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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
03 September 2020 (03.09.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/176712 A1

(51) International Patent Classification:
C12N 15/867 (2006.01) *A61P 7/06* (2006.01)
A61P 7/00 (2006.01)

(21) International Application Number:
PCT/US2020/020053

(22) International Filing Date:
27 February 2020 (27.02.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/812,173 28 February 2019 (28.02.2019) US
62/846,504 10 May 2019 (10.05.2019) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, California 94607-5200 (US).

(72) Inventors: **KOHN, Donald B.**; c/o The Regents of the University of California, 1111 Franklin Street, Twelfth Floor, Oakland, California 94607-5200 (US). **MORGAN, Richard A.**; c/o The Regents of the University of California, 1111 Franklin Street, Twelfth Floor, Oakland, California 94607-5200 (US). **HOLLIS, Roger P.**; c/o The Regents of the University of California, 1111 Franklin Street, Twelfth Floor, Oakland, California 94607-5200 (US).

(74) Agent: **HUNTER, Tom** et al.; Weaver Austin Villeneuve & Sampson LLP, P.O. Box 70250, Oakland, California 94612-0250 (US).

(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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(54) Title: AUGMENTATIONS TO LENTIVIRAL VECTORS (CCLC-MGATA/ANK-CORE LCR-BETA-AS3-FB) TO INCREASE EXPRESSION

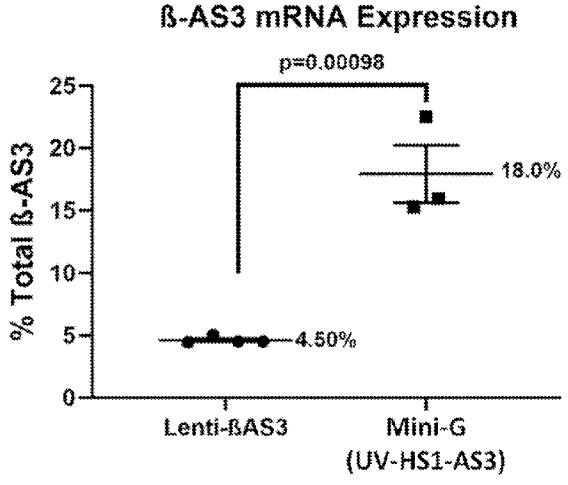


Fig. 9

(57) Abstract: In certain embodiments a lentiviral vector having an LCR comprising HS1 ENCODE core (EC1) sequence (SEQ ID NO:1), and one or more of an HS2 core sequence (ecHS2), an HS3 core sequence (ecHS3), an HS4 core sequence (ecHS4), a full length HS2, a full length HS3, and/or a full length HS4 sequence is provided. In certain embodiments the vector comprises a modified βAS3-globin transgene, where transgene comprises a codon optimized exon 1, and/or a codon optimized exon 2, and/or a codon optimized exon 3.

[Continued on next page]

WO 2020/176712 A1

(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, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

AUGMENTATIONS TO LENTIVIRAL VECTORS (CCLC-MGATA/ANK-CORE LCR-BETA-AS3-FB) TO INCREASE EXPRESSION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of and priority to USSN 62/812,173, filed on February 28, 2019, and to USSN 62/846,504, filed on May 10, 2019, both of which are incorporated herein by reference in their entirety for all purposes.

STATEMENT OF GOVERNMENTAL SUPPORT

[Not Applicable]

INCORPORATION BY REFERENCE OF SEQUENCE LISTING PROVIDED AS A TEXT FILE

[0002] A Sequence Listing is provided herewith as a text file, "UCLAP208P2US_ST25.txt" created on May 10, 2019 and having a size of 107,770 bytes. The contents of the text file are incorporated by reference herein in their entirety.

BACKGROUND

[0003] Sickle cell disease (SCD) is one of the most common monogenic disorders worldwide and is a major cause of morbidity and early mortality (Hoffman *et al.* (2009) Hematology: Basic Principles and Practice. 5th ed. London, United Kingdom, Churchill Livingstone). SCD affects approximately 80,000 Americans, and causes significant neurologic, pulmonary, and renal injury, as well as severe acute and chronic pain that adversely impacts quality of life. It is estimated that approximately 240,000 children are born annually in Africa with SCD and 80% die by their second birthday. The average lifespan of subjects with SCD in the United States is approximately 40 years and this has remained unchanged over the last 3-4 decades.

[0004] SCD is caused by a single amino acid change in β-globin (Glu 6 to Val 6) which leads to hemoglobin polymerization and red blood cell (rbc) sickling. SCD typically results in continual low-grade ischemia and episodic exacerbations or "crises" resulting in tissue ischemia, organ damage, and premature death.

[0005] Although SCD is well characterized, there is still no ideal long-term treatment. Current therapies are based on induction of fetal hemoglobin (HbF) to inhibit polymerization

of sickle hemoglobin (HbS) (Voskaridou *et al.* (2010) *Blood,* 115(12): 2354-2363) and cell dehydration (Eaton and Hofrichter (1987) *Blood*, 70(5): 1245-1266) or reduction of the percentage of HbS by transfusions (Stamatoyannopoulos *et al.*, eds. (2001) *Molecular Basis of Blood Diseases. 3rd ed.* Philadelphia, Pennsylvania, USA: WB Saunders). Allogeneic human stem cell transplantation (HSCT) from bone marrow (BM) or umbilical cord blood (UCB) or mobilized peripheral blood stem cells (mPBSC) is a potentially curative therapy, although only a small percentage of patients have undergone this procedure, mostly children with severe symptoms who had HLA-matched sibling donors (Bolaños-Meade and Brodsky (2009) *Curr. Opin. Oncol.* 21(2): 158-161; Rees *et al.* (2010) *Lancet*, 376(9757): 2018-2031; Shenoy (2011) *Hematology Am Soc Hematol Educ Program.* 2011: 273-279).

[0006] Transplantation of allogeneic cells carries the risk of graft-versus host disease (GvHD), which can be a cause of extensive morbidity. HSCT using UCB from matched unrelated donors holds reduced risk of acute or chronic GvHD compared with using BM; however, there is a higher probability of engraftment failure using UCB as a result of its lower cell dose and immunologic immaturity (Kamani *et al.* (2012) *Biol. Blood Marrow Transplant.* 18(8): 1265-1272; Locatelli and Pagliara (2012) *Pediatr. Blood Cancer.* 59(2): 372-376).

[0007] Gene therapy with autologous human stem cells (HSCs) is an alternative to allogeneic HSCT, since it avoids the limitations of finding a matched donor and the risks of GvHD and graft rejection. For gene therapy application in SCD patients,one source for autologous HSC would be BM, due to the complications previously described when G-CSF was used to collect autologous peripheral blood stem cells (PBSCs) in SCD patients (Abboud *et al.* (1998) *Lancet*351(9107): 959; Adler *et al.* (2001) *Blood,* 97(10): 3313-3314; Fitzhugh *et al.* (2009) *Cytotherapy,* 11(4): 464-471). However, more recently, plerixafor, an immunostimulat, can be used to mobilize hematopoietic stem cells into the blood stream. The stem cells can then be extracted from the blood and use. Although general anesthesia imposes a risk for SCD patients as well, current best medical practices can minimize these (Neumayr *et al.* (1998) *Am. J. Hematol.* 57(2): 101-108).

[0008] The development of integrating vectors for β-globin gene transfer has been challenging due to the complex regulatory elements needed for high-level, erythroid-specific expression (Lisowski a& Sadelain (2008) *Br. J. Haematol.* 141(3): 335-345). γ-Retroviral vectors were unable to transfer these β-globin expression cassettes intact (Gelinas *et al.* (1989) *Adv. Exp. Med. Biol.* 271: 135-148; Gelinas *et al.* (1989) *Prog. Clin. Biol. Res.* 316B:

235-249). In contrast, lentiviral vectors (LV) can transfer β-globin cassettes intact with relatively high efficiency, although the titers of these vectors are reduced compared with those of vectors bearing simpler cassettes (*see, e.g.,* May *et al.* (2000) *Nature* 406(6791): 82-86; Pawliuk *et al.* (2001) *Science*, 294(5550): 2368-2371). In the last decade, many groups have developed different β-globin LV for targeting β-hemoglobinopathies, with successful therapeutic results following transplantation of *ex vivo*- modified HSC in mouse models (May *et al.* (2000) *Nature*406(6791): 82-86; Pawliuk *et al.* (2001) *Science*, 294(5550): 2368-2371; Levasseur *et al.* (2003) *Blood*, 102(13):4312-4319; Hanawa *et al.* (2004) Blood, 104(8): 2281-2290; Puthenveetil *et al.* (2004) Blood, 104(12): 3445-3453; Miccio *et al.* (2008) *Proc. Natl. Acad. Sci. USA,* 105(30):10547-10552; Pestina *et al.* (2008) *Mol. Ther.* 17(2): 245-252). Recently, Bluebird Bio, updated trial results with longer follow-up from lentiviral vector treated patients with beta-thalassemia and sickle cell disease. Across three small studies testing the gene therapy in the two blood diseases, patients given LentiGlobin saw their levels of the crucial oxygen-carrying protein hemoglobin rise to approach normal, eliminating the need for blood transfusions in most over the studied period.

[0009] Sickle patients with hereditary persistence of fetal hemoglobin (HbF) (HPFH) have improved survival and amelioration of clinical symptoms, with maximal clinical benefits observed when the HbF is elevated above threshold values (*e.g.*, 8%-15% of the total cellular Hb) (Voskaridou *et al.* (2010) *Blood,* 115(12): 2354-2363; Platt *et al.* (1994) *N. Engl. J. Med.* 330(23): 1639-1644). Therefore, some gene therapy strategies have employed viral vectors carrying the human γ-globin gene (*HBG1/2*). However, these constructs expressed HbF poorly in adult erythroid cells, since fetal-specific transcription factors are required for high-level expression of the γ-globin gene (Chakalova *et al.* (2005) *Blood*105(5): 2154-2160; Russell (2007) *Eur. J. Haematol.* 79(6): 516-525). These limitations have been overcome by embedding the exons encoding human γ-globin within the human β-globin gene 5′ promoter and 3′ enhancer elements (Hanawa *et al.* (2004) Blood, 104(8): 2281-2290; Persons *et al.* (2002) *Blood*, 101(6): 2175-2183; Perumbeti *et al.* (2009) *Blood*, 114(6): 1174-1185). Breda *et al.* (2012) *PLoS One*, 7(3): e32345 used an LV vector encoding the human hemoglobin (HBB) gene to increase the expression of normal HbA in $CD34^+$-derived erythroid cells from SCD patients, however, the expression level needed when the *HBB* gene is used would be higher than would be required for *HBG1/2* gene expression to achieve therapeutic benefits in SCD patients.

[0010] Another approach is to modify β-globin genes to confer antisickling activity by substituting key amino acids from γ-globin. The modified β-globin cassette should yield

the necessary high-level, erythroid-specific expression in adult erythroid cells. Pawliuk *et al.* (2001) *Science*, 294(5550): 2368-2371 designed an LV carrying a human β-globin gene with the amino acid modification T87Q. The glutamine at position 87 of γ-globin has been implicated in the anti-sickling activity of HbF (Nagel *et al.* (1979) *Proc. Natl. Acad. Sci., USA,* 76(2): 670-672). This anti-sickling construct corrected SCD in 2 murine models of the disease, and a similar LV has been used in a clinical trial for β-thalassemia and SCD in France (Cavazzana-Calvo *et al.* (2010) *Nature*, 467(7313): 318-322).

[0011] Townes and colleagues have taken a similar approach, developing a recombinant human anti-sickling β-globin gene (HBBAS3) encoding a β-globin protein (HbAS3) that has 3 amino substitutions compared with the original (HbA): T87Q for blocking the lateral contact with the canonical Val 6 of HbS, E22A to disrupt axial contacts (McCune *et al.* (1994) *Proc. Natl. Acad. Sci. USA,* 91(21): 9852-9856) and G16D, which confers a competitive advantage over sickle-β-globin chains for interaction with the α-globin polypeptide. Functional analysis of the purified HbAS3 protein demonstrated that this recombinant protein had potent activity to inhibit HbS tetramer polymerization (Levasseur *et al.* (2004) *J. Biol. Chem.* 279(26): 27518-27524.). Levasseur *et al.* (2003) *Blood*, 102(13): 4312-4319, showed efficient transduction of BM stem cells from a murine model of SCD with a self-inactivating (SIN) LV carrying the HBBAS3 transgene that resulted in normalized rbc physiology and prevented the pathological manifestations of SCD.

[0012] Unfortunately, current β-globin expression vectors, suffer from low vector titer and sub-optimal gene transfer to hematopoietic stem cells, representing a major barrier toward the effective implementation of this gene therapy strategy to the clinic.

SUMMARY

[0013] Various embodiments cnetemplated herein may include, but need not be limited to, one or more of the following:

[0014] Embodiment 1: A recombinant lentiviral vector (LV) comprising:

[0015] an expression cassette comprising a nucleic acid construct comprising:

[0016] a human β-globin locus control region (LCR) comprising an HS1 ENCODE core (EC1) sequence (SEQ ID NO:1), and one or more of an HS2 core sequence (ecHS2 (SEQ ID NO: 2)), an HS3 core sequence (ecHS3 (SEQ ID NO:3)), an HS4 core sequence (ecHS4 (SEQ ID NO:4)), a full length HS2, a full length HS3, and/or a full length HS4.

[0017] Embodiment 2: The vector of embodiment 1, wherein said human β-globin locus control region (LCR) comprises an HS2 core sequence (ecHS2 (SEQ ID NO: 2)).

[0018] Embodiment 3: The vector of embodiment 1, wherein said human β-globin locus control region (LCR) comprises a full length HS2 sequence.

[0019] Embodiment 4: The vector according to any one of embodiments 1-3, wherein said human β-globin locus control region (LCR) comprises an HS3 core sequence (ecHS3 (SEQ ID NO:3)).

[0020] Embodiment 5: The vector according to any one of embodiments 1-3, wherein said human β-globin locus control region (LCR) comprises a full length HS3 sequence.

[0021] Embodiment 6: The vector according to any one of embodiments 1-5, wherein said human β-globin locus control region (LCR) comprises an HS4 core sequence (ecHS4 (SEQ ID NO:4).

[0022] Embodiment 7: The vector according to any one of embodiments 1-5, wherein said human β-globin locus control region (LCR) comprises a full length HS4.

[0023] Embodiment 8: The vector of embodiment 1, wherein said vector comprises the LCR shown in SEQ ID NO:6.

[0024] Embodiment 9: The vector according to any one of embodiments 1-8, wherein said heterologous gene comprises a recombinant human beta globin gene encoding a beta globin polypeptide.

[0025] Embodiment 10: The vector of embodiment 9, wherein said human beta globin gene comprises a wild-type beta globin gene.

[0026] Embodiment 11: The vector of embodiment 9, wherein said human beta globin gene comprises an anti-sickling human beta globin gene encoding an anti-sickling beta globin polypeptide.

[0027] Embodiment 12: The vector of embodiment 11, wherein said anti-sickling human beta globin gene encoding an anti-sickling-beta globin polypeptide comprise one or more mutations selected from the group consisting of Gly16Asp, Glu22Ala and Thr87Gln.

[0028] Embodiment 13: The vector of embodiment 12, wherein said beta globin gene comprises the mutation Gly16Asp.

[0029] Embodiment 14: The vector according to any one of embodiments 12-13, wherein said beta globin gene comprises the mutation Glu22Ala.

[0030] Embodiment 15: The vector according to any one of embodiments 12-14, wherein said beta globin gene comprises the mutation Thr87Gln.

[0031] Embodiment 16: The vector of embodiment 12, wherein said anti-sickling human β-globin gene comprises about 2.3 kb of recombinant human β-globin gene including exons and introns under the control of said human β-globin locus control region.

[0032] Embodiment 17: The vector according to any one of embodiments 12-16, wherein said β-globin gene comprises β-globin intron 2 with a 375 bp RsaI deletion from IVS2.

[0033] Embodiment 18: The vector according to any one of embodiments 12-17, wherein said β-globin gene comprises an SspI (S) to RsaI (R) deletion (~220bp).

[0034] Embodiment 19: The vector of embodiment 11, wherein said anti-sickling human beta globin gene comprises a modified βAS3-globin transgene, said transgene comprising a codon optimized exon 1 (SEQ ID NO:7) (as shown within SEQ ID NO:10), and/or a codon optimized exon 2 (SEQ ID NOs:5 or 8) (as shown within SEQ ID NO:11), and/or comprising a codon optimized exon 3 (SEQ ID NO:9) (as shown within SEQ ID NO:12).

[0035] Embodiment 20: The vector of embodiment 19, wherein said βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO:5).

[0036] Embodiment 21: The vector of embodiment 19, wherein said βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO:8).

[0037] Embodiment 22: The vector according to any one of embodiments 19-21, wherein said βAS3-globin transgene comprises a codon optimized exon 1 (SEQ ID NO:7).

[0038] Embodiment 23: The vector according to any one of embodiments 19-22, wherein said βAS3-globin transgene comprises a codon optimized exon 3 (SEQ ID NO:9).

[0039] Embodiment 24: The vector according to any one of embodiments 1-23, wherein said vector comprises a human Ankyrin insulator element.

[0040] Embodiment 25: The vector according to any one of embodiments 1-24, further comprising an insulator in the 3' LTR.

[0041] Embodiment 26: The vector of embodiment 25, wherein said insulator comprises FB (FII/BEAD-A), a 77 bp insulator element, which contains the minimal CTCF binding site enhancer-blocking components of the chicken β-globin 5' DnaseI-hypersensitive site 4 (5' HS4).

[0042] Embodiment 27: The vector according to any one of embodiments 1-26, wherein said vector comprises a ψ region vector genome packaging signal.

[0043] Embodiment 28: The vector according to any one of embodiments 1-27, wherein the 5' LTR comprises a CMV enhancer/promoter.

[0044] Embodiment 29: The vector according to any one of embodiments 1-28, wherein said vector comprises a Rev Responsive Element (RRE).

[0045] Embodiment 30: The vector according to any one of embodiments 1-29, wherein said vector comprises a central polypurine tract.

[0046] Embodiment 31: The vector according to any one of embodiments 1-30, wherein said vector comprises a post-translational regulatory element.

[0047] Embodiment 32: The vector of embodiment 31, wherein the posttranscriptional regulatory element is modified Woodchuck Post-transcriptional Regulatory Element (WPRE).

[0048] Embodiment 33: The vector of embodiment 1, wherein said vector comprises the nucleic acid sequence of SEQ ID NO:6.

[0049] Embodiment 34: The vector according to any one of embodiments 1-33, wherein said vector is incapable of reconstituting a wild-type lentivirus through recombination.

[0050] Embodiment 35: The vector according to any one of embodiments 1-33, wherein said vector is incapable of reconstituting a wild-type lentivirus through recombination.

[0051] Embodiment 36: The vector embodiment 1, wherein said vector comprises the features of pUV-AS3 (Figure 3).

[0052] Embodiment 37: The vector embodiment 1, wherein said vector comprises the nucleotide sequence of pUV-AS3 (SEQ ID NO: 13).

[0053] Embodiment 38: The vector embodiment 1, wherein said vector comprises the features of pUV-AS3(coX2) (Figure 4).

[0054] Embodiment 39: The vector embodiment 1, wherein said vector comprises the nucleotide sequence of pUV-AS3(coX2) (SEQ ID NO:14).

[0055] Embodiment 40: The vector embodiment 1, wherein said vector comprises the features of pUV-HS1-AS3 (Figure 5).

[0056] Embodiment 41: The vector embodiment 1, wherein said vector comprises the nucleotide sequence of pUV-HS1-AS3 (SEQ ID NO:15).

[0057] Embodiment 42: The vector embodiment 1, wherein said vector comprises the features of pUV+HS1-AS3(coX2) (Figure 6).

[0058] Embodiment 43: The vector embodiment 1, wherein said vectors comprises the nucleotide sequence of pUV+HS1-AS3(coX2) (SEQ ID NO: 16).

[0059] Embodiment 44: A host cell transduced with a vector according to any one of embodiments 1-43.

[0060] Embodiment 45: The host cell of embodiment 44, wherein the cell is a stem cell.

[0061] Embodiment 46: The host cell of embodiment 45, wherein said cell is a stem cell derived from bone marrow, and/or from umbilical cord blood, and/or from peripheral blood.

[0062] Embodiment 47: The host cell of embodiment 44, wherein the cell is a 293T cell.

[0063] Embodiment 48: The host cell of embodiment 44, wherein the cell is a human hematopoietic progenitor cell.

[0064] Embodiment 49: The host cell of embodiment 48, wherein the human hematopoietic progenitor cell is a CD34+ cell.

[0065] Embodiment 50: A method of treating a hemoglobinopathy, in a subject, said method comprising:

[0066] transducing a stem cell and/or progenitor cell from said subject with a vector according to any one of embodiments 1-43; and

[0067] transplanting said transduced cell or cells derived therefrom into said subject where said cells or derivatives therefrom express said anti-sickling human beta globin gene.

[0068] Embodiment 51: The method of embodiment 50, wherein the cell is a stem cell.

[0069] Embodiment 52: The host cell of embodiment 50, wherein said cell is a stem cell derived from bone marrow.

[0070] Embodiment 53: The method of embodiment 50, wherein the cell is a human hematopoietic stem and progenitor cell.

[0071] Embodiment 54: The method of embodiment 53, wherein the human hematopoietic progenitor cell is a $CD34^+$ cell.

[0072] Embodiment 55: The method according to any one of embodiments 50-54, wherein said hemoglobinopathy is sickle cell disease.

[0073] Embodiment 56: The method according to any one of embodiments 50-54, wherein said hemoglobinopathy is β-thalassemia.

[0074] Embodiment 57: A nucleic acid encoding a modified βAS3-globin transgene, said transgene comprising a codon optimized exon 1 (SEQ ID NO:7) (as shown within SEQ ID NO:10), and/or a codon optimized exon 2 (SEQ ID NOs:5 or 8), and/or comprising a codon optimized exon 3 (SEQ ID NO:9) (as shown within SEQ ID NO:12).

[0075] Embodiment 58: The nucleic acid of embodiment 57, wherein said βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO: 5).

[0076] Embodiment 59: The nucleic acid of embodiment 57, wherein said βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO: 8).

[0077] Embodiment 60: The nucleic acid according to any one of embodiments 57-59, wherein said βAS3-globin transgene comprises a codon optimized exon 1 (SEQ ID NO: 7).

[0078] Embodiment 61: The nucleic acid according to any one of embodiments 57-60, wherein said βAS3-globin transgene comprises a codon optimized exon 3 (SEQ ID NO:9).

Definitions.

[0079] An "HS core sequence" as used herein refers to a reduced β-globin locous control region (LCR) hypersensitivity site (HS) sequence as defined herein (*e.g.*, (HS2 (~420 bp), HS3 (~340bp), and/or HS4 (~410 bp)). Full-length HS sequences refers to LCR HS2,

HS3, and HS4 as previously defined (*e.g.*, HS2 (~1.20kb), HS3 (~1.28kb), and HS4 (~1.1kb)) (*see, e.g.*, Forrester *et al.* (1986) *Proc. Natl. Acad. Sci. USA,* 83: 1359-1363).

[0080] "Recombinant" is used consistently with its usage in the art to refer to a nucleic acid sequence that comprises portions that do not naturally occur together as part of a single sequence or that have been rearranged relative to a naturally occurring sequence. A recombinant nucleic acid is created by a process that involves the hand of man and/or is generated from a nucleic acid that was created by hand of man (*e.g.*, by one or more cycles of replication, amplification, transcription, *etc.*). A recombinant virus is one that comprises a recombinant nucleic acid. A recombinant cell is one that comprises a recombinant nucleic acid.

[0081] As used herein, the term "recombinant lentiviral vector" or "recombinant LV) refers to an artificially created polynucleotide vector assembled from an LV and a plurality of additional segments as a result of human intervention and manipulation.

[0082] By "globin nucleic acid molecule" is meant a nucleic acid molecule that encodes a globin polypeptide. In various embodiments the globin nucleic acid molecule may include regulatory sequences upstream and/or downstream of the coding sequence.

[0083] By "globin polypeptide" is meant a protein having at least 85%, or at least 90%, or at least 95%, or at least 98% amino acid sequence identity to a human beta, delta or gamma globin.

[0084] The term "therapeutic functional globin gene" refers to a nucleotide sequence the expression of which leads to a globin that does not produce a hemoglobinopathy phenotype, and which is effective to provide therapeutic benefits to an individual with a defective globin gene. The functional globin gene may encode a wild-type globin appropriate for a mammalian individual to be treated, or it may be a mutant form of globin, preferably one which provides for superior properties, for example superior oxygen transport properties or anti-sickling properties. The functional globin gene includes both exons and introns, as well as globin promoters and splice donors/acceptors.

[0085] By "an effective amount" is meant the amount of a required agent or composition comprising the agent to ameliorate or eliminate symptoms of a disease relative to an untreated patient. The effective amount of composition(s) used to practice the methods described herein for therapeutic treatment of a disease varies depending upon the manner of administration, the age, body weight, and general health of the subject. Ultimately, the

attending physician or veterinarian will decide the appropriate amount and dosage regimen. Such amount is referred to as an "effective" amount.

[0086] The term "sequence identity" refers to the extent to which two optimally aligned polynucleotide or polypeptide sequences are invariant throughout a window of alignment of components, *e.g.*, nucleotides or amino acids. "Identity" can be readily calculated by known methods including, but not limited to, those described in: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993); Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991); and the like.

[0087] An "identity fraction" for aligned segments of a test sequence and a reference sequence is the number of identical components which are shared by the two aligned sequences divided by the total number of components in the reference sequence segment, *i.e.*, the entire reference sequence or a smaller defined part of the reference sequence. Percent sequence identity is typically represented as the identity fraction multiplied by 100. As used herein, the term "percent sequence identity" or "percent identity" refers to the percentage of identical nucleotides in a linear polynucleotide sequence of a reference ("query") polynucleotide molecule (or its complementary strand) as compared to a test ("subject") polynucleotide molecule (or its complementary strand) when the two sequences are optimally aligned (with appropriate nucleotide insertions, deletions, or gaps totaling less than 20 percent of the reference sequence over the window of comparison).

[0088] Optimal alignment of sequences for aligning a comparison window is well known to those skilled in the art and may be conducted by tools such as the local homology algorithm of Smith and Waterman, the homology alignment algorithm of Needleman and Wunsch, the search for similarity method of Pearson and Lipman, and optionally by computerized implementations of these algorithms such as GAP, BESTFIT, FASTA, and TFASTA available as part of the GCG® Wisconsin Package® (Accelrys Inc., Burlington, Mass.). The comparison of one or more polynucleotide sequences may be to a full-length polynucleotide sequence or a portion thereof, or to a longer polynucleotide sequence. In various embodiments "percent identity" may also be determined using BLASTX version 2.0 for translated nucleotide sequences and BLASTN version 2.0 for polynucleotide sequences.

[0089] The percent of sequence identity can be determined using the "Best Fit" or "Gap" program of the Sequence Analysis Software Package.TM. (Version 10; Genetics Computer Group, Inc., Madison, Wis.). "Gap" utilizes the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443-453, to find the alignment of two sequences that maximizes the number of matches and minimizes the number of gaps. "BestFit" performs an optimal alignment of the best segment of similarity between two sequences and inserts gaps to maximize the number of matches using the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math.,* 2: 482-489, Smith *et al.* (1983) *Nucleic Acids Res.* 11: 2205-2220. Useful methods for determining sequence identity are also disclosed in Guide to Huge Computers (Martin J. Bishop, ed., Academic Press, San Diego (1994)), and Carillo *et al.* (1988) *Appl. Math,* 48: 1073). Certain illustrative computer programs for determining sequence identity include, but are not limited to, the Basic Local Alignment Search Tool (BLAST) programs, that are publicly available from National Center Biotechnology Information (NCBI) at the National Library of Medicine, National Institute of Health, Bethesda, Md. 20894; see BLAST Manual, Altschul *et al.*, NCBI, NLM, NIH; (Altschul *et al.*, J. Mol. Biol. 215:403-410 (1990)); version 2.0 or higher of BLAST programs allows the introduction of gaps (deletions and insertions) into alignments; for peptide sequence, BLASTX can be used to determine sequence identity; and for polynucleotide sequence, BLASTN can be used to determine sequence identity.

BRIEF DESCRIPTION OF THE DRAWINGS

[0090] Figure 1, panels A-B, illustrates the structure of a Core-ASB-FB and +EC1 lentiviral vectors (Panel A) and expression levels produced by each of these vectors when transfected into healthy donor hematopoietic stem and progenitor cells (HSPCs) panel B).

[0091] Figure 2, panels A and B, illustrates the structure of a modified βAS3-globin transgene comprising a codon optimized exon 1, or a codon optimized exon 1 and 2, or a codon optimized exon 1, 2 and 3 (panel A) and levels of expression produced by each of these constructs compared to Core-AS3-FB (panel B).

[0092] Figure 3 illustrates the structure of the pUV-AS3 vector and "*" illustrates changes from the GLOBE1 vector. pUV-AS3 is shown as SEQ ID NO:13 in Table 1. The cPPT and surrounding modifications (from GLOBE1) include: 1) GTTAAC (SEQ ID NO:17) is now ATCTCGACACAAATGGCAGTATTCATCCACAA SEQ ID NO:18) (at position 4,924 in SEQ ID NO:13); 2) An additional **A** (at position 5,062 in SEQ ID NO:13); and 3) ATCGGTACGTAC (SEQ ID NO:18) is now

CGGGTTTATTACAGGGACAGCAGA (SEQ ID NO:19) (at position 5,074 in SEQ ID NO:13). 3'PPT and surrounding modification: 4) GGGGGATCCACTAGTTCTAGAG CGGCCAAATGGCGGCCGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGA TCTTAGCCACTTTTTAAAAGAAAAGGGGGGAC (SEQ ID NO:20) is now TAAAAGAAAAGGGGGGAC (SEQ ID NO:21) (at position 7,847 in SEQ ID NO:13). 3'LTR modifications: 5) CAAGTAGTGTGTGCCCGTCTGTTG TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCT CTAGC (SEQ ID NO:22) (U5 region) is replaced by CAAGTAGTGTGTTGGT TTTTTGTGTGTGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGG GCAGCACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG CGGTGGGCTCTATGG (SEQ ID NO:23) (rabbit beta globin DSE and bovine growth hormone polyA) at position 8,014 of SEQ ID NO:13. SV40 origin of replication modification. SV40 origin of replication modification: 6) ATCCCGCCCCTAACT CCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGC AGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTT TTTTGGAGGCCTAGG (SEQ ID NO:24) is replaced by ATCCCGCCCCTAAC TCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTCTCACTACTTCTGGAATAGCT CAGAGGCCGAGGCGGCCTCGGCCTCTGCATAAATAAAAAAAATTAGTCGAGGCT TTTTTGGAGGCCTAGG (SEQ ID NO:25) at position 8,387 of SEQ ID NO:13. In addition the antibiotic resistance gene was changed from ampicillin to kanamycin resistance.

[0093] Figure 4 illustrates the structure of the pUV-AS3(coX2) vector.

[0094] Figure 5 illustrates the structure of the pUV-HS1-AS3 vector.

[0095] Figure 6 illustrates the structure of the pUV+AS3(coX2) vector.

[0096] Figure 7 shows that the unconcentrated vector titer of the pUV-HS1-AS3 (Mini-G) is ~10 fold higher than the parental vector. Titer was determined by the number of infectious particles, transducing units (TU), per ml on HT-29 cells. Readout is through detection of PSI sequences by ddPCR 3 days post transduction.

[0097] Figure 8, panels A-B, shows that Gene transfer is significantly enhanced for the shorter globin vector Mini-G (UV-HS1-AS3) in primary human hematopoietic stem and progenitor ($CD34^+$) cells (HSPCs). Panel A demonstrates how gene transfer continues to increase with Mini-G, compared to the parental vector that displays little increase in gene transfer. Panel B focusses on a single vector concentration ($2x10^7$ TU/ml) that is typically

used for clinical vector transductions, here the gene transfer is significantly improved and is ~10 fold higher than the parental vector.

[0098] Figure 9 shows that expression of AS3-globin is ~5-fold higher in erythrocytes differentiated *in vitro* from transduced HSPCs. Globin expression was assayed at the RNA level using ddPCR probe/primers. The difference in expression is caused by the enhanced gene transfer of Mini-G (UV-HS1-AS3).

DETAILED DESCRIPTION

[0099] It is believed that autologous stem cell gene therapy for sickle cell disease (SCD) or other hemoglobinopathies (*e.g.,* β-thallasemia, *etc.*) has the potential to treat these illnesses without the need for immune suppression of current allogeneic hematopoietic stem cell transplantation (HSCT) approaches. In particular, it is believed that autologous stem cell gene therapy that introduces, for example, anti-sickling human beta globin into hematopoietic cells (or progenitors thereof) can provide effective therapy for SCD (including, for example, normalized red blood cell (RBC) physiology and prevention of the manifestations of SCD) or certain other hemoglobinopathies.

[0100] Current β-globin expression vectors, however, suffer from low vector titer and sub-optimal gene transfer to hematopoietic stem cells, representing a major barrier toward the effective implementation of this gene therapy strategy to the clinic. Without being bound to a particular theory, it is believed that the predominant factor most likely affecting vector performance is overall vector length.

[0101] To address this concern, we previously developed a reduced length globin vectors that produced at ~3-fold higher titer with superior gene transfer to hematopoietic stem cells and comparable expression of BAS3 when compared to antecedent vectors. These vectors were characterized by the incorporation of reduceded length β-globin locous control region (LCR) hypersensitivity site (HS) sequences (*e.g.*, (HS2 (~420 bp), HS3 (~340bp), and/or HS4 (~410 bp) as describedin PCT Publication No: WO/2018/106724 (PCT/US17/64766) which is incorporated herein by reference for the sequence of the reduced length hypersensitivity sites and vectors describe therein.

[0102] Previous reports have demonstrated that the full length HS1 element could increase expression per vector genome when combined with the full length HS 2, 3, and 4 elements{Lisowski & Sadelain (2007) *Blood,* 110: 4175-4178). We hypothesized that the HS1 ENCODE core (EC1) could similarly improve expression. Thus, as an alternative

strategy to increase normalized expression, we again deployed ENCODE to redefine the boundaries of the HS1 element and designed a derivative of Core-AS3-FB containing the HS Core 1 and HS Core 2, 3, and 4 (called +EC1 (encode core 1)) to determine the influence of EC1 on BASB-globin expression.

[0103] Healthy donor HSPCs were transduced with Core-ASB-FB and +EC1 (Figure 1, Panel A) at equal MOI and cultured under erythroid culture conditions for 14 days. From three independent experiments, a 1.6-fold increase in expression per vector genome was observed when comparing Core-ASB-FB(+EC1) to its parental construct ($p<0.05$) (Figure 1, Panel B).

[0104] Accordingly, in certain embodments, a recombinant lentiviral vector (LV) is provided where the vector comprisess an expression cassette comprising a human β-globin locus control region (LCR) comprising an HS1 ENCODE core (EC1) sequence (SEQ ID NO:1), and one or more of an HS2 core sequence (ecHS2 (SEQ ID NO: 2)), an HS3 core sequence (ecHS3 (SEQ ID NO:3)), an HS4 core sequence (ecHS4 (SEQ ID NO:4) (*see, e.g.,* sequences in Table 1), a full length HS2, a full length HS3, and/or a full length HS4; and a heterologous gene to be expressed by the construct operably linked to the human β-globin locus control region.

[0105] In certain embodiments the human β-globin locus control region (LCR) comprises an HS2 core sequence (ecHS2 (SEQ ID NO: 2)). In certain embodiments the human β-globin locus control region (LCR) comprises a full length HS2 sequence. In certain embodiments the human β-globin locus control region (LCR) comprises an HS3 core sequence (ecHS3 (SEQ ID NO:3)). In certain embodiments the human β-globin locus control region (LCR) comprises a full length HS3 sequence. In certain embodiments the human β-globin locus control region (LCR) comprises an HS4 core sequence (ecHS4 (SEQ ID NO:4). In certain embodiments the human β-globin locus control region (LCR) comprises a full length HS4. In certain embodiments the vector comprises the LCR shown in SEQ ID NO:6.

Table 1. Nucleic acid sequences of ecHS1, ecHS2, ecHS3, ecHS4, a full vector with HS1, a vector comprising codon optimized exon 1, codon optimized exon 1, codon optimized exon 2, codon optimized exon 3, a vector comprising codon optimized exons 1and 2, a vector comprising codon optimized exons 1, 2, and 3. In the codon optimized exons, the case letters are codon optimized sequence and the lower case match wildtype sequence

ecHS1 (SEQ ID NO:1)	Catcaataattctagccccacaggagtttgttctgaaagtaaacttccacaaccgcaagcttattgaggcta aggcatctgtgaaggaaagaaacatctcctctaaaccactatgctgctagagcctcttttctgtactcaagc ctcattcagacactagtgtcaccagtctcctcatatacctattgtattttcttcttcttgctggtttagtcatgttt tctgggagcttaggggcttattttattttgttttgttttctaatcaacagagatgggcaaacccattatttttttc

	tttagacttgggatggtgatagctgggcagcgtcagaaactgtgtgtggatatagataagagctcaggact atgctgagctgtgatgagggaggggcctagctaaaggcagtgagagtcagaatgctcctgctattgccttct cagtccccacgcttggtttctacacaagtagatacatagaaaaggctataggttagtgtttgagagtcct gcatgattagttgctcagaaatgcccgataaatatgttatgtgtgtttatgtatatatatgttttatatgtgtgt gtgtgtgtgttgtgtttacaaatatgtgattatcatcaaaacgtgaggg
ecHS2 (SEQ ID NO:2)	Tacgtatatgtgtatatatatatatatattcaggaaataatatattctagaatatgtcacattctgtctcaggc atccattttctttatgatgccgtttgaggtggagttttagtcaggtggtcagcttctccttttttttgccatctgcc ctgtaagcatcctgctggggacccagataggagtcatcactctaggctgagaacatctgggcacacaccct aagcctcagcatgactcatcatgactcagcattgctgtgcttgagccagaaggtttgcttagaaggttacac agaaccagaaggcgggggtggggcactgaccccgacaggggcctggccagaactgctcatgcttggacta tgggaggtcactaatggagacacacagaaatgtaacaggaactaaggaaaaactgaagctt
ecHS3 (ecCORE HS3) (SEQ ID NO:3)	tgggggtataggggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcccccaccttt tcccatgtctgccctctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccacca acctgacctaaccacctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccc tgaccccaccccacccctccctggaacctctgatagacacatctggcacaccagctcgcaaagtcaccgtga gggtcttgtgtttgctgagtcaaaattccttgaaatccaagtccttagagactcc
ecHS4 (ecCORE HS4) (SEQ ID NO:4)	caggcttggattcaaagctcctgactttctgtctagtgtatgtgcagtgagccccttttcctctaactgaaaga aggaaaaaaaaatggaacccaaaatattctacatagtttccatgtcacagccagggctgggcagtctcctg ttatttcttttaaaataaatatatcatttaaatgcataaataagcaaaccctgctcgggaatgggagggaga gtctctggagtccacccccttctcggccctggctctgcagatagtgctatcaaagccctgacagagccctgccc attgctgggccttggagtgagtcagcctagtagaggaggcagggcaagccatctcatagctgctgagtggga gagagaaaagggctcattgtctataaactcaggtcatggctattcttat
coX2 (codon optimized exon 2) (SEQ ID NO:5)	cctGAAGTTCTCaGGaTCCACGTGCAGCTTGTCGCAGTGCAGCTCGCTCAGCTGGG CGAAGGTGCCCTTCAGGTTGTCCAGGTGGGCCAGGCCGTCGCTGAAGGCGCCCA GCACCTTCTTGCCGTGGGCCTTCACCTTGGGGTTGCCCATCACGGCGTCGGGGGT GCTCAGGTCGCCGAAGCTCTCGAAGAAGCGCTGGGTCCAGGGGTACACCACCAG cagc
Full vector with HS1 (SEQ ID NO:6)	gcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgt atgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcg cgcaattaaccctcactaaagggaacaaaagctggagctgcaagcttggccattgcatacgttgtatccat atcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagttatt aatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaa tggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaac gccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatca agtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccca gtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgc ggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattg acgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccat tgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggt ctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaat aaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcag acccttttagtcagtgtggaaaatctctagcagtggcgcccgaacagggacttgaaagcgaaagggaaac cagaggagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgac tggtgagtacgccaaaaattttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtat

	taagcgggggagaattagatcgcgatgggaaaaaattcggttaaggccaggggaaagaaaaaatataa attaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacat cagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagatc attatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctt tagacaagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcag acctggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaa ccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagagcagtgggaat aggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaa cagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagata cctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttgcaccactgctgtgccttgg aatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagag aaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatga acaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggta tataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtga atagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacagg cccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgattagtgaacggatct cgacggtatcgatctcgacacaaatggcagtattcatccacaattttaaaagaaaaggggggattggggg gtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaaca aattacaaaaattcaaaattttcgggtttattacagggacagcagagatccagtttgggtcgaggatatcgg atcggaattctctagatgatcaggatccctcgagcccttatcgatcacgagactagcctcgactactagtgg agatcccccgggctgcagagccagaagcaccataagggacatgataagggagccagcagacctctgatct cttcctgaatgctaatcttaaacatcctgaggaagaatgggacttccatttggggtgggcctatgatagggt aataagacagtagtgaatatcaagctacaaaaagccccctttcaaattcttctcagtcctaacttttcatact aagcccagtccttccaaagcagactgtgaaagagtgatagttccgggagactagcactgcagattccgggt cactgtgagtgggggaggcagggaagaagggctcacaggacagtcaaaccatgccccctgttttttccttct tcaagtagacctctataagacaacagagacaactaaggctgagtggccaggcgaggagaaaccatctcg ccgtaaaacatggaaggaacacttcaggggaaaggtggtatctctaagcaagagaactgagtggagtca aggctgagagatgcaggataagcaaatgggtagtgaaaagacattcatgaggacagctaaaacaataag taatgtaaaatacagcatagcaaaactttaacctccaaatcaagcctctacttgaatccttttctgagggat gaataaggcataggcatcaggggctgttgccaatgtgcattagctgtttgcagcctcaccttctttcatggag tttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcactgacctccca cattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttttttattag gcagaatccagatgctcaaggcccttcataatatcccccagtttagtagttggacttagggaacaaaggaa cctttaatagaaattggacagcaagaaagcgagcttagtgatacttgtgggccagggcattagccacacca gccaccactttctgataggcagcctgcactggtggggtgaattctttgccaaagtgatgggccagcacacag accagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaagggcctag cttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggattgtagctg ctattagcaatatgaaacctcttacatcagttacaatttatatgcagaaatatttatatgcagaaatattgct attgccttaacccagaaattatcactgttattctttagaatggtgcaaagaggcatgatacattgtatcattat tgccctgaaagaaagagattagggaaagtattagaaataagataaacaaaaaagtatattaaaagaaga aagcatttttttaaaattacaaatgcaaaattaccctgatttggtcaatatgtgtaccctgttacttctccccttc ctatgacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcaccctg aagttctcaggatccacgtgcagcttgtcacagtgcagctcactcagctgggcaaaggtgcccttgaggttg tccaggtgagccaggccatcactaaaggcaccgagcactttcttgccatgagccttcaccttagggttgccc ataacagcatcaggagtggacagatccccaaaggactcaaagaacctctgggtccaagggtagaccacc agcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagagtc ttctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccag

	ggcctcaccaccaacggcatccacgttcaccttgtcccacagggcagtaacggcagacttctcctcaggag tcaggtgcaccatggtgtctgtttgaggttgctagtgaacacagttgtgtcagaagcaaatgtaagcaatag atggctctgccctgacttttatgcccagccctggctcctgccctccctgctcctgggagtagattggccaaccc tagggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcactggct taggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaatt gctactaaaaacatcctcctttgcaagtgtatttacccgacgcgtcggcgataagcttgatccatcgatcatc aataattctagccccacaggagtttgttctgaaagtaaacttccacaaccgcaagcttattgaggctaaggc atctgtgaaggaaagaaacatctcctctaaaccactatgctgctagagcctcttttctgtactcaagcctcat tcagacactagtgtcaccagtctcctcatatacctattgtattttcttcttcttgctggtttagtcatgttttctgg gagcttaggggcttatttttatttttgttttgttttctaatcaacagagatgggcaaacccattatttttttctttag acttgggatggtgatagctgggcagcgtcagaaactgtgtgtggatatagataagagctcaggactatgct gagctgtgatgagggaggggcctagctaaaggcagtgagagtcagaatgctcctgctattgccttctcagt ccccacgcttggtttctacacaagtagatacatagaaaaggctataggttagtgtttgagagtcctgcatga ttagttgctcagaaatgcccgataaatatgttatgtgtgtttatgtatatatatgttttatatgtgtgtgtgtg tgttgtgtttacaaatatgtgattatcatcaaaacgtgagggtacgtatatgtgtatatatatatatatattca ggaaataatatattctagaatatgtcacattctgtctcaggcatccattttctttatgatgccgtttgaggtgg agttttagtcaggtggtcagcttctccttttttttgccatctgccctgtaagcatcctgctggggacccagata ggagtcatcactctaggctgagaacatctgggcacacaccctaagcctcagcatgactcatcatgactcag cattgctgtgcttgagccagaaggtttgcttagaaggttacacagaaccagaaggcgggggtggggcactg accccgacaggggcctggccagaactgctcatgcttggactatgggaggtcactaatggagacacacaga aatgtaacaggaactaaggaaaaactgaagctttgggggtataggggagcagtcccatgtagtagtagaa tgaaaaatgctgctatgctgtgcctcccccaccttccatgtctgccctctactcatggtctatctctcctggc tcctgggagtcatggactccacccagcaccaccaacctgacctaaccacctatctgagcctgccagcctata acccatctgggccctgatagctggtggccagccctgaccccaccccacccctccctggaacctctgatagaca catctggcacaccagctcgcaaagtcaccgtgagggtcttgtgtttgctgagtcaaaattccttgaaatcca agtccttagagactcccaggcttggattcaaagctcctgactttctgtctagtgtatgtgcagtgagccccttt tcctctaactgaaagaaggaaaaaaaaatggaacccaaaatattctacatagtttccatgtcacagccagg gctgggcagtctcctgttatttcttttaaaataaatatatcatttaaatgcataaataagcaaaccctgctcg ggaatgggagggagagtctctggagtccacccctctcggccctggctctgcagatagtgctatcaaagcc ctgacagagccctgcccattgctgggccttggagtgagtcagcctagtagaggaggcagggcaagccatctc atagctgctgagtgggagagagaaaagggctcattgtctataaactcaggtcatggctattcttatggccta ctcgaccacgagggaattccgataatcaacctctggattacaaaatttgtgaaagattgactggtattctta actatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctattgcttcccgtatggc tttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgt ggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctccttt ccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggac aggggctcggctgttgggcactgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctc gcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggacc ttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatc tccctttgggccgcctccccgcatcgataccgtcgacctcgagacctagaaaaacatggccaattcgagctc ggtacctttaagaccaatgacttacaaggcagctgtagatcttagccacttttaaaagaaaaggggggac tggaagggctaattcactcccaacgaagacaagatcccagggatgtacgtccctaacccgctaggggggca gcacccaggcctgcactgccgcctgccggcaggggtccagtcctgctttttgcttgtactgggtctctctggtt agaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaataaagcttgc cttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttttag tcagtgtggaaaatctctagcagtagtagttcatgtcatcttattattcagtatttataacttgcaaagaaatg aatatcagagagtgagaggaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaa atttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgt

	ctggctctagctatcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgc cccatggctgactaatttttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagta gtgaggaggctttttttggaggcctaggcttttgcgtcgagacgtacccaattcgccctatagtgagtcgtatt acgcgcgctcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgcc ttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagt tgcgcagcctgaatggcgaatggcgcgacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggtt acgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgc cacgttcgccggctttccccgtcaagctctaaatcgggggctcccttttagggttccgatttagtgctttacggc acctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttc gcccttttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctat ctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaaca aaaatttaacgcgaattttaacaaaatattaacgtttacaatttcccaggtggcacttttcggggaaatgtgc gcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataa atgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgc ggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggt gcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacg ttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagag caactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatctt acggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaactt acttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactc gccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgt agcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaat agactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgc tgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccct cccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctga gataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaa aacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacgt gagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcg cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctacc aactcttttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtag ttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctg ctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcgg tcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacc tacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcg gcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgt cgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaa cgccagcaacgcggccatttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatc ccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccga gcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttg gccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtga
Codon optimized exon 1 SEQ ID NO:7)	atgGTGCACCTGACCCCCGAGGAGAAGAGCGCCGTGACCGCCCTGT GGGACAAGGTGAACGTGGACGCCGTGGGCGGCGAGGCCCTGggca g
Codon optimized exon 2	cctgaagttctcggggtccacgtgcagcttgtcgcagtgcagctcgctcagctgggcgaaggtgcccttcag gttgtccaggtgggccaggccgtcgctgaaggcgcccagcaccttcttgccgtgggccttcaccttggggtt

SEQ ID NO:8)	gcccatcacggcgtcgggggtgctcaggtcgccgaagctctcgaagaagcgctgggtccaggggtacacc accagcagc Note -- this optimization in vector with exons 1, 2, and 3 all codon optimized. Other vectors use optimized exon 2 coEX2 (SEQ ID NO:5)
Codon optimized exon 3 SEQ ID NO:9)	ctcCTGGGCAACGTGCTGGTGTGCGTGCTGGCCCACCACTTCGGCA AGGAGTTCACCCCCCCCGTGCAGGCCGCCTACCAGAAGGTGGTGG CCGGCGTGGCCAACGCCCTGGCCCACAAGTACCACtaa
Codon optimized exon 1 full vector (SEQ ID NO:10)	gcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgt atgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcg cgcaattaaccctcactaaagggaacaaaagctggagctgcaagcttggccattgcatacgttgtatccat atcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagttatt aatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaa tggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaac gccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatca agtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccca gtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgc ggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattg acgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccat tgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggt ctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaat aaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcag acccttttagtcagtgtggaaaatctctagcagtggcgcccgaacagggacttgaaagcgaaagggaaac cagaggagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgac tggtgagtacgccaaaaattttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtat taagcgggggagaattagatcgcgatgggaaaaaattcggttaaggccaggggaaagaaaaaatataa attaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacat cagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagatc attatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctt tagacaagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcag acctggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaa ccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagagcagtgggaat aggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaa cagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagata cctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttgcaccactgctgtgccttgg aatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagag aaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatga acaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggta tataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtga atagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacggg cccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgattagtgaacggatct cgacggtatcgatctcgacacaaatggcagtattcatccacaattttaaaagaaaaggggggattggggg gtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaaca aattacaaaaattcaaaattttcgggtttattacagggacagcagagatccagtttgggtcgaggatatcgg atcggaattctctagatgatcaggatccctcgagcccttatcgatcacgagactagcctcgactactagtgg agatcccccgggctgcagagccagaagcaccataagggacatgataagggagccagcagacctctgatct

	cttcctgaatgctaatcttaaacatcctgaggaagaatgggacttccatttggggtgggcctatgatagggt aataagacagtagtgaatatcaagctacaaaaagccccctttcaaattcttctcagtcctaacttttcatact aagcccagtccttccaaagcagactgtgaaagagtgatagttccgggagactagcactgcagattccgggt cactgtgagtgggggaggcagggaagaagggctcacaggacagtcaaaccatgccccctgtttttccttct tcaagtagacctctataagacaacagagacaactaaggctgagtggccaggcgaggagaaaccatctcg ccgtaaaacatggaaggaacacttcaggggaaaggtggtatctctaagcaagagaactgagtggagtca aggctgagagatgcaggataagcaaatgggtagtgaaaagacattcatgaggacagctaaaacaataag taatgtaaaatacagcatagcaaaactttaacctccaaatcaagcctctacttgaatccttttctgagggat gaataaggcataggcatcaggggctgttgccaatgtgcattagctgtttgcagcctcaccttctttcatggag tttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcactgacctccca cattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttttttattag gcagaatccagatgctcaaggcccttcataatatcccccagtttagtagttggacttagggaacaaaggaa cctttaatagaaattggacagcaagaaagcgagcttagtgatacttgtgggccagggcattagccacacca gccaccactttctgataggcagcctgcactggtggggtgaattctttgccaaagtgatgggccagcacacag accagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaagggcctag cttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggattgtagctg ctattagcaatatgaaacctcttacatcagttacaatttatatgcagaaatatttatatgcagaaatattgct attgccttaacccagaaattatcactgttattctttagaatggtgcaaagaggcatgatacattgtatcattat tgccctgaaagaaagagattagggaaagtattagaaataagataaacaaaaaagtatattaaaagaaga aagcatttttttaaaattacaaatgcaaaattaccctgatttggtcaatatgtgtaccctgttacttctccccttc ctatgacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcaccctg aagttctcaggatccacgtgcagcttgtcacagtgcagctcactcagctgggcaaaggtgcccttgaggttg tccaggtgagccaggccatcactaaaggcaccgagcactttcttgccatgagccttcaccttagggttgccc ataacagcatcaggagtggacagatccccaaaggactcaaagaacctctgggtccaagggtagaccacc agcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagagtc ttctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccag ggcctcgccgcccacggcgtccacgttcaccttgtcccacagggcggtcacggcgctcttctcctcgggggt caggtgcaccatggtgtctgtttgaggttgctagtgaacacagttgtgtcagaagcaaatgtaagcaatag atggctctgccctgacttttatgcccagccctggctcctgccctccctgctcctgggagtagattggccaaccc tagggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcactggct taggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaatt gctactaaaaacatcctcctttgcaagtgtatttacccgacgcgtcggcgataagcttgatccatcgattacg tatatgtgtatatatatatatatattcaggaaataatatattctagaatatgtcacattctgtctcaggcatcc attttctttatgatgccgtttgaggtggagttttagtcaggtggtcagcttctccttttttttgccatctgccctgt aagcatcctgctggggacccagataggagtcatcactctaggctgagaacatctgggcacacaccctaagc ctcagcatgactcatcatgactcagcattgctgtgcttgagccagaaggtttgcttagaaggttacacagaa ccagaaggcgggggtggggcactgaccccgacaggggcctggccagaactgctcatgcttggactatggg aggtcactaatggagacacacagaaatgtaacaggaactaaggaaaaactgaagctttgggggtatagg ggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcccccaccttttcccatgtctgcc ctctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgacctaac cacctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccctgaccccacccc accctccctggaacctctgatagacacatctggcacaccagctcgcaaagtcaccgtgagggtcttgtgttt gctgagtcaaaattccttgaaatccaagtccttagagactcccaggcttggattcaaagctcctgactttctg tctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaaaaaaaatggaacccaaaatatt ctacatagtttccatgtcacagccagggctgggcagtctcctgttatttcttttaaaataaatatatcatttaa atgcataaataagcaaaccctgctcgggaatgggagggagagtctctggagtccacccctttctcggccctg gctctgcagatagtgctatcaaagccctgacagagccctgcccattgctgggccttggagtgagtcagccta gtagagaggcagggcaagccatctcatagctgctgagtgggagagagaaaagggctcattgtctataaac

	tcaggtcatggctattcttatggcctactcgaccacgagggaattccgataatcaacctctggattacaaaa tttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgccttt gtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgagg agttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggg gcattgccaccacctgtcagctcctttccgggactttcgctttcccctccctattgccacggcggaactcatc gccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcgggg aaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtc ccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcg ccttcgccctcagacgagtcggatctccctttgggccgcctccccgcatcgataccgtcgacctcgagaccta gaaaaacatggccaattcgagctcggtacctttaagaccaatgacttacaaggcagctgtagatcttagcc acttttaaaagaaaaggggggactggaagggctaattcactcccaacgaagacaagatcccagggatgt acgtccctaacccgctagggggcagcacccaggcctgcactgccgcctgccggcaggggtccagtcctgct ttttgcttgtactgggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaaccca ctgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaa ctagagatccctcagacccttttagtcagtgtggaaaatctctagcagtagtagttcatgtcatcttattattc agtatttataacttgcaaagaaatgaatatcagagagtgagaggaacttgtttattgcagcttataatggtt acaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtc caaactcatcaatgtatcttatcatgtctggctctagctatcccgcccctaactccgcccatcccgcccctaac tccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcct cggcctctgagctattccagaagtagtgaggaggctttttggaggcctaggcttttgcgtcgagacgtaccc aattcgccctatagtgagtcgtattacgcgcgctcactggccgtcgttttacaacgtcgtgactgggaaaac cctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcc cgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcgacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctc ctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttt agggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgg gccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttcca aactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctatt ggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttccc aggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatc cgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacat ttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaag taaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatc cttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtatt atcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagta ctcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataacca tgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgc acaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacga cgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactactta ctctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcg gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcag cactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggat gaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagttta ctcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataa tctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaagg atcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtg gtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatacca aatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcg

	ctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacg atagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcga acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggaga aaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggg aaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtc agggggggcggagcctatggaaaaacgccagcaacgcggccatttttacggttcctggccttttgctggcctt ttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgata ccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaata cgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaa gcgggcagtga
Codon optimized exon 1 and 2 full vector (SEQ ID NO:11)	gcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgt atgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcg cgcaattaaccctcactaaagggaacaaaagctggagctgcaagcttggccattgcatacgttgtatccat atcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagttatt aatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaa tggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaac gccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatca agtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccca gtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgc ggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattg acgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccat tgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggt ctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaat aaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcag acccttttagtcagtgtggaaaatctctagcagtggcgcccgaacagggacttgaaagcgaaagggaaac cagaggagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgac tggtgagtacgccaaaaattttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtat taagcgggggagaattagatcgcgatgggaaaaaattcggttaaggccaggggaaagaaaaaatataa attaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacat cagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagatc attatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctt tagacaagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcag acctggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaa ccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagagcagtgggaat aggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaa cagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagata cctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttgcaccactgctgtgccttgg aatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagag aaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatga acaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggta tataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtga atagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacggg cccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgattagtgaacggatct cgacggtatcgatctcgacacaaatggcagtattcatccacaattttaaaagaaaaggggggattggggg gtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaaca aattacaaaaattcaaaattttcgggtttattacagggacagcagagatccagtttgggtcgaggatatcgg atcggaattctctagatgatcaggatccctcgagcccttatcgatcacgagactagcctcgactactagtgg

	agatcccccgggctgcagagccagaagcaccataagggacatgataagggagccagcagacctctgatct cttcctgaatgctaatcttaaacatcctgaggaagaatgggacttccatttggggtgggcctatgatagggt aataagacagtagtgaatatcaagctacaaaaagccccctttcaaattcttctcagtcctaacttttcatact aagcccagtccttccaaagcagactgtgaaagagtgatagttccgggagactagcactgcagattccgggt cactgtgagtgggggaggcagggaagaagggctcacaggacagtcaaaccatgccccctgtttttccttct tcaagtagacctctataagacaacagagacaactaaggctgagtggccaggcgaggagaaaccatctcg ccgtaaaacatggaaggaacacttcaggggaaaggtggtatctctaagcaagagaactgagtggagtca aggctgagagatgcaggataagcaaatgggtagtgaaaagacattcatgaggacagctaaaacaataag taatgtaaaatacagcatagcaaaactttaacctccaaatcaagcctctacttgaatccttttctgagggat gaataaggcataggcatcaggggctgttgccaatgtgcattagctgtttgcagcctcaccttctttcatggag tttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcactgacctccca cattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttttttattag gcagaatccagatgctcaaggcccttcataatatcccccagtttagtagttggacttagggaacaaaggaa cctttaatagaaattggacagcaagaaagcgagcttagtgatacttgtgggccagggcattagccacacca gccaccactttctgataggcagcctgcactggtggggtgaattctttgccaaagtgatgggccagcacacag accagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaagggcctag cttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggattgtagctg ctattagcaatatgaaacctcttacatcagttacaatttatatgcagaaatatttatatgcagaaatattgct attgccttaacccagaaattatcactgttattctttagaatggtgcaaagaggcatgatacattgtatcattat tgccctgaaagaaagagattagggaaagtattagaaataagataaacaaaaaagtatattaaaagaaga aagcatttttttaaaattacaaatgcaaaattaccctgatttggtcaatatgtgtaccctgttacttctccccttc ctatgacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcaccctg aagttctcaggatccacgtgcagcttgtcgcagtgcagctcgctcagctgggcgaaggtgcccttcaggttg tccaggtgggccaggccgtcgctgaaggcgcccagcaccttcttgccgtgggccttcaccttggggttgccc atcacggcgtcgggggtgctcaggtcgccgaagctctcgaagaagcgctgggtccaggggtacaccacca gcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagagtct tctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccagg gcctcgccgcccacggcgtccacgttcaccttgtcccacagggcggtcacggcgctcttctcctcgggggtc aggtgcaccatggtgtctgtttgaggttgctagtgaacacagttgtgtcagaagcaaatgtaagcaatagat ggctctgccctgacttttatgcccagccctggctcctgccctccctgctcctgggagtagattggccaacccta gggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcactggctta ggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaattgc tactaaaaacatcctcctttgcaagtgtatttacccgacgcgtcggcgataagcttgatccatcgattacgta tatgtgtatatatatatatatattcaggaaataatatattctagaatatgtcacattctgtctcaggcatccat tttctttatgatgccgtttgaggtggagttttagtcaggtggtcagcttctccttttttttgccatctgccctgta agcatcctgctggggacccagataggagtcatcactctaggctgagaacatctgggcacacaccctaagcc tcagcatgactcatcatgactcagcattgctgtgcttgagccagaaggtttgcttagaaggttacacagaac cagaaggcgggggtggggcactgaccccgacaggggcctggccagaactgctcatgcttggactatggga ggtcactaatggagacacacagaaatgtaacaggaactaaggaaaaactgaagctttgggggtataggg gagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcccccacctttcccatgtctgccc tctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgacctaacc acctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccctgaccccacccca ccctccctggaacctctgatagacacatctggcacaccagctcgcaaagtcaccgtgagggtcttgtgtttg ctgagtcaaaattccttgaaatccaagtccttagagactcccaggcttggattcaaagctcctgactttctgt ctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaaaaaaaatggaacccaaaatattc tacatagtttccatgtcacagccagggctgggcagtctcctgttatttcttttaaaataaatatatcatttaaa tgcataaataagcaaaccctgctcgggaatgggagggagagtctctggagtccacccccttctcggccctgg ctctgcagatagtgctatcaaagccctgacagagccctgcccattgctgggccttggagtgagtcagcctag

	tagagaggcagggcaagccatctcatagctgctgagtgggagagagaaaagggctcattgtctataaact caggtcatggctattcttatggcctactcgaccacgagggaattccgataatcaacctctggattacaaaat ttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttg tatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgagga gttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggc attgccaccacctgtcagctcctttccgggactttcgctttcccctccctattgccacggcggaactcatcgc cgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaa atcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtccc ttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgcc ttcgccctcagacgagtcggatctccctttgggccgcctccccgcatcgataccgtcgacctcgagacctag aaaaacatggccaattcgagctcggtacctttaagaccaatgacttacaaggcagctgtagatcttagcca ctttttaaaagaaaaggggggactggaagggctaattcactcccaacgaagacaagatcccagggatgta cgtccctaacccgctagggggcagcacccaggcctgcactgccgcctgccggcaggggtccagtcctgcttt ttgcttgtactgggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccact gcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaact agagatccctcagacccttttagtcagtgtggaaaatctctagcagtagtagttcatgtcatcttattattcag tatttataacttgcaaagaaatgaatatcagagagtgagaggaacttgtttattgcagcttataatggttac aaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtcca aactcatcaatgtatcttatcatgtctggctctagctatcccgcccctaactccgcccatcccgcccctaactc cgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctcg gcctctgagctattccagaagtagtgaggaggctttttggaggcctaggcttttgcgtcgagacgtaccca attcgccctatagtgagtcgtattacgcgcgctcactggccgtcgttttacaacgtcgtgactgggaaaaccc tggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccg caccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcgacgcgccctgtagcggcgcat taagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctt tcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttttag ggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggc catcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaa actggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattg gttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttccca ggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatcc gctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacattt ccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagt aaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatcc ttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtatta tcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtac tcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccat gagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgc acaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacga cgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactactta ctctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcg gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcag cactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggat gaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagttta ctcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataa tctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaagg atcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtg gtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatacca

	aatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcg ctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacg atagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcga acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggaga aaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggg aaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtc agggggcggagcctatggaaaaacgccagcaacgcggccatttttacggttcctggccttttgctggcctt ttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgata ccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaata cgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaa gcgggcagtga
Codon optimized exon 1, 2, and 3 full vector (SEQ ID NO:12)	gcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgt atgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcg cgcaattaaccctcactaaagggaacaaaagctggagctgcaagcttggccattgcatacgttgtatccat atcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagttatt aatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaa tggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaac gccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatca agtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgccca gtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgc ggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattg acgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccat tgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggt ctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaat aaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcag acccttttagtcagtgtggaaaatctctagcagtggcgcccgaacagggacttgaaagcgaaagggaaac cagaggagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgac tggtgagtacgccaaaaattttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtat taagcgggggagaattagatcgcgatgggaaaaaattcggttaaggccaggggaaagaaaaaatataa attaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacat cagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagatc attatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctt tagacaagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcag acctggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaa ccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagagcagtgggaat aggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaa cagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagata cctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttgcaccactgctgtgccttgg aatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagag aaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatga acaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggta tataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtga atagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacagg cccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgattagtgaacggatct cgacggtatcgatctcgacacaaatggcagtattcatccacaattttaaaagaaaaggggggattggggg gtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaaca aattacaaaaattcaaaattttcgggtttattacagggacagcagagatccagtttgggtcgaggatatcgg

	atcggaattctctagatgatcaggatccctcgagcccttatcgatcacgagactagcctcgactactagtgg agatcccccgggctgcagagccagaagcaccataagggacatgataagggagccagcagacctctgatct cttcctgaatgctaatcttaaacatcctgaggaagaatgggacttccatttggggtgggcctatgatagggt aataagacagtagtgaatatcaagctacaaaaagccccctttcaaattcttctcagtcctaacttttcatact aagcccagtccttccaaagcagactgtgaaagagtgatagttccgggagactagcactgcagattccgggt cactgtgagtgggggaggcagggaagaagggctcacaggacagtcaaaccatgccccctgtttttccttct tcaagtagacctctataagacaacagagacaactaaggctgagtggccaggcgaggagaaaccatctcg ccgtaaaacatggaaggaacacttcaggggaaaggtggtatctctaagcaagagaactgagtggagtca aggctgagagatgcaggataagcaaatgggtagtgaaaagacattcatgaggacagctaaaacaataag taatgtaaaatacagcatagcaaaactttaacctccaaatcaagcctctacttgaatccttttctgagggat gaataaggcataggcatcaggggctgttgccaatgtgcattagctgtttgcagcctcaccttctttcatggag tttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcactgacctccca cattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttttttattag gcagaatccagatgctcaaggcccttcataatatcccccagtttagtagttggacttagggaacaaaggaa cctttaatagaaattggacagcaagaaagcgagcttagtggtacttgtgggccagggcgttggccacgccg gccaccaccttctggtaggcggcctgcacgggggggtgaactccttgccgaagtggtgggccagcacgc acaccagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaagggcct agcttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggattgtagc tgctattagcaatatgaaacctcttacatcagttacaatttatatgcagaaatatttatatgcagaaatattg ctattgccttaacccagaaattatcactgttattctttagaatggtgcaaagaggcatgatacattgtatcatt attgccctgaaagaaagagattagggaaagtattagaaataagataaacaaaaaagtatattaaaagaa gaaagcatttttttaaaattacaaatgcaaaattaccctgatttggtcaatatgtgtaccctgttacttctcccc ttcctatgacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcaccc tgaagttctcggggtccacgtgcagcttgtcgcagtgcagctcgctcagctgggcgaaggtgcccttcaggt tgtccaggtgggccaggccgtcgctgaaggcgcccagcaccttcttgccgtgggccttcaccttggggttgc ccatcacggcgtcgggggtgctcaggtcgccgaagctctcgaagaagcgctgggtccaggggtacaccac cagcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagagt cttctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgccca gggcctcgccgcccacggcgtccacgttcaccttgtcccacagggcggtcacggcgctcttctcctcggggg tcaggtgcaccatggtgtctgtttgaggttgctagtgaacacagttgtgtcagaagcaaatgtaagcaatag atggctctgccctgacttttatgcccagccctggctcctgccctccctgctcctgggagtagattggccaaccc tagggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcactggct taggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaatt gctactaaaaacatcctcctttgcaagtgtatttacccgacgcgtcggcgataagcttgatccatcgattacg tatatgtgtatatatatatatatattcaggaaataatatattctagaatatgtcacattctgtctcaggcatcc attttctttatgatgccgtttgaggtggagttttagtcaggtggtcagcttctccttttttttgccatctgccctgt aagcatcctgctggggacccagataggagtcatcactctaggctgagaacatctgggcacacaccctaagc ctcagcatgactcatcatgactcagcattgctgtgcttgagccagaaggtttgcttagaaggttacacagaa ccagaaggcgggggtggggcactgaccccgacaggggcctggccagaactgctcatgcttggactatggg aggtcactaatggagacacacagaaatgtaacaggaactaaggaaaaactgaagctttgggggtatagg ggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcccccacctttcccatgtctgcc ctctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgacctaac cacctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccctgaccccacccc accctccctggaacctctgatagacacatctggcacaccagctcgcaaagtcaccgtgagggtcttgtgttt gctgagtcaaaattccttgaaatccaagtccttagagactcccaggcttggattcaaagctcctgactttctg tctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaaaaaaaatggaacccaaaatatt ctacatagtttccatgtcacagccagggctgggcagtctcctgttatttcttttaaaataaatatatcatttaa atgcataaataagcaaaccctgctcgggaatgggagggagagtctctggagtccacccctttctcggccctg

	gctctgcagatagtgctatcaaagccctgacagagccctgcccattgctgggccttggagtgagtcagccta gtagagaggcagggcaagccatctcatagctgctgagtgggagagagaaaagggctcattgtctataaac tcaggtcatggctattcttatggcctactcgaccacgagggaattccgataatcaacctctggattacaaaa tttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgccttt gtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgagg agttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggg gcattgccaccacctgtcagctcctttccgggactttcgctttcccctccctattgccacggcggaactcatc gccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcgggg aaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtc ccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcg ccttcgccctcagacgagtcggatctccctttgggccgcctccccgcatcgataccgtcgacctcgagaccta gaaaaacatggccaattcgagctcggtacctttaagaccaatgacttacaaggcagctgtagatcttagcc acttttttaaaagaaaaggggggactggaagggctaattcactcccaacgaagacaagatcccagggatgt acgtccctaacccgctaggggcagcacccaggcctgcactgccgcctgccggcaggggtccagtcctgct ttttgcttgtactgggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaaccca ctgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaa ctagagatccctcagacccttttagtcagtgtggaaaatctctagcagtagtagttcatgtcatcttattattc agtatttataacttgcaaagaaatgaatatcagagagtgagaggaacttgtttattgcagcttataatggtt acaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtc caaactcatcaatgtatcttatcatgtctggctctagctatcccgcccctaactccgcccatcccgcccctaac tccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcct cggcctctgagctattccagaagtagtgaggaggctttttggaggcctaggcttttgcgtcgagacgtaccc aattcgccctatagtgagtcgtattacgcgcgctcactggccgtcgttttacaacgtcgtgactgggaaaac cctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcc cgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcgacgcgccctgtagcggcgc attaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctc ctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttt agggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgg gccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttcca aactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctatt ggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttccc aggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatc cgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacat ttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaag taaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatc cttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtatt atcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagta ctcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataacca tgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgc acaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacga cgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactactta ctctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcg gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcag cactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggat gaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttta ctcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataa tctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaagg atcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtg

	gtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatacca aatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcg ctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacg atagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcga acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggaga aaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggg aaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtc agggggggcggagcctatggaaaaacgccagcaacgcggccatttttacggttcctggccttttgctggcctt ttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgata ccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaata cgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaa gcgggcagtga
pUV-AS3 from junction marker (SEQ ID NO:13)	ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTA CGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTC CTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCC CGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGT GCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGT TCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGA CGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGG ATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAAC AAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTT AGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTT ATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATA ACCCTGATAAATGCTTCAATAATAGCACCTAGATCAAGAGACAGG ATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGG TTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGC ACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTC AGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGG TGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCT GGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCAC TGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGC AGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCAT CATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTAC CTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAC GTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACG AAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTC AAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGG CGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTC TGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCA GGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGG CGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCC GATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCT GAATTATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTAC GCATCTGTGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGG GAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTC AAATATGTATCCGCTCATGACCAAAATCCCTTAACGTGAGTTTTC GTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTC TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAA AAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTA

	CCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATA CCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCA AGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTT ACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTT GGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCT GAACGGGGGGGTCGTGCACACAGCCCAGCTTGGAGCGAACGACC TACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGC CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCG GCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGA AACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGAC TTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTAT GGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT GCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCT GTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGC CGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC GGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCC GATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGC GGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTA GGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGT GGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGA CCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACA AAAGCTGGAGCTGCAAGCTTGGCCATTGCATACGTTGTATCCATA TCATAATATGTACATTTATATTGGCTCATGTCCAACATTACCGCCA TGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGG GGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAAC TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCC CATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAG GGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCC CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC AGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGT ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATC AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTC CACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA CGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAA ATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGC TCGTTTAGTGAACCGGGGTCTCTCTGGTTAGACCAGATCTGAGCC TGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAA TAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGT GTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGT GGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGA AAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCT GAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTAC GCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGT GCGAGAGCGTCAGTATTAAGCGGGGGAGAATTAGATCGCGATGG GAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAATT AAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAG TTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATAC TGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTA GATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAA GGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAG

	GAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGC TGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGA AGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGA GTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAA AAAGAGCAGTGGGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAG CAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTA CAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAAT TTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACA GTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGA AAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTC TGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGTTG GAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGAT GGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACT CCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAA GAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTT AACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATA GTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTACTTTCTA TAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGA CCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATA GAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATT AGTGAACGGATCTCGACGGTATCGATCTCGACACAAATGGCAGTA TTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGGTACAG TGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACAAA CTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGG TTTATTACAGGGACAGCAGACATGAGGACAGCTAAAACAATAAG TAATGTAAAATACAGCATAGCAAAACTTTAACCTCCAAATCAAGC CTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATC AGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCT TTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAG CTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCT TTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATG AAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCC TTCATAATATCCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGG AACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTAGTGAT ACTTGTGGGCCAGGGCATTAGCCACACCAGCCACCACTTTCTGAT AGGCAGCCTGCACTGGTGGGGTGAATTCTTTGCCAAAGTGATGGG CCAGCACACAGACCAGCACGTTGCCCAGGAGCTGTGGGAGGAAG ATAAGAGGTATGAACATGATTAGCAAAAGGGCCTAGCTTGGACT CAGAATAATCCAGCCTTATCCCAACCATAAAATAAAAGCAGAAT GGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACA TCAGTTACAATTTATATGCAGAAATACCCTGTTACTTCTCCCCTTC CTATGACATGAACTTAACCATAGAAAAGAAGGGGAAAGAAAACA TCAAGGGTCCCATAGACTCACCCTGAAGTTCTCAGGATCCACGTG CAGCTTGTCACAGTGCAGCTCACTCAGCTGGGCAAAGGTGCCCTT GAGGTTGTCCAGGTGAGCCAGGCCATCACTAAAGGCACCGAGCA CTTTCTTGCCATGAGCCTTCACCTTAGGGTTGCCCATAACAGCATC AGGAGTGGACAGATCCCCAAAGGACTCAAAGAACCTCTGGGTCC AAGGGTAGACCACCAGCAGCCTAAGGGTGGGAAAATAGACCAAT AGGCAGAGAGAGTCAGTGCCTATCAGAAACCCAAGAGTCTTCTCT GTCTCCACATGCCCAGTTTCTATTGGTCTCCTTAAACCTGTCTTGT AACCTTGATACCAACCTGCCCAGGGCCTCACCACCAACGGCATCC

	ACGTTCACCTTGTCCCACAGGGCAGTAACGGCAGACTTCTCCTCA GGAGTCAGGTGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACA CAGTTGTGTCAGAAGCAAATGTAAGCAATAGATGGCTCTGCCCTG ACTTTTATGCCCAGCCCTGGCTCCTGCCCTCCCTGCTCCTGGGAGT AGATTGGCCAACCCTAGGGTGTGGCTCCACAGGGTGAGGTCTAAG TGATGACAGCCGTACCTGTCCTTGGCTCTTCTGGCACTGGCTTAGG AGTTGGACTTCAAACCCTCAGCCCTCCCTCTAAGATATATCTCTTG GCCCCATACCATCAGTACAAATTGCTACTAAAAACATCCTCCTTT GCAAGTGTATTTACTACGTATATGTGTATATATATATATATATTCA GGAAATAATATATTCTAGAATATGTCACATTCTGTCTCAGGCATC CATTTTCTTTATGATGCCGTTTGAGGTGGAGTTTTAGTCAGGTGGT CAGCTTCTCCTTTTTTTTGCCATCTGCCCTGTAAGCATCCTGCTGG GGACCCAGATAGGAGTCATCACTCTAGGCTGAGAACATCTGGGC ACACACCCTAAGCCTCAGCATGACTCATCATGACTCAGCATTGCT GTGCTTGAGCCAGAAGGTTTGCTTAGAAGGTTACACAGAACCAGA AGGCGGGGGTGGGGCACTGACCCCGACAGGGGCCTGGCCAGAAC TGCTCATGCTTGGACTATGGGAGGTCACTAATGGAGACACACAGA AATGTAACAGGAACTAAGGAAAAACTGAAGCTTTGGGGGTATAG GGGAGCAGTCCCATGTAGTAGTAGAATGAAAAATGCTGCTATGCT GTGCCTCCCCCACCTTTCCCATGTCTGCCCTCTACTCATGGTCTAT CTCTCCTGGCTCCTGGGAGTCATGGACTCCACCCAGCACCACCAA CCTGACCTAACCACCTATCTGAGCCTGCCAGCCTATAACCCATCT GGGCCCTGATAGCTGGTGGCCAGCCCTGACCCCACCCCACCCTCC CTGGAACCTCTGATAGACACATCTGGCACACCAGCTCGCAAAGTC ACCGTGAGGGTCTTGTGTTTGCTGAGTCAAAATTCCTTGAAATCC AAGTCCTTAGAGACTCCCAGGCTTGGATTCAAAGCTCCTGACTTT CTGTCTAGTGTATGTGCAGTGAGCCCCTTTTCCTCTAACTGAAAGA AGGAAAAAAAAAATGGAACCCAAAATATTCTACATAGTTTCCATGT CACAGCCAGGGCTGGGCAGTCTCCTGTTATTTCTTTTAAAATAAA TATATCATTTAAATGCATAAATAAGCAAACCCTGCTCGGGAATGG GAGGGAGAGTCTCTGGAGTCCACCCCTTCTCGGCCCTGGCTCTGC AGATAGTGCTATCAAAGCCCTGACAGAGCCCTGCCCATTGCTGGG CCTTGGAGTGAGTCAGCCTAGTAGAGAGGCAGGGCAAGCCATCT CATAGCTGCTGAGTGGGAGAGAGAAAAGGGCTCATTGTCTATAA ACTCAGGTCATGGCTATTCTTATTAAAAGAAAAGGGGGGACTGGA AGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTG TACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTC TGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCC TTGAGTGCTTCAAGTAGTGTGTTGGTTTTTTGTGTGTGCATTGTCT GAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGCACAG CAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG CGGTGGGCTCTATGGAGTAGTAGTTCATGTCATCTTATTATTCAGT ATTTATAACTTGCAAAGAAATGAATATCAGAGAGTGAGAGGAAC TTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATC ACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTG GTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGCTCTAGCT ATCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTT CCGCCCATTCTCCGCCCCATGGCTCTCACTACTTCTGGAATAGCTC AGAGGCCGAGGCGGCCTCGGCCTCTGCATAAATAAAAAAAATTA GTCGAGGCTTTTTTGGAGGCCTAGGGACGTACCCAATTCGCCCTA TAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACG

	TCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGC AGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCG CACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATG GG
pUV-AS3(coX2) from junction marker (SEQ ID NO:14)	acgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgcc agcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagct ctaaatcgggggctcccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagg gtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttcttt aatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggat tttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaata ttaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaata cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatagcacctagatcaagag acaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtgga gaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagc gcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcag cgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaa gggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaag tatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaag cgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacga agagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgagcatgcccgacggcgaggat ctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcg actgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagag cttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgcct tctatcgccttcttgacgagttcttctgaattattaacgcttacaatttcctgatgcggtattttctccttacgca tctgtgcggtatttcacaccgcatcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttattt ttctaaatacattcaaatatgtatccgctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtc agaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgcaaaca aaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaact ggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaac tctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgt gtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggtc gtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaa agcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggaga gcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactt gagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttta cggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgta ttaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcga ggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggc acgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattag gcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcaca caggaaacagctatgaccatgattacgccaagcgcgcaattaaccctcactaaagggaacaaaagctgg agctgcaagcttggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaaca ttaccgccatgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccc atatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgccc attgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtgga gtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtc

	aatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtac atctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggt ttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacg ggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggagg tctatataagcagagctcgtttagtgaaccgGGGTCTCTCTGGTTAGACCAGATCTGAGCCT GGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCC TTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGA TCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCagtggcgcccgaacagggact tgaaagcgaaagggaaaccagaggagctctcTCGACGCAGGACTCGGCTTGCTGAAGCGC GCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACT AGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGG GGAGaattagatcgcgatgggaaaaaattcggttaaggccaggggaaagaaaaaatataaattaaa acatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacatcagaa ggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagatcattata taatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagac aagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcGGCCGCTgatcttcagacc tggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaacca ttaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagaGCAGTGGGAAT aggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaa cagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagata cctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttgcaccactgctgtgccttgg aatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagag aaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatga acaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggta tataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtga atagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacagg cccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgattagtgaacggatC TCGACGGTATCGATCTCGAcacaaatggcagtattcatccacaattttaaaagaaaaggggggat tgggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattaca aaaacaaattacaaaaattcaaaattttcgggtttattacagggacagcagacatgaggacagctaaaac aataagtaatgtaaaatacagcatagcaaaactttaacctccaaatcaagcctctacttgaatccttttctg agggatgaataaggcataggcatcaggggctgttgccaatgtgcattagctgtttgcagcctcaccttctttc atggagtttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcactga cctcccacattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttt ttattaggcagaatccagatgctcaaggcccttcataatatcccccagtttagtagttggacttagggaaca aaggaacctttaatagaaattggacagcaagaaagcgagcttagtgatacttgtgggccagggcattagc cacaccagccaccactttctgataggcagcctgcactggtggggtgaattctttgccaaagtgatgggccag cacacagaccagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaag ggcctagcttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggatt gtagctgctattagcaatatgaaacctcttacatcagttacaatttatatgcagaaataccctgttacttctcc ccttcctatgacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcac cctGAAGTTCTCaGGaTCCACGTGCAGCTTGTCGCAGTGCAGCTCGCTCAGCTGGG CGAAGGTGCCCTTCAGGTTGTCCAGGTGGGCCAGGCCGTCGCTGAAGGCGCCCA GCACCTTCTTGCCGTGGGCCTTCACCTTGGGGTTGCCCATCACGGCGTCGGGGGT GCTCAGGTCGCCGAAGCTCTCGAAGAAGCGCTGGGTCCAGGGGTACACCACCAG cagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagagtcttc tctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccaggg

	cctcaccaccaacggcatccacgttcaccttgtcccacagggcagtaacggcagacttctcctcaggagtca ggtgcaccatggtgtctgtttgaggttgctagtgaacacagttgtgtcagaagcaaatgtaagcaatagatg gctctgccctgacttttatgcccagccctggctcctgccctccctgctcctgggagtagattggccaaccctag ggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcactggcttag gagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaattgct actaaaaacatcctcctttgcaagtgtatttacTACGTaTATGTGTATATATATATATATATTCA GGAAATAATATATTCTAGAATATGTCACATTCTGTCTCAGGCATCCATTTTCTTTAT GATGCCGTTTGAGGTGGAGTTTTAGTCAGGTGGTCAGCTTCTCCTTTTTTTTGCCA TCTGCCCTGTAAGCATCCTGCTGGGGACCCAGATAGGAGTCATCACTCTAGGCTG AGAACATCTGGGCACACACCCTAAGCCTCAGCATGACTCATCATGACTCAGCATT GCTGTGCTTGAGCCAGAAGGTTTGCTTAGAAGGTTACACAGAACCAGAAGGCGG GGGTGGGGCACTGACCCCGACAGGGGCCTGGCCAGAACTGCTCATGCTTGGACT ATGGGAGGTCACTAATGGAGACACACAGAAATGTAACAGGAACTAAGGAAAAA CTGAAGCTTtgggggtataggggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgc ctcccccacctttcccatgtctgccctctactcatggtctatctctcctggctcctgggagtcatggactccacc cagcaccaccaacctgacctaaccacctatctgagcctgccagcctataacccatctgggccctgatagctg gtggccagccctgaccccaccccacccctccctggaacctctgatagacacatctggcacaccagctcgcaa agtcaccgtgagggtcttgtgtttgctgagtcaaaattccttgaaatccaagtccttagagactcccaggctt ggattcaaagctcctgactttctgtctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaa aaaaaatggaacccaaaatattctacatagtttccatgtcacagccagggctgggcagtctcctgttatttct tttaaaataaatatatcatttaaatgcataaataagcaaaccctgctcgggaatgggagggagagtctctg gagtccacccttctcggccctggctctgcagatagtgctatcaaagccctgacagagccctgcccattgct gggccttggagtgagtcagcctagtagagaggcagggcaagccatctcatagctgctgagtgggagagag aaaagggctcattgtctataaactcaggtcatggctattcttattaaaagaaaaggggggactggaagggc taattcactcccaacgaagacaagatctgctttttgcttgtactGGGTCTCTCTGGTTAGACCAGA TCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATA AAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTTGGTTTTTTGTGTGTGCATTGTCTG AGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGCACAGCAAGGGGGA GGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGagta gtagttcatgtcatcttattattcagtatttataacttgcaaagaaatgaatatcagagagtgagaggaactt gtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttca ctgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggctctagctatcccgcccctaa ctccgcccATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTC TCACTACTTCTGGAATAGCTCAGAGGCCGAGGCGGCCTCGGCCTCTGCATAAATA AAAAAAATTAGTCGAGGCTTTTTTGGAGGCCTAGGgacgtacccaattcgccctatagtga gtcgtattacgcgcgctcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaactt aatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcc caacagttgcgcagcctgaatggcgaatggg
pUV-HS1-AS3 from junction marker (SEQ ID NO:15)	acgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgcc agcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagct ctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagg gtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttcttt aatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggat tttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaata ttaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaata cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatagcacctagatcaagag

	acaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtgga gaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagc gcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcag cgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaa gggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaag tatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaag cgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacga agagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgagcatgcccgacggcgaggat ctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcg actgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagag cttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgcct tctatcgccttcttgacgagttcttctgaattattaacgcttacaatttcctgatgcggtattttctccttacgca tctgtgcggtatttcacaccgcatcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttattt ttctaaatacattcaaatatgtatccgctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtc agaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgcaaaca aaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaact ggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaac tctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgt gtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggtc gtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaa agcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggaga gcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactt gagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttta cggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgta ttaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcga ggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggc acgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattag gcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcaca caggaaacagctatgaccatgattacgccaagcgcgcaattaaccctcactaaagggaacaaaagctgg agctgcaagcttggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaaca ttaccgccatgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccc atatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgccc attgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtgga gtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtc aatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtac atctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggt ttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacg ggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggagg tctatataagcagagctcgtttagtgaaccgGGGTCTCTCTGGTTAGACCAGATCTGAGCCT GGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCC TTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGA TCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCagtggcgcccgaacagggact tgaaagcgaaagggaaaccagaggagctctcTCGACGCAGGACTCGGCTTGCTGAAGCGC GCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACT AGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGG GGAGaattagatcgcgatgggaaaaaattcggttaaggccaggggaaagaaaaatataaattaaa acatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacatcagaa ggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagatcattata

	taatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagac aagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcGGCCGCTgatcttcagacc tggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaacca ttaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagaGCAGTGGGAAT aggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaa cagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagata cctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttgcaccactgctgtgccttgg aatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagag aaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatga acaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggta tataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtga atagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacagg cccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgattagtgaacggatC TCGACGGTATCGATCTCGAcacaaatggcagtattcatccacaattttaaaagaaaaggggggat tgggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattaca aaaacaaattacaaaaattcaaaattttcgggtttattacagggacagcagacatgaggacagctaaaac aataagtaatgtaaaatacagcatagcaaaactttaacctccaaatcaagcctctacttgaatccttttctg agggatgaataaggcataggcatcaggggctgttgccaatgtgcattagctgtttgcagcctcaccttctttc atggagtttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcactga cctcccacattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttt ttattaggcagaatccagatgctcaaggcccttcataatatcccccagtttagtagttggacttagggaaca aaggaacctttaatagaaattggacagcaagaaagcgagcttagtgatacttgtgggccagggcattagc cacaccagccaccactttctgataggcagcctgcactggtggggtgaattctttgccaaagtgatgggccag cacacagaccagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaag ggcctagcttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggatt gtagctgctattagcaatatgaaacctcttacatcagttacaatttatatgcagaaataccctgttacttctcc ccttcctatgacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcac cctgaagttctcaggatccacgtgcagcttgtcacagtgcagctcactcagctgggcaaaggtgcccttgag gttgtccaggtgagccaggccatcactaaaggcaccgagcactttcttgccatgagccttcaccttagggtt gcccataacagcatcaggagtggacagatccccaaaggactcaaagaacctctgggtccaagggtagac caccagcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaag agtcttctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgc ccagggcctcaccaccaacggcatccacgttcaccttgtcccacagggcagtaacggcagacttctcctca ggagtcaggtgcaccatggtgtctgtttgaggttgctagtgaacacagttgtgtcagaagcaaatgtaagca atagatggctctgccctgacttttatgcccagccctggctcctgccctccctgctcctgggagtagattggcc aaccctagggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggca ctggcttaggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagta caaattgctactaaaaacatcctcctttgcaagtgtatttaccatcaataattctagccccacaggagtttgtt ctgaaagtaaacttccacaaccgcaagcttattgaggctaaggcatctgtgaaggaaagaaacatctcctc taaaccactatgctgctagagcctcttttctgtactcaagcctcattcagacactagtgtcaccagtctcctca tatacctattgtattttcttcttcttgctggtttagtcatgttttctgggagcttaggggcttattttattttgttttg ttttctaatcaacagagatgggcaaacccattatttttttctttagacttgggatggtgatagctgggcagcgt cagaaactgtgtgtggatatagataagagctcaggactatgctgagctgtgatgagggaggggcctagcta aaggcagtgagagtcagaatgctcctgctattgccttctcagtccccacgcttggtttctacacaagtagata catagaaaaggctataggttagtgtttgagagtcctgcatgattagttgctcagaaatgcccgataaatatg ttatgtgtgtttatgtatatatatgttttatatgtgtgtgtgtgtgtgttgtgtttacaaatatgtgattatcatca aaacgtgagggTACGTaTATGTGTATATATATATATATATTCAGGAAATAATATATTCT

	AGAATATGTCACATTCTGTCTCAGGCATCCATTTTCTTTATGATGCCGTTTGAGGT GGAGTTTTAGTCAGGTGGTCAGCTTCTCCTTTTTTTTGCCATCTGCCCTGTAAGCA TCCTGCTGGGGACCCAGATAGGAGTCATCACTCTAGGCTGAGAACATCTGGGCA CACACCCTAAGCCTCAGCATGACTCATCATGACTCAGCATTGCTGTGCTTGAGCCA GAAGGTTTGCTTAGAAGGTTACACAGAACCAGAAGGCGGGGGTGGGGCACTGA CCCCGACAGGGGCCTGGCCAGAACTGCTCATGCTTGGACTATGGGAGGTCACTA ATGGAGACACACAGAAATGTAACAGGAACTAAGGAAAAACTGAAGCTTtgggggta taggggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcccccacctttcccatgt ctgccctctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgac ctaaccacctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccctgacccc accccaccctccctggaacctctgatagacacatctggcacaccagctcgcaaagtcaccgtgagggtctt gtgtttgctgagtcaaaattccttgaaatccaagtccttagagactcccaggcttggattcaaagctcctgac tttctgtctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaaaaaaaatggaacccaaa atattctacatagtttccatgtcacagccagggctgggcagtctcctgttatttcttttaaaataaatatatca tttaaatgcataaataagcaaaccctgctcgggaatgggagggagagtctctggagtccacccccttctcgg ccctggctctgcagatagtgctatcaaagccctgacagagccctgcccattgctgggccttggagtgagtca gcctagtagaggaggcagggcaagccatctcatagctgctgagtgggagagagaaaagggctcattgtcta taaactcaggtcatggctattcttattaaaagaaaaggggggactggaagggctaattcactcccaacgaa gacaagatctgctttttgcttgtactGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAG CTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGT GCTTCAAGTAGTGTGTTGGTTTTTTGTGTGTGCATTGTCTGAGTAGGTGTCATTCT ATTCTGGGGGGTGGGGTGGGGCAGCACAGCAAGGGGGAGGATTGGGAAGACA ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGagtagtagttcatgtcatcttatta ttcagtatttataacttgcaaagaaatgaatatcagagagtgagaggaacttgtttattgcagcttataatg gttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttg tccaaactcatcaatgtatcttatcatgtctggctctagctatcccgcccctaactccgcccATCCCGCCC CTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTCTCACTACTTCTGGAAT AGCTCAGAGGCCGAGGCGGCCTCGGCCTCTGCATAAATAAAAAAAATTAGTCGA GGCTTTTTTGGAGGCCTAGGgacgtacccaattcgccctatagtgagtcgtattacgcgcgctcac tggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatcc ccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctga atggcgaatggg
pUV+HS1-AS3(coX2) from junction marker (SEQ ID NO:16)	acgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgcc agcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagct ctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagg gtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttcttt aatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggat tttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaata ttaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaata cattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatagcacctagatcaagag acaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtgga gaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggctgtcagc gcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcag cgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaa gggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaag tatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaag cgaaacatcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacga agagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgagcatgcccgacggcgaggat

	ctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtggaaaatggccgcttttctggattcatcg actgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctacccgtgatattgctgaagag cttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcagcgcatcgcct tctatcgccttcttgacgagttcttctgaattattaacgcttacaatttcctgatgcggtattttctccttacgca tctgtgcggtatttcacaccgcatcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttattt ttctaaatacattcaaatatgtatccgctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtc agaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgcaaaca aaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaact ggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaac tctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgt gtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggtc gtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaa agcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggaga gcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactt gagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttta cggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgta ttaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcga ggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggc acgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattag gcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcaca caggaaacagctatgaccatgattacgccaagcgcgcaattaaccctcactaaagggaacaaaagctgg agctgcaagcttggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaaca ttaccgccatgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccc atatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgccc attgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtgga gtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtc aatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtac atctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggt ttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacg ggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggagg tctatataagcagagctcgtttagtgaaccgGGGTCTCTCTGGTTAGACCAGATCTGAGCCT GGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCC TTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGA TCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCagtggcgcccgaacagggact tgaaagcgaaagggaaaccagaggagctctcTCGACGCAGGACTCGGCTTGCTGAAGCGC GCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACT AGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGG GGAGaattagatcgcgatgggaaaaaattcggttaaggccaggggaaagaaaaatataaattaaa acatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacatcagaa ggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagatcattata taatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagac aagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcGGCCGCTgatcttcagacc tggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaacca ttaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaaaaagaGCAGTGGGAAT aggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaa cagcatctgttgcaactcacagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagata cctaaaggatcaacagctcctggggatttggggttgctctggaaaactcatttgcaccactgctgtgccttgg

	aatgctagttggagtaataaatctctggaacagatttggaatcacacgacctggatggagtgggacagag aaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatga acaagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggta tataaaattattcataatgatagtaggaggcttggtaggtttaagaatagtttttgctgtactttctatagtga atagagttaggcagggatattcaccattatcgtttcagacccacctcccaaccccgaggggacccgacagg cccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgattagtgaacggatC TCGACGGTATCGATCTCGAcacaaatggcagtattcatccacaattttaaaagaaaaggggggat tgggggtacagtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattaca aaaacaaattacaaaaattcaaaattttcgggtttattacagggacagcagacatgaggacagctaaaac aataagtaatgtaaaatacagcatagcaaaactttaacctccaaatcaagcctctacttgaatccttttctg agggatgaataaggcataggcatcaggggctgttgccaatgtgcattagctgtttgcagcctcaccttctttc atggagtttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcactga cctcccacattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttt ttattaggcagaatccagatgctcaaggcccttcataatatcccccagtttagtagttggacttagggaaca aaggaacctttaatagaaattggacagcaagaaagcgagcttagtgatacttgtgggccagggcattagc cacaccagccaccactttctgataggcagcctgcactggtggggtgaattctttgccaaagtgatgggccag cacacagaccagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaag ggcctagcttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggatt gtagctgctattagcaatatgaaacctcttacatcagttacaatttatatgcagaaataccctgttacttctcc ccttcctatgacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcac cctGAAGTTCTCaGGaTCCACGTGCAGCTTGTCGCAGTGCAGCTCGCTCAGCTGGG CGAAGGTGCCCTTCAGGTTGTCCAGGTGGGCCAGGCCGTCGCTGAAGGCGCCCA GCACCTTCTTGCCGTGGGCCTTCACCTTGGGGTTGCCCATCACGGCGTCGGGGGT GCTCAGGTCGCCGAAGCTCTCGAAGAAGCGCTGGGTCCAGGGGTACACCACCAG cagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagagtcttc tctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccaggg cctcaccaccaacggcatccacgttcaccttgtcccacagggcagtaacggcagacttctcctcaggagtca ggtgcaccatggtgtctgtttgaggttgctagtgaacacagttgtgtcagaagcaaatgtaagcaatagatg gctctgccctgacttttatgcccagccctggctcctgccctccctgctcctgggagtagattggccaaccctag ggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcactggcttag gagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaattgct actaaaaacatcctcctttgcaagtgtatttaccatcaataattctagccccacaggagtttgttctgaaagt aaacttccacaaccgcaagcttattgaggctaaggcatctgtgaaggaaagaaacatctcctctaaaccac tatgctgctagagcctcttttctgtactcaagcctcattcagacactagtgtcaccagtctcctcatatacctat tgtattttcttcttcttgctggtttagtcatgttttctgggagcttaggggcttattttattttgttttgttttctaat caacagagatgggcaaacccattatttttttctttagacttgggatggtgatagctgggcagcgtcagaaac tgtgtgtggatatagataagagctcaggactatgctgagctgtgatgagggaggggcctagctaaaggca gtgagagtcagaatgctcctgctattgccttctcagtccccacgcttggtttctacacaagtagatacataga aaaggctataggttagtgtttgagagtcctgcatgattagttgctcagaaatgcccgataaatatgttatgtg tgtttatgtatatatatgttttatatgtgtgtgtgtgtgttgtgtttacaaatatgtgattatcatcaaaacgt gagggTACGTaTATGTGTATATATATATATATATTCAGGAAATAATATATTCTAGAA TATGTCACATTCTGTCTCAGGCATCCATTTTCTTTATGATGCCGTTTGAGGTGGAG TTTTAGTCAGGTGGTCAGCTTCTCCTTTTTTTTGCCATCTGCCCTGTAAGCATCCTG CTGGGGACCCAGATAGGAGTCATCACTCTAGGCTGAGAACATCTGGGCACACAC CCTAAGCCTCAGCATGACTCATCATGACTCAGCATTGCTGTGCTTGAGCCAGAAG GTTTGCTTAGAAGGTTACACAGAACCAGAAGGCGGGGGTGGGGCACTGACCCCG ACAGGGGCCTGGCCAGAACTGCTCATGCTTGGACTATGGGAGGTCACTAATGGA GACACACAGAAATGTAACAGGAACTAAGGAAAAACTGAAGCTTtgggggtatagggg

	agcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcccccacctttcccatgtctgccct ctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgacctaacca cctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccctgaccccaccccac cctccctggaacctctgatagacacatctggcacaccagctcgcaaagtcaccgtgagggtcttgtgtttgct gagtcaaaattccttgaaatccaagtccttagagactcccaggcttggattcaaagctcctgactttctgtct agtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaaaaaaaatggaacccaaaatattcta catagtttccatgtcacagccagggctgggcagtctcctgttatttcttttaaaataaatatatcatttaaatg cataaataagcaaaccctgctcgggaatgggagggagagtctctggagtccacccctttctcggccctggct ctgcagatagtgctatcaaagccctgacagagccctgcccattgctgggccttggagtgagtcagcctagta gagaggcagggcaagccatctcatagctgctgagtgggagagagaaaagggctcattgtctataaactca ggtcatggctattcttattaaaagaaaaggggggactggaagggctaattcactcccaacgaagacaaga tctgctttttgcttgtactGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCT GGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTC AAGTAGTGTGTTGGTTTTTTGTGTGTGCATTGTCTGAGTAGGTGTCATTCTATTCT GGGGGGTGGGGTGGGGCAGCACAGCAAGGGGGAGGATTGGGAAGACAATAGC AGGCATGCTGGGGATGCGGTGGGCTCTATGGagtagtagttcatgtcatcttattattcagta tttataacttgcaaagaaatgaatatcagagagtgagaggaacttgtttattgcagcttataatggttacaa ataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaa ctcatcaatgtatcttatcatgtctggctctagctatcccgcccctaactccgcccATCCCGCCCCTAAC TCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTCTCACTACTTCTGGAATAGCTC AGAGGCCGAGGCGGCCTCGGCCTCTGCATAAATAAAAAAAATTAGTCGAGGCTT TTTTGGAGGCCTAGGgacgtacccaattcgccctatagtgagtcgtattacgcgcgctcactggccgt cgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatcccccttttc gccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcg aatggg

[0106] In certain embodiments the human β -globin gene in the vectors contemplated herein comprises an anti-sickling human β -globin gene encoding an anti-sickling β -globin polypeptide. In certain embodiments the anti-sickling version of a human β -globin gene used in the vector comprises one, two, or three mutations selected from the group consisting of Gly16Asp, Glu22Ala and Thr87Gln (*see, e.g.,* Levasseur (2004) *J. Biol. Chem.* 279(26): 27518-27524). Without being bound to a particular theory, it is believed the Glu22Ala mutation increases affinity to α-chain, the Thr87Gln mutation blocks lateral contact with Val6 of βS protein, and the Gly16Asp mutation decreases axial contact between globin chains.

[0107] However, we also tested the usefulness of codon optimization on BASB-globin transgene expression. The BAS3-globin transgene was codon optimized for expression using the J-CAT software (Grote *et al.* (2005) *Nucleic Acids. Res.,* 33: 526-531) and care was taken to retain the three anti-sickling mutations (Gly16Asp, Glu22Ala, and Thr87Gln.). The optimized BASB-globin exons had 84% sequence homology at the nucleotide level when compared to the original unmodified sequences. Three derivatives of Core-ASB-FB were

made that contained, 1). Codon-optimized exon one with unmodified exons two and three, 2) codon-optimized exons one and two with unmodified exon three, and 3), codon-optimized exons one, two, and three with no unmodified exons (*see, e.g.*, Figure 2, panel A). To determine if codon optimization could enhance expression, HSPCs derived from multiple healthy donors were transduced with Core-AS3-FB derivatives at equal MOI and subjected to in vitro erythroid differentiation for 14 days. Enhancement of expression by about 1.6-fold was observed in derivatives that contained a codon optimized exon two ($p<0.05$) when compared to the parental construct (*see, e.g.*, Figure 2, panel B). The additional optimization of exons one or three did not result in increased expression per vector genome when compared to the appropriate control (*see, e.g.*, Figure 2, panel B). Taken together, these data provide strategies to increase expression per vector genome.

[0108] Accordingly in various embodiments the heterologous gene in the vector described above comprises an anti-sickling human beta globin gene comprises a modified βAS3-globin transgene, said transgene comprising a codon optimized exon 1 (SEQ ID NO:7) (as shown within SEQ ID NO:10), and/or a codon optimized exon 2 (SEQ ID NO:8) (as shown within SEQ ID NO:11), and/or comprising a codon optimized exon 3 (SEQ ID NO:9) (as shown within SEQ ID NO:12). In certain embodiments the βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO:8). In certain embodiments the βAS3-globin transgene comprises a codon optimized exon 1 (SEQ ID NO:7). In certain embodiments the βAS3-globin transgene comprises a codon optimized exon 3 (SEQ ID NO:9). See Table 1 for referenced sequences.

[0109] In certain embodiments the vectors described herein comprise a human Ankyrin insulator element

[0110] In certain embodiments the vectors described herein comprise a murine GATA1.

[0111] In various embodiments, the LVs described herein can have additional safety features that can include, for example, the presence of an insulator (*e.g.*, an FB insulator in the 3'LTR). Additionally, or alternatively, in certain embodiments, the HIV LTR has been substituted with an alternative promoter (*e.g.*, a CMV) to yield a higher titer vector without the inclusion of the HIV TAT protein during packaging. Other strong promoters (*e.g.*, RSV, and the like can also be used).

[0112] As shown above, the vectors described herein are effective to transduce cells at high titer and to also provide high levels of expression.

[0113] In view of these results, it is believed that LVs described herein, *e.g.*, recombinant TAT-independent, SIN LVs that express a human beta-globin gene can be used to effectively treat hemoglobinopathies in subjects (*e.g.*, human and non-human mammals). Such hemoglobinopathies include, but are not limited to sickle cell disease (SCD) and β-thalassemia. It is believed these vectors can be used for the modification of stem cells (*e.g.*, hematopoietic stem and progenitor cells) that can be introduced into a subject in need thereof for the treatment of, *e.g.*, SCD or β-thalassemia. Moreover, it appears that the resulting cells will produce enough of the transgenic β-globin protein to demonstrate significant improvement in subject health. It is also believed the vectors can be directly administered to a subject to achieve *in vivo* transduction of the target (*e.g.*, hematopoietic stem or progenitor cells) and thereby also effect a treatment of subjects in need thereof.

[0114] As noted above, in various embodiments the LVs described herein can comprise various safety features. For example, the HIV LTR has been substituted with a CMV promoter to yield higher titer vector without the inclusion of the HIV TAT protein during packaging. In certain embodiments an insulator (*e.g.*, the FB insulator) is introduced into the 3'LTR for safety. The LVs are also constructed to provide efficient transduction and high titer.

[0115] It will be appreciated that the foregoing elements are illustrative and need not be limiting. In view of the teachings provided herein, suitable substitutions for these elements will be recognized by one of skill in the art and are contemplated within the scope of the teachings provided herein.

Anti-sickling β-globin gene and expression cassette.

[0116] As indicated above, in various embodiments the LV described herein comprise an expression cassette encoding a wild-type β-globin gene, or an anti-sickling human β-globin gene. On illustrative, but non-limiting cassette is βAS3 which comprises an ~2.3 kb recombinant human β-globin gene (exons and introns) with three amino acid substitutions (Thr87Gln; Gly16Asp; and Glu22Ala) under the control of transcriptional control elements (*e.g.*, the human β-globin gene 5' promoter (*e.g.*, ~266 bp), the human β-globin 3' enhancer (*e.g.*, ~260 bp), β-globin intron 2 with a ~375 bp RsaI deletion from IVS2, and a ~3.4 kb composite human β-globin locus control region (*e.g.*, HS2 ~1203 bp; HS3 ~1213 bp; HS4 ~954 bp). One embodiment of a βAS3 cassette is described by Levasseur (2003) *Blood* 102: 4312-4319.

[0117] In certain embodiments the β-globin gene comprises a *Ssp*I (S) to *Rsa*I (R) deletion (~220bp), *e.g.*, as described by Antoniou *et al.* 1998) *Nucl. Acids Res.*, 26(3): 721-729.

[0118] The βAS3 cassette, however, is illustrative and need not be limiting. Using the teaching provided herein, numerous variations will be available to one of skill in the art. Such variations include, for example, use of a gene encoding a wild-type β-globin, use of a gene comprising one or two mutations selected from the group consisting of Thr87Gln, Gly16Asp, and Glu22Ala, and/or further or alternative mutations to the β-globin to further enhance non-sickling properties, alterations in the transcriptional control elements (*e.g.*, promoter and/or enhancer), variations on the intron size/structure, and the like.

[0119] As noted above, in certain embodiments, the βAS3 cassette comprises a codon optimized exon 1, and/or a codon optimized exon 2, and/or a codon optimized exon 3. In certain embodiments only exon 2 is codon optimized.

TAT-Independent and Self inactivating lentiviral vectors.

[0120] To further improve safety, in various embodiments, the lentiviral vectors described herein comprise a TAT-independent, self-inactivating (SIN) configuration. Thus, in various embodiments it is desirable to employ in the LVs described herein an LTR region that has reduced promoter activity relative to wild-type LTR. Such constructs can be provided that are effectively "self-inactivating" (SIN) which provides a biosafety feature. SIN vectors are ones in which the production of full-length vector RNA in transduced cells is greatly reduced or abolished altogether. This feature minimizes the risk that replication-competent recombinants (RCRs) will emerge. Furthermore, it reduces the risk that that cellular coding sequences located adjacent to the vector integration site will be aberrantly expressed.

[0121] Furthermore, a SIN design reduces the possibility of interference between the LTR and the promoter that is driving the expression of the transgene. SIN LVs can often permit full activity of the internal promoter.

[0122] The SIN design increases the biosafety of the LVs. The majority of the HIV LTR is comprised of the U3 sequences. The U3 region contains the enhancer and promoter elements that modulate basal and induced expression of the HIV genome in infected cells and in response to cell activation. Several of these promoter elements are essential for viral replication. Some of the enhancer elements are highly conserved among viral isolates and

have been implicated as critical virulence factors in viral pathogenesis. The enhancer elements may act to influence replication rates in the different cellular target of the virus

[0123] As viral transcription starts at the 3' end of the U3 region of the 5' LTR, those sequences are not part of the viral mRNA and a copy thereof from the 3' LTR acts as template for the generation of both LTR's in the integrated provirus. If the 3' copy of the U3 region is altered in a retroviral vector construct, the vector RNA is still produced from the intact 5' LTR in producer cells, but cannot be regenerated in target cells. Transduction of such a vector results in the inactivation of both LTR's in the progeny virus. Thus, the retrovirus is self-inactivating (SIN) and those vectors are known as SIN transfer vectors.

[0124] In certain embodiments self-inactivation is achieved through the introduction of a deletion in the U3 region of the 3' LTR of the vector DNA, *i.e.*, the DNA used to produce the vector RNA. During RT, this deletion is transferred to the 5' LTR of the proviral DNA. Typically, it is desirable to eliminate enough of the U3 sequence to greatly diminish or abolish altogether the transcriptional activity of the LTR, thereby greatly diminishing or abolishing the production of full-length vector RNA in transduced cells. However, it is generally desirable to retain those elements of the LTR that are involved in polyadenylation of the viral RNA, a function typically spread out over U3, R and U5. Accordingly, in certain embodiments, it is desirable to eliminate as many of the transcriptionally important motifs from the LTR as possible while sparing the polyadenylation determinants.

[0125] The SIN design is described in detail in Zufferey *et al.* (1998) *J Virol.* 72(12): 9873-9880, and in U.S. Patent No: 5,994,136. As described therein, there are, however, limits to the extent of the deletion at the 3' LTR. First, the 5' end of the U3 region serves another essential function in vector transfer, being required for integration (terminal dinucleotide+att sequence). Thus, the terminal dinucleotide and the att sequence may represent the 5' boundary of the U3 sequences which can be deleted. In addition, some loosely defined regions may influence the activity of the downstream polyadenylation site in the R region. Excessive deletion of U3 sequence from the 3'LTR may decrease polyadenylation of vector transcripts with adverse consequences both on the titer of the vector in producer cells and the transgene expression in target cells.

[0126] Additional SIN designs are described in U.S. Patent Publication No: 2003/0039636. As described therein, in certain embodiments, the lentiviral sequences removed from the LTRs are replaced with comparable sequences from a non-lentiviral retrovirus, thereby forming hybrid LTRs. In particular, the lentiviral R region within the

LTR can be replaced in whole or in part by the R region from a non-lentiviral retrovirus. In certain embodiments, the lentiviral TAR sequence, a sequence which interacts with TAT protein to enhance viral replication, is removed, preferably in whole, from the R region. The TAR sequence is then replaced with a comparable portion of the R region from a non-lentiviral retrovirus, thereby forming a hybrid R region. The LTRs can be further modified to remove and/or replace with non-lentiviral sequences all or a portion of the lentiviral U3 and U5 regions.

[0127] Accordingly, in certain embodiments, the SIN configuration provides a retroviral LTR comprising a hybrid lentiviral R region that lacks all or a portion of its TAR sequence, thereby eliminating any possible activation by TAT, wherein the TAR sequence or portion thereof is replaced by a comparable portion of the R region from a non-lentiviral retrovirus, thereby forming a hybrid R region. In a particular embodiment, the retroviral LTR comprises a hybrid R region, wherein the hybrid R region comprises a portion of the HIV R region (*e.g.*, a portion comprising or consisting of the nucleotide sequence shown in SEQ ID NO: 10 in US 2003/0039636) lacking the TAR sequence, and a portion of the MoMSV R region (*e.g.*, a portion comprising or consisting of the nucleotide sequence shown in SEQ ID NO: 9 in 2003/0039636) comparable to the TAR sequence lacking from the HIV R region. In another particular embodiment, the entire hybrid R region comprises or consists of the nucleotide sequence shown in SEQ ID NO: 11 in 2003/0039636.

[0128] Suitable lentiviruses from which the R region can be derived include, for example, HIV (HIV-1 and HIV-2), EIV, SIV and FIV. Suitable retroviruses from which non-lentiviral sequences can be derived include, for example, MoMSV, MoMLV, Friend, MSCV, RSV and Spumaviruses. In one illustrative embodiment, the lentivirus is HIV and the non-lentiviral retrovirus is MoMSV.

[0129] In another embodiment described in US 2003/0039636, the LTR comprising a hybrid R region is a left (5') LTR and further comprises a promoter sequence upstream from the hybrid R region. Preferred promoters are non-lentiviral in origin and include, for example, the U3 region from a non-lentiviral retrovirus (*e.g.*, the MoMSV U3 region). In one particular embodiment, the U3 region comprises the nucleotide sequence shown in SEQ ID NO: 12 in US 2003/0039636. In another embodiment, the left (5') LTR further comprises a lentiviral U5 region downstream from the hybrid R region. In one embodiment, the U5 region is the HIV U5 region including the HIV att site necessary for genomic integration. In another embodiment, the U5 region comprises the nucleotide sequence shown in SEQ ID

NO: 13 in US 2003/0039636. In yet another embodiment, the entire left (5') hybrid LTR comprises the nucleotide sequence shown in SEQ ID NO: 1 in US 2003/0039636.

[0130] In another illustrative embodiment, the LTR comprising a hybrid R region is a right (3') LTR and further comprises a modified (*e.g.*, truncated) lentiviral U3 region upstream from the hybrid R region. The modified lentiviral U3 region can include the att sequence, but lack any sequences having promoter activity, thereby causing the vector to be SIN in that viral transcription cannot go beyond the first round of replication following chromosomal integration. In a particular embodiment, the modified lentiviral U3 region upstream from the hybrid R region consists of the 3' end of a lentiviral (*e.g.*, HIV) U3 region up to and including the lentiviral U3 att site. In one embodiment, the U3 region comprises the nucleotide sequence shown in SEQ ID NO: 15 in US 2003/0039636. In another embodiment, the right (3') LTR further comprises a polyadenylation sequence downstream from the hybrid R region. In another embodiment, the polyadenylation sequence comprises the nucleotide sequence shown in SEQ ID NO: 16 in US 2003/0039636. In yet another embodiment, the entire right (5') LTR comprises the nucleotide sequence shown in SEQ ID NO: 2 or 17 of US 2003/0039636.

[0131] Thus, in the case of HIV based LV, it has been discovered that such vectors tolerate significant U3 deletions, including the removal of the LTR TATA box (*e.g.*, deletions from -418 to -18), without significant reductions in vector titers. These deletions render the LTR region substantially transcriptionally inactive in that the transcriptional ability of the LTR in reduced to about 90% or lower.

[0132] It has also been demonstrated that the *trans*-acting function of Tat becomes dispensable if part of the upstream LTR in the transfer vector construct is replaced by constitutively active promoter sequences (*see, e.g.,* Dull *et al.* (1998) *J Virol.* 72(11): 8463–8471. Furthermore, we show that the expression of *rev* in *trans* allows the production of high-titer HIV-derived vector stocks from a packaging construct which contains only *gag* and *pol*. This design makes the expression of the packaging functions conditional on complementation available only in producer cells. The resulting gene delivery system, conserves only three of the nine genes of HIV-1 and relies on four separate transcriptional units for the production of transducing particles.

[0133] In one embodiments illustrated in Example 1, the cassette expressing an anti-sickling β-globin (*e.g.*, βAS3) is placed in the pCCL LV backbone, which is a SIN vector with the CMV enhancer/promoter substituted in the 5' LTR.

[0134] It will be recognized that the CMV promoter typically provides a high level of non-tissue specific expression. Other promoters with similar constitutive activity include, but are not limited to the RSV promoter, and the SV40 promoter. Mammalian promoters such as the beta-actin promoter, ubiquitin C promoter, elongation factor 1αpromoter, tubulin promoter, *etc.*, may also be used.

[0135] The foregoing SIN configurations are illustrative and non-limiting. Numerous SIN configurations are known to those of skill in the art. As indicated above, in certain embodiments, the LTR transcription is reduced by about 95% to about 99%. In certain embodiments LTR may be rendered at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95% at least about 96%, at least about 97%, at least about 98%, or at least about 99% transcriptionally inactive.

Insulator element

[0136] In certain embodiments, to further enhance biosafety, insulators are inserted into the lentiviral vectors described herein. Insulators are DNA sequence elements present throughout the genome. They bind proteins that modify chromatin and alter regional gene expression. The placement of insulators in the vectors described herein offer various potential benefits including, *inter alia*: 1) Shielding of the vector from positional effect variegation of expression by flanking chromosomes (*i.e.,* barrier activity); and 2) Shielding flanking chromosomes from insertional *trans*-activation of gene expression by the vector (enhancer blocking). Thus, insulators can help to preserve the independent function of genes or transcription units embedded in a genome or genetic context in which their expression may otherwise be influenced by regulatory signals within the genome or genetic context (see, *e.g.*, Burgess-Beusse *et al.* (2002) *Proc. Natl. Acad. Sci. USA,* 99: 16433; and Zhan *et al.* (2001) *Hum. Genet.*, 109: 471). In the present context insulators may contribute to protecting lentivirus-expressed sequences from integration site effects, which may be mediated by *cis*-acting elements present in genomic DNA and lead to deregulated expression of transferred sequences. In various embodiments LVs are provided in which an insulator sequence is inserted into one or both LTRs or elsewhere in the region of the vector that integrates into the cellular genome.

[0137] The first and best characterized vertebrate chromatin insulator is located within the chicken β-globin locus control region. This element, which contains a DNase-I hypersensitive site-4 (cHS4), appears to constitute the 5′ boundary of the chicken β-globin locus (Prioleau *et al.* (1999) *EMBO J.* 18: 4035-4048). A 1.2-kb fragment containing the

cHS4 element displays classic insulator activities, including the ability to block the interaction of globin gene promoters and enhancers in cell lines (Chung *et al.* (1993) *Cell,* 74: 505-514), and the ability to protect expression cassettes in *Drosophila* (*Id.*), transformed cell lines (Pikaart *et al.* (1998) *Genes Dev.* 12: 2852-2862), and transgenic mammals (Wang *et al.* (1997) *Nat. Biotechnol.,* 15: 239-243; Taboit-Dameron *et al.* (1999) *Transgenic Res.,* 8: 223-235) from position effects. Much of this activity is contained in a 250-bp fragment. Within this stretch is a 49-bp cHS4 core (Chung *et al.* (1997) *Proc. Natl. Acad. Sci., USA,* 94: 575-580) that interacts with the zinc finger DNA binding protein CTCF implicated in enhancer-blocking assays (Bell *et al.* (1999) *Cell,* 98: 387-396).

[0138] One illustrative and suitable insulator is FB (FII/BEAD-A), a 77 bp insulator element, that contains the minimal CTCF binding site enhancer-blocking components of the chicken β-globin 5' HS4 insulators and a homologous region from the human T-cell receptor alpha/delta blocking element alpha/delta I (BEAD-I) insulator described by Ramezani *et al.* (2008) *Stem Cell* 26: 3257-3266. The FB "synthetic" insulator has full enhancer blocking activity. This insulator is illustrative and non-limiting. Other suitable insulators may be used including, for example, the full-length chicken beta-globin HS4 or insulator sub-fragments thereof, the ankyrin gene insulator, and other synthetic insulator elements.

Packaging signal.

[0139] In various embodiments the vectors described herein further comprise a packaging signal. A "packaging signal," "packaging sequence," or "psi sequence" is any nucleic acid sequence sufficient to direct packaging of a nucleic acid whose sequence comprises the packaging signal into a retroviral particle. The term includes naturally occurring packaging sequences and also engineered variants thereof. Packaging signals of a number of different retroviruses, including lentiviruses, are known in the art.

Rev Responsive Element (RRE).

[0140] In certain embodiments the lentiviral vectors described herein comprise a Rev response element (RRE) to enhance nuclear export of unspliced RNA. RREs are well known to those of skill in the art. Illustrative RREs include, but are not limited to RREs such as that located at positions 7622-8459 in the HIV NL4-3 genome (Genbank accession number AF003887) as well as RREs from other strains of HIV or other retroviruses. Such sequences are readily available from Genbank or from the database with URL hiv-web.lanl.gov/content/index.

Central PolyPurine Tract (cPPT).

[0141] In various embodiments the lentiviral vectors described herein further include a central polypurine tract. Insertion of a fragment containing the central polypurine tract (cPPT) in lentiviral (*e.g.*, HIV-1) vector constructs is known to enhance transduction efficiency drastically, reportedly by facilitating the nuclear import of viral cDNA through a central DNA flap.

Expression-Stimulating Posttranscriptional Regulatory Element (PRE)

[0142] In certain embodiments the lentiviral vectors (LVs) described herein may comprise any of a variety of posttranscriptional regulatory elements (PREs) whose presence within a transcript increases expression of the heterologous nucleic acid (*e.g.*, βAS3) at the protein level. PREs may be particularly useful in certain embodiments, especially those that involve lentiviral constructs with modest promoters.

[0143] One type of PRE is an intron positioned within the expression cassette, which can stimulate gene expression. However, introns can be spliced out during the life cycle events of a lentivirus. Hence, if introns are used as PRE's they are typically placed in an opposite orientation to the vector genomic transcript.

[0144] Posttranscriptional regulatory elements that do not rely on splicing events offer the advantage of not being removed during the viral life cycle. Some examples are the posttranscriptional processing element of herpes simplex virus, the posttranscriptional regulatory element of the hepatitis B virus (HPRE) and the woodchuck hepatitis virus (WPRE). Of these the WPRE is typically preferred as it contains an additional cis-acting element not found in the HPRE. This regulatory element is typically positioned within the vector so as to be included in the RNA transcript of the transgene, but outside of stop codon of the transgene translational unit.

[0145] The WPRE is characterized and described in U.S. Pat. No: 6,136,597. As described therein, the WPRE is an RNA export element that mediates efficient transport of RNA from the nucleus to the cytoplasm. It enhances the expression of transgenes by insertion of a *cis*-acting nucleic acid sequence, such that the element and the transgene are contained within a single transcript. Presence of the WPRE in the sense orientation was shown to increase transgene expression by up to 7- to 10-fold. Retroviral vectors transfer sequences in the form of cDNAs instead of complete intron-containing genes as introns are generally spliced out during the sequence of events leading to the formation of the retroviral particle. Introns mediate the interaction of primary transcripts with the splicing machinery.

Because the processing of RNAs by the splicing machinery facilitates their cytoplasmic export, due to a coupling between the splicing and transport machineries, cDNAs are often inefficiently expressed. Thus, the inclusion of the WPRE in a vector results in enhanced expression of transgenes.

Illustrative, but non-limiting embodiments.

[0146] Figures 2-6 schematically illustrate various features certain illustrative embodiments of the vectors described herein. It is noted that in various embgodiments the IVS2 sequence comprises a deletion. As shown in Table 2 the underlined region shows the entire deletion, bold and underlined shows the additional 371 bp deletion that was added in the pUV-AS3 (Figure 3), pUV-AS3(coX2) (Figure 4), pUV-HS1-AS3 (Figure 5), and pUV+HS1-AS3(coX2) (Figure 6). There was a smaller deletion in our older vectors and now there is an additional larger 371 bp deletion in our new vectors (591 bp total), the additional deleted region is the one in bold and underlined.

Table 2. Illustration of IVS2 deletion.

IVS2 Deletion
CTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGCAAAAGGGCCTAGCTT GGACTCAGAATAATCCAGCCTTATCCCAACCATAAAATAAAAGCAGAATGGTA GCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACATCAGTTACAATTT ATATGCAGAAATATTTATATGCAGAGATATTGCTATTGCCTTAACCCAGAAATT ATCACTGTTATTCTTTAGAATGGTGCAAAGAGGCATGATACATTGTATCATTAT TGCCCTGAAAGAAAGAGATTAGGGAAAGTATTAGAAATAAGATAAACAAAAA AGTATATTAAAAGAAGAAAGCATTTTTTAAAATTACAAATGCAAAATTACCCT GATTTGGTCAATATGTGTAC**ACATATTAAAACATTACACTTTAACCCATAAA** **TATGTATAATGATTATGTATCAATTAAAAATAAAAGAAAATAAAGTAGGGA** **GATTATGAATATGCAAATAAGCACACATATATTCCAAATAGTAATGTACTA** **GGCAGACTGTGTAAAGTTTTTTTTTAAGTTACTTAATGTATCTCAGAGATA** **TTTCCTTTTGTTATACACAATGTTAAGGCATTAAGTATAATAGTAAAAATT** **GCGGAGAAGAAAAAAAAAGAAAGCAAGAATTAAACAAAAGAAAACAATTG** **TTATGAACAGCAAATAAAAGAAACTAAAACGATCCTGAGACTTCCACACTG** **ATGCAATCATTCGTCTGTTTCCCATTCTAAACTG**TACCCTGTTACTTATCCCC TTCCTATGACATGAACTTAACCATAGAAAAGAAGGGGAAAGAAAACATCAAG CGTCCCATAGACTCAC

[0147] In certain embodiments the vectors contemplated herein comprises the features of pUV-AS3 (Figure 3), pUV-AS3(coX2) (Figure 4), pUV-HS1-AS3 (Figure 5), or pUV+HS1-AS3(coX2) (Figure 6).

[0148] In certain embodiments the vectors contemplated herein comprise the nucleotide sequence of pUV-AS3 (SEQ ID NO: 13), pUV-AS3(coX2) (SEQ ID NO:14), pUV-HS1-AS3 (SEQ ID NO:15), and pUV+HS1-AS3(coX2) (SEQ ID NO: 16).

Improved titer, gene transfer, and increased total expression.

[0149] As shown in Figure 7, unconcentrated vector titer of the pUV-HS1-AS3 (Mini-G) is ~10 fold higher than the parental vector. Titer was determined by the number of infectious particles, transducing units (TU), per ml on HT-29 cells.

[0150] Gene transfer was significantly enhanced for the shorter globin vector Mini-G (pUV-HS1-AS3) in primary human hematopoietic stem and progenitor ($CD34^+$) cells (HSPCs). Figure 8, panel A demonstrates how gene transfer continues to increase with Mini-G, compared to the parental vector that displays little increase in gene transfer. Figure 8, panel B focusses on a single vector concentration ($2x10^7$ TU/ml) that is typically used for clinical vector transductions, here the gene transfer is significantly improved and is ~10 fold higher than the parental vector.

[0151] Additionally, Mini-G provided increased total expression. Figure 9 shows that expression of AS3-globin is ~5-fold higher in erythrocytes differentiated *in vitro* from transduced HSPCs. Globin expression was assayed at the RNA level using ddPCR probe/primers. Without being bound by a particular theory, it is believed the difference in expression is caused by the enhanced gene transfer of Mini-G.

Transduced Host Cells and Methods of cell transduction.

[0152] The recombinant lentiviral vectors (LV) and resulting virus described herein are capable of transferring a heterologous nucleic acid (*e.g.*, a nucleic acid encoding an anti-sickling β-globin) sequence into a mammalian cell. In various embodiments, for delivery to cells, vectors described herein are preferably used in conjunction with a suitable packaging cell line or co-transfected into cells *in vitro* along with other vector plasmids containing the necessary retroviral genes (*e.g.*, gag and pol) to form replication incompetent virions capable of packaging the vectors of the present invention and infecting cells.

[0153] The recombinant LVs and resulting virus described herein are capable of transferring a nucleic acid (*e.g.*, a nucleic acid encoding an anti-sickling β-globin or other

sequence) into a mammalian cell. For delivery to cells, various vectors described herein are preferably used in conjunction with a suitable packaging cell line or co-transfected into cells *in vitro* along with other vector plasmids containing the necessary retroviral genes (*e.g.*, gag and pol) to form replication incompetent virions capable of packaging the vectors of the present invention and infecting cells.

[0154] In certain embodiments the vectors are introduced via transfection into the packaging cell line. The packaging cell line produces viral particles that contain the vector genome. Methods for transfection are well known by those of skill in the art. After cotransfection of the packaging vectors and the transfer vector to the packaging cell line, the recombinant virus is recovered from the culture media and titered by standard methods used by those of skill in the art. Thus, the packaging constructs can be introduced into human cell lines by calcium phosphate transfection, lipofection or electroporation, generally together with or without a dominant selectable marker, such as neomycin, DHFR, Glutamine synthetase, followed by selection in the presence of the appropriate drug and isolation of clones. In certain embodiments the selectable marker gene can be linked physically to the packaging genes in the construct.

[0155] Stable cell lines wherein the packaging functions are configured to be expressed by a suitable packaging cell are known (*see, e.g.,* U.S. Patent No. 5,686,279, which describes packaging cells). In general, for the production of virus particles, one may employ any cell that is compatible with the expression of lentiviral Gag and Pol genes, or any cell that can be engineered to support such expression. For example, producer cells such as 293T cells and HT1080 cells may be used.

[0156] The packaging cells with a lentiviral vector incorporated therein form producer cells. Producer cells are thus cells or cell-lines that can produce or release packaged infectious viral particles carrying the therapeutic gene of interest (*e.g.*, modified β-globin). These cells can further be anchorage dependent which means that these cells will grow, survive, or maintain function optimally when attached to a surface such as glass or plastic. Some examples of anchorage dependent cell lines used as lentiviral vector packaging cell lines when the vector is replication competent are HeLa or 293 cells and PERC.6 cells.

[0157] Accordingly, in certain embodiments, methods are provided of delivering a gene to a cell which is then integrated into the genome of the cell, comprising contacting the cell with a virion containing a lentiviral vector described herein. The cell (*e.g.*, in the form of tissue or an organ) can be contacted (*e.g.*, infected) with the virion ex vivo and then delivered

to a subject (*e.g.*, a mammal, animal or human) in which the gene (*e.g.*, anti-sickling β-globin) will be expressed. In various embodiments the cell can be autologous to the subject (*i.e.*, from the subject) or it can be non-autologous (*i.e.*, allogeneic or xenogenic) to the subject. Moreover, because the vectors described herein are capable of being delivered to both dividing and non-dividing cells, the cells can be from a wide variety including, for example, bone marrow cells, mesenchymal stem cells (*e.g.*, obtained from adipose tissue), and other primary cells derived from human and animal sources. Alternatively, the virion can be directly administered *in vivo* to a subject or a localized area of a subject (*e.g.*, bone marrow).

[0158] Of course, as noted above, the lentivectors described herein will be particularly useful in the transduction of human hematopoietic progenitor cells or a hematopoietic stem cells, obtained either from the bone marrow, the peripheral blood or the umbilical cord blood, as well as in the transduction of a $CD4^+$ T cell, a peripheral blood B or T lymphocyte cell, and the like. In certain embodiments particularly preferred targets are $CD34^+$ hematopoetic stem and progenitor cells.

Gene therapy.

[0159] In still other embodiments, methods are provide for transducing a human hematopoietic stem cell. In certain embodiments the methods involve contacting a population of human cells that include hematopoietic stem cells with one of the foregoing lentivectors under conditions to effect the transduction of a human hematopoietic progenitor cell in said population by the vector. The stem cells may be transduced *in vivo* or *in vitro*, depending on the ultimate application. Even in the context of human gene therapy, such as gene therapy of human stem cells, one may transduce the stem cell *in vivo* or, alternatively, transduce *in vitro* followed by infusion of the transduced stem cell into a human subject. In one aspect of this embodiment, the human stem cell can be removed from a human, *e.g.*, a human patient, using methods well known to those of skill in the art and transduced as noted above. The transduced stem cells are then reintroduced into the same or a different human.

Stem cell/progenitor cell gene therapy.

[0160] In various embodiments the lentivectors described herein are particularly useful for the transduction of human hematopoietic progenitor cells or haematopoietic stem cells (HSCs), obtained either from the bone marrow, the peripheral blood or the umbilical cord blood, as well as in the transduction of a $CD4^+$ T cell, a peripheral blood B or T

lymphocyte cell, and the like. In certain embodiments particularly preferred targets are $CD34^+$ hematopoietic stem and progenitor cells.

[0161] When cells, for instance $CD34^+$ cells, dendritic cells, peripheral blood cells or tumor cells are transduced *ex vivo*, the vector particles are incubated with the cells using a dose generally in the order of between 1 to 50 multiplicities of infection (MOI) which also corresponds to 1×10^5 to 50×10^5 transducing units of the viral vector per 10^5 cells. This can include amounts of vector corresponding to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, and 50 MOI. Typically, the amount of vector may be expressed in terms of HT-29 transducing units (TU).

[0162] In certain embodiments cell-based therapies involve providing stem cells and/or hematopoietic precursors, transduce the cells with the lentivirus encoding, *e.g.*, an anti-sickling human β-globin, and then introduce the transformed cells into a subject in need thereof (*e.g.*, a subject with the sickle cell mutation).

[0163] In certain embodiments the methods involve isolating population of cells, *e.g.*, stem cells from a subject, optionally expand the cells in tissue culture, and administer the lentiviral vector whose presence within a cell results in production of an anti-sickling β-globin in the cells *in vitro*. The cells are then returned to the subject, where, for example, they may provide a population of red blood cells that produce the anti-sickling β globin.

[0164] In some illustrative, but non-limiting, embodiments, a population of cells, which may be cells from a cell line or from an individual other than the subject, can be used. Methods of isolating stem cells, immune system cells, *etc.*, from a subject and returning them to the subject are well known in the art. Such methods are used, *e.g.*, for bone marrow transplant, peripheral blood stem cell transplant, *etc.*, in patients undergoing chemotherapy.

[0165] Where stem cells are to be used, it will be recognized that such cells can be derived from a number of sources including bone marrow (BM), cord blood (CB), mobilized peripheral blood stem cells (mPBSC), and the like. In certain embodiments the use of induced pluripotent stem cells (IPSCs) is contemplated. Methods of isolating hematopoietic stem cells (HSCs), transducing such cells and introducing them into a mammalian subject are well known to those of skill in the art.

[0166] In certain embodiments a lentiviral vector described herein (*see, e.g.,* Figure **4**) is used in stem cell gene therapy for SCD by introducing the βAS3 anti-sickling β-globin gene into the bone marrow stem cells of patients with sickle cell disease followed by autologous transplantation.

Direct introduction of vector.

[0167] In certain embodiments direct treatment of a subject by direct introduction of the vector(s) described herein is contemplated. The lentiviral compositions may be formulated for delivery by any available route including, but not limited to parenteral (*e.g.*, intravenous), intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, rectal, and vaginal. Commonly used routes of delivery include inhalation, parenteral, and transmucosal.

[0168] In various embodiments pharmaceutical compositions can include an LV in combination with a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[0169] In some embodiments, active agents, *i.e.*, a lentiviral described herein and/or other agents to be administered together the vector, are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such compositions will be apparent to those skilled in the art. Suitable materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomes can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811. In some embodiments the composition is targeted to particular cell types or to cells that are infected by a virus. For example, compositions can be targeted using monoclonal antibodies to cell surface markers, *e.g.*, endogenous markers or viral antigens expressed on the surface of infected cells.

[0170] It is advantageous to formulate compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit comprising a predetermined quantity of a LV calculated to produce the desired therapeutic effect in association with a pharmaceutical carrier.

[0171] A unit dose need not be administered as a single injection but may comprise continuous infusion over a set period of time. Unit dose of the LV described herein may conveniently be described in terms of transducing units (T.U.) of lentivector, as defined by titering the vector on a cell line such as HeLa or 293. In certain embodiments unit doses can range from 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{13} T.U. and higher.

[0172] Pharmaceutical compositions can be administered at various intervals and over different periods of time as required, *e.g.*, one time per week for between about 1 to about 10 weeks; between about 2 to about 8 weeks; between about 3 to about 7 weeks; about 4 weeks; about 5 weeks; about 6 weeks, *etc.* It may be necessary to administer the therapeutic composition on an indefinite basis. The skilled artisan will appreciate that certain factors can influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Treatment of a subject with a LV can include a single treatment or, in many cases, can include a series of treatments.

[0173] Illustrative, but non-limiting, doses for administration of gene therapy vectors and methods for determining suitable doses are known in the art. It is furthermore understood that appropriate doses of a LV may depend upon the particular recipient and the mode of administration. The appropriate dose level for any particular subject may depend upon a variety of factors including the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate: of excretion, other administered therapeutic agents, and the like.

[0174] In certain embodiments lentiviral gene therapy vectors described herein can be delivered to a subject by, for example, intravenous injection, local administration, or by stereotactic injection (see, *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA,* 91: 3054). In certain embodiments vectors may be delivered orally or inhalationally and may be encapsulated or otherwise manipulated to protect them from degradation, enhance uptake into tissues or cells, *etc.* Pharmaceutical preparations can include a LV in an acceptable diluent, or can comprise a slow release matrix in which a LV is imbedded. Alternatively or additionally, where a vector can be produced intact from recombinant cells, as is the case for retroviral or lentiviral vectors as described herein, a pharmaceutical preparation can include one or more cells which produce vectors. Pharmaceutical compositions comprising a LV described herein can be included in a container, pack, or dispenser, optionally together with instructions for administration.

[0175] The foregoing compositions, methods and uses are intended to be illustrative and not limiting. Using the teachings provided herein other variations on the compositions, methods and uses will be readily available to one of skill in the art.

[0176] The approach to generate reduced length enhance regions is superior to previous strategies for generating tissue-specific enhancers for, among other reasons: 1) The cost of goods is decreased due to a low number of outputs required to be tested, 2) Strength of synthetic enhancers may be superior to those produced with current methods, or they may be less active but more suitable for LV-mediated delivery, and 3). Enhancers can be of minimal length.

[0177] Additionally, without being bound to a particular theory, it is believed the enhancer mapping strategy described herein can be modified to generate genome-wide enhancer maps using a similar cloning strategy and sonicated human genomic DNA and that the mapping strategies can be used to generate synthetic enhancers responsive to an array of distinct cellular perturbations.

[0178] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

SEQUENCE LISTING

SEQ ID NO:1 ecHS1

catcaataattctagccccacaggagtttgttctgaaagtaaacttccacaaccgcaagcttattgaggctaaggcatctgtgaaggaaa
gaaacatctcctctaaaccactatgctgctagagcctcttttctgtactcaagcctcattcagacactagtgtcaccagtctcctcatataccc
tattgtattttcttcttcttgctggtttagtcatgttttctgggagcttaggggcttattttatttttgttttgttttctaatcaacagagatgggcaaac
ccattatttttttctttagacttgggatggtgatagctgggcagcgtcagaaactgtgtgtggatatagataagagctcaggactatgctga
gctgtgatgagggaggggcctagctaaaggcagtgagagtcagaatgctcctgctattgccttctcagtccccacgcttggtttctacac
aagtagatacatagaaaaggctataggttagtgtttgagagtcctgcatgattagttgctcagaaatgcccgataaatatgttatgtgtgttt
atgtatatatatgttttatatgtgtgtgtgtgtgttgtgtttacaaatatgtgattatcatcaaaacgtgaggg

SEQ ID NO:2 ecHS2

tacgtatatgtgtatatatatatatattcaggaaataatatattctagaatatgtcacattctgtctcaggcatccattttctttatgatgccgtt
tgaggtggagttttagtcaggtggtcagcttctccttttttttgccatctgccctgtaagcatcctgctggggacccagataggagtcatca
ctctaggctgagaacatctgggcacacaccctaagcctcagcatgactcatcatgactcagcattgctgtgcttgagccagaaggtttg
cttagaaggttacacagaaccagaaggcgggggtggggcactgaccccgacaggggcctggccagaactgctcatgcttggactat
gggaggtcactaatggagacacacagaaatgtaacaggaactaaggaaaaactgaagctt

SEQ ID NO:3 ecHS3

tgggggtatagggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcccccacctttcccatgtctgccctcta
ctcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgacctaaccacctatctgagcctgccagc
ctataacccatctgggccctgatagctggtggccagccctgaccccaccccacccctccctggaacctctgatagacacatctggcaca
ccagctcgcaaagtcaccgtgagggtcttgtgtttgctgagtcaaaattccttgaaatccaagtccttagagactcc

SEQ ID NO:4 ecHS4

Caggcttggattcaaagctcctgactttctgtctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaaaaaaaaatggaa
cccaaaatattctacatagtttccatgtcacagccagggctgggcagtctcctgttatttcttttaaaataaatatatcatttaaatgcataaat
aagcaaaccctgctcgggaatgggagggagagtctctggagtccacccctctcggccctggctctgcagatagtgctatcaaagcc
ctgacagagccctgcccattgctgggccttggagtgagtcagcctagtagaggaggcagggcaagccatctcatagctgctgagtggg
agagagaaaagggctcattgtctataaactcaggtcatggctattcttat

SEQ ID NO:5 coX2 (codon optimized exon 2)

CCTGAAGTTCTCAGGATCCACGTGCAGCTTGTCGCAGTGCAGCTCGCTCAGCTGG
GCGAAGGTGCCCTTCAGGTTGTCCAGGTGGGCCAGGCCGTCGCTGAAGGCGCCC
AGCACCTTCTTGCCGTGGGCCTTCACCTTGGGGTTGCCCATCACGGCGTCGGGGG
TGCTCAGGTCGCCGAAGCTCTCGAAGAAGCGCTGGGTCCAGGGGTACACCACCA
GCAGC

SEQ ID NO:6 full vector with HS1

gcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtga
gcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcgcgcaattaaccctcactaaagggaacaaaagctgg
agctgcaagcttggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattg
attattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc
ccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattg
acgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaat
gacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctat
taccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgt
caatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggc
gtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggtctctctggttagaccagatctgagcctgggagctctctg

gctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaa
ctagagatccctcagacccttttagtcagtgtggaaaatctctagcagtggcgcccgaacagggacttgaaagcgaaagggaaacca
gaggagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgactggtgagtacgccaaaa
attttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtattaagcgggggagaattagatcgcgatgggaaaa
aattcggttaaggccaggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaat
cctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagat
cattatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagacaagatagaggaa
gagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcagacctggaggaggagatatgagggacaattggaga
agtgaattatataaatataaagtagtaaaaattgaaccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaa
aaagagcagtgggaataggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg
tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaacagcatctgttgcaactca
cagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctggggatttggggtt
gctctggaaaactcatttgcaccactgctgtgccttggaatgctagttggagtaataaatctctggaacagatttggaatcacacgacctg
gatggagtgggacagagaaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatgaa
caagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggtatataaaattattcataatgatag
taggaggcttggtaggtttaagaatagtttttgctgtactttctatagtgaatagagttaggcagggatattcaccattatcgtttcagaccca
cctcccaaccccgaggggacccgacaggcccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgatt
agtgaacggatctcgacggtatcgatctcgacacaaatggcagtattcatccacaattttaaaagaaaaggggggattgggggtaca
gtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaacaaattacaaaaattcaaaattttcgg
gtttattacagggacagcagagatccagtttgggtcgaggatatcggatcggaattctctagatgatcaggatccctcgagcccttatcg
atcacgagactagcctcgactactagtggagatcccccgggctgcagagccagaagcaccataagggacatgataagggagccag
cagacctctgatctcttcctgaatgctaatcttaaacatcctgaggaagaatgggacttccatttggggtgggcctatgatagggtaataa
gacagtagtgaatatcaagctacaaaaagcccccttttcaaattcttctcagtcctaacttttcatactaagcccagtccttccaaagcagac
tgtgaaagagtgatagttccgggagactagcactgcagattccgggtcactgtgagtgggggaggcagggaagaagggctcacag
gacagtcaaaccatgccccctgttttttccttcttcaagtagacctctataagacaacagagacaactaaggctgagtggccaggcgagg
agaaaccatctcgccgtaaaacatggaaggaacacttcaggggaaaggtggtatctctaagcaagagaactgagtggagtcaaggct
gagagatgcaggataagcaaatgggtagtgaaaagacattcatgaggacagctaaaacaataagtaatgtaaaatacagcatagcaa
aactttaacctccaaatcaagcctctacttgaatccttttctgagggatgaataaggcataggcatcaggggctgttgccaatgtgcatta
gctgtttgcagcctcaccttctttcatggagtttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcact
gacctcccacattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttttttattaggcagaatccaga
tgctcaaggcccttcataatatcccccagtttagtagttggacttagggaacaaaggaaccttttaatagaaattggacagcaagaaagcg
agcttagtgatacttgtgggccagggcattagccacaccagccaccactttctgataggcagcctgcactggtggggtgaattctttgcc
aaagtgatgggccagcacacagaccagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaagg
gcctagcttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggattgtagctgctattagcaatatg
aaacctcttacatcagttacaatttatatgcagaaatatttatatgcagaaatattgctattgccttaacccagaaattatcactgttattcttta
gaatggtgcaaagaggcatgatacattgtatcattattgccctgaaagaaagagattagggaaagtattagaaataagataaacaaaaa
agtatattaaaagaagaaagcatttttaaaattacaaatgcaaaattaccctgatttggtcaatatgtgtaccctgttacttctccccttcctat
gacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcaccctgaagttctcaggatccacgtgca
gcttgtcacagtgcagctcactcagctgggcaaaggtgcccttgaggttgtccaggtgagccaggccatcactaaaggcaccgagca
ctttcttgccatgagccttcaccttagggttgcccataacagcatcaggagtggacagatccccaaaggactcaaagaacctctgggtc
caagggtagaccaccagcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagagtctt
ctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccagggcctcaccaccaacggca
tccacgttcaccttgtcccacagggcagtaacggcagacttctcctcaggagtcaggtgcaccatggtgtctgtttgaggttgctagtga
acacagttgtgtcagaagcaaatgtaagcaatagatggctctgccctgacttttatgcccagccctggctcctgccctccctgctcctgg
gagtagattggccaaccctagggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcactgg
cttaggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaattgctactaaaaacatcct
cctttgcaagtgtatttacccgacgcgtcggcgataagcttgatccatcgatcatcaataattctagccccacaggagtttgttctgaaagt
aaacttccacaaccgcaagcttattgaggctaaggcatctgtgaaggaaagaaacatctcctctaaaccactatgctgctagagcctctt
ttctgtactcaagcctcattcagacactagtgtcaccagtctcctcatatacctattgtattttcttcttcttgctggtttagtcatgttttctggga
gcttaggggcttattttatttttgttttgttttctaatcaacagagatgggcaaacccattatttttttctttagacttgggatggtgatagctgggc
agcgtcagaaactgtgtgtggatatagataagagctcaggactatgctgagctgtgatgagggaggggcctagctaaaggcagtgag

agtcagaatgctcctgctattgccttctcagtccccacgcttggtttctacacaagtagatacatagaaaaggctataggttagtgtttgag
agtcctgcatgattagttgctcagaaatgcccgataaatatgttatgtgtgtttatgtatatatatgttttatatgtgtgtgtgtgtgttgtgttt
acaaatatgtgattatcatcaaaacgtgagggtacgtatatgtgtatatatatatatattcaggaaataatatattctagaatatgtcacatt
ctgtctcaggcatccatttttctttatgatgccgtttgaggtggagttttagtcaggtggtcagcttctccttttttttgccatctgccctgtaagc
atcctgctggggacccagataggagtcatcactctaggctgagaacatctgggcacacaccctaagcctcagcatgactcatcatgac
tcagcattgctgtgcttgagccagaaggtttgcttagaaggttacacagaaccagaaggcgggggtggggcactgaccccgacagg
ggcctggccagaactgctcatgcttggactatgggaggtcactaatggagacacacagaaatgtaacaggaactaaggaaaaactga
agctttgggggtataggggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcccccacctttcccatgtctgcc
ctctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgacctaaccacctatctgagcctgc
cagcctataacccatctgggccctgatagctggtggccagccctgaccccaccccacccctccctggaacctctgatagacacatctgg
cacaccagctcgcaaagtcaccgtgagggtcttgtgtttgctgagtcaaaattccttgaaatccaagtccttagagactcccaggcttgg
attcaaagctcctgactttctgtctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaaaaaaaatggaacccaaaatat
tctacatagtttccatgtcacagccagggctgggcagtctcctgttatttctttaaaataaatatatcatttaaatgcataaataagcaaacc
ctgctcgggaatgggagggagagtctctggagtccacccccttctcggccctggctctgcagatagtgctatcaaagccctgacagagc
cctgcccattgctgggccttggagtgagtcagcctagtagagaggcagggcaagccatctcatagctgctgagtgggagagagaaa
agggctcattgtctataaactcaggtcatggctattcttatggcctactcgaccacgagggaattccgataatcaacctctggattacaaa
atttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctattgcttccc
gtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtg
cactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttcccctcccctattg
ccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcgg
ggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaat
ccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctccctt
tgggccgcctccccgcatcgataccgtcgacctcgagacctagaaaaacatggccaattcgagctcggtacctttaagaccaatgactt
acaaggcagctgtagatcttagccacttttaaaagaaaaggggggactggaagggctaattcactcccaacgaagacaagatccca
gggatgtacgtccctaacccgctaggggcagcacccaggcctgcactgccgcctgccggcaggggtccagtcctgctttttgcttgt
actgggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgcttaagcctcaataaagcttgcct
tgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttttagtcagtgtggaaaatctctag
cagtagtagttcatgtcatcttattattcagtatttataacttgcaaagaaatgaatatcagagagtgagaggaacttgtttattgcagcttata
atggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgta
tcttatcatgtctggctctagctatcccgccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatgg
ctgactaatttttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggaggctttttggaggcctag
gcttttgcgtcgagacgtacccaattcgccctatagtgagtcgtattacgcgcgctcactggccgtcgttttacaacgtcgtgactgggaa
aaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcc
cttcccaacagttgcgcagcctgaatggcgaatggcgcgacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacg
cgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttcccc
gtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttc
acgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaa
caacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaa
tttaacgcgaatttaacaaaatattaacgtttacaatttcccaggtggcacttttcggggaaatgtgcgcggaaccccctatttgtttatttttct
aaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacat
ttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatc
agttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatga
tgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctca
gaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataacca
tgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcat
gtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggca
acaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgca
ggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgca
gcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacag
atcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttta
atttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgt

agaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggttt
gtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagcc
gtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcga
taagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacaca
gcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggaga
aaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctt
tatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagc
aacgcggccatttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattac
cgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgccc
aatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtga

SEQ ID NO: 7 codon optimized exon 1

atgGTGCACCTGACCCCCGAGGAGAAGAGCGCCGTGACCGCCCTGTGGGACAAGG
TGAACGTGGACGCCGTGGGCGGCGAGGCCCTGggcag

SEQ ID NO:8 codon optimized exon 2 (in vector with optimized codons 1, 2, and 3)

cctgaagttctcggggtccacgtgcagcttgtcgcagtgcagctcgctcagctgggcgaaggtgcccttcaggttgtccaggtgggcc
aggccgtcgctgaaggcgcccagcaccttcttgccgtgggccttcaccttggggttgcccatcacggcgtcgggggtgctcaggtcg
ccgaagctctcgaagaagcgctgggtccaggggtacaccaccagcagc

SEQ ID NO:9 codon optimized exon 3

ctcCTGGGCAACGTGCTGGTGTGCGTGCTGGCCCACCACTTCGGCAAGGAGTTCAC
CCCCCCCGTGCAGGCCGCCTACCAGAAGGTGGTGGCCGGCGTGGCCAACGCCCT
GGCCCACAAGTACCACtaa

SEQ ID NO:10 vector comprising codon optimized exon 1

gcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtga
gcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcgcgcaattaaccctcactaaagggaacaaaagctgg
agctgcaagcttggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattg
attattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc
ccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattg
acgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaat
gacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctat
taccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgt
caatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggc
gtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggtctctctggttagaccagatctgagcctgggagctctctg
gctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaa
ctagagatccctcagacccttttagtcagtgtggaaaatctctagcagtggcgcccgaacagggacttgaaagcgaaagggaaacca
gaggagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgactggtgagtacgccaaaa
attttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtattaagcgggggagaattagatcgcgatgggaaaa
aattcggttaaggccaggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaat
cctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagat
cattatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagacaagatagaggaa
gagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcagacctggaggaggagatatgagggacaattggaga
agtgaattatataaatataaagtagtaaaaattgaaccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaa
aaagagcagtgggaataggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg
tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaacagcatctgttgcaactca
cagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctggggatttggggtt
gctctggaaaactcatttgcaccactgctgtgccttggaatgctagttggagtaataaatctctggaacagatttggaatcacacgacctg

gatggagtgggacagagaaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatgaa
caagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggtatataaaattattcataatgatag
taggaggcttggtaggtttaagaatagtttttgctgtactttctatagtgaatagagttaggcagggatattcaccattatcgtttcagaccca
cctcccaaccccgaggggacccgacgggcccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgatt
agtgaacggatctcgacggtatcgatctcgacacaaatggcagtattcatccacaattttaaaagaaaaggggggattgggggtaca
gtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaacaaattacaaaaattcaaaattttcgg
gtttattacagggacagcagagatccagtttgggtcgaggatatcggatcggaattctctagatgatcaggatccctcgagcccttatcg
atcacgagactagcctcgactactagtggagatcccccgggctgcagagccagaagcaccataagggacatgataagggagccag
cagacctctgatctcttcctgaatgctaatcttaaacatcctgaggaagaatgggacttccatttggggtgggcctatgatagggtaataa
gacagtagtgaatatcaagctacaaaaagcccccttttcaaattcttctcagtcctaacttttcatactaagcccagtccttccaaagcagac
tgtgaaagagtgatagttccgggagactagcactgcagattccgggtcactgtgagtgggggaggcagggaagaagggctcacag
gacagtcaaaccatgccccctgtttttccttcttcaagtagacctctataagacaacagagacaactaaggctgagtggccaggcgagg
agaaaccatctcgccgtaaaacatggaaggaacacttcaggggaaaggtggtatctctaagcaagagaactgagtggagtcaaggct
gagagatgcaggataagcaaatgggtagtgaaaagacattcatgaggacagctaaaacaataagtaatgtaaaatacagcatagcaa
aactttaacctccaaatcaagcctctacttgaatccttttctgagggatgaataaggcataggcatcaggggctgttgccaatgtgcatta
gctgtttgcagcctcaccttctttcatggagtttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcact
gacctcccacattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttttttattaggcagaatccaga
tgctcaaggcccttcataatatcccccagtttagtagttggacttagggaacaaaggaaccttttaatagaaattggacagcaagaaagcg
agcttagtgatacttgtgggccagggcattagccacaccagccaccactttctgataggcagcctgcactggtggggtgaattctttgcc
aaagtgatgggccagcacacagaccagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaagg
gcctagcttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggattgtagctgctattagcaatatg
aaacctcttacatcagttacaatttatatgcagaaatatttatatgcagaaatattgctattgccttaacccagaaattatcactgttattcttta
gaatggtgcaaagaggcatgatacattgtatcattattgccctgaaagaaagagattagggaaagtattagaaataagataaacaaaaa
agtatattaaaagaagaaagcatttttttaaaattacaaatgcaaaattaccctgatttggtcaatatgtgtaccctgttacttctccccttcctat
gacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcaccctgaagttctcaggatccacgtgca
gcttgtcacagtgcagctcactcagctgggcaaaggtgcccttgaggttgtccaggtgagccaggccatcactaaaggcaccgagca
ctttcttgccatgagccttcaccttagggttgcccataacagcatcaggagtggacagatccccaaaggactcaaagaacctctgggtc
caagggtagaccaccagcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagagtctt
ctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccagggcctcgccgcccacggc
gtccacgttcaccttgtcccacagggcggtcacggcgctcttctcctcgggggtcaggtgcaccatggtgtctgtttgaggttgctagtg
aacacagttgtgtcagaagcaaatgtaagcaatagatggctctgccctgacttttatgcccagccctggctcctgccctccctgctcctg
ggagtagattggccaaccctagggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcactg
gcttaggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaattgctactaaaaacatcc
tcctttgcaagtgtatttacccgacgcgtcggcgataagcttgatccatcgattacgtatatgtgtatatatatatatatattcaggaaataat
atattctagaatatgtcacattctgtctcaggcatccatttctttatgatgccgtttgaggtggagttttagtcaggtggtcagcttctccttttt
tttgccatctgccctgtaagcatcctgctggggacccagataggagtcatcactctaggctgagaacatctgggcacacaccctaagcc
tcagcatgactcatcatgactcagcattgctgtgcttgagccagaaggtttgcttagaaggttacacagaaccagaaggcgggggtgg
ggcactgaccccgacaggggcctggccagaactgctcatgcttggactatgggaggtcactaatggagacacacagaaatgtaaca
ggaactaaggaaaaactgaagctttgggggtatagggggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctcc
cccacctttcccatgtctgccctctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgacct
aaccacctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccctgaccccaccccacccctccctggaa
cctctgatagacacatctggcacaccagctcgcaaagtcaccgtgagggtcttgtgtttgctgagtcaaaattccttgaaatccaagtcct
tagagactcccaggcttggattcaaagctcctgactttctgtctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaaa
aaaaatggaacccaaaatattctacatagtttccatgtcacagccagggctgggcagtctcctgttatttctttttaaaataaatatatcattta
aatgcataaataagcaaaccctgctcgggaatgggagggagagtctctggagtccacccctttctcggccctggctctgcagatagtgc
tatcaaagccctgacagagccctgcccattgctgggccttggagtgagtcagcctagtagagaggcagggcaagccatctcatagct
gctgagtgggagagagaaaagggctcattgtctataaactcaggtcatggctattcttatggcctactcgaccacgagggaattccgat
aatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcct
ttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtca
ggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggact
ttcgctttcccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactga

caattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgct
acgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctc
agacgagtcggatctccctttgggccgcctccccgcatcgataccgtcgacctcgagacctagaaaaacatggccaattcgagctcg
gtacctttaagaccaatgacttacaaggcagctgtagatcttagccacttttaaaagaaaaggggggactggaagggctaattcactcc
caacgaagacaagatcccagggatgtacgtccctaacccgctaggggcagcacccaggcctgcactgccgcctgccggcaggg
gtccagtcctgctttttgcttgtactgggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgctt
aagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttta
gtcagtgtggaaaatctctagcagtagtagttcatgtcatcttattattcagtatttataacttgcaaagaaatgaatatcagagagtgagag
gaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtg
gtttgtccaaactcatcaatgtatcttatcatgtctggctctagctatcccgcccctaactccgcccatcccgcccctaactccgcccagtt
ccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtg
aggaggctttttggaggcctaggcttttgcgtcgagacgtacccaattcgccctatagtgagtcgtattacgcgcgctcactggccgtc
gttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagc
gaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcgacgcgccctgtagcggcgcattaa
gcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttc
tcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaa
aaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaata
gtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggtta
aaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcccaggtggcacttttcggggaaatgtgc
gcggaaccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaa
aggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggt
gaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttc
gccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaac
tcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaaga
gaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccg
cttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgac
accacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagac
tggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtga
gcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggca
actatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatac
tttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttc
gttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaa
aaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagata
ccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgtt
accagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcggg
ctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaa
gcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagctt
ccaggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggc
ggagcctatggaaaaacgccagcaacgcggccatttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtga
gcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggttt
cccgactggaaagcgggcagtga

SEQ ID NO:11 codon optimized exon 1 and 2

gcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtga
gcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcgcgcaattaaccctcactaaagggaacaaaagctgg
agctgcaagcttggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattg
attattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc
ccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattg
acgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgcccccctattgacgtcaat
gacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctat

taccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgt
caatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggc
gtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggtctctctggttagaccagatctgagcctgggagctctctg
gctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaa
ctagagatccctcagacccttttagtcagtgtggaaaatctctagcagtggcgcccgaacagggacttgaaagcgaaagggaaacca
gaggagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgactggtgagtacgccaaaa
attttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtattaagcgggggagaattagatcgcgatgggaaaa
aattcggttaaggccaggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaat
cctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagat
cattatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagacaagatagaggaa
gagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcagacctggaggaggagatatgagggacaattggaga
agtgaattatataaatataaagtagtaaaaattgaaccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaa
aaagagcagtgggaataggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg
tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaacagcatctgttgcaactca
cagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctggggatttggggtt
gctctggaaaactcatttgcaccactgctgtgccttggaatgctagttggagtaataaatctctggaacagatttggaatcacacgacctg
gatggagtgggacagagaaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatgaa
caagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggtatataaaattattcataatgatag
taggaggcttggtaggtttaagaatagtttttgctgtactttctatagtgaatagagttaggcagggatattcaccattatcgtttcagaccca
cctcccaaccccgaggggacccgacgggcccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgatt
agtgaacggatctcgacggtatcgatctcgacacaaatggcagtattcatccacaattttaaaagaaaaggggggattgggggtaca
gtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaacaaattacaaaaattcaaaattttcgg
gtttattacagggacagcagagatccagtttgggtcgaggatatcggatcggaattctctagatgatcaggatccctcgagcccttatcg
atcacgagactagcctcgactactagtggagatcccccgggctgcagagccagaagcaccataagggacatgataagggagccag
cagacctctgatctcttcctgaatgctaatcttaaacatcctgaggaagaatgggacttccatttggggtgggcctatgatagggtaataa
gacagtagtgaatatcaagctacaaaaagcccccttttcaaattcttctcagtcctaacttttcatactaagcccagtccttccaaagcagac
tgtgaaagagtgatagttccgggagactagcactgcagattccgggtcactgtgagtgggggaggcagggaagaagggctcacag
gacagtcaaaccatgccccctgtttttccttcttcaagtagacctctataagacaacagagacaactaaggctgagtggccaggcgagg
agaaaccatctcgccgtaaaacatggaaggaacacttcaggggaaaggtggtatctctaagcaagagaactgagtggagtcaaggct
gagagatgcaggataagcaaatgggtagtgaaaagacattcatgaggacagctaaaacaataagtaatgtaaaatacagcatagcaa
aactttaacctccaaatcaagcctctacttgaatccttttctgagggatgaataaggcataggcatcaggggctgttgccaatgtgcatta
gctgtttgcagcctcaccttctttcatggagtttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcact
gacctcccacattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgttttttattaggcagaatccaga
tgctcaaggcccttcataatatcccccagtttagtagttggacttagggaacaaaggaaccttaatagaaattggacagcaagaaagcg
agcttagtgatacttgtgggccagggcattagccacaccagccaccactttctgataggcagcctgcactggtggggtgaattctttgcc
aaagtgatgggccagcacacagaccagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaaagg
gcctagcttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggattgtagctgctattagcaatatg
aaacctcttacatcagttacaatttatatgcagaaatatttatatgcagaaatattgctattgccttaacccagaaattatcactgttattcttta
gaatggtgcaaagaggcatgatacattgtatcattattgccctgaaagaaagagattagggaaagtattagaaataagataaacaaaaa
agtatattaaaagaagaaagcatttttaaaattacaaatgcaaaattaccctgatttggtcaatatgtgtaccctgttacttctccccttcctat
gacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcaccctgaagttctcaggatccacgtgca
gcttgtcgcagtgcagctcgctcagctgggcgaaggtgcccttcaggttgtccaggtgggccaggccgtcgctgaaggcgcccagc
accttcttgccgtgggccttcaccttggggttgcccatcacggcgtcgggggtgctcaggtcgccgaagctctcgaagaagcgctgg
gtccaggggtacaccaccagcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaagag
tcttctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccagggcctcgccgcccacg
gcgtccacgttcaccttgtcccacagggcggtcacggcgctcttctcctcgggggtcaggtgcaccatggtgtctgtttgaggttgctag
tgaacacagttgtgtcagaagcaaatgtaagcaatagatggctctgccctgacttttatgcccagccctggctcctgccctccctgctcct
gggagtagattggccaaccctagggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggcact
ggcttaggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaattgctactaaaaacatc
ctcctttgcaagtgtatttacccgacgcgtcggcgataagcttgatccatcgattacgtatatgtgtatatatatatatatattcaggaaataa
tatattctagaatatgtcacattctgtctcaggcatccattttctttatgatgccgtttgaggtggagttttagtcaggtggtcagcttctccttttt

ttttgccatctgccctgtaagcatcctgctggggacccagataggagtcatcactctaggctgagaacatctgggcacacaccctaagc
ctcagcatgactcatcatgactcagcattgctgtgcttgagccagaaggtttgcttagaaggttacacagaaccagaaggcgggggtg
gggcactgaccccgacaggggcctggccagaactgctcatgcttggactatgggaggtcactaatggagacacacagaaatgtaac
aggaactaaggaaaaactgaagctttgggggtataggggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcctc
ccccacctttcccatgtctgccctctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctgacc
taaccacctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccctgaccccaccccaccctccctgga
acctctgatagacacatctggcacaccagctcgcaaagtcaccgtgagggtcttgtgtttgctgagtcaaaattccttgaaatccaagtc
cttagagactcccaggcttggattcaaagctcctgactttctgtctagtgtatgtgcagtgagccccttttcctctaactgaaagaaggaaa
aaaaaatggaacccaaaatattctacatagtttccatgtcacagccagggctgggcagtctcctgttatttcttttaaaataaatatatcattt
aaatgcataaataagcaaaccctgctcgggaatgggagggagagtctctggagtccacccccttctcggccctggctctgcagatagtg
ctatcaaagccctgacagagccctgcccattgctgggccttggagtgagtcagcctagtagaggcagggcaagccatctcatagc
tgctgagtgggagagagaaaagggctcattgtctataaactcaggtcatggctattcttatggcctactcgaccacgagggaattccgat
aatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcct
ttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtca
ggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggact
ttcgctttcccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactga
caattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgct
acgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctc
agacgagtcggatctccctttgggccgcctccccgcatcgataccgtcgacctcgagacctagaaaaacatggccaattcgagctcg
gtacctttaagaccaatgacttacaaggcagctgtagatcttagccacttttaaaagaaaaggggggactggaagggctaattcactcc
caacgaagacaagatcccagggatgtacgtccctaacccgctaggggcagcacccaggcctgcactgccgcctgccggcaggg
gtccagtcctgctttttgcttgtactgggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactgctt
aagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttta
gtcagtgtggaaaatctctagcagtagtagttcatgtcatcttattattcagtatttataacttgcaaagaaatgaatatcagagagtgagag
gaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtg
gtttgtccaaactcatcaatgtatcttatcatgtctggctctagctatcccgcccctaactccgcccatcccgcccctaactccgcccagtt
ccgcccattctccgccccatggctgactaatttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtg
aggaggctttttggaggcctaggcttttgcgtcgagacgtacccaattcgccctatagtgagtcgtattacgcgcgctcactggccgtc
gttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagc
gaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcgacgcgccctgtagcggcgcattaa
gcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttc
tcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgaccccaa
aaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaata
gtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggtta
aaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcccaggtggcacttttcggggaaatgtgc
gcggaaccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaa
aggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggt
gaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagtttc
gccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaac
tcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaaga
gaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccg
ctttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgac
accacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagac
tggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtga
gcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggca
actatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatac
tttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttc
gttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaa
aaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagata
ccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgtt
accagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcggg

ctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaa
gcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagctt
ccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggc
ggagcctatggaaaaacgccagcaacgcggccatttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtga
gcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggttt
cccgactggaaagcgggcagtga

SEQ ID NO:12 codon optimized exon 1, 2 and 3

gcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtga
gcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcgcgcaattaaccctcactaaagggaacaaaagctgg
agctgcaagcttggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattg
attattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggc
ccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattg
acgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaat
gacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctat
taccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgt
caatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggc
gtgtacggtgggaggtctatataagcagagctcgtttagtgaaccggggtctctctggttagaccagatctgagcctgggagctctctg
gctaactagggaacccactgcttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaa
ctagagatccctcagacccttttagtcagtgtggaaaatctctagcagtggcgcccgaacagggacttgaaagcgaaagggaaacca
gaggagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgactggtgagtacgccaaaa
attttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtattaagcgggggagaattagatcgcgatgggaaaa
aattcggttaaggccaggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaat
cctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagat
cattatataatacagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagacaagatagaggaa
gagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcagacctggaggaggagatatgagggacaattggaga
agtgaattatataaatataaagtagtaaaaattgaaccattaggagtagcacccaccaaggcaaagagaagagtggtgcagagagaaa
aaagagcagtgggaataggagctttgttccttgggttcttgggagcagcaggaagcactatgggcgcagcgtcaatgacgctgacgg
tacaggccagacaattattgtctggtatagtgcagcagcagaacaatttgctgagggctattgaggcgcaacagcatctgttgcaactca
cagtctggggcatcaagcagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctggggatttggggtt
gctctggaaaactcatttgcaccactgctgtgccttggaatgctagttggagtaataaatctctggaacagatttggaatcacacgacctg
gatggagtgggacagagaaattaacaattacacaagcttaatacactccttaattgaagaatcgcaaaaccagcaagaaaagaatgaa
caagaattattggaattagataaatgggcaagtttgtggaattggtttaacataacaaattggctgtggtatataaaattattcataatgatag
taggaggcttggtaggtttaagaatagtttttgctgtactttctatagtgaatagagttaggcagggatattcaccattatcgtttcagaccca
cctcccaaccccgaggggacccgacaggcccgaaggaatagaagaagaaggtggagagagagacagagacagatccattcgatt
agtgaacggatctcgacggtatcgatctcgacacaaatggcagtattcatccacaatttttaaaagaaaaggggggattgggggtaca
gtgcaggggaaagaatagtagacataatagcaacagacatacaaactaaagaattacaaaaacaaattacaaaaattcaaaattttcgg
gtttattacagggacagcagagatccagtttgggtcgaggatatcggatcggaattctctagatgatcaggatccctcgagcccttatcg
atcacgagactagcctcgactactagtggagatcccccgggctgcagagccagaagcaccataagggacatgataagggagccag
cagacctctgatctcttcctgaatgctaatcttaaacatcctgaggaagaatgggacttccatttggggtgggcctatgatagggtaataa
gacagtagtgaatatcaagctacaaaaagcccccttttcaaattcttctcagtcctaacttttcatactaagcccagtccttccaaagcagac
tgtgaaagagtgatagttccgggagactagcactgcagattccgggtcactgtgagtgggggaggcagggaagaagggctcacag
gacagtcaaaccatgccccctgtttttccttcttcaagtagacctctataagacaacagagacaactaaggctgagtggccaggcgagg
agaaaccatctcgccgtaaaacatggaaggaacacttcaggggaaaggtggtatctctaagcaagagaactgagtggagtcaaggct
gagagatgcaggataagcaaatgggtagtgaaaagacattcatgaggacagctaaaacaataagtaatgtaaaatacagcatagcaa
aactttaacctccaaatcaagcctctacttgaatccttttctgagggatgaataaggcataggcatcaggggctgttgccaatgtgcatta
gctgtttgcagcctcaccttctttcatggagtttaagatatagtgtattttcccaaggtttgaactagctcttcatttctttatgttttaaatgcact
gacctcccacattccctttttagtaaaatattcagaaataatttaaatacatcattgcaatgaaaataaatgtttttttattaggcagaatccaga
tgctcaaggcccttcataatatcccccagtttagtagttggacttagggaacaaaggaaccttttaatagaaattggacagcaagaaagcg

agcttagtggtacttgtgggccagggcgttggccacgccggccaccaccttctggtaggcggcctgcacgggggggtgaactcctt
gccgaagtggtgggccagcacgcacaccagcacgttgcccaggagctgtgggaggaagataagaggtatgaacatgattagcaaa
agggcctagcttggactcagaataatccagccttatcccaaccataaaataaaagcagaatggtagctggattgtagctgctattagcaa
tatgaaacctcttacatcagttacaatttatatgcagaaatatttatatgcagaaatattgctattgccttaacccagaaattatcactgttattc
tttagaatggtgcaaagaggcatgatacattgtatcattattgccctgaaagaaagagattagggaaagtattagaaataagataaacaa
aaaagtatattaaaagaagaaagcatttttttaaaattacaaatgcaaaattaccctgatttggtcaatatgtgtaccctgttacttctccccttc
ctatgacatgaacttaaccatagaaaagaaggggaaagaaaacatcaagggtcccatagactcaccctgaagttctcggggtccacgt
gcagcttgtcgcagtgcagctcgctcagctgggcgaaggtgcccttcaggttgtccaggtgggccaggccgtcgctgaaggcgccc
agcaccttcttgccgtgggccttcaccttggggttgcccatcacggcgtcgggggtgctcaggtcgccgaagctctcgaagaagcgct
gggtccaggggtacaccaccagcagcctaagggtgggaaaatagaccaataggcagagagagtcagtgcctatcagaaacccaag
agtcttctctgtctccacatgcccagtttctattggtctccttaaacctgtcttgtaaccttgataccaacctgcccagggcctcgccgccca
cggcgtccacgttcaccttgtcccacagggcggtcacggcgctcttctcctcgggggtcaggtgcaccatggtgtctgtttgaggttgct
agtgaacacagttgtgtcagaagcaaatgtaagcaatagatggctctgccctgacttttatgcccagccctggctcctgccctccctgct
cctgggagtagattggccaaccctagggtgtggctccacagggtgaggtctaagtgatgacagccgtacctgtccttggctcttctggc
actggcttaggagttggacttcaaaccctcagccctccctctaagatatatctcttggccccataccatcagtacaaattgctactaaaaac
atcctcctttgcaagtgtatttacccgacgcgtcggcgataagcttgatccatcgattacgtatatgtgtatatatatatatattcaggaaa
taatatattctagaatatgtcacattctgtctcaggcatccattttctttatgatgccgtttgaggtggagttttagtcaggtggtcagcttctcc
ttttttttgccatctgccctgtaagcatcctgctggggacccagataggagtcatcactctaggctgagaacatctgggcacacaccctaa
gcctcagcatgactcatcatgactcagcattgctgtgcttgagccagaaggtttgcttagaaggttacacagaaccagaaggcgggggg
tggggcactgaccccgacaggggcctggccagaactgctcatgcttggactatgggaggtcactaatggagacacacagaaatgta
acaggaactaaggaaaaactgaagctttgggggtataggggagcagtcccatgtagtagtagaatgaaaaatgctgctatgctgtgcc
tcccccacctttcccatgtctgccctctactcatggtctatctctcctggctcctgggagtcatggactccacccagcaccaccaacctga
cctaaccacctatctgagcctgccagcctataacccatctgggccctgatagctggtggccagccctgaccccaccccacccctccctg
gaacctctgatagacacatctggcacaccagctcgcaaagtcaccgtgagggtcttgtgtttgctgagtcaaaattccttgaaatccaag
tccttagagactcccaggcttggattcaaagctcctgactttctgtctagtgtatgtgcagtgagccccttttcctctaactgaaagaagga
aaaaaaaatggaacccaaaatattctacatagtttccatgtcacagccagggctgggcagtctcctgttatttcttttaaaataaatatatca
tttaaatgcataaataagcaaaccctgctcgggaatgggagggagagtctctggagtccacccctctcggccctggctctgcagatag
tgctatcaaagccctgacagagccctgcccattgctgggccttggagtgagtcagcctagtagaggcagggcaagccatctcata
gctgctgagtgggagagagaaaagggctcattgtctataaactcaggtcatggctattcttatggcctactcgaccacgagggaattcc
gataatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaat
gcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgtt
gtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgg
gactttcgctttcccctcccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggca
ctgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttc
tgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgc
cctcagacgagtcggatctccctttgggccgcctccccgcatcgataccgtcgacctcgagacctagaaaaacatggccaattcgagc
tcggtacctttaagaccaatgacttacaaggcagctgtagatcttagccacttttaaaagaaaaggggggactggaagggctaattcac
tcccaacgaagacaagatcccagggatgtacgtccctaacccgctaggggcagcacccaggcctgcactgccgcctgccggcag
gggtccagtcctgctttttgcttgtactgggtctctctggttagaccagatctgagcctgggagctctctggctaactagggaacccactg
cttaagcctcaataaagcttgccttgagtgcttcaagtagtgtgtgcccgtctgttgtgtgactctggtaactagagatccctcagacccttt
tagtcagtgtggaaaatctctagcagtagtagttcatgtcatcttattattcagtatttataacttgcaaagaaatgaatatcagagagtgag
aggaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttg
tggtttgtccaaactcatcaatgtatcttatcatgtctggctctagctatcccgcccctaactccgcccatcccgcccctaactccgcccag
ttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagt
gaggaggcttttttggaggcctaggcttttgcgtcgagacgtacccaattcgccctatagtgagtcgtattacgcgcgctcactggccgt
cgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatag
cgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcgacgcgccctgtagcggcgcatta
agcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttt
ctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctcccttagggttccgatttagtgctttacggcacctcgacccca
aaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaat
agtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggtt

aaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcccaggtggcacttttcggggaaatgtg
cgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaa
aaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctgg
tgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttc
gccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaac
tcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaaga
gaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccg
ctttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgac
accacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagac
tggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtga
gcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggca
actatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatac
tttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagtttc
gttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaa
aaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagata
ccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgtt
accagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcggg
ctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaa
gcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagctt
ccaggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggc
ggagcctatggaaaaacgccagcaacgcggccatttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcc
cctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtga
gcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggttt
cccgactggaaagcgggcagtga

SEQ ID NO:13 pUV-AS3 from junction marker

ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCC
TTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTT
AGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGT
GATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT
TGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA
CCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATT
GGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATT
AACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT
GTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGA
TAAATGCTTCAATAATAGCACCTAGATCAAGAGACAGGATGAGGATCGTTTCGC
ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGG
CTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGT
TCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGG
TGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGAC
GGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTG
GCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCT
GCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATC
CGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA
CTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGG
GGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCG
AGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAA
TGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTAT
CAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGG

GCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCG
CCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTATTAACGCTTACAATTTCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCA
AATATGTATCCGCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG
CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG
CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGC
AGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA
CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTG
CCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGA
TAAGGCGCAGCGGTCGGGCTGAACGGGGGGGTCGTGCACACAGCCCAGCTTGGA
GCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGC
CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGG
AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAG
TCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGG
CCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGAC
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAC
CGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCC
CGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCA
TTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG
TGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC
GCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTGCAAGCTTGGCCAT
TGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTCCAACA
TTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGG
GGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACG
TATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGT
ATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC
GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC
ATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT
TACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAC
TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCA
AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT
GAACCGGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAAC
TAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGT
GTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGT
CAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGG
GAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGC
AAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGG
CTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTA
GATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAA
TTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCT
GGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCA
TCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCC
TCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACA
AGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCT
GATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATA

TAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAA
GAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCT
TGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGAC
GGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTG
AGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAG
CAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTC
CTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGA
ATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGAT
GGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGA
AGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATA
AATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAA
ATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTA
CTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCC
ACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGT
GGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCTCGACGGTAT
CGATCTCGACACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGA
TTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATAC
AAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTA
CAGGGACAGCAGACATGAGGACAGCTAAAACAATAAGTAATGTAAAATACAGC
ATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGA
TGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGC
CTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAG
CTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAA
ATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTA
GGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGA
CTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTA
GTGATACTTGTGGGCCAGGGCATTAGCCACACCAGCCACCACTTTCTGATAGGCA
GCCTGCACTGGTGGGGTGAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACC
AGCACGTTGCCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGC
AAAAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAATA
AAAGCAGAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACA
TCAGTTACAATTTATATGCAGAAATACCCTGTTACTTCTCCCCTTCCTATGACATG
AACTTAACCATAGAAAAGAAGGGGAAAGAAAACATCAAGGGTCCCATAGACTC
ACCCTGAAGTTCTCAGGATCCACGTGCAGCTTGTCACAGTGCAGCTCACTCAGCT
GGGCAAAGGTGCCCTTGAGGTTGTCCAGGTGAGCCAGGCCATCACTAAAGGCAC
CGAGCACTTTCTTGCCATGAGCCTTCACCTTAGGGTTGCCCATAACAGCATCAGG
AGTGGACAGATCCCCAAAGGACTCAAAGAACCTCTGGGTCCAAGGGTAGACCAC
CAGCAGCCTAAGGGTGGGAAAATAGACCAATAGGCAGAGAGAGTCAGTGCCTAT
CAGAAACCCAAGAGTCTTCTCTGTCTCCACATGCCCAGTTTCTATTGGTCTCCTTA
AACCTGTCTTGTAACCTTGATACCAACCTGCCCAGGGCCTCACCACCAACGGCAT
CCACGTTCACCTTGTCCCACAGGGCAGTAACGGCAGACTTCTCCTCAGGAGTCAG
GTGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAA
ATGTAAGCAATAGATGGCTCTGCCCTGACTTTTATGCCCAGCCCTGGCTCCTGCC
CTCCCTGCTCCTGGGAGTAGATTGGCCAACCCTAGGGTGTGGCTCCACAGGGTGA
GGTCTAAGTGATGACAGCCGTACCTGTCCTTGGCTCTTCTGGCACTGGCTTAGGA
GTTGGACTTCAAACCCTCAGCCCTCCCTCTAAGATATATCTCTTGGCCCCATACCA
TCAGTACAAATTGCTACTAAAAACATCCTCCTTTGCAAGTGTATTTACTACGTAT
ATGTGTATATATATATATATATTCAGGAAATAATATATTCTAGAATATGTCACAT
TCTGTCTCAGGCATCCATTTTCTTTATGATGCCGTTTGAGGTGGAGTTTTAGTCAG
GTGGTCAGCTTCTCCTTTTTTTTGCCATCTGCCCTGTAAGCATCCTGCTGGGGACC
CAGATAGGAGTCATCACTCTAGGCTGAGAACATCTGGGCACACACCCTAAGCCT

CAGCATGACTCATCATGACTCAGCATTGCTGTGCTTGAGCCAGAAGGTTTGCTTA
GAAGGTTACACAGAACCAGAAGGCGGGGGTGGGGCACTGACCCCGACAGGGGC
CTGGCCAGAACTGCTCATGCTTGGACTATGGGAGGTCACTAATGGAGACACACA
GAAATGTAACAGGAACTAAGGAAAAACTGAAGCTTTGGGGGTATAGGGGAGCA
GTCCCATGTAGTAGTAGAATGAAAAATGCTGCTATGCTGTGCCTCCCCCACCTTT
CCCATGTCTGCCCTCTACTCATGGTCTATCTCTCCTGGCTCCTGGGAGTCATGGAC
TCCACCCAGCACCACCAACCTGACCTAACCACCTATCTGAGCCTGCCAGCCTATA
ACCCATCTGGGCCCTGATAGCTGGTGGCCAGCCCTGACCCCACCCCACCCTCCCT
GGAACCTCTGATAGACACATCTGGCACACCAGCTCGCAAAGTCACCGTGAGGGT
CTTGTGTTTGCTGAGTCAAAATTCCTTGAAATCCAAGTCCTTAGAGACTCCCAGG
CTTGGATTCAAAGCTCCTGACTTTCTGTCTAGTGTATGTGCAGTGAGCCCCTTTTC
CTCTAACTGAAAGAAGGAAAAAAAAATGGAACCCAAAATATTCTACATAGTTTC
CATGTCACAGCCAGGGCTGGGCAGTCTCCTGTTATTTCTTTTAAAATAAATATAT
CATTTAAATGCATAAATAAGCAAACCCTGCTCGGGAATGGGAGGGAGAGTCTCT
GGAGTCCACCCCTTCTCGGCCCTGGCTCTGCAGATAGTGCTATCAAAGCCCTGAC
AGAGCCCTGCCCATTGCTGGGCCTTGGAGTGAGTCAGCCTAGTAGAGAGGCAGG
GCAAGCCATCTCATAGCTGCTGAGTGGGAGAGAGAAAAGGGCTCATTGTCTATA
AACTCAGGTCATGGCTATTCTTATTAAAAGAAAAGGGGGGACTGGAAGGGCTAA
TTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTGTACTGGGTCTCTCTGGTTA
GACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGC
CTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTTGGTTTTTTGTGTGTGCA
TTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGCACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATG
GAGTAGTAGTTCATGTCATCTTATTATTCAGTATTTATAACTTGCAAAGAAATGA
ATATCAGAGAGTGAGAGGAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAA
GCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGT
GGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGCTCTAGCTATCCCGCCC
CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCA
TGGCTCTCACTACTTCTGGAATAGCTCAGAGGCCGAGGCGGCCTCGGCCTCTGCA
TAAATAAAAAAAATTAGTCGAGGCTTTTTTGGAGGCCTAGGGACGTACCCAATTC
GCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGT
GACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTT
TCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTT
GCGCAGCCTGAATGGCGAATGGG

SEQ ID NO:14 pUV-AS3(coX2) from junction marker

ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCC
TTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTT
AGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGT
GATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT
TGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA
CCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATT
GGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATT
AACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT
GTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGA
TAAATGCTTCAATAATAGCACCTAGATCAAGAGACAGGATGAGGATCGTTTCGC
ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGG
CTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGT
TCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGG

TGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGAC
GGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTG
GCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCT
GCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATC
CGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA
CTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGG
GGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCG
AGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAA
TGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTAT
CAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGG
GCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCG
CCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTATTAACGCTTACAATTTCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCA
AATATGTATCCGCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG
CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG
CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGC
AGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA
CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTG
CCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGA
TAAGGCGCAGCGGTCGGGCTGAACGGGGGGGTCGTGCACACAGCCCAGCTTGGA
GCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGC
CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGG
AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAG
TCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGG
CCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGAC
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAC
CGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCC
CGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCA
TTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG
TGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC
GCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTGCAAGCTTGGCCAT
TGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTCCAACA
TTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGG
GGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACG
TATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGT
ATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC
GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC
ATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT
TACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAC
TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCA
AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT
GAACCGGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAAC
TAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGT
GTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGT
CAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGG

GAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGC
AAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGG
CTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTA
GATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAA
TTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCT
GGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCA
TCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCC
TCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACA
AGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCT
GATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATA
TAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAA
GAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCT
TGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGAC
GGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTG
AGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAG
CAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTC
CTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGA
ATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGAT
GGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGA
AGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATA
AATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAA
ATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTA
CTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCC
ACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGT
GGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCTCGACGGTAT
CGATCTCGACACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGA
TTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATAC
AAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTA
CAGGGACAGCAGACATGAGGACAGCTAAAACAATAAGTAATGTAAAATACAGC
ATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGA
TGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGC
CTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAG
CTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAA
ATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTA
GGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGA
CTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTA
GTGATACTTGTGGGCCAGGGCATTAGCCACACCAGCCACCACTTTCTGATAGGCA
GCCTGCACTGGTGGGGTGAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACC
AGCACGTTGCCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGC
AAAAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAATA
AAAGCAGAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACA
TCAGTTACAATTTATATGCAGAAATACCCTGTTACTTCTCCCCTTCCTATGACATG
AACTTAACCATAGAAAAGAAGGGGAAAGAAAACATCAAGGGTCCCATAGACTC
ACCCTGAAGTTCTCAGGATCCACGTGCAGCTTGTCGCAGTGCAGCTCGCTCAGCT
GGGCGAAGGTGCCCTTCAGGTTGTCCAGGTGGGCCAGGCCGTCGCTGAAGGCGC
CCAGCACCTTCTTGCCGTGGGCCTTCACCTTGGGGTTGCCCATCACGGCGTCGGG
GGTGCTCAGGTCGCCGAAGCTCTCGAAGAAGCGCTGGGTCCAGGGGTACACCAC
CAGCAGCCTAAGGGTGGGAAAATAGACCAATAGGCAGAGAGAGTCAGTGCCTAT
CAGAAACCCAAGAGTCTTCTCTGTCTCCACATGCCCAGTTTCTATTGGTCTCCTTA
AACCTGTCTTGTAACCTTGATACCAACCTGCCCAGGGCCTCACCACCAACGGCAT
CCACGTTCACCTTGTCCCACAGGGCAGTAACGGCAGACTTCTCCTCAGGAGTCAG

GTGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAA
ATGTAAGCAATAGATGGCTCTGCCCTGACTTTTATGCCCAGCCCTGGCTCCTGCC
CTCCCTGCTCCTGGGAGTAGATTGGCCAACCCTAGGGTGTGGCTCCACAGGGTGA
GGTCTAAGTGATGACAGCCGTACCTGTCCTTGGCTCTTCTGGCACTGGCTTAGGA
GTTGGACTTCAAACCCTCAGCCCTCCCTCTAAGATATATCTCTTGGCCCCATACCA
TCAGTACAAATTGCTACTAAAAACATCCTCCTTTGCAAGTGTATTTACTACGTAT
ATGTGTATATATATATATATATTCAGGAAATAATATATTCTAGAATATGTCACAT
TCTGTCTCAGGCATCCATTTTCTTTATGATGCCGTTTGAGGTGGAGTTTTAGTCAG
GTGGTCAGCTTCTCCTTTTTTTTGCCATCTGCCCTGTAAGCATCCTGCTGGGGACC
CAGATAGGAGTCATCACTCTAGGCTGAGAACATCTGGGCACACACCCTAAGCCT
CAGCATGACTCATCATGACTCAGCATTGCTGTGCTTGAGCCAGAAGGTTTGCTTA
GAAGGTTACACAGAACCAGAAGGCGGGGGTGGGGCACTGACCCCGACAGGGGC
CTGGCCAGAACTGCTCATGCTTGGACTATGGGAGGTCACTAATGGAGACACACA
GAAATGTAACAGGAACTAAGGAAAAACTGAAGCTTTGGGGGTATAGGGGAGCA
GTCCCATGTAGTAGTAGAATGAAAAATGCTGCTATGCTGTGCCTCCCCCACCTTT
CCCATGTCTGCCCTCTACTCATGGTCTATCTCTCCTGGCTCCTGGGAGTCATGGAC
TCCACCCAGCACCACCAACCTGACCTAACCACCTATCTGAGCCTGCCAGCCTATA
ACCCATCTGGGCCCTGATAGCTGGTGGCCAGCCCTGACCCCACCCCACCCTCCCT
GGAACCTCTGATAGACACATCTGGCACACCAGCTCGCAAAGTCACCGTGAGGGT
CTTGTGTTTGCTGAGTCAAAATTCCTTGAAATCCAAGTCCTTAGAGACTCCCAGG
CTTGGATTCAAAGCTCCTGACTTTCTGTCTAGTGTATGTGCAGTGAGCCCCTTTTC
CTCTAACTGAAAGAAGGAAAAAAAAATGGAACCCAAAATATTCTACATAGTTTC
CATGTCACAGCCAGGGCTGGGCAGTCTCCTGTTATTTCTTTTAAAATAAATATAT
CATTTAAATGCATAAATAAGCAAACCCTGCTCGGGAATGGGAGGGAGAGTCTCT
GGAGTCCACCCCTTCTCGGCCCTGGCTCTGCAGATAGTGCTATCAAAGCCCTGAC
AGAGCCCTGCCCATTGCTGGGCCTTGGAGTGAGTCAGCCTAGTAGAGAGGCAGG
GCAAGCCATCTCATAGCTGCTGAGTGGGAGAGAGAAAAGGGCTCATTGTCTATA
AACTCAGGTCATGGCTATTCTTATTAAAAGAAAAGGGGGGACTGGAAGGGCTAA
TTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTGTACTGGGTCTCTCTGGTTA
GACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGC
CTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTTGGTTTTTTGTGTGTGCA
TTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGCACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATG
GAGTAGTAGTTCATGTCATCTTATTATTCAGTATTTATAACTTGCAAAGAAATGA
ATATCAGAGAGTGAGAGGAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAA
GCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGT
GGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGCTCTAGCTATCCCGCCC
CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCA
TGGCTCTCACTACTTCTGGAATAGCTCAGAGGCCGAGGCGGCCTCGGCCTCTGCA
TAAATAAAAAAAATTAGTCGAGGCTTTTTTGGAGGCCTAGGGACGTACCCAATTC
GCCCTATAGTGAGTCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGT
GACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTT
TCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTT
GCGCAGCCTGAATGGCGAATGGG

SEQ ID NO:15 pUV-HS1-AS3 from junction marker

ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCC
TTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTT
AGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGT

GATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT
TGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA
CCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATT
GGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATT
AACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT
GTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGA
TAAATGCTTCAATAATAGCACCTAGATCAAGAGACAGGATGAGGATCGTTTCGC
ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGG
CTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGT
TCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGG
TGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGAC
GGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTG
GCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCT
GCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATC
CGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA
CTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGG
GGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCG
AGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAA
TGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTAT
CAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGG
GCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCG
CCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTATTAACGCTTACAATTTCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCA
AATATGTATCCGCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG
CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG
CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGC
AGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA
CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTG
CCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGA
TAAGGCGCAGCGGTCGGGCTGAACGGGGGGGTCGTGCACACAGCCCAGCTTGGA
GCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGC
CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGG
AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAG
TCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGG
CCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGAC
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAC
CGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCC
CGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCA
TTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG
TGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC
GCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTGCAAGCTTGGCCAT
TGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTCCAACA
TTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGG
GGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACG
TATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGT
ATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC

GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC
ATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT
TACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAC
TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCA
AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT
GAACCGGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAAC
TAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGT
GTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGT
CAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGG
GAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGC
AAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGG
CTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTA
GATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAA
TTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCT
GGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCA
TCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCC
TCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACA
AGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCT
GATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATA
TAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAA
GAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCT
TGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGAC
GGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTG
AGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAG
CAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTC
CTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGA
ATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGAT
GGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGA
AGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATA
AATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAA
ATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTA
CTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCC
ACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGT
GGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCTCGACGGTAT
CGATCTCGACACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGA
TTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATAC
AAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTA
CAGGGACAGCAGACATGAGGACAGCTAAAACAATAAGTAATGTAAAATACAGC
ATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGA
TGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGC
CTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAG
CTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAA
ATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTA
GGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGA
CTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTA
GTGATACTTGTGGGCCAGGGCATTAGCCACACCAGCCACCACTTTCTGATAGGCA
GCCTGCACTGGTGGGGTGAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACC
AGCACGTTGCCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGC
AAAAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAATA
AAAGCAGAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACA

TCAGTTACAATTTATATGCAGAAATACCCTGTTACTTCTCCCCTTCCTATGACATG
AACTTAACCATAGAAAAGAAGGGGAAAGAAAACATCAAGGGTCCCATAGACTC
ACCCTGAAGTTCTCAGGATCCACGTGCAGCTTGTCACAGTGCAGCTCACTCAGCT
GGGCAAAGGTGCCCTTGAGGTTGTCCAGGTGAGCCAGGCCATCACTAAAGGCAC
CGAGCACTTTCTTGCCATGAGCCTTCACCTTAGGGTTGCCCATAACAGCATCAGG
AGTGGACAGATCCCCAAAGGACTCAAAGAACCTCTGGGTCCAAGGGTAGACCAC
CAGCAGCCTAAGGGTGGGAAAATAGACCAATAGGCAGAGAGAGTCAGTGCCTAT
CAGAAACCCAAGAGTCTTCTCTGTCTCCACATGCCCAGTTTCTATTGGTCTCCTTA
AACCTGTCTTGTAACCTTGATACCAACCTGCCCAGGGCCTCACCACCAACGGCAT
CCACGTTCACCTTGTCCCACAGGGCAGTAACGGCAGACTTCTCCTCAGGAGTCAG
GTGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAA
ATGTAAGCAATAGATGGCTCTGCCCTGACTTTTATGCCCAGCCCTGGCTCCTGCC
CTCCCTGCTCCTGGGAGTAGATTGGCCAACCCTAGGGTGTGGCTCCACAGGGTGA
GGTCTAAGTGATGACAGCCGTACCTGTCCTTGGCTCTTCTGGCACTGGCTTAGGA
GTTGGACTTCAAACCCTCAGCCCTCCCTCTAAGATATATCTCTTGGCCCCATACCA
TCAGTACAAATTGCTACTAAAAACATCCTCCTTTGCAAGTGTATTTACCATCAAT
AATTCTAGCCCCACAGGAGTTTGTTCTGAAAGTAAACTTCCACAACCGCAAGCTT
ATTGAGGCTAAGGCATCTGTGAAGGAAAGAAACATCTCCTCTAAACCACTATGCT
GCTAGAGCCTCTTTTCTGTACTCAAGCCTCATTCAGACACTAGTGTCACCAGTCTC
CTCATATACCTATTGTATTTTCTTCTTCTTGCTGGTTTAGTCATGTTTTCTGGGAGC
TTAGGGGCTTATTTTATTTTGTTTTGTTTTCTAATCAACAGAGATGGGCAAACCCA
TTATTTTTTTCTTTAGACTTGGGATGGTGATAGCTGGGCAGCGTCAGAAACTGTGT
GTGGATATAGATAAGAGCTCAGGACTATGCTGAGCTGTGATGAGGGAGGGGCCT
AGCTAAAGGCAGTGAGAGTCAGAATGCTCCTGCTATTGCCTTCTCAGTCCCCACG
CTTGGTTTCTACACAAGTAGATACATAGAAAAGGCTATAGGTTAGTGTTTGAGAG
TCCTGCATGATTAGTTGCTCAGAAATGCCCGATAAATATGTTATGTGTGTTTATGT
ATATATATGTTTTATATGTGTGTGTGTGTGTGTTGTGTTTACAAATATGTGATTAT
CATCAAAACGTGAGGGTACGTATATGTGTATATATATATATATATTCAGGAAATA
ATATATTCTAGAATATGTCACATTCTGTCTCAGGCATCCATTTTCTTTATGATGCC
GTTTGAGGTGGAGTTTTAGTCAGGTGGTCAGCTTCTCCTTTTTTTTGCCATCTGCC
CTGTAAGCATCCTGCTGGGGACCCAGATAGGAGTCATCACTCTAGGCTGAGAAC
ATCTGGGCACACACCCTAAGCCTCAGCATGACTCATCATGACTCAGCATTGCTGT
GCTTGAGCCAGAAGGTTTGCTTAGAAGGTTACACAGAACCAGAAGGCGGGGGTG
GGGCACTGACCCCGACAGGGGCCTGGCCAGAACTGCTCATGCTTGGACTATGGG
AGGTCACTAATGGAGACACACAGAAATGTAACAGGAACTAAGGAAAAACTGAA
GCTTTGGGGGTATAGGGGAGCAGTCCCATGTAGTAGTAGAATGAAAAATGCTGC
TATGCTGTGCCTCCCCCACCTTTCCCATGTCTGCCCTCTACTCATGGTCTATCTCTC
CTGGCTCCTGGGAGTCATGGACTCCACCCAGCACCACCAACCTGACCTAACCACC
TATCTGAGCCTGCCAGCCTATAACCCATCTGGGCCCTGATAGCTGGTGGCCAGCC
CTGACCCCACCCCACCCTCCCTGGAACCTCTGATAGACACATCTGGCACACCAGC
TCGCAAAGTCACCGTGAGGGTCTTGTGTTTGCTGAGTCAAAATTCCTTGAAATCC
AAGTCCTTAGAGACTCCCAGGCTTGGATTCAAAGCTCCTGACTTTCTGTCTAGTG
TATGTGCAGTGAGCCCCTTTTCCTCTAACTGAAAGAAGGAAAAAAAAATGGAAC
CCAAAATATTCTACATAGTTTCCATGTCACAGCCAGGGCTGGGCAGTCTCCTGTT
ATTTCTTTTAAAATAAATATATCATTTAAATGCATAAATAAGCAAACCCTGCTCG
GGAATGGGAGGGAGAGTCTCTGGAGTCCACCCCTTCTCGGCCCTGGCTCTGCAGA
TAGTGCTATCAAAGCCCTGACAGAGCCCTGCCCATTGCTGGGCCTTGGAGTGAGT
CAGCCTAGTAGAGAGGCAGGGCAAGCCATCTCATAGCTGCTGAGTGGGAGAGAG
AAAAGGGCTCATTGTCTATAAACTCAGGTCATGGCTATTCTTATTAAAAGAAAAG
GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTT
GTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACT

AGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTG
TGTTGGTTTTTTGTGTGTGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGCACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCT
GGGGATGCGGTGGGCTCTATGGAGTAGTAGTTCATGTCATCTTATTATTCAGTAT
TTATAACTTGCAAAGAAATGAATATCAGAGAGTGAGAGGAACTTGTTTATTGCA
GCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCA
TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCAT
GTCTGGCTCTAGCTATCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCC
AGTTCCGCCCATTCTCCGCCCCATGGCTCTCACTACTTCTGGAATAGCTCAGAGG
CCGAGGCGGCCTCGGCCTCTGCATAAATAAAAAAAATTAGTCGAGGCTTTTTTGG
AGGCCTAGGGACGTACCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCAC
TGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA
TCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGC
ACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGG

SEQ ID NO:16 pUV+HS1-AS3(coX2) from junction marker

ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCC
TTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTT
AGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGT
GATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT
TGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA
CCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATT
GGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATT
AACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT
GTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGA
TAAATGCTTCAATAATAGCACCTAGATCAAGAGACAGGATGAGGATCGTTTCGC
ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGG
CTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGT
TCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGG
TGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGAC
GGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTG
GCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCT
GCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATC
CGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA
CTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGG
GGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCG
AGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAA
TGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTAT
CAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGG
GCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCG
CCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTATTAACGCTTACAATTTCCTG
ATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCA
AATATGTATCCGCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG
CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTG
CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGC
AGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAA
CTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTG

CCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGA
TAAGGCGCAGCGGTCGGGCTGAACGGGGGGGTCGTGCACACAGCCCAGCTTGGA
GCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGC
CACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGG
AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAG
TCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGG
CCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGAC
CGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAC
CGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCC
CGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCA
TTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG
TGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGC
GCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTGCAAGCTTGGCCAT
TGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTCCAACA
TTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGG
GGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACG
TATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGT
ATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC
GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC
ATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT
TACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAC
TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCA
AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT
GAACCGGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAAC
TAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGT
GTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGT
CAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGG
GAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGC
AAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGG
CTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGGAGAATTA
GATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAA
TTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCT
GGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCA
TCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCC
TCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACA
AGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCT
GATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATA
TAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAA
GAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCCT
TGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGAC
GGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTG
AGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAG
CAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTC
CTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGA
ATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGAT
GGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGA
AGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATA

AATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAA
ATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTA
CTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCC
ACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGT
GGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCTCGACGGTAT
CGATCTCGACACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGA
TTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATAC
AAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTA
CAGGGACAGCAGACATGAGGACAGCTAAAACAATAAGTAATGTAAAATACAGC
ATAGCAAAACTTTAACCTCCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGA
TGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGC
CTCACCTTCTTTCATGGAGTTTAAGATATAGTGTATTTTCCCAAGGTTTGAACTAG
CTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCCACATTCCCTTTTTAGTAAA
ATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTA
GGCAGAATCCAGATGCTCAAGGCCCTTCATAATATCCCCCAGTTTAGTAGTTGGA
CTTAGGGAACAAAGGAACCTTTAATAGAAATTGGACAGCAAGAAAGCGAGCTTA
GTGATACTTGTGGGCCAGGGCATTAGCCACACCAGCCACCACTTTCTGATAGGCA
GCCTGCACTGGTGGGGTGAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACC
AGCACGTTGCCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGC
AAAAGGGCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAATA
AAAGCAGAATGGTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACA
TCAGTTACAATTTATATGCAGAAATACCCTGTTACTTCTCCCCTTCCTATGACATG
AACTTAACCATAGAAAAGAAGGGGAAAGAAAACATCAAGGGTCCCATAGACTC
ACCCTGAAGTTCTCAGGATCCACGTGCAGCTTGTCGCAGTGCAGCTCGCTCAGCT
GGGCGAAGGTGCCCTTCAGGTTGTCCAGGTGGGCCAGGCCGTCGCTGAAGGCGC
CCAGCACCTTCTTGCCGTGGGCCTTCACCTTGGGGTTGCCCATCACGGCGTCGGG
GGTGCTCAGGTCGCCGAAGCTCTCGAAGAAGCGCTGGGTCCAGGGGTACACCAC
CAGCAGCCTAAGGGTGGGAAAATAGACCAATAGGCAGAGAGAGTCAGTGCCTAT
CAGAAACCCAAGAGTCTTCTCTGTCTCCACATGCCCAGTTTCTATTGGTCTCCTTA
AACCTGTCTTGTAACCTTGATACCAACCTGCCCAGGGCCTCACCACCAACGGCAT
CCACGTTCACCTTGTCCCACAGGGCAGTAACGGCAGACTTCTCCTCAGGAGTCAG
GTGCACCATGGTGTCTGTTTGAGGTTGCTAGTGAACACAGTTGTGTCAGAAGCAA
ATGTAAGCAATAGATGGCTCTGCCCTGACTTTTATGCCCAGCCCTGGCTCCTGCC
CTCCCTGCTCCTGGGAGTAGATTGGCCAACCCTAGGGTGTGGCTCCACAGGGTGA
GGTCTAAGTGATGACAGCCGTACCTGTCCTTGGCTCTTCTGGCACTGGCTTAGGA
GTTGGACTTCAAACCCTCAGCCCTCCCTCTAAGATATATCTCTTGGCCCCATACCA
TCAGTACAAATTGCTACTAAAAACATCCTCCTTTGCAAGTGTATTTACCATCAAT
AATTCTAGCCCCACAGGAGTTTGTTCTGAAAGTAAACTTCCACAACCGCAAGCTT
ATTGAGGCTAAGGCATCTGTGAAGGAAAGAAACATCTCCTCTAAACCACTATGCT
GCTAGAGCCTCTTTTCTGTACTCAAGCCTCATTCAGACACTAGTGTCACCAGTCTC
CTCATATACCTATTGTATTTTCTTCTTCTTGCTGGTTTAGTCATGTTTTCTGGGAGC
TTAGGGGCTTATTTTATTTTGTTTTGTTTTCTAATCAACAGAGATGGGCAAACCCA
TTATTTTTTTCTTTAGACTTGGGATGGTGATAGCTGGGCAGCGTCAGAAACTGTGT
GTGGATATAGATAAGAGCTCAGGACTATGCTGAGCTGTGATGAGGGAGGGGCCT
AGCTAAAGGCAGTGAGAGTCAGAATGCTCCTGCTATTGCCTTCTCAGTCCCCACG
CTTGGTTTCTACACAAGTAGATACATAGAAAAGGCTATAGGTTAGTGTTTGAGAG
TCCTGCATGATTAGTTGCTCAGAAATGCCCGATAAATATGTTATGTGTGTTTATGT
ATATATATGTTTTATATGTGTGTGTGTGTGTGTTGTGTTTACAAATATGTGATTAT
CATCAAAACGTGAGGGTACGTATATGTGTATATATATATATATATTCAGGAAATA
ATATATTCTAGAATATGTCACATTCTGTCTCAGGCATCCATTTTCTTTATGATGCC
GTTTGAGGTGGAGTTTTAGTCAGGTGGTCAGCTTCTCCTTTTTTTTGCCATCTGCC

CTGTAAGCATCCTGCTGGGGACCCAGATAGGAGTCATCACTCTAGGCTGAGAAC
ATCTGGGCACACACCCTAAGCCTCAGCATGACTCATCATGACTCAGCATTGCTGT
GCTTGAGCCAGAAGGTTTGCTTAGAAGGTTACACAGAACCAGAAGGCGGGGGTG
GGGCACTGACCCCGACAGGGGCCTGGCCAGAACTGCTCATGCTTGGACTATGGG
AGGTCACTAATGGAGACACACAGAAATGTAACAGGAACTAAGGAAAAACTGAA
GCTTTGGGGGTATAGGGGAGCAGTCCCATGTAGTAGTAGAATGAAAAATGCTGC
TATGCTGTGCCTCCCCCACCTTTCCCATGTCTGCCCTCTACTCATGGTCTATCTCTC
CTGGCTCCTGGGAGTCATGGACTCCACCCAGCACCACCAACCTGACCTAACCACC
TATCTGAGCCTGCCAGCCTATAACCCATCTGGGCCCTGATAGCTGGTGGCCAGCC
CTGACCCCACCCCACCCTCCCTGGAACCTCTGATAGACACATCTGGCACACCAGC
TCGCAAAGTCACCGTGAGGGTCTTGTGTTTGCTGAGTCAAAATTCCTTGAAATCC
AAGTCCTTAGAGACTCCCAGGCTTGGATTCAAAGCTCCTGACTTTCTGTCTAGTG
TATGTGCAGTGAGCCCCTTTTCCTCTAACTGAAAGAAGGAAAAAAAAATGGAAC
CCAAAATATTCTACATAGTTTCCATGTCACAGCCAGGGCTGGGCAGTCTCCTGTT
ATTTCTTTTAAAATAAATATATCATTTAAATGCATAAATAAGCAAACCCTGCTCG
GGAATGGGAGGGAGAGTCTCTGGAGTCCACCCCTTCTCGGCCCTGGCTCTGCAGA
TAGTGCTATCAAAGCCCTGACAGAGCCCTGCCCATTGCTGGGCCTTGGAGTGAGT
CAGCCTAGTAGAGAGGCAGGGCAAGCCATCTCATAGCTGCTGAGTGGGAGAGAG
AAAAGGGCTCATTGTCTATAAACTCAGGTCATGGCTATTCTTATTAAAAGAAAAG
GGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTT
GTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACT
AGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTG
TGTTGGTTTTTTGTGTGTGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGCACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCT
GGGGATGCGGTGGGCTCTATGGAGTAGTAGTTCATGTCATCTTATTATTCAGTAT
TTATAACTTGCAAAGAAATGAATATCAGAGAGTGAGAGGAACTTGTTTATTGCA
GCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCA
TTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCAT
GTCTGGCTCTAGCTATCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCC
AGTTCCGCCCATTCTCCGCCCCATGGCTCTCACTACTTCTGGAATAGCTCAGAGG
CCGAGGCGGCCTCGGCCTCTGCATAAATAAAAAAAATTAGTCGAGGCTTTTTTGG
AGGCCTAGGGACGTACCCAATTCGCCCTATAGTGAGTCGTATTACGCGCGCTCAC
TGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA
TCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGC
ACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGG

CLAIMS

What is claimed is:

1. A recombinant lentiviral vector (LV) comprising:

an expression cassette comprising a nucleic acid construct comprising:

a human β-globin locus control region (LCR) comprising an HS1 ENCODE core (EC1) sequence (SEQ ID NO:1), and one or more of an HS2 core sequence (ecHS2 (SEQ ID NO: 2)), an HS3 core sequence (ecHS3 (SEQ ID NO:3)), an HS4 core sequence (ecHS4 (SEQ ID NO:4)), a full length HS2, a full length HS3, and/or a full length HS4.

2. The vector of claim 1, wherein said human β-globin locus control region (LCR) comprises an HS2 core sequence (ecHS2 (SEQ ID NO: 2)).

3. The vector of claim 1, wherein said human β-globin locus control region (LCR) comprises a full length HS2 sequence.

4. The vector according to any one of claims 1-3, wherein said human β-globin locus control region (LCR) comprises an HS3 core sequence (ecHS3 (SEQ ID NO:3)).

5. The vector according to any one of claims 1-3, wherein said human β-globin locus control region (LCR) comprises a full length HS3 sequence.

6. The vector according to any one of claims 1-5, wherein said human β-globin locus control region (LCR) comprises an HS4 core sequence (ecHS4 (SEQ ID NO:4).

7. The vector according to any one of claims 1-5, wherein said human β-globin locus control region (LCR) comprises a full length HS4.

8. The vector of claim 1, wherein said vector comprises the LCR shown in SEQ ID NO:6.

9. The vector according to any one of claims 1-8, wherein said heterologous gene comprises a recombinant human beta globin gene encoding a beta globin polypeptide.

10. The vector of claim 9, wherein said human beta globin gene comprises a wild-type beta globin gene.

11. The vector of claim 9, wherein said human beta globin gene comprises an anti-sickling human beta globin gene encoding an anti-sickling beta globin polypeptide.

12. The vector of claim 11, wherein said anti-sickling human beta globin gene encoding an anti-sickling-beta globin polypeptide comprise one or more mutations selected from the group consisting of Gly16Asp, Glu22Ala and Thr87Gln.

13. The vector of claim 12, wherein said beta globin gene comprises the mutation Gly16Asp.

14. The vector according to any one of claims 12-13, wherein said beta globin gene comprises the mutation Glu22Ala.

15. The vector according to any one of claims 12-14, wherein said beta globin gene comprises the mutation Thr87Gln.

16. The vector of claim 12, wherein said anti-sickling human β-globin gene comprises about 2.3 kb of recombinant human β-globin gene including exons and introns under the control of said human β-globin locus control region.

17. The vector according to any one of claims 12-16, wherein said β-globin gene comprises β-globin intron 2 with a 375 bp RsaI deletion from IVS2.

18. The vector according to any one of claims 12-17, wherein said β-globin gene comprises an SspI (S) to RsaI (R) deletion (~220bp).

19. The vector of claim 11, wherein said anti-sickling human beta globin gene comprises a modified βAS3-globin transgene, said transgene comprising a codon optimized exon 1 (SEQ ID NO:7) (as shown within SEQ ID NO:10), and/or a codon optimized exon 2 (SEQ ID NOs:5 or 8) (as shown within SEQ ID NO:11), and/or comprising a codon optimized exon 3 (SEQ ID NO:9) (as shown within SEQ ID NO:12).

20. The vector of claim 19, wherein said βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO:5).

21. The vector of claim 19, wherein said βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO:8).

22. The vector according to any one of claims 19-21, wherein said βAS3-globin transgene comprises a codon optimized exon 1 (SEQ ID NO:7).

23. The vector according to any one of claims 19-22, wherein said βAS3-globin transgene comprises a codon optimized exon 3 (SEQ ID NO:9).

24. The vector according to any one of claims 1-23, wherein said vector comprises a human Ankyrin insulator element.

25. The vector according to any one of claims 1-24, further comprising an insulator in the 3' LTR.

26. The vector of claim 25, wherein said insulator comprises FB (FII/BEAD-A), a 77 bp insulator element, which contains the minimal CTCF binding site enhancer-blocking components of the chicken β-globin 5' DnaseI-hypersensitive site 4 (5' HS4).

27. The vector according to any one of claims 1-26, wherein said vector comprises a ψ region vector genome packaging signal.

28. The vector according to any one of claims 1-27, wherein the 5' LTR comprises a CMV enhancer/promoter.

29. The vector according to any one of claims 1-28, wherein said vector comprises a Rev Responsive Element (RRE).

30. The vector according to any one of claims 1-29, wherein said vector comprises a central polypurine tract.

31. The vector according to any one of claims 1-30, wherein said vector comprises a post-translational regulatory element.

32. The vector of claim 31, wherein the posttranscriptional regulatory element is modified Woodchuck Post-transcriptional Regulatory Element (WPRE).

33. The vector of claim 1, wherein said vector comprises the nucleic acid sequence of SEQ ID NO:6.

34. The vector according to any one of claims 1-33, wherein said vector is incapable of reconstituting a wild-type lentivirus through recombination.

35. The vector according to any one of claims 1-33, wherein said vector is incapable of reconstituting a wild-type lentivirus through recombination.

36. The vector claim 1, wherein said vector comprises the features of pUV-AS3 (Figure 3).

37. The vector claim 1, wherein said vector comprises the nucleotide sequence of pUV-AS3 (SEQ ID NO: 13).

38. The vector claim 1, wherein said vector comprises the features of pUV-AS3(coX2) (Figure 4).

39. The vector claim 1, wherein said vector comprises the nucleotide sequence of pUV-AS3(coX2) (SEQ ID NO:14).

40. The vector claim 1, wherein said vector comprises the features of pUV-HS1-AS3 (Figure 5).

41. The vector claim 1, wherein said vector comprises the nucleotide sequence of pUV-HS1-AS3 (SEQ ID NO:15).

42. The vector claim 1, wherein said vector comprises the features of pUV+HS1-AS3(coX2) (Figure 6).

43. The vector claim 1, wherein said vectors comprises the nucleotide sequence of pUV+HS1-AS3(coX2) (SEQ ID NO: 16).

44. A host cell transduced with a vector according to any one of claims 1-43.

45. The host cell of claim 44, wherein the cell is a stem cell.

46. The host cell of claim 45, wherein said cell is a stem cell derived from bone marrow, and/or from umbilical cord blood, and/or from peripheral blood.

47. The host cell of claim 44, wherein the cell is a 293T cell.

48. The host cell of claim 44, wherein the cell is a human hematopoietic progenitor cell.

49. The host cell of claim 48, wherein the human hematopoietic progenitor cell is a CD34+ cell.

50. A method of treating a hemoglobinopathy, in a subject, said method comprising:

transducing a stem cell and/or progenitor cell from said subject with a vector according to any one of claims 1-43; and

transplanting said transduced cell or cells derived therefrom into said subject where said cells or derivatives therefrom express said anti-sickling human beta globin gene.

51. The method of claim 50, wherein the cell is a stem cell.

52. The host cell of claim 50, wherein said cell is a stem cell derived from bone marrow.

53. The method of claim 50, wherein the cell is a human hematopoietic stem and progenitor cell.

54. The method of claim 53, wherein the human hematopoietic progenitor cell is a $CD34^+$ cell.

55. The method according to any one of claims 50-54, wherein said hemoglobinopathy is sickle cell disease.

56. The method according to any one of claims 50-54, wherein said hemoglobinopathy is β-thalassemia.

57. A nucleic acid encoding a modified βAS3-globin transgene, said transgene comprising a codon optimized exon 1 (SEQ ID NO:7) (as shown within SEQ ID NO:10), and/or a codon optimized exon 2 (SEQ ID NOs:5 or 8), and/or comprising a codon optimized exon 3 (SEQ ID NO:9) (as shown within SEQ ID NO:12).

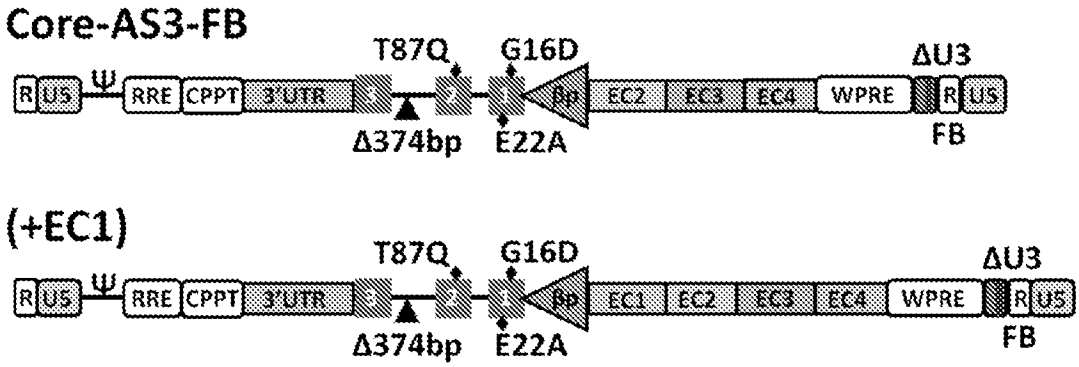
58. The nucleic acid of claim 57, wherein said βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO: 5).

59. The nucleic acid of claim 57, wherein said βAS3-globin transgene comprises a codon optimized exon 2 (SEQ ID NO: 8).

60. The nucleic acid according to any one of claims 57-59, wherein said βAS3-globin transgene comprises a codon optimized exon 1 (SEQ ID NO: 7).

61. The nucleic acid according to any one of claims 57-60, wherein said βAS3-globin transgene comprises a codon optimized exon 3 (SEQ ID NO:9).

(A)


Core-AS3-FB
T87Q
G16D
ΔU3
R
U5
Ψ
RRE
CPPT
3'UTR
βp
EC2
EC3
EC4
WPRE
R
U5
Δ374bp
E22A
FB
(+EC1)
T87Q
G16D
ΔU3
R
U5
Ψ
RRE
CPPT
3'UTR
βp
EC1
EC2
EC3
EC4
WPRE
R
U5
Δ374bp
E22A
FB

(B)

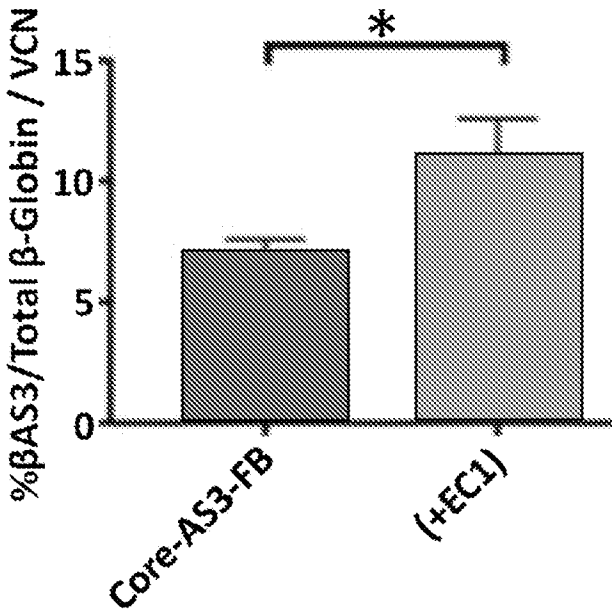
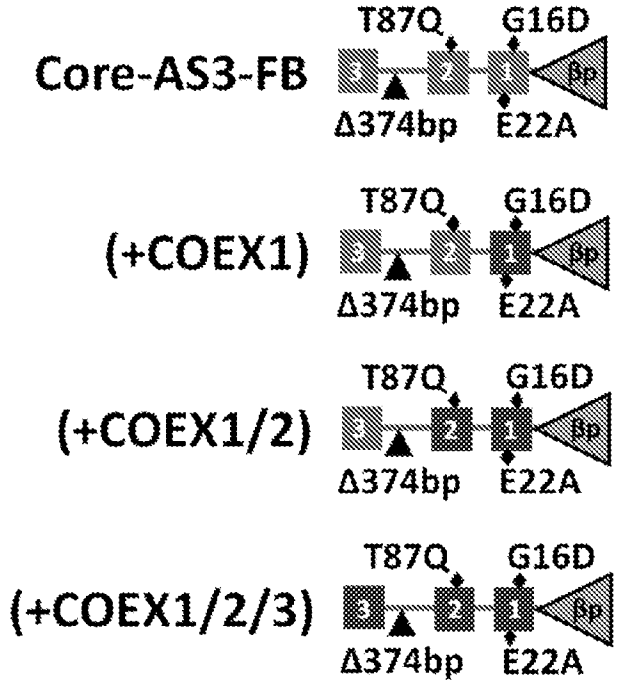

*
%βAS3/Total β-Globin / VCN
15
10
5
0
Core-AS3-FB
(+EC1)

Fig. 1

(A)


Core-AS3-FB
T87Q
G16D
Δ374bp
E22A
βp
(+COEX1)
(+COEX1/2)
(+COEX1/2/3)

(B)

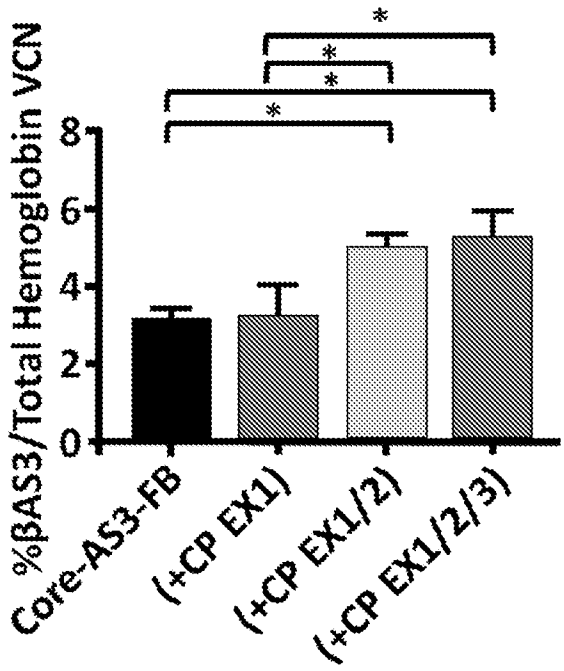

%βAS3/Total Hemoglobin VCN
0
2
4
6
8
*
Core-AS3-FB
(+CP EX1)
(+CP EX1/2)
(+CP EX1/2/3)

Fig. 2

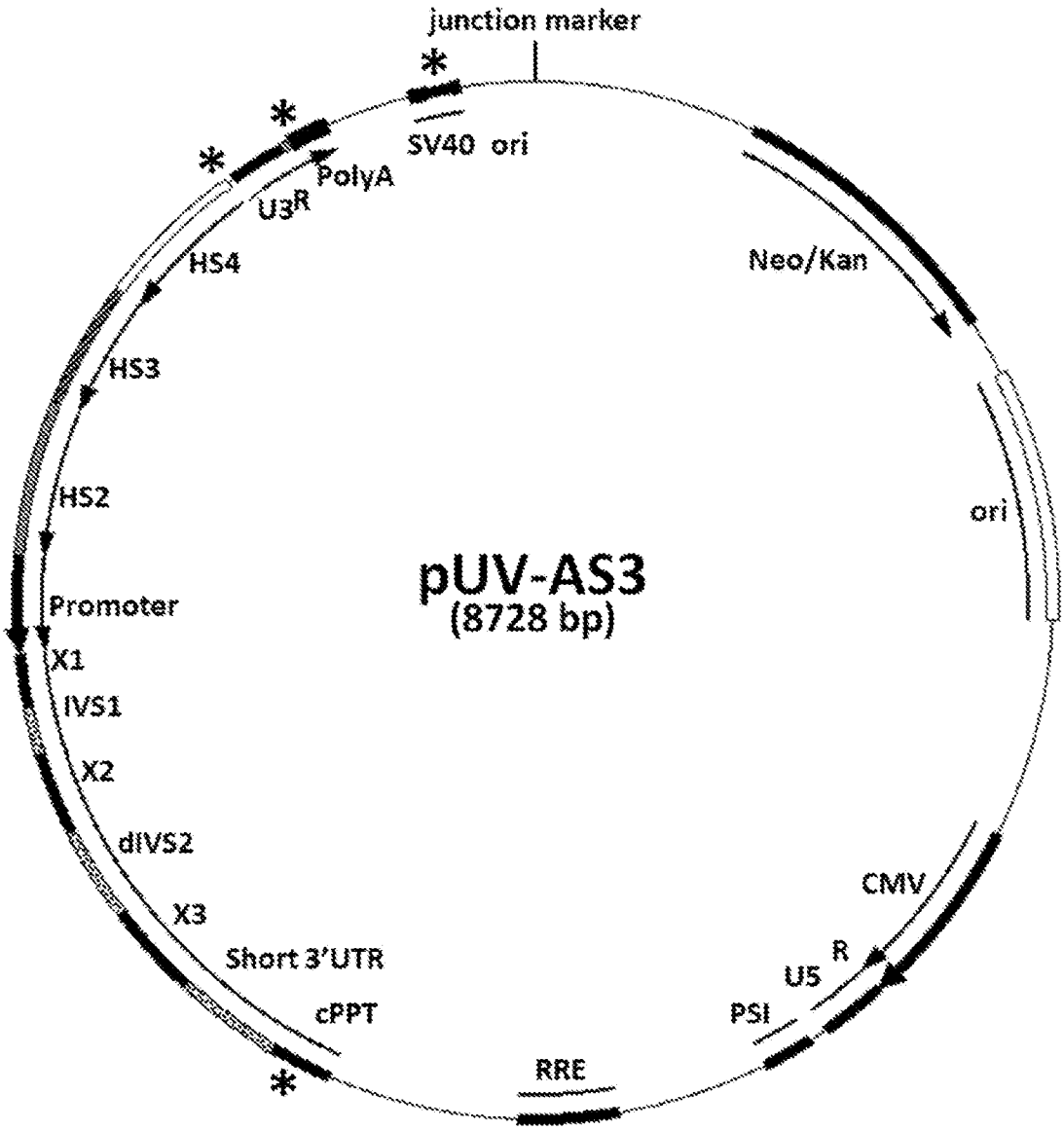

junction marker
SV40 ori
PolyA
U3R
HS4
HS3
HS2
Promoter
X1
IVS1
X2
dIVS2
X3
Short 3'UTR
cPPT
RRE
PSI
U5
R
CMV
ori
Neo/Kan
pUV-AS3
(8728 bp)

Fig. 3

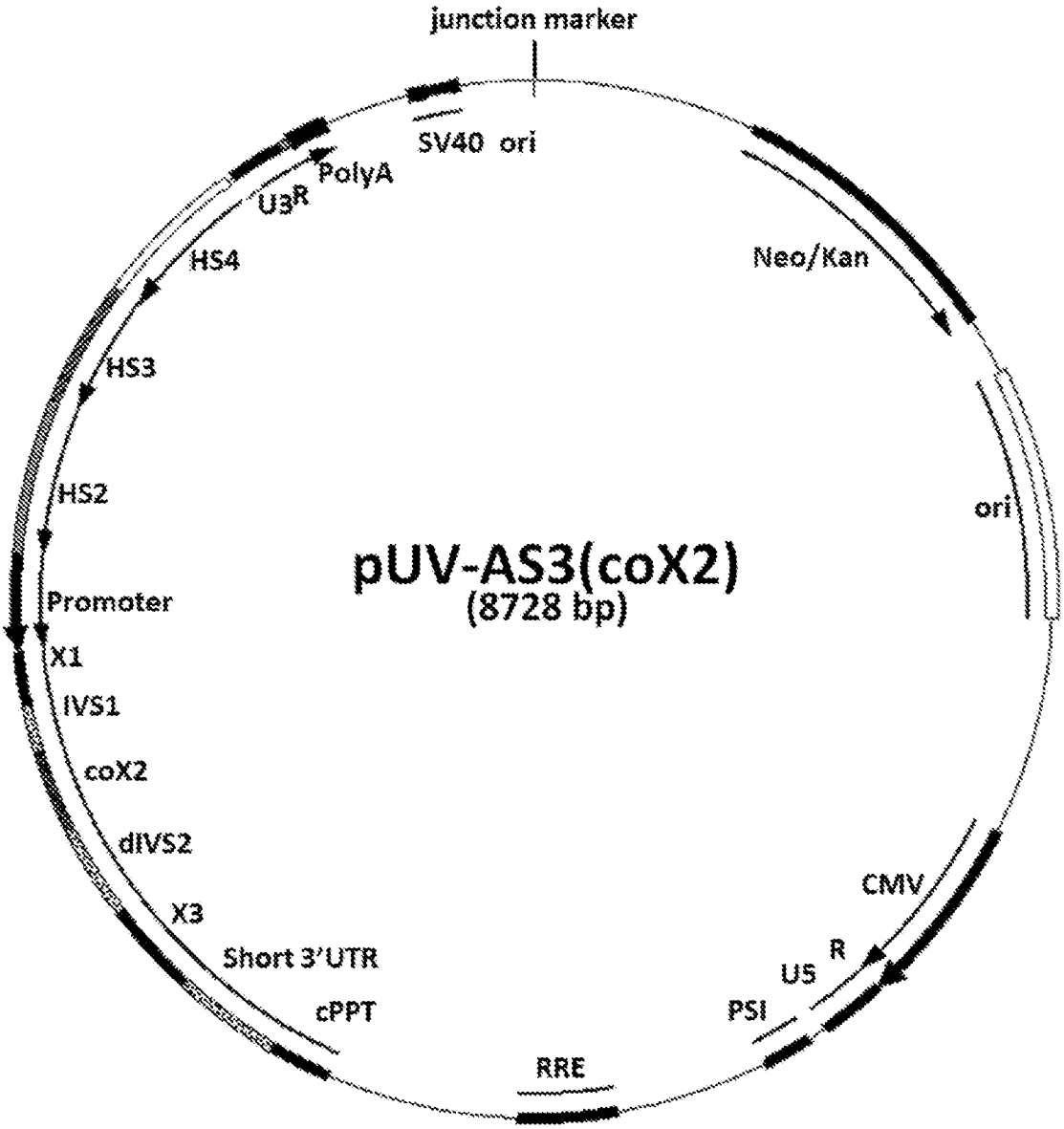

junction marker
SV40 ori
PolyA
U3R
HS4
HS3
HS2
Promoter
X1
IVS1
coX2
dIVS2
X3
Short 3'UTR
cPPT
RRE
PSI
U5
R
CMV
ori
Neo/Kan
pUV-AS3(coX2)
(8728 bp)

Fig. 4

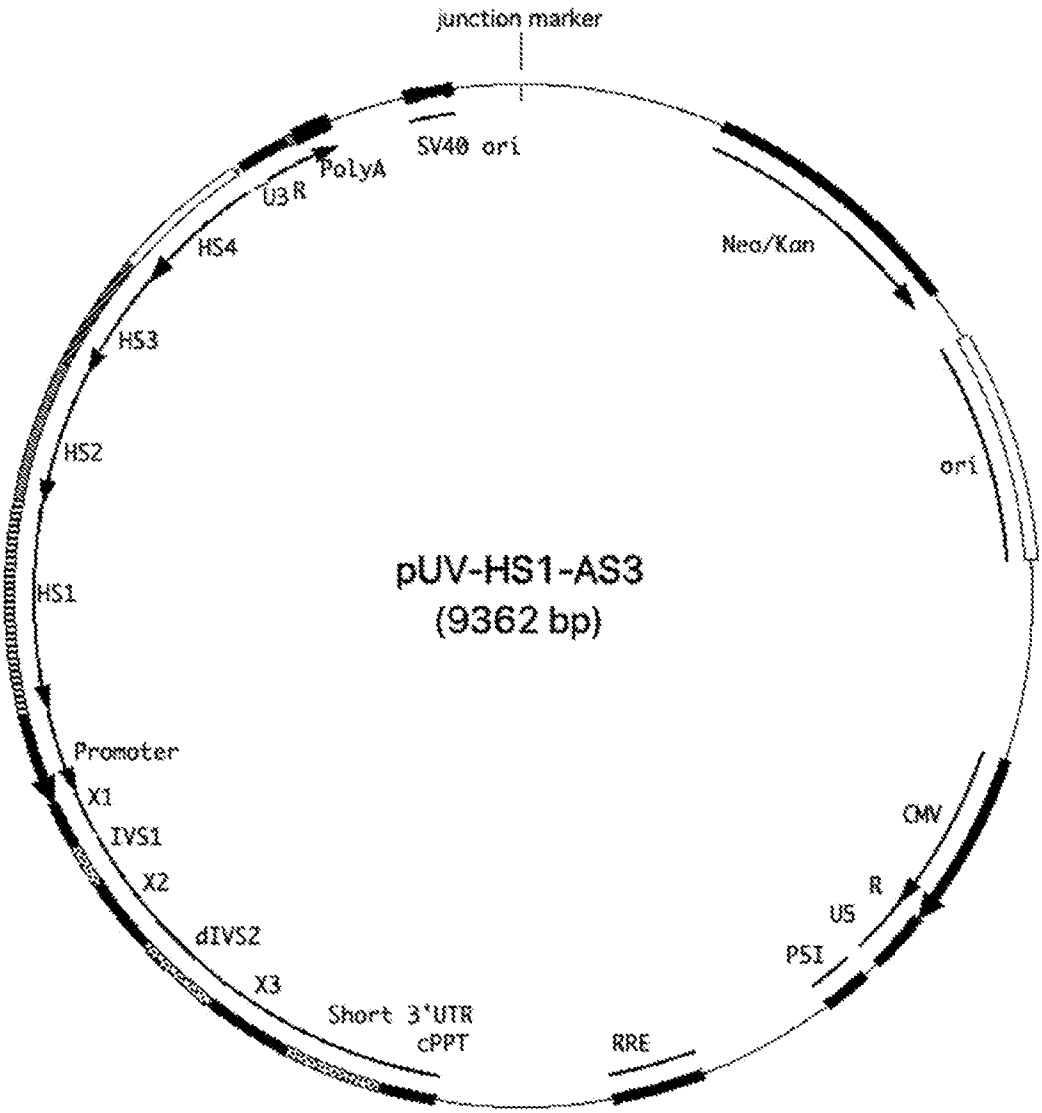

junction marker
SV40 ori
PolyA
U3 R
HS4
HS3
HS2
HS1
Promoter
X1
IVS1
X2
dIVS2
X3
Short 3'UTR
cPPT
RRE
PSI
U5
R
CMV
ori
Neo/Kan
pUV-HS1-AS3
(9362 bp)

Fig. 5

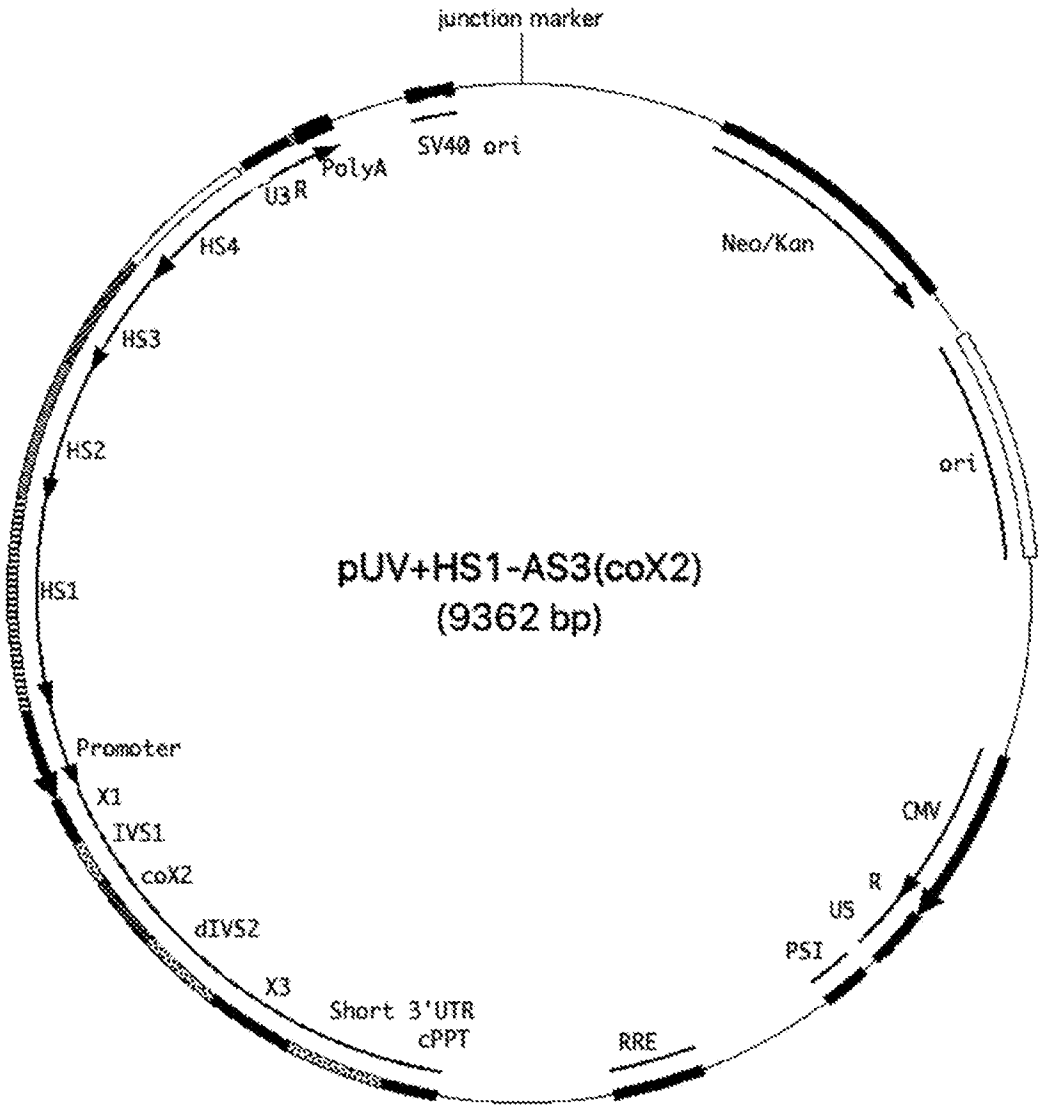

junction marker
SV40 ori
PolyA
U3
R
HS4
HS3
HS2
HS1
Promoter
X1
IVS1
coX2
dIVS2
X3
Short 3'UTR
cPPT
RRE
PSI
US
R
CMV
ori
Neo/Kan
pUV+HS1-AS3(coX2)
(9362 bp)

Fig. 6

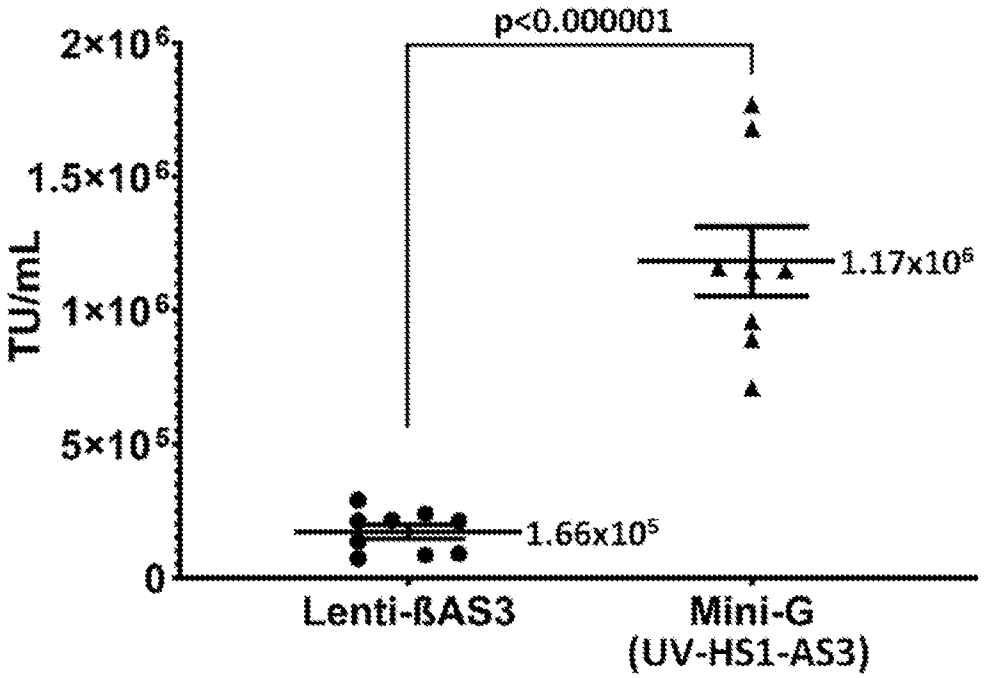

p<0.000001
2×10^6
1.5×10^6
1×10^6
5×10^5
0
TU/mL
1.17x10^6
1.66x10^5
Lenti-ßAS3
Mini-G
(UV-HS1-AS3)

Fig. 7

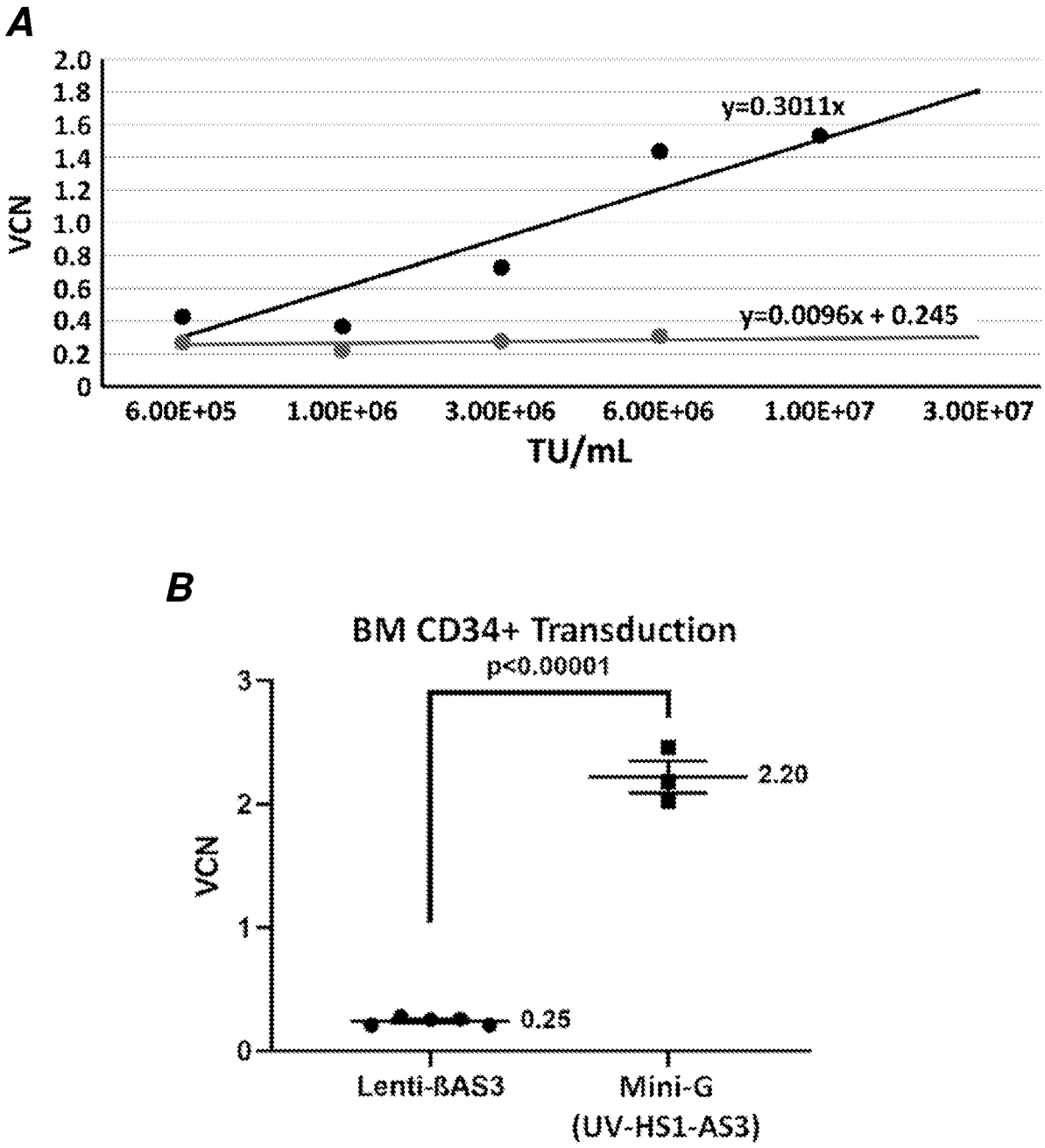

A
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
VCN
y=0.3011x
y=0.0096x + 0.245
6.00E+05
1.00E+06
3.00E+06
6.00E+06
1.00E+07
3.00E+07
TU/mL
B
BM CD34+ Transduction
p<0.00001
3
2
1
0
VCN
2.20
0.25
Lenti-βAS3
Mini-G
(UV-HS1-AS3)

Fig. 8

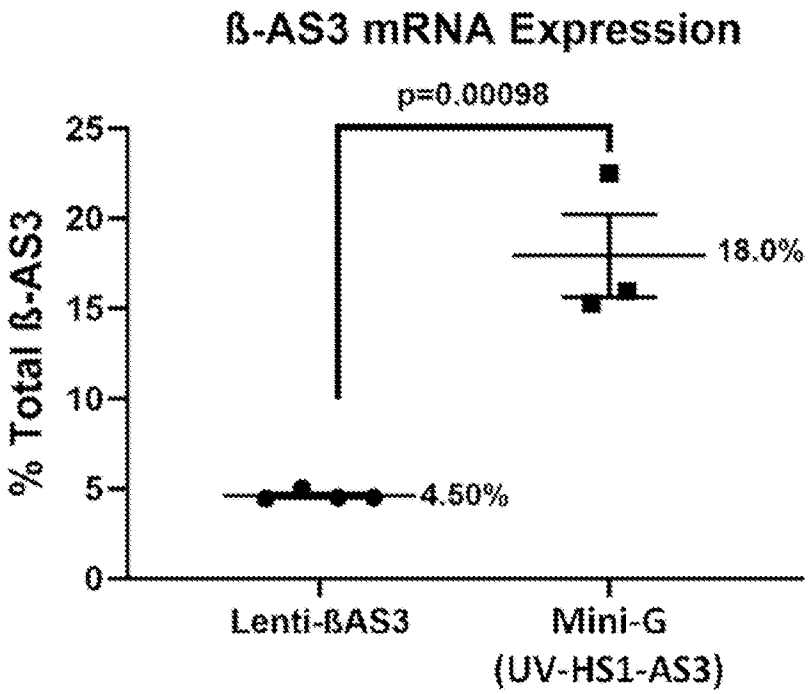

ß-AS3 mRNA Expression
p=0.00098
% Total ß-AS3
25
20
15
10
5
0
18.0%
4.50%
Lenti-ßAS3
Mini-G
(UV-HS1-AS3)

Fig. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/20053

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 15/867, A61P 7/00, A61P 7/06 (2020.01)

CPC - C12N 15/8673, A61K 48/0066, C07K 14/805, C12N 2740/16043, C12N 2830/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2014/043131 A1 (REGENTS OF THE UNIVERSITY OF CALIFORNIA) 20 March 2014 (20.03.2014) Fig. 1	1-2
A	US 2015/0133528 A1 (RANA THERAPEUTICS INC.) 14 May 2015 (14.05.2015) para [0056], SEQ ID NO: 5	1-2

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 June 2020	23 JUL 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/20053

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-7, 9-32, 34-35, 44-56 and 61
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

---------- see extra sheet ----------

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Claims 1 and 2, limited to SEQ ID NOs: 1 and 2

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US 20/20053

Continuation of Box No. III, Observations where unity of invention is lacking:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I+: Claims 1-5, 8, 33, 36-43, directed to a recombinant lentiviral vector (rLV) nucleic acid comprising: an expression cassette comprising a nucleic acid construct comprising a human .beta.-globin locus control region (LCR) comprising an HS 1 ENCODE core (EC1) sequence (SEQ ID NO:1) and an HS2 core sequence (ecHS2 (SEQ ID NO: 2). The rLV will be searched to the extent that the rLV comprises EC1 (SEQ ID NO: 1) and ecHS2 (SEQ ID NO: 2). It is believed that claims 1 and 2, limited to a rLV comprising an LCR further comprising EC1-ecHS2 (SEQ ID NOs: 1 and 2) encompass this first named invention, and thus these claims will be searched without fee to the extent that rLV encompasses said elements. Additional rLV nucleic acids will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected recombinant vectors. Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be a rLV nucleic acid comprising a human .beta.-globin locus control region (LCR) comprising an EC1 sequence (SEQ ID NO:1), and HS2 core sequences ecHS2, and ecHS3 (SEQ ID NOs: 2-3)) (claims 1-2, 4).

Group II+: Claims 57-60, directed to a nucleic acid comprising one to three codon optimized exons encoding a modified PAS3-globin transgene. Group II+ will be searched upon payment of additional fees. The modified .beta.AS3-globin transgene may be searched, for example, to the extent that the transgene encompasses a codon optimized exon 1 (SEQ ID NO:7) (as shown within SEQ ID NO:10). It is believed that claims 57 and 60, limited to a modified PAS3-globin transgene comprising, read on this exemplary invention. Additional .beta.AS3-globin transgene(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be a modified .beta.AS3-globin transgene comprising codon optimized exons 1 (SEQ ID NO: 1), exon 2 (SEQ ID NO: 5), and exon 3 (SEQ ID NO:9) (Claims 57, 58 and 60).

The inventions listed as Groups I+ and II+ do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features

No technical features are shared between the nucleic acid sequences of the rLV nucleic acid sequences of Group I+, accordingly, these groups lack unity a priori.

No technical features are shared between the modified PAS3-globin transgene nucleic acid sequences of Group II+, accordingly, these groups lack unity a priori.

Additionally, even if the inventions listed as Group I+ and Group II+ were considered to share technical features, these shared technical features are previously disclosed by the prior art, as further discussed below.

Groups I+ requires recombinant lentiviral vector (rLV) nucleic acid comprising: an expression cassette comprising a human P-globin locus control region, not required by Group II+.

Group II+ requires a modified PAS3-globin transgene, not required by Group I+.

Common Technical Features

The common technical feature shared by Groups I+ and II+, is a nucleic acid encoding a recombinant globin expression vector.

However, this shared technical feature does not represent a contribution over prior art, because the shared technical feature is made obvious by WO2014043131A1 to The Regents Of The University Of California (hereinafter 'Univ. California'). Univ. California discloses a lentiviral expression vector with control regions and exons from a human globin gene (compare Fig 1 of Univ. of California to Fig. 1 A of the applicant.).

As the technical feature was known in the art at the time of the invention, this cannot be considered a special technical feature that would otherwise unify the inventions.

Groups I+ and II+ therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.