COMMENT                                    0.843305 #217#)) (CObjectList))
COMMENT                              #219=(CGroupWidget
COMMENT                                 (CWidget 22 (7 22 0) 1 2 0 0 Nil -1033 100)
COMMENT                                 (CObjectList
COMMENT                                  #220=(CLabel
COMMENT                                    (CWidget 0 (0 0) 1 2 0 0 #157# 61689908 100)
COMMENT                                    (LOGPEN 0 0 0) 1
COMMENT                                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                     "Arial") 2.53336 0.666667 0 "HS3" "@N" 1
COMMENT                                    -12.4348 3.48466 9.09696 -8.60083 1.35613
COMMENT                                    0.843305
COMMENT                                    #221=(CWideLine
COMMENT                                          (CWidget 0 (3 #3# 0) 1 2 0 0 #173#
COMMENT                                           543452769 100) (LOGPEN 0 0 0) 1
COMMENT                                          (LOGBRUSH 0 128 0) 1 0.880211 1.65192 1
COMMENT                                          0.082322)) #221#) (CObjectList))
COMMENT                              #222=(CGroupWidget
COMMENT                                 (CWidget 25 (7 25 0) 1 2 0 0 Nil -477 100)
COMMENT                                 (CObjectList
COMMENT                                  #223=(CScratch
COMMENT                                    (CWidget 0 (3 #10# 0) 1 2 0 0 #176#
COMMENT                                     2082537934 100) (LOGPEN 0 14 32768) 8 1
COMMENT                                    3.40977 1.9 0.082322 1)
COMMENT                                  #224=(CLabel
COMMENT                                    (CWidget 0 (0 0) 1 2 0 0 #157# 268697892 100)
COMMENT                                    (LOGPEN 0 0 0) 1
COMMENT                                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                     "Arial") 2.53336 0.666667 0
COMMENT                                    "\247-globin polyA" "@N" 1 7.10156 -12.1963
COMMENT                                    0.96022 -8.77349 4.53561 0.843305 #223#))
COMMENT                                 (CObjectList))
COMMENT                              #225=(CGroupWidget
COMMENT                                 (CWidget 29 (7 29 0) 1 2 0 0 Nil -811 100)
COMMENT                                 (CObjectList
COMMENT                                  #226=(CWideArrow
COMMENT                                    (CWideLine
COMMENT                                     (CWidget 0 (3 #12# 0) 1 2 0 0 #161#
COMMENT                                      507308978 100) (LOGPEN 0 0 6723840) 1
COMMENT                                     (LOGBRUSH 0 10079334 0) 0.835356 4.70216
COMMENT                                     4.85111 1 0.082322) 0.8 1.8 0)
COMMENT                                  #227=(CLabel
COMMENT                                    (CWidget 0 (0 0) 1 2 0 0 #157# 83484000 100)
COMMENT                                    (LOGPEN 0 0 0) 1
COMMENT                                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                     "Arial") 2.53336 0.666667 0 "RSV" "@N" 1
COMMENT                                    12.8905 0 -8.44706 -4.78211 1.4359 0.843305
COMMENT                                    #226#)) (CObjectList))
COMMENT                              #228=(CGroupWidget
COMMENT                                 (CWidget 38 (7 38 0) 1 2 0 0 Nil -1255 100)
COMMENT                                 (CObjectList
COMMENT                                  #229=(CScratch
COMMENT                                    (CWidget 0 (3 #15# 0) 1 2 0 0 #176# 268697822
COMMENT                                     100) (LOGPEN 0 14 32768) 8 1 1.06961 1.9
COMMENT                                    0.082322 1)
COMMENT                                  #230=(CLabel
COMMENT                                    (CWidget 0 (0 0) 1 2 0 0 #157# 61695732 100)
COMMENT                                    (LOGPEN 0 0 0) 1
COMMENT                                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                     "Arial") 2.53336 0.666667 0 "SA A" "@N" 1
COMMENT                                    -12.3403 12.1963 12.1485 2.03443 1.64103
COMMENT                                    0.843305 #229#)
COMMENT                                  #231=(CScratch
```

```
COMMENT                    (CWidget 0 (3 #16# 0) 1 2 0 0 #176# 268697807
COMMENT                     100) (LOGPEN 0 14 32768) 8 1 1.09465 1.9
COMMENT                    0.082322 1)
COMMENT                #232=(CLabel (CWidget 0 (0 0) 1 2 0 0 #157# 0 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SA B" "@N" 1
COMMENT                    -12.3464 10.454 11.4863 3.38088 1.64103
COMMENT                    0.843305 #231#)
COMMENT                #233=(CScratch
COMMENT                    (CWidget 0 (3 #17# 0) 1 2 0 0 #176# 268697792
COMMENT                     100) (LOGPEN 0 14 32768) 8 1 1.47536 1.9
COMMENT                    0.082322 1)
COMMENT                #234=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #157# 268889064 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SA C" "@N" 1
COMMENT                    -13.0385 0 5.59896 4.74734 1.64103 0.843305
COMMENT                    #233#)
COMMENT                #235=(CScratch
COMMENT                    (CWidget 0 (3 #18# 0) 1 2 0 0 #176# 268697602
COMMENT                     100) (LOGPEN 0 14 32768) 8 1 2.79215 1.9
COMMENT                    0.082322 1)
COMMENT                #236=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #157# 425726885 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SA D" "@N" 1
COMMENT                    -7.54504 -10.454 2.42732 1.97573 1.64103
COMMENT                    0.843305 #235#)
COMMENT                #237=(CScratch
COMMENT                    (CWidget 0 (3 #19# 0) 1 2 0 0 #176# 268697762
COMMENT                     100) (LOGPEN 0 14 32768) 8 1 1.19416 1.9
COMMENT                    0.082322 1)
COMMENT                #238=(CLabel (CWidget 0 (0 0) 1 2 0 0 #157# 0 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SD A" "@N" 1
COMMENT                    -12.6922 8.71165 11.3022 4.69315 1.64103
COMMENT                    0.843305 #237#)
COMMENT                #239=(CScratch
COMMENT                    (CWidget 0 (3 #20# 0) 1 2 0 0 #176# 268697747
COMMENT                     100) (LOGPEN 0 14 32768) 8 1 1.24231 1.9
COMMENT                    0.082322 1)
COMMENT                #240=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #157# 61691700 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SD B" "@N" 1
COMMENT                    -12.743 6.96932 10.9213 5.7547 1.64103
COMMENT                    0.843305 #239#)
COMMENT                #241=(CScratch
COMMENT                    (CWidget 0 (3 #21# 0) 1 2 0 0 #176# 268697732
COMMENT                     100) (LOGPEN 0 14 32768) 8 1 1.25258 1.9
COMMENT                    0.082322 1)
COMMENT                #242=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #157# 61694360 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SD C" "@N" 1
COMMENT                    -12.6455 5.22699 9.50654 6.42879 1.64103
COMMENT                    0.843305 #241#)
COMMENT                #243=(CScratch
COMMENT                    (CWidget 0 (3 #22# 0) 1 2 0 0 #176# 1 100)
COMMENT                    (LOGPEN 0 14 32768) 8 1 1.26799 1.9 0.082322
COMMENT                    1)
COMMENT                #244=(CLabel (CWidget 0 (0 0) 1 2 0 0 #157# 0 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SD D" "@N" 1
COMMENT                    -12.9363 1.74233 6.42714 5.90802 1.64103
COMMENT                    0.843305 #243#)
COMMENT                #245=(CScratch
COMMENT                    (CWidget 0 (3 #23# 0) 1 2 0 0 #176# 1 100)
COMMENT                    (LOGPEN 0 14 32768) 8 1 1.54406 1.9 0.082322
COMMENT                    1)
COMMENT                #246=(CLabel
COMMENT                    (CWidget 0 (0 0) 1 2 0 0 #157# 61691924 100)
COMMENT                    (LOGPEN 0 0 0) 1
COMMENT                    (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                     "Arial") 2.53336 0.666667 0 "SD E" "@N" 1
COMMENT                    -13.0333 -1.74233 5.29447 3.68924 1.59544
COMMENT                    0.843305 #245#)
COMMENT                #247=(CScratch
COMMENT                    (CWidget 0 (3 #24# 0) 1 2 0 0 #176# 1 100)
COMMENT                    (LOGPEN 0 14 32768) 8 1 1.54663 1.9 0.082322
```

```
COMMENT                                 1)
COMMENT                          #248=(CLabel
COMMENT                                (CWidget 0 (0 0) 1 2 0 0 #157# 268889064 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "SD F" "@N" 1
COMMENT                                -13.1063 -3.48466 5.1744 2.30065 1.56125
COMMENT                                0.843305 #247#)) (CObjectList))
COMMENT                    #249=(CGroupWidget
COMMENT                          (CWidget 50 (7 50 0) 1 2 0 0 Nil -700 100)
COMMENT                          (CObjectList
COMMENT                           #250=(CLine
COMMENT                                (CWidget 0 (3 #11# 0) 1 2 0 0 #163#
COMMENT                                 1920099654 100) (LOGPEN 0 14 6723840) 8
COMMENT                                0.835356 3.33722 3.42197)
COMMENT                           #251=(CLabel
COMMENT                                (CWidget 0 (0 0) 1 2 0 0 #157# 61693492 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "3' UTR" "@N" 1
COMMENT                                5.11639 -13.9386 -0.00360714 -7.79171 2.11966
COMMENT                                0.843305 #250#)) (CObjectList))
COMMENT                    #252=(CGroupWidget
COMMENT                          (CWidget 61 (7 61 0) 1 2 0 0 Nil -144 100)
COMMENT                          (CObjectList
COMMENT                           #253=(CWideArrow
COMMENT                                (CWideLine
COMMENT                                 (CWidget 0 (3 #5# 0) 1 2 0 0 #160# 0 100)
COMMENT                                 (LOGPEN 0 0 3355545) 1
COMMENT                                 (LOGBRUSH 0 6724095 0) 1 2.94623 3.0894 1
COMMENT                                 0.082322) 0.8 1.8 1)
COMMENT                           #254=(CLabel
COMMENT                                (CWidget 0 (0 0) 1 2 0 0 #157# 444 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "Exon 2" "@N" 1
COMMENT                                -3.86727 -13.9386 1.63466 -5.65615 2.19943
COMMENT                                0.843305 #253#)
COMMENT                           #255=(CWideArrow
COMMENT                                (CWideLine
COMMENT                                 (CWidget 0 (3 #6# 0) 1 2 0 0 #160# 0 100)
COMMENT                                 (LOGPEN 0 0 3355545) 1
COMMENT                                 (LOGBRUSH 0 6724095 0) 0.835356 3.2544
COMMENT                                 3.33722 1 0.082322) 0.8 1.8 1)
COMMENT                           #256=(CLabel
COMMENT                                (CWidget 0 (0 0) 1 2 0 0 #157# 61692344 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "Exon 3" "@N" 1
COMMENT                                4.32194 -15.681 0.409159 -6.71877 2.19943
COMMENT                                0.843305 #255#)
COMMENT                           #257=(CWideArrow
COMMENT                                (CWideLine
COMMENT                                 (CWidget 0 (3 #7# 0) 1 2 0 0 #160#
COMMENT                                  2082602632 100) (LOGPEN 0 0 3355545) 1
COMMENT                                 (LOGBRUSH 0 6724095 0) 0.835356 2.80114
COMMENT                                 2.86277 1 0.082322) 0.8 1.8 1)
COMMENT                           #258=(CLabel (CWidget 0 (0 0) 1 2 0 0 #157# 1 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "Exon 1" "@N" 1
COMMENT                                -5.92224 -12.1963 2.4893 -4.63224 2.19943
COMMENT                                0.843305 #257#)) (CObjectList))
COMMENT                    #259=(CGroupWidget
COMMENT                          (CWidget 85 (7 85 0) 1 2 0 0 Nil -588 100)
COMMENT                          (CObjectList
COMMENT                           #260=(CLabel
COMMENT                                (CWidget 0 (0 0) 1 2 0 0 #157# 425726837 100)
COMMENT                                (LOGPEN 0 0 0) 1
COMMENT                                (LOGFONT 73 28 0 0 700 0 0 0 0 3 2 1 34
COMMENT                                 "Arial") 2.53336 0.666667 0 "HS2" "@N" 1
COMMENT                                -11.4615 -6.96932 5.0729 -8.26929 1.35613
COMMENT                                0.843305
COMMENT                                #261=(CWideLine
COMMENT                                      (CWidget 0 (3 #4# 0) 1 2 0 0 #161#
COMMENT                                       543452769 100) (LOGPEN 0 0 6723840) 1
COMMENT                                      (LOGBRUSH 0 10079334 0) 1 1.65577
COMMENT                                      2.57707 1 0.082322)) #261#)
COMMENT                           (CObjectList))) (CObjectList))
COMMENT                #262=(CGroupWidget (CWidget 14 (16 0) 1 2 0 0 Nil -366 100)
COMMENT                      (CObjectList) (CObjectList))
COMMENT                #263=(CGroupWidget (CWidget 11 (0 0) 1 2 0 0 Nil -1033 100)
COMMENT                      (CObjectList) (CObjectList))
COMMENT                #264=(CGroupWidget (CWidget 12 (0 0) 1 2 0 0 Nil -1144 100)
COMMENT                      (CObjectList) (CObjectList))) (CObjectList)))
COMMENT        (CSeqView 10 10 (CObjectList) (CObList) 1 (CObList)) (CObList) 291
COMMENT        (CStringList) 260 1 (CObList)))
```

```
FEATURES             Location/Qualifiers
     misc_feature    2447..2627
                     /vntifkey="21"
                     /label=RU5
     misc_feature    8509..8743
                     /vntifkey="21"
                     /label=sin18\RU5
     misc_feature    2783..3147
                     /vntifkey="21"
                     /label=GAG
     misc_marker     7199..8400
                     /vntifkey="22"
                     /label=HS3
     misc_difference 575B..7192
                     /vntifkey="85"
                     /label=HS2
     exon            complement(4960..5182)
                     /vntifkey="61"
                     /label=Exon\2
     exon            complement(4574..4702)
                     /vntifkey="61"
                     /label=Exon\3
     exon            complement(5313..5408)
                     /vntifkey="61"
                     /label=Exon\1
     misc_feature    3887..3898
                     /vntifkey="21"
                     /label=HIV-SA
     misc_binding    3293..3534
                     /vntifkey="20"
                     /label=RRE
     polyA_signal    4461..4466
                     /vntifkey="25"
                     /label=§-globin\polyA
     3'UTR           4442..4573
                     /vntifkey="50"
                     /label=3'\UTR
     promoter        2216..2447
                     /vntifkey="29"
                     /label=RSV
     CDS             132..992
                     /vntifkey="4"
                     /label=Amp\r
     misc_feature    1..1999
                     /vntifkey="21"
                     /label=pUC19
     splicing_signal 8106..8107
                     /vntifkey="38"
                     /label=SA_A
     splicing_signal 8067..8068
                     /vntifkey="38"
                     /label=SA_B
     splicing_signal 7474..7475
                     /vntifkey="38"
                     /label=SA_C
     splicing_signal 5423..5424
                     /vntifkey="38"
                     /label=SA_D
     splicing_signal 7912..7913
                     /vntifkey="38"
                     /label=SD_A
     splicing_signal 7837..7838
                     /vntifkey="38"
                     /label=SD_B
     splicing_signal 7821..7822
                     /vntifkey="38"
                     /label=SD_C
     splicing_signal 7797..7798
                     /vntifkey="38"
                     /label=SD_D
     splicing_signal 7367..7368
                     /vntifkey="38"
                     /label=SD_E
     splicing_signal 7363..7364
                     /vntifkey="38"
                     /label=SD_F
BASE COUNT      2712 a      2253 c      2273 g      2533 t
ORIGIN
        1 taggtggcac ttttcgggga aatgtgcgcg gaacccctat ttgtttattt ttctaaatac
       61 attcaaatat gtatccgctc atgagacaat aaccctgata aatgcttcaa taatattgaa
      121 aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt tttgcggcat
      181 tttgccttcc tgtttttgct cacccagaaa cgctggtgaa agtaaaagat gctgaagatc
      241 agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag atccttgaga
      301 gttttcgccc cgaagaacgt tttccaatga tgagcacttt taaagttctg ctatgtggcg
      361 cggtattatc ccgtattgac gccgggcaag agcaactcgg tcgccgcata cactattctc
      421 agaatgactt ggttgagtac tcaccagtca cagaaaagca tcttacggat ggcatgacag
      481 taagagaatt atgcagtgct gccataacca tgagtgataa cactgcggcc aacttacttc
```

```
 541 tgacaacgat cggaggaccg aaggagctaa ccgctttttt gcacaacatg ggggatcatg
 601 taactcgcct tgatcgttgg gaaccggagc tgaatgaagc cataccaaac gacgagcgtg
 661 acaccacgat gcctgtagca atggcaacaa cgttgcgcaa actattaact ggcgaactac
 721 ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa gttgcaggac
 781 cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct ggagccggtg
 841 agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc tcccgtatcg
 901 tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga cagatcgctg
 961 agataggtgc ctcactgatt aagcattggt aactgtcaga ccaagtttac tcatatatac
1021 tttagattga tttaaaactt catttttaat ttaaaaggat ctaggtgaag atccttttg
1081 ataatctcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg tcagaccccg
1141 tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc
1201 aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag ctaccaactc
1261 tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtt cttctagtgt
1321 agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac ctcgctctgc
1381 taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc gggttggact
1441 caagacgata gttaccggat aaggcgcagc ggtcgggctg aacgggggt tcgtgcacac
1501 agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt gagctatgag
1561 aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc ggcagggtcg
1621 gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg
1681 tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca gggggcggga
1741 gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt tgctggcctt
1801 ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt attaccgcct
1861 ttgagtgagc tgataccgct cgccgcagcc gaacgaccga gcgcagcgag tcagtgagcg
1921 aggaagcgga agagcgccca atacgcaaac cgcctctccc cgcgcgttgg ccgattcatt
1981 aatgcagctg gcacgacagg tttcccgact ggaaagcggg cagtgagcgc aacgcaatta
2041 atgtgagtta gctcactcat taggcacccc aggctttaca ctttatgctt ccggctcgta
2101 tgttgtgtgg aattgtgagc ggataacaat ttcacacagg aaacagctat gaccatgatt
2161 acgccaagcg cgcaattaac cctcactaaa gggaacaaaa gctggagctg caagcttaat
2221 gtagtcttat gcaatactct tgtagtcttg caacatggta acgatgagtt agcaacatgc
2281 cttacaagga gagaaaaagc accgtgcatg ccgattggtg gaagtaaggt ggtacgatcg
2341 tgccttatta ggaaggcaac agacgggtct gacatggatt ggacgaacca ctgaattgcc
2401 gcattgcaga gatattgtat ttaagtgcct agctcgatac ataaacgggt ctctctggtt
2461 agaccagatc tgagcctggg agctctctgg ctaactaggg aacccactgc ttaagcctca
2521 ataaagcttg ccttgagtgc ttcaagtagt gtgtgcccgt ctgttgtgtg actctggtaa
2581 ctagagatcc ctcagaccct tttagtcagt gtggaaaatc tctagcagtg gcgcccgaac
2641 agggacttga aagcgaaagg gaaaccagag gagctctctc gacgcaggac tcggcttgct
2701 gaagcgcgca cggcaagagg cgaggggcgg cgactggtga gtacgccaaa aattttgact
2761 agcggaggct agaaggagag agatgggtgc gagagcgtca gtattaagcg ggggagaatt
2821 agatcgcgat gggaaaaaat tcggttaagg ccaggggaa agaaaaaata taaattaaaa
2881 catatagtat gggcaagcag ggagctagaa cgattcgcag ttaatcctgg cctgttagaa
2941 acatcagaag gctgtagaca aatactggga cagctacaac catcccttca gacaggatca
3001 gaagaactta gatcattata taatacagta gcaaccctct attgtgtgca tcaaaggata
3061 gagataaaag acaccaagga agctttagac aagatagagg aagagcaaaa caaaagtaag
3121 accaccgcac agcaagcggc cgctgatctt cagacctgga ggaggagata tgagggacaa
3181 ttggagaagt gaattatata aatataaagt agtaaaaatt gaaccattag gagtagcacc
3241 caccaaggca aagagaagag tggtgcagag agaaaaaaga gcagtgggaa taggagcttt
3301 gttccttggg ttcttgggag cagcaggaag cactatgggc gcagcgtcaa tgacgctgac
3361 ggtacaggcc agacaattat tgtctggtat agtgcagcag cagaacaatt tgctgagggc
3421 tattgaggcg caacagcatc tgttgcaact cacagtctgg ggcatcaagc agctccaggc
3481 aagaatcctg gctgtggaaa gatacctaaa ggatcaacag ctcctgggga tttggggttg
3541 ctctggaaaa ctcatttgca ccactgctgt gccttggaat gctagttgga gtaataaatc
3601 tctggaacag atttggaatc acacgacctg gatggagtgg gacagagaaa ttaacaatta
3661 cacaagctta atacactcct taattgaaga atcgcaaaac cagcaagaaa agaatgaaca
3721 agaattattg gaattagata aatgggcaag tttgtggaat tggtttaaca taacaaattg
3781 gctgtggtat ataaaattat tcataatgat agtaggaggc ttggtaggtt taagaatagt
3841 ttttgctgta ctttctatag tgaatagagt taggcaggga tattcaccat tatcgtttca
3901 gacccacctc ccaaccccga ggggacccga caggcccgaa ggaatagaag aagaaggtgg
3961 agagagagac agagacagat ccattcgatt agtgaacgga tctcgacggt atcggttaac
4021 ttttaaaaga aaaggggga ttgggggta cagtgcaggg gaaagaatag tagacataat
4081 agcaacagac atacaaacta aagaattaca aaaacaaatt acaaaattca aaattttatc
4141 ggtacgtacc atgaggacag ctaaaacaat aagtaatgta aaatacagca tagcaaaact
4201 ttaacctcca aatcaagcct ctacttgaat ccttttctga gggatgaata aggcataggc
4261 atcaggggct gttgccaatg tgcattagct gtttgcagcc tcaccttctt tcatggagtt
4321 taagatatag tgtattttcc caaggtttga actagctctt catttcttta tgttttaaat
4381 gcactgacct cccacattcc ctttttagta aaatattcag aaataattta aatacatcat
4441 tgcaatgaaa ataaatgttt tttattaggc agaatccaga tgctcaaggc ccttcataat
4501 atcccccagt ttagtagttg gacttaggga acaaaggaac ctttaataga aattggacag
4561 caagaaagcg agcttagtga tacttgtggg ccagggcatt agccacacca gccaccactt
4621 tctgataggc agcctgcact ggtggggtga attctttgcc aaagtgatgg gccagcacac
4681 agaccagcac gttgcccagg agctgtggga ggaagataag aggtatgaac atgattagca
4741 aaagggccta gcttggactc agaataatcc agccttatcc caaccataaa ataaaagcag
4801 aatggtagct ggattgtagc tgctattagc aatatgaaac ctcttacatc agttacaatt
4861 tatatgcaga aatacctgt tacttctccc cttcctatga catgaactta accatagaaa
4921 agaaggggaa agaaaacatc aagggtccca tagactcacc ctgaagttct caggatccac
4981 gtgcagcttg tcacagtgca gctcactcag tgtggcaaag gtgcccttga ggttgtccag
5041 gtgagccagg ccatcactaa aggcaccgag cactttcttg ccatgagcct tcaccttagg
5101 gttgcccata acagcatcag gagtggacag atccccaaag gactcaaaga acctctgggt
5161 ccaagggtag accaccagca gcctaagggt gggaaaatag accaataggc agagagagtc
5221 agtgcctatc agaaacccaa gagtcttctc tgtctccaca tgcccagttt ctattggtct
5281 ccttaaacct gtcttgtaac cttgatacca acctgcccag ggcctcacca ccaacttcat
5341 ccacgttcac cttgccccac agggcagtaa cggcagactt ctcctcagga gtcaggtgca
5401 ccatggtgtc tgtttgaggt tgctagtgaa cacagttgtg tcagaagcaa atgtaagcaa
5461 tagatggctc tgccctgact tttatgccca gccctggctc ctgccctccc tgctcctggg
5521 agtagattgg ccaaccctag ggtgtggctc cacagggtga ggtctaagtg atgacagccg
5581 tacctgtcct tggctctttct ggcactggct taggagttgg acttcaaacc ctcagccctc
5641 cctctaagat atatctcttg gccccatacc atcagtacaa attgctacta aaaacatcct
5701 cctttgcaag tgtatttaca cggtatcgat aagcttgata tcgaattcct gcagcccccct
```

```
5761 tttgccacct agctgtccag gggtgcctta aaatggcaaa caaggtttgt tttcttttcc
5821 tgttttcatg ccttcctctt ccatatcctt gtttcatatt aatacatgtg tatagatcct
5881 aaaaatctat acacatgtat taataaagcc tgattctgcc gcttctaggt atagaggcca
5941 cctgcaagat aaatatttga ttcacaataa ctaatcattc tatggcaatt gataacaaca
6001 aatatatata tatatatata tacgtatatg tgtatatata tatatatata ttcaggaaat
6061 aatatattct agaatatgtc acattctgtc tcaggcatcc attttcttta tgatgccgtt
6121 tgaggtggag ttttagtcag gtggtcagct tctccttttt tttgccatct gccctgtaag
6181 catcctgctg gggacccaga taggagtcat cactctaggc tgagaacatc tgggcacaca
6241 ccctaagcct cagcatgact catcatgact cagcattgct gtgcttgagc cagaaggttt
6301 gcttagaagg ttacacagaa ccagaaggcg ggggtggggc actgaccccg acaggggcct
6361 ggccagaact gctcatgctt ggactatggg aggtcactaa tggagacaca cagaaatgta
6421 acaggaacta aggaaaaact gaagcttatt taatcagaga tgaggatgct ggaagggata
6481 gagggagctg agcttgtaaa aagtatagta atcattcagc aaatggtttt gaagcacctg
6541 ctggatgcta aacactattt tcagtgcttg aatcataaat aagaataaaa catgtatctt
6601 attccccaca agagtccaag taaaaaataa cagttaatta taatgtgctc tgtcccccag
6661 gctggagtgc agtggcacga tctcagctca ctgcaacctc cgcctcccgg gttcaagcaa
6721 ttctcctgcc tcagccaccc taatagctgg gattacaggt gcacaccacc atgccaggct
6781 aatttttgta ctttttgtag aggttttgta ctttttgtag aggcagggta tcaccatgtt
6841 gtccaagatg gtcttgaact cctgagctcc aagcagtcca cccacctcag cctcccaaag
6901 tgctgggatt acaggtgtga gacaccatgc ccagattttc catatttaat agaggtattt
6961 atgggatggg ggaaaagaat gtttctctca ctgtggatta ttttagagag tggagaatgg
7021 tcaagatttt tttaaaaatt aagaaaacat aagttggacc ttgagaaatg aaaatttatt
7081 tttttgttgg aggataccca ttctctatct cccatcaggg caagctgtaa ggaactggct
7141 aagacacagt gagacagagt gacttagtct tagaggcccc actggtacga cggtcaccaa
7201 gctttcatta aaaaaagtct aaccagctgc attcgacttt gactgcagca gctggttaga
7261 aggttctact ggaggagggt cccagcccat tgctaaatta acatcaggct ctgagactgg
7321 cagtatatct ctaacagtgg ttgatgctat cttctggaac ttgcctgcta cattgagacc
7381 actgacccat acataggaag cccatagctc tgtcctgaac tgttaggcca ctggtccaga
7441 gagtgtgcat ctcctttgat cctcataata accctatgag atagacacaa ttattactct
7501 tactttatag atgatgatcc tgaaaacata ggagtcaagg cacttgcccc tagctggggg
7561 tataggggag cagtcccatg tagtagtaga atgaaaaatg ctgctatgct gtgcctcccc
7621 caccttccc atgtctgccc tctactcatg gtctatctct cctggctcct gggagtcatg
7681 gactccaccc agcaccacca acctgaccta accacctatc tgagcctgcc agcctataac
7741 ccatctgggc cctgatagct ggtggccagc cctgacccca ccccaccctc cctggaacct
7801 ctgatagaca catctggcac accagctcgc aaagtcaccg tgagggtctt gtgtttgctg
7861 agtcaaaatt ccttgaaatc caagtcctta gagactcctg ctcccaaatt tacagtcata
7921 gacttcttca tggctgtctc ctttatccac agaatgattc ctttgcttca ttgccccatc
7981 catctgatcc tcctcatcag tgcagcacag ggcccatgag cagtagctgc agagtctcac
8041 ataggtctgg cactgcctct gacatgtccg accttaggca aatgcttgac tcttctgagc
8101 tcagtcttgt catggcaaaa taaagataat aatagtgttt ttttatggag ttagcgtgag
8161 gatggaaaac aatagcaaaa ttgattagac tataaaaggt ctcaacaaat agtagtagat
8221 tttatcatcc attaatcctt ccctctcctc tcttactcat cccatcacgt atgcctctta
8281 attttccctt acctataata agagttattc ctcttattat attcttctta tagtgattct
8341 ggatattaaa gtgggaatga ggggcaggcc actaacgaag aagatgtttc tcaaagaagc
8401 gggggatcca ctagttctag agcggccaaa tggcggccgt acctttaaga ccaatgactt
8461 acaaggcagc tgtagatctt agccactttt taaaagaaaa gggggactg gaagggctaa
8521 ttcactccca acgaagacaa gatctgcttt ttgcttgtac tgggtctctc tggttagacc
8581 agatctgagc ctgggagctc tctggctaac tagggaaccc actgcttaag cctcaataaa
8641 gcttgccttg agtgcttcaa gtagtgtgtg cccgtctgtt gtgtgactct ggtaactaga
8701 gatccctcag acccttttag tcagtgtgga aaatctctag cagtagtagt tcatgtcatc
8761 ttattattca gtatttataa cttgcaaaga aatgaatatc agagagtgag aggaacttgt
8821 ttattgcagc ttataatggt tacaaataaa gcaatagcat cacaaatttc acaaataaag
8881 catttttttc actgcattct agttgtggtt tgtccaaact catcaatgta tcttatcatg
8941 tctggctcta gctatcccgc ccctaactcc gcccatcccg cccctaactc cgcccagttc
9001 cgcccattct ccgccccatg gctgactaat ttttttttatt tatgcagagg ccgaggccgc
9061 ctcggcctct gagctattcc agaagtagtg aggaggcttt tttggaggcc tagggacgta
9121 cccaattcgc cctatagtga gtcgtattac gcgcgctcac tggccgtcgt tttacaacgt
9181 cgtgactggg aaaaccctgg cgttacccaa cttaatcgcc ttgcagcaca tccccctttc
9241 gccagctggc gtaatagcga agaggcccgc accgatcgcc cttcccaaca gttgcgcagc
9301 ctgaatggcg aatgggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt
9361 acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc
9421 ccttcctttc tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg ggggctccct
9481 ttagggttcc gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat
9541 ggttcacgta gtgggccatc gccctgatag acggtttttc gcccttgac gttggagtcc
9601 acgttcttta atagtggact cttgttccaa actggaacaa cactcaaccc tatctcggtc
9661 tattcttttg atttataagg gattttgccg atttcggcct attggttaaa aaatgagctg
9721 atttaacaaa aatttaacgc gaattttaac aaaatattaa cgcttacaat t
```

Appendix - SEQ ID NO3

```
LOCUS       #743.pCCLsin.PPT        7830 bp    DNA     circular     20-APR-2012
SOURCE
  ORGANISM
COMMENT     http://www.informaxinc.com/
COMMENT     This file is created by Vector NTI
            http://www.invitrogen.com/
COMMENT     ORIGDB|GenBank
COMMENT     VNTDATE|623879904|
COMMENT     VNTDBDATE|623880825|
COMMENT     LSOWNER|
COMMENT     VNTNAME|#743.pCCLsin.PPT.hPGK.GFP.Wpre_mut_AMP_splicing_aggiornata_SIMPLIFIED|
COMMENT     VNTAUTHORNAME|Demo User|
COMMENT     VNTOAUTHORNAME|IRCC - ROSSO 3|
COMMENT     Vector_NTI_Display_Data_(Do_Not_Edit!)
COMMENT     (SXF
COMMENT      (CGexDoc
COMMENT       "#743.pCCLsin.PPT.hPGK.GFP.Wpre_mut_AMP_splicing_aggiornata_SIMPLIFIED" 0
COMMENT       7830
COMMENT       (CDBMol 0 0 1 1 1 0 0 0 0 "" "" 0 0 0 0 (CObList) (CObList) (CObList)
COMMENT        (CObList) -1 "")
COMMENT       (CDocSetData 1 0 0 0 1 1000 "MAIN" 1 1 1 1 0 0 1 1 0 1 10 6 40 50 0 1 0
COMMENT        (CHomObj 0 0 0 3 75) (CWordArray) (CWordArray)
COMMENT        (CStringList "AvrII" "SacII") (CStringList "atg")
COMMENT        (CStringList "taa" "tga" "tag") (CObList) 1 "{(0,1),2}" 0 0 "" 0
COMMENT        4294967295 0 1 0 0 0 0 1 "MAIN" 0 0 30 0
COMMENT        (CProteinMotifSearchObject 100 10 1 1 1 1 1 0 1 0 0 0 0 0))
COMMENT       (CMolPar 1 0 0 0 4294967295 1 7830 0 0 0 0 0 0 0 0) (CStringList)
COMMENT       (CStringList) (CObList) (COAPar 25 250 50 0 6 4 3 7)
COMMENT       (COAPar 25 250 50 0 6 4 3 7) (COAPar 25 250 50 0 6 4 3 7) (CObList)
COMMENT       (CObList
COMMENT        #0=(CFSignal (CObList) "hPGK" 29 0 0 4625 5140 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "4622..5137")
COMMENT        #1=(CFSignal (CObList) "dR3RU5" 19 0 0 6571 6804 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "6568..6801")
COMMENT        #2=(CFSignal (CObList) "RU5" 19 0 0 2891 3072 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "2891..3072")
COMMENT        #3=(CFSignal (CObList) "eGFP" 4 0 0 5161 5880 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "5158..5877")
COMMENT        #4=(CFSignal (CObList) "MUTATED Woodchuck hepatitis B virus PRE" 21 0 0
COMMENT            5896 6487 0 (CStringList)
COMMENT            (CStringList "/db_xref=\"ID_TheraBank_407\"") 1 1 1 1 "5893..6484")
COMMENT        #5=(CFSignal (CObList) "SD5" 21 0 0 2974 2975 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #6=(CFSignal (CObList) "SA4" 21 0 0 3068 3069 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #7=(CFSignal (CObList) "SA3" 21 0 0 3071 3072 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #8=(CFSignal (CObList) "SD1" 21 0 0 3181 3182 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #9=(CFSignal (CObList) "SA1" 21 0 0 3130 3131 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #10=(CFSignal (CObList) "SA9" 21 0 0 3434 3435 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #11=(CFSignal (CObList) "SD2" 21 0 0 3560 3561 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #12=(CFSignal (CObList) "SA7" 21 0 0 3600 3601 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #13=(CFSignal (CObList) "SD3" 21 0 0 3923 3924 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #14=(CFSignal (CObList) "SA21" 21 0 0 3932 3933 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #15=(CFSignal (CObList) "SA20" 21 0 0 3936 3937 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #16=(CFSignal (CObList) "SA6" 21 0 0 3950 3951 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #17=(CFSignal (CObList) "SA5" 21 0 0 4072 4073 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #18=(CFSignal (CObList) "SA2" 21 0 0 4344 4345 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #19=(CFSignal (CObList) "SD6" 21 0 0 4350 4351 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #20=(CFSignal (CObList) "SA10" 21 0 0 4364 4365 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #21=(CFSignal (CObList) "SA11" 21 0 0 4376 4377 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #22=(CFSignal (CObList) "SD4" 21 0 0 4453 4454 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #23=(CFSignal (CObList) "SD14" 21 0 0 6503 6504 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")
COMMENT        #24=(CFSignal (CObList) "SD15" 21 0 0 6523 6524 0 (CStringList)
COMMENT            (CStringList) 1 1 1 1 "")) (CObList) (CObList) (CObList) (CObList)
COMMENT       (CObList) (CObList)
```

Appendix - SEQ ID NO3

```
COMMENT         (CTextView 0
COMMENT          #25=(CGroupPar (CParagraph 0 (0 0) 1 2 0 0 180)
COMMENT               (CObjectList
COMMENT                #26=(CRefLinePar
COMMENT                     (CLinePar (CParagraph 0 (0 0) 0 2 0 0 233)
COMMENT
"#743.pCCLsin.PPT.hPGK.GFP.Wpre_mut_AMP_splicing_aggiornata_SIMPLIFIED"
COMMENT                      2) 5 "" 0 4)
COMMENT                #27=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 1 (0 0) 1 1 0 0 178)
COMMENT                      (CObjectList
COMMENT                       #28=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "DNA
'#743.pCCLsin.PPT.hPGK.GFP.Wpre_mut_AMP_splicing_aggiornata_SIMPLIFIED'"
COMMENT                            1)
COMMENT                       #29=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Currently local object. Original author: IRCC - ROSSO 3"
COMMENT                           1)
COMMENT                       #30=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Created: 04/20/12 07:58PM" 1)
COMMENT                       #31=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Last Modified: 04/20/12 08:13PM" 1)
COMMENT                       #32=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "length: 7830 bp" 1)
COMMENT                       #33=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "storage type: Basic" 1)
COMMENT                       #34=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "form: Circular" 1))) "General Description")
COMMENT                #35=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 2 (0 0) 1 1 0 0 178)
COMMENT                      (CObjectList
COMMENT                       #36=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Original Source Database: GenBank" 1)
COMMENT                       #37=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "Modification Date in the Original DB: 21-OCT-2005" 1)))
COMMENT                   "Standard Fields")
COMMENT                #38=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 5 (0 0) 1 1 0 0 178)
COMMENT                      (CObjectList
COMMENT                       #39=(CLinePar (CParagraph 0 (0 0) 1 2 1 0 180)
COMMENT                            "IRCC - ROSSO 3" 1))) "Original Author")
COMMENT                #40=(CRefLinePar
COMMENT                     (CLinePar (CParagraph 0 (0 0) 0 2 0 0 233) "Comments" 2) 1 ""
COMMENT                     0 0)
COMMENT                #41=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 8 (0 0) 1 2 0 0 178) (CObjectList))
COMMENT                     "Annotations")
COMMENT                #42=(CFolderPar
COMMENT                     (CGroupPar (CParagraph 12 (6 0) 1 2 0 0 178)
COMMENT                      (CObjectList
COMMENT                       #43=(CFolderPar
COMMENT                            (CGroupPar (CParagraph 4 (7 4 0) 1 1 1 0 178)
COMMENT                             (CObjectList
COMMENT                              #44=(CFolderPar
COMMENT                                   (CGroupPar
COMMENT                                    (CParagraph 5161 (3 #3# 0) 1 2 2 0 327)
COMMENT                                    (CObjectList
COMMENT                                     #45=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Start: 5161  End: 5880" 1)
COMMENT                                     #46=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Original Location Description:" 1)
COMMENT                                     #47=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "   5158..5877" 1))) "eGFP")))
COMMENT                          "CDS (1 total)")
COMMENT                       #48=(CFolderPar
COMMENT                            (CGroupPar (CParagraph 19 (7 19 0) 1 1 1 0 178)
COMMENT                             (CObjectList
COMMENT                              #49=(CFolderPar
COMMENT                                   (CGroupPar
COMMENT                                    (CParagraph 2891 (3 #2# 0) 1 2 2 0 194)
COMMENT                                    (CObjectList
COMMENT                                     #50=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Start: 2891  End: 3072" 1)
COMMENT                                     #51=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "Original Location Description:" 1)
COMMENT                                     #52=(CLinePar
COMMENT                                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                          "   2891..3072" 1))) "RU5")
COMMENT                               #53=(CFolderPar
COMMENT                                    (CGroupPar
COMMENT                                     (CParagraph 6571 (3 #1# 0) 1 2 2 0 194)
```

```
                    Appendix - SEQ ID NO3
COMMENT                (CObjectList
COMMENT                 #54=(CLinePar
COMMENT                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                      "Start: 6571  End: 6804" 1)
COMMENT                 #55=(CLinePar
COMMENT                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                      "Original Location Description:" 1)
COMMENT                 #56=(CLinePar
COMMENT                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                      "    6568..6801" 1))) "dR3RU5")))
COMMENT          "LTR (2 total)")
COMMENT       #57=(CFolderPar
COMMENT            (CGroupPar (CParagraph 21 (7 21 0) 1 2 1 0 178)
COMMENT             (CObjectList
COMMENT              #58=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 2974 (3 #5# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #59=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 2974  End: 2975" 1))) "SD5")
COMMENT              #60=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3068 (3 #6# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #61=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3068  End: 3069" 1))) "SA4")
COMMENT              #62=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3071 (3 #7# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #63=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3071  End: 3072" 1))) "SA3")
COMMENT              #64=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3130 (3 #9# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #65=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3130  End: 3131" 1))) "SA1")
COMMENT              #66=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3181 (3 #8# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #67=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3181  End: 3182" 1))) "SD1")
COMMENT              #68=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3434 (3 #10# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #69=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3434  End: 3435" 1))) "SA9")
COMMENT              #70=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3560 (3 #11# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #71=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3560  End: 3561" 1))) "SD2")
COMMENT              #72=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3600 (3 #12# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #73=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3600  End: 3601" 1))) "SA7")
COMMENT              #74=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3923 (3 #13# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #75=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3923  End: 3924" 1))) "SD3")
COMMENT              #76=(CFolderPar
COMMENT                   (CGroupPar
COMMENT                    (CParagraph 3932 (3 #14# 0) 1 2 2 0 194)
COMMENT                    (CObjectList
COMMENT                     #77=(CLinePar
COMMENT                          (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                          "Start: 3932  End: 3933" 1))) "SA21")
COMMENT              #78=(CFolderPar
COMMENT                   (CGroupPar
```

```
                                    Appendix - SEQ ID NO3
COMMENT                                (CParagraph 3936 (3 #15# 0) 1 2 2 0 194)
COMMENT                                (CObjectList
COMMENT                                 #79=(CLinePar
COMMENT                                      (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                      "Start: 3936  End: 3937" 1))) "SA20")
COMMENT                           #80=(CFolderPar
COMMENT                                (CGroupPar
COMMENT                                 (CParagraph 3950 (3 #16# 0) 1 2 2 0 194)
COMMENT                                 (CObjectList
COMMENT                                  #81=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 3950  End: 3951" 1))) "SA6")
COMMENT                           #82=(CFolderPar
COMMENT                                (CGroupPar
COMMENT                                 (CParagraph 4072 (3 #17# 0) 1 2 2 0 194)
COMMENT                                 (CObjectList
COMMENT                                  #83=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 4072  End: 4073" 1))) "SA5")
COMMENT                           #84=(CFolderPar
COMMENT                                (CGroupPar
COMMENT                                 (CParagraph 4344 (3 #18# 0) 1 2 2 0 194)
COMMENT                                 (CObjectList
COMMENT                                  #85=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 4344  End: 4345" 1))) "SA2")
COMMENT                           #86=(CFolderPar
COMMENT                                (CGroupPar
COMMENT                                 (CParagraph 4350 (3 #19# 0) 1 2 2 0 194)
COMMENT                                 (CObjectList
COMMENT                                  #87=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 4350  End: 4351" 1))) "SD6")
COMMENT                           #88=(CFolderPar
COMMENT                                (CGroupPar
COMMENT                                 (CParagraph 4364 (3 #20# 0) 1 2 2 0 194)
COMMENT                                 (CObjectList
COMMENT                                  #89=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 4364  End: 4365" 1))) "SA10")
COMMENT                           #90=(CFolderPar
COMMENT                                (CGroupPar
COMMENT                                 (CParagraph 4376 (3 #21# 0) 1 2 2 0 194)
COMMENT                                 (CObjectList
COMMENT                                  #91=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 4376  End: 4377" 1))) "SA11")
COMMENT                           #92=(CFolderPar
COMMENT                                (CGroupPar
COMMENT                                 (CParagraph 4453 (3 #22# 0) 1 2 2 0 194)
COMMENT                                 (CObjectList
COMMENT                                  #93=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 4453  End: 4454" 1))) "SD4")
COMMENT                           #94=(CFolderPar
COMMENT                                (CGroupPar
COMMENT                                 (CParagraph 5896 (3 #4# 0) 1 2 2 0 194)
COMMENT                                 (CObjectList
COMMENT                                  #95=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Start: 5896  End: 6487" 1)
COMMENT                                  #96=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Original Location Description:" 1)
COMMENT                                  #97=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "    5893..6484" 1)
COMMENT                                  #98=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "Qualifiers:" 1)
COMMENT                                  #99=(CLinePar
COMMENT                                       (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                       "    /db_xref=<a
HREF=\"genbank:ID_TheraBank_407\">ID_TheraBank_407</a>"
COMMENT                                       1)))
COMMENT                                "MUTATED Woodchuck hepatitis B virus PRE")
COMMENT                           #100=(CFolderPar
COMMENT                                 (CGroupPar
COMMENT                                  (CParagraph 6503 (3 #23# 0) 1 2 2 0 194)
COMMENT                                  (CObjectList
COMMENT                                   #101=(CLinePar
COMMENT                                         (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                         "Start: 6503  End: 6504" 1))) "SD14")
COMMENT                           #102=(CFolderPar
COMMENT                                 (CGroupPar
COMMENT                                  (CParagraph 6523 (3 #24# 0) 1 2 2 1 194)
```

```
                                    Appendix - SEQ ID NO3
COMMENT                                 (CObjectList
COMMENT                                  #103=(CLinePar
COMMENT                                        (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                        "Start: 6523  End: 6524" 1))) "SD15")))
COMMENT                           "Misc. Feature (21 total)")
COMMENT                      #104=(CFolderPar
COMMENT                            (CGroupPar (CParagraph 29 (7 29 0) 1 1 1 0 178)
COMMENT                             (CObjectList
COMMENT                              #105=(CFolderPar
COMMENT                                    (CGroupPar
COMMENT                                     (CParagraph 4625 (3 #0# 0) 1 2 2 0 194)
COMMENT                                     (CObjectList
COMMENT                                      #106=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "Start: 4625  End: 5140" 1)
COMMENT                                      #107=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "Original Location Description:" 1)
COMMENT                                      #108=(CLinePar
COMMENT                                            (CParagraph 0 (0 0) 1 2 3 0 180)
COMMENT                                            "   4622..5137" 1))) "hPGK")))
COMMENT                           "Promoter Eukaryotic (1 total)"))) "Feature Map"))))
COMMENT       (CGraphView
COMMENT        (CWStyleSheet
COMMENT         (CObjectList
COMMENT          #109=(CWidgetStyle "RSite Label" 1 (LOGPEN 0 0 8437824) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman")
COMMENT                0.611111 0 1 5 "@N (@S)" 0)
COMMENT          #110=(CWidgetStyle "Signal Label" 1 (LOGPEN 0 0 4227264) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.611111 0 1 1
COMMENT                "@N" 0)
COMMENT          #111=(CWidgetStyle "Molecule Label 2" 0 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 49 "Courier New") 0.888889 0
COMMENT                1 16 "@L bp" 0)
COMMENT          #112=(CWidgetStyle "Molecule Label 1" 0 0 1
COMMENT                (LOGFONT 0 0 0 0 700 0 0 0 0 3 2 1 34 "Arial") 0.833333 0 1 1
COMMENT                "@N" 0)
COMMENT          #113=(CWidgetStyle "Shape 3" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 0 26367 0)
COMMENT                0 0 1 (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT          #114=(CWidgetStyle "Shape 1" 1 (LOGPEN 0 0 0) 1 1
COMMENT                (LOGBRUSH 0 16728256 0) 0 0 1 (LOGSHAPE 8 1 0 0 0))
COMMENT          #115=(CWidgetStyle "Axis" 1 (LOGPEN 0 0 0) 2 1 (LOGBRUSH 0 12632256 0)
COMMENT                0 0 0)
COMMENT          #116=(CWidgetStyle "Line 2" 1 (LOGPEN 0 0 16711680) 8 0 0 0 0)
COMMENT          #117=(CWidgetStyle "RSite" 1 (LOGPEN 0 0 0) 8 0 0 0 0)
COMMENT          #118=(CWidgetStyle "Short Signal" 1 (LOGPEN 0 0 8388608) 10 0 0 0 0)
COMMENT          #119=(CWidgetStyle "Uniq RSite Label" 1 (LOGPEN 0 0 8421376) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman")
COMMENT                0.611111 128 1 5 "@N (@S)" 0)
COMMENT          #120=(CWidgetStyle "Vanilla" 1 (LOGPEN 0 0 0) 1 1
COMMENT                (LOGBRUSH 0 16777215 0) 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 7 48 2 18 "Times New Roman") 0.8 0
COMMENT                1 2 "?" 0)
COMMENT          #121=(CWidgetStyle "Mark 1" 0 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 2 7 48 2 2 "Windings") 0.7 0 1 2 "?"
COMMENT                0)
COMMENT          #122=(CWidgetStyle "Motif Label" 1 (LOGPEN 0 0 16744512) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.611111 8388608
COMMENT                1 65535 "@N (@H)" 0)
COMMENT          #123=(CWidgetStyle "Fragment Label 2" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 49 "Courier New") 1.05 0 1 48
COMMENT                "@F bp (molecule @L bp)" 0)
COMMENT          #124=(CWidgetStyle "Fragment Label 1" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.91 0 1 1
COMMENT                "Fragment of @N" 0)
COMMENT          #125=(CWidgetStyle "Shape 4" 1 (LOGPEN 0 0 0) 1 1
COMMENT                (LOGBRUSH 2 8388608 5) 0 0 0)
COMMENT          #126=(CWidgetStyle "Shape 2" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 0 128 0) 0
COMMENT                0 0)
COMMENT          #127=(CWidgetStyle "Shape 0" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 0 0 0) 0 0
COMMENT                0)
COMMENT          #128=(CWidgetStyle "ORF" 1 (LOGPEN 0 0 16384) 8 0 0 0 1
COMMENT                (LOGSHAPE 7 0.2 3.41182 2.86186 0.609808))
COMMENT          #129=(CWidgetStyle "Line 4" 1 (LOGPEN 0 0 32768) 8 0 0 0 0)
COMMENT          #130=(CWidgetStyle "Line 3" 1 (LOGPEN 0 0 16711680) 8 0 0 0 0)
COMMENT          #131=(CWidgetStyle "Line 1" 1 (LOGPEN 0 0 16711680) 1 0 0 0 0)
COMMENT          #132=(CWidgetStyle "Short Promoter" 1 (LOGPEN 0 0 128) 6 0 0 0 0)
COMMENT          #133=(CWidgetStyle "Motif" 1 (LOGPEN 0 0 0) 1 0 0 0 0)
COMMENT          #134=(CWidgetStyle "Line 0" 1 (LOGPEN 0 0 0) 8 0 0 0 0)
COMMENT          #135=(CWidgetStyle "Void" 0 0 0 0 0)
COMMENT          #136=(CWidgetStyle "General Label" 1 (LOGPEN 0 0 0) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman") 0.91 0
COMMENT                1 3 "@T @N " 0)
COMMENT          #137=(CWidgetStyle "Position" 1 (LOGPEN 0 0 0) 1 0 0 0 0)
COMMENT          #138=(CWidgetStyle "Annotation" 0 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman") 0.91 0
```

Appendix - SEQ ID NO3

```
COMMENT                0 0)
COMMENT          #139=(CWidgetStyle "Position Label" 1 (LOGPEN 0 0 8388608) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 1 0 0 3 2 1 34 "Arial") 0.63 8388608 1 1
COMMENT                "@N" 0)
COMMENT          #140=(CWidgetStyle "Range" 1 (LOGPEN 0 0 0) 1 1
COMMENT                (LOGBRUSH 0 16777215 0) 0 0 0)
COMMENT          #141=(CWidgetStyle "Range Label" 1 (LOGPEN 0 0 8388608) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 1 0 0 3 2 1 34 "Arial") 0.63 8388608 1 1
COMMENT                "@N" 0)
COMMENT          #142=(CWidgetStyle "ORF Label" 1 (LOGPEN 0 0 49216) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 18 "Times New Roman")
COMMENT                0.611111 0 1 65535 "@N" 0)
COMMENT          #143=(CWidgetStyle "CDS Label" 1 (LOGPEN 0 0 4227264) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.555556 255 1 1
COMMENT                "@N" 0)
COMMENT          #144=(CWidgetStyle "Shape 5" 1 (LOGPEN 0 0 0) 3 1
COMMENT                (LOGBRUSH 0 16777113 0) 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 7 48 2 50 "Arial") 0.9 0 0 1
COMMENT                (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT          #145=(CWidgetStyle "CDS" 1 (LOGPEN 0 0 0) 1 1 (LOGBRUSH 2 39423 3) 0 0
COMMENT                1 (LOGSHAPE 9 1 0.8 1.8 0))
COMMENT          #146=(CWidgetStyle "Label 2" 1 (LOGPEN 0 0 4227264) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.944444 8388608
COMMENT                1 1 "@N" 0)
COMMENT          #147=(CWidgetStyle "Label 3" 1 (LOGPEN 0 0 8421376) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 700 255 0 0 0 3 2 1 34 "Arial") 0.833333 255 1
COMMENT                5 "@N (@S)" 0)
COMMENT          #148=(CWidgetStyle "Label 4" 1 (LOGPEN 0 0 8437824) 1 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.722222 0 1 5
COMMENT                "@N (@S)" 0)
COMMENT          #149=(CWidgetStyle "Splicing_Signal Label" 1 (LOGPEN 0 0 39372) 8 0 1
COMMENT                (LOGFONT 0 0 0 0 400 0 0 0 0 3 2 1 34 "Arial") 0.611111 0 1 1
COMMENT                "@N" 0)) 0.164644 1.74233 0.164644 2.53336
COMMENT          (27 (CShapeMapEntry 2 "Line 2" 1 "Signal Label") 38
COMMENT           (CShapeMapEntry 0 "Shape 3" 1 "Splicing_Signal Label") 70
COMMENT           (CShapeMapEntry 0 "Unique RSite" 1 "Uniq RSite Label") 67
COMMENT           (CShapeMapEntry 0 "ORF" 0 "ORF Label")) 47.532 47.532 74.0391 74.0391
COMMENT          0.1 -7830) 1 0 1 1 1
COMMENT         (mapper: 15.6234 -14.7018 74.0391 74.0391 0.01 10 14 7830 7830 1 0 0)
COMMENT         #150=(CGroupWidget (CWidget 0 (0 0) 1 2 0 0 Nil 0 100)
COMMENT              (CObjectList
COMMENT               #151=(CGroupWidget (CWidget 1 (0 0) 1 2 0 0 Nil -970 100)
COMMENT                     (CObjectList
COMMENT                      #152=(CAxis
COMMENT                           (CWideLine
COMMENT                            (CWidget 0 (0 0) 1 2 0 0 #115# 17221828 0)
COMMENT                            (LOGPEN 0 2 0) 2 (LOGBRUSH 0 12632256 0) 1 6.27499
COMMENT                            6.27299 1 0.0214037) 0.0801173)
COMMENT                      #153=(CLabel (CWidget 1001 (0 0) 1 2 0 0 #112# 0 100)
COMMENT                           (LOGPEN 0 0 4227264) 1
COMMENT                           (LOGFONT 78 29 0 0 700 0 0 0 0 3 2 1 34 "Arial")
COMMENT                           2.53336 0.833333 0
COMMENT                           "#743.pCCLsin.PPT.hPGK.GFP.Wpre mut AMP splicing aggiornata
SIMPLIFIED"
COMMENT                           "@N" 1 0 0.5 0 -2.91312 29.3359 1.0535 Nil)
COMMENT                      #154=(CLabel
COMMENT                           (CWidget 1002 (0 0) 1 2 0 0 #111# 1852404556 100)
COMMENT                           (LOGPEN 0 0 4227264) 1
COMMENT                           (LOGFONT 83 31 0 0 400 0 0 0 0 3 2 1 49
COMMENT                            "Courier New") 2.53336 0.888889 0 "7830 bp" "@L bp"
COMMENT                           16 0 -0.8 0 -4.1798 3.10647 1.14804 Nil))
COMMENT                     (CObjectList))
COMMENT               #155=(CGroupWidget (CWidget 10 (6 0) 1 2 0 0 Nil -782 100)
COMMENT                     (CObjectList
COMMENT                      #156=(CGroupWidget
COMMENT                           (CWidget 4 (7 4 0) 1 2 0 0 Nil -782 100)
COMMENT                           (CObjectList
COMMENT                            #157=(CWideArrow
COMMENT                                 (CWideLine
COMMENT                                  (CWidget 0 (3 #3# 0) 1 2 0 0 #113# 0 100)
COMMENT                                  (LOGPEN 0 0 0) 1 (LOGBRUSH 0 26367 0) 1
COMMENT                                  1.56229 2.13913 1 0.082322) 0.8 1.8 0)
COMMENT                            #158=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                                 (LOGPEN 0 0 4227264) 1
COMMENT                                 (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                                  "Arial") 2.53336 0.611111 0 "eGFP" "@N" 1 0
COMMENT                                 0 6.77312 1.97573 1.67479 0.756357 #157#))
COMMENT                           (CObjectList))
COMMENT                      #159=(CGroupWidget
COMMENT                           (CWidget 19 (7 19 0) 1 2 0 0 Nil -876 100)
COMMENT                           (CObjectList
COMMENT                            #160=(CWideArrow
COMMENT                                 (CWideLine
COMMENT                                  (CWidget 0 (3 #1# 0) 1 2 0 0 #125#
COMMENT                                   1667321128 100) (LOGPEN 0 0 0) 1
COMMENT                                  (LOGBRUSH 2 8388608 5) 1 0.822004 1.00948 1
```

```
                    Appendix - SEQ ID NO3
COMMENT                  0.082322) 0.8 1.8 0)
COMMENT               #161=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0 "dR3RU5" "@N" 1
COMMENT                     0 0 11.4032 1.97573 2.3366 0.756357 #160#)
COMMENT               #162=(CWideArrow
COMMENT                     (CWideLine
COMMENT                      (CWidget 0 (3 #2# 0) 1 2 0 0 #125# 0 100)
COMMENT                      (LOGPEN 0 0 0) 1 (LOGBRUSH 2 8388608 5) 1
COMMENT                      3.81198 3.9578 1 0.082322) 0.8 1.8 0)
COMMENT               #163=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0 "RU5" "@N" 1 0 0
COMMENT                     -4.7195 3.24241 1.24259 0.756357 #162#))
COMMENT               (CObjectList))
COMMENT          #164=(CGroupWidget
COMMENT               (CWidget 21 (7 21 0) 1 2 0 0 Nil -688 100)
COMMENT               (CObjectList
COMMENT                #165=(CWideArrow
COMMENT                     (CWideLine
COMMENT                      (CWidget 0 (3 #4# 0) 1 2 0 0 #114#
COMMENT                       1952543860 100) (LOGPEN 0 0 0) 1
COMMENT                      (LOGBRUSH 0 16728256 0) 1 1.07598 1.55027 1
COMMENT                      0.082322) 0.8 1.8 0)
COMMENT                #166=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0
COMMENT                     "MUTATED Woodchuck hepatitis B virus PRE"
COMMENT                     "@N" 1 0 0 14.6543 5.77577 12.4934 0.756357
COMMENT                     #165#)
COMMENT                #167=(CScratch
COMMENT                     (CWidget 0 (3 #5# 0) 1 2 0 0 #118# 7407 100)
COMMENT                     (LOGPEN 0 14 8388608) 10 1 3.8913 1.9
COMMENT                     0.082322 1)
COMMENT                #168=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0 "SD5" "@N" 1 0 0
COMMENT                     -4.74716 1.97573 1.24259 0.756357 #167#)
COMMENT                #169=(CScratch
COMMENT                     (CWidget 0 (3 #6# 0) 1 2 0 0 #118# 7392 100)
COMMENT                     (LOGPEN 0 14 8388608) 10 1 3.81599 1.9
COMMENT                     0.082322 1)
COMMENT                #170=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0 "SA4" "@N" 1 0 0
COMMENT                     -4.42223 4.50909 1.24259 0.756357 #169#)
COMMENT                #171=(CScratch
COMMENT                     (CWidget 0 (3 #7# 0) 1 2 0 0 #118# 7377 100)
COMMENT                     (LOGPEN 0 14 8388608) 10 1 3.81358 1.9
COMMENT                     0.082322 1)
COMMENT                #172=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0 "SA3" "@N" 1 0 0
COMMENT                     -4.41186 5.77577 1.24259 0.756357 #171#)
COMMENT                #173=(CScratch
COMMENT                     (CWidget 0 (3 #8# 0) 1 2 0 0 #118# 7362 100)
COMMENT                     (LOGPEN 0 14 8388608) 10 1 3.72545 1.9
COMMENT                     0.082322 1)
COMMENT                #174=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0 "SD1" "@N" 1 0 0
COMMENT                     -4.03163 8.30913 1.24259 0.756357 #173#)
COMMENT                #175=(CScratch
COMMENT                     (CWidget 0 (3 #9# 0) 1 2 0 0 #118# 7347 100)
COMMENT                     (LOGPEN 0 14 8388608) 10 1 3.76631 1.9
COMMENT                     0.082322 1)
COMMENT                #176=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0 "SA1" "@N" 1 0 0
COMMENT                     -4.20792 7.04245 1.24259 0.756357 #175#)
COMMENT                #177=(CScratch
COMMENT                     (CWidget 0 (3 #10# 0) 1 2 0 0 #118# 7332 100)
COMMENT                     (LOGPEN 0 14 8388608) 10 1 3.52276 1.9
COMMENT                     0.082322 1)
COMMENT                #178=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                     (LOGPEN 0 0 4227264) 1
COMMENT                     (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                      "Arial") 2.53336 0.611111 0 "SA9" "@N" 1 0 0
```

Appendix - SEQ ID NO3

```
COMMENT          -3.15709 9.57581 1.24259 0.756357 #177#)
COMMENT      #179=(CScratch
COMMENT          (CWidget 0 (3 #11# 0) 1 2 0 0 #118# 7317 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 3.42181 1.9
COMMENT          0.082322 1)
COMMENT      #180=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SD2" "@N" 1 0 0
COMMENT          -2.72156 10.8425 1.24259 0.756357 #179#)
COMMENT      #181=(CScratch
COMMENT          (CWidget 0 (3 #12# 0) 1 2 0 0 #118# 7302 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 3.38976 1.9
COMMENT          0.082322 1)
COMMENT      #182=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 1 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SA7" "@N" 1 0 0
COMMENT          -2.58329 12.1092 1.24259 0.756357 #181#)
COMMENT      #183=(CScratch
COMMENT          (CWidget 0 (3 #13# 0) 1 2 0 0 #118# 7287 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 3.13098 1.9
COMMENT          0.082322 1)
COMMENT      #184=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SD3" "@N" 1 0 0
COMMENT          0.670072 13.3758 1.24259 0.756357 #183#)
COMMENT      #185=(CScratch
COMMENT          (CWidget 0 (3 #14# 0) 1 2 0 0 #118# 7272 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 3.12377 1.9
COMMENT          0.082322 1)
COMMENT      #186=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SA21" "@N" 1 0
COMMENT          0 0.872057 12.1092 1.58025 0.756357 #185#)
COMMENT      #187=(CScratch
COMMENT          (CWidget 0 (3 #15# 0) 1 2 0 0 #118# 7257 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 3.12057 1.9
COMMENT          0.082322 1)
COMMENT      #188=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SA20" "@N" 1 0
COMMENT          0 0.886793 10.8425 1.58025 0.756357 #187#)
COMMENT      #189=(CScratch
COMMENT          (CWidget 0 (3 #16# 0) 1 2 0 0 #118# 7242 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 3.10935 1.9
COMMENT          0.082322 1)
COMMENT      #190=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 1 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SA6" "@N" 1 0 0
COMMENT          0.769538 9.57581 1.24259 0.756357 #189#)
COMMENT      #191=(CScratch
COMMENT          (CWidget 0 (3 #17# 0) 1 2 0 0 #118# 7227 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 3.01161 1.9
COMMENT          0.082322 1)
COMMENT      #192=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SA5" "@N" 1 0 0
COMMENT          1.21898 8.30913 1.24259 0.756357 #191#)
COMMENT      #193=(CScratch
COMMENT          (CWidget 0 (3 #18# 0) 1 2 0 0 #118# 7212 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 2.79369 1.9
COMMENT          0.082322 1)
COMMENT      #194=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SA2" "@N" 1 0 0
COMMENT          2.22101 7.04245 1.24259 0.756357 #193#)
COMMENT      #195=(CScratch
COMMENT          (CWidget 0 (3 #19# 0) 1 2 0 0 #118# 7197 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 2.78888 1.9
COMMENT          0.082322 1)
COMMENT      #196=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 1 100)
COMMENT          (LOGPEN 0 0 4227264) 1
COMMENT          (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT           "Arial") 2.53336 0.611111 0 "SD6" "@N" 1 0 0
COMMENT          2.24312 5.77577 1.24259 0.756357 #195#)
COMMENT      #197=(CScratch
COMMENT          (CWidget 0 (3 #20# 0) 1 2 0 0 #118# 7182 100)
COMMENT          (LOGPEN 0 14 8388608) 10 1 2.77767 1.9
COMMENT          0.082322 1)
```

Appendix - SEQ ID NO3

```
COMMENT                                #198=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                                      (LOGPEN 0 0 4227264) 1
COMMENT                                      (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                                       "Arial") 2.53336 0.611111 0 "SA10" "@N" 1 0
COMMENT                                      0 2.46352 4.50909 1.58025 0.756357 #197#)
COMMENT                                #199=(CScratch
COMMENT                                      (CWidget 0 (3 #21# 0) 1 2 0 0 #118# 7167 100)
COMMENT                                      (LOGPEN 0 14 8388608) 10 1 2.76805 1.9
COMMENT                                      0.082322 1)
COMMENT                                #200=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                                      (LOGPEN 0 0 4227264) 1
COMMENT                                      (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                                       "Arial") 2.53336 0.611111 0 "SA11" "@N" 1 0
COMMENT                                      0 2.50773 3.24241 1.58025 0.756357 #199#)
COMMENT                                #201=(CScratch
COMMENT                                      (CWidget 0 (3 #22# 0) 1 2 0 0 #118# 7152 100)
COMMENT                                      (LOGPEN 0 14 8388608) 10 1 2.70636 1.9
COMMENT                                      0.082322 1)
COMMENT                                #202=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                                      (LOGPEN 0 0 4227264) 1
COMMENT                                      (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                                       "Arial") 2.53336 0.611111 0 "SD4" "@N" 1 0 0
COMMENT                                      2.62256 1.97573 1.24259 0.756357 #201#)
COMMENT                                #203=(CScratch
COMMENT                                      (CWidget 0 (3 #23# 0) 1 2 0 0 #118# 7137 100)
COMMENT                                      (LOGPEN 0 14 8388608) 10 1 1.06396 1.9
COMMENT                                      0.082322 1)
COMMENT                                #204=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 1 100)
COMMENT                                      (LOGPEN 0 0 4227264) 1
COMMENT                                      (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                                       "Arial") 2.53336 0.611111 0 "SD14" "@N" 1 0
COMMENT                                      0 10.3435 4.50909 1.58025 0.756357 #203#)
COMMENT                                #205=(CScratch
COMMENT                                      (CWidget 0 (3 #24# 0) 1 2 0 0 #118# 7122 100)
COMMENT                                      (LOGPEN 0 14 8388608) 10 1 1.04793 1.9
COMMENT                                      0.082322 1)
COMMENT                                #206=(CLabel
COMMENT                                      (CWidget 0 (0 0) 1 2 0 0 #110# 39423 100)
COMMENT                                      (LOGPEN 0 0 4227264) 1
COMMENT                                      (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                                       "Arial") 2.53336 0.611111 0 "SD15" "@N" 1 0
COMMENT                                      0 10.4172 3.24241 1.58025 0.756357 #205#))
COMMENT                                (CObjectList))
COMMENT                          #207=(CGroupWidget
COMMENT                                (CWidget 29 (7 29 0) 1 2 0 0 Nil -970 100)
COMMENT                                (CObjectList
COMMENT                                 #208=(CWideArrow
COMMENT                                       (CWideLine
COMMENT                                        (CWidget 0 (3 #0# 0) 1 2 0 0 #114# 0 100)
COMMENT                                        (LOGPEN 0 0 0) 1 (LOGBRUSH 0 16728256 0) 1
COMMENT                                        2.15516 2.56856 1 0.082322) 0.8 1.8 0)
COMMENT                                 #209=(CLabel (CWidget 0 (0 0) 1 2 0 0 #110# 0 100)
COMMENT                                       (LOGPEN 0 0 4227264) 1
COMMENT                                       (LOGFONT 57 21 0 0 400 0 0 0 0 3 2 1 34
COMMENT                                        "Arial") 2.53336 0.611111 0 "hPGK" "@N" 1 0
COMMENT                                       0 4.44302 1.97573 1.71531 0.756357 #208#))
COMMENT                                (CObjectList))) (CObjectList))
COMMENT                     #210=(CGroupWidget (CWidget 14 (16 0) 1 2 0 0 Nil -595 100)
COMMENT                           (CObjectList) (CObjectList))
COMMENT                     #211=(CGroupWidget (CWidget 11 (0 0) 1 2 0 0 Nil -501 100)
COMMENT                           (CObjectList) (CObjectList))
COMMENT                     #212=(CGroupWidget (CWidget 12 (0 0) 1 2 0 0 Nil -407 100)
COMMENT                           (CObjectList) (CObjectList))) (CObjectList)))
COMMENT          (CSeqView 10 6 (CObjectList) (CObList) 1 (CObList)) (CObList) -2147483648
COMMENT          (CStringList) 2147483648 268632118 (CObList)))
FEATURES             Location/Qualifiers
     promoter        4625..5140
                     /vntifkey="29"
                     /label=hPGK
     LTR             6571..6804
                     /vntifkey="19"
                     /label=dR3RU5
     LTR             2891..3072
                     /vntifkey="19"
                     /label=RU5
     CDS             5161..5880
                     /vntifkey="4"
                     /label=eGFP
     misc_feature    5896..6487
                     /db_xref="ID_TheraBank_407"
                     /vntifkey="21"
                     /label=MUTATED\Woodchuck\hepatitis\B\virus\PRE
     misc_feature    3181..3182
                     /vntifkey="21"
                     /label=SD1
     misc_feature    3130..3131
```

Appendix - SEQ ID NO3

```
                     /vntifkey="21"
                     /label=SA1
     misc_feature    3434..3435
                     /vntifkey="21"
                     /label=SA9
     misc_feature    3560..3561
                     /vntifkey="21"
                     /label=SD2
     misc_feature    3932..3933
                     /vntifkey="21"
                     /label=SA21
     misc_feature    3923..3924
                     /vntifkey="21"
                     /label=SD3
     misc_feature    3936..3937
                     /vntifkey="21"
                     /label=SA20
     misc_feature    3950..3951
                     /vntifkey="21"
                     /label=SA6
     misc_feature    4072..4073
                     /vntifkey="21"
                     /label=SA5
     misc_feature    4344..4345
                     /vntifkey="21"
                     /label=SA2
     misc_feature    4350..4351
                     /vntifkey="21"
                     /label=SD6
     misc_feature    4364..4365
                     /vntifkey="21"
                     /label=SA10
     misc_feature    4376..4377
                     /vntifkey="21"
                     /label=SA11
     misc_feature    4453..4454
                     /vntifkey="21"
                     /label=SD4
     misc_feature    6503..6504
                     /vntifkey="21"
                     /label=SD14
     misc_feature    6523..6524
                     /vntifkey="21"
                     /label=SD15
     misc_feature    3071..3072
                     /vntifkey="21"
                     /label=SA3
     misc_feature    3068..3069
                     /vntifkey="21"
                     /label=SA4
     misc_feature    2974..2975
                     /vntifkey="21"
                     /label=SD5
     misc_feature    3600..3601
                     /vntifkey="21"
                     /label=SA7
BASE COUNT     1957 a      1973 c      2023 g      1877 t
ORIGIN
        1 caggtggcac ttttcgggga aatgtgcgcg gaacccctat ttgtttattt ttctaaatac
       61 attcaaatat gtatccgctc atgagacaat aaccctgata aatgcttcaa taatattgaa
      121 aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt tttgcggcat
      181 tttgccttcc tgtttttgct cacccagaaa cgctggtgaa agtaaaagat gctgaagatc
      241 agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag atccttgaga
      301 gttttcgccc cgaagaacgt tttccaatga tgagcacttt taaagttctg ctatgtggcg
      361 cggtattatc ccgtattgac gccgggcaag agcaactcgg tcgccgcata cactattctc
      421 agaatgactt ggttgagtac tcaccagtca cagaaaagca tcttacggat ggcatgacag
      481 taagagaatt atgcagtgct gccataacca tgagtgataa cactgcggcc aacttacttc
      541 tgacaacgat cggaggaccg aaggagctaa ccgcttttt gcacaacatg ggggatcatg
      601 taactcgcct tgatcgttgg gaaccggagc tgaatgaagc cataccaaac gacgagcgtg
      661 acaccacgat gcctgtagca atggcaacaa cgttgcgcaa actattaact ggcgaactac
      721 ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa gttgcaggac
      781 cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct ggagccggtg
      841 agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc tcccgtatcg
      901 tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga cagatcgctg
      961 agataggtgc ctcactgatt aagcattggt aactgtcaga ccaagtttac tcatatatac
     1021 tttagattga tttaaaactt catttttaat ttaaaaggat ctaggtgaag atccttttg
     1081 ataatctcat gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg tcagaccccg
     1141 tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc
     1201 aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag ctaccaactc
     1261 tttttccgaa ggtaactggc ttcagcagag cgcagatacc aaatactgtc cttctagtgt
     1321 agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac ctcgctctgc
     1381 taatcctgtt accagtggct gctgccagtg gcgataagtc gtgtcttacc gggttggact
     1441 caagacgata gttaccggat aaggcgcagc ggtcgggctg aacgggggt tcgtgcacac
     1501 agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt gagctatgag
     1561 aaagcgccac gcttcccgaa gggagaaagg cggacaggta tccggtaagc ggcagggtcg
     1621 gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg
```

Appendix - SEQ ID NO3

1681 tcgggtttcg ccacctctga cttgagcgtc gatttttgtg atgctcgtca gggggggcgga
1741 gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt tgctggcctt
1801 ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt attaccgcct
1861 ttgagtgagc tgataccgct cgccgcagcc gaacgaccga gcgcagcgag tcagtgagcg
1921 aggaagcgga agagcgccca atacgcaaac cgcctctccc cgcgcgttgg ccgattcatt
1981 aatgcagctg gcacgacagg tttcccgact ggaaagcggg cagtgagcgc aacgcaatta
2041 atgtgagtta gctcactcat taggcacccc aggctttaca ctttatgctt ccggctcgta
2101 tgttgtgtgg aattgtgagc ggataacaat ttcacacagg aaacagctat gaccatgatt
2161 acgccaagcg cgcaattaac cctcactaaa gggaacaaaa gctggagctg caagcttggc
2221 cattgcatac gttgtatcca tatcataata tgtacattta tattggctca tgtccaacat
2281 taccgccatg ttgacattga ttattgacta gttattaata gtaatcaatt acggggtcat
2341 tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaat ggcccgcctg
2401 gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatgtt cccatagtaa
2461 cgccaatagg gactttccat tgacgtcaat gggtggagta tttacggtaa actgcccact
2521 tggcagtaca tcaagtgtat catatgccaa gtacgccccc tattgacgtc aatgacggta
2581 aatggcccgc ctggcattat gcccagtaca tgaccttatg ggactttcct acttggcagt
2641 acatctacgt attagtcatc gctattacca tggtgatgcg gttttggcag tacatcaatg
2701 ggcgtggata gcggtttgac tcacggggat ttccaagtct ccaccccatt gacgtcaatg
2761 ggagtttgtt ttggcaccaa aatcaacggg actttccaaa atgtcgtaac aactccgccc
2821 cattgacgca aatgggcggt aggcgtgtac ggtgggaggt ctatataagc agagctcgtt
2881 tagtgaaccg gggtctctct ggttagacca gatctgagcc tgggagctct ctggctaact
2941 agggaaccca ctgcttaagc ctcaataaag cttgccttga gtgcttcaag tagtgtgtgc
3001 ccgtctgttg tgtgactctg gtaactagag atcccctcaga ccctttttagt cagtgtggaa
3061 aatctctagc agtggcgccc gaacagggac ttgaaagcga aagggaaacc agaggagctc
3121 tctcgacgca ggactcggct tgctgaagcg cgcacggcaa gaggcgaggg gcggcgactg
3181 gtgagtacgc caaaaatttt gactagcgga ggctagaagg agagagatgg gtgcgagagc
3241 gtcagtatta agcgggggag aattagatcg cgatgggaaa aaattcggtt aaggccaggg
3301 ggaaagaaaa aatataaatt aaaacatata gtatgggcaa gcagggagct agaacgattc
3361 gcagttaatc ctggcctgtt agaaacatca gaaggctgta gacaaatact gggacagcta
3421 caaccatccc ttcagacagg atcagaagaa cttagatcat tatataatac agtagcaacc
3481 ctctattgtg tgcatcaaag gatagagata aaagacacca aggaagcttt agacaagata
3541 gaggaagagc aaaacaaaag taagaccacc gcacagcaag cggccgctga tcttcagacc
3601 tggaggagga gatatgaggg acaattggag aagtgaatta tataaatata aagtagtaaa
3661 aattgaacca ttaggagtag cacccaccaa ggcaaagaga agagtggtgc agagagaaaa
3721 aagagcagtg ggaataggag ctttgttcct tgggttcttg ggagcagcag gaagcactat
3781 gggcgcagcc tcaatgacgc tgacggtaca ggccagacaa ttattgtctg gtatagtgca
3841 gcagcagaac aatttgctga gggctattga ggcgcaacag catctgttgc aactcacagt
3901 ctggggcatc aagcagctcc aggcaagaat cctggctgtg gaaagatacc taaaggatca
3961 acagctcctg gggatttggg gttgctctgg aaaactcatt tgcaccactg ctgtgccttg
4021 gaatgctagt tggagtaata aatctctgga acagatttgg aatcacacga cctggatgga
4081 gtgggacaga gaaattaaca attacacaag cttaatacac tccttaattg aagaatcgca
4141 aaaccagcaa gaaaagaatg aacaagaatt attggaatta gataaatggg caagtttgtg
4201 gaattggttt aacataacaa attggctgtg gtatataaaa ttattcataa tgatagtagg
4261 aggcttggta ggtttaagaa tagtttttgc tgtactttct atagtgaata gagttaggca
4321 gggatattca ccattatcgt ttcagaccca cctcccaacc ccgaggggac ccgacaggcc
4381 cgaaggaata gaagaagaag gtggagagag agacagagac agatccattc gattagtgaa
4441 cggatctcga cggtatcggt taacttttaa aagaaaaggg gggattgggg ggtacagtgc
4501 aggggaaaga atagtagaca taatagcaac agacatacaa actaaagaat tacaaaaaca
4561 aattacaaaa attcaaaatt ttatcgatca cgagactagc ctcgagaagc ttgatatcga
4621 attcccacgg ggttggggtt gcgcctttc caaggcagcc ctgggtttgc gcagggacgc
4681 ggctgctctg ggcgtggttc cgggaaacgc agcggcgccg accctgggtc tcgcacattc
4741 ttcacgtccg ttcgcagcgt cacccggatc ttcgccgcta cccttgtggg ccccccggcg
4801 acgcttcctg ctccgcccct aagtcgggaa ggttccttgc ggttcgcggc gtgccggacg
4861 tgacaaacgg aagccgcacg tctcactagt accctcgcag acggacagcg ccagggagca
4921 atggcagcgc gccgaccgcg atgggctgtg gccaatagcg gctgctcagc ggggcgcgcc
4981 gagagcagcg gccgggaagg ggcggtgcgg gaggcggggt gtggggcggt agtgtgggcc
5041 ctgttcctgc ccgcgcggtg ttccgcattc tgcaagcctc cggagcgcac gtcggcagtc
5101 ggctcccctcg ttgaccgaat caccgacctc tctccccagg gggatccacc ggtcgccacc
5161 atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac
5221 ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac
5281 ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc
5341 ctcgtgacca ccctgaccta cggcgtgcag tgcttcagcc gctaccccga ccacatgaag
5401 cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc
5461 ttcaaggacg acggcaacta caagacccgc gccgaggtga agttcgaggg cgacaccctg
5521 gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac
5581 aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac
5641 ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc
5701 gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac
5761 tacctgagca cccagtccgc cctgagcaaa gaccccaacg agaagcgcga tcacatggtc
5821 ctgctggagt tcgtgaccgc cgccgggatc actctcggca tggacgagct gtacaagtaa
5881 agcggccgcg tcgacaatca acctctggat tacaaaattt gtgaaagatt gactggtatt
5941 cttaactatg ttgctccttt tacgctatgt ggatacgctg ctttaatgcc tttgtatcat
6001 gctattgctt cccgtatggc tttcattttc tcctccttgt ataaatcctg gttgctgtct
6061 ctttatgagg agttgtggcc cgttgtcagg caacgtggcg tggtgtgcac tgtgtttgct
6121 gacgcaaccc ccactggttg gggcattgcc accacctgtc agctcctttc cgggactttc
6181 gctttccccc tccctattgc cacggcggaa ctcatcgccg cctgccttgc ccgctgctgg
6241 acaggggctc ggctgttggg cactgacaat tccgtggtgt tgtcggggaa atcatcgtcc
6301 tttccttggc tgctcgcctg tgttgccacc tggattctgc gcgggacgtc cttctgctac
6361 gtcccttcgg ccctcaatcc agcggacctt ccttcccgcg gcctgctgcc ggctctgcgg
6421 cctcttccgc gtcttcgcct tcgccctcag acgagtcgga tctccctttg ggccgcctcc
6481 ccgcctggaa ttcgagctcg gtacctttaa gaccaatgac ttacaaggca gctgtagatc
6541 ttagccactt tttaaaagaa aaggggggac tggaagggct aattcactcc caacgaagac
6601 aagatctgct ttttgcttgt actgggtctc tctggttaga ccagatctga gcctgggagc
6661 tctctggcta actagggaac ccactgctta agcctcaata aagcttgcct tgagtgcttc
6721 aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc agacccttttt
6781 agtcagtgtg gaaaatctct agcagtagta gttcatgtca tcttattatt cagtatttat

Appendix - SEQ ID NO3

```
6841 aacttgcaaa gaaatgaata tcagagagtg agaggaactt gtttattgca gcttataatg
6901 gttacaaata aagcaatagc atcacaaatt tcacaaataa agcatttttt tcactgcatt
6961 ctagttgtgg tttgtccaaa ctcatcaatg tatcttatca tgtctggctc tagctatccc
7021 gcccctaact ccgcccagtt ccgcccattc tccgccccat ggctgactaa ttttttttat
7081 ttatgcagag gccgaggccg cctcggcctc tgagctattc cagaagtagt gaggaggctt
7141 ttttggaggc ctaggctttt gcgtcgagac gtacccaatt cgccctatag tgagtcgtat
7201 tacgcgcgct cactggccgt cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc
7261 caacttaatc gccttgcagc acatcccccc ttcgccagct ggcgtaatag cgaagaggcc
7321 cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggcg cgacgcgccc
7381 tgtagcggcg cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt
7441 gccagcgccc tagcgcccgc tcctttcgct ttcttccctt cctttctcgc cacgttcgcc
7501 ggctttcccc gtcaagctct aaatcggggg ctccctttag ggttccgatt tagtgcttta
7561 cggcacctcg accccaaaaa acttgattag ggtgatggtt cacgtagtgg gccatcgccc
7621 tgatagacgg tttttcgccc tttgacgttg gagtccacgt tctttaatag tggactcttg
7681 ttccaaactg gaacaacact caaccctatc tcggtctatt cttttgattt ataagggatt
7741 ttgccgattt cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat
7801 tttaacaaaa tattaacgtt tacaatttcc
```

//

Claims

1. Use of a lentiviral vector containing a lentiviral backbone in which at least two of the splice sites have been eliminated to improve the safety profile of the lentiviral vector.

2. Use according to claim 1 wherein the ability of the lentiviral vector to generate a lentiviral sequence fused to a cellular transcript is reduced.

3. A polynucleotide sequence comprising a lentiviral nucleotide sequence wherein at least one of the following splice sites (forward and/or reverse) is inactivated:

SPLICE ACCEPTOR GROUP 1

SA1 - corresponding to nucleotides 3127-3128 of SEQ ID NO:1 or nucleotides 3130-3131 of SEQ ID NO:3

SA2 – corresponding to nucleotides 4341-4342 of SEQ ID NO:1 or nucleotides 4344-4345 of SEQ ID NO:3.

SA3 – corresponding to nucleotides 3071-3072 of SEQ ID NO:1 or nucleotides 3071-3072 of SEQ ID NO:3.

SA4 – corresponding to nucleotides 3068-3069 of SEQ ID NO:1 or nucleotides 3068-3069 of SEQ ID NO:3.

SA5 - corresponding to nucleotides 4069-4070 of SEQ ID NO:1 or nucleotides 4072-4073 of SEQ ID NO:3.

SA6 - corresponding to nucleotides 3947-3948 of SEQ ID NO:1 or nucleotides 3950-3951 of SEQ ID NO:3.

SA7 - corresponding to nucleotides 3597-3598 (complement) of SEQ ID NO:1 or nucleotides 3600-3601 (complement) of SEQ ID NO:3.

SA9 - corresponding to nucleotides 3431-3432 of SEQ ID NO:1 or nucleotides 3434-3435 of SEQ ID NO:3.

SA10 - corresponding to nucleotides 4361-4362 of SEQ ID NO:1 or nucleotides 4364-4365 of SEQ ID NO:3.

SA11 - corresponding to nucleotides 4373-4374 of SEQ ID NO:1 or nucleotides 4376-4377 of SEQ ID NO:3.

SA20 - corresponding to nucleotides 3933-3934 (complement) of SEQ ID NO:1 or nucleotides 3936-3937 (complement) of SEQ ID NO:3.

SA21 - corresponding to nucleotides 3929-3930 (complement) of SEQ ID NO:1 or nucleotides 3932-3933 (complement) of SEQ ID NO:3.

SPLICE DONOR GROUP 1

SD1 - corresponding to nucleotides 3178-3179 of SEQ ID NO:1 or nucleotides 3181-3182 of SEQ ID NO:3.

SD2 - corresponding to nucleotides 3557-3558 of SEQ ID NO:1 or nucleotides 3560-3561 of SEQ ID NO:3.

SD3 - corresponding to nucleotides 3920-3921 of SEQ ID NO:1 or nucleotides 3923-3924 of SEQ ID NO:3.

SD4 - corresponding to nucleotides 4450-4451 of SEQ ID NO:1 or nucleotides 4453-4454 of SEQ ID NO:3.

SD5- corresponding to nucleotides 2974-2975 (complement) of SEQ ID NO:1 or nucleotides 2974-2975 (complement) of SEQ ID NO:3.

SD6- corresponding to nucleotide 4347-4348 (complement) of SEQ ID NO:1 or nucleotides 4350-4351 (complement) of SEQ ID NO:3.

SD14- corresponding to nucleotides 6500-6501(complement) of SEQ ID NO:1 or nucleotides 6503-6504 (complement) of SEQ ID NO:3.

SD15- corresponding to nucleotides 6520-6521(complement) of SEQ ID NO:1 or nucleotides 6523-6524 (complement) of SEQ ID NO:3.

SPLICE ACCEPTOR GROUP 2

SA1 - corresponding to nucleotides 3040-3041 of SEQ ID NO:1 or nucleotides 3040-3041 of SEQ ID NO:3.

SA4 - corresponding to nucleotides 3077-3078 of SEQ ID NO:1 or nucleotides 3077-3078 of SEQ ID NO:3.

SA5 - corresponding to nucleotides 3089-3090 of SEQ ID NO:1 or nucleotides 3089-3090 of SEQ ID NO:3.

SA6 - corresponding to nucleotides 3108-3109 of SEQ ID NO:1 or nucleotides 3108-3109 of SEQ ID NO:3.

SA8 - corresponding to nucleotides 3130-3131 of SEQ ID NO:1 or nucleotides 3133-3134 of SEQ ID NO:3.

4. The polynucleotide according to claim 3 wherein at least one of the nucleotides corresponding to the splice site is replaced by another nucleotide.

5. The polynucleotide according to claim 4 wherein G is changed to A.

6. The polynucleotide according to any one of claims 3 to 5 wherein at least 2, 3, 4, 5, 6, 7, 8, 9,10,11, 12 or 13 of the splice sites are inactivated.

7. The polynucleotide according to any one of claims 3 to 6 for use in claim 1 or 2.

8. A polynucleotide sequence comprising a HS3 region of b-globin locus control region nucleotide sequence wherein at least one of the following splice sites (forward and/or reverse) is inactivated:

Splice acceptor	corresponds to site shown in
SA A	nucleotides 8106-8107SEQ ID NO: 2
SA B	nucleotides 8067-8068
SA C	nucleotides 7474-7475
SA D	nucleotides 5423-5424

Splice donor	corresponds to site shown in
SD A	nucleotides 7912-7913
SD B	nucleotides 7837-7838
SD C	nucleotides 7821-7822
SD D	nucleotides 7797-7798
SD E	nucleotides 7367-7668
SD F	nucleotides 7363-73642

9. A vector comprising the polynucleotide sequence of any one of claims 3 to 8.

10. A vector according to claim 9 wherein the vector is a lentiviral vector.

11. A packaging, producer or host cell comprising the polynucleotide sequence or vector of any one of claims 3 to 10.

10. A pharmaceutical composition comprising a polynucleotide sequence, vector or cell according to any one of claims 3 to 11.

11. A vector comprising an miRNA target sequence wherien said miRNA target sequence is positioned upstream of a splice donor site or downstream of a splice acceptor site, wherein said splice donor or splice acceptor site is responsible for splicing events that generate unwanted fusion transcripts comprisng vector sequences and cellular mRNAs, wherein said miRNA target sequence casues degradation of said unwanted fusion transcripst in a cell comprising a corresponding endogenous miRNA.

12. A vector according to claim 11 wherein the miRNA target sequence is recognised by enodenous miRNA expressed in hematopoietic or hepatic cells such that the fusion transcript is selectively degraded in said cells.

13. A vector according to claim 11 or 12 wherein the miRNA target sequence is one targeted by hsa-mir-142as (also called hsa-mir-142-3p) miRNA, let-7a, mir-15a, mir-16, mir-17-5p, mir-19, mir-142-5p, mir-145 and/or mir-218 miRNA.

14 A vector according to any one of claims 11 to 13 wherein the vector is a lentiviral vector.

15. A vector according to any one of claims 11 to 14 wherin the splice acceptor or splice donor site is a splice site defined in claim 3 or 7.

16. A vector according to any one of claims 11 to 15 for use in therapy.

17. A method of preventing expression of unwanted transcripts comprisng vector sequences and cellular mRNAs in a subject comprising administering a vector according to any one of claims 11 to 16 to said subject..

FIGURE 1

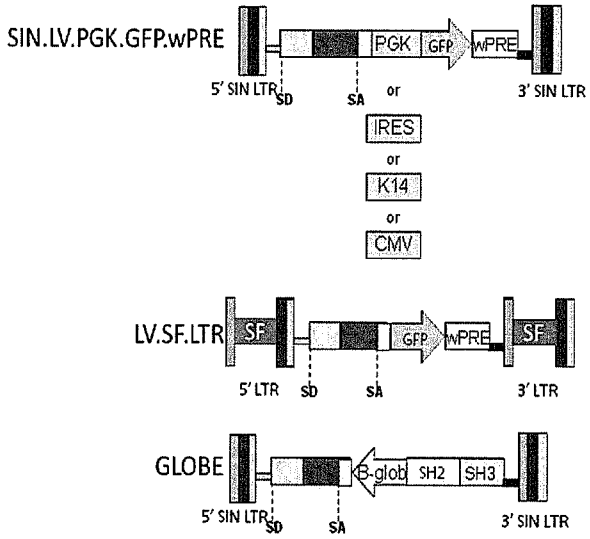
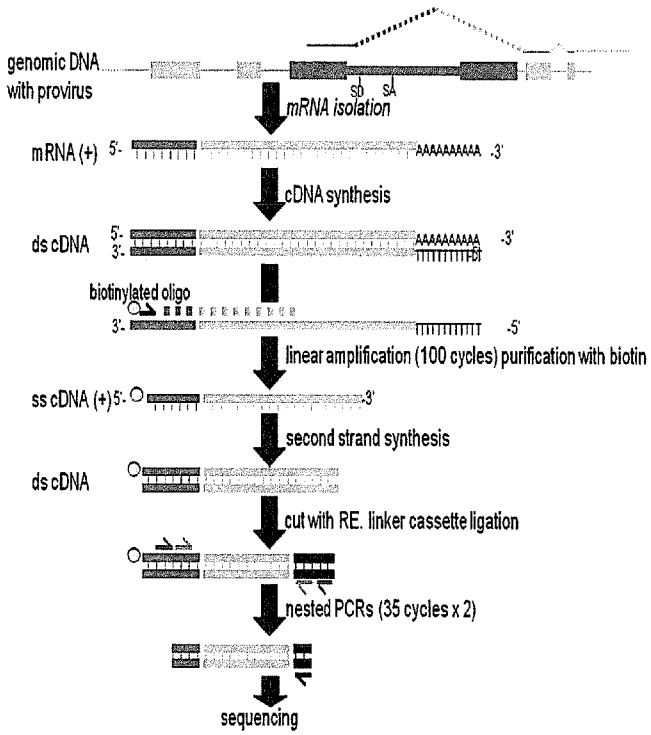
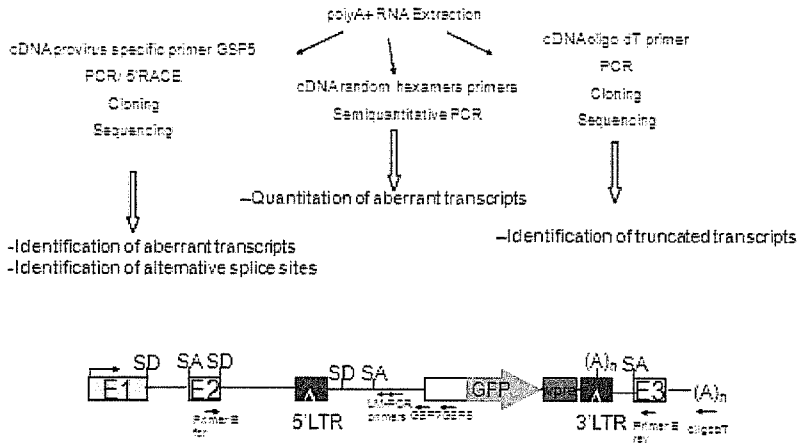

SIN.LV.PGK.GFP.wPRE
PGK
GFP
wPRE
5' SIN LTR
SD
SA
or
3' SIN LTR
IRES
or
K14
or
CMV
LV.SF.LTR
SF
GFP
wPRE
SF
5' LTR
SD
SA
3' LTR
GLOBE
B-glob
SH2
SH3
5' SIN LTR
SD
SA
3' SIN LTR

FIGURE 2

A)


genomic DNA
with provirus
SD
SA
mRNA isolation
mRNA (+) 5'-
AAAAAAAAA -3'
cDNA synthesis
ds cDNA
5'-
3'-
AAAAAAAAA
-3'
biotinylated oligo
3'-
-5'
linear amplification (100 cycles) purification with biotin
ss cDNA (+) 5'-
-3'
second strand synthesis
ds cDNA
cut with RE. linker cassette ligation
nested PCRs (35 cycles x 2)
sequencing

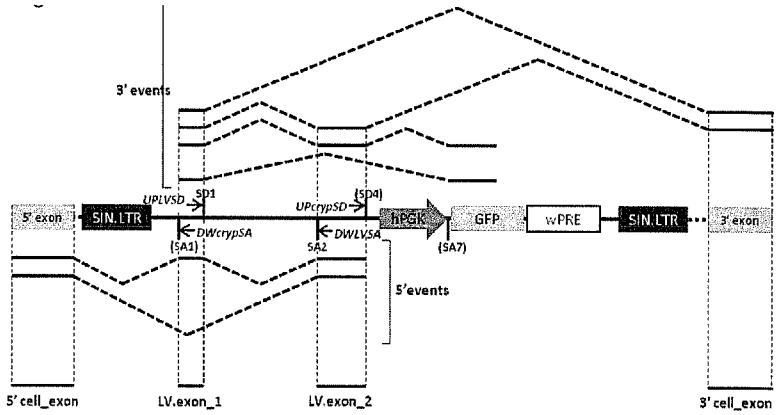
B)


polyA+ RNA Extraction
cDNA provirus specific primer GSP5
PCR/ 5'RACE
Cloning
Sequencing
cDNA random hexamers primers
Semiquantitative PCR
cDNA oligo dT primer
PCR
Cloning
Sequencing
-Quantitation of aberrant transcripts
-Identification of aberrant transcripts
-Identification of alternative splice sites
-Identification of truncated transcripts

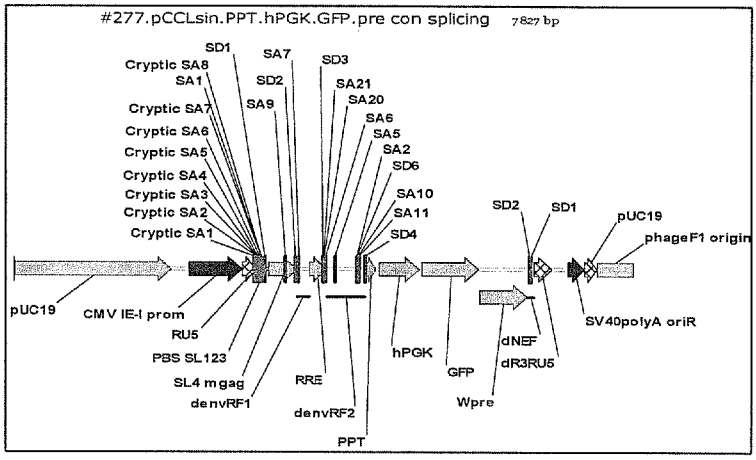
SD
SA SD
E1
E2
SD SA
5'LTR
GFP
(A)n SA
E3
(A)n
3'LTR
oligodT

FIGURE 3

A)


3' events
UPLVSD
SD1
UPcrypSD
(SD4)
5' exon
SIN.LTR
DWcrypSA
(SA1)
SA2
DWLVSA
hPGK
GFP
WPRE
SIN.LTR
3' exon
(SA7)
5'events
5' cell_exon
LV.exon_1
LV.exon_2
3' cell_exon

B)


#277.pCCLsin.PPT.hPGK.GFP.pre con splicing 7827 bp
SD1
SA7
SD3
Cryptic SA8
SA1
SD2
SA21
Cryptic SA7
SA9
SA20
Cryptic SA6
SA6
Cryptic SA5
SA5
SA2
Cryptic SA4
SD6
Cryptic SA3
SA10
Cryptic SA2
SA11
Cryptic SA1
SD4
SD2
SD1
pUC19
phageF1 origin
pUC19
CMV IE-I prom
RU5
PBS SL123
SL4 mgag
denvRF1
RRE
denvRF2
PPT
hPGK
GFP
Wpre
dNEF
dR3RU5
SV40polyA oriR

C)

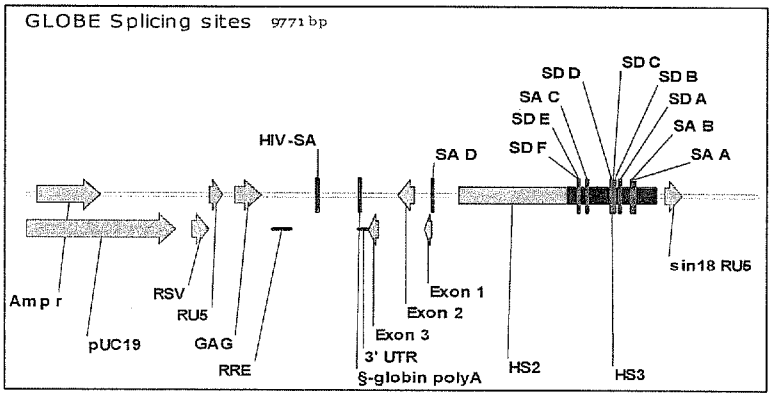
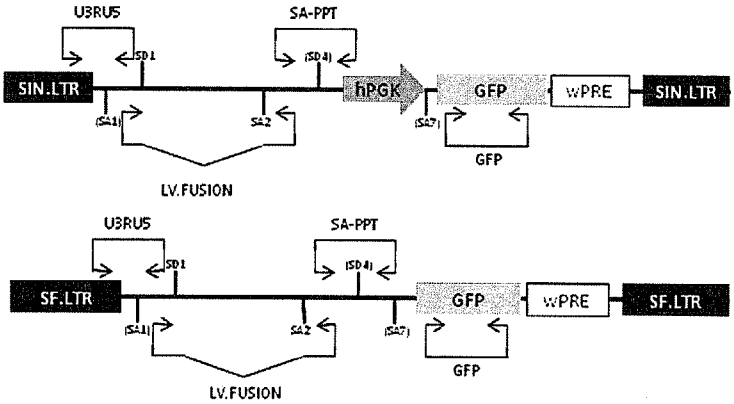

GLOBE Splicing sites 9771 bp
SD D
SD C
SD B
SA C
SD A
SD E
SA B
HIV-SA
SD F
SA A
SA D
sin18 RU5
Amp r
RSV
RU5
Exon 1
Exon 2
Exon 3
pUC19
GAG
3' UTR
RRE
§-globin polyA
HS2
HS3

FIGURE 4

A)


U3RU5
SA-PPT
SD1
SD4
SIN.LTR
hPGK
GFP
WPRE
SIN.LTR
SA1
SA2
SA7
GFP
LV.FUSION
U3RU5
SA-PPT
SD1
SD4
SF.LTR
GFP
WPRE
SF.LTR
SA1
SA2
SA7
GFP
LV.FUSION

B)

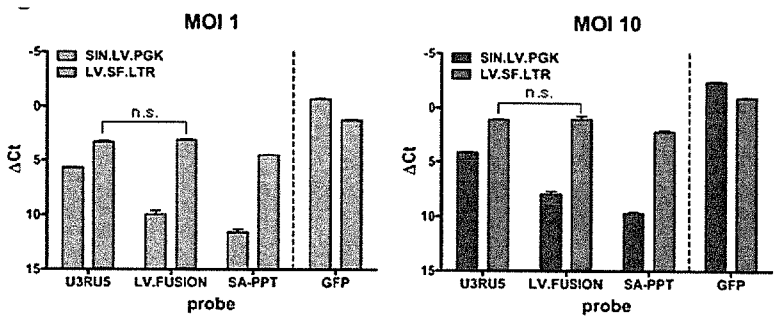

MOI 1
MOI 10
SIN.LV.PGK
LV.SF.LTR
n.s.
ΔCt
-5
0
5
10
15
U3RU5
LV.FUSION
SA-PPT
GFP
probe

FIGURE 5

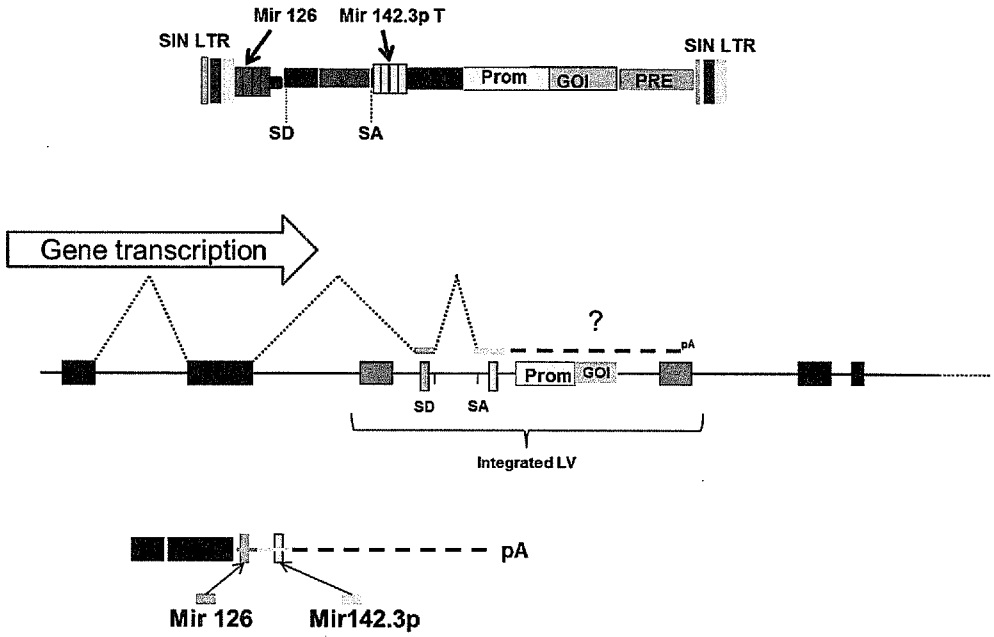

Mir 126
Mir 142.3p T
SIN LTR
SIN LTR
Prom
GOI
PRE
SD
SA
Gene transcription
?
pA
Prom
GOI
SD
SA
Integrated LV
pA
Mir 126
Mir142.3p

Figure 6

Cesana D. et al., Figure 1

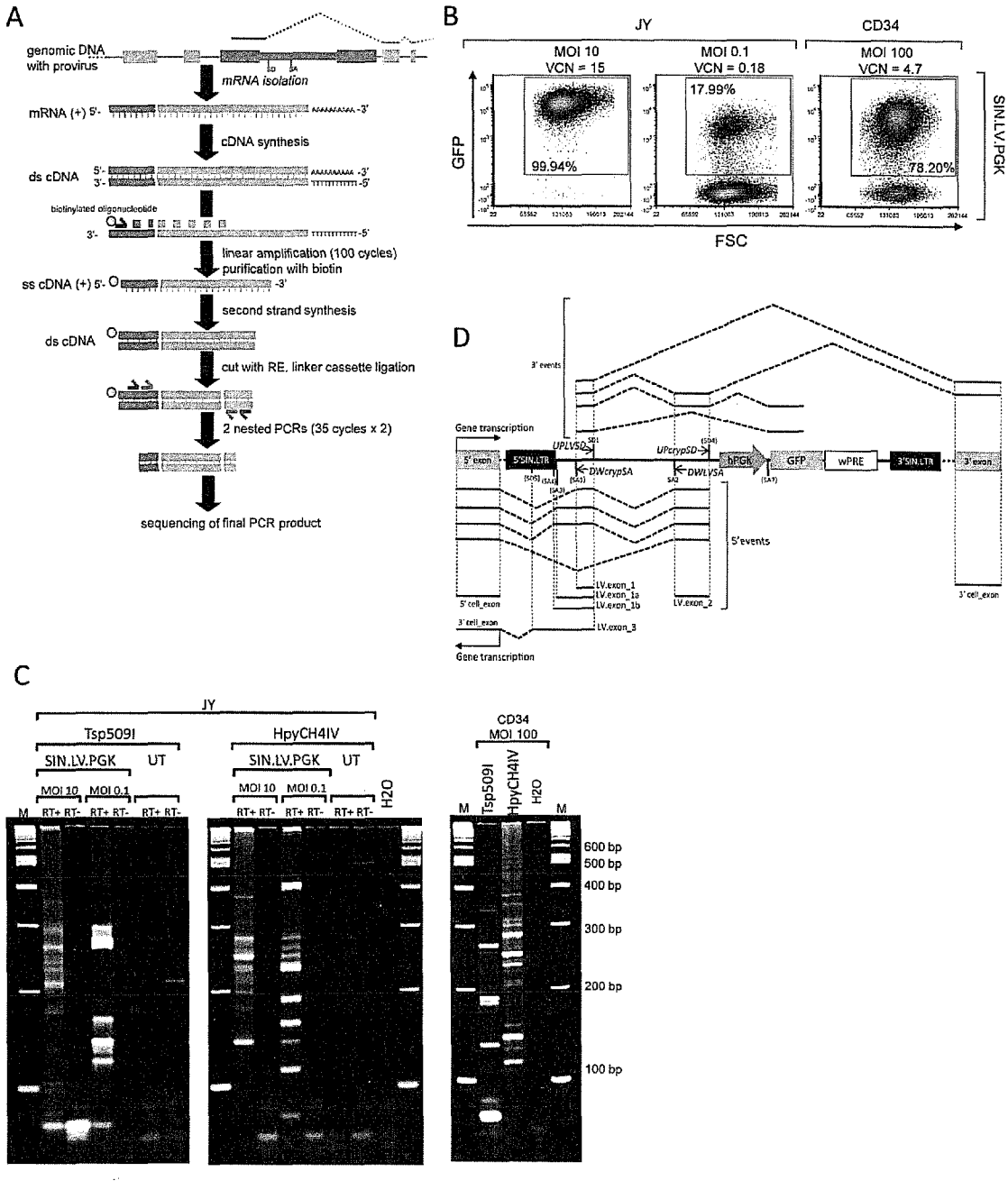
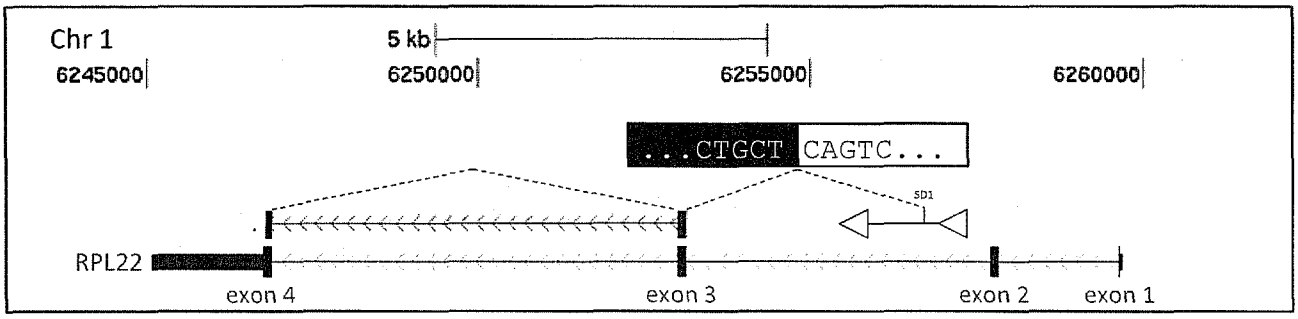
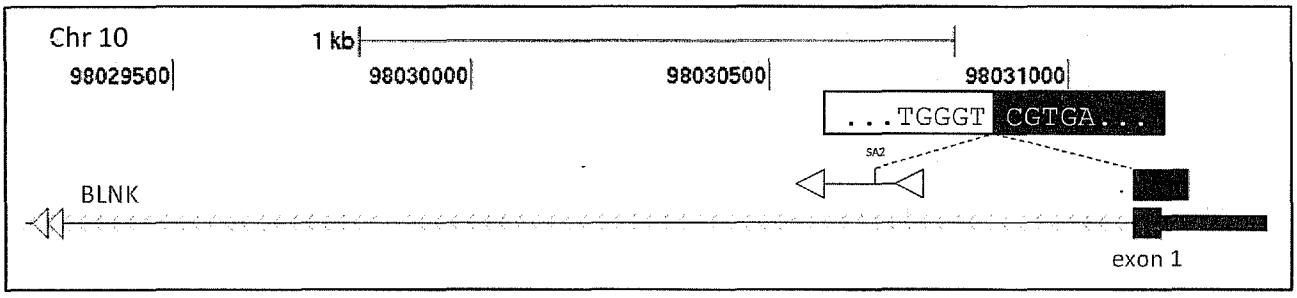


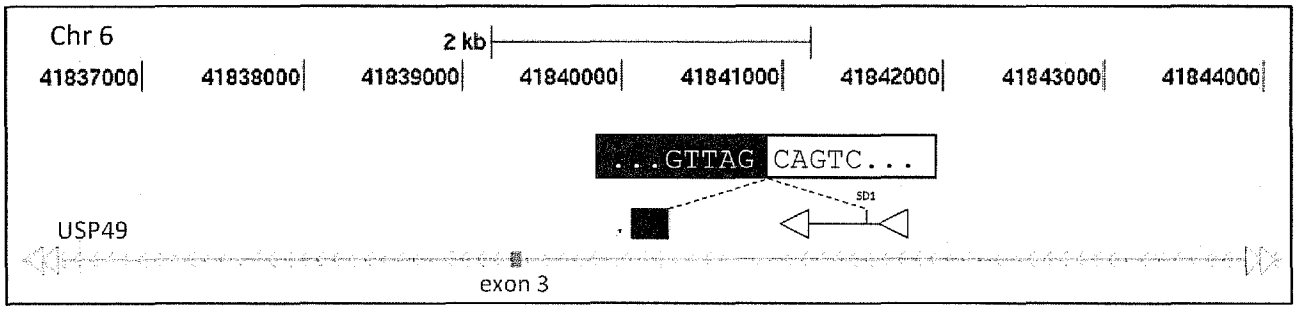
Figure 7



7029215 – 5' cell_exon



6433432 – 3' cell intron



6433403 – 3' intergenic

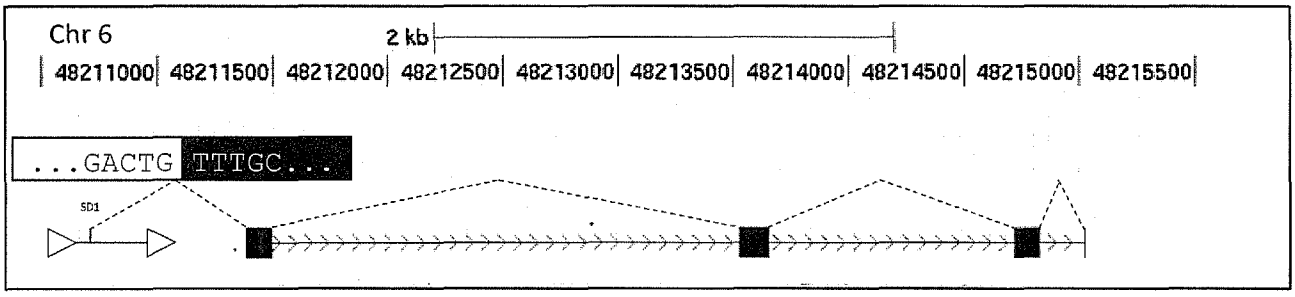


Figure 8

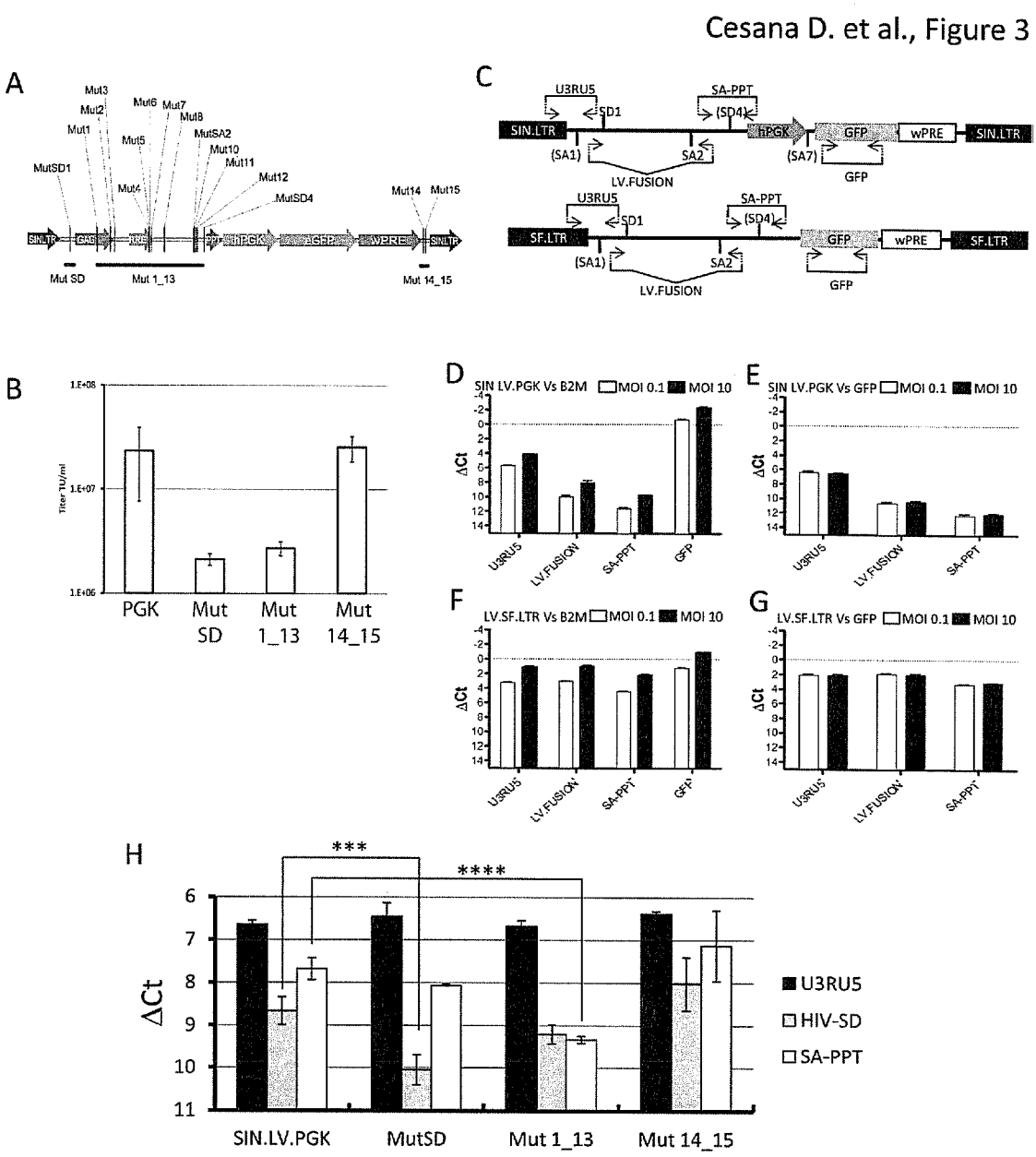
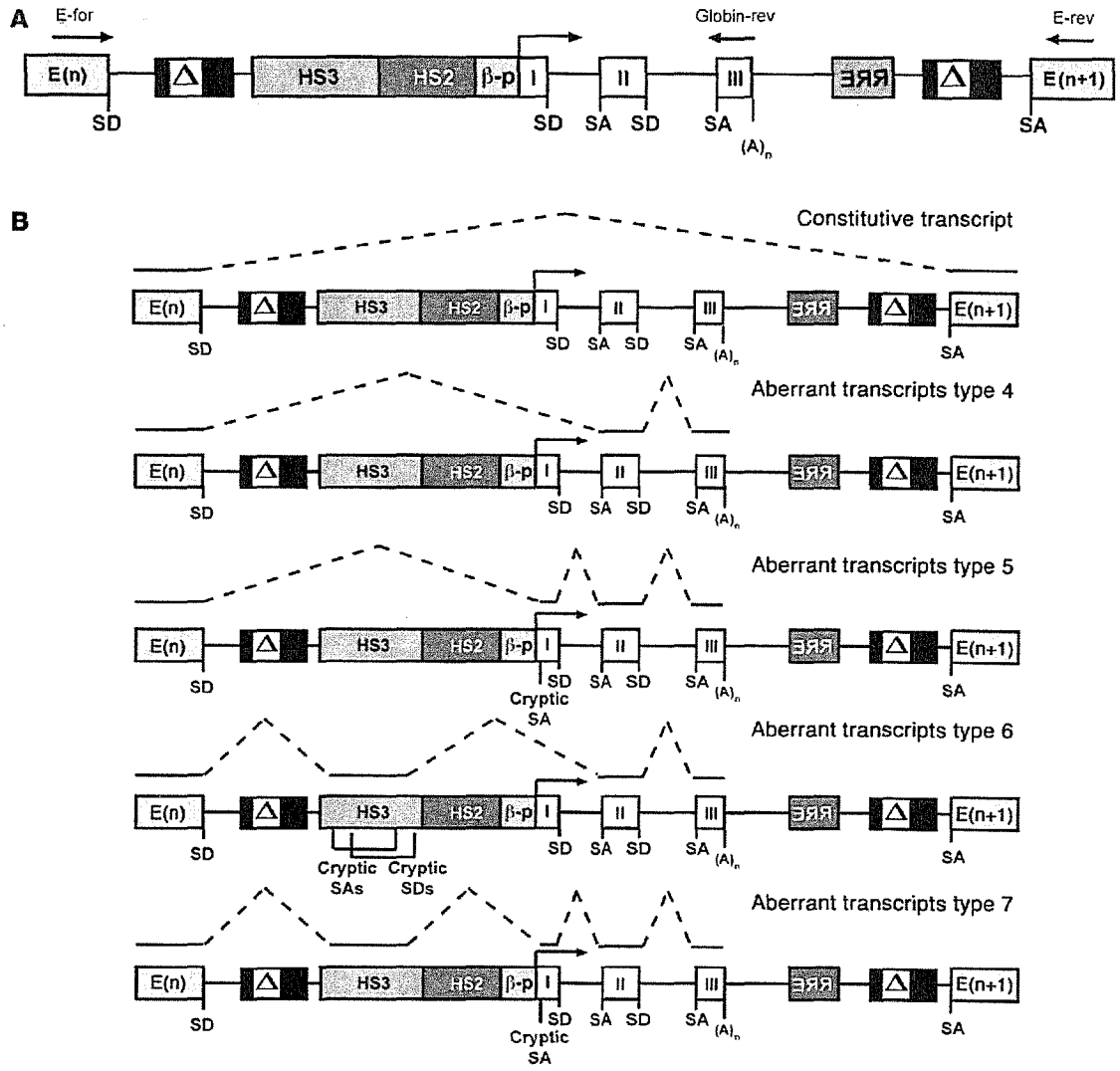

Cesana D. et al., Figure 3
A
Mut3
Mut2
Mut1
Mut6
Mut7
Mut8
Mut5
MutSA2
Mut10
Mut11
MutSD1
Mut12
Mut4
MutSD4
Mut14
Mut15
SIN.LTR
GAG
RRE
cPPT
hPGK
eGFP
wPRE
SIN.LTR
Mut SD
Mut 1_13
Mut 14_15
C
U3RU5
SA-PPT
SD1
(SD4)
SIN.LTR
hPGK
GFP
wPRE
SIN.LTR
(SA1)
SA2
(SA7)
LV.FUSION
GFP
U3RU5
SA-PPT
SD1
(SD4)
SF.LTR
GFP
wPRE
SF.LTR
(SA1)
SA2
LV.FUSION
GFP
B
Titer TU/ml
1.E+08
1.E+07
1.E+06
PGK
Mut SD
Mut 1_13
Mut 14_15
D
SIN LV.PGK Vs B2M
MOI 0.1
MOI 10
ΔCt
U3RU5
LV.FUSION
SA-PPT
GFP
E
SIN LV.PGK Vs GFP
MOI 0.1
MOI 10
ΔCt
U3RU5
LV.FUSION
SA-PPT
F
LV.SF.LTR Vs B2M
MOI 0.1
MOI 10
ΔCt
U3RU5
LV.FUSION
SA-PPT
GFP
G
LV.SF.LTR Vs GFP
MOI 0.1
MOI 10
ΔCt
U3RU5
LV.FUSION
SA-PPT
H
***
****
ΔCt
6
7
8
9
10
11
U3RU5
HIV-SD
SA-PPT
SIN.LV.PGK
MutSD
Mut 1_13
Mut 14_15

Figure 9



C

HS3

ACTTCTTTGAGAAACATCTTCTTCGTTAGTGGCCTGCCCCTCATTCCCACTTTAATATCCAGAATCACTATAAGAAGAATATAATAAGAGGAA

TAACTCTTATTATAGGTAAGGGAAAATTAAGAGGCATACGTGATGGGATGAGTAAGAGAGGAGAGGGAAGGATTAATGGATGATAAAATCTAC

TACTATTTGTTGAGACCTTTTATAGTCTAATCAATTTTGCTATTGTTTTCCATCCTCACGCTAACTCCATAAAAAAACACTATTATTATCTTT

ATTTTGCCATGACAAGACTGAGCTCAGAAGAGTCAAGCATTTGCCTAAGGTCGGACATGTCAGAGGCAGTGCCAGACCTATGTGAGACTCTGC

AGCTACTGCTCATGGGCCCTGTGCTGCACTGATGAGGAGGATCAGATGGATGGGGCAATGAAGCAAAGGAATCATTCTGTGGATAAAGGAGAC

AGCCATGAAGAAGTCTATGACTCAAATTTGGGAGCAGGAGTCTCTAAGGACTTCGATTTCAAGGAATTTTGACTCAGCAAACACAAGACCCT

CACGAGACTTTGCGAGCTGAGTGCCAGATGTGTCTATCAGAGATCCAGGGAGGGTGGGGTGGGGTCAGGGCTGGCCACCAGCTATCAGGG

CCCAGATGGGTTATAGGCTGGCAGGCTCAGATAGGTGGTTAGGTCAGGTTGGTGGTGCTGGGTGGAGTCCATGACTCCCAGGAGCCAGGAGAG

ATAGACCATGAGTAGAGGGCAGACATGGGAAAGGTGGGGGAGGCACAGCATAGCAGCATTTTTCATTCTACTACTACATGGGACTGCTCCCCT

ATACCCCCAGCTAGGGGCAAGTGCCTTGACTCCTATGTTTTCAGGATCATCATCTATAAAGTAAGAGTAATAATTGTGTCTATCTCATAGGGT

TATTATGAGGATCAAAGGAGATGCACACTCTCTGGACCAGTGGCCTAACAGTTCAGGACAGAGCTATGGGCTTCCTATGTATGGGTCAGTGGT

CTCAATGTACAGTAAGTTCCAGAAGATAGCATCAACCAC

β-Globin 5'UTR

ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCC

TGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAG