

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2020319168 B2**

(54) Title
Regulatable expression systems

(51) International Patent Classification(s)
C12N 15/85 (2006.01)

(21) Application No: **2020319168**

(22) Date of Filing: **2020.07.24**

(87) WIPO No: **WO21/014428**

(30) Priority Data

(31) Number
62/878,514
63/013,283

(32) Date
2019.07.25
2020.04.21

(33) Country
US
US

(43) Publication Date: **2021.01.28**

(44) Accepted Journal Date: **2024.02.29**

(71) Applicant(s)
Novartis AG

(72) Inventor(s)
BEIBEL, Martin;GUBSER KELLER, Caroline;LUKASHEV, Dmitriy;RENAUD, Nicole;RUDINSKIY, Nikita;SIVASANKARAN, Rajeev

(74) Agent / Attorney
Davies Collison Cave Pty Ltd, Level 15 1 Nicholson Street, MELBOURNE, VIC, 3000, AU

(56) Related Art
WO 2016/034680 A1
WO 2018/009547 A1

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

28 January 2021 (28.01.2021)



(10) International Publication Number

WO 2021/014428 A1

(51) International Patent Classification:

C12N 15/85 (2006.01)

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) International Application Number:

PCT/IB2020/057038

(22) International Filing Date:

24 July 2020 (24.07.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/878,514 25 July 2019 (25.07.2019) US

63/013,283 21 April 2020 (21.04.2020) US

(71) Applicant: **NOVARTIS AG** [CH/CH]; Lichtstrasse 35, 4056 Basel (CH).

(72) Inventors: **BEIBEL, Martin**; Novartis Pharma AG, Postfach, 4002 Basel (CH). **GUBSER KELLER, Caroline**; Novartis Pharma AG, Postfach, 4002 Basel (CH). **LUKASHEV, Dmitriy**; Novartis Institutes for BioMedical Research, Inc., 250 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **RENAUD, Nicole**; Novartis Institutes for BioMedical Research, Inc., 250 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **RUDINSKIY, Nikita**; Novartis Institutes for BioMedical Research, Inc., 250 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **SIVASANKARAN, Rajeev**; Novartis Institutes for BioMedical Research, Inc., 250 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

(74) Agent: **NOVARTIS AG**; Lichtstrasse 35, 4056 Basel (CH).

(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, IT, 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.

(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,

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

— with international search report (Art. 21(3))

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

— in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE

(54) Title: REGULATABLE EXPRESSION SYSTEMS

(57) Abstract: Provided herein are compositions comprising minigenes comprising splice modulator binding sequences, for regulatable gene expression, and systems and methods of use thereof.

WO 2021/014428 A1

REGULATABLE EXPRESSION SYSTEMS

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on July 22, 2020, is named PAT058643-WO-PCT_SL.txt and is 108,491 bytes in size.

FIELD OF THE DISCLOSURE

Disclosed herein are compositions comprising minigenes for regulatable gene expression and systems and methods of use thereof.

BACKGROUND

Gene therapy methods that deliver genetic material (e.g., heterologous nucleic acids) into target cells in order to increase the expression of desired gene products may support this therapeutic objective. Viruses have evolved to become highly efficient at nucleic acid delivery to specific cell types while avoiding immunosurveillance by an infected host. Robbins et al., (1998) *Pharmacol. Ther.*, 80(1):35-47. These properties make viruses attractive as delivery vehicles, or vectors, for gene therapy. Several types of viruses, including retrovirus, adenovirus, adeno-associated virus (AAV), and herpes simplex virus, have been modified in the laboratory for use in gene therapy applications. Lunstrom et al., (2018) *Diseases*, 6(2): 42. In particular, vectors derived from Adeno-Associated Viruses (AAVs) may effectively deliver genetic material because (i) they are able to infect (transduce) a wide variety of non-dividing and dividing cell types including muscle fibers and neurons; (ii) they are devoid of the virus structural genes, thereby eliminating the natural host cell responses to virus infection, e.g., interferon-mediated responses; (iii) wild-type viruses have never been associated with any pathology in humans; (iv) in contrast to wild type AAVs, which are capable of integrating into the host cell genome, replication-deficient AAV vectors generally persist as episomes, thus limiting the risk of insertional mutagenesis or activation of oncogenes; and (v) in contrast to other vector systems, AAV vectors do not trigger a significant immune response (see ii), thus granting long-term expression of, e.g., therapeutic heterologous nucleic acid(s) (provided their gene products are not rejected). Wold et al., (2013) *Curr. Gene Ther.*, 13(6):421-33; Lee et al., (2017) *Genes Dis.*, 4(2): 43–63.

AAV is a member of the parvoviridae family. The AAV genome comprises a linear single-stranded DNA molecule which typically contains approximately 4.7 kilobases (kb) and two major open reading frames encoding the non-structural Rep (replication) and structural Cap (capsid) proteins. Flanking the AAV coding regions are two cis-acting inverted terminal repeat (ITR) sequences, which are typically approximately 145 nucleotides in length and have interrupted palindromic sequences that can fold into hairpin structures that function as primers during initiation of DNA replication. In addition to their role in

DNA replication, the ITR sequences have been shown to contribute to viral integration, rescue from the host genome, and encapsidation of viral nucleic acid into mature virions. Muzyczka et al., (1992) *Curr. Top. Micro. Immunol.*, 158:97-129.

Many proteins have been developed which are important scientific research tools or medications for preventing or treating diseases. While viral vectors such as AAVs are desirable for their ability to transduce a variety of cell types and deliver the heterologous nucleic acids encoding these proteins to a variety of target tissue types, side effects can occur upon expression of the proteins, varying from, for example, a loss of drug efficacy to serious toxicities. It is desirable to develop strategies to modulate the expression level of the therapeutic proteins, e.g., to modulate the timing or location of expression of therapeutic proteins and/or levels of the therapeutic proteins to increase efficacy and/or decrease side effects.

Accordingly, the present disclosure provides, in part, minigene nucleotide sequences that are useful to control expression of proteins using a small-molecule to turn off or turn on expression of the protein of interest. The disclosure also provides vectors, recombinant viruses and pharmaceutical compositions comprising such minigene sequences, and contemplates their use in methods regulating gene expression.

SUMMARY

In a first aspect, provided is a nucleic acid molecule including a minigene linked to a transgene encoding a protein of interest, wherein the minigene includes: A first exon; A first intron; A second exon; A second intron; and A third exon; wherein said second exon includes a splice modulator binding sequence and wherein, in the presence of a splice modulator, said second exon is included in an mRNA product of the nucleic acid, and in the absence of said splice modulator, said second exon is not included in an mRNA product of the nucleic acid.

In embodiments, the third exon includes a stop codon that is in frame in the mRNA product of the nucleic acid produced in the absence of the splice modulator and which is not in frame in the mRNA product of the nucleic acid produced in the presence of the splice modulator.

In embodiments, the second exon includes a stop codon that is in frame in the mRNA product of the nucleic acid produced in the presence of the splice modulator.

In embodiments, the first exon and the third exon do not comprise a start codon. In some embodiments, the second exon comprises a start codon.

In embodiments of any of the aforementioned aspects and embodiments, the nucleic acid includes a sequence encoding a protease cleavage site disposed between the minigene and the transgene.

In embodiments said protease cleavage site is cleaved by a mammalian protease.

In embodiments the mammalian protease is furin, PCSK1, PCSK5, PCSK6, PCSK7, cathepsin B, Granzyme B, Factor XA, Enterokinase, genenase, sortase, precision protease, thrombin, TEV protease, or elastase 1.

In embodiments of any of the aforementioned aspects and embodiments, the protease cleavage site includes a polypeptide having an cleavage motif selected from the group consisting of RX(K/R)R consensus motif, RXXX[KR]R consensus motif, RRX consensus motif, RNRR (SEQ ID NO: 39), I-E-P-D-X consensus motif (SEQ ID NO: 35), Glu/Asp-Gly-Arg, Asp-Asp-Asp-Asp-Lys (SEQ ID NO: 36), Pro-Gly-Ala-Ala-His-Tyr (SEQ ID NO: 37), LPXTG/A consensus motif, Leu-Glu-Val-Phe-Gln-Gly-Pro (SEQ ID NO: 38), Leu-Val-Pro-Arg-Gly-Ser (SEQ ID NO: 40), E-N-L-Y-F-Q-G (SEQ ID NO: 41), and [AGSV]-x (SEQ ID NO: 42). In embodiments said cleavage site is cleaved by furin. In embodiments, the protease cleavage site cleaved by furin is RNRR (SEQ ID NO: 39); RTKR (SEQ ID NO: 43); GTGAEDPRPSRKRRSLGDVG (SEQ ID NO: 45); GTGAEDPRPSRKRR (SEQ ID NO: 47); LQWLEQQVAKRRTRK (SEQ ID NO: 49); GTGAEDPRPSRKRRSLGG (SEQ ID NO: 51); GTGAEDPRPSRKRRSLG (SEQ ID NO: 53); SLNLTESHNSRKRR (SEQ ID NO: 55); or CKINGYPKRGRKRR (SEQ ID NO: 57). In embodiments the protease cleavage site cleaved by furin includes RNRR (SEQ ID NO: 39). In embodiments the sequence encoding the protease cleave site includes, e.g., consists of, CGCAACCGCCGC (SEQ ID NO: 19).

In embodiments including in any of the aforementioned aspects and embodiments the nucleic acid includes a sequence encoding a self-cleaving peptide disposed between the minigene and the transgene, optionally wherein the self-cleaving peptide cleaves within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids of the N-terminus of the protein of interest. In embodiments, the self-cleaving peptide is a 2A peptide, optionally selected from a T2A peptide, a P2A peptide, a E2A peptide and a F2A peptide, e.g., includes a T2A peptide, e.g., wherein the self-cleaving peptide includes EGRGSLLTCDVEENPGP (SEQ ID NO: 61), optionally wherein the self-cleaving peptide includes (GSG)EGRGSLLTCDVEENPGP (SEQ ID NO: 59).

In embodiments including in any of the aforementioned aspects and embodiments the splice modulator binding sequence is located at the 3' terminus of the second exon.

In embodiments including in any of the aforementioned aspects and embodiments the splice modulator binding sequence includes, e.g., consists of, AGA and the splice modulator is 5-(1H-Pyrazol-4-yl)-2-(6-((2,2,6,6-tetramethylpiperidin-4-yl)oxy)pyridazin-3-yl)phenol (LMI070).

In embodiments including in any of the aforementioned aspects and embodiments the second exon includes, e.g., consists of a sequence selected from:

CCTTGCTATCCCTGTCTTCTGTAGCTATTCTGAAACCATCAACAAAGGAGCACACCATTCCATCAGCA
AAAGA (SEQ ID NO: 1);

GTAATTAGCTGAGAAGGAAGATCTGAAGGTTTAACGAGAGAGGGCGAGAGATACAAAATATCTGCTA
GGAGA (SEQ ID NO: 2);

GGATTGTTTGTATTCTGCCAATGATTTGTGAGACAGTCTGTTCCCCACATCCTCGTCAACAGA (SEQ
ID NO: 3);

CTTTCTGACATCTTAACGAGGCAATACAGAGAGACGAATTTTCATCAGTTTGTTTCAGGGAGACACATA
TAACAAAAGA (SEQ ID NO: 4);

ATCCATACATACTTAATGCTGAAATGTGAAGGGCTGAGAAAAAAGAAAAGA (SEQ ID NO: 5);

AATTGGAAACATCGAGGGAAAATGGGCTTTTTATTATTAACAAAACCTCAGTATTATCACTTAGAAA
CCTGAAATTGAACTCCAAAAGCCAAAGA (SEQ ID NO: 6);

AAGAATGTTCTTTTGTGAAGAATGACTTAAGGAAGATTCATGATGACTGAGTGTGCCCGTGTGGAA
CTTTAGGACATAGATGCACTCCTACAGA (SEQ ID NO: 7);

TTGTCCTTCACTCCGTACTCCAGTTGGCCAAGCATAGGTCGCATGCCAGGGTCAAGGAGACTAAGG
GAGA (SEQ ID NO: 8);

GACATACAGACATGGCAGCCCCTAGCATGTGTATCCTAAGA (SEQ ID NO: 9);

ACATACAGACATGGCAGCCCCTAGCATGTGTATCCTAAGA (SEQ ID NO: 10);

AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGAATGAGCAGAAAATCATTTTCAGGGCCTGTTCTCT
ATGTCCTTGCTATCCCTGTCTTCTGTAGCTATTCTGAAACCATCAACAAAGGAGCACACCATTCCATC
AGCAAAAAGA (SEQ ID NO: 80) and

A fragment or mutant of any of (a) to (k) having at least 90%, at least 95% at least 96%, at least 97%, at
least 98% or at least 99% identity thereto.

In embodiments including in any of the aforementioned aspects and embodiments the second exon
includes a sequence derived from an exon of SNX7, optionally wherein the sequence is derived a cryptic
exon of SNX7.

In embodiments including in any of the aforementioned aspects and embodiments the second exon
includes, e.g., consists of,

AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGATTGAGCAGAAAATCATTTTCAGGGCCTGTTCTCT
ATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCTGAAACCATCAACAAAGGAGCACACCATTCCATCA
GCAAAAGA (SEQ ID NO: 16);

a fragment of SEQ ID NO: 16; or

a mutant sequence of SEQ ID NO: 16 or a fragment thereof having at least 90%, at least 95% at least 96%, at least 97%, at least 98% or at least 99% identity thereto.

In embodiments including in any of the aforementioned aspects and embodiments the second exon includes, e.g., consists of,

AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGATTGAGCAGAAAATCATTTTCAGGGCCTGTTCTCT
ATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCTGAAACCATCAACAAAGGAGCACACC**ATGG**CATCA
GCAAAAGA (SEQ ID NO: 98);

a fragment of SEQ ID NO: 98; or

a mutant sequence of SEQ ID NO: 98 or a fragment thereof having at least 90%, at least 95% at least 96%, at least 97%, at least 98% or at least 99% identity thereto.

In embodiments including in any of the aforementioned aspects and embodiments the second exon consists of $3n-1$ nucleotides, where n is an integer.

In embodiments including in any of the aforementioned aspects and embodiments the first exon includes:

One or more, e.g., three, GAA repeats (SEQ ID NO: 69) (for example, includes GAAGAAGAA (SEQ ID NO: 69));

A Kozak sequence (e.g., a Kozak sequence including GCCACC (SEQ ID NO: 70)); or

Both (a) and (b).

In embodiments including in any of the aforementioned aspects and embodiments the first exon includes, e.g., consists of,

GAAGAAGAAGATATCAAGTTAGCATTTACAGATTTGGCTGAGGAGAAGAACAG (SEQ ID NO: 96);

a fragment of SEQ ID NO: 96; or

a mutant sequence of SEQ ID NO: 96 or a fragment thereof having at least 90%, at least 95% at least 96%, at least 97%, at least 98% or at least 99% identity thereto.

In embodiments including in any of the aforementioned aspects and embodiments the first intron includes, e.g., consists of,

GTAATTAGTGTTGTTTGATATTGCTTCATTTTAAAGTTATTTGCTCATTAGCATTTGATATTGCTTTCT
ATTGATTGTCCTAACTACTCCTCTTTCTCCTCTCCCTTCTCCATTTTGAAG (SEQ ID NO: 97);

a fragment of SEQ ID NO: 97; or

a mutant sequence of SEQ ID NO: 97 or a fragment thereof having at least 90%, at least 95% at least 96%, at least 97%, at least 98% or at least 99% identity thereto.

In embodiments including in any of the aforementioned aspects and embodiments the minigene has been modified to:

Remove or mutate all but a single start codon, e.g., an ATG start codon;

Remove or mutate all cryptic splice donor and splice acceptor sequences other than those at the termini of the first exon, the second exon and the third exon.

In embodiments, the minigene has a single start codon disposed within the first exon. In embodiments, the minigene has a single start codon disposed within the second exon.

In embodiments including in any of the aforementioned aspects and embodiments the minigene includes fewer than 2000, fewer than 1900, fewer than 1800, fewer than 1700, fewer than 1600, fewer than 1500, fewer than 1400, fewer than 1300, fewer than 1200, fewer than 1100, or fewer than 1000, fewer than 900, fewer than 800, fewer than 700, fewer than 600, fewer than 500 nucleotides.

In embodiments including in any of the aforementioned aspects and embodiments the minigene includes between about 2500 and about 500 nucleotides, e.g., between about 2000 and about 600 nucleotides, e.g., between about 1500 and about 700 nucleotides, e.g., between about 1200 and about 800 nucleotides, between about 1100 and about 900 nucleotides, between about 800 and about 500 nucleotides, between about 800 and about 600 nucleotides.

In embodiments including in any of the aforementioned aspects and embodiments the minigene includes, e.g., consists of, SEQ ID NO: 71 or SEQ ID NO: 94, or a sequence with at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identity thereto, or a functional fragment thereof.

In an aspect, disclosed herein is a nucleic acid molecule, including (a) a transgene encoding a protein of interest, and (b) a minigene including, e.g., consisting of, SEQ ID NO: 71 or SEQ ID NO: 94, or a sequence with at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identity thereto, or a functional fragment thereof.

In embodiments including in any of the aforementioned aspects and embodiments, the nucleic acid molecule further includes a sequence encoding a furin cleavage site, said sequence including SEQ ID NO: 19, and a sequence encoding a self-cleaving peptide, said sequence including SEQ ID NO: 20, optionally wherein the minigene is disposed 5' to the sequence encoding the furin cleavage site (e.g., immediately 5' to the sequence encoding the furin cleavage site), the sequence encoding the furin cleavage site is disposed 5' to the sequence encoding the self-cleaving peptide (e.g., immediately 5' to the sequence encoding the self-cleaving peptide), and the sequence encoding the self-cleaving peptide is disposed 5' to the transgene (e.g., immediately 5' to the transgene).

In embodiments including in any of the aforementioned aspects and embodiments, the nucleic acid molecule further including a promoter operably linked to the minigene and transgene, optionally wherein said promoter is disposed 5' to the minigene.

In embodiments including in any of the aforementioned aspects and embodiments the promoter is a JeT promoter, a CBA promoter, a PGK promoter, or a synapsin promoter, or any promoter that does not include an intron.

In embodiments including in any of the aforementioned aspects and embodiments, the nucleic acid molecule further includes a post-transcriptional regulatory element.

In embodiments including in any of the aforementioned aspects and embodiments the post-transcriptional regulatory element (PRE) includes a PRE derived from hepatitis B (HPRE), bat (BPRE), ground squirrel (GSPRE), arctic squirrel (ASPRE), duck (DPRE), chimpanzee (CPRE) and woolly monkey (WMPRE) or woodchuck (WPRE), optionally wherein said post-transcriptional regulatory element is disposed 3' to the transgene.

In embodiments including in any of the aforementioned aspects and embodiments the post-transcriptional regulatory element includes SEQ ID NO: 72, SEQ ID NO: 73 or SEQ ID NO:88.

In embodiments including in any of the aforementioned aspects and embodiments, the nucleic acid molecule further includes a polyadenylation signal (polyA), optionally wherein said polyA is disposed 3' to the transgene.

In embodiments including in any of the aforementioned aspects and embodiments the poly A signal is an SV40 polyA, human growth hormone (HGH) polyA, or bovine growth hormone (BGH) polyA, a beta-globin polyA, an alpha-globin polyA, an ovalbumin polyA, a kappa-light chain polyA, and a synthetic polyA.

In embodiments including in any of the aforementioned aspects and embodiments the polyA includes, e.g., consists of, SEQ ID NO: 22.

In another aspect, disclosed herein is a vector including a nucleic acid according to any one of the previous aspects and embodiments. In embodiments, the vector is a DNA vector, optionally a circular vector, optionally a plasmid. In embodiments, the vector is double stranded or single stranded, e.g., is double stranded.

In embodiments, the vector is a viral vector. In embodiments, the viral vector is an adeno-associated viral (AAV) vector, chimeric AAV vector, adenoviral vector, retroviral vector, lentiviral vector, DNA viral vector, herpes simplex viral vector, baculoviral vector, or any mutant or derivative thereof. In embodiments, the viral vector is a recombinant AAV vector, optionally a self-complementary AAV (scAAV) vector. In embodiments, the viral vector is a recombinant AAV vector, optionally a single-stranded AAV (ssAAV) vector. In embodiments, the recombinant AAV vector includes one or more inverted terminal repeats (ITRs), optionally wherein the ITRs are AAV2 ITRs, optionally wherein the AAV vector includes two ITRs, optionally wherein the two ITRs include SEQ ID NO: 12 and SEQ ID NO: 23.

In embodiments, including in any of the previous aspects and embodiments, the vector includes, e.g. from 5' to 3':

an ITR, optionally an AAV2 ITR, optionally, wherein the ITR has been modified to include a deletion of a terminal resolution site, optionally including SEQ ID NO: 12;

a promoter, optionally a JeT promoter including or consisting of SEQ ID NO: 13;

a nucleic acid molecule of any one of aspects 1-28;

a polyA signal, optionally including or consisting of SEQ ID NO: 22; and

an ITR, optionally an AAV2 ITR, optionally including or consisting of SEQ ID NO: 23.

In an aspect, provided herein is a recombinant virus including the nucleic acid or vector of any of the previous aspects and embodiments. In embodiments, the recombinant virus is an adeno-associated virus (AAV), chimeric AAV, adenovirus, retrovirus, lentivirus, DNA virus, herpes simplex virus, baculovirus, or any mutant or derivative thereof. In embodiments, the virus is an AAV. In embodiments, the AAV includes one or more of an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV 8, AAV9, AAV10, and AAV11, AAV12, AAVrh8, AAVrh10, AAVrh36, AAVrh37, AAV-DJ, AAV-DJ/8, AAV.Anc80, AAV.Anc80L65, AAV-PHP.B, AAV-PHP.B2, AAV-PHP.B3, AAV-PHP.A, AAV-PHP.eB, and AAV-PHP.S capsid serotype, or a variant thereof, e.g., a combination of capsids from more than one AAV serotype. In embodiments, the AAV includes an AAV9 capsid serotype or any mutant or derivative thereof. In embodiments, the virus includes AAV9 capsid proteins VP1, VP2, and VP3, e.g., as encoded by SEQ ID NO: 74, SEQ ID NO: 75, and SEQ ID NO: 76, respectively, or including an amino acid sequence of SEQ ID NO: 77, SEQ ID NO:

78, SEQ and ID NO: 79, respectively. In embodiments, the AAV includes a self-complementary AAV (scAAV) vector. In embodiments, the AAV includes a single-stranded AAV (ssAAV) vector.

In another aspect, provided herein is a cell including the nucleic acid molecule, the vector, or the recombinant virus of any one the previous aspects and embodiments. In embodiments, the cell is a human cell. In embodiments, the cell is a neuron or astrocyte.

In an aspect, provided herein is a cell, including a cell of any previous cell aspect and embodiments, wherein when the cell includes a splice modulator, e.g., LMI070, the level of expression of the protein of interest is greater, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, than the level of expression of the protein of interest when the cell does not include said splice modulator, optionally wherein the level of expression when the cell does not include said splice modulator is undetectable.

In an aspect, provided herein is a cell, including a cell of any previous cell aspect and embodiments, wherein when the cell does not include a splice modulator, e.g., LMI070, the level of expression of the protein of interest is greater, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, than the level of expression of the protein of interest when the cell includes said splice modulator, optionally wherein the level of expression when the cell includes said splice modulator is undetectable.

In an aspect, provided herein is a method of conditionally expressing a protein of interest, said method including: contacting an expression system (e.g. a cell, e.g., a cell of any one of the previous aspects and embodiments) including the nucleic acid molecule, the vector, or the recombinant virus of any previous aspect and embodiment, with a splice modulator, e.g., LMI070, wherein:

in the presence of said splice modulator, expression of said protein of interest is increased, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, relative to the level of expression of said protein of interest in the absence of said splice modulator; and

in the absence of said splice modulator, expression of said protein of interest is substantially decreased, e.g., e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold less, relative to the level of expression of said protein of interest in the presence of the splice modulator.

In an aspect, provided herein is a method of conditionally expressing a protein of interest, said method including: contacting an expression system (e.g. a cell, e.g., a cell of any one of the previous aspects and embodiments) including the nucleic acid molecule, the vector, or the recombinant virus of any previous aspect and embodiment, with a splice modulator, e.g., LMI070, wherein:

in the absence of said splice modulator, expression of said protein of interest is increased, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, relative to the level of expression of said protein of interest in the presence of said splice modulator; and

in the presence of said splice modulator, expression of said protein of interest is substantially decreased, e.g., e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold less, relative to the level of expression of said protein of interest in the absence of the splice modulator.

In an aspect, provided herein is a pharmaceutical composition including a nucleic acid molecule, a vector, a recombinant virus, or a cell of any of the previous aspects and embodiments.

In an aspect, provided herein is a method of treating a subject in need of a gene therapy, said method including administering to said subject a nucleic acid molecule, a vector, a recombinant virus, a cell or a pharmaceutical composition of any of the previous aspects and embodiments. In embodiments, the method further includes administering to the subject an amount of a splice modulator, e.g., LMI070, effective to cause at least a 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold increase or decrease in expression of the protein of interest, relative to the expression level of the protein of interest in the absence of the splice modulator.

In an aspect, provided herein is a kit including a nucleic acid molecule, a vector, a recombinant virus, a cell or a pharmaceutical composition of any of the previous aspects and embodiments; and a splice modulator.

In an aspect, provided herein is a nucleic acid molecule, a vector, a recombinant virus, a cell or a pharmaceutical composition of any of the previous aspects and embodiments, for use in a method of conditionally expressing a protein of interest, said method including: contacting an expression system (e.g. a cell, e.g., a cell of any one of aspects 53-57) including the nucleic acid molecule of any one of aspects 1-2 and 4-36, the vector of any one of aspects 37-45 or the recombinant virus of any one of aspects 46-52, with a splice modulator, e.g., LMI070, wherein:

in the presence of said splice modulator, expression of said protein of interest is increased, e.g., is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, relative to the level of expression of said protein of interest in the absence of said splice modulator; and

in the absence of said splice modulator, expression of said protein of interest is substantially decreased, e.g., is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold less, relative to the level of expression of said protein of interest in the presence of the splice modulator.

In an aspect, provided herein is a nucleic acid molecule, a vector, a recombinant virus, a cell or a pharmaceutical composition of any of the previous aspects and embodiments, for use in a method of conditionally expressing a protein of interest, said method including: contacting an expression system (e.g. a cell, e.g., a cell of any one of aspects 53-57) including the nucleic acid molecule of any one of aspects 1 or 3-36, the vector of any one of aspects 37-45 or the recombinant virus of any one of aspects 46-52, with a splice modulator, e.g., LMI070, wherein:

in the absence of said splice modulator, expression of said protein of interest is increased, e.g., is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, relative to the level of expression of said protein of interest in the presence of said splice modulator; and

in the presence of said splice modulator, expression of said protein of interest is substantially decreased, e.g., is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold less, relative to the level of expression of said protein of interest in the absence of the splice modulator.

In an aspect, provided herein is a nucleic acid molecule, a vector, a recombinant virus, a cell or a pharmaceutical composition of any of the previous aspects and embodiments, for use in a method of treating a subject in need of a gene therapy.

In an aspect, provided herein is a nucleic acid molecule, a vector, a recombinant virus, a cell or a pharmaceutical composition of any of the previous aspects and embodiments, or the nucleic acid, vector, recombinant virus, cell, or pharmaceutical composition for use according to any one of aspects 64-66, wherein the transgene encodes a protein of a genome editing system (for example, an RNA-guided nuclease such as a Cas9 protein, a zinc finger nuclease or a TALEN), an antibody or antibody fragment, or a therapeutic protein (for example, protein selected from progranulin, SMN, MeCP2, CLN2, CLN3, CLN4, CLN5, CLN6, CLN7, CLN8).

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1A. Describes the concept of a splice modulator-mediated “ON-switch”. In an ON-switch system, exon C contains a premature termination (stop) codon that is in frame with the coding sequence initiated by the start codon located in exon A when exon B is excluded. When a splice modulator such as LMI070 is included the transcript now includes frame-shifting exon B, thereby restoring an uninterrupted open reading frame which leads to transgene expression.

FIG. 1B. Describes the concept of a splice modulator-mediated “OFF-switch”. In an OFF-switch system, exon A is spliced to exon C, which leads to transgene expression. When a splice modulator such as LMI070 is present, exon B, which contains a premature termination (stop) codon, is included, resulting in termination of translation.

Figure 2A. Design of AAV vector with SNX7 minigene-based switch. Fig. 2A shows a schematic diagram of the SNX7 locus containing a splice modulator (LMI070) exonic target binding site at chromosome: GRCh37:1:99204216:99204359:1
(AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGAATGAGCAGAAAATCATTTTCAGGGCCTGTTCTC
TATGTCCTTGCTATCCCTGTCTTCTGTAGCTATTCTGAAACCATCAACAAAGGAGCACACCATTCCAT

CAGCAAAAGA (SEQ ID NO: 80)), as well as an intronic sequence downstream of exon 8 at chromosome:GRCh37:1:99203793:99203946:1
 (CTTCCAGAGGAGATTGGAAACTTGAAGATAAAGTGAATGTGCTAATAATGCCCTGAAAGCAGATTGGGAGAGATGGAAACAAAATATGCAAAATGATATCAAGTTAGCATTTACAGATATGGCTGAGGAGAAATATCCATTATTATGAACAG (SEQ ID NO: 99)), and 21,251 nucleotides upstream of exon 9 at chromosome:GRCh37:1:99225610:99225687:1
 (TGCCTTGCTACGTGGGAGTCATTCCTTACATCACAGACCAACCTTCACTTGGGAAGAAGCCTCTGAAGATAAACCTTAA (SEQ ID NO: 100))

Figure 2B. Design of AAV vector with SNX7 minigene-based switch. Fig. 2B shows the construction of the non-naturally occurring SNX7 minigene using exon 8 (called exon A), a 270 nucleotide intron (AB), an exon comprising a splice modulator (e.g., LMI070) binding site at its 3' end (called exon B), a 407 nucleotide intron fragment (shortened from 21,251 nt; BC), and exon 9 (called exon C). Additional modifications were made to the minigene to improve its performance, such as: 1) a Kozak consensus sequence and ATG codon (GCCACCATG) was inserted at position 65 in exon A; 2) All other ATG sequences in the minigene were replaced with TTG; 3) a TA at position 20 of exon A was replaced with AG to make GAAGAAGAA sequence (SEQ ID NO: 69); 4) 1 nt was removed from exon B to create frame shift (number of nucleotides = 3n-1) in ORF; 5) T was inserted at position 4 of exon C to create frame shift in ORF resulting in multiple stop codons; 6) TAC at position 9 of exon C was changed to TAA to create earlier termination codon; 7) CAG at position 34 of exon C was changed to ACC to mutate a potential cryptic splice site; 8) CTCT at position 60 of exon C was changed to TAGC to create a Nhe I restriction site; and 9) TAA at the end of exon C was removed to create continuous ORF.

FIG. 2C shows the construction of a scAAV vector comprising the SNX7 minigene ON switch. The scAAV was created by combining, AAV2 ITR containing a deletion of trs, followed by a JeT promoter, followed by the SNX7 minigene (see above, Fig. 2B), followed by a coding sequence for a furin cleavage site (RNRR (SEQ ID NO: 39)) added to the end of exon C, followed by coding sequence for a T2A peptide, followed by a transgene sequence (here, a coding sequence for EGFP without the first ATG); followed by a SV40 late polyadenylation signal, followed by an AAV2 ITR.

FIG. 3 shows the regulation of GFP expression using SNX7 minigene-based ON-switch (FIG. 3A) and OFF-switch (FIG. 3B), and the mRNA expression products in the absence of splice modulator ("no LMI070") and in the presence of splice modulator ("Plus LMI070"). Figure discloses SEQ ID NOS 108-111, respectively, in order of appearance.

FIG. 4. Regulation of GFP expression by SNX7 switch in HEK293 cells. Fig. 4A shows GFP expression in HEK293 cells transfected with pSNX7-GFP (vector comprising an ON-switch) at various concentrations of splice modulator (LMI070). Fig. 4B plots GFP expression measured by mean fluorescence intensity as a function of LMI070 concentration. Fig. 4C plots quantitation of mRNA transcripts containing exon B or having direct exon A-to-exon C splicing at various concentrations of splice modulator.

FIG. 5. Regulation of GFP expression by SNX7 switch in rat cortical neurons. Fig. 5A shows GFP expression levels in primary rat neurons transfected with pSNX7-GFP (vector comprising an ON-switch) at various concentrations of splice modulator (LMI070). Fig. 5B plots quantitation of mRNA transcripts containing exon B or having direct exon A-to-exon C splicing at various concentrations of splice modulator in rat cortical neurons.

FIG. 6. AAV vectors comprising a human progranulin (PRGN) transgene under the control of SNX7 ON-switch. Fig. 6A shows 1) schematic diagram ssAAV vector comprising a neuron-specific promoter (human Synapsin promoter) and containing an SNX7 ON-switch minigene. Fig. 6B shows hPRGN expression in primary rat neurons transfected with the vectors described in Fig. 6A (Syn_SNX) in the presence or absence of splice modulator, compared with hPRGN expression levels from vectors which do not comprise the SNX7-based switch ("Syn"). Fig. 6C shows mRNA expression levels for mRNA that includes exon B and mRNA that has direct exon A to exon C splicing, in the presence and absence of splice modulator.

FIG 7A. depicts study plan of timecourse *in vivo* testing AAV vector containing SNX7 switch (version 1). Single stranded AAV9 containing hPGRN expression cassette under control of synapsin promoter with SNX7 switch was injected ICV in P0 neonatal mice. After 4 weeks, mice received oral administration of 30 mg/kg LMI070 and mice were taken down at different time points starting 24 hours post administration.

FIG 7B. demonstrates that oral administration of LMI070 switches on transgene expression in mouse brain in time-dependent manner in mice previously administered the AAV factor described in Fig. 7A. Graph demonstrates TR-FRET measurement of hPGRN expression in brain after indicated times post LMI070 delivery.

FIG 8A. depicts study plan of dose-response *in vivo* testing AAV vector containing SNX7 switch (version 1). Single stranded AAV9 containing hPGRN expression cassette under control of synapsin promoter with SNX7 switch was injected ICV in P0 neonatal mice. After 4 weeks, mice received oral administration of different doses LMI070 and mice were taken down at different time points starting 12 hours post administration. **FIG 8B** demonstrates that oral administration of LMI070 switches on transgene expression in mouse brain in dose-dependent manner in mice previously administered the AAV vector described in Fig. 8A. Graph demonstrates TR-FRET measurement of hPGRN expression in brain upon indicated doses of LMI070 and after indicated times post LMI070 delivery.

FIG. 9 shows comparison of the first version of SNX7 minigene and the modified SNX7 minigene (version 2), which has reduced size and reduced peptide expression in the absence of LMI070. Figure discloses SEQ ID NOS 108 and 112-113, respectively, in order of appearance.

FIG. 10 shows that the modified SNX7 minigene (version 2) is more sensitive than the previous version of SNX7 minigene in response to LMI070.

Detailed Description

The disclosed compositions and methods may be understood more readily by reference to the following detailed description taken in connection with the accompanying figures, which form a part of this disclosure.

Throughout this text, the descriptions refer to compositions and methods of using the compositions. Where the disclosure discloses or claims a feature or embodiment associated with a composition, such a feature or embodiment is equally applicable to the methods of using, or uses of the composition. Likewise, where the disclosure discloses or claims a feature or embodiment associated with a method of using a composition, such a feature or embodiment is equally applicable to the composition. When a range of values is expressed, it includes embodiments using any particular value within the range. Further, reference to values stated in ranges includes each and every value within that range. All ranges are inclusive of their endpoints and combinable. When values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another embodiment. Reference to a particular numerical value includes at least that particular value, unless the context clearly dictates otherwise. The use of "or" will mean "and/or" unless the specific context of its use dictates otherwise. All references cited herein are incorporated by reference for any purpose. Where a reference and the specification conflict, the specification will control. It is to be appreciated that certain features of the disclosed compositions and methods, which are, for clarity, disclosed herein in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the disclosed compositions and methods that are, for brevity, disclosed in the context of a single embodiment, may also be provided separately or in any sub-combination.

Definitions

As used herein, the singular forms "a," "an," and "the" include plural forms unless the context clearly dictates otherwise. The term "about" or "approximately," when used in the context of numerical values and ranges, refers to values or ranges that approximate or are close to the recited values or ranges such that the embodiment may perform as intended, as is apparent to the skilled person from the teachings contained herein. In some embodiments, about means plus or minus 10% of a numerical amount.

The terms "polynucleotide" and "nucleic acid" are used interchangeably herein and refer to a polymeric form of nucleotides of any length. They may include one or more of ribonucleotides or deoxyribonucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases, e.g. locked nucleic acids (LNA), peptide nucleic acids (PNA).

The terms “peptide,” “polypeptide,” and “protein” are used interchangeably, and refer to a compound comprised of amino acid residues covalently linked by peptide bonds. A protein or peptide typically contains at least two amino acids or amino acid variants, and no limitation is placed on the maximum number of amino acids that can comprise a protein’s or peptide’s sequence. Polypeptides include any peptide or protein comprising two or more amino acids or variants joined to each other by peptide bonds. The terms include, for example, biologically active fragments, substantially homologous polypeptides, oligopeptides, homodimers, heterodimers, variants of polypeptides, modified polypeptides, derivatives, analogs, fusion proteins, among others. A polypeptide includes a natural peptide, a recombinant peptide, or a combination thereof.

The term “sequence identity” and “sequence homology” are used interchangeably herein, and as used in connection with a polynucleotide or polypeptide, refers to the percentage of bases or amino acids that are the same, and are in the same relative position, when comparing or aligning two sequences of polynucleotides or polypeptides. Sequence identity can be determined in a number of different manners. For instance, sequences may be aligned using various methods and computer programs (e.g., BLAST, T-COFFEE, MUSCLE, MAFFT, etc.). See, e.g., Altschul et al., (1990) *J. Mol. Biol.*, 215:403-10.

The term “isolated” in reference to a nucleic acid or protein discussed herein refers to a nucleic acid or protein that has been separated from one or more of the components normally found associated with it in the natural environment. The separation may comprise removal from a larger nucleic acid (e.g., from a gene or chromosome) or from other proteins or molecules normally in contact with the nucleic acid or protein. The term encompasses but does not require complete isolation.

As used herein, an isolated nucleic acid comprising a “heterologous nucleic acid sequence” refers to an isolated nucleic acid comprising a portion (i.e., the heterologous nucleic acid portion) that is not normally found operably linked to one or more other components of the isolated nucleic acid in a natural context. For instance, the heterologous nucleic acid may comprise a nucleic acid sequence not originally found in a cell, bacterial cell, virus, or organism from which other components of the isolated nucleic acid (e.g., the promoter) naturally derive or where the other components of the isolated nucleic acid (e.g., the promoter) are not naturally found operatively linked with the heterologous nucleic acid in the cell, bacterial cell, virus, or organism. In some embodiments the heterologous nucleic acid includes a transgene. As used herein, a “transgene” is a nucleic acid sequence that encodes a molecule of interest (for example, a therapeutic protein, a reporter protein or a therapeutic RNA molecule) that is not originally associated with one or more components of the nucleic acid molecule. In some embodiments, the heterologous nucleic acid sequence encodes a human protein. In some embodiments, the heterologous nucleic acid sequence encodes an RNA sequence, e.g., a shRNA.

A DNA sequence or DNA polynucleotide sequence that “encodes” a particular RNA is a sequence of DNA that is capable of being transcribed into RNA. A DNA polynucleotide may encode an RNA (mRNA) that is translated into protein, or a DNA polynucleotide may encode an RNA that is not

translated into protein (e.g. tRNA, rRNA, or a guide RNA; also called “non-coding” RNA or “ncRNA”). A DNA sequence or DNA polynucleotide sequence may also “encode” a particular polypeptide or protein sequence, wherein, for example, the DNA directly encodes an mRNA that can be translated into the polypeptide or protein sequence. A “protein coding sequence” or a sequence that encodes a particular protein or polypeptide is a nucleic acid sequence that is capable of being transcribed into mRNA (in the case of DNA) and translated (in the case of mRNA) into a polypeptide in vitro or in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence may be determined by a start codon at the 5' terminus (N-terminus) and a translation stop nonsense codon at the 3' terminus (C-terminus). A coding sequence can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic nucleic acids. A transcription termination sequence will usually be located 3' to the coding sequence.

The term “promoter” or “promoter sequence” as used herein is a DNA regulatory sequence capable of facilitating transcription (e.g., capable of causing detectable levels of transcription and/or increasing the detectable level of transcription over the level provided in the absence of the promoter) of an operatively linked coding or non-coding sequence, e.g., of a downstream (3' direction) coding or non-coding sequence, e.g., through binding RNA polymerase. In some embodiments, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements to initiate transcription at levels detectable above background. In some embodiments, a promoter sequence may comprise a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. In addition to sequences sufficient to initiate transcription, a promoter may also include sequences of other regulatory elements that are involved in modulating transcription (e.g., enhancers, Kozak sequences and introns). Various promoters, including inducible promoters and constitutive promoters, may be used to drive the vectors disclosed herein. Examples of promoters known in the art that may be used in some embodiments, e.g., in viral vectors disclosed herein, include the CMV promoter, CBA promoter, smCBA promoter and those promoters derived from an immunoglobulin gene, SV40, or other tissue specific genes (e.g: RLBP1, RPE, VMD2). In addition, standard techniques are known in the art for creating functional promoters by mixing and matching known regulatory elements. Fragments of promoters, e.g., those that retain at least minimum number of bases or elements to initiate transcription at levels detectable above background, may also be used.

In some embodiments, a promoter can be a constitutively active promoter (i.e., a promoter that constitutively drives expression in any cell type and/or under any conditions). In other embodiments, a promoter can be a constitutively active promoter in a particular tissue context, e.g., in neurons, in cardiac cells, etc.. In other embodiments, a promoter can be an inducible promoter (i.e., a promoter whose activity is controlled by an external stimulus, e.g., the presence of a particular temperature, compound, or protein). In some embodiments, a promoter may be a spatially restricted promoter that can drive activity

or not depending on the physical context in which the promoter is found. Non-limiting examples of spatially restricted promoters include tissue specific promoter, cell type specific promoter, etc. In some embodiments, a promoter may be a temporally restricted promoter that drives expression depending on the temporal context in which the promoter is found. For example, a temporally restricted promoter may drive expression only at specific stages of embryonic development or during specific stages of a biological process. Non-limiting examples of temporally restricted promoters include hair follicle cycle promoters in mice.

In some embodiments, the promoter is tissue-specific such that, in a multi-cellular organism, the promoter drives expression only in a subset of specific cells. For example, tissue-specific promoters include, but are not limited to, neuron-specific promoters, adipocyte-specific promoters, cardiomyocyte-specific promoters, smooth muscle-specific promoters, photoreceptor-specific promoters, etc.. A neuron-specific promoter refers to a promoter that, when administered e.g., peripherally, directly into the central nervous system (CNS), or delivered to neuronal cells, including in vitro, ex vivo, or in vivo, preferentially drives or regulates expression of an operatively-linked heterologous nucleic acid, e.g., one encoding a protein or peptide or shRNA of interest, in neurons as compared to expression in non-neuronal cells.

The terms “DNA regulatory sequences,” “control elements,” and “regulatory elements,” used interchangeably herein, refer to transcriptional and translational control sequences, such as promoters, enhancers, silencers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate transcription of a non-coding sequence (e.g., a short hairpin RNA) or a coding sequence (e.g., PGRN) and/or regulate translation of an encoded polypeptide.

The terms “polyadenylation (polyA) signal sequence” and “polyadenylation sequence” refer to a regulatory element that provides a signal for transcription termination and addition of an adenosine homopolymeric chain to the 3' end of an RNA transcript. The polyadenylation signal may comprise a termination signal (e.g., an AAUAAA sequence or other non-canonical sequences) and optionally flanking auxiliary elements (e.g., a GU-rich element) and/or other elements associated with efficient cleavage and polyadenylation. The polyadenylation sequence may comprise a series of adenosines attached by polyadenylation to the 3' end of an mRNA. Specific polyA signal sequences may include the poly(A) signal of SEQ ID NO:22 or of SEQ ID NO: 89. In some embodiments, DNA regulatory sequences or control elements are tissue-specific regulatory sequences.

The term “post-transcriptional regulatory element” (“PRE”) refers to one or more regulatory elements that, when transcribed into mRNA, regulate gene expression at the level of the mRNA transcript. Examples of such post-transcriptional regulatory elements may include sequences that encode micro-RNA binding sites, RNA binding protein binding sites, etc. Examples of post-transcriptional regulatory element that may be used with the nucleic acid molecules and vectors disclosed herein include the woodchuck hepatitis post-transcriptional regulatory element (WPRE), the hepatitis post-transcriptional regulatory element (HPRE). Exemplary PREs may also include the PRE disclosed as SEQ ID NO: 88.

Examples PREs may also include the PRE disclosed as SEQ ID NO: 72 or the PRE disclosed as SEQ ID NO: 73.

The term "intron" refers to nucleic acid sequence(s), e.g., those within an open reading frame, that are noncoding for one or more amino acids of a polypeptide transcript (e.g., protein of interest) expressed from the nucleic acid. Intronic sequences may be transcribed from DNA into RNA (i.e., may be present in the pre-mRNA), but may be removed before the protein is expressed from the mature mRNA, e.g., through splicing.

The term "exon" refers to nucleic acid sequence(s), e.g., those within an open reading frame, that are coding for one or more amino acids of a transcript (e.g., a protein of interest) expressed from a nucleic acid. Exonic sequences may be transcribed from DNA into RNA (i.e., may be present in the pre-mRNA), and also may be present in a mature mRNA (i.e., the processed form of RNA (e.g., after splicing)) that is translated to a polypeptide.

As used herein, processes conducted "in vitro" refer to processes which are performed outside of the normal biological environment, for example, studies performed in a test tube, a flask, a petri dish, in artificial culture medium. Processes conducted "in vivo" refer to processes performed within living organisms or cells, for example, studies performed in cell cultures or in mice. Studies performed "ex vivo" refer to studies done in or on tissue from an organism in an external environment, e.g., with minimal alteration of natural conditions, e.g., allowing for manipulation of an organism's cells or tissues under more controlled conditions than may be possible in in vivo experiments.

The term "naturally-occurring" or "unmodified" as used herein as applied to, e.g., a nucleic acid, a polypeptide, a cell, or an organism, is one found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (such as a virus) is naturally occurring whether present in that organism or isolated from one or more components of the organism.

In some embodiments, a "vector" is any genetic element (e.g., DNA, RNA, or a mixture thereof) that contains a nucleic acid of interest (e.g., a transgene) that is capable of being expressed in a host cell, e.g., a nucleic acid of interest within a larger nucleic acid sequence or structure suitable for delivery to a cell, tissue, and/or organism, such as a plasmid, phage, transposon, cosmid, chromosome, virus, virion, etc. For instance, a vector may comprise an insert (e.g., a heterologous nucleic acid comprising a transgene encoding a gene to be expressed or an open reading frame of that gene) and one or more additional elements, e.g., a minigene as described herein and/or elements suitable for delivering or controlling expression of the insert. The vector may be capable of replication and/or expression, e.g., when associated with the proper control elements, and it may be capable of transferring genetic information between cells. In some embodiments, a vector may be a vector suitable for expression in a host cell, e.g., an AAV vector. In some embodiments, a vector may be a plasmid suitable for expression and/or replication, e.g., in a cell or bioreactor. In some embodiments, vectors designed specifically for the

expression of a heterologous nucleic acid sequence, e.g., a transgene encoding a protein of interest, shRNA, and the like, in the target cell may be referred to as expression vectors, and generally have a promoter sequence that drives expression of the transgene. In other embodiments, vectors, e.g., transcription vectors, may be capable of being transcribed but not translated: they can be replicated in a target cell but not expressed. Transcription vectors may be used to amplify their insert.

The term "expression vector" refers to a vector comprising a polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence to be expressed. An expression vector may comprise sufficient cis-acting elements for expression, alone or in combination with other elements for expression supplied by the host cell or in an in vitro expression system. Expression vectors include, e.g., cosmids, plasmids (e.g., naked or contained in liposomes) and viruses (e.g., lentiviruses, retroviruses, adenoviruses, and adeno-associated viruses) that incorporate the recombinant polynucleotide.

The term "plasmid" refers to a nonchromosomal (and typically double-stranded) DNA sequence comprising an intact "replicon" such that the plasmid is replicated in a host cell. A plasmid may be a circular nucleic acid. When the plasmid is placed within a unicellular organism, the characteristics of that organism are changed or transformed as a result of the DNA of the plasmid. For example, a plasmid carrying the gene for tetracycline resistance (TcR) transforms a cell previously sensitive to tetracycline into one which is resistant to it. Exemplary plasmids useful in some embodiments for the viral vectors disclosed herein include SEQ ID NO: 92.

The term "recombinant virus" as used herein is intended to refer to a non-wild-type and/or artificially produced recombinant virus (e.g., a parvovirus, adenovirus, lentivirus or adeno-associated virus etc.) that comprises a transgene or other heterologous nucleic acid. The recombinant virus may comprise a recombinant viral genome (e.g. comprising a minigene as described herein and a transgene) packaged within a viral (e.g.: AAV) capsid. A specific type of recombinant virus may be a "recombinant adeno-associated virus", or "rAAV". The recombinant viral genome packaged in the viral capsid may be a viral vector. In some embodiments, the recombinant viruses disclosed herein comprise viral vectors (e.g., comprising a minigene and transgene of interest, e.g., as described herein). Examples of viral vectors include but are not limited to an adeno-associated viral (AAV) vector, a chimeric AAV vector, an adenoviral vector, a retroviral vector, a lentiviral vector, a DNA viral vector, a herpes simplex viral vector, a baculoviral vector, or any mutant or derivative thereof.

In another embodiment, the term "transfection" is used to refer to the uptake of foreign DNA by a cell, such that the cell has been "transfected" once the exogenous DNA has been introduced inside the cell membrane. See, e.g., Graham et al., (1973) *Virology*, 52:456; Sambrook et al., (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York; Davis et al., (1986) *Basic Methods in Molecular Biology*, Elsevier; Chu et al., (1981) *Gene*, 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells. In some embodiments, the term

“transduction” is used to refer to the uptake of foreign DNA by a cell, where the foreign DNA is provided by a virus or a viral vector. Consequently, a cell has been “transduced” when exogenous DNA has been introduced inside the cell membrane. In some embodiments, the term “transformation” is used to refer to the uptake of foreign DNA by bacterial cells.

As used herein, the term “cell line” refers to a population of cells capable of continuous or prolonged growth and division in vitro. In certain circumstances, spontaneous or induced changes can occur in karyotype during storage or transfer of such clonal populations. Therefore, cells derived from the cell line referred to may not be precisely identical to the ancestral cells or cultures, and the cell line referred to includes such variants.

The term “operably linked” refers to a functional relationship between two or more polynucleotide (e.g., DNA) segments. Typically, the term refers to the functional relationship of a transcriptional regulatory sequence and a sequence to be transcribed. For example, a promoter or enhancer sequence is operably linked to a coding sequence if it, e.g., stimulates or modulates the transcription of the coding sequence in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory sequences that are operably linked to a sequence are contiguous to that sequence or are separated by short spacer sequences, i.e., they are cis-acting. However, some transcriptional regulatory sequences, such as enhancers, need not be physically contiguous or located in close proximity to the coding sequences whose transcription they enhance.

As used herein, the term “AAV vector” refers to a vector derived from or comprising one or more nucleic acid sequences derived from an adeno-associated virus serotype, including without limitation, an AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8 or AAV-9 viral vector. AAV vectors may have one or more of the AAV wild-type genes deleted in whole or part, e.g., the rep and/or cap genes, while retaining, e.g., functional flanking inverted terminal repeat (“ITR”) sequences. In some embodiments, an AAV vector may be packaged in a protein shell or capsid, e.g., comprising one or more AAV capsid proteins, which may provide a vehicle for delivery of vector nucleic acid to the nucleus of target cells. In some embodiments, an AAV vector comprises one or more AAV ITR sequences (e.g., AAV2 ITR sequences). In some embodiments, an AAV vector comprises one or more AAV ITR sequences (e.g., AAV2 ITR sequences) but does not contain any additional viral nucleic acid sequence. In some embodiments, the AAV vector components (e.g., ITRs) are derived from a different serotype virus than the rAAV capsid (for example, the AAV vector may comprise ITRs derived from AAV2 and the AAV vector may be packaged into an AAV9 capsid). Embodiments of these vector constructs are provided, e.g., in WO/2019/094253 (PCT/US2018/058744), which is incorporated herein by reference in its entirety.

In some embodiments, an “scAAV” is a self-complementary adeno-associated virus (scAAV). scAAV is termed “self-complementary” because at least a portion of the vector (e.g., at least a portion of the coding region) of the scAAV forms an intra-molecular double-stranded DNA. In some embodiments,

the rAAV is an scAAV. In some embodiments, a viral vector is engineered from a naturally occurring adeno-associated virus (AAV) to provide an scAAV for use in gene therapy. Embodiments of these vector constructs and methods of preparing and purifying them are provided, e.g., in WO/2019/094253 (PCT/US2018/058744), which is incorporated herein by reference in its entirety.

In some embodiments, an "ssAAV" is a single-stranded adeno-associated virus (ssAAV). ssAAV is termed "single-stranded" because at least a portion of the vector (e.g., at least a portion of the coding region) of the ssAAV is single-stranded DNA. In some embodiments, the rAAV is an ssAAV. In some embodiments, a viral vector is engineered from a naturally occurring adeno-associated virus (AAV) to provide an ssAAV for use in gene therapy.

As used herein, an "virus" or "virion" indicates a viral particle, comprising a viral vector, e.g., alone or in combination with one or more additional components such as one or more viral capsids. For instance, an AAV virus may comprise, e.g., a linear, single-stranded AAV nucleic acid genome associated with an AAV capsid protein coat.

In some embodiments, terms such as "virus," "virion," "AAV virus," "recombinant AAV virion," "rAAV virion," "AAV vector particle," "full capsids," "full particles," and the like refer to infectious, replication-defective virus, e.g., those comprising an AAV protein shell encapsidating a heterologous nucleotide sequence of interest, e.g., in a viral vector which is flanked on one or both sides by AAV ITRs. A rAAV virion may be produced in a suitable host cell which comprises sequences, e.g., one or more plasmids, specifying an AAV vector, alone or in combination with nucleic acids encoding AAV helper functions and accessory functions (such as cap genes), e.g., on the same or additional plasmids. In some embodiments, the host cell is rendered capable of encoding AAV polypeptides that provide for packaging the AAV vector (containing a recombinant nucleotide sequence of interest) into infectious recombinant virion particles for subsequent gene delivery.

The terms "inverted terminal repeat" or "ITR" refer to a stretch of nucleotide sequences that can form a T-shaped palindromic structure, e.g., in adeno-associated viruses (AAV) and/or recombinant adeno-associated viral vectors (rAAV). Muzyczka et al., (2001) *Fields Virology*, Chapter 29, Lippincott Williams & Wilkins. In recombinant AAV vectors, these sequences may play a functional role in genome packaging and in second-strand synthesis.

The term "host cell" denotes a cell comprising an exogenous nucleic acid of interest, for example, one or more microorganism, yeast cell, insect cell, or mammalian cell. For instance, the host cell may comprise an AAV helper construct, an AAV vector plasmid, an accessory function vector, and/or other transfer DNA. The term includes the progeny of the original cell which has been transfected. The progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

The term "AAV helper function" refers to an AAV-derived coding sequences which can be expressed to provide AAV gene products, e.g., those that function in trans for productive AAV replication. For instance, AAV helper functions may include both of the major AAV open reading frames (ORFs), rep and cap. The Rep expression products have been shown to possess many functions, including, among others: recognition, binding and nicking of the AAV origin of DNA replication; DNA helicase activity; and modulation of transcription from AAV (or other heterologous) promoters. The Cap expression products supply necessary packaging functions. AAV helper functions may be used herein to complement AAV functions in trans that are missing from AAV vectors.

The term "AAV helper construct" refers generally to a nucleic acid molecule that includes nucleotide sequences providing or encoding proteins or nucleic acids that provide AAV functions deleted from an AAV vector, e.g. a vector for delivery of a nucleotide sequence of interest to a target cell or tissue. AAV helper constructs are commonly used to provide transient expression of AAV rep and/or cap genes to complement missing AAV functions for AAV replication. Typically, helper constructs lack AAV ITRs and can neither replicate nor package themselves. AAV helper constructs may be in the form of a plasmid, phage, transposon, cosmid, virus, or virion. A number of AAV helper constructs have been disclosed, such as the commonly used plasmids pAAV/Ad and pIM29+45 which encode both Rep and Cap expression products. See, e.g., Samulski et al., (1989) *J. Virol.*, 63:3822-3828; McCarty et al., (1991) *J. Virol.*, 65:2936-2945. A number of other vectors have been disclosed which encode Rep and/or Cap expression products. See, e.g., U.S. Pat. Nos. 5,139,941 and 6,376,237. Embodiments of these vector constructs and methods of preparing and purifying them are provided, e.g., in WO/2019/094253 (PCT/US2018/058744), which is incorporated herein by reference in its entirety.

A "minigene" as the term is used herein refers to a nucleic acid sequence comprising a plurality of introns and exons and at least one splice modulator binding site. In embodiments, the presence or absence of the splice modulator during expression from the heterologous nucleic acid sequence modulates the number of exons present in the mature mRNA. Minigenes are described more fully herein.

A "splice modulator" is a molecule which binds to a splice modulator binding site and modulates the splicing of a pre-mRNA molecule, for example, a pre-mRNA molecule produced from a nucleic acid molecule described herein. In embodiments, the splice modulator increases the inclusion of an exon in the mature mRNA molecule. In other embodiments, the splice modulator decreases the inclusion of an exon in the mature mRNA molecule.

A "splice modulator binding sequence" is a sequence of nucleic acids which is recognized by a splice modulator. The term should be understood to encompass both the sequence found in a pre-mRNA as well as the sequence found in the DNA from which the pre-mRNA was produced. In exemplary embodiments, the splice modulator is a compound described herein, e.g., LMI070, and the splice modulator binding site includes the sequence AGA. In embodiments, the splice modulator binding site is disposed at or near, e.g., at, the 3' end of an exon of a minigene described herein.

A “pre-mRNA” is the first form of RNA created through transcription of DNA (e.g., of a nucleic acid molecule described herein) that has not yet undergone further processing, such as, for example, splicing. Thus, a pre-mRNA can include both introns and exons. Pre-mRNA molecules are further processed, e.g., through splicing, to form the “mature-RNA” or “mRNA.”

The nucleic acid sequences, minigenes, vectors, and methods disclosed herein relate to minigenes and regulatable expression systems comprising said minigenes, uses of splice modulators in combination with such minigenes and expression systems to control transgene expression, other uses therefor, and combinations thereof, for example, those that (1) drive expression of a transgene sequence in the presence of a splice modulator and reduce expression of the transgene sequence in the absence of a splice modulator (ON-switch) and (2) drive expression of a transgene sequence in the absence of a splice modulator and reduce expression of the transgene sequence in the presence of a splice modulator (OFF-switch). For instance, the nucleic acid sequences, vectors, and methods disclosed herein may drive expression of human PGRN or other therapeutic protein sequence in a splice-modulator-dependent fashion.

Nucleic Acid Molecules

1. Disclosed herein are nucleic acid molecules comprising a transgene encoding a molecule of interest (e.g., a protein of interest) wherein the transgene is operably linked to a minigene, e.g., as described herein.

Minigenes

The nucleic acid molecules and other aspects disclosed herein include a minigene. Exemplary minigenes of the invention are depicted in Figure 1A (on switch) and Figure 1B (off switch). Disclosed herein are minigenes which are nucleic acid sequences comprising a plurality of introns and exons and at least one splice modulator binding site. In embodiments, the minigene is operably linked to a transgene. Minigenes as described herein are used in conjunction with one or more splice modulators to control (e.g., turn on or turn off) expression of a molecule of interest from a transgene that is associated with the minigene.

In aspects, a minigene comprises: a first exon; a first intron; a second exon; a second intron; and a third exon; wherein said second exon comprises a splice modulator binding sequence and wherein, in the presence of a splice modulator, said second exon is included in an mRNA product of the nucleic acid, and in the absence of said splice modulator, said second exon is not included in an mRNA product of the nucleic acid.

In aspects, the third exon of the minigene includes a stop codon that is in frame in the mRNA product of the nucleic acid produced in the absence of the splice modulator and which is not in frame in the mRNA product of the nucleic acid produced in the presence of the splice modulator. Thus, in the

absence of a splice modulator, translation of a sequence encoding a molecule of interest (e.g., a protein of interest) disposed downstream of the minigene is reduced, for example, due to premature termination of translation by inclusion of the exon comprising the in-frame stop codon, whereas in the presence of the splice modulator, the stop codon is out of frame and translation of the molecule of interest is increased. Such aspects are thus referred to herein as “on-switch” minigenes since the presence of the splice modulator turns “on” (e.g., increases) expression of the molecule of interest.

In other embodiments, the second exon comprises a stop codon that is in frame in the mRNA product of the nucleic acid produced in the presence of the splice modulator. Thus, in the presence of a splice modulator, the exon comprising the stop codon is included in the transcript, and translation of a sequence encoding a molecule of interest (e.g., a protein of interest) disposed downstream of the minigene is decreased, whereas in the absence of the splice modulator, the exon comprising the stop codon is not present in the mRNA and expression of the molecule of interest is increased. Such aspects are thus referred to herein as “off-switch” minigenes since the presence of the splice modulator turns “off” (e.g., decreases) expression of the molecule of interest.

Without being bound by theory, it is recognized herein that vectors may have limited coding capacity (i.e., in order to be functional, their size may be limited). Thus, contemplated herein are minigenes which comprises fewer than 2000, fewer than 1900, fewer than 1800, fewer than 1700, fewer than 1600, fewer than 1500, fewer than 1400, fewer than 1300, fewer than 1200, fewer than 1100, fewer than 1000, fewer than 900, fewer than 800, fewer than 700, fewer than 600, or fewer than 500 nucleotides. Also contemplated herein are minigenes which comprise between about 2500 and about 500 nucleotides, e.g., between about 2000 and about 600 nucleotides, e.g., between about 1500 and about 700 nucleotides, e.g., between about 1200 and about 800 nucleotides, e.g. between about 1100 and about 900 nucleotides. Without being bound by theory, minigenes having such length can be included by a vector comprising a transgene and the resulting vector is of appropriate size to be functional, e.g., in a host cell. In embodiments, the sequences of the minigene are of human origin or are derived from sequences of human origin. Where the reference sequences of human origin which are identified as comprising a splice modulator binding sequence are longer than the lengths contemplated herein, such sequences may be shortened such as, for example, by deleting intronic or exonic sequence.

In embodiments, the minigenes described herein may be further modified. Such modifications are designed to improve one or more properties of the minigene. For example, a sequence derived from a human genome sequence may be included in a minigene may be further modified to mutate or remove one or more start codons (e.g., ATG sequences); remove or mutate all unwanted potential splice acceptor or splice donor sequences; include 1 or more, e.g., 2, 3, 4, 5, or 6 GAA repeats (SEQ ID NO: 101) (e.g., include GAAGAAGAA: SEQ ID NO: 69); include a Kozak sequence (e.g., a Kozak sequence of GCCACC: SEQ ID NO: 70); or any combination of modifications thereof.

Splice Modulator Binding Sequences

The aspects of the invention include minigenes comprising at least one exon comprising a splice modulator binding sequence. In aspects, the splice modulator binding sequence is disposed at or near the '3 end of an exon of the minigene. In aspects, the splice modulator binding sequence is disposed at the '3 end of an exon of the minigene. In aspects, the splice modulator binding site is derived from a sequence of the human genome. The methods described herein, e.g., in the Examples, are used to identify splice modulator binding sites recognized by splice modulators. Table 1 below, lists exemplary sequences of exons comprising a splice modulator binding site (e.g., the sequence AGA) at the '3 end of exon. Such splice modulator binding sequences are recognized by splice modulators described herein such as LMI070. Figure 2 shows the design of a minigene derived from SNX7.

Table 1. Sequences of top 10 exonic targets of LMI070 (e.g., comprising a sequence –AGA at or near the 3' end of the exon) as identified by RNAseq.

SYMBOL	CHR.	START	END	STRAND	EXON LENGTH	SEQUENCE	SEQ ID NO:
ARISJ	CHR4	114894 796	11489 4867	-	72	GTAATTAGCTGAGAAGGAAGATCTG AAGGTTTAAACGAGAGAGGGCGAGAG ATACAAAATATCTGCTAGGAGA	2
GXYLT1	CHR12	424889 53	42489 016	-	64	GGATTGTTTGTATTCTGCCAATGAT TTGTGAGACAGTCTGTTCCCCACATC CTCGTCAACAGA	3
HSD17B4	CHR5	118792 986	11879 3063	+	78	CTTCTGACATCTTAACGAGGCAATA CAGAGAGACGAATTTTCATCAGTTTG TTCAGGGAGACACATATAACAAAAG A	4
IFT57	CHR3	107911 323	10791 1373	-	51	ATCCATACATACTTAATGCTGAAATG TGAAGGGCTGAGAAAAAGAAAAGA	5
MARCH7	CHR2	160619 771	16061 9867	+	97	AATTGGAACATCGAGGGAAAATGG GCTTTTTATTATTAACAAAACCTCA GTATTATCACTTAGAAACCTGAAATT GAACTCCAAAAGCCAAAGA	6
SNX24	CHR5	122233 837	12223 3931	+	95	AAGAATGTTCTTTTGTGAAGAATGA CTTAAGGAAGATTCATGATGACTGA GTGTGCCCGTGTGGAACCTTAGGAC ATAGATGCACTCCTACAGA	7
SNX7	CHR1	992042 87	99204 359	+	73	CCTTGCTATCCCTGTCTTCTGTAGCT ATTCTGAAACCATCAACAAAGGAGCA CACCATTCCATCAGCAAAAGA	1
STRADB	CHR2	202335 765	20233 5834	+	70	TTGTCCTTCACTCCGTACTCCAGTTG GCCAAGCATAGGTGCGATGCCAGG GTCAAGGAGACTAAGGGAGA	8

VDAC2	CHR10	769901 68	76990 208	+	41	GACATACAGACATGGCAGCCCCTAG CATGTGTATCCTAAGA	9
VDAC2	CHR10	769901 69	76990 208	+	40	ACATACAGACATGGCAGCCCCTAGC ATGTGTATCCTAAGA	10

SEQ ID NO: 80 is the full genomic sequence (144nt) of the cryptic exon comprising a splice modulator binding site, located between exon 8 and 9 of the *snx7* locus comprising SEQ ID NO: 1.

AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGAATGAGCAGAAAATCATTTTCAGGGCCTGTTCTCT
ATGTCCTTGCTATCCCTGTCTTCTGTAGCTATTCTGAAACCATCAACAAAGGAGCACACCATTCCATC
AGCAAAAGA (SEQ ID NO: 80).

SEQ ID NO:16 is derived from SEQ ID NO: 80, with the modifications to create frameshift in ORF and removed start codons to avoid leaking expression.

AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGATTGAGCAGAAAATCATTTTCAGGGCCTGTTCTCT
ATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCTGAAACCATCAACAAAGGAGCACACCATTCCATCA
GCAAAAGA (SEQ ID NO: 16).

In embodiments, the minigene comprises an exon sequence, e.g., a second exon sequence, derived from any one of SEQ ID NO:1 to SEQ ID NO: 10 or SEQ ID NO: 80. In embodiments, an exon of the minigene, e.g., the second exon, includes or consists of any one of SEQ ID NO:1 to SEQ ID NO: 10 or SEQ ID NO: 80. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 1, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 1 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 2, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 2 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 3, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 3 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 4, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 4 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 5, or a sequence having at least 90%, 95%, 97%, 98%, or 99%

identity thereto; or a fragment of SEQ ID NO: 5 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 6, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 6 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 7, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 7 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 8, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 8 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 9, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 9 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 10, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 10 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 80, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 80 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In some embodiments, the minigenes described herein include an exon, e.g., a second exon, comprising or consisting of SEQ ID NO: 16, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or a fragment of SEQ ID NO: 16 comprising at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, 95%, 97%, 98%, or 99% of the nucleotides of the sequence. In embodiments, the second exon consists of SEQ ID NO: 16.

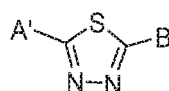
In some embodiments, the second exon is modified to consist of $3n-1$ nucleotides, where n is any integer, such that inclusion of the second exon in the mRNA results in a frame shift relative to the mRNA which does not include the second exon.

Splice Modulators

A “splice modulator” as term is used herein refers to a compound which is capable of mediating alternative splicing. In exemplary embodiments the splice modulator modulates (e.g.,

increases) the inclusion of an exon in an mRNA product. In exemplary embodiments, the splice modulator modulates (e.g., increases) the inclusion of an exon in an mRNA product by binding to a splice modulator binding sequence (e.g., the sequence AGA, e.g., the sequence AGA at the 3' end of the exon that is modulated).

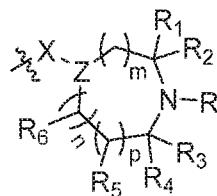
In aspects of the invention, the splice modulator is a compound described herein. In a first splice modulator aspect, the splice modulator is a compound according to Formula (I):



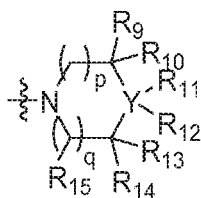
(I)

or pharmaceutically acceptable salts thereof, wherein A' is phenyl which is substituted with 0, 1, 2, or 3 substituents independently selected from C₁-C₄alkyl, wherein 2 C₁-C₄alkyl groups can combine with the atoms to which they are bound to form a 5-6 membered ring and is substituted with 0 or 1 substituents selected from oxo, oxime and, hydroxy, haloC₁-C₄alkyl, dihaloC₁-C₄alkyl, trihaloC₁-C₄alkyl, C₁-C₄alkoxy, C₁-C₄alkoxy- C₃-C₇cycloalkyl, haloC₁-C₄alkoxy, dihaloC₁-C₄alkoxy, trihaloC₁-C₄alkoxy, hydroxy, cyano, halogen, amino, mono- and di-C₁-C₄alkylamino, heteroaryl, C₁-C₄alkyl substituted with hydroxy, C₁-C₄alkoxy substituted with aryl, amino, -C(O)NH C₁-C₄alkyl – heteroaryl, -NHC(O)- C₁-C₄alkyl- heteroaryl, C₁-C₄alkyl C(O)NH- heteroaryl, C₁-C₄alkyl NHC(O)- heteroaryl, 3-7 membered cycloalkyl, 5-7 membered cycloalkenyl or 5, 6 or 9 membered heterocycle containing 1 or 2 heteroatoms, independently, selected from S, O and N, wherein heteroaryl has 5, 6 or 9 ring atoms, 1, 2 or 3 ring heteroatoms selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from oxo, hydroxy, nitro, halogen, C₁-C₄alkyl, C₁-C₄alkenyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, C₁-C₄alkyl-OH, trihaloC₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, -C(O)NH₂, -NH₂, -NO₂, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, 4-7member heterocycleC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl; or A' is 6 member heteroaryl having 1-3 ring nitrogen atoms, which 6 member heteroaryl is substituted by phenyl or a heteroaryl having 5 or 6 ring atoms, 1 or 2 ring heteroatoms independently selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from C₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and

di-C₁-C₄alkylamino C₁-C₄alkyl; or A' is bicyclic heteroaryl having 9 to 10 ring atoms and 1, 2, or 3 ring heteroatoms independently selected from N, O or S, which bicyclic heteroaryl is substituted with 0, 1, or 2 substituents independently selected from oxo, cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy and C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino; B is a group of the formula:



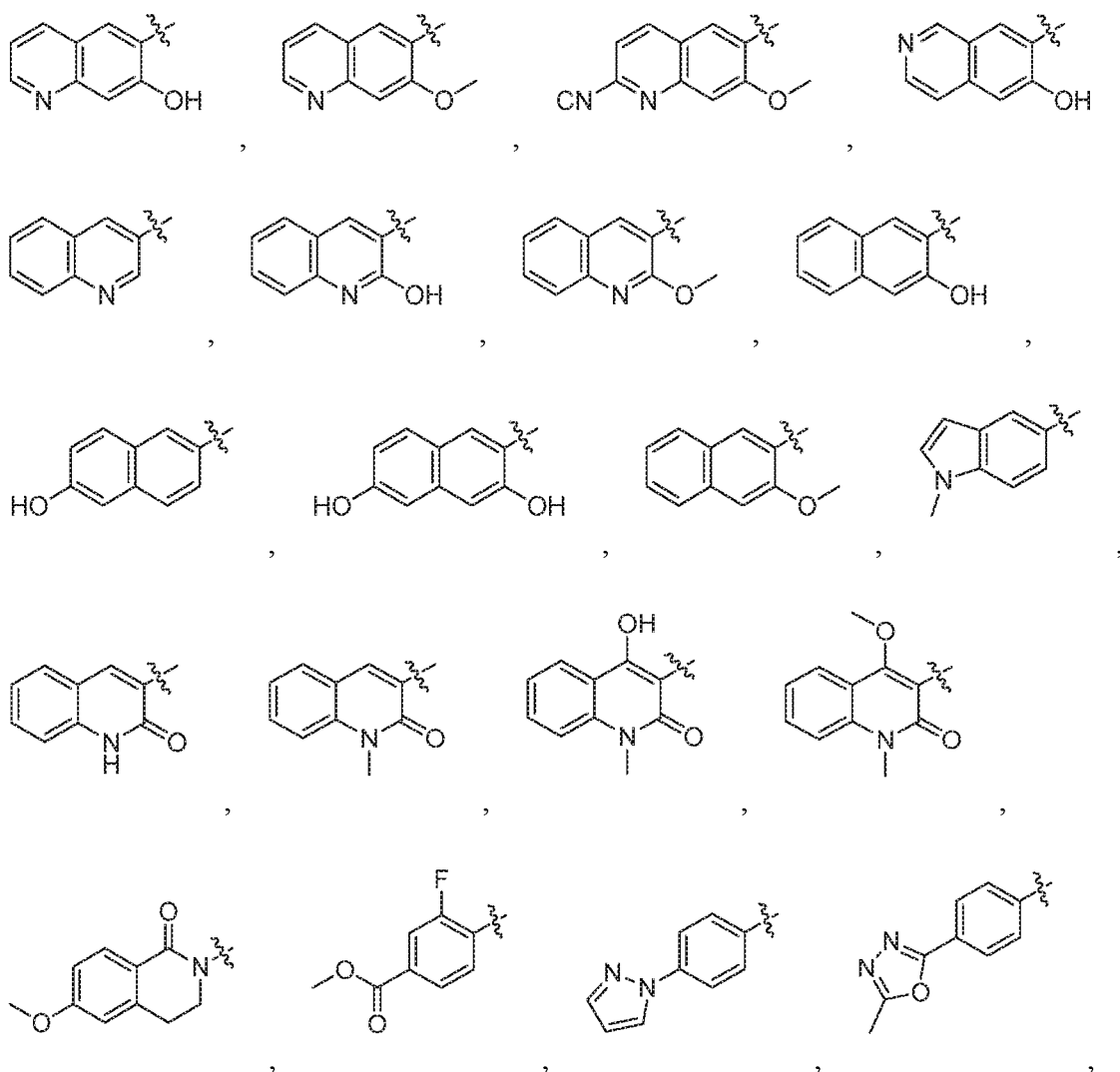
wherein m, n and p are independently selected from 0 or 1; R, R₁, R₂, R₃, and R₄ are independently selected from the group consisting of hydrogen, C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄alkylamino; R₅ and R₆ are independently selected from hydrogen and fluorine; or R and R₃, taken in combination form a fused 5 or 6 member heterocyclic ring having 0 or 1 additional ring heteroatoms selected from N, O or S; R₁ and R₃, taken in combination form a C₁-C₃alkylene group; R₁ and R₅, taken in combination form a C₁-C₃alkylene group; R₃ and R₄, taken in combination with the carbon atom to which they attach, form a spirocyclic C₃-C₆cycloalkyl; X is CR_A[•]R_B[•], NR₇ or a bond; R₇ is hydrogen, or C₁-C₄alkyl; R_A[•] and R_B[•] are independently selected from hydrogen and C₁-C₄alkyl, or R_A[•] and R_B[•], taken in combination, form a divalent C₂-C₅alkylene group; Z is CR₈ or N; when Z is N, X is a bond; R₈ is hydrogen or taken in combination with R₆ form a double bond; or B is a group of the formula:

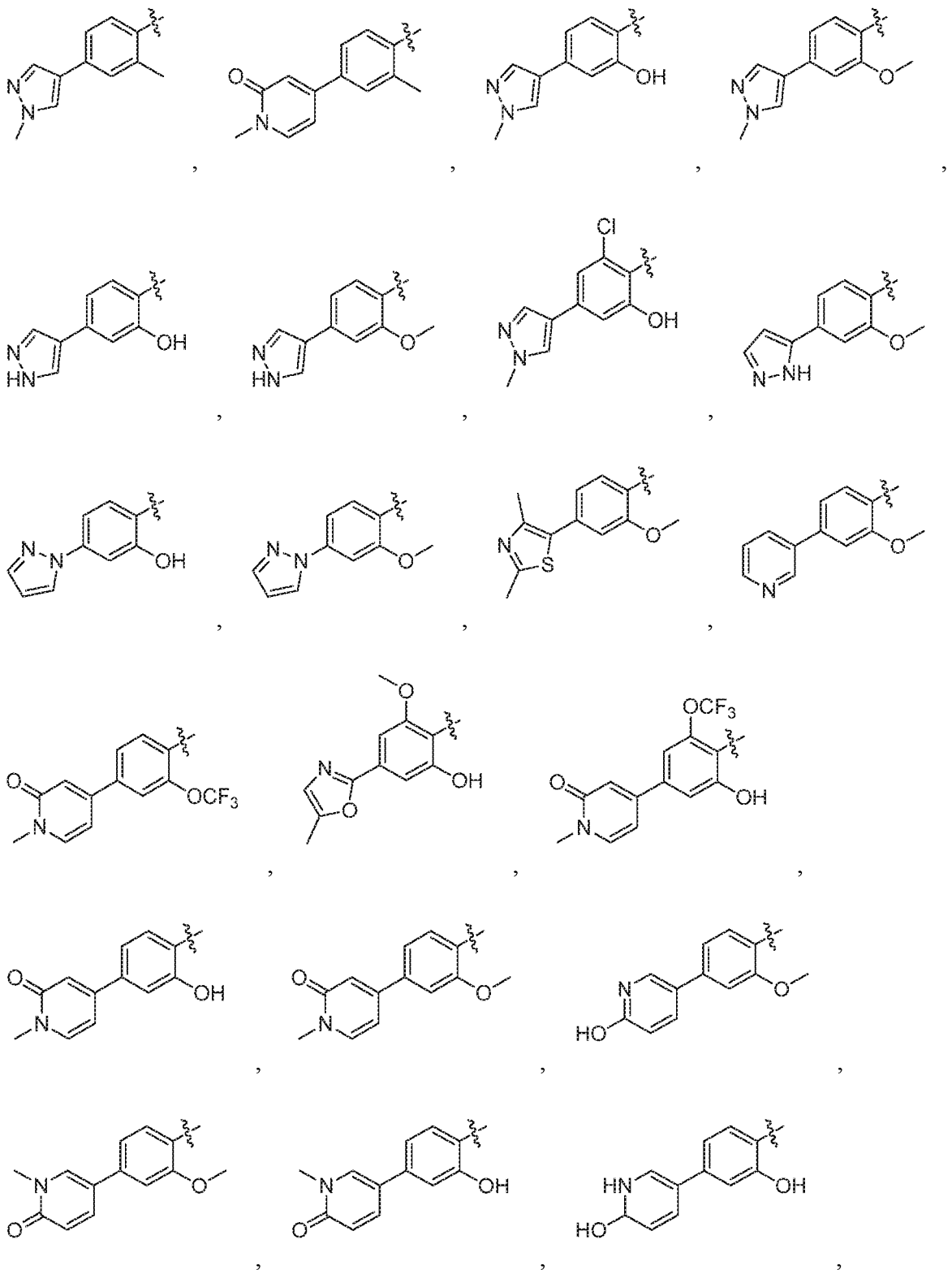


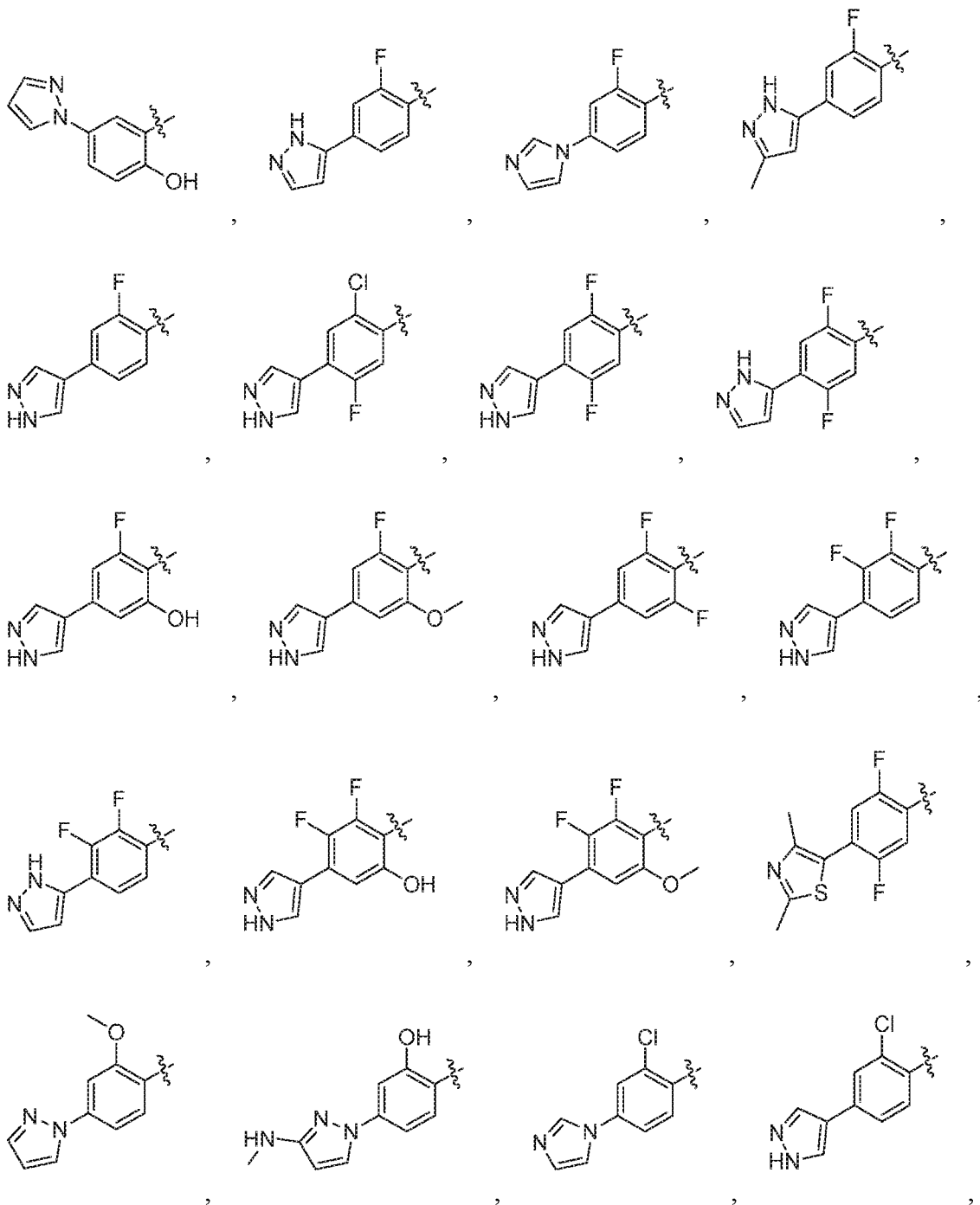
wherein Y is C or O and when Y is O R₁₁ and R₁₂ are both absent; p and q are independently selected from the group consisting of 0, 1, and 2; R₉ and R₁₃ are independently selected from hydrogen and C₁-C₄alkyl; R₁₀ and R₁₄ are independently selected from hydrogen, amino, mono- and di-C₁-C₄akylamino and C₁-C₄alkyl, which alkyl is optionally substituted with

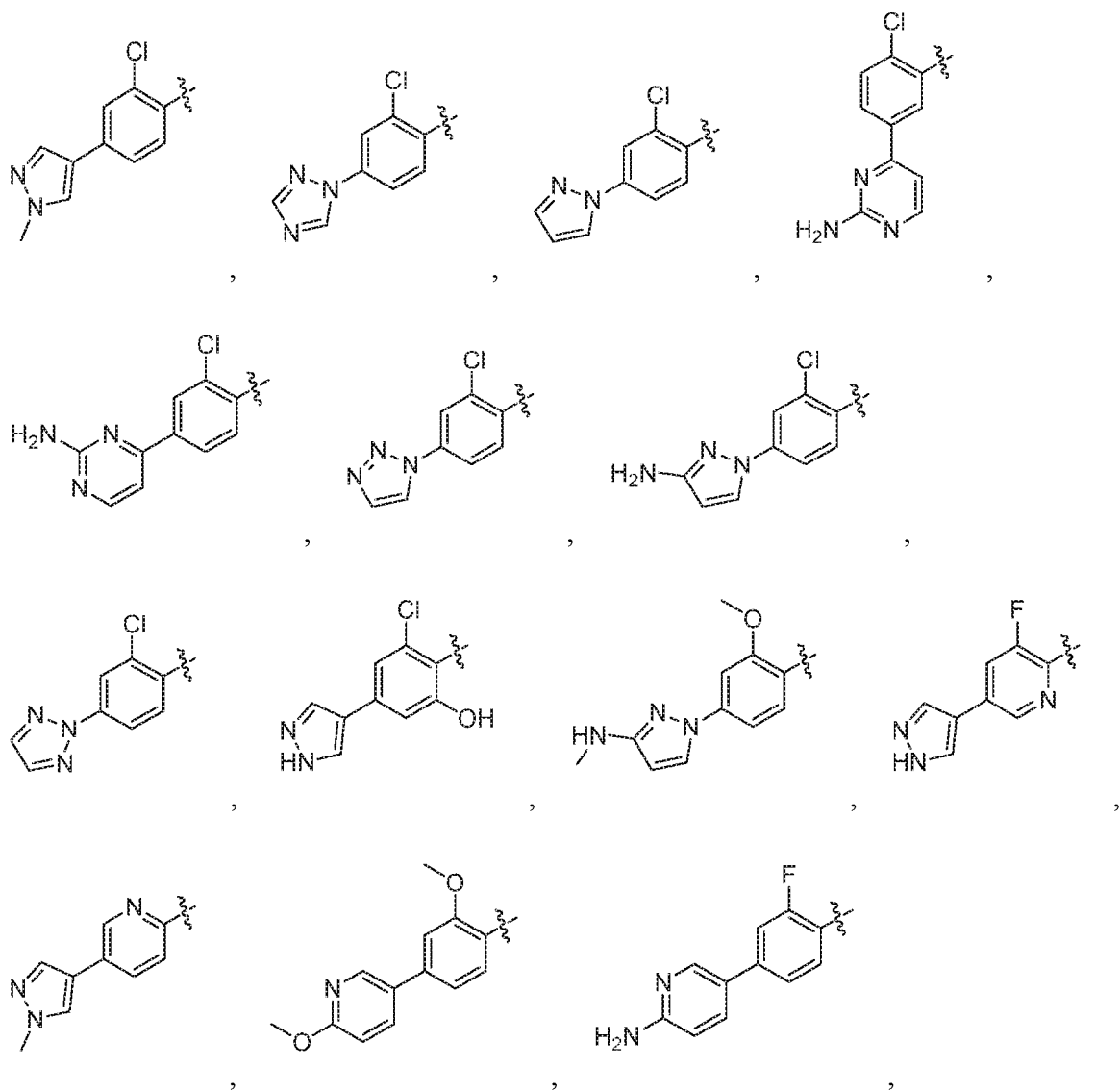
hydroxy, amino or mono- and di-C₁-C₄akylamino; R₁₁ is hydrogen, C₁-C₄alkyl, amino or mono- and di-C₁-C₄akylamino; R₁₂ is hydrogen or C₁-C₄alkyl; or R₉ and R₁₁, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups; or R₁₁ and R₁₂, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups.

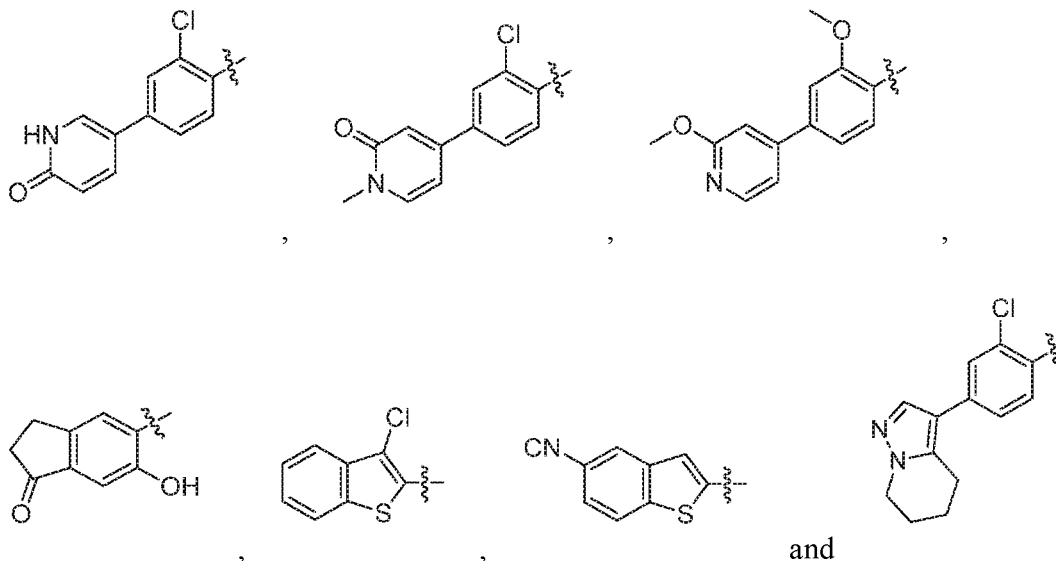
In a second splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to the first splice modulator aspect wherein A' is selected from:



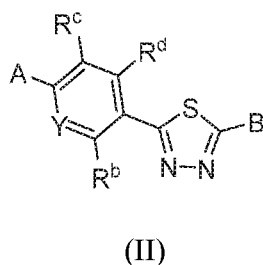






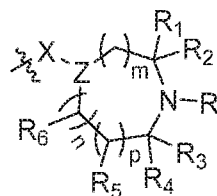


In a third splice modulator aspect, the splice modulator is a compound according to Formula (II):

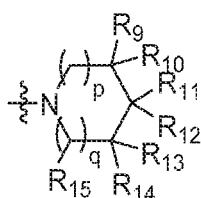


or pharmaceutically acceptable salts thereof, wherein Y is N or C-R^a; R^a is hydrogen or C₁-C₄alkyl; R^b is hydrogen, C₁-C₄alkyl, C₁-C₄alkoxy, hydroxy, cyano, halogen, trihalo C₁-C₄alkyl or trihalo C₁-C₄alkoxy; R^c and R^d are each, independently, hydrogen, C₁-C₄alkyl, C₁-C₄alkoxy, hydroxy, trihalo C₁-C₄alkyl, trihalo C₁-C₄alkoxy or heteroaryl; A is 6 member heteroaryl having 1-3 ring nitrogen atoms, which 6 member heteroaryl is substituted with 0, 1, or 2 substituents independently selected from oxo, C₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl; or A is 5 member heteroaryl having 1-3 ring heteroatoms independently selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from C₁-C₄alkyl, hydroxyl, mono- and di-C₁-C₄alkylamino, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl; or A and R^c, together with the atoms to which they are bound, form a 6 member aryl with 0, 1, or 2 substituents independently

selected from cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy and C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino; B is a group of the formula:



wherein m, n and p are independently selected from 0 or 1; R, R₁, R₂, R₃, and R₄ are independently selected from the group consisting of hydrogen, C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄alkylamino; R₅ and R₆ are independently selected from hydrogen and fluorine; or R and R₃, taken in combination form a fused 5 or 6 member heterocyclic ring having 0 or 1 additional ring heteroatoms selected from N, O or S; R₁ and R₃, taken in combination form a C₁-C₃alkylene group; R₁ and R₅, taken in combination form a C₁-C₃alkylene group; R₃ and R₄, taken in combination with the carbon atom to which they attach, form a spirocyclic C₃-C₆cycloalkyl; X is CR_{A'}R_{B'}, NR₇ or a bond; R₇ is hydrogen, or C₁-C₄alkyl; R_{A'} and R_{B'} are independently selected from hydrogen and C₁-C₄alkyl, or R_{A'} and R_{B'}, taken in combination, form a divalent C₂-C₅alkylene group; Z is CR₈ or N; when Z is N, X is a bond; R₈ is hydrogen or taken in combination with R₆ form a double bond; or B is a group of the formula:

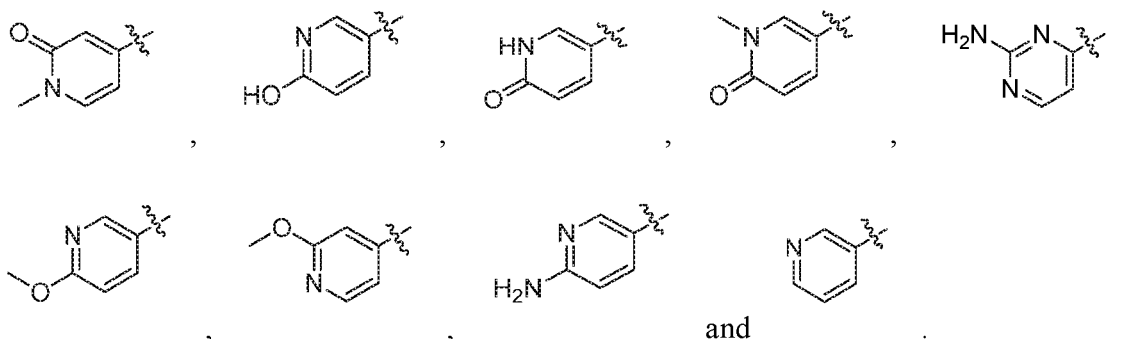


wherein p and q are independently selected from the group consisting of 0, 1, and 2; R₉ and R₁₃ are independently selected from hydrogen and C₁-C₄alkyl; R₁₀ and R₁₄ are independently selected from hydrogen, amino, mono- and di-C₁-C₄alkylamino and C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄alkylamino; R₁₁ is hydrogen, C₁-C₄alkyl, amino or mono- and di-C₁-C₄alkylamino; R₁₂ is hydrogen or C₁-C₄alkyl; or R₉ and R₁₁, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally

substituted with 1-3 C₁-C₄alkyl groups; or R₁₁ and R₁₂, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups.

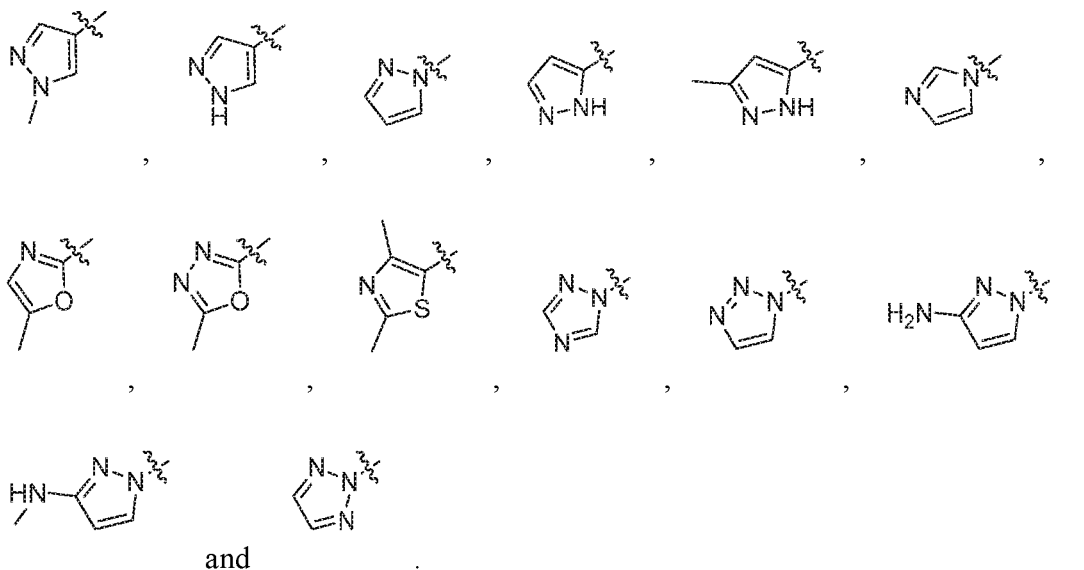
In a fourth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to the third splice modulator aspect, wherein A is 6 member heteroaryl having 1-3 ring nitrogen atoms, which 6 member heteroaryl is substituted with 0, 1, or 2 substituents independently selected from oxo, C₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl.

In a fifth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the third or fourth splice modulator aspects, wherein A is selected from:

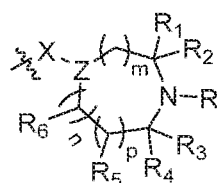


In a sixth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to the third splice modulator aspect, wherein A is 5 member heteroaryl having 1-3 ring heteroatoms independently selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from C₁-C₄alkyl, hydroxyl, mono- and di-C₁-C₄alkylamino, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl.

In a seventh splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the third or sixth splice modulator aspects, wherein A is selected from:

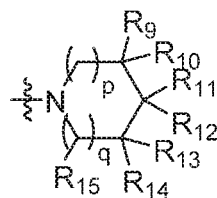


In an eighth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the first through seventh splice modulator aspects, wherein B is a group of the formula:



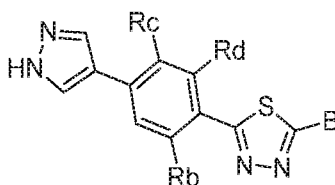
wherein m, n and p are independently selected from 0 or 1; R, R₁, R₂, R₃, and R₄ are independently selected from the group consisting of hydrogen, C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₅ and R₆ are hydrogen; or R and R₃, taken in combination form a fused 5 or 6 member heterocyclic ring having 0 or 1 additional ring heteroatoms selected from N, O or S; R₁ and R₃, taken in combination form a C₁-C₃alkylene group; R₁ and R₅, taken in combination form a C₁-C₃alkylene group; R₃ and R₄, taken in combination with the carbon atom to which they attach, form a spirocyclicC₃-C₆cycloalkyl; X is CR_{A'}R_{B'}, O, NR₇ or a bond; R_{A'} and R_{B'} are independently selected from hydrogen and C₁-C₄alkyl, or R_{A'} and R_{B'}, taken in combination, form a divalent C₂-C₅alkylene group; Z is CR₈ or N; when Z is N, X is a bond; R₈ is hydrogen or taken in combination with R₆ form a double bond.

In a ninth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the first through seventh splice modulator aspects, wherein B is a group of the formula:



wherein p and q are independently selected from the group consisting of 0, 1, and 2; R₉ and R₁₃ are independently selected from hydrogen and C₁-C₄alkyl; R₁₀ and R₁₄ are independently selected from hydrogen, amino, mono- and di-C₁-C₄akylamino and C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₁₁ is hydrogen, C₁-C₄alkyl, amino or mono- and di-C₁-C₄akylamino; R₁₂ is hydrogen or C₁-C₄alkyl; or R₉ and R₁₁, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups; or R₁₁ and R₁₂, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups.

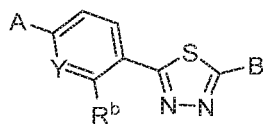
In a tenth splice modulator aspect, the splice modulator is a compound according to Formula (III):



(III)

or pharmaceutically acceptable salt thereof, wherein R^b is hydrogen or hydroxy; R^c is hydrogen or halogen; and R^d is halogen.

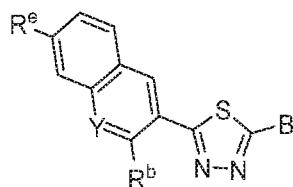
In an eleventh splice modulator aspect, the splice modulator is a compound according to Formula (IV):



(IV)

or pharmaceutically acceptable salt thereof, wherein R^b is hydroxyl, methoxy, trifluoromethyl or trifluoromethoxy.

In a twelfth splice modulator aspect, the splice modulator is a compound according to Formula (V):



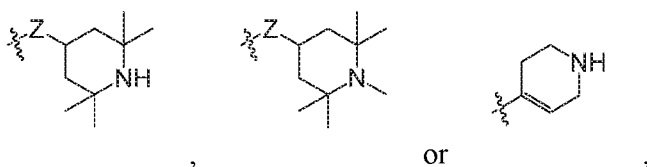
(V)

or pharmaceutically acceptable salt thereof, wherein R^b is hydroxyl, methoxy, trifluoromethyl or trifluoromethoxy; and R^e is hydrogen, hydroxy or methoxy.

In a thirteenth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the third through ninth or eleventh through twelfth splice modulator aspects, wherein Y is N.

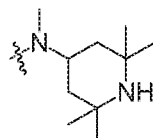
In a fourteenth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the third through ninth or eleventh through twelfth splice modulator aspects, wherein Y is CH.

In a fifteenth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the first through eighth or tenth through fourteenth splice modulator aspects, wherein B is selected from



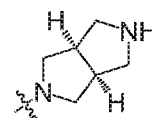
wherein Z is NH or N(Me).

In a sixteenth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salts thereof, according of any one of the first through eighth or tenth

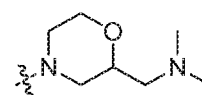
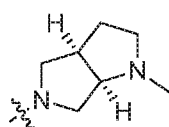
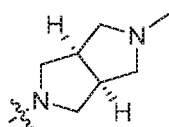


through fifteenth splice modulator aspects, wherein B is

In a seventeenth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according of any one of the first through seventh or

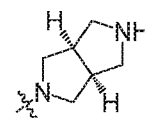


ninth through fourteenth splice modulator aspects, wherein B is selected from



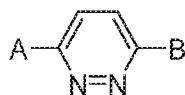
and

In an eighteenth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according of any one of the first through seventh, ninth



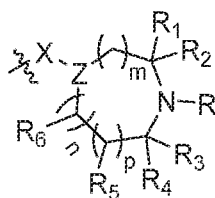
through fourteenth or seventeenth splice modulator aspects wherein B is

In a nineteenth splice modulator aspect, the splice modulator is a compound according to Formula (VI):

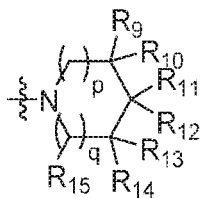


(VI)

or pharmaceutically acceptable salt thereof, wherein A is bicyclic heteroaryl or heterocycle having 9 or 10 ring atoms and 1 or 2 ring N atoms and 0 or 1 O atoms, which bicyclic heteroaryl or heterocycle is substituted with 0, 1, 2, 3, 4 or 5 substituents independently selected from -C(O)NH₂, -C(O)O-C₁-C₄alkyl, aryl, oxo, cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C₁-C₄alkyl, C₁-C₄alkyl aryl, C₁-C₄alkyl heterocyclyl, C₁-C₄alkyl heteroaryl, C₁-C₄alkoxy aryl, C₁-C₄alkoxy heterocyclyl, C₁-C₄alkoxy heteroaryl, C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino; and B is a group of the formula:

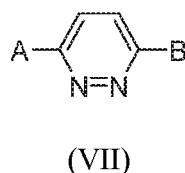


wherein m, n and p are independently selected from 0 or 1; R, R₁, R₂, R₃, and R₄ are independently selected from the group consisting of hydrogen, C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄alkylamino; R₅ and R₆ are independently selected from hydrogen and fluorine; or R and R₃, taken in combination form a fused 5 or 6 member heterocyclic ring having 0 or 1 additional ring heteroatoms selected from N, O or S; R₁ and R₃, taken in combination form a C₁-C₃alkylene group; R₁ and R₅, taken in combination form a C₁-C₃alkylene group; R₃ and R₄, taken in combination with the carbon atom to which they attach, form a spirocyclic C₃-C₆cycloalkyl; X is CR_AR_B, O, NR₇ or a bond; R₇ is hydrogen, or C₁-C₄alkyl; R_A and R_B are independently selected from hydrogen and C₁-C₄alkyl, or R_A and R_B, taken in combination, form a divalent C₂-C₅alkylene group; Z is CR₈ or N; when Z is N, X is a bond; R₈ is hydrogen or taken in combination with R₆ form a double bond; or B is a group of the formula:

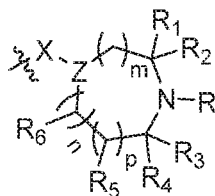


wherein p and q are independently selected from the group consisting of 0, 1, and 2; R₉ and R₁₃ are independently selected from hydrogen and C₁-C₄alkyl; R₁₀ and R₁₄ are independently selected from hydrogen, amino, mono- and di-C₁-C₄akylamino and C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₁₁ is hydrogen, C₁-C₄alkyl, amino or mono- and di-C₁-C₄akylamino; R₁₂ is hydrogen or C₁-C₄alkyl; or R₉ and R₁₁, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups; or R₁₁ and R₁₂, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups.

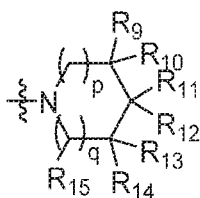
In a twentieth splice modulator aspect, the splice modulator is a compound according to Formula (VII):



or pharmaceutically acceptable salt thereof, wherein A is bicyclic heteroaryl having 10 ring atoms and 1 or 2 ring N atoms, which bicyclic heteroaryl is substituted with 0, 1, or 2 substituents independently selected from oxo, cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C₁-C₄alkyl, C₁-C₄alkyl aryl, C₁-C₄alkyl heterocyclyl, C₁-C₄alkyl heteroaryl, C₁-C₄alkoxy aryl, C₁-C₄alkoxy heterocyclyl, C₁-C₄alkoxy heteroaryl, C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono-and di-C₁-C₄alkylamino; and B is a group of the formula:

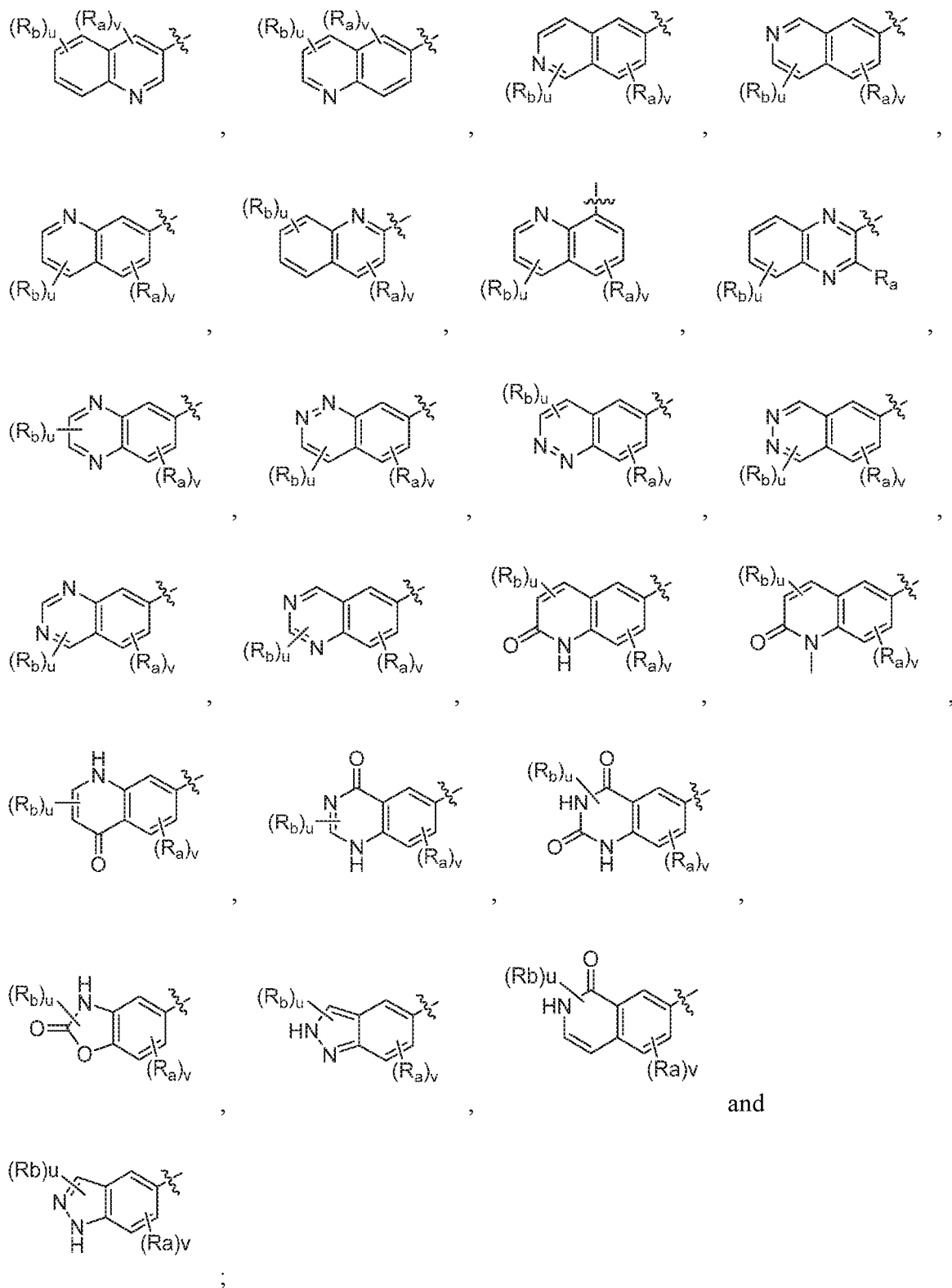


wherein m, n and p are independently selected from 0 or 1; R, R₁, R₂, R₃, and R₄ are independently selected from the group consisting of hydrogen, C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₅ and R₆ are independently selected from hydrogen and fluorine; or R and R₃, taken in combination form a fused 5 or 6 member heterocyclic ring having 0 or 1 additional ring heteroatoms selected from N, O or S; R₁ and R₃, taken in combination form a C₁-C₃alkylene group; R₁ and R₅, taken in combination form a C₁-C₃alkylene group; R₃ and R₄, taken in combination with the carbon atom to which they attach, form a spirocyclic C₃-C₆cycloalkyl; X is CR_AR_B, O, NR₇ or a bond; R₇ is hydrogen, or C₁-C₄alkyl; R_A and R_B are independently selected from hydrogen and C₁-C₄alkyl, or R_A and R_B, taken in combination, form a divalent C₂-C₅alkylene group; Z is CR₈ or N; when Z is N, X is a bond; R₈ is hydrogen or taken in combination with R₆ form a double bond; or B is a group of the formula:



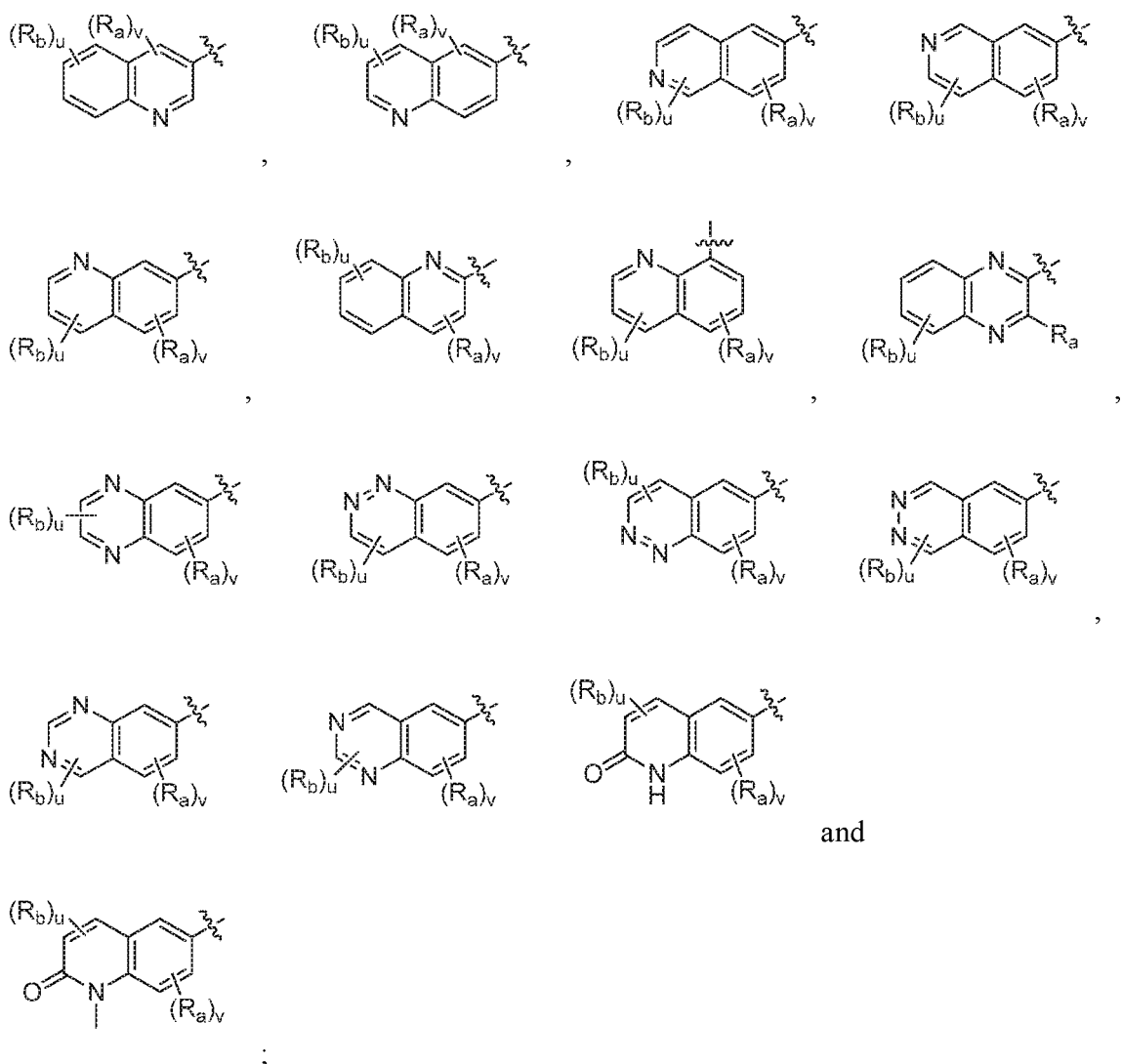
wherein p and q are independently selected from the group consisting of 0, 1, and 2; R₉ and R₁₃ are independently selected from hydrogen and C₁-C₄alkyl; R₁₀ and R₁₄ are independently selected from hydrogen, amino, mono- and di-C₁-C₄akylamino and C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₁₁ is hydrogen, C₁-C₄alkyl, amino or mono- and di-C₁-C₄akylamino; R₁₂ is hydrogen or C₁-C₄alkyl; or R₉ and R₁₁, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups; or R₁₁ and R₁₂, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups.

In a twenty-first splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth or twentieth splice modulator aspects, wherein A is selected from:



wherein u and v are each, independently, 0, 1, 2 or 3; and each R_a and R_b are, independently, selected from cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C₁-C₄alkyl, C₁-C₄alkyl aryl, C₁-C₄alkyl heterocyclyl, C₁-C₄alkyl heteroaryl, C₁-C₄alkoxy aryl, C₁-C₄alkoxy heterocyclyl, C₁-C₄alkoxy heteroaryl, and C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino.

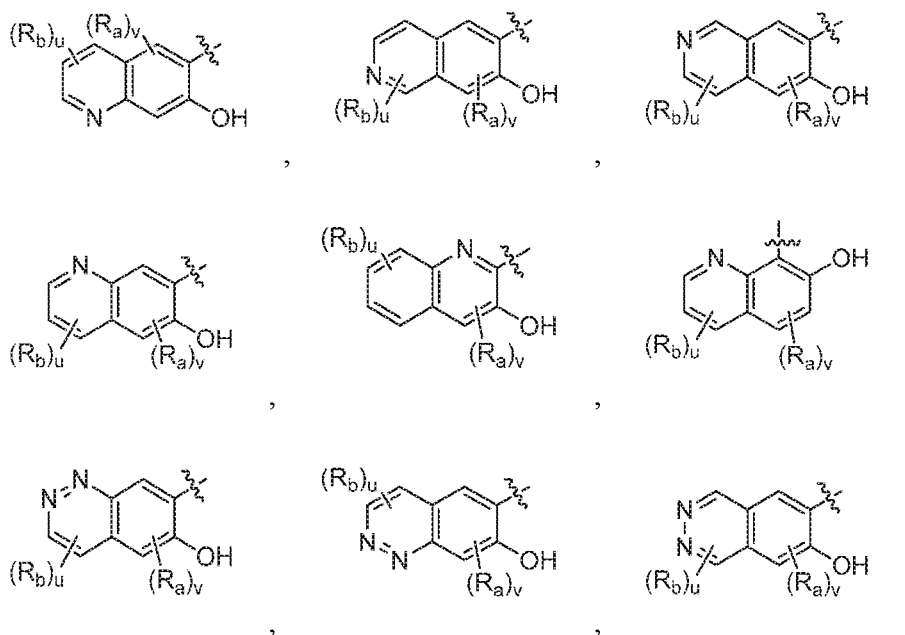
In a twenty-second splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through twenty-first splice modulator aspect, wherein A is selected from:

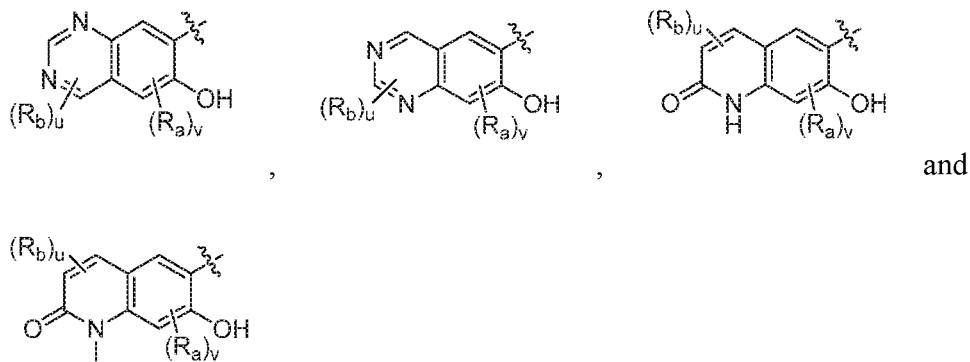


wherein u and v are each, independently, 0, 1, 2 or 3; and each R_a and R_b are, independently, selected from, cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C₁-C₄alkyl, C₁-C₄alkyl aryl, C₁-C₄alkyl heterocyclyl, C₁-C₄alkyl heteroaryl, C₁-C₄alkoxy aryl, C₁-C₄alkoxy heterocyclyl, C₁-C₄alkoxy heteroaryl, and C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino.

In another splice modulator aspect, provided herein are compounds or pharmaceutically acceptable salts thereof, according to any one of the nineteenth through twenty-second splice modulator aspects, wherein A is substituted in the ortho position with a hydroxyl group.

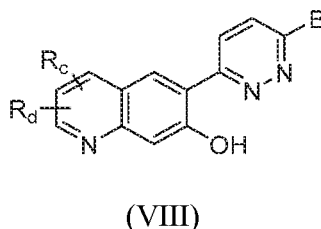
In a twenty-third splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through twenty-second splice modulator aspects, wherein A is selected from:





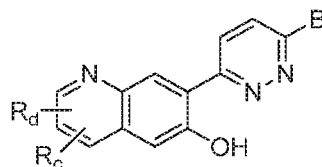
In a twenty-fourth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through twenty-third splice modulator aspect, wherein A has a single N atom.

In a twenty-fifth splice modulator aspect, the splice modulator is a compound according to Formula (VIII):



or pharmaceutically acceptable salt thereof, wherein R_c and R_d are each, independently, selected from hydrogen, cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C₁-C₄alkyl, C₁-C₄alkyl aryl, C₁-C₄alkyl heterocyclyl, C₁-C₄alkyl heteroaryl, C₁-C₄alkoxy aryl, C₁-C₄alkoxy heterocyclyl, C₁-C₄alkoxy heteroaryl, C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino.

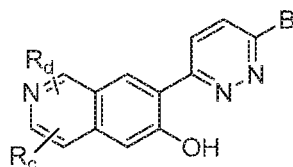
In a twenty-sixth splice modulator aspect, the splice modulator is a compound according to Formula (IX):



(IX)

or pharmaceutically acceptable salt thereof, wherein R_c and R_d are each, independently, selected from hydrogen, cyano, halogen, hydroxy, C_1 - C_4 alkyl, C_2 - C_4 alkenyl, C_2 - C_4 alkynyl, C_1 - C_4 alkoxy, C_3 - C_7 cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C_1 - C_4 alkyl, C_1 - C_4 alkyl aryl, C_1 - C_4 alkyl heterocyclyl, C_1 - C_4 alkyl heteroaryl, C_1 - C_4 alkoxy aryl, C_1 - C_4 alkoxy heterocyclyl, C_1 - C_4 alkoxy heteroaryl, C_1 - C_4 alkoxy substituted with hydroxy, C_1 - C_4 alkoxy, amino and mono- and di- C_1 - C_4 alkylamino.

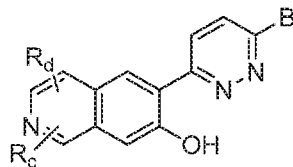
In a twenty-seventh splice modulator aspect, the splice modulator is a compound according to Formula (X):



(X)

or pharmaceutically acceptable salt thereof, wherein R_c and R_d are each, independently, selected from hydrogen, cyano, halogen, hydroxy, C_1 - C_4 alkyl, C_2 - C_4 alkenyl, C_2 - C_4 alkynyl, C_1 - C_4 alkoxy, C_3 - C_7 cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C_1 - C_4 alkyl, C_1 - C_4 alkyl aryl, C_1 - C_4 alkyl heterocyclyl, C_1 - C_4 alkyl heteroaryl, C_1 - C_4 alkoxy aryl, C_1 - C_4 alkoxy heterocyclyl, C_1 - C_4 alkoxy heteroaryl, C_1 - C_4 alkoxy substituted with hydroxy, C_1 - C_4 alkoxy, amino and mono- and di- C_1 - C_4 alkylamino.

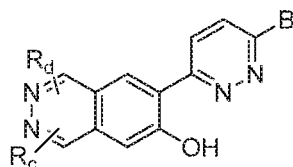
In a twenty-eighth splice modulator aspect, the splice modulator is a compound according to Formula (XI):



(XI)

or pharmaceutically acceptable salt thereof, wherein R_c and R_d are each, independently, selected from hydrogen, cyano, halogen, hydroxy, C_1 - C_4 alkyl, C_2 - C_4 alkenyl, C_2 - C_4 alkynyl, C_1 - C_4 alkoxy, C_3 - C_7 cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C_1 - C_4 alkyl, C_1 - C_4 alkyl aryl, C_1 - C_4 alkyl heterocyclyl, C_1 - C_4 alkyl heteroaryl, C_1 - C_4 alkoxy aryl, C_1 - C_4 alkoxy heterocyclyl, C_1 - C_4 alkoxy heteroaryl, C_1 - C_4 alkoxy substituted with hydroxy, C_1 - C_4 alkoxy, amino and mono- and di- C_1 - C_4 alkylamino.

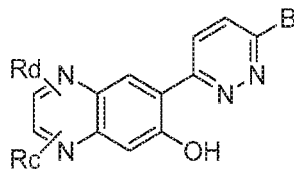
In a twenty-ninth splice modulator aspect, the splice modulator is a compound according to Formula (XII):



(XII)

or pharmaceutically acceptable salt thereof, wherein R_c and R_d are each, independently, selected from hydrogen, cyano, halogen, hydroxy, C_1 - C_4 alkyl, C_2 - C_4 alkenyl, C_2 - C_4 alkynyl, C_1 - C_4 alkoxy, C_3 - C_7 cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C_1 - C_4 alkyl, C_1 - C_4 alkyl aryl, C_1 - C_4 alkyl heterocyclyl, C_1 - C_4 alkyl heteroaryl, C_1 - C_4 alkoxy aryl, C_1 - C_4 alkoxy heterocyclyl, C_1 - C_4 alkoxy heteroaryl, C_1 - C_4 alkoxy substituted with hydroxy, C_1 - C_4 alkoxy, amino and mono- and di- C_1 - C_4 alkylamino.

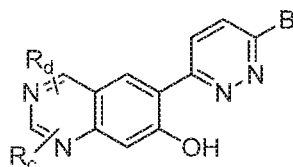
In a thirtieth splice modulator aspect, the splice modulator is a compound according to Formula (XIII):



(XIII)

or pharmaceutically acceptable salt thereof, wherein R_c and R_d are each, independently, selected from hydrogen, cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C₁-C₄alkyl, C₁-C₄alkyl aryl, C₁-C₄alkyl heterocyclyl, C₁-C₄alkyl heteroaryl, C₁-C₄alkoxy aryl, C₁-C₄alkoxy heterocyclyl, C₁-C₄alkoxy heteroaryl, C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino.

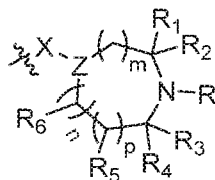
In a thirty-first splice modulator aspect, the splice modulator is a compound according to Formula (XIV):



(XIV)

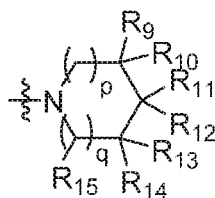
or pharmaceutically acceptable salt thereof, wherein R_c and R_d are each, independently, selected from hydrogen, cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, heterocyclyl, heteroaryl, heterocyclyl C₁-C₄alkyl, C₁-C₄alkyl aryl, C₁-C₄alkyl heterocyclyl, C₁-C₄alkyl heteroaryl, C₁-C₄alkoxy aryl, C₁-C₄alkoxy heterocyclyl, C₁-C₄alkoxy heteroaryl, C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino.

In a thirty-second splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through thirty-first splice modulator aspects, wherein B is a group of the formula:



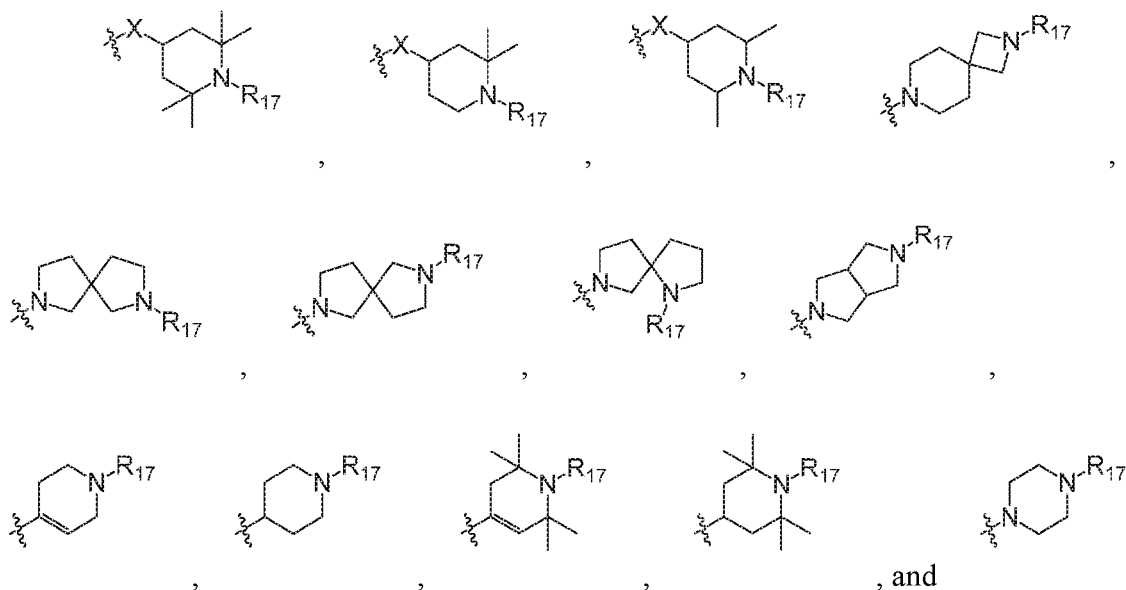
wherein m, n and p are independently selected from 0 or 1; R, R₁, R₂, R₃, and R₄ are independently selected from the group consisting of hydrogen, C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₅ and R₆ are hydrogen; or R and R₃, taken in combination form a fused 5 or 6 member heterocyclic ring having 0 or 1 additional ring heteroatoms selected from N, O or S; R₁ and R₃, taken in combination form a C₁-C₃alkylene group; R₁ and R₅, taken in combination form a C₁-C₃alkylene group; R₃ and R₄, taken in combination with the carbon atom to which they attach, form a spirocyclic C₃-C₆cycloalkyl; X is CR_AR_B, O, NR₇ or a bond; R_A and R_B are independently selected from hydrogen and C₁-C₄alkyl, or R_A and R_B, taken in combination, form a divalent C₂-C₅alkylene group; Z is CR₈ or N; when Z is N, X is a bond; R₈ is hydrogen or taken in combination with R₆ form a double bond.

In a thirty-third splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through thirty-second splice modulator aspects, wherein B is a group of the formula:



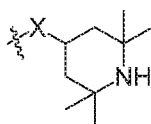
wherein p and q are independently selected from the group consisting of 0, 1, and 2; R₉ and R₁₃ are independently selected from hydrogen and C₁-C₄alkyl; R₁₀ and R₁₄ are independently selected from hydrogen, amino, mono- and di-C₁-C₄akylamino and C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₁₁ is hydrogen, C₁-C₄alkyl, amino or mono- and di-C₁-C₄akylamino; R₁₂ is hydrogen or C₁-C₄alkyl; or R₉ and R₁₁, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups; or R₁₁ and R₁₂, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups.

In a thirty-fourth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through thirty-third splice modulator aspects, wherein B is selected from the group consisting of:



wherein X is O or N(Me) or NH; and R₁₇ is hydrogen or methyl.

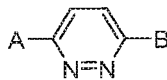
In a thirty-fifth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through thirty-fourth splice modulator aspects, wherein B is:



In a thirty-sixth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through thirty-fifth splice modulator aspects, wherein X is -O-.

In a thirty-seventh splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to any one of the nineteenth through thirty-sixth splice modulator aspects, wherein X is N(Me).

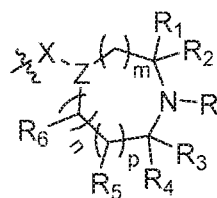
In a thirty-eighth splice modulator aspect, the splice modulator is a compounds according to Formula (XV):



(XV)

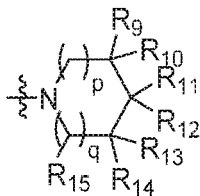
or pharmaceutically acceptable salt thereof, wherein A is 2-hydroxy-phenyl which is substituted with 0, 1, 2, or 3 substituents independently selected from C₁-C₄alkyl, wherein 2 C₁-C₄alkyl groups can combine with the atoms to which they are bound to form a 5-6 membered ring and is substituted with 0 or 1 substituents selected from oxo, oxime and hydroxy, haloC₁-C₄alkyl, dihaloC₁-C₄alkyl, trihaloC₁-C₄alkyl, C₁-C₄alkoxy, C₁-C₄alkoxy- C₃-C₇cycloalkyl, haloC₁-C₄alkoxy, dihaloC₁-C₄alkoxy, trihaloC₁-C₄alkoxy, hydroxy, cyano, halogen, amino, mono- and di-C₁-C₄alkylamino, heteroaryl, C₁-C₄alkyl substituted with hydroxy, C₁-C₄alkoxy substituted with aryl, amino, -C(O)NH C₁-C₄alkyl – heteroaryl, -NHC(O)- C₁-C₄alkyl- heteroaryl, C₁-C₄alkyl C(O)NH- heteroaryl, C₁-C₄alkyl NHC(O)- heteroaryl, 3-7 membered cycloalkyl, 5-7 membered cycloalkenyl or 5, 6 or 9 membered heterocycle containing 1 or 2 heteroatoms, independently, selected from S, O and N, wherein heteroaryl has 5, 6 or 9 ring atoms, 1, 2 or 3 ring heteroatoms selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from oxo, hydroxy, nitro, halogen, C₁-C₄alkyl, C₁-C₄alkenyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, C₁-C₄alkyl-OH, trihaloC₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, -C(O)NH₂, -NH₂, -NO₂, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, 4-7member heterocycleC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl; or A is 2-naphthyl optionally substituted at the 3 position with hydroxy and additionally substituted with 0, 1, or 2 substituents selected from hydroxy, cyano, halogen, C₁-C₄alkyl, C₂-C₄alkenyl, C₁-C₅alkoxy, wherein the alkoxy is unsubstituted or substituted with hydroxy, C₁-C₄alkoxy, amino, N(H)C(O)C₁-C₄alkyl, N(H)C(O)₂ C₁-C₄alkyl, alkylene 4 to 7 member heterocycle, 4 to 7 member heterocycle and mono-and di-C₁-C₄alkylamino; or A is 6 member heteroaryl having 1-3 ring nitrogen atoms, which 6 member heteroaryl is substituted by phenyl or a heteroaryl having 5 or 6 ring atoms, 1 or 2 ring heteroatoms independently selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from C₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and

di-C₁-C₄alkylaminoC₁-C₄alkyl; or A is bicyclic heteroaryl having 9 to 10 ring atoms and 1, 2, or 3 ring heteroatoms independently selected from N, O or S, which bicyclic heteroaryl is substituted with 0, 1, or 2 substituents independently selected from cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy and C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino; or A is tricyclic heteroaryl having 12 or 13 ring atoms and 1, 2, or 3 ring heteroatoms independently selected from N, O or S, which tricyclic heteroaryl is substituted with 0, 1, or 2 substituents independently selected from cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy, C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino, mono- and di-C₁-C₄alkylamino and heteroaryl, wherein said heteroaryl has 5, 6 or 9 ring atoms, 1, 2 or 3 ring heteroatoms selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from oxo, hydroxy, nitro, halogen, C₁-C₄alkyl, C₁-C₄alkenyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, C₁-C₄alkyl-OH, trihaloC₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, -C(O)NH₂, -NH₂, -NO₂, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, 4-7member heterocycleC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl; B is a group of the formula:



wherein m, n and p are independently selected from 0 or 1; R, R₁, R₂, R₃, and R₄ are independently selected from the group consisting of hydrogen, C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄alkylamino; R₅ and R₆ are independently selected from hydrogen and fluorine; or R and R₃, taken in combination form a fused 5 or 6 member heterocyclic ring having 0 or 1 additional ring heteroatoms selected from N, O or S; R₁ and R₃, taken in combination form a C₁-C₃alkylene group; R₁ and R₅, taken in combination form a C₁-C₃alkylene group; R₃ and R₄, taken in combination with the carbon atom to which they attach, form a spirocyclicC₃-C₆cycloalkyl; X is CR_AR_B, O, NR₇ or a bond; R₇ is hydrogen, or C₁-C₄alkyl; R_A and R_B are independently selected from hydrogen and C₁-C₄alkyl, or R_A and R_B, taken in combination, form a divalent C₂-C₅alkylene group; Z is CR₈ or N; when Z

is N, X is a bond; R₈ is hydrogen or taken in combination with R₆ form a double bond; or B is a group of the formula:



wherein p and q are independently selected from the group consisting of 0, 1, and 2; R₉ and R₁₃ are independently selected from hydrogen and C₁-C₄alkyl; R₁₀ and R₁₄ are independently selected from hydrogen, amino, mono- and di-C₁-C₄akylamino and C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₁₁ is hydrogen, C₁-C₄alkyl, amino or mono- and di-C₁-C₄akylamino; R₁₂ is hydrogen or C₁-C₄alkyl; or R₉ and R₁₁, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups; or R₁₁ and R₁₂, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups.

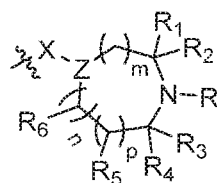
In a thirty-ninth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to the thirty-eighth splice modulator aspect, wherein A is 6 member heteroaryl having 1-3 ring nitrogen atoms, which 6 member heteroaryl is substituted by phenyl or a heteroaryl having 5 or 6 ring atoms, 1 or 2 ring heteroatoms independently selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from C₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl; or A is bicyclic heteroaryl having 9 to 10 ring atoms and 1, 2, or 3 ring heteroatoms independently selected from N, O or S, which heteroaryl is substituted with 0, 1, or 2 substituents independently selected from cyano, halogen, hydroxy, C₁-C₄alkyl, C₂-C₄alkenyl, C₂-C₄alkynyl, C₁-C₄alkoxy and C₁-C₄alkoxy substituted with hydroxy, C₁-C₄alkoxy, amino and mono- and di-C₁-C₄alkylamino.

In a fortieth splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to the thirty-eighth splice modulator aspect, wherein A is 2-hydroxy-phenyl which is substituted with 0, 1, 2, or 3 substituents independently

selected from C₁-C₄alkyl, haloC₁-C₄alkyl C₁-C₄alkoxy, hydroxy, cyano, halogen, amino, mono- and di-C₁-C₄alkylamino, heteroaryl and C₁-C₄alkyl substituted with hydroxy or amino, which heteroaryl has 5 or 6 ring atoms, 1 or 2 ring heteroatoms selected from N, O and S and substituted with 0, 1, or 2 substituents independently selected from C₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, 4-7member heterocycleC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl.

In a forty-first splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to the thirty-eighth splice modulator aspect, wherein A is 2-naphthyl optionally substituted at the 3 position with hydroxy and additionally substituted with 0, 1, or 2 substituents selected from hydroxy, cyano, halogen, C₁-C₄alkyl, C₂-C₄alkenyl, C₁-C₄alkoxy, wherein the alkoxy is unsubstituted or substituted with hydroxy, C₁-C₄alkoxy, amino, N(H)C(O)C₁-C₄alkyl, N(H)C(O)₂ C₁-C₄alkyl, 4 to 7 member heterocycle and mono- and di-C₁-C₄alkylamino; or

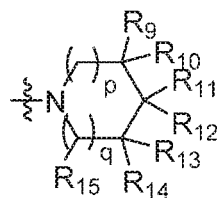
In a forty-second splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to the thirty-eighth through forty-first splice modulator aspects, wherein B is a group of the formula:



wherein m, n and p are independently selected from 0 or 1; R, R₁, R₂, R₃, and R₄ are independently selected from the group consisting of hydrogen, C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄alkylamino; R₅ and R₆ are hydrogen; or R and R₃, taken in combination form a fused 5 or 6 member heterocyclic ring having 0 or 1 additional ring heteroatoms selected from N, O or S; R₁ and R₃, taken in combination form a C₁-C₃alkylene group; R₁ and R₅, taken in combination form a C₁-C₃alkylene group; R₃ and R₄, taken in combination with the carbon atom to which they attach, form a spirocyclicC₃-C₆cycloalkyl; X is CR_AR_B, O, NR₇ or a bond; R_A and R_B are independently selected from hydrogen and C₁-C₄alkyl, or R_A and R_B, taken in combination, form a divalent C₂-

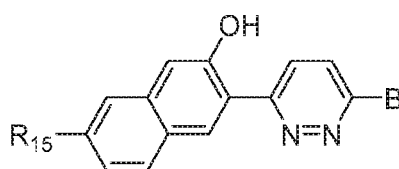
C₅alkylene group; Z is CR₈ or N; when Z is N, X is a bond; R₈ is hydrogen or taken in combination with R₆ form a double bond.

In a forty-third splice modulator aspect, the splice modulator is a compound or pharmaceutically acceptable salt thereof, according to the thirty-eighth through forty-first splice modulator aspects, wherein B is a group of the formula:



wherein p and q are independently selected from the group consisting of 0, 1, and 2; R₉ and R₁₃ are independently selected from hydrogen and C₁-C₄alkyl; R₁₀ and R₁₄ are independently selected from hydrogen, amino, mono- and di-C₁-C₄akylamino and C₁-C₄alkyl, which alkyl is optionally substituted with hydroxy, amino or mono- and di-C₁-C₄akylamino; R₁₁ is hydrogen, C₁-C₄alkyl, amino or mono- and di-C₁-C₄akylamino; R₁₂ is hydrogen or C₁-C₄alkyl; or R₉ and R₁₁, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups; or R₁₁ and R₁₂, taken in combination form a saturated azacycle having 4 to 7 ring atoms which is optionally substituted with 1-3 C₁-C₄alkyl groups.

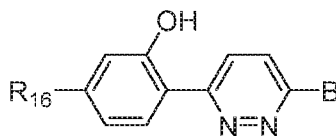
In a forty-fourth splice modulator aspect, the splice modulator is a compound according to Formula (XVI):



(XVI)

or pharmaceutically acceptable salt thereof, wherein R₁₅ is hydrogen, hydroxyl, C₁-C₄alkoxy, which alkoxy is optionally substituted with hydroxy, methoxy, amino, mono- and dimethylamino or morpholine.

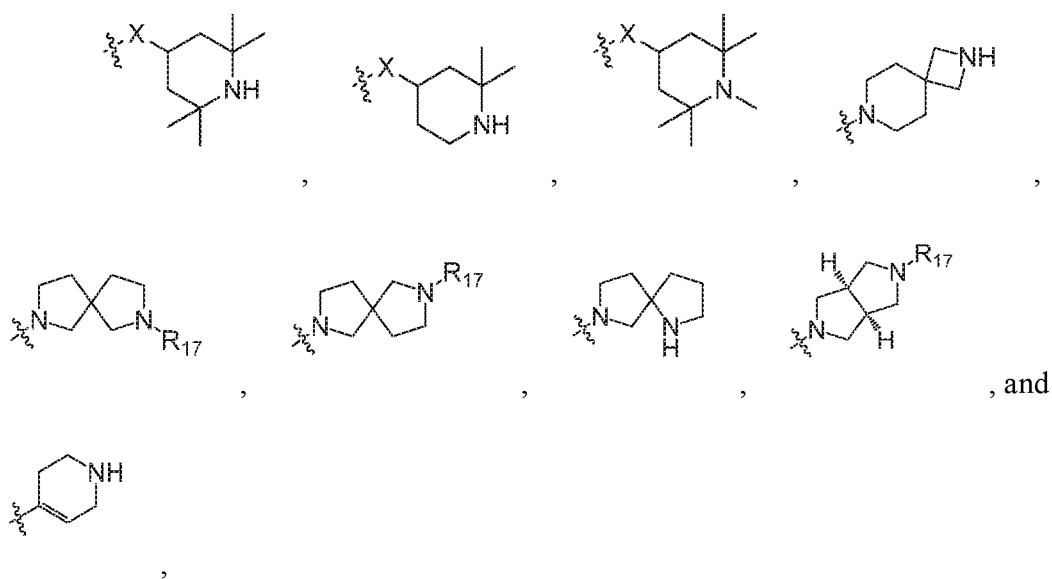
In a forty-fifth splice modulator aspect, the splice modulator is a compound according to Formula (XVII):



(XVII)

or pharmaceutically acceptable salt thereof, wherein R₁₆ is a 5 member heteroaryl having one ring nitrogen atom and 0 or 1 additional ring heteroatom selected from N, O or S, wherein the heteroaryl is optionally substituted with C₁-C₄alkyl.

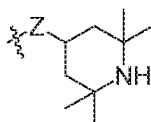
In a forty-sixth splice modulator aspect, the splice modulator is a compound according to of the thirty-eighth through forty-first, forty-fourth and forty-fifth splice modulator aspects, wherein B is selected from the group consisting of



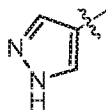
wherein X is O or N(Me); and R₁₇ is hydrogen or methyl.

In a forty-seventh splice modulator aspect, the splice modulator is a compound according to the thirty-eighth through forty-second and forty-fourth through forty-fifth splice modulator aspects, wherein X is -O-.

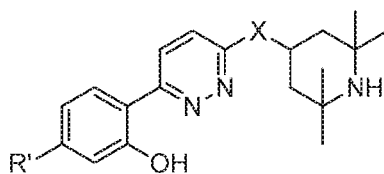
In a forty-eighth splice modulator aspect, the splice modulator is a compound according to the thirty-eighth through forty-second and forty-fourth through forty-fifth splice modulator aspects, wherein B is:



In a forty-ninth splice modulator aspect, the splice modulator is a compound according to the forty-fifth through forty-eighth splice modulator aspects, wherein R₁₆ is:



In a fiftieth splice modulator aspect, the splice modulator is a compound according to Formula (XVIII):



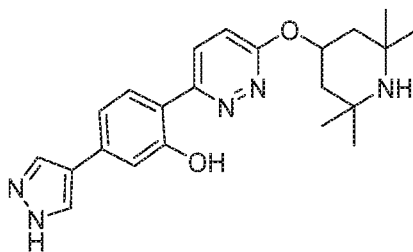
(XVIII)

or pharmaceutically acceptable salt thereof, wherein X is -O- or



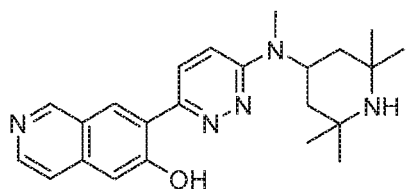
; R' is a 5-membered heteroaryl optionally substituted with 0, 1, or 2 groups selected from oxo, hydroxy, nitro, halogen, C₁-C₄alkyl, C₁-C₄alkenyl, C₁-C₄alkoxy, C₃-C₇cycloalkyl, C₁-C₄alkyl-OH, trihaloC₁-C₄alkyl, mono- and di-C₁-C₄alkylamino, -C(O)NH₂, -NH₂, -NO₂, hydroxyC₁-C₄alkylamino, hydroxyC₁-C₄alkyl, 4-7member heterocycleC₁-C₄alkyl, aminoC₁-C₄alkyl and mono- and di-C₁-C₄alkylaminoC₁-C₄alkyl.

In certain embodiments, the splice modulator is 5-(1H-Pyrazol-4-yl)-2-(6-((2,2,6,6-tetramethylpiperidin-4-yl)oxy)pyridazin-3-yl)phenol (LMI070; branaplam) having the following structure,



or a pharmaceutically acceptable salt thereof.

In certain embodiments, the splice modulator is splice modulator 2, wherein the compound is 7-(6-(methyl(2,2,6,6-tetramethylpiperidin-4-yl)amino)pyridazin-3-yl)isoquinolin-6-ol having the following structure,



or a pharmaceutically acceptable salt thereof.

Additional splice modulators and splice modulator binding sequences bound by those modulators are described in, for example, patent application publications US2012/0083495, WO2014/028459, WO2015/017589, WO2014/116845, WO2017/100726, WO2018/098446, WO2018/226622, WO2019/005993, WO2019/005980, and WO2019028440, the contents of which are hereby incorporated herein by reference in their entireties, and the splice modulators and splice modulator binding sequences described therein are contemplated for use in the methods, minigenes and other aspects and embodiments described herein.

Cleavage Sites

In aspects, the nucleic acid molecule of the invention includes one or more sequences encoding a cleavage site, which serves the function of cleaving the sequence (e.g., all the sequence or substantially all the sequence) encoded by the minigene from the sequence (e.g., protein of interest) encoded by the transgene. In aspects, the cleavage site can either be a self-cleavage site, a protease cleavage site or any combination thereof. The cleavage site can be designed to be cleaved by any site-

specific protease that is expressed in a cell of interest (either through recombinant expression or endogenous expression) at adequate levels to cleave off the sequence encoded by the one or more exons of the minigene from the protein of interest. In important aspects of the invention, the protease cleavage site is chosen to correspond to a protease natively (or by virtue of cell engineering) to be present in a cellular compartment relevant to the expression of the protein of interest. I.e., the intracellular trafficking of the protease should overlap or partially overlap with the intracellular trafficking of the protein of interest. For example, if the protein of interest is located at the cell surface, the enzyme to cleave it can be added exogenous to the cell.

If the protein of interest resides in or passes through the endosomal/lysosomal system a protease cleavage site for an enzyme resident in those compartments can be used. Such protease/consensus motifs include, e.g.,

Furin: RX(K/R)R consensus motif 1

Furin: RNRR (SEQ ID NO: 39)

PCSK1: RX(K/R)R consensus motif

PCSK5: RX(K/R)R consensus motif

PCSK6: RX(K/R)R consensus motif

PCSK7: RXXX[KR]R consensus motif

Cathepsin B : RRX

Granzyme B : I-E-P-D-X (SEQ ID NO: 35)

Factor XA: Ile – Glu/Asp-Gly-Arg

Enterokinase: Asp-Asp-Asp-Asp-Lys (SEQ ID NO: 36)

Genenase: Pro-Gly-Ala-Ala-His-Tyr (SEQ ID NO: 37)

Sortase: LPXTG/A

PreScission protease: Leu-Glu-Val-Phe-Gln-Gly-Pro (SEQ ID NO: 38)

Thrombin: Leu-Val-Pro-Arg-Gly-Ser (SEQ ID NO: 40)

TEV protease: E-N-L-Y-F-Q-G (SEQ ID NO: 41)

Elastase 1 [AGSV]-x (SEQ ID NO: 42)

In some embodiments, the nucleic acid described herein includes a sequence encoding a furin cleavage site. In some embodiments, the nucleic acids described herein include a sequence encoding

any one of the furin cleavage sites listed in Table 20. In embodiments, the furin cleavage site is SEQ ID NO: 39. In some embodiments, the nucleic acids described herein include a sequence encoding a furin cleavage site that includes or consists of SEQ ID NO: 39, for example, the sequence encoding a cleavage site includes or consists of SEQ ID NO: 19.

In some embodiments, the nucleic acids described herein include a sequence encoding a furin cleavage site selected from RNRR (SEQ ID NO: 39) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; RTKR (SEQ ID NO: 43) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; GTGAEDPRPSRKRRSLGDVG (SEQ ID NO: 45) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; GTGAEDPRPSRKRR (SEQ ID NO: 47) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; LQWLEQQVAKRRTKR (SEQ ID NO: 49) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; GTGAEDPRPSRKRRSLGG (SEQ ID NO: 51) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; GTGAEDPRPSRKRRSLG (SEQ ID NO: 53) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; SLNLTESHNSRKRR (SEQ ID NO: 55) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto; or CKINGYPKRGRKRR (SEQ ID NO: 57) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto.

In some embodiments, the nucleic acids described herein include a sequence encoding a furin cleavage site selected from RNRR (SEQ ID NO: 39); RTKR (SEQ ID NO: 43); GTGAEDPRPSRKRRSLGDVG (SEQ ID NO: 45); GTGAEDPRPSRKRR (SEQ ID NO: 47); LQWLEQQVAKRRTKR (SEQ ID NO: 49); GTGAEDPRPSRKRRSLGG (SEQ ID NO: 51); GTGAEDPRPSRKRRSLG (SEQ ID NO: 53); SLNLTESHNSRKRR (SEQ ID NO: 55); and CKINGYPKRGRKRR (SEQ ID NO: 57).

In some embodiments, the nucleic acids described herein include a sequence encoding a furin cleavage site selected from RNRR (SEQ ID NO: 39) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto. In some embodiments, the nucleic acids described herein include SEQ ID NO: 19, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto. In some embodiments, the nucleic acids described herein include SEQ ID NO: 19.

In some embodiments, the nucleic acids described herein include a sequence encoding a furin cleavage site selected from GTGAEDPRPSRKRRSLGDVG (SEQ ID NO: 45) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto, or GTGAEDPRPSRKRR (SEQ ID NO: 47) or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto. In some embodiments, the nucleic acids described herein include SEQ ID NO: 46 or SEQ ID NO: 48, or a sequence having at least 90%, 95%, 97%, 98%, or 99% identity thereto.

In some embodiments, the nucleic acids described herein include a sequence encoding a furin cleavage site selected from GTGAEDPRPSRKRRSLGDVG (SEQ ID NO: 45) or GTGAEDPRPSRKRR (SEQ ID

NO: 47). In some embodiments, the nucleic acids described herein include SEQ ID NO: 46 or SEQ ID NO: 48

In some embodiments, the nucleic acids described herein include a sequence encoding the furin cleavage site of GTGAEDPRPSRKRRSLGDVG (SEQ ID NO: 45).

Table 20. Exemplary furin cleavage sites and nucleic acid sequences encoding them.

	Amino acid sequence	Nucleic acid sequence
Furin cleavage site0	RNRR (SEQ ID NO: 39)	cgcaaccgcccgc (SEQ ID NO: 19)
Furin cleavage site1	RTKR (SEQ ID NO: 43)	cgtactaaaaga (SEQ ID NO: 44)
Furin cleavage site2	GTGAEDPRPSRKRRSLGDVG (SEQ ID NO: 45)	ggaaccggcgcggaagacccccggccctccaggaag cgaaggtccctcggagacgtgggt (SEQ ID NO: 46)
Furin cleavage site3	GTGAEDPRPSRKRR (SEQ ID NO: 47)	ggaaccggcgcggaagacccccggccctccaggaag cgaagg (SEQ ID NO: 48)
Furin cleavage site4	LQWLEQQVAKRRTKR (SEQ ID NO: 49)	ctgcaatggctggagcagcaggtggcgaagcggagaa ctaagcgg (SEQ ID NO: 50)
Furin cleavage site5	GTGAEDPRPSRKRRSLGG (SEQ ID NO: 51)	ggcacaggtgccgaggaccctcggccaagccgcaaaa ggaggtcacttggcggc (SEQ ID NO: 52)
Furin cleavage site6	GTGAEDPRPSRKRRSLG (SEQ ID NO: 53)	ggaaccggagcagaagatcccagaccaagccggaaa aggcgtccctgggt (SEQ ID NO: 54)
Furin cleavage site7	SLNLTESHNSRKRR (SEQ ID NO: 55)	agtctcaattgactgagtcacacaattccaggaagaaaa gg (SEQ ID NO: 56)
Furin cleavage site8	CKINGYPKRGRKRR (SEQ ID NO: 57)	tgcaagatcaacggctaccctaagaggggagcagaaagc ggcgg (SEQ ID NO: 58)

In some embodiments, the nucleic acid sequence comprising a minigene and a transgene, e.g., described herein, can include one or more sequences encoding a peptide cleavage sites (e.g., an self-cleaving peptide or a substrate for an intracellular protease). In embodiments, the sequence encoding a peptide cleavage site is disposed between the minigene and the transgene. Examples of self-cleaving peptide cleavage sites sequences include the following, wherein the GSG residues in parentheses are optional:

Table 21. Exemplary self-cleaving peptide sequences and nucleic acid sequences encoding them (GSG sequence in each is optional).

	Amino acid sequence	Nucleic acid sequence
T2A	(GSG) E G R G S L L T C G D V E E N P G P (SEQ ID NO: 59)	GGCAGCGGCGAAGGCCGCGGCAGCCT GCTGACCTGCGGCGATGTGGAAGAAAA CCCGGGCCCG (SEQ ID NO: 20)
T2A (without GSG)	E G R G S L L T C G D V E E N P G P (SEQ ID NO: 61)	GAAGGCCGCGGCAGCCTGCTGACCTG CGGCGATGTGGAAGAAAACCCGGGCC C (SEQ ID NO: 62)
P2A	(GSG) A T N F S L L K Q A G D V E E N P G P (SEQ ID NO: 63)	(ggcagcggc)gccaccaactcagcctgctgaagcagg ccggcgacgtggaggagaaccccgcccc (SEQ ID NO: 64)
E2A	(GSG) Q C T N Y A L L K L A G D V E S N P G P (SEQ ID NO: 65)	(ggcagcggc)cagtgcaccaactacgccctgctgaagc tggccggcgacgtggagagcaaccccgcccc (SEQ ID NO: 66)
F2A	(GSG) V K Q T L N F D L L K L A G D V E S N P G P (SEQ ID NO: 67)	(ggcagcggc)gtgaagcagaccctgaactcgacctgc tgaagctggccggcgacgtggagagcaaccccgcccc c (SEQ ID NO: 68)

In some embodiments, the nucleic acid molecule includes a sequence encoding a protease cleavage site, such as a furin cleavage site, and a sequence encoding a self-cleaving peptide, for example a 2A peptide, for example a T2A peptide. In embodiments, the nucleic acid comprises the sequence encoding the furin cleavage site 5' to the sequence encoding the 2A-encoding sequence. In embodiments, the furin cleavage site comprises or consists of SEQ ID NO: 39 and the T2A sequence comprises or consists of SEQ ID NO: 59 or SEQ ID NO: 61. In embodiments, the sequence encoding the furin cleavage site is or comprises SEQ ID NO: 19 and the sequence encoding the peptide cleavage site is or comprises SEQ ID NO: 20 or SEQ ID NO: 62. In embodiments, the sequence encoding the 2A sequence is disposed immediately 5' of the transgene (e.g., the sequence encoding the protein of interest), such that upon cleavage, fewer than 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 amino acids of the minigene, furin cleavage site and/or 2A peptide are left on the protein of interest.

Promoters

All cells in the animal or human body contain the same DNA, yet different cells in different tissues express, on the one hand, a set of common genes, and on the other, a set of genes that vary depending on the type of tissue and the stage of development. Without being bound by theory, any promoter that does not contain an intron can be used in the various aspects and embodiments (e.g., in the nucleic acid molecules) described herein. Exemplary promoters that can be used with the various aspects and embodiments described herein include, but are not limited to, the cytomegalovirus (CMV) promoter, the CAG promoter, the SV40 promoter, the JeT promoter, the PGK promoter and the chicken beta-actin promoter (CBA) promoter. In embodiments, the promoter is active in more than one cell type. In other

embodiments, the promoter is active in one cell type (e.g., cell-specific) or in cell types of one tissue (e.g., tissue-specific), such as, for example, central nervous tissue (e.g., brain tissue). In embodiments, the promoter is neuron specific. Examples of neuron specific promoters that can be used in the various aspects and embodiments described herein include, but are not limited to, isolated or synthetic neuron-specific promoters and functional fragments thereof used in vectors and other nucleic acids to drive expression of an operatively linked minigene and transgene, e.g., promoters derived from neuron-specific enolase (NSE) (see, e.g., EMBL HSENO2, X51956); an aromatic amino acid decarboxylase (AADC) promoter; a neurofilament promoter (see, e.g., GenBank HUMNFL, L04147); a synapsin promoter (see, e.g., GenBank HUMSYNIB, M55301); a thy-1 promoter (see, e.g., Chen et al., (1987) *Cell*, 51:7-19; Llewellyn et al. (2010) *Nat. Med.*, 16(10):1161-1166); a serotonin receptor promoter (see, e.g., GenBank S62283); a tyrosine hydroxylase promoter (TH) (see, e.g., Oh et al., (2009) *Gene Ther.*, 16:437; Sasaoka et al., (1992) *Mol. Brain Res.*, 16:274; Boundy et al., (1998) *J. Neurosci.*, 18:9989; and Kaneda et al., (1991) *Neuron*, 6:583-594); a GnRH promoter (see, e.g., Radovick et al., (1991) *Proc. Natl. Acad. Sci. USA*, 88:3402-3406); an L7 promoter (see, e.g., Oberdick et al., (1990) *Science*, 248:223-226); a DNMT promoter (see, e.g., Bartge et al., (1988) *Proc. Natl. Acad. Sci. USA*, 85:3648-3652); an enkephalin promoter (see, e.g., Comb et al., (1988) *EMBO J.*, 17:3793-3805); a myelin basic protein (MBP) promoter; a Ca²⁺-calmodulin-dependent protein kinase II-alpha (CamKIM) promoter (see, e.g., Mayford et al., (1996) *Proc. Natl. Acad. Sci. USA*, 93:13250; and Casanova et al., (2001) *Genesis*, 31:37); a CMV enhancer/platelet-derived growth factor-p promoter (see, e.g., Liu et al., (2004) *Gene Ther.*, 11:52-60); and the like. In some embodiments, portions or all of the minimal human synapsin 1 promoter (SYN) are used. Kugler et al., (2003) *Gene Ther.*, 10(4): 337-47; Thiel et al., (1991) *Proc. Natl. Acad. Sci. USA*, 88(8) 3431-5; Castle et al., (2016) *Methods Mol. Biol.*, 1382: 133-49; McLean et al., (2014) *Neurosci. Lett.*, 576: 73-78; Kugler et al., (2003) *Virology*, 311(1): 89-95.

In some embodiments, a tissue- or cell-specific promoter is configured to provide higher expression of an operatively linked minigene and/or transgene in a neuronal cell or tissue relative to that in a non-neuronal cell. In some embodiments, the neuron specific promoter is configured to provide higher expression of an operatively linked minigene and/or transgene in a neuron relative to that in a non-neuronal cell. Examples of neuronal cells or tissue include those comprising neurons, as well as Schwann cells, glial cells, astrocytes, etc. Examples of non-neuronal cells include, but are not limited to, hepatic cells, cardiomyocytes, red blood cells, epithelial cells etc. Higher levels of expression of an operatively linked minigene and/or transgene may include an increase in the number of RNA transcripts produced from transcription of the minigene and/or transgene. In some embodiments, the number of RNA transcripts produced may be measured by PCR. In some other embodiments, the number of RNA transcripts produced may be measured by RT-PCR, e.g., qPCR. In some embodiments, the number of RNA transcripts produced may be measured by sequencing. In some embodiments, the number of RNA transcripts produced may be measured by single-molecule Fluorescence In-Situ Hybridization (FISH). In some embodiments, the number of RNA transcripts produced may be measured by Northern blot

analysis. Higher levels of expression of an operatively linked minigene and/or transgene may alternatively or in addition include an increase in the amount of protein produced, when the minigene and/or transgene encodes a protein of interest. In some embodiments, the amount of protein produced may be measured by an enzyme-linked immunosorbent assay (ELISA). In some embodiments, the amount of protein produced may be measured by Western blot analysis. In some embodiments, the amount of protein produced may be measured by immunostaining. In some embodiments, the amount of protein produced may be measured by time-resolved Forster Resonance Energy Transfer (TR-FRET). In some embodiments, the amount of protein produced may be measured by immunohistochemistry (IHC). In some embodiments, the level of expression is measured by more than one of these or other methods.

In aspects and embodiments, the promoter is a JeT promoter comprising SEQ ID NO: 13. In aspects and embodiments, the promoter is a human synapsin promoter comprising SEQ ID NO: 86.

Poly A signal sequence

In various embodiments, the nucleic acids, vectors and other compositions disclosed herein may comprise one or more polyadenylation (PolyA) signal sequences. The polyadenylation signal sequences may comprise a central sequence (e.g., AAUAAA) flanked by auxiliary sequence elements. Without being bound by theory, the sequence may signal the end of the transcript and serve as the site where a homopolymeric A sequence is added on the 3' end by polyadenylate polymerase.

Polyadenylation signal sequences known in the art are contemplated, including but not limiting to the SV40 polyA, the human growth hormone (HGH) polyA, the bovine growth hormone (BGH) polyA, the beta-globin polyA, the alpha-globin polyA, the ovalbumin polyA, the kappa-light chain polyA, and a synthetic polyA. PolyA signal sequences may be used in the nucleic acids and other compositions disclosed herein. In some embodiments, the polyA sequence in the transgene or nucleic acid sequence consists of SEQ ID NO: 22 or a functional fragment thereof. In some embodiments, the transgene or nucleic acid sequence comprises a sequence having at least about 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO: 22 or a functional fragment thereof. In some embodiments, the polyA sequence in the transgene or nucleic acid consists of a sequence of at least about 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO: 22 or a functional fragment thereof. In some embodiments, the polyA sequence in the transgene or nucleic acid sequence consists of SEQ ID NO: 89 or a functional fragment thereof. In some embodiments, the transgene or nucleic acid sequence comprises a sequence having at least about 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO: 89 or a functional fragment thereof. In some embodiments, the polyA sequence in the transgene or nucleic acid consists of a sequence of at least about 80, 85, 90, 95, 98, or 99% identity to SEQ ID NO: 89 or a functional fragment thereof.

Post-Transcriptional Regulatory Elements

In various embodiments, the nucleic acids, transgenes, and other compositions disclosed herein may comprise one or more post-transcriptional regulatory elements (PREs), e.g., those that can enhance

or otherwise improve expression of the transgene. Without being bound by the theory, PREs may enhance expression by enabling stability and 3' end formation of mRNA, and/or may facilitate the nucleocytoplasmic export of unspliced mRNAs. PREs may also comprise binding sites for RNA-binding proteins (RBPs) or microRNAs.

Exemplary PREs include but are not limited to a PRE from the Hepatitis B virus (HPRE), bat virus (BPRE), ground squirrel virus (GSPRE), arctic squirrel virus (ASPRE), duck virus (DPRE), chimpanzee virus (CPRE) woolly monkey virus (WMPRE) or woodchuck virus (WPRE). In some embodiments, the nucleic acid or transgene comprises a PRE. In certain embodiments, the PRE comprises the HPRE.

In some embodiments, a synthetic PRE is used. An example sequence of a synthetic PRE includes the sequence of the HPRE-NOX SEQ ID NO: 88, or a fragment thereof. In some embodiments, PREs may be disposed downstream (or 3' to) a promoter element.

Exemplary PREs also include, but are not limited to, a PRE comprising, e.g., consisting of, SEQ ID NO: 72, or a fragment thereof. Exemplary PREs also include, but are not limited to, a PRE comprising, e.g., consisting of, SEQ ID NO: 73, or a fragment thereof.

ACAGGCCTATTGATTGGAAAGTATGTCAACGAATTGTGGGTCTTTTGGGGTTTGCTGCCCCCT
TTTACGCAATGTGGATATCCTGCTTTAATGCCTTTATATGCATGTATACAAGCAAAACAGGCTTTTACT
TTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGTATCTGACCCTTTACCCCGTTGCTCGGCAACG
GCCTGGTCTGTGCCAAGTGTGGTCTGACGCAACCCCCACTGGTTGGGGCTTGCCATAGGCCATCA
GCGCATGCGTGGAACCTTTGTGTCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTT
GCTCGCAGCAGGTCTGGAGCGAACTCATCGGGACTGACAATTCTGTCTGTCTCTCCCGCAAGTAT
ACATCGTTTCCAGGGCTGCTAGGCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTAC
GTCCCGTCTGGCGCTGAATCCCGCGGACGACCCCTCCCGGGGCGCTTGGGGCTCTACCGCCCGCT
TCTCCGTCTGCCGTACCGACCGACCGGGGCGCACCTCTCTTTACGCGGACTCCCGTCTGTGCC
TTCTCATCTGCCGACCGTGTGCACTTCGCTTCACCTCTGCACGTGCGATGGAGACCACCGTGAAC
GCCCACCGGAACCTGCCCAAGGTCTTGCATAAGAGGACTCTTGGACTTTCAGCAATGTC (SEQ ID
NO: 72)

AACAGGCCTATTGATTGGAAAGTATGTCAACGAATTGTGGGTCTTTTGGGGTTTGCTGCCCC
TTTTACGCAATGTGGATATCCTGCTTTAATGCCTTTATATGCATGTATACAAGCAAAACAGGCTTTTAC
TTTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGTATCTGACCCTTTACCCCGTTGCTCGGCAAC
GGCCTGGTCTGTGCCAAGTGTGGTCTGACGCAACCCCCACTGGTTGGGGCTTGCCATAGGCCATC
AGCGCATGCGTGGAACCTTTGTGTCTCCTCTGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTT
TGCTCGCAGCAGGTCTGGAGCGAACTCATCGGGACTGACAATTCTGTCTGTCTCTCCCGCAAGTA
TACATCGTTTCCAGGGCTGCTAGGCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTAC
GTCCCGTCTGGCGCTGAATCCCGCGGACGACCCCTCCCGGGGCGCTTGGGGCTCTACCGCCCGCT
TCTCCGTCTGCCGTACCGACCGACCGGGGCGCACCTCTCTTTACGCGGACTCCCGTCTGTGCC

TTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCTCTGCACGTCGCATGGAGACCACCGTGAAC
GCCCACCAGAACCTGCCCAAGGTCTTGCATAAGAGGACTCTTGGACTTTCAGCAATGTC (SEQ ID
NO: 73)

Exemplary PREs also include a PRE comprising or consisting of sequence with at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to a PRE described herein, e.g., to SEQ ID NO: 88, SEQ ID NO: 72 or SEQ ID NO: 73.

In some embodiments, PREs may be disposed downstream (or 3' to) a transgene sequence or protein-coding sequence. In some embodiments, PREs may be disposed upstream of (or 5' to) a polyA sequence. In some embodiments, PREs may be disposed upstream of (or 5' to) a transgene sequence or protein-coding sequence.

Transgenes

In various embodiments, the minigenes and other regulatory elements disclosed herein may be used to regulate expression of an operably linked transgene. In some embodiments, the transgene encodes a protein such as an antibody or functional binding fragment, a receptor, an enzyme, etc. In some embodiments, the transgene encodes a therapeutic nucleic acid such as an shRNA, siRNA, gRNA for use in CRISPR, etc. In some embodiments, more than one transgene may be used (e.g., a nucleic acid or vector may encode more than one protein or RNA that provides therapeutic benefits). Examples of methods to increase levels of these functional polypeptides or nucleic acids in cells include transfection or transduction of a nucleic acid sequence encoding the polypeptide of interest, e.g., in a nucleic acid or vector disclosed herein, e.g., an AAV viral vector.

i. Proteins

In various embodiments, the minigenes and other regulatory elements disclosed herein may be used to regulate (e.g., turn on or turn off, in the presence or absence of a splice modulator) expression of polypeptides. Without being bound by theory, increases in the level of polypeptides in may provide therapeutic effects by providing for a polypeptide whose expression is reduced or missing in a subject patient's tissue. Without being bound by theory, controlling the timing or location of expression, e.g., by application or withdrawal of a splice modulator, may improve the effectiveness and/or safety of such therapeutic protein by ensuring expression only when it is wanted. Exemplary polypeptides that may be regulated by the minigenes described herein include but are not limited to superoxide dismutase, aromatic acid decarboxylase (AADC), survival of motor neuron (SMN) protein, progranulin (PRGN), a Cas9 protein, a zinc finger nuclease or a TALEN, or a therapeutic protein such as, for example, a protein selected from MeCP2, CLN2, CLN3, CLN4, CLN5, CLN6, CLN7, and CLN8, or a protein related to spinocerebellar ataxia (SCA), optionally any of SCA1-SCA29.

In various embodiments, the minigenes and other regulatory elements disclosed herein may be used to regulate expression of progranulin (PGRN). Without being bound by theory, increases in the level of functional PGRN polypeptides in neurons may provide therapeutic effects, e.g., in the treatment of FTD. Without being bound by theory, PGRN is typically observed in humans as a ubiquitously expressed, 88 kDa secreted glycoprotein. It is encoded by the human granulin gene (GRN). Exemplary nucleic acids encoding the progranulin protein include NG_007886.1 and NM_002087.3 as defined by RefSeqGene, and NC_000017.11 and NC_000017.10 as defined by NCBI Reference Sequences. Exemplary progranulin polypeptide sequences include NP_002078.1. In some embodiments, the progranulin polypeptide contains seven granulin-like domains, which consist of highly conserved tandem repeats of a rare 12 cysteinyl motif (SEQ ID NO: 102) connected by linker sequences.

In some embodiments, peptide fragments of PGRN and nucleic acids encoding them are encompassed by the term PGRN to the extent they retain one or more function of PGRN. Cleavage of PGRN to form granulins (GRNs) or epithelins may produce proteins with different function and are outside the meaning of fragments of PGRN as used herein. In some embodiments, a nucleic acid, vector, or other composition disclosed herein comprises a transgene sequence encoding a human protein. In some embodiments, the transgene sequence encodes PGRN. In some embodiments, the transgene sequence encodes a human progranulin (hPGRN) protein. In some embodiments, the transgene sequence encodes a codon-optimized version of the hPGRN protein. In some embodiments, the transgene sequence comprises a sequence of SEQ ID NO: 87 or a functional fragment thereof, e.g., a fragment capable of providing detectable changes in one or more of the functions provided by intact PGRN. In some embodiments, the transgene sequence comprises a sequence with at least 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80%, 75%, or 70% sequence identity (or any percentage in between) to SEQ ID NO: 87. In some embodiments, the hPGRN encoded by the transgene comprises an amino acid sequence of SEQ ID NO: 87. In some embodiments, the hPGRN encoded by the heterologous nucleic acid sequence comprises a sequence with at least 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80%, 75%, or 70% sequence identity comprises an amino acid sequence of SEQ ID NO: 81.

	Sequence
hPGRN Nucleic acid SEQ ID NO: 87	TGGACCCTGGTGAGCTGGGTGGCCTTAACAGCAGGGCTGGTGGCTGGAAC GCGGTGCCCAGATGGTCAGTTCTGCCCTGTGGCCTGCTGCCTGGACCCCG GAGGAGCCAGCTACAGCTGCTGCCGTCCCCTTCTGGACAAATGGCCCACA AACTGAGCAGGCATCTGGGTGGCCCCTGCCAGGTTGATGCCCACTGCTC TGCCGGCCACTCCTGCATCTTTACCGTCTCAGGGACTTCCAGTTGCTGCCC CTTCCCAGAGGCCGTGGCATGCGGGGATGGCCATCACTGCTGCCACGG GGCTTCCACTGCAGTGCAGACGGGCGATCCTGCTTCCAAAGATCAGGTAA CAACTCCGTGGGTGCCATCCAGTGCCCTGATAGTCAGTTCTGAATGCCCGG ACTTCTCCACGTGCTGTGTTATGGTCGATGGCTCCTGGGGGTGCTGCCCA TGCCCCAGGCTTCCTGCTGTGAAGACAGGGTGCACTGCTGTCCGCACGGT GCCTTCTGCGACCTGGTTCACACCCGCTGCATCACACCCACGGGCACCCA

	CCCCCTGGCAAAGAAGCTCCCTGCCAGAGGACTAACAGGGCAGTGGCCT TGTCCAGCTCGGTCATGTGTCCGGACGCACGGTCCCGGTGCCCTGATGGT TCTACCTGCTGTGAGCTGCCCAGTGGGAAGTATGGCTGCTGCCCAATGCC CAACGCCACCTGCTGCTCCGATCACCTGCACTGCTGCCCCCAAGACACTGT GTGTGACCTGATCCAGAGTAAGTGCCTCTCCAAGGAGAACGCTACCACGG ACCTCCTCACTAAGCTGCCTGCGCACACAGTGGGGGATGTGAAATGTGACA TGGAGGTGAGCTGCCAGATGGCTATACCTGCTGCCGTCTACAGTCGGGG GCCTGGGGCTGCTGCCCTTTTACCCAGGCTGTGTGCTGTGAGGACCACAT AACTGCTGTCCCGCGGGGTTTACGTGTGACACGCAGAAGGGTACCTGTG AACAGGGGCCCCACCAGGTGCCCTGGATGGAGAAGGCCCCAGCTCACCTC AGCCTGCCAGACCCACAAGCCTTGAAGAGAGATGTCCCCTGTGATAATGTC AGCAGCTGTCCCTCCTCCGATACCTGCTGCCAACTCACGTCTGGGGAGTG GGGCTGCTGTCCAATCCCAGAGGCTGTCTGCTGCTCGGACCACCAGCACT GCTGCCCCCAGGGCTACACGTGTGTAGCTGAGGGGCAGTGTACGCGAGG AAGCGAGATCGTGGCTGGACTGGAGAAGATGCCTGCCCGCCGGGCTTCCT TATCCACCCCAGAGACATCGGCTGTGACCAGCACACCAGCTGCCCGGTG GGGCAGACCTGCTGCCCGAGCCTGGGTGGGAGCTGGGCCTGCTGCCAGT TGCCCCATGCTGTGTGCTGCGAGGATCGCCAGCACTGCTGCCCGGCTGGC TACACCTGCAACGTGAAGGCTCGATCCTGCGAGAAGGAAGTGGTCTCTGC CCAGCCTGCCACCTTCCTGGCCCGTAGCCCTCACGTGGGTGTGAAGGACG TGGAGTGTGGGGAAGGACACTTCTGCCATGATAACCAGACCTGCTGCCGA GACAACCGACAGGGCTGGGCCTGCTGTCCCTACCGCCAGGGCGTCTGTTG TGCTGATCGGCGCCACTGCTGTCCCTGCTGGCTTCCGCTGCGCAGCCAGGG GTACCAAGTGTTTGCGCAGGGAGGCCCGCGCTGGGACGCCCCCTTTGAGG GACCCAGCCTTGAGACAGCTGCTGTAG
hPGRN Amino acid SEQ ID NO: 81	MWTLVSWWALTAGLVAGTRCPDQGFCPVACCLDPGGASYSCCRPLLDKWPTTLRHL GGPCQVDAHCSAGHSIFTVSGTSSCCPFPEAVACGDGHHCCPRGFHCSADGRSCF QRSGNNSVGAIQCPDSQFECPDFSTCCVMVDGSWGCCPMPQASCCEDRVHCCPHG AFCDLVHTRCITPTGTHPLAKKLPQRTNRAVALSSVMCPDARSRCPDGSTCCCLPS GKYGCCPMPNATCCSDHLHCCPQDVTCDLIQSKCLSKENATDILLTKLPAHTVGDVKG DMEVSCPDGYTCCRLQSGAWGCCPFTQAVCCEDHIHCCPAGFTCDTQKGTCEQGPH QVPWMEKAPAHLSLPDPQALKRDVPCDNVSSCPSSDTCCQLTSGEWGCCPIPEAVCC SDHQHCCPQGYTCVAEGQCQRGSEIVAGLEKMPARRASLSHPRDIGCDQHTSCPVGQ TCCPSLGGSWACCQLPHAVCCEDRQHCCPAGYTCNVKARSCEKEVVSAQPATFLARS PHVGKDVCEGEGHFCHDNQTCCRDNRQGWACCPYRQGVCCADRRHCCPAGFRCA ARGTKCLRREAPRWDAPLRDPALRQLL

ii. RNA

In various embodiments, the isolated nucleic acids, vectors, and other compositions disclosed herein may comprise a transgene sequence encoding a sequence that provides neuronal tissue-specific therapeutic effects without requiring protein translation. In some embodiments, the promoters, silencers, regulatory elements, and other nucleic acid elements disclosed herein may be used to regulate neuronal tissue or neuron-specific expression of RNA. In some embodiments, the transgene sequence encodes a ribonucleic acid providing a particular therapeutic function. In some embodiments, the transgene sequence encodes a siRNA. In some embodiments, the transgene sequence encodes a shRNA. In some

embodiments, the transgene sequence encodes an miRNA. In some embodiments, the transgene sequence encodes a tRNA.

iii. Antibody

In various embodiments, the promoters, silencers, and other regulatory elements disclosed herein may be used to regulate neuronal tissue or neuron-specific expression of antibodies or fragments thereof. For instance, in some embodiments, the transgene sequence encodes an antibody. In some embodiments, the transgene sequence encodes a fragment of an antibody, e.g., one that retains antigen-binding capabilities. In some embodiments, the transgene sequence encodes a light chain of an antibody. In some embodiments, the transgene sequence encodes a heavy chain of an antibody. In some embodiments, the transgene sequence encodes a V_H . In some embodiments, the transgene sequence encodes a V_L . In some embodiments, the transgene sequence encodes a V_H . In some embodiments, the transgene sequence encodes a Fab. In some embodiments, the transgene sequence encodes a scFv. In some embodiments, the transgene sequence encodes an enzyme with neuron-specific function.

iii. More than one component

In various embodiments, the promoters, silencers, and other regulatory elements disclosed herein may be used to regulate neuronal tissue or neuron-specific expression of more than one transgene. In some embodiments, the transgene sequences encode both an RNA and a polypeptide. In some embodiments, the transgene sequence encodes components of a CRISPR/Cas system. In some embodiments, the transgene sequence encodes a Cas9 protein. In some embodiments, the transgene sequence encodes a Cpf1 protein. In some embodiments, the transgene sequence encodes a CRISPR RNA (crRNA). In some embodiments, the transgene encodes a transactivating crRNA (tracrRNA).

In some embodiments, a nucleic acid, vector, or other composition disclosed herein comprises a minigene (e.g., as described herein), a transgene sequence encoding hPGRN, a PRE, and a polyA signal sequence, e.g., present in that order from 5' to 3'. In some embodiments, the nucleic acid, vector, or other composition comprises, from 5' to 3', a promoter, a minigene, a sequence encoding a protease cleavage site (e.g., a furin cleavage site), a sequence encoding a self-cleaving peptide (e.g., a T2A peptide), a transgene sequence encoding hPGRN, a PRE, and a polyA signal sequence. In some embodiments, the nucleic acid, vector, or other composition comprises, from 5' to 3', a promoter, a minigene comprising SEQ ID NO: 16 (e.g., a minigene comprising or consisting of SEQ ID NO: 71 or SEQ ID NO: 94), a sequence encoding a furin cleavage site comprising or consisting of SEQ ID NO: 19, a sequence encoding a self cleaving T2A peptide comprising or consisting of SEQ ID NO: 20, a transgene encoding PRGN (e.g., SEQ ID NO: 87), a PRE sequence comprising SEQ ID NO: 88, and a polyA sequence (e.g., a polyA comprising or consisting of SEQ ID NO: 89).

In any of the aforementioned aspects and embodiments, the nucleic acid sequences contemplated may be DNA, RNA, or modified versions thereof. Modified nucleic acids may be

distinguished from naturally occurring nucleic acids by modifications to the backbone of the polynucleotide chain, for example, peptide nucleic acids (PNA), morpholinos, locked nucleic acids (LNA), glycol nucleic acids (GNA) and threose nucleic acid (TNA). Modified nucleic acids may also include analogs with modifications to the four nucleobases. In some embodiments, the nucleic acids are PNAs. In some embodiments, the nucleic acids are LNAs. In some embodiments, the nucleic acids are morpholinos. In some embodiments, the nucleic acids are in a single-stranded form. In some embodiments, the nucleic acids are in double-stranded form. In some embodiments, the nucleic acids are linear. In some embodiments, the nucleic acids are circular. In some embodiments, the nucleic acids are plasmids.

Viral vectors

Also disclosed herein are vectors comprising the nucleic acids (e.g., minigenes, transgenes, other nucleic acid components such as promoters, PREs and polyAs, and combinations thereof) discussed herein. In some embodiments, a vector may serve to deliver a transgene to a target cell and/or to increase expression of that transgene in a target cell. In various embodiments, the vector may be used to regulate expression of proteins, antibodies or functional binding fragments, enzymes, etc., and/or nucleic acids, e.g., shRNA, siRNA, gRNA for use in CRISPR, etc., through use in combination with a splice modulator

For instance, a vector may comprise a “on-switch” minigene linked to a transgene encoding a therapeutic protein and/or RNA and, upon addition of a splice modulator, increase the expression of that transgene. In other embodiments, a vector may comprise an “off-switch” minigene linked to a transgene encoding a therapeutic protein and/or RNA and, upon addition of a splice modulator, decrease the expression of that transgene. In some embodiments, the vector may comprise a DNA or RNA (or a mixture thereof) sequence that comprises an insert (e.g., at least one open reading frame of a transgene sequence) and one or more additional elements. The vector may serve to transfer genetic information to another cell. Vectors may be used for cloning, e.g., as cloning vectors or plasmids. Vectors may also be designed specifically for other purposes, such as cellular infection, e.g., in a human neuronal cell, to drive expression, e.g., therapeutic protein and/or RNA expression. In some embodiments, vectors comprising the nucleic acids disclosed herein are contemplated. The vectors may be a DNA vector, a circular vector, or a plasmid. In some embodiments, the vector is double stranded. In other embodiments the vector is single stranded.

In some embodiments, the vector is a viral vector. In some embodiments, the vector is a viral vector used to deliver transgene sequence(s) to neuronal cells or tissue. Examples of viruses used for vectors include but are not limited to retroviruses, adenoviruses, lentiviruses, adeno-associated viruses, and other hybrid viruses. In some embodiments, the viral vector is an adeno-associated viral (AAV) vector, chimeric AAV vector, adenoviral vector, retroviral vector, lentiviral vector, DNA viral vector, herpes simplex viral vector, baculoviral vector, or any mutant or derivative thereof.

Without being bound by theory, viral vectors disclosed herein may insert their genomes into the host cell that they infect, thus delivering its nucleic acid sequence to the host. The viral genome inserted may be episomal or may be integrated into the chromosomes of the host cell at a site that may be random or targeted. In an embodiment, the vector is a viral vector used to deliver transgene sequences to cells. Examples of viruses used for vectors include but are not limited to retroviruses, adenoviruses, lentiviruses, adeno-associated viruses, and other hybrid viruses. Warnock et al., (2011) *Methods Mol. Biol.*, 737:1-25. Lentivirus is a genus of retroviruses that can integrate significant amounts of viral DNA into a host cell, making them an efficient method of gene delivery. On the other hand, adenoviruses introduce genetic material that is not integrate into the chromosome of the host cell, thus reducing the risk of disrupting the host cell. In some embodiments, the viral vector is an adeno-associated viral (AAV) vector, chimeric AAV vector, adenoviral vector, retroviral vector, lentiviral vector, DNA viral vector, herpes simplex viral vector, baculoviral vector, or any mutant or derivative thereof.

In some embodiments, the vector comprising the transgene is or is derived from an adeno-associated virus (AAV). In some embodiments, the vector is a recombinant adeno-associated viral vector (rAAV). The rAAV genomes may comprise one or more AAV ITRs flanking a minigene and transgene sequence encoding a polypeptide (including, but not limited to, a hPGRN polypeptide) or encoding siRNA, shRNA, antisense, and/or miRNA directed at mutated proteins or control sequences of their genes. The minigene and transgene sequences are operatively linked, and may be linked by sequence encoding one or more protease cleavage sites or sequences encoding one or more self-cleaving peptides, or combinations thereof. In embodiments, the vectors additionally comprise other transcriptional control elements such as those disclosed herein, e.g., promoter, enhancer, , PRE, and/or polyA sequences that are functional in target cells to drive expression of the transgene sequence. The transgene sequence may also include intron sequences to facilitate processing of an RNA transcript when expressed in mammalian cells.

In various embodiments, the AAV vector, e.g., the rAAV vector, is a self-complementary AAV vector (scAAV). As used herein, "self-complementary" means the coding region has been designed to form an intra-molecular double-stranded template, e.g., in one or more inverted terminal repeats (ITRs). Without being bound by theory, a rate-limiting step for AAV genome often involves the second-strand synthesis since the typical AAV genome is a single-stranded DNA template. Ferrari et al, (1996) *J. Virology*, 70(5): 3227-34; Fisher et al, (1996) *J. Virology*, 70(1): 520-32. However, for scAAV genomes, upon infection, the two complementary halves of scAAV may associate to form one double stranded DNA (dsDNA) unit that is ready for replication and transcription rather than waiting for cell mediated synthesis of the second strand. In some embodiments, the rAAV vector disclosed herein is a scAAV vector and provides for faster and/or increased expression.

In some embodiments, the rAAV vectors disclosed herein lack one or more (e.g., all) AAV rep and/or cap genes. An AAV vector may comprise (e.g., in its ITRs) nucleic acid sequences (e.g., DNA)

from any suitable AAV serotype. Suitable AAV serotypes include, but are not limited to, AAV serotypes AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10, AAV-11, AAV-12, AAVrh8, AAVrh10, AAV.Anc80, AAV.Anc80L65, AAV-DJ, and AAV-DJ/8, AAVrh37, AAV-DJ, AAV-DJ/8, AAV-PHP.B, AAV-PHP.B2, AAV-PHP.B3, AAV-PHP.A, AAV-PHP.eB, and AAV-PHP.S. For instance, an AAV vector, e.g., an scAAV vector, may comprise nucleic acid sequences from an AAV2, e.g., ITR sequences from an AAV2. An AAV vector, e.g., an scAAV vector, may also comprise nucleic acids from more than one serotype. The nucleotide sequences of the genomes of the AAV serotypes are known in the art. For example, the complete genome of AAV1 is provided in GenBank Accession No. NC_002077; the complete genome of AAV2 is provided in GenBank Accession No. NC 001401 and Srivastava et al., *Virology*, 45: 555-564 {1983}; the complete genome of AAV3 is provided in GenBank Accession No. NC_1829; the complete genome of AAV4 is provided in GenBank Accession No. NC_001829; the AAV5 genome is provided in GenBank Accession No. AF085716; the complete genome of AAV-6 is provided in GenBank Accession No. NC_001862; at least portions of AAV7 and AAV8 genomes are provided in GenBank Accession Nos. AX753246 and AX753249, respectively; the AAV9 genome is provided in Gao et al., *J. Virol.*, 78: 6381-6388 (2004); the AAV10 genome is provided in Williams, (2006) *Mol. Ther.*, 13(1): 67-76; and the AAV11 genome is provided in Mori et al., (2004) *Virology*, 330(2): 375-383.

In some embodiments, functional inverted terminal repeat (ITR) sequences may be used to support, e.g., the rescue, replication and packaging of the AAV virion. Thus, an AAV vector disclosed herein may include sequences that in cis provide for replication and packaging (e.g., functional ITRs) of the virus. The ITRs can be but need not be the wild-type nucleotide sequences, and may be altered, e.g., by the insertion, deletion or substitution of nucleotides, so long as the sequences provide for functional rescue, replication and packaging. The ITRs may be from any AAV serotype for which a recombinant virus can be derived including, but not limited to, AAV serotypes AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10, and AAV-11. The nucleotide sequences of the genomes of the AAV serotypes are known in the art. For example, the complete genome of AAV-1 is provided in GenBank Accession No. NC_002077; the complete genome of AAV-2 is provided in GenBank Accession No. NC 001401 and Srivastava et al., *Virology*, 45: 555-564 {1983}; the complete genome of AAV-3 is provided in GenBank Accession No. NC_1829; the complete genome of AAV-4 is provided in GenBank Accession No. NC_001829; the AAV-5 genome is provided in GenBank Accession No. AF085716; the complete genome of AAV-6 is provided in GenBank Accession No. NC_001862; at least portions of AAV-7 and AAV-8 genomes are provided in GenBank Accession Nos. AX753246 and AX753249, respectively; the AAV-9 genome is provided in Gao et al., (2004) *J. Virol.*, 78: 6381-6388; the AAV-10 genome is provided in Williams, (2006) *Mol. Ther.*, 13(1): 67-76; and the AAV-11 genome is provided in Mori et al., (2004) *Virology*, 330(2): 375-383. In one embodiment, the vector is an AAV-9 vector, with AAV-2 derived ITRs.

In some embodiments, the rAAV vector disclosed herein comprise one or more ITRs, e.g., two ITRs, with one upstream and the other downstream of a transgene (e.g., encoding hPGRN) and/or the other nucleic acid elements discussed above. In some embodiments, a nucleic acid disclosed herein, e.g., in an scAAV vector, comprises a first ITR that is disposed 5' and a second ITR that is disposed 3' to the promoter, minigene, transgene, post-transcriptional regulatory element, and/or polyA, e.g., wherein the ITRs are independently 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 50, 100, 150, 200, 250 nucleotides 5' and/or 3' of the other elements. An ITR sequence may be wild-type, or it may comprise one or more mutations, e.g., as long as it retains one or more function of a wild-type ITR. In some embodiments, wild-type ITR may be modified to comprise a deletion of a terminal resolution site. In some embodiments, an scAAV as disclosed herein may comprise two ITR sequences, where both are wild-type, variant, or modified AAV ITR sequences. In some embodiments, at least one ITR sequence is a wild-type, variant or modified AAV ITR sequence. In some embodiments, the two ITR sequences are both wild-type, variant or modified AAV ITR sequences. In some embodiments, the "left" or 5'- ITR is a modified AAV ITR sequence that allows for production of self-complementary genomes, and the "right" or 3'-ITR is a wild-type AAV ITR sequence. In some embodiments, the "right" or 3'-ITR is a modified AAV ITR sequence that allows for the production of self-complementary genomes, and the "left" or 5'- ITR is a wild-type AAV ITR sequence. In some embodiments, the ITR sequences are wild-type, variant, or modified AAV2 ITR sequences. In some embodiments, at least one ITR sequence is a wild-type, variant or modified AAV2 ITR sequence. In some embodiments, the two ITR sequences are both wild-type, variant or modified AAV2 ITR sequences. In some embodiments, the "left" or 5'- ITR is a modified AAV2 ITR sequence that allows for production of self-complementary genomes, and the "right" or 3'-ITR is a wild-type AAV2 ITR sequence. In some embodiments, the "right" or 3'-ITR is a modified AAV2 ITR sequence that allows for the production of self-complementary genomes, and the "left" or 5'- ITR is a wild-type AAV2 ITR sequence. Exemplary sequences that may be used for one or more of the ITRs are described herein. In some embodiments, the AAV vector comprises SEQ ID NO: 12 and SEQ ID NO: 23. In some embodiments, the AAV vector comprises SEQ ID NO: 85 and SEQ ID NO: 90. Embodiments of AAV ITRs provided in WO/2019/094253 (PCT/US2018/058744), which is incorporated herein by reference in its entirety, may also be used for any AAV ITR disclosed herein.

In various embodiments, a vector disclosed herein may comprise a minigene and a nucleic acid sequence encoding a hPGRN disclosed herein. In some embodiments, addition of a splice modulator increases the expression of a functional PRGN polypeptide in a targeted cell. In other embodiments, addition of a splice modulator decreases expression of a functional PRGN polypeptide in a targeted cell. In some embodiments, the vector is a viral vector. In some embodiments, the vector comprising the transgene encoding hPGRN is or is derived from an AAV. In some embodiments, the vector is an rAAV. In various embodiments, the AAV vector comprising the transgene encoding a hPGRN disclosed herein, e.g., the rAAV vector, is an scAAV. The rAAV genomes may comprise one or more AAV ITRs flanking a transgene sequence encoding hPGRN. The transgene sequence may be operatively linked to

transcriptional control elements such as those disclosed herein, e.g., promoter, enhancer, PRE, and/or polyA sequences that are functional in target cells to drive expression of the transgene sequence.

In some embodiments, the rAAV vector lacks one or more (e.g., all) AAV rep and/or cap genes. An AAV vector may comprise (e.g., in its ITRs) nucleic acid sequences (e.g., DNA) from any suitable AAV serotype. Suitable AAV serotypes include, but are not limited to, AAV serotypes AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10 and AAV-11. For instance, an AAV vector, e.g., an scAAV vector, may comprise nucleic acid sequences from an AAV-2, e.g., ITR sequences from an AAV-2. An AAV vector, e.g., an scAAV vector, may also comprise nucleic acids from more than one serotype. GenBank Accession No. NC 001401 and Srivastava et al., *Viol.*, 45: 555-564 {1983}; GenBank Accession No. NC_1829; GenBank Accession No. NC_001829; GenBank Accession No. AF085716; GenBank Accession No. NC_001862; GenBank Accession Nos. AX753246 and AX753249; Gao et al., *J. Virol.*, 78: 6381-6388 (2004); Williams, (2006) *Mol. Ther.*, 13(1): 67-76; and Mori et al., (2004) *Virology*, 330(2): 375-383.

In some embodiments, functional inverted terminal repeat (ITR) sequences in a viral vector comprising the transgene encoding a hPGRN disclosed herein may be used to support, e.g., the rescue, replication and packaging of the AAV virion. Thus, an AAV vector disclosed herein may include sequences that in cis provide for replication and packaging (e.g., functional ITRs) of the virus. The ITRs need not be the wild-type nucleotide sequences, and may be altered, e.g., by the insertion, deletion or substitution of nucleotides, so long as the sequences provide for functional rescue, replication and packaging. The ITRs may be from any AAV serotype for which a recombinant virus can be derived including, but not limited to, AAV serotypes AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10 and AAV-11. GenBank Accession No. NC_002077; GenBank Accession No. NC 001401 and Srivastava et al., *Viol.*, 45: 555-564 {1983}; GenBank Accession No. NC_1829; GenBank Accession No. NC_001829; GenBank Accession No. AF085716; GenBank Accession No. NC_001862; GenBank Accession Nos. AX753246 and AX753249, respectively; Gao et al., (2004) *J. Virol.*, 78: 6381-6388; Williams, (2006) *Mol. Ther.*, 13(1): 67-76; and Mori et al., (2004) *Virology*, 330(2): 375-383. In one embodiment, the vector is an AAV-9 vector, with AAV-2 derived ITRs.

In some embodiments, the AAV viral vector comprises a sequence of SEQ ID NO: 91. In some embodiments, the AAV viral vector comprises a sequence of SEQ ID NO: 11. In each of these embodiments, the transgene sequence may be replaced with a sequence encoding an alternate molecule of interest, e.g., as described herein.

In some embodiments, a vector or nucleic acid sequence disclosed herein forms a cloning vector or an expression vector. In such embodiments, the vector may comprise other components that facilitate replication or maintenance of the vector. In some embodiments, the vector further comprises a selectable marker for clonal selection. In some embodiments, the selectable marker in the vector comprises a prokaryotic or eukaryotic antibiotic resistance gene. In some embodiments, the selectable marker in the

vector comprises a kanamycin resistance gene. In some embodiments, the selectable marker in the vector comprises an ampicillin resistance gene. In some embodiments, the vector further comprises a puromycin resistance gene. In some embodiments, the selectable marker in the vector comprises a hygromycin resistance gene. In some embodiments, the vector (e.g., plasmid) comprises a nucleic acid sequence of SEQ ID NO: 92.

Exemplary AAV vector sequence comprising a minigene and transgene encoding EGFP:

```
CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCT
TTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGAATTCGGGCG
GAGTTAGGGCGGAGCCAATCAGCGTGCGCCGTTCCGAAAGTTGCCTTTTATGGCTGGGCG
GAGAATGGGCGGTGAACGCCGATGATTATATAAGGACGCGCCGGGTGTGGCACAGCTAG
TTCCGTGCGAGCCGGGATTTGGGTGCGGTTCTTGTTTGTGGATCCCTGTGATCGTCACTT
GACACCGGTCTTCCAGAGGAGATTGGAAAACCTTGAAGAAGAAGTGGATTGTGCTAATATTG
CCCTGAAAGCAGCCACCATTGGATTGGGAGAGTTGGAAACAAAATTTGCAAATTGATATCAA
GTTAGCATTACAGATTTGGCTGAGGAGAATATCCATTATTTTGAACAGGTAATTAGTGTTG
TTTGATATTGCTTCATTTTAAAGTTATTTGCTCATTTACTTTTGGTCCGTCCATTGTTGAAAG
AGTGTATTAAAGAACAAAGTGTACATTCTATTGCCTCTCTGGTAGCTTGTTTTGTTGAAGT
TGTCAGTTACCATTTGGTTTTGTTTATCCTCAGTTTGTGTTTTGGATTTGGATTCTTCAAAA
GCATTTGATATTGCTTTCTATTGATTGTCCTAACTACTCCTCTTTCCTCTCCCTTCTCCATTT
TTGAAGAGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGATTGAGCAGAAAATCATTTCAG
GGCCTGTTCTCTATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCTGAAACCATCAACAAAG
GAGCACACCATTCCATCAGCAAAAGAGTAACAACATCTTTTTTTAAGTTCATTTTGTTTTTCA
GTTGATTGTATTTCAATTTTTTTTACAGCTGACTTTTCTCAGAGAAGTTTTTTTTTTTATTGTAAA
CATACTTTTTCTAGAAAGTATATTTTAAATAACATCTTTAACCTTATCTCTGGCTGAATTATT
GAATATTTGAAATTATTACATTAACAAAATTTTGTCTTACAGCAGTGGTCCCCAACCTTCTTA
GCAGTAGCATCCCTCATTAAAGAATTAATAATTTGTAGAAATTGACAAGGATTCTGACAAGCTG
TTGGGAGAGAAGAATAGAGCAGATTGCAGTAGGAACAGTTGTGTTAGAAATTTATTAATCCTT
TAACACTGAAAGTAACTATTGTTGATTGCCTCTTGGTGTGTTTCCATTATTCAGTGCTCTTG
CTAAGTGGGAGTCATTCCCTTACATCAACCACCAACCTTCACTTGGAAGAAGCTAGCGAAGA
TAAACCTCGCAACCGCCGCGGCAGCGGCGAAGGCCGCGGCAGCCTGCTGACCTGCGGC
GATGTGGAAGAAAACCCGGGCCCCGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGGTGGT
GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCG
AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGC
AAGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGACCTACGGCGTGCAGTGCTTC
AGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGC
TACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG
GAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTAT
ATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTTCAAGATCCGCCACAACATC
GAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGG
CCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCT
CGGCATGGACGAGCTGTACAAGTAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTAT
TTGTAACCATTAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTC
AGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTTAAATCGATCTGAGGAACCCCTAGTGATGG
```

AGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTC
 GCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGG
 GAGTGG (SEQ ID NO: 11)

Table 2 and Table 3 describe exemplary sequences of the nucleic acids, vectors, and minigenes.

Table 2. Exemplary minigene and AAV vector sequences having a SNX7-derived minigene.

Sequence description	SEQUENCE	SEQ ID NO:
AAV2 5'-ITR with trs deletion	CTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGCCC GGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAG CGAGCGCGCAGAGAGGGAGTGG	12
JeT promoter	AATTCGGGCGGAGTTAGGGCGGAGCCAATCAGCGTGCGCCG TTCCGAAAGTTGCCTTTTATGGCTGGGCGGAGAATGGGCGGT GAACGCCGATGATTATATAAGGACGCGCCGGGTGTGGCACA GCTAGTTCCGTCGCAGCCGGGATTTGGGTCGCGGTTCTTGTTT GTGGATCCCTGTGATCGTCACTTGACA	13
SNX Minigene (Version 1) First Exon (purine-rich exonic splicing enhancer in italics; Kozak in bold; start codon underlined)	CTTCCAGAGGAGATTGGAAA <u>CTTGAAGAAG</u> AGTGGATTGT GCTAATATTGCCCTGAAAGCAG CCACC ATGATTGGGAGAG TTGGAAACAAAATTTGCAAATTGATATCAAGTTAGCATTTAC AGATTTGGCTGAGGAGAATATCCATTATTTTGAACAG	14
SNX7 Minigene (Version 1) first intron	GTAATTAGTGTTGTTTGATATTGCTTCATTTTAAAGTTATTTG CTCATTTACTTTTGGTCCGTCCATTGTTGAAAGAGTGATTAA AGAACAAGTGTCACATTCTATTGCCTCTCTGGTAGCTTGGTTT TGTTGAAGTTGTCAGTTACCATTGTTTGTGTTTATCCTCAGT TTGTTGTTTGGATTGATTCTTCAAAGCATTGATATTGC TTTCTATTGATTGTCCTAACTACTCCTCTTTCCTCTCCCTTCTC CATTTTGAAG	15
SNX7 Minigene (Version 1) second exon	AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGATTGAGCA GAAAATCATTTCAAGGCCTGTTCTCTATTGTCCTTGCTATCCT GTCTTCTGTAGCTATCTGAAACCATCAACAAAGGAGCACACC ATCCATCAGCAAAAGA	16
SNX7 Minigene (both Version 1 and Version 2) second intron	GTAACAACATCTTTTTTAAAGTTCATTTTGTTTTTCAGTTGATT GTATTTCAATTTTTTACAGCTGACTTTTCTCAGAGAAGTTTT TTTTTTATTGTAAACATACTTTTCTAGAAAGTATATTTAAA ATAACATCTTTAACCTTATCTCTGGCTGAATTATTGAATATTT GAAATTATTACATTAACAAAATTTGTCTTACAGCAGTGGTC CCCAACCTTCTTAGCAGTAGCATCCCTCATTAAGAATTAAAA TTTGTAGAAATTGACAAGGATTCTGACAAGCTGTTGGGAGAG AAGAATAGAGCAGATTGCAGTAGGAACAGTTGTGTTAGAAT	17

	TTATTAATCCTTTAACA CTGAAAGTAAACTATTGTTGATTGCC TCTTGGTGTGTTTCCATTATTCAG	
SNX7 Minigene (Both Version 1 and Version 2) third exon	TGCTCTTGCTAAGTGGGAGTCATTCTTACATCAACCACCAA CCTTCACTTGGAAGAAGCTAGCGAAGATAAACCT	18
Full SNX7 minigene (Version 1) sequence	CTTCCAGAGGAGATTGGAAA ACTTGAAGAAGAAGTGGATTG TGCTAATATTGCCCTGAAAGCAGCCACCATGGATTGGGAGAG TTGGAAACAAAATTTGCAAATTGATATCAAGTTAGCATTTAC AGATTTGGCTGAGGAGAATATCCATTATTTGAACAGGTAAT TAGTGTTGTTTGATATTGCTTCATTTTAAAGTTATTTGCTCAT TTACTTTTGGTCCGTCCTATTGTTGAAAGAGTGTATTAAGAA CAAGTGTACATTCTATTGCCTCTCTGGTAGCTTGGTTTTGTT GAAGTTGTCAGTTACCATTTGGTTTTGTTTATCCTCAGTTTGT TGTTTTGGATTTGGATTCTTCAAAAGCATTTGATATTGCTTTC TATTGATTGTCCTAACTACTCCTCTTCTCCTCTCCCTCTCCATT TTTGAAGAGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGA TTGAGCAGAAAATCATTTTCAGGGCCTGTTCTCTATTGTCCTTG CTATCCTGTCTTCTGTAGCTATCTGAAACCATCAACAAAGGA GCACACCATTCCATCAGCAAAAAGAGTAACAACATCTTTTTTT AAGTTCATTTTGTTTTTTCAGTTGATTGTATTTCATTTTTTTAC AGCTGACTTTTCTCAGAGAAGTTTTTTTTTTTATTGTAAACATA CTTTTTCTAGAAAGTATATTTTAAAATAACATCTTTAACCTTA TCTCTGGCTGAATTATTGAATATTTGAAATTATTACATTAACA AAATTTTGTCTTACAGCAGTGGTCCCCAACCTTCTTAGCAGT AGCATCCCTCATTAAGAATTAATAATTTGTAGAAATTGACAAG GATTCTGACAAGCTGTTGGGAGAGAAGAATAGAGCAGATTG CAGTAGGAACAGTTGTGTTAGAATTTATTAATCCTTTAACAC TGAAAGTAACTATTGTTGATTGCCTCTTGGTGTGTTTCCATT ATTCAGTGCTCTTGCTAAGTGGGAGTCATTCTTACATCAAC CACCAACCTTCACTTGGAAGAAGCTAGCGAAGATAAACCT	71
Sequence encoding furin cleavage sequence	CGCAACCGCCGC	19
Sequence encoding T2A peptide	GGCAGCGGCGAAGGCCGCGGCAGCCTGCTGACCTGCGGCCGA TGTGGAAGAAAACCCGGGCCCCG	20
Sequence encoding EGFP (transgene)	GTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT CCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCA GCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTG CCCTGGCCCAACCCTCGTGACCACCCTGACCTACGGCGTGACG TGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTC TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC	21

	TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGG CACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATC ATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAA GATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCG ACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG CTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTG AGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCT GGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA GCTGTACAAGTAA	
SV40 polyA	TGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAA CCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCA TTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTT TTTAA	22
AAV2 3'-ITR	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCG CTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGAC GCCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCG CGCAGAGAGGGAGTGG	23
SNX Minigene (Version 2) First Exon	GAAGAAGAAGATATCAAGTTAGCATTTACAGATTTGGCTGAGGAGA AGAACAG	96
SNX7 Minigene (Version 2) first intron	GTAATTAGTGTTGTTTGATATTGCTTCATTTTAAAGTTATTTGCTCAT TTAGCATTTGATATTGCTTTCTATTGATTGTCCTAACTACTCCTCTT TCCTCTCCCTTCTCCATTTTTGAAG	97
SNX7 Minigene (Version 2) second exon (start codon in Bold ; modified nucleotides in <i>Italic</i>)	AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGATTGAGCAGAAA ATCATTTTCAGGGCCTGTTCTCTATTGTCCTTGCTATCCTGTCTTCT GTAGCTATCTGAAACCATCAACAAAGGAGCACACC ATGG CATCAG CAAAAGA	98
Full SNX7 minigene (Version 2) sequence	GAAGAAGAAGATATCAAGTTAGCATTTACAGATTTGGCTGAGGAG AAGAACAGGTAATTAGTGTTGTTTGATATTGCTTCATTTTAAAGTTA TTTGCTCATTTAGCATTTGATATTGCTTTCTATTGATTGTCCTAACTA CTCCTCTTTCCTCTCCCTTCTCCATTTTTGAAGAGTTTGCAAAGGAA GGAAAGGAGCAGAGACTTGATTGAGCAGAAAATCATTTTCAGGGCC TGTTCTCTATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCTGAAACC ATCAACAAAGGAGCACACCATGGCATCAGCAAAAGAGTAACAACA TCTTTTTTTAAGTTCATTTTGTTTTTCAGTTGATTGTATTCAATTTT TTACAGCTGACTTTTCTCAGAGAAGTTTTTTTTTTTATTGTAAACATA CTTTTTCTAGAAAGTATATTTTAAAATAACATCTTTAACCTTATCTC TGGCTGAATTATTGAATATTTGAAATTATTACATTAACAAAATTTTG TCTTACAGCAGTGGTCCCCAACCTTCTTAGCAGTAGCATCCCTCATT AAGAATTAAAATTTGTAGAAATTGACAAGGATTCTGACAAGCTGTT GGGAGAGAAGAATAGAGCAGATTGCAGTAGGAACAGTTGTGTTAG AATTTATTAATCCTTTAACACTGAAAGTAAACTATTGTTGATTGCCT CTTGGTGTGTTTCCATTATTCAGTGCTCTTGCTAAGTGGGAGTCATT CCTTACATCAACCACCAACCTTCACTTGGAAGAAGCTAGCGAAGAT AAACCT	94

Table 3. Sequence of exemplary components, and plasmid encoding a single-stranded AAV comprising a minigene and transgene encoding human progranulin

Sequence description	SEQUENCE	SEQ ID NO:
AAV2 5'-ITR with trs deletion	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCC GGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCCT CAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCC ATCACTAGGGGTTTCCT	85
Human Synapsin promoter	AACCGAGTATCTGCAGAGGGCCCTGCGTATGAGTGCAAGTG GGTTTTAGGACCAGGATGAGGCGGGGTGGGGGTGCCTACCT GACGACCGACCCCGACCCACTGGACAAGCACCCAACCCCCA TTCCCCAAATTGCGCATCCCCTATCAGAGAGGGGGAGGGGA AACAGGATGCGGCGAGGCGCGTGCGCACTGCCAGCTTCAGC ACCGCGGACAGTGCCTTCGCCCCCGCCTGGCGGCGCGCGCCA CCGCCGCTCAGCACTGAAGGCGCGCTGACGTCACCTCGCCGG TCCCCCGCAAACCTCCCTTCCCGGCCACCTTGGTCGCGTCCG CGCCGCCGCCGGCCAGCCGGACCGCACACGCGAGGCGCG AGATAGGGGGGCACGGGCGCGACCATCTGCGCTGCGGCGCC GGCGACTCAGCGCTGCCTCAGTCTGCGGTGGGCAGCGGAGG AGTCGTGTCGTGCCTGAGAGCGCAGCTGT	86
Full SNX7 minigene sequence (version 1)	CTTCCAGAGGAGATTGGAAAACCTGAAGAAGAAGTGGATTGT GCTAATATTGCCCTGAAAGCAGCCACCATTGGATTGGGAGAG TTGGAAACAAAATTTGCAAATTGATATCAAGTTAGCATTTAC AGATTTGGCTGAGGAGAATATCCATTATTTTGAACAGGTAAT TAGTGTTGTTTGATATTGCTTCATTTTAAAGTTATTTGCTCAT TTACTTTTGGTCCGTCCATTGTTGAAAGAGTGTATTAAAGAA CAAGTGTACATTCTATTGCCTCTCTGGTAGCTTGGTTTTGTT GAAGTTGTCAGTTACCATTTGGTTTTGTTTATCCTCAGTTTGT TGTTTTGGATTTGGATTCTTCAAAAGCATTTGATATTGCTTTC TATTGATTGTCCTAACTACTCCTCTTCTCCTCTCCCTTCTCCATT TTTGAAGAGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGA TTGAGCAGAAAATCATTTACAGGGCCTGTTCTCTATTGTCCTTG CTATCCTGTCTTCTGTAGCTATCTGAAACCATCAACAAAGGA GCACACCATTCCATCAGCAAAAGAGTAACAACATCTTTTTTT AAGTTCATTTTGTTCAGTTGATTGTATTCAATTTTTTTTAC AGCTGACTTTTCTCAGAGAAGTTTTTTTTTTTATTGTAAACATA CTTTTTCTAGAAAGTATATTTTAAATAACATCTTTAACCTTA TCTCTGGCTGAATTATTGAATATTTGAAATTATTACATTAACA AAATTTTGTCTTACAGCAGTGGTCCCCAACCTTCTTAGCAGT AGCATCCCTCATTAAGAATTAAATTTGTAGAAATTGACAAG GATTCTGACAAGCTGTTGGGAGAGAAGAATAGAGCAGATTG CAGTAGGAACAGTTGTGTTAGAATTTATTAATCCTTTAACAC TGAAAGTAAACTATTGTTGATTGCCTCTTGGTGTGTTCCATT	71

	ATTCAGTGCTCTTGCTAAGTGGGAGTCATTTCCTTACATCAAC CACCAACCTTCACTTGGAAGAAGCTAGCGAAGATAAACCT	
Sequence encoding furin cleavage sequence	CGCAACCGCCGC	19
Sequence encoding T2A peptide	GGCAGCGGCGAAGGCCGCGGCAGCCTGCTGACCTGCGGCGA TGTGGAAGAAAACCCGGGCCCCG	20
Sequence encoding human progranulin (transgene)	TGGACCCTGGTGAGCTGGGTGGCCTTAACAGCAGGGCTGGTG GCTGGAACGCGGTGCCCAGATGGTCAGTTCTGCCCTGTGGCC TGCTGCCTGGACCCCGGAGGAGCCAGCTACAGCTGCTGCCGT CCCCTTCTGGACAAATGGCCCACAACACTGAGCAGGCATCTG GGTGGCCCCTGCCAGGTTGATGCCCACTGCTCTGCCGGCCAC TCCTGCATCTTTACCGTCTCAGGGACTTCCAGTTGCTGCCCC TCCCAGAGGCCGTGGCATGCGGGGATGGCCATCACTGCTGCC CACGGGGCTTCCACTGCAGTGCAGACGGGCGATCCTGCTTCC AAAGATCAGGTAACAACACTCCGTGGGTGCCATCCAGTGCCCTG ATAGTCAGTTCGAATGCCCGGACTTCTCCACGTGCTGTGTTA TGGTCGATGGCTCCTGGGGGTGCTGCCCCATGCCCCAGGCTT CCTGCTGTGAAGACAGGGTGCAGTGTGCTGCGCACGGTGCTT TCTGCGACCTGGTTCACACCCGCTGCATCACACCCACGGGCA CCCACCCCTGGCAAAGAAGCTCCCTGCCAGAGGACTAAC AGGGCAGTGGCCTTGTCCAGCTCGGTGATGTGTCCGGACGCA CGGTCCCGGTGCCCTGATGGTTCTACCTGCTGTGAGCTGCCC AGTGGGAAGTATGGCTGCTGCCCAATGCCCAACGCCACCTGC TGCTCCGATCACCTGCACTGCTGCCCCAAGACACTGTGTGT GACCTGATCCAGAGTAAGTGCCTCTCCAAGGAGAACGCTACC ACGGACCTCCTCACTAAGCTGCCTGCGCACACAGTGGGGGAT GTGAAATGTGACATGGAGGTGAGCTGCCAGATGGCTATAC CTGCTGCCGTCTACAGTCGGGGGCTGGGGCTGCTGCCCTTT TACCCAGGCTGTGTGCTGTGAGGACCACATACTGCTGTCC CGCGGGGTTTACGTGTGACACGCAGAAGGGTACCTGTGAAC AGGGGCCCCACCAGGTGCCCTGGATGGAGAAGGCCCCAGCT CACCTCAGCCTGCCAGACCCACAAGCCTTGAAGAGAGATGTC CCCTGTGATAATGTCAGCAGCTGTCCCTCCTCCGATACCTGCT GCCAACTCACGTCTGGGGAGTGGGGCTGCTGTCCAATCCCAG AGGCTGTCTGCTGCTCGGACCACCAGCACTGCTGCCCCAGG GCTACACGTGTGTAGCTGAGGGGCAGTGTGAGCGAGGAAGC GAGATCGTGGCTGGACTGGAGAAGATGCCTGCCCCGCCGGC TTCCTTATCCCACCCCAGAGACATCGGCTGTGACCAGCACAC CAGCTGCCCCGTGGGGCAGACCTGCTGCCCCGAGCCTGGGTG GGAGCTGGGCCTGCTGCCAGTTGCCCCATGCTGTGTGCTGCG AGGATCGCCAGCACTGCTGCCCGGCTGGCTACACCTGCAACG TGAAGGCTCGATCCTGCGAGAAGGAAGTGGTCTCTGCCCAGC CTGCCACCTTCTGGCCCGTAGCCCTCACGTGGGTGTGAAGG ACGTGGAGTGTGGGGAAGGACACTTCTGCCATGATAACCAG	87

	ACCTGCTGCCGAGACAACCGACAGGGCTGGGCCTGCTGTCCC TACCGCCAGGGCGTCTGTTGTGCTGATCGGCGCCACTGCTGT CCTGCTGGCTTCCGCTGCGCAGCCAGGGGTACCAAGTGTGTTG CGCAGGGAGGCCCGCGCTGGGACGCCCTTTGAGGGACCC AGCCTTGAGACAGCTGCTGTAG	
HPRE-NOX	AACAGGCCTATTGATTGGAAAGTATGTCAACGAATTGTGGGT CTTTTGGGGTTTGCTGCCCCCTTTTACGCAATGTGGATATCCTG CTTTAATGCCTTTATATGCATGTATACAAGCAAAACAGGCTT TTACTTTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGT ATCTGACCCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGT GCCAAGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCTTGG CCATAGGCCATCAGCGCATGCGTGGAACCTTTGTGTCTCCTC TGCCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCG CAGCAGGTCTGGAGCGAACTCATCGGGACTGACAATTCTGT CGTGCTCTCCCGCAAGTATACATCGTTTCCAGGGCTGCTAGG CTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTA CGTCCCGTCGGCGCTGAATCCCGCGGACGACCCCTCCCGGGG CCGCTTGGGGCTCTACCGCCCGCTTCTCCGTCTGCCGTACCG ACCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCCGTC TGTGCCTTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCT CTGCACGTGCGATGGAGACCACCGTGAACGCCCACCGGAAC CTGCCCAAGGTCTTGATAAGAGGACTCTTGGACTTTCAGCA ATGTC	88
HGH polyA	ACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGG CCCTGGAAGTTGCCACTCCAGTGCCCAACAGCCTTGTCTTAA TAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTA TAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGC AAGTTGGGAAGACAACCTGTAGGGCCTGCGGGGTCTATTGG GAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCA ATCTCCGCCTCCTGGGTCAAGCGATTCTCCTGCCTCAGCCTC CCGAGTTGTTGGGATTCCAGGCATGCATGACCAGGCTCAGCT AATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGC CAGGCTGGTCTCCAACCTCCTAATCTCAGGTGATCTACCCACC TTGGCCTCCCAAATTGCTGGGATTACAGGCGTGAACCACTGC TCCCTTCCCTGTCCTT	89
AAV2 3'-ITR	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCG CTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGAC GCCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCG CGCAGCTGCCTGCAGG	90
Full rAAV genome sequence	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCC GGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCT CAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCC ATCACTAGGGGTTCTGCGGCCGCACGCGTAACCGAGTATCT GCAGAGGGCCCTGCGTATGAGTGCAAGTGGGTTTTAGGACC	91

	AGGATGAGGCGGGGTGGGGGTGCCTACCTGACGACCGACCC CGACCCACTGGACAAGCACCCAACCCCCATTCCCCAAATTGC GCATCCCCTATCAGAGAGGGGGAGGGGAAACAGGATGCGGC GAGGCGCGTGCGCACTGCCAGCTTCAGCACCGCGGACAGTG CCTTCGCCCCCGCCTGGCGGCGCGCGCCACCGCCGCCTCAGC ACTGAAGGCGCGCTGACGTCACTCGCCGGTCCCCGCAAATC CCCCTTCCCGGCCACCTTGGTCGCGTCCGCGCCGCCGCCGGC CCAGCCGGACCGCACCCACGCGAGGCGCGAGATAGGGGGGCA CGGGCGCGACCATCTGCGCTGCGGCGCCGGCGACTCAGCGCT GCCTCAGTCTGCGGTGGGCAGCGGAGGAGTCGTGTCGTGCCT GAGAGCGCAGCTGTGAATTCCTTCCAGAGGAGATTGGA AAA CTTGAAGAAGAAGTGGATTGTGCTAATATTGCCCTGAAAGCA GCCACCATGGATTGGGAGAGTTGGAAACAAAATTTGCAAAT TGATATCAAGTTAGCATTTACAGATTTGGCTGAGGAGAATAT CCATTATTTTGAACAGGTAATTAGTGTTGTTTGATATTGCTTC ATTTTAAAGTTATTTGCTCATTTACTTTTGGTCCGTCCATTGTT GAAAGAGTGTATTAAAGAACAAGTGTCACATTCTATTGCCTC TCTGGTAGCTTGGTTTTGTTGAAGTTGTCAGTTACCATTTGGT TTTGTTTATCCTCAGTTTGTGTTTTTGGATTTGGATTCTTCAAA AGCATTTGATATTGCTTTCTATTGATTGTCCTAACTACTCCTC TTTCTCTCCCTTCTCCATTTTTGAAGAGTTTGCAAAGGAAGG AAAGGAGCAGAGACTTGATTGAGCAGAAAATCATTTCAGGG CCTGTTCTCTATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCT GAAACCATCAACAAAGGAGCACACCATTCCATCAGCAAAAG AGTAACAACATCTTTTTTTAAGTTCATTTTGTTTTTCAGTTGA TTGTATTTCAATTTTTTTTACAGCTGACTTTTCTCAGAGAAGTT TTTTTTTTATTGTAAACATACTTTTTCTAGAAAGTATATTTTA AAATAACATCTTTAACCTTATCTCTGGCTGAATTATTGAATAT TTGAAATTATTACATTAACAAAATTTTGTCTTACAGCAGTGG TCCCCAACCTTCTTAGCAGTAGCATCCCTCATTAGAATTAA AATTTGTAGAAATTGACAAGGATTCTGACAAGCTGTTGGGAG AGAAGAATAGAGCAGATTGCAGTAGGAACAGTTGTGTTAGA ATTTATTAATCCTTTAACACTGAAAGTAAACTATTGTTGATTG CCTCTTGGTGTGTTTCCATTATTCAGTGCTCTTGCTAAGTGGG AGTCATTCCCTACATCAACCACCAACCTTCACTTGGAAGAAG CTAGCGAAGATAAACCTCGCAACCGCCGCGGCAGCGGCGAA GGCCGCGGCAGCCTGCTGACCTGCGGCGATGTGGAAGAAAA CCCGGGCCCCGTGGACCCTGGTGAGCTGGGTGGCCTTAACAGC AGGGCTGGTGGCTGGAACGCGGTGCCCAGATGGTCAGTTCTG CCCTGTGGCCTGCTGCCTGGACCCCGGAGGAGCCAGCTACAG CTGCTGCCGTCCCCTTCTGGACAAATGGCCCACAACACTGAG CAGGCATCTGGGTGGCCCCCTGCCAGGTTGATGCCCAGTGCTC TGCCGGCCACTCCTGCATCTTTACCGTCTCAGGGACTTCCAGT TGCTGCCCCTTCCCAGAGGCCGTGGCATGCGGGGATGGCCAT CACTGCTGCCCACGGGGCTTCCACTGCAGTGACAGCGGGCGA TCCTGCTTCCAAAGATCAGGTAACAACTCCGTGGGTGCCATC CAGTGCCCTGATAGTCAGTTCGAATGCCCGGACTTCTCCACG	
--	--	--

	TGCTGTGTTATGGTCGATGGCTCCTGGGGGTGCTGCCCCATG CCCCAGGCTTCCTGCTGTGAAGACAGGGTGCACTGCTGTCCG CACGGTGCCTTCTGCGACCTGGTTTACACCCCGCTGCATCACA CCCACGGGCACCCACCCCTGGCAAAGAAGCTCCCTGCCAG AGGACTAACAGGGCAGTGGCCTTGTCCAGCTCGGTCATGTGT CCGGACGCACGGTCCCGGTGCCCTGATGGTTCTACCTGCTGT GAGCTGCCCAGTGGGAAGTATGGCTGCTGCCCAATGCCAAC GCCACCTGCTGCTCCGATCACCTGCACTGCTGCCCCCAAGAC ACTGTGTGTGACCTGATCCAGAGTAAGTGCCTCTCCAAGGAG AACGCTACCACGGACCTCCTACTAAGCTGCCTGCGCACACA GTGGGGGATGTGAAATGTGACATGGAGGTGAGCTGCCCAGA TGGCTATACCTGCTGCCGTCTACAGTCGGGGGCCTGGGGCTG CTGCCCTTTTACCCAGGCTGTGTGCTGTGAGGACCACATACA CTGCTGTCCCGCGGGGTTTACGTGTGACACGCAGAAGGGTAC CTGTGAACAGGGGCCCCACCAGGTGCCCTGGATGGAGAAGG CCCCAGCTCACCTCAGCCTGCCAGACCCACAAGCCTTGAAGA GAGATGTCCCCTGTGATAATGTCAGCAGCTGTCCCTCCTCCG ATACCTGCTGCCAACTCACGTCTGGGGAGTGGGGCTGCTGTC CAATCCCAGAGGCTGTCTGCTGCTCGGACCACCAGCACTGCT GCCCCCAGGGCTACACGTGTGTAGCTGAGGGGCAGTGTGAG CGAGGAAGCGAGATCGTGGCTGGACTGGAGAAGATGCCTGC CCGCCGGGCTTCCTTATCCACCCCAAGACATCGGCTGTGA CCAGCACACCAGCTGCCCGGTGGGGCAGACCTGCTGCCCGA GCCTGGGTGGGAGCTGGGCCTGCTGCCAGTTGCCCCATGCTG TGTGCTGCGAGGATCGCCAGCACTGCTGCCCGGCTGGCTACA CCTGCAACGTGAAGGCTCGATCCTGCGAGAAGGAAGTGGTC TCTGCCCAGCCTGCCACCTTCCTGGCCCGTAGCCCTCACGTG GGTGTGAAGGACGTGGAGTGTGGGGAAGGACACTTCTGCCA TGATAACCAGACCTGCTGCCGAGACAACCGACAGGGCTGGG CCTGCTGTCCCTACCGCCAGGGCGTCTGTTGTGCTGATCGGC GCCACTGCTGTCCTGCTGGCTTCCGCTGCGCAGCCAGGGGTA CCAAGTGTGCGCAGGGAGGCCCGCGCTGGGACGCCCTT TGAGGGACCCAGCCTTGAGACAGCTGCTGTAGGTCGACTAA ACAGGCCTATTGATTGAAAGTATGTCAACGAATTGTGGGTC TTTTGGGGTTTTGCTGCCCCTTTTACGCAATGTGGATATCCTGC TTTAATGCCTTTATATGCATGTATACAAGCAAAACAGGCTTT TACTTTCTCGCCAACTTACAAGGCCTTTCTAAGTAAACAGTA TCTGACCCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGC CAAGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCTTGCC ATAGGCCATCAGCGCATGCGTGGAACCTTTGTGTCTCCTCTG CCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCGC AGCAGGTCTGGAGCGAACTCATCGGGACTGACAATTCTGTG GTGCTCTCCCGCAAGTATACATCGTTTCCAGGGCTGCTAGGC TGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTAC GTCCCGTCGGCGCTGAATCCCGCGGACGACCCCTCCCGGGG CGCTTGGGGCTCTACCGCCCGCTTCTCCGTCTGCCGTACCGA CCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCCGTCT	
--	--	--

	GTGCCTTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCTC TGCACGTCGCATGGAGACCACCGTGAACGCCACCGGAACC TGCCCAAGGTCTTGCATAAGAGGACTCTTGGACTTTTCAGCAA TGTCAACTCGAGAGATCTACGGGTGGCATCCCTGTGACCCCT CCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGC CCACCAGCCTTGTCTAATAAAATTAAGTTGCATCATTTTGTCT TGACTAGGTGTCCTTCTATAATATTATGGGGTGGAGGGGGGT GGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGG CCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCA CAATCTTGGCTCACTGCAATCTCCGCCTCCTGGGTTCAAGCG ATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCAT GCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGAC GGGGTTTCACCATATTGGCCAGGCTGGTCTCCAACTCCTAAT CTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGGGAT TACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTT GTAGGTAAACCACGTGCGGACCGAGCGGCCGCAGGAACCCCT AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCT CACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCT TTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGC CTGCAGG	
Plasmid Sequence (beta lactamase coding sequence is highlighted gray; pUC origin of replication is underlined	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCC GGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCT CAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCC ATCACTAGGGGTTCTGCGGCCGCACGCGTAACCGAGTATCT GCAGAGGGCCCTGCGTATGAGTGCAAGTGGGTTTTAGGACC AGGATGAGGCGGGGTGGGGGTGCCTACCTGACGACCGACCC CGACCCACTGGACAAGCACCCAACCCCATTTCCCAAATTGC GCATCCCCTATCAGAGAGGGGGAGGGGAAACAGGATGCGGC GAGGCGCGTGCGCACTGCCAGCTTCAGCACCGCGGACAGTG CCTTCGCCCCCGCCTGGCGGCGCGCGCCACCGCCGCCTCAGC ACTGAAGGCGCGCTGACGTCACTCGCCGGTCCCCCGCAAAT CCCCTTCCCGGCCACCTTGGTCGCGTCCGCGCCGCCGCCGGC CCAGCCGGACCGCACCAACGCGAGGCGCGAGATAGGGGGGCA CGGGCGCGACCATCTGCGCTGCGGCGCCGGCGACTCAGCGCT GCCTCAGTCTGCGGTGGGCAGCGGAGGAGTCGTGTCGTGCCT GAGAGCGCAGCTGTGAATTCCTTCCAGAGGAGATTGGAAAA CTTGAAGAAGAAGTGGATTGTGCTAATATTGCCCTGAAAGCA GCCACC<u>ATG</u>GATTGGGAGAGTTGGAAACAAAATTTGCAAAT TGATATCAAGTTAGCATTTACAGATTTGGCTGAGGAGAATAT CCATTATTTTGAACAGGTAATTAGTGTTGTTTGATATTGCTTC ATTTTAAAGTTATTTGCTCATTTACTTTTGGTCCGTCCATTGTT GAAAGAGTGTATTAAAGAACAAGTGTACATTCTATTGCCTC TCTGGTAGCTTGGTTTTGTTGAAGTTGTCAGTTACCATTTGGT TTTGTTTATCCTCAGTTTGTGTTTTGGATTGGAATCTTCAAA AGCATTTGATATTGCTTTCTATTGATTGTCCTAACTACTCCTC TTTCTCTCCCTTCTCCATTTTGAAGAGTTTGCAAAGGAAGG	92

	AAAGGAGCAGAGACTTGATTGAGCAGAAAATCATTTCAGGG CCTGTTCTCTATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCT GAAACCATCAACAAAGGAGCACACCATTCCATCAGCAAAAG AGTAACAACATCTTTTTTTAAGTTCATTTTGTTTTTCAGTTGA TTGTATTTCAATTTTTTTTACAGCTGACTTTTCTCAGAGAAGTT TTTTTTTTATTGTAAACATACTTTTTCTAGAAAGTATATTTTA AAATAACATCTTTAACCTTATCTCTGGCTGAATTATTGAATAT TTGAAATTATTACATTAACAAAATTTTGTCTTACAGCAGTGG TCCCCAACCTTCTTAGCAGTAGCATCCCTCATTAAGAATTAA AATTTGTAGAAATTGACAAGGATTCTGACAAGCTGTTGGGAG AGAAGAATAGAGCAGATTGCAGTAGGAACAGTTGTGTTAGA ATTTATTAATCCTTTAACACTGAAAGTAACTATTGTTGATTG CCTCTTGGTGTGTTCCATTATTCAGTGCTCTTGCTAAGTGGG AGTCATTCCTTACATCAACCACCAACCTTCACTTGGAAGAAG CTAGCGAAGATAAACCTCGCAACCGCCGCGCAGCGGCGAA GGCCGCGGCAGCCTGCTGACCTGCGGCGATGTGGAAGAAAA CCCGGGCCCGTGGACCCTGGTGAGCTGGGTGGCCTTAACAGC AGGGCTGGTGGCTGGAACGCGGTGCCCAGATGGTCAGTTCTG CCCTGTGGCCTGCTGCCTGGACCCCGGAGGAGCCAGCTACAG CTGCTGCCGTCCCCTTCTGGACAAATGGCCCACAACACTGAG CAGGCATCTGGGTGGCCCCTGCCAGGTTGATGCCCACTGCTC TGCCGGCCACTCCTGCATCTTTACCGTCTCAGGGACTTCCAGT TGCTGCCCCTTCCCAGAGGCCGTGGCATGCGGGGATGGCCAT CACTGCTGCCCACGGGGCTTCCACTGCAGTGCAGACGGGCGA TCCTGCTTCCAAAGATCAGGTAACAACCTCCGTGGGTGCCATC CAGTGCCCTGATAGTCAGTTCGAATGCCCGGACTTCTCCACG TGCTGTGTTATGGTCGATGGCTCCTGGGGGTGCTGCCCCATG CCCCAGGCTTCCTGCTGTGAAGACAGGGTGCACTGCTGTCCG CACGGTGCCTTCTGCGACCTGGTTCACACCCGCTGCATCACA CCCACGGGCACCCACCCCTGGCAAAGAAGCTCCCTGCCCAG AGGACTAACAGGGCAGTGGCCTTGTCCAGCTCGGTCATGTGT CCGGACGCACGGTCCCGGTGCCCTGATGGTTCTACCTGCTGT GAGCTGCCCAGTGGGAAGTATGGCTGCTGCCCAATGCCCAAC GCCACCTGCTGCTCCGATCACCTGCACTGCTGCCCCAAGAC ACTGTGTGTGACCTGATCCAGAGTAAGTGCCTCTCCAAGGAG AACGCTACCACGGACCTCCTCACTAAGCTGCCTGCGCACACA GTGGGGGATGTGAAATGTGACATGGAGGTGAGCTGCCCAGA TGGCTATACCTGCTGCCGTCTACAGTCGGGGGCCTGGGGCTG CTGCCCTTTTACCCAGGCTGTGTGCTGTGAGGACCACATACA CTGCTGTCCC GCGGGGTTTACGTGTGACACGCAGAAGGGTAC CTGTGAACAGGGGGCCCCACCAGGTGCCCTGGATGGAGAAGG CCCCAGCTCACCTCAGCCTGCCAGACCCACAAGCCTTGAAGA GAGATGTCCCCTGTGATAATGTCAGCAGCTGTCCCTCCTCCG ATACCTGCTGCCAACTCACGTCTGGGGAGTGGGGCTGCTGTC CAATCCCAGAGGCTGTCTGCTGCTCGGACCACCAGCACTGCT GCCCCAGGGCTACACGTGTGTAGCTGAGGGGCAGTGTGAG CGAGGAAGCGAGATCGTGGCTGGACTGGAGAAGATGCCTGC	
--	--	--

	CCGCCGGGCTTCCTTATCCCACCCCAGAGACATCGGCTGTGA CCAGCACACCAGCTGCCCCGGTGGGGCAGACCTGCTGCCCCGA GCCTGGGTGGGAGCTGGGCCTGCTGCCAGTTGCCCCATGCTG TGTGCTGCGAGGATCGCCAGCACTGCTGCCCCGGCTGGCTACA CCTGCAACGTGAAGGCTCGATCCTGCGAGAAGGAAGTGGTC TCTGCCCAGCCTGCCACCTTCCTGGCCCGTAGCCCTCACGTG GGTGTGAAGGACGTGGAGTGTGGGGAAGGACACTTCTGCCA TGATAACCAGACCTGCTGCCGAGACAACCGACAGGGCTGGG CCTGCTGTCCCTACCGCCAGGGCGTCTGTTGTGCTGATCGGC GCCACTGCTGTCCTGCTGGCTTCCGCTGCGCAGCCAGGGGTA CCAAGTGTGTTGCGCAGGGAGGCCCGCGCTGGGACGCCCTT TGAGGGACCCAGCCTTGAGACAGCTGCTGTAGGTCGACTAA ACAGGCCTATTGATTGGAAAGTATGTCAACGAATTGTGGGTC TTTTGGGGTGTGCTGCCCCCTTTACGCAATGTGGATATCCTGC TTTAATGCCTTTATATGCATGTATAACAAGCAAAACAGGCTTT TACTTTCTCGCCAACTTACAAGGCCTTCTAAGTAAACAGTA TCTGACCCTTTACCCCGTTGCTCGGCAACGGCCTGGTCTGTGC CAAGTGTGTTGCTGACGCAACCCCCACTGGTTGGGGCTTGGCC ATAGGCCATCAGCGCATGCGTGGAACCTTTGTGTCTCCTCTG CCGATCCATACTGCGGAACTCCTAGCCGCTTGTTTTGCTCGC AGCAGGTCTGGAGCGAACTCATCGGGACTGACAATTCTGTG GTGCTCTCCCGCAAGTATACATCGTTTCCAGGGCTGCTAGGC TGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTTTAC GTCCCGTCGGCGCTGAATCCCGCGGACGACCCCTCCCGGGGC CGCTTGGGGCTCTACCGCCCGCTTCTCCGTCTGCCGTACCGA CCGACCACGGGGCGCACCTCTCTTTACGCGGACTCCCGTCT GTGCCTTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCTC TGCACGTGCGATGGAGACCACCGTGAACGCCACCGGAACC TGCCCAAGGTCTTGCATAAGAGGACTCTTGGACTTTCAGCAA TGTCAACTCGAGAGATCTACGGGTGGCATCCCTGTGACCCCT CCCCAGTGCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGC CCACCAGCCTTGTCTAATAAAAATTAAGTTGCATCATTTTGTG TGAAGGTGTCCTTCTATAATATTATGGGGTGGAGGGGGGT GGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGG CCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCA CAATCTTGGCTCACTGCAATCTCCGCTCCTGGGTTCAAGCG ATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCAT GCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGAC GGGGTTTACCATAATTGGCCAGGCTGGTCTCCAACCTCCTAAT CTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGGGAT TACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTT GTAGGTAACCACGTGCGGACCGAGCGGCCGCAGGAACCCCT AGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCT CACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCT TTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGC CTGCAGGAAGCTGTAAGCTCTAGGAGATCCGAACCAGATAA GTGAAATCTAGTTCCAAACTATTTTGTCAATTTTAATTTTCGT	
--	---	--

ATTAGCTTACGACGCTACACCCAGTTCCTCATCTATTTTGTAC
 TCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCC
 AGTTCCCAACTATTTTGTCCGCCACAGCGGGGCATTTTCTT
 CCTGTTATGTTTTTAATCAAACATCCTGCCAACTCCATGTGAC
 AAACCGTCATCTTCGGCTACTTTTTCTCTGTACAGAATGAA
 AATTTTTCTGTATCTCTTCGTTATTAATGTTTGTAAATTGACT
 GAATATCAACGCTTATTTGCAGCCTGAATGGCGAATGGACGC
 GCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTAC
 GCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGC
 TCCTTTCGCTTCTTCCCTTCCTTCTCGCCACGTTCCGCCGGCT
 TTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTC
 GATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATT
 AGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGG
 TTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGG
 ACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGT
 CTATTCCTTTGATTTATAAGGGATTTTGCCGATTTCCGGCCTAT
 TGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAAT
 TTTAACAAAAATTAACGTTTACAATTTTCAGGTGGCACTTTTC
 GGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTCTAAA
 TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGAT
 AAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTC
 AACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTG
 CCTTCCTGTTTTTGCTACCCAGAAACGCTGGTGAAAGTAAA
 AGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCG
 AACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCC
 CCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGC
 TATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGC
 AACTCGGTCCGCCCATACACTATTCTCAGAATGACTTGGTTG
 AGTACTACCAGTCACAGAAAAGCATCTTACGGATGGCATG
 ACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGAT
 AACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCG
 AAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTA
 ACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATA
 CCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGC
 AACAAACGTTGCGCAAACCTATTAACCTGGCGAACTACTTACTCT
 AGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATA
 AAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCT
 GGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTC
 GCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCC
 GTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGG
 ATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTG
 ATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATA
 CTTTAGATTGATTTAAAACCTTCATTTTTTAATTTAAAAGGATCT
 AGGTGAAGATCCTTTTTGATAATCTCATGACCAAAAATCCCTT
 AACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAA
 AGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAAT
 CTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGT

<u>TTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGT</u> <u>AACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCT</u> <u>AGTGTAGCCGTAGTTAGGCCACCACTTCAAGAAGCTCTGTAGC</u> <u>ACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAAGTGGCT</u> <u>GCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCA</u> <u>AGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC</u> <u>GGGGGGTTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCT</u> <u>ACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGC</u> <u>GCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGT</u> <u>AAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC</u> <u>CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTC</u> <u>GCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG</u> <u>GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTT</u> <u>TTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCT</u> <u>TTCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGC</u> <u>CTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGA</u> <u>GCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGA</u> <u>TGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCG</u> <u>CAGACCAGCCGCGTAACCTGGCAAAATCGGTTACGGTTGAGT</u> <u>AATAAATGGATGCCCTGCGTAAGCGGGTGTGGGCGGACAAT</u> <u>AAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAAT</u> <u>AAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGT</u> <u>CCAGTTATGCTGTGAAAAAGCATACTGGACTTTTGTTATGGC</u> <u>TAAAGCAAACCTCTTCATTTTCTGAAGTGCAAATTGCCCGTCG</u> <u>TATTAAAGAGGGGCGTGGCCAAGGGCATGGTAAAGACTATA</u> <u>TTGCGGCGTGTGTGACAATTTACCGAACAACTCCGCGGCCGG</u> <u>GAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACT</u> <u>TGGGTCGATATCAAAGTGCATCACTTCTTCCCGTATGCCCAA</u> <u>CTTTGTATAGAGAGCCACTGCGGGATCGTCACCGTAATCTGC</u> <u>TTGCACGTAGATCACATAAGCACCAAGCGCGTTGGCCTCATG</u> <u>CTTGAGGAGATTGATGAGCGCGGTGGCAATGCCCTGCCTCCG</u> <u>GTGCTCGCCGGAGACTGCGAGATCATAGATATAGATCTCACT</u> <u>ACGCGGCTGCTCAAACCTGGGCAGAACGTAAGCCGCGAGAG</u> <u>CGCCAACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGATG</u> <u>AATGTCTTACTACGGAGCAAGTTCCCGAGGTAATCGGAGTCC</u> <u>GGCTGATGTTGGGAGTAGGTGGCTACGTCTCCGAACACGAC</u> <u>CCGAAAAGATCAAGAGCAGCCCGCATGGATTTGACTTGGTC</u> <u>AGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTGAGCC</u> <u>ACCTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATC</u> <u>GTTGCTGCTGCGTAACATCGTTGCTGCTCCATAACATCAAAC</u> <u>ATCGACCCACGGCGTAACGCGCTTGCTGCTTGGATGCCCGAG</u> <u>GCATAGACTGTACAAAAAACAGTCATAACAAGCCATGAAA</u> <u>ACCGCCACTGCGCCGTTACCACCGCTGCGTTCGGTCAAGGTT</u> <u>CTGGACCAGTTGCGTGAGCGCATACGCTACTTGCAATTACAGT</u> <u>TTACGAACCGAACAGGCTTATGTCAACTGGGTTCGTGCCTTC</u> <u>ATCCGTTTCCACGGTGTGCGTCACCCGGCAACCTTGGGCAGC</u> <u>AGCGAAGTCGAGGCATTTCTGTCTTGCTGGCGAACGAGCGC</u>	
--	--

	AAGGTTTCGGTCTCCACGCATCGTCAGGCATTGGCGGCCTTG CTGTTCTTCTACGGCAAGGTGCTGTGCACGGATCTGCCCTGG CTTCAGGAGATCGGAAGACCTCGGCCGTCGCGGCGCTTGCCG GTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTT CTGGAAGGCGAGCATCGTTTGTTCGCCCAGGACTCTAGCTAT AGTTCTAGTGGTTGGCTACAGCTTGCATG	
Plasmid comprising SNX7 minigene (grey), JeT promoter (underline), Furin/T2A site (bold), and GFP (italicized)	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCGGGC GTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGC CAGAGAGGGAGTGGAATTCGGGCGGAGTTAGGGCGGAGCCAATCA GCGTGCGCCGTTCGGAAGTTGCCTTTATGGCTGGGCGGAGAATG GGCGGTGAACGCCGATGATTATATAAGGACGCGCCGGGTGTGGCA CAGCTAGTTCCGTCGCAGCCGGGATTTGGGTGCGCGTTCCTGTGTTGT GGATCCCTGTGATCGTCACTTGACACCGGTCTTCCAGAGGAGATTG GAAAACCTTGAAGAAGAAGTGGATTGTGCTAATATTGCCCTGAAAGC AGCCACCATGGATTGGGAGAGTTGGAAACAAAATTTGCAAAATTGAT ATCAAGTTAGCATTTACAGATTGGCTGAGGAGAAATATCCATTATTT TGAACAGGTAAATTAGTGTGTGTTGATATTGCTTCATTTAAAGTTAT TTGCTCATTTACTTTTGGTCCGTCCTATTGTTGAAAGAGTGTATTAAA GAACAAGTGTACATTCTATIGCCCTCTCTGGTAGCTTGGTTTTGTGTG AAGTTGTCAGTTACCATTTGGTTTTGTGTTATCCTCAGTTGTGTGTTT GGATTGGATTCTTCAAAAAGCATTGATATTGCTTCTATTGATTGT CCTAACTACTCCTCTTTCCTCTCCCTTCTCCATTTTGAAGAGTTTG AAAGGAAGGAAAGGAGCAGAGACTTGATTGAGCAGAAAATCATT CAGGGCCTGTCTCTATTGTCTTGTCTATCTTGTCTGTAGCTATC TGAAACCATCAACAAAGGAGCACACCATTCCATCAGCAAAAGAGT AACAACATCTTTTAAAGTTTCAATTTGTGTTTTCAGTTGATTGTATTT CAATTTTTTTACAGCTGACTTTTCTCAGAGAAGTTTTTTTTTATGT AAACATACTTTTCTAGAAAGTATATTTTAAAATAACATCTTTAAACC TTATCTCTGGCTGAATTATTGAATATTGAAATTATTACATTAACAA AATTTGTCTTACAGCAGTGGTCCCCAACCTTCTTAGCAGTAGCATC CCTCATTAAGAATTAAAAATTTGTAGAAATTGACAAGGATTCTGACA AGCTGTTGGGAGAGAAGAATAGAGCAGATTGCAGTAGGAACAGT GTGTTAGAATTTATTAATCTTTTAACACTGAAAGTAACTATTGTGT ATTGCCCTTGGTGTGTTTCCATTATTCACTGCTCTTGCTAAGTGGG AGTCATTCTTACATCAACCACCAACCTTCACTTGGGAAGAAGCTAG CGAAGATAAACCT TCGCAACCGCCGCGGCAGCGGCGAAGGCCGC GGCAGCCTGCTGACCTGCGGCGATGTGGAAGAAAACCCGGGCC CGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTG GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGC GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATC TGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACC CTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAAG CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC ATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACG GCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCG TGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCC CCGTGCTGCTGCCCACAACCACTACCTGAGCACCCAGTCCGCCCTGA	93

	GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTG TGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAT GCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTA TAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTAT GTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAATCGATCT GAGGAACCCCTAGTGATGGAGTTGGCCACTCCCCCCCCCCCCCGGC GATTCTCTTGTTTGCTCCAGACTCTCAGGCAATGACCTGATAGCCTT TGTAAGAGACCTCTCAAAAATAGCTACCCTCTCCGGCATGAATTTAT CAGCTAGAACGGTTGAATATCATATTGATGGTGATTTGACTGTCTCC GGCCTTTCTCACCCGTTTGAATCTTTACCTACACATTACTCAGGCAT TGCATTTAAAATATATGAGGGTTCTAAAAATTTTATCCTTGCGTTG AAATAAAGGCTTCTCCCGCAAAAGTATTACAGGGTCATAATGTTTT TGGTACAACCGATTTAGCTTTATGCTCTGAGGCTTTATTGCTTAATT TTGCTAATTCTTTGCCTTGCTGTATGATTTATTGGATGTTGGAATC GCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACAC CGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTT AAGCCAGCCCCGACACCCGCCAACACTATGGTGCACTCTCAGTACA ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAA CACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGC TTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGG TTTTACCGTTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGA TACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAG ACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT GTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAA TAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGA GCCATATTCAACGGGAAACGTCGAGGCCGCGATTAAATTCCAACAT GGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGG CAATCAGGTGCGACAATCTATCGCTTGTATGGGAAGCCCGATGCGC CAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGT TACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCA CTTCCGACCATCAAGCATTTTATCCGTA CTCTGATGATGCATGGTT ACTCACCCTGCGATCCCCGAAAAACAGCGTTCCAGGTATTAGAA GAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGT TCCTGCGCCGGTTGCACTCGATTCTGTTTGTAAATTGTCCTTTTAAC AGCGATCGCGTATTTTCGCTCGCTCAGGCGCAATCACGAATGAATA ACGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTG GCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTC TCACCGGATTCAAGTCGTCATCATGGTGATTTCTCACTTGATAACCT TATTTTGGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGA GTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAAC GCCTCGGTGAGTTTTCTCCTTCAATACAGAAACGGCTTTTTCAAAAA TATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGAT GCTCGATGAGTTTTTCTAACTGTCAGACCAAGTTTACTCATATATAC TTTAGATTGATTTAAACTTCATTTTTAATTTAAAAGGATCTAGGTG AAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTT TTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCT TCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAA AAAACCACCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTA CCAACCTTTTTCCGAAGGTAACCTGGCTTCAGCAGAGCGCAGATAC CAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCCTTCAAG AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACC AGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGAC	
--	--	--

	TCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG GGGGGTTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCG AACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCC CGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGG AACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTA TCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGAT TTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAG CAACGCGGCCTTTTTACGGTTTCTGGCCTTTTGCTGGCCTTTTGCTC ACATGTTCTTTCTGCGTTATCCCCTGATTCTGTGGATAACCGTATT ACCGCCTTTGAGTGAGCTGATACCGCTCGCCGACCCGAACGACCG AGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATAC GCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCAATTAATGCAGCTG GCGTAATAGCGAAGAGGCCCCGACCGATCGCCCTTCCCAACAGTTG CGCAGCCTGAATGGCGAATGGCGATTCCGTTGCAATGGCTGGCGGT AATATTGTTCTGGATATTACCAGCAAGGCCGATAGTTTGAGTTCTTC TACTCAGGCAAGTGATGTTATTACTAATCAAAGAAGTATTGCGACA ACGGTTAATTTGCGTGATGGACAGACTCTTTTACTCGGTGGCCTCAC TGATTATAAAAAACACTTCTCAGGATTCTGGCGTACCGTTCTGTCTA AAATCCCTTTAATCGGCCTCCTGTTTAGCTCCCGCTCTGATTCTAAC GAGGAAAGCACGTTATACGTGCTCGTCAAAGCAACCATAGTACGCG CCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCA GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCT TTCTTCCCTTCCTTTCTCGCCACGTTCCGCGGCTTTCCCCGTCAGCT CTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCA CCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGG CCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCAC GTTCTTTAATAGTGACTCTTGTTCCAAACTGGAACAACACTCAACC CTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTTCG GCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGA ATTTTAACAAAAATTTAACGCTTACAATTTAAATATTTGCTTATACA ATCTTCTGTTTTTGGGGCTTTTCTGATTATCAACCGGGGTACATAT GATTGACATGCTAGTTTTACGATTACCGTTTCATCGCC	
Full SNX7 minigene (version 2)	GAAGAAGAAGATATCAAGTTAGCATTTACAGATTTGGCTGAGGAG AAGAACAGGTAATTAGTGTTGTTTGATATTGCTTCATTTTAAAGTTA TTTGCTCATTTAGCATTTGATATTGCTTTCTATTGATTGTCCTAACTA CTCCTCTTTCCTCTCCCTTCTCCATTTTGAAGAGTTTGCAAAGGAA GGAAAGGAGCAGAGACTTGATTGAGCAGAAAAATCATTTACGGGCC TGTTCTCTATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCTGAAACC ATCAACAAAGGAGCACACCATGGCATCAGCAAAAGAGTAACAACA TCTTTTTTTAAGTTCATTTTGTTCAGTTGATTGATTCAATTTTT TTACAGCTGACTTTTCTCAGAGAAGTTTTTTTTTTTATTGTAAACATA CTTTTTCTAGAAAGTATATTTTAAAAATAACATCTTTAACCTTATCTC TGGCTGAATTATTGAATATTTGAAATTATTACATTAACAAAATTTG TCTTACAGCAGTGGTCCCCAACCTTCTTAGCAGTAGCATCCCTCATT AAGAATTAATAATTTGTAGAAATTGACAAGGATTCTGACAAGCTGTT GGGAGAGAAGAATAGAGCAGATTGCAGTAGGAACAGTTGTGTTAG AATTTATTAATCCTTTAACACTGAAAGTAAACTATTGTTGATTGCCT CTTGGTGTGTTTCCATTATTCAGTGCTCTTGCTAAGTGGGAGTCATT	94

	CCTTACATCAACCACCAACCTTCACTTGGAAGAAGCTAGCGAAGAT AAACCT	
Plasmid comprising modified SNX7 minigene version 2 (grey), JeT promoter (underline), Furin/T2A site (bold), and luciferase (italicized)	GCCCAATACGCAAACCGCCTCTCCCCGCGCCTTGCCGATTTCATTA ATGCAGCTGATTCTAACGAGGAAAGCACGTTATACGTGCTCGTCAA AGCAACCATAGTACGCGCCCTGTAGCGGCGCATTAAAGCGGCGG GTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCT AGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTTCG CGCTTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCC GATTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGG TGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA ACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTATA AGGGATTTTGCCGATTTTCGGCTATTGGTTAAAAAATGAGCTGATT AACAAAAATTTAACGCGAATTTTAAACAAAATATTAACGCTTACAAT TTAAATATTTGCTTATACAATCTTCCTGTTTTTGGGGCTTTTCTGATT ATCAACCGGGGTACATATGATTGACATGCTAGTTTTACGATTACCG TTCATCGCCCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAA AGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGA GCGAGCGCGCAGAGAGGGAGTGGAATTCAATTCGGGCGGAGTTAG GGCGGAGCCAATCAGCGTGCGCCGTTCGAAAGTTGCCTTTTATGG CTGGGCGGAGAATGGGCGGTGAACGCCGATGATTATATAAGGACG CGCCGGGTGTGGCACAGCTAGTTCCGTCGCAGCCGGGATTTGGGTC GCGGTTCTTGTTTGTGGATCCCTGTGATCGTCACTTGACACCGGTGA AGAAGAAGATATCAAGTTAGCATTACAGATTGGCTGAGGAGAA GAACAGGTAATTAGTGTGTGTGATATTGCTTCATTTTAAAGTTATT TGCTCATTTAGCATTIGATATTGCTTTCTATIGATTGTCTAACTACT CCCTCTTCCCTCTCCCTTCTCCATTTTGAAGAGTTTGCAGAAAGGAAGG AAAGGAGCAGAGACTTGATTGAGCAGAAAAATCATTTCAGGGCCTGT TCTCTATTGTCCTTGCTATCCTGCTCTCTGTAAGCTATCTGAAACCATC AACAAAGGAGCACACCATGGCATCAGCAAAAGAGTAACAACATCT TTTTTTAAGTTTCATTTTGTTTTTCAGTTGATTGTAATTCATTTTTTTA CAGCTGACTTTTCTCAGAGAAAGTTTTTTTTTTATTGTAAACATACTTT TTCTAGAAAGTATATTTTAAAAATAACATCTTTAACCTTATCTCTGGC TGAATTATTGAATATTTGAAATTAATACATTAACAAAAATTTTGICTT ACAGCAGTGGTCCCCAACCTTCTTAGCAGTAGCATCCCTCATTAAG AATTAAAAATTGTAGAAATIGACAAGGATTCGACAAGCTGTGTGGG AGAGAAGAAATAGAGCAGATTGCAGTAGGAACAGTTGTGTAGAAAT TTATTAATCCTTTAAACACTGAAAGTAACTATGTGTGATTGCCTCTT GGTGTGTTTCCATTATTCAGTGTCTTTGCTAAGTGGGAGTCATTCCT TACATCAACCACCAACCTTCACTTGGAAGAAGCTAGCGAAGATAAA CC TCGCAACCGCCGCGGCAGCGGCGAAGGCCGCGGCAGCCTGC TGACCTGCGGCGATGTGGAAGAAAACCGGGCCCCGGTCTTCACA CTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTG GACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTCAGAATCTCG GGGTGTCCGTAACCTCCGATCCAAAGGATTGTCTGAGCGGTGAAAATGG GCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGC GACCAAATGGGCCAGATCGAAAAAATTTTAAAGGTGGTGTACCCTGTGG ATGATCATCACTTAAAGGTGATCCTGCACTATGGCACACTGGTAATCGAC GGGGTTACGCCGAACATGATCGACTATTCGGACGGCCGTATGAAGGC ATCGCCGTGTTCGACGGCAAAAAGATCACTGTAAACAGGGACCCTGTGGA ACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCTCCCT	95

	GCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGA ACGCATTCTGGCGTAACCTGCCCCGCTGGGCCTCCCAACGGGGCCCTCC TCCCCCTCCTTGACCAAGCTTATCGATACCGTCGACTAGAGCTCGCT GATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGC CCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGT CCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGG TGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGG GAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAGAGATCGATC TGAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCT CGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCCCCGACGCCCGG GCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGG GAGTGGCCCCCCCCCCCCCCCCCCCCGGCGATTCTCTTGTGTTGCTCCA GACTCTCAGGCAATGACCTGATAGCCTTTGTAGAGACCTCTCAAAA ATAGCTACCCTCTCCGGCATGAATTTATCAGCTAGAACGGTTGAAT ATCATATTGATGGTGATTTGACTGTCTCCGGCCTTTCTCACCCGTTT GAATCTTTACCTACACATTACTCAGGCATTGCATTTAAATATATGA GGGTTCTAAAAATTTTTATCCTTGCGTTGAAATAAAGGCTTCTCCCG CAAAAGTATTACAGGGTCATAATGTTTTTGGTACAACCGATTTAGC TTTATGCTCTGAGGCTTTATTGCTTAATTTTGCTAATTCTTTGCCTTG CCTGTATGATTTATTGGATGTTGGAATCGCCTGATGCGGTATTTTCT CCTTACGCATCTGTGCGGTATTTACACCCGCATATGGTGCCTCTCA GTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCC GCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCA TCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTC AGAGGTTTTACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCT CGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGTTT CTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCC TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGA GACAATAACCCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA GTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCG GCATTTTGCTTCTCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGT AAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGA ACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAA GAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGC GGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTGCGCGC ATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAG AAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGC TGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACA ACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGACAACATGG GGGATCATGTAACCTGCCTTGATCGTTGGGAACCGGAGCTGAATGA AGCCATAACAAACGACGAGCGTGACACCACGATGCCTGTAGCAAT GGCAACAACGTTGCGCAAACTATTAAGTGGCGAACTACTTACTCTA GCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTG CAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCT GATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAG CACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAC GACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGC TGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGTGTACAGACAA GTTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTTAATTT AAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAA TCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAA AAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTG	
--	--	--

	CTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTGTTG CCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCA GCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTT AGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCT CTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGT GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCA GCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGA GCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGA GAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCG GTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCA GGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCT CTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCT TTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT CTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCG CCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC GGAAGAGC	
--	--	--

In various embodiments, a minigene or vector disclosed herein may be used to increase in the levels of functional polypeptide, e.g., the level of hPGRN, in response to the presence or absence of splice modulator. In some embodiments, a vector disclosed herein exhibits higher expression of the transgene sequence in the presence of a splice modulator compared to the expression of the same vector in the absence of the splice modulator. In some embodiments, the level of expression of the molecule of interest in the presence of the splice modulator is greater, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 fold greater, than the level of expression of the molecule of interest in the absence of the splice modulator. In some embodiments, the level of expression of the molecule of interest in the absence of the splice modulator is greater, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 fold greater, than the level of expression of the molecule of interest in the presence of the splice modulator. In some embodiments, the increase in expression of the transgene sequence is measured by an increase in the number of RNA transcripts of the transgene sequence. In some embodiments, the increase in expression of the transgene sequence is measured by PCR. In some embodiments, the increase in expression of the transgene sequence is measured by RT-PCR. In some embodiments, the increase in expression of the transgene sequence is measured by qPCR. In some embodiments, the increase in expression of the transgene sequence is measured by qRT-PCR. In some embodiments, the increase in expression of the transgene sequence is measured by sequencing. In some embodiments, the increase in expression of the transgene sequence is measured by Northern blot analysis. In some embodiments, the increase in expression of the transgene sequence is measured by single-molecule Fluorescence In-Situ Hybridization (FISH). In some embodiments, the increase in expression of the transgene sequence is measured by an increase in the amount of protein encoded by the transgene produced. In some embodiments, the increase in expression of the transgene sequence is measured by an enzyme-linked immunosorbent assay (ELISA). In some embodiments, the increase in

expression of the transgene sequence is measured by Western blot analysis. In some embodiments, the increase in expression of the transgene sequence is measured by immunostaining. In some embodiments, the increase in expression of the transgene sequence is measured by more than one of the above listed methods. In some embodiments, the increase in expression of the transgene sequence is measured by the amount of mRNA which includes the second exon. In some embodiments, the increase in expression of the transgene sequence is measured by the amount of mRNA which includes a direct first exon to third exon splice. Exemplary polypeptides produced in the presence or absence of splice modulator from vectors incorporating either an on-switch minigene or an off-switch minigene are depicted in Figure 3 (in each case, prior to cleavage of the protease cleavage site and/or self-cleaving peptide sequence).

Recombinant virus

In various embodiments, the nucleic acids and vectors discussed herein may be present in one or more virus particle, such as a recombinant virus particle. Recombinant viruses are viruses generated by recombinant means. Various different viral types may be used, e.g., retroviruses, adenovirus, lentivirus, AAV, murine leukemia viruses, etc. Without being bound by theory, vectors delivered from retroviruses such as the lentivirus may provide for long-term gene transfer since they allow long-term, stable integration of a transgene and its propagation in daughter cells and may also provide low immunogenicity. Other suitable retroviruses include gammaretroviruses. Exemplary gammaretroviral vectors include Murine Leukemia Virus (MLV), Spleen-Focus Forming Virus (SFFV), and Myeloproliferative Sarcoma Virus (MPSV), and vectors derived therefrom. Other gammaretroviral vectors are described, e.g., in Tobias Maetzig et al., "Gammaretroviral Vectors: Biology, Technology and Application" *Viruses*. 2011 Jun; 3(6): 677–713. In some embodiments, the virus is a recombinant adenovirus comprising a nucleic acid or vector disclosed herein. In some embodiments, the virus is a recombinant AAV comprising a nucleic acid or vector disclosed herein.

In some embodiments, the nucleic acids or vectors disclosed herein are for use in the manufacture of a recombinant virus. In some embodiments, the nucleic acids or vectors disclosed herein are for use in the manufacture of an rAAV. Thus, also disclosed herein, in various embodiments, are virus compositions (also referred to as virions), e.g., rAAV virus compositions comprising a viral vector or nucleic acid disclosed above. In some embodiments, the recombinant virus is an adeno-associated virus (AAV) or any mutant or derivative thereof. In some embodiments, the recombinant virus is a chimeric AAV or any mutant or derivative thereof. In some embodiments, the recombinant virus is an adenovirus or any mutant or derivative thereof. In some embodiments, the recombinant virus is a retrovirus or any mutant or derivative thereof. In some embodiments, the recombinant virus is a lentivirus or any mutant or derivative thereof. In some embodiments, the recombinant virus is a DNA virus or any mutant or derivative thereof. In some embodiments, the recombinant virus is a herpes simplex virus or any mutant or derivative thereof. In some embodiments, the recombinant virus is a baculovirus or any mutant or derivative thereof.

In some embodiments, an AAV disclosed herein may comprise one or more AAV capsid proteins. AAV capsid proteins may be from any AAV serotype for which a recombinant virus can be derived including, but not limited to, AAV serotypes AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10, AAV-11, AAV-12, AAVrh8, AAVrh10, AAV-DJ, AAV-DJ/8, AAV-PHP.B, AAV-PHP.B2, AAV-PHP.B3, AAV-PHP.A, AAV-PHP.eB, and AAV-PHP.S. In some embodiments, one or more capsid protein in an AAV is from an AAV-9. Without being bound by theory, typically in AAV, three capsid proteins, VP1, VP2 and VP3 multimerize to form the capsid. The polypeptide sequences of capsid proteins are known in the art, and can also be derived from the genome of the AAV. These can be used as exemplary capsids in the AAV virus compositions disclosed herein. For example, the complete genome of AAV-1 is provided in GenBank Accession No. NC_002077; the complete genome of AAV-2 is provided in GenBank Accession No. NC_001401 and Srivastava et al., *Virol.*, 45: 555-564 (1983); the complete genome of AAV-3 is provided in GenBank Accession No. NC_01829; the complete genome of AAV-4 is provided in GenBank Accession No. NC_001829; the AAV-5 genome is provided in GenBank Accession No. AF085716; the complete genome of AAV-6 is provided in GenBank Accession No. NC_001862; at least portions of AAV-7 and AAV-8 genomes are provided in GenBank Accession Nos. AX753246 and AX753249, respectively; the AAV-9 genome is provided in Gao et al., *J. Virol.*, 78: 6381-6388 (2004); the AAV-10 genome is provided in Williams, (2006) *Mol. Ther.*, 13(1): 67-76; and the AAV-11 genome is provided in Mori et al., (2004) *Virology*, 330(2): 375-383. Capsid proteins AAV-PHP.B, AAV-PHP.B2, AAV-PHP.B3, AAV-PHP.A, AAV-PHP.eB, or AAV-PHP.S are provided in Deverman et al., (2016) *Nat. Biotech.*, 34: 204-209 and Chan et al., (2017) *Nat. Neurosci.*, 20: 1172-1179. In some embodiments, the recombinant virus is an AAV comprising one or more AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV 8, AAV9, AAV10, and AAV11, AAV 12, AAVrh8, AAVrh10, AAV-DJ, AAV-DJ/8, AAV-PHP.B, AAV-PHP.B2, AAV-PHP.B3, AAV-PHP.A, AAV-PHP.eB, or AAV-PHP.S capsid serotype, or a functional variant thereof. In some embodiments, the recombinant virus is an AAV comprising a combination of capsids from more than one AAV serotype.

In some embodiments, AAV compositions disclosed herein comprise one or more cis-acting sequences directing viral DNA replication (rep), encapsidation/packaging and host cell chromosome integration are contained within the ITRs. In some embodiments, one or more of these sequences may also be present in trans rather than cis, e.g., on a separate plasmid during the virus manufacturing process in a host cell. Typically, three AAV promoters (named p5, p19, and p40 for their relative map locations) drive the expression of the two AAV internal open reading frames encoding rep and cap genes in wild-type virus. In some embodiments, one or more of these promoters and/or open reading frames are present in cis in an AAV vector and/or AAV virion disclosed herein, or are present on separate plasmids during the AAV virus manufacturing process, e.g., in a host cell producing the virus. The two rep promoters (p5 and p19), coupled with the differential splicing of the single AAV intron (at nucleotides 2107 and 2227), may result in the production of four rep proteins (rep 78, rep 68, rep 52, and rep 40) from the rep gene. Rep proteins possess multiple enzymatic properties that are ultimately responsible for

replicating the viral genome. The cap gene is typically expressed from the p40 promoter and it encodes the three capsid proteins VP1, VP2, and VP3. Alternative splicing and non-consensus translational start sites are responsible for the production of the three related capsid proteins. A single consensus polyadenylation site is located at map position 95 of the AAV genome. The life cycle and genetics of AAV are reviewed in Muzyczka, (1992) *Curr. Topics Microbiol. Imm.*, 158: 97-129.

In some embodiments, the AAV capsid proteins VP1, VP2, VP3 used in the AAV disclosed herein are encoded by or comprise the following sequences:

VP1 nucleic acid (SEQ ID NO: 74):

atggctgccgatgggtatctccagattggctcaggacaaccttagtaaggaattcgcgagtggtgggcttgaacctggagcccccaacccaa
ggcaaatcaacaatcaagacaacgctcgaggtctgtgcttcgggttacaataccttggacccggcaacggactcgacaagggggagccg
gtcaacgcagcagacgcggcgccctcgagcagcacaaggcctacgaccagcagctcaaggccggagacaacccgtacctaagtaacaacc
acgcccagcgcgagttccaggagcggctcaaaagaagatcgtctttgggggcaacctgggagcagctctccaggccaaaagaggctctt
gaacctctggtctggttaggaagcggctaagacggctcctggaagaagaggcctgtagagcagctcctcaggaaacggactcctccgccc
tattggcaaatcgggtgcacagcccgtaaaaagagactcaatttcggctcagactggcgacacagagtcagtcaccagacctcaaccaatcgga
gaacctcccgagccccctcagggtgtggatcttacaatggcttcagggtgtggcgacacagagtcagtcaccagacctcaaccaatcgga
gggtagttcctcgggaatttggcattgcgattccaatggctggggacagagtcacaccagcaccggaacctgggcccctgcccacctaca
acaatcacctctacaagcaaatctcaacagcacatctggaggatctcaaatgacaacgcctactcggctacagacccccctgggggtatttga
cttaacagattccactgccattctcaccacgtgactggcagcagctcatcaacaacaactgggattccggcctaagcagactcaactcaagctct
tcaacattcaggtaaaagagggtacggacaacaatggagtaagaccatcgccaataacctaccagcaggtccagggtctcagggactcagac
tatcagctcccgtacgtgctcgggtcggctcagagggtcgcctcccgccgttccagcggacgtttcatgattcctcagtaagggtatctgacgctta
atgatggaagccaggccgtgggtcgtctcctttactgctggaatatttccgctgcaaatgctaagaacgggtaacaactccagttcagctacg
agtttgagaacgtaccttccatagcagctacgctcagcacaagcctggaccgactaatgaatccactcatgaccaatactgtactatctctcaa
agactattaacggttctggacagaatcaacaacgctaaaattcagttgtggcggaccagcaaatggctgtccagggaagaaactacatacct
ggaccagctaccgacaacaacgtgtctcaaccactgtgactcaaaaacaacaacagcgaatttctggcctggagcttctctgggtctcaatgg
acgtaatatgcttgatgaatcctggacctgctatggccagccacaagaaggagaggaccgtttcttcttctgtctggtatcttaattttggcaacaag
gaactggaagagacaacgttgatgctggacaaaagtcagtataaccaacgaagaagaattaaaactactaaccggtagcaacggagtccat
ggacaagtggccacaaccaccagagtgcccaagcagaggcgacacgggtgggttcaaaaccaaggaatactccgggtatggtttggcag
gacagagatgtgtacctgcaaggacccatttgggcaaaaattcctcacaggacggcaacttccaccttctcggctgatggagggtttggaatga
agacccgctcctcagatcctcatcaaaaacacacctgtacctgagatcctccaacggccttcaacaaggacaagctgaactcttcatcccc
agtattctactggccaagtgcagctggagatcagtgaggagctgcagaaggaaaacagcaagcgtggaacccggagatccagtaacttcca
actattacaagtctaataatgttgaattgtgttaatactgaagggtatatagtgaacccgccccattggcaccagatacctgactcgtaatctgtaa

VP2 nucleic acid (SEQ ID NO: 75):

acggctcctggaagaagaggcctgtagagcagctcctcaggaaacggactcctccgcccgtattggcaaatcgggtgcacagcccgtaaaa
agagactcaattcgggtcagactggcgacacagagtcagtcaccagacctcaaccaatcgagaaacctcccgagccccctcagggtgtggatct
cttacaatggctcagggtgggtggcaccagtgccagacaataacgaagggtccgatggagtggttagttcctcgggaaattggcattgcgattccc
aatggctgggggacagagtcacaccagcaccggaacctgggcccctgcccacctacaacaatcacctctacaagcaaatctcaacagca
catctggaggatcttcaaatgacaacgcctactcggctacagcaccctgggggtatttgaacttcaacagattccactgtccacttctcaccacgtg
actggcagcagactcatcaacaacaactgggattccggcctaagcagactcaactcaagctctcaacattcagggtcaagagggttacggacaac
aatggagtaagaccatcgccaataacctaccagcaggtccagggtctcagggactcagactatcagctcccgtacgtgctcgggtcggctcac
gagggtcgtcctcccgccgttccagcggacgttttcatgattcctcagtagcgggtatctgacgcttaatatggaagccaggccgtgggtcgtctcct
ttactgcctggaatatttccgctgcaaatgctaagaacgggtaacaactccagttcagctacgagtttgagaacgtaccttccatagcagctacgc
tcacagccaaagcctggaccgactaatgaatccactcatgaccaatactgtactatctctaaagactattaacggttctggacagaatcaaca
acgctaaaattcagtggtggcggaccagcaacatggctgtccagggaagaaactacatacctggaccagctaccgacaacaacggttctcaa
ccactgtgactcaaaaacaacagcgaatttctggcctggagcttcttctgggtctcaatggacgtaatactgtatgaatcctggacctgctat
ggccagccacaagaaggagaggaccgttttcttcttctgtggtatcttaattttggcaacaaggaaactggaagagacaacgttgatgcggac
aaagtcatgataaccaacgaagaagaattaaaactactaaccggtagcaacggagtcctatggacaagtggccacaaccaccagagtgcc
ccaagcacaggcgacagaccggctgggttcaaaaccaaggaatactccgggtatggtttggcaggacagagatgtgtacctgcaaggacccattt

gggccaaaattcctcacacggacggcaactttacccttctccgctgatgggaggggttgaatgaagcaccgcctcctcagatcctcatcaaaaa
 cacacgtgtacgtcgatcctcaacggccttcaacaaggacaagctgaactcttcatcaccagttatctactggccaagtcagcgtggagatc
 gagtgggagctgcagaaggaaaacagcaagcgtggaacccggagatccagtacactccaactattacaagtctaataatgtgaattgtctgta
 atactgaagggtatatagtgaacccgccccattggcaccagatacctgactcgtaatctgtaa

VP3 nucleic acid (SEQ ID NO: 76):

atggcttcagggtggcgccaccagtggcagacaataacgaaggtgcgatggagtggttagttcctcgggaattggcattgcgattcccaatgg
 ctgggggacagagtcaccaccagcaccgaacctgggacctgccacctacaacaatcacctctacaagcaaattccaacagcacatctg
 gaggtatctcaaatgacaacgctactctggctacagcaccctgggggtattttgactcaacagattccactgccactctcaccacgtgactggc
 agcgtactcatcaacaacactggggtattccggcctaagcgtactcaactcaagctcttaacattcagggtcaagaggttacggacaacaatgga
 gtcaagaccatcgccaataaccttaccagcaggtccagggtctcacggactcagactatcagctccgtacgtgctcgggtcggctcacgagggc
 tgcctccgctgctccagcggacgttttcatgattcctcagtaacgggtatctgacgcttaatgatggaagccaggccgtgggtcgtctctttactgc
 ctggaatatttccgctgcgaatgctaagaacgggtaacaactccagttcagctacgagtttgagaacgtaccttccatagcagctacgctcacag
 ccaagcctggaccgactaataatccactcatcgaccaatactgtactatctctcaagactattaacgggttctggacagaatacaaaacgcta
 aaattcagtggtggcggaccagcaacatggctgtccagggaagaaactacatacctggaccagctaccgacaacaacgtgtctcaaccactgt
 gactcaaaacaacaacagcgaatttctggcctggagctcttctgggctctaatggacgtaataagcttgatgaatcctggacctgctatggccag
 ccacaaagaaggagaggaccgtttcttcttctgtctggtatcttaattttggcaacaaggaactggaagagacaacgtggatgctggacaaagtc
 atgataaccaacgaagaagaataaaaactactaaccggtagcaacggagtcctatggacaagtgccacaaaccaccagagtgcccaagc
 acaggcgagaccggctgggttcaaaacgaaggaatacttccgggtatggttggcaggacagagatgtgtacctgcaaggaccatttgggcca
 aaattcctcacacggacggcaactttacccttctccgctgatgggaggggttgaatgaagcaccgcctcctcagatcctcatcaaaaaacacacc
 tgtacctgcggtatcctcaacggccttcaacaaggacaagctgaactcttcatcaccagttatctactggccaagtcagcgtggagatcgagtg
 gagctgcagaaggaaaacagcaagcgtggaacccggagatccagtacactccaactattacaagtctaataatgtgaatttctgttaatactg
 aagggtgatatagtgaacccgccccattggcaccagatacctgactcgtaatctgtaa

VP1 Protein (SEQ ID NO: 77):

MAADGYLPDWLEDNLSEGIREWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPNGLDKGEPVNA
 ADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKEDTSFGGNLGRAVFQAKKRLLLEPLGLVEEAA
 KTAGPKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDPQPIGEPPAAPSGVGSLTMAS
 GGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVITSTRTWALPTYNNHLYKQISNSTSGGSSNDN
 AYFGYSTPWGYFDFNRFHCHFSRPRDWQRLINNNWGFPRKRLNFKLFNIQVKEVTDNNGVKTIANNLTST
 VQVFTDSYQLPYVLGSAHEGCLPPFPADVFMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNN
 FQFSYEFENVFPHSSYAHSQSLDRLMNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIP
 GPSYRQQRVSTTVTQNNNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQG
 TGRDNVDADKVMITNEEEIKTTNPVATESYQVATNHQSAQAQAQTGWVQNQGILPGMVWQDRDVLQ
 GPIWAKIPHTDGNFHPSPMLMGFGMKHPPQILIKNTVPADPPTAFNKDKLNSFITQYSTGQVSVEIEWE
 LQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL

VP2 Protein (SEQ ID NO: 78):

TAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDPQPIGEPPAAPSGVGSLTMAS
 GGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVITSTRTWALPTYNNHLYKQISNSTSGGSSNDN
 AYFGYSTPWGYFDFNRFHCHFSRPRDWQRLINNNWGFPRKRLNFKLFNIQVKEVTDNNGVKTIANNLTST
 VQVFTDSYQLPYVLGSAHEGCLPPFPADVFMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNN
 FQFSYEFENVFPHSSYAHSQSLDRLMNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIP
 GPSYRQQRVSTTVTQNNNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQG
 TGRDNVDADKVMITNEEEIKTTNPVATESYQVATNHQSAQAQAQTGWVQNQGILPGMVWQDRDVLQ
 GPIWAKIPHTDGNFHPSPMLMGFGMKHPPQILIKNTVPADPPTAFNKDKLNSFITQYSTGQVSVEIEWE
 LQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL

VP3 Protein (SEQ ID NO: 79):

MASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVITSTRTWALPTYNNHLYKQISNSTSGGSS
 NDNAYFGYSTPWGYFDFNRFHCHFSRPRDWQRLINNNWGFPRKRLNFKLFNIQVKEVTDNNGVKTIANNL

TSTVQVFTDSYQLPYVLGSAHEGCLPPFPADVFMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRT
 GNNFQFSYEFENVPFHSSYAHSQSLDRLMNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGR
 NYIPGPSYRQQRVSTTVTQNNNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFG
 KQGTGRDNVDADKVMITNEEEIKTTNPVATESYGGVATNHQSAQAQAQTGWVQNQGILPGMVWQDRD
 VYLQGPWAKIPHTDGNFHPSPLMGGFGMKHPPQILIKNTPVPADPPTAFNKDKLNSFITQYSTGQVSVE
 IEWELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL

In one embodiment, the recombinant virus is an AAV comprising an AAV9 capsid serotype or any mutant or derivative thereof. In some embodiments, the recombinant virus comprises AAV9 capsid proteins VP1, VP2 and VP3. In some embodiments, the recombinant virus is a scAAV.

In some embodiments, a recombinant virus may be used to increase the levels of functional polypeptides in specific cell types. In some embodiments, the virus disclosed herein exhibits higher expression of the transgene sequence in a specific tissue type as compared to the expression of the same virus in a different tissue type. In some embodiments, the virus exhibits higher expression of the transgene sequence in a neuronal tissue, fluid or cell as compared to the expression of the same virus in a non-neuronal tissue, fluid or cell. In some embodiments, a vector disclosed herein exhibits higher expression of the transgene sequence in the presence of a splice modulator compared to the expression of the same vector in the absence of the splice modulator. In some embodiments, the level of expression of the molecule of interest from the recombinant virus in the presence of the splice modulator is greater, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 fold greater, than the level of expression of the molecule of interest from the recombinant virus in the absence of the splice modulator. In some embodiments, the level of expression of the molecule of interest from the recombinant virus in the absence of the splice modulator is greater, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 fold greater, than the level of expression of the molecule of interest from the recombinant virus in the presence of the splice modulator. In some embodiments, the increase in expression of the transgene sequence is measured by an increase in the number of RNA transcripts of the transgene sequence. In some embodiments, the increase in expression of the transgene sequence is measured by PCR. In some embodiments, the increase in expression of the transgene sequence is measured by RT-PCR. In some embodiments, the increase in expression of the transgene sequence is measured by qPCR. In some embodiments, the increase in expression of the transgene sequence is measured by qRT-PCR. In some embodiments, the increase in expression of the transgene sequence is measured by sequencing. In some embodiments, the increase in expression of the transgene sequence is measured by Northern blot analysis. In some embodiments, the increase in expression of the transgene sequence is measured by single-molecule Fluorescence In-Situ Hybridization (FISH). In some embodiments, the increase in expression of the transgene sequence is measured by an increase in the amount of protein encoded by the transgene produced. In some embodiments, the increase in expression of the transgene sequence is measured by an enzyme-linked immunosorbent assay (ELISA). In some embodiments, the increase in expression of the transgene sequence is measured by Western blot analysis. In some embodiments, the increase in expression of the transgene sequence is measured by immunostaining. In some

embodiments, the increase in expression of the transgene sequence is measured by more than one of the above listed methods. In some embodiments, the increase in expression of the transgene sequence is measured by the amount of mRNA which includes the second exon. In some embodiments, the increase in expression of the transgene sequence is measured by the amount of mRNA which includes a direct first exon to third exon splice. Exemplary polypeptides produced in the presence or absence of splice modulator from vectors incorporating either an on-switch minigene or an off-switch minigene are depicted in Figure 3 (in each case, prior to cleavage of the protease cleavage site and/or self-cleaving peptide sequence). It is contemplated that once the polypeptide comprising the protease cleavage site and/or self-cleaving peptide sequence, the sequence(s) are cleaved such that the protein of interest is produced without (or with fewer than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids of) heterologous sequence derived from the minigene or cleavage sequences.

In various embodiments, the target cells of this disclosure may be any mammalian cell type. In some aspects of this disclosure, the nucleic acids and vectors regulate expression in a neuronal tissue or fluid or cell. In some embodiments, the neuronal tissue is the brain. In some embodiments, the neuronal tissue is the frontal lobe of the brain. In some embodiments, the neuronal tissue is the temporal lobe of the brain. In some embodiments, the neuronal tissue is the central nervous system. In some embodiments, the neuronal tissue is the spinal cord. In some embodiments, the neuronal cell is a human neuronal cell. In some embodiments, the neuronal cell is a neuron. In some embodiments, the neuronal cell is an astrocyte. In some embodiments, the neuronal fluid is cerebrospinal fluid. In some embodiments, a non-neuronal tissue is the liver. In some embodiments, the non-neuronal fluid is plasma. In some embodiments, a non-neuronal cell is a hepatocyte. In some embodiments, a non-neuronal cell is a stellate fat storing cell. In some embodiments, a non-neuronal cell is a Kupffer cell. In some embodiments, a non-neuronal cell is a liver endothelial cell. In some embodiments, the non-neuronal fluid is plasma. In some embodiments, the non-neuronal fluid is serum. In some embodiments, the non-neuronal fluid is blood.

Methods of Producing Recombinant Virus

Also disclosed herein, in various embodiments, are methods of producing recombinant virus comprising neuron specific promoters. In some embodiments, nucleic acid sequences, e.g., plasmids encoding an AAV or other viral genome, are used to produce the recombinant virus. In some embodiments, nucleic acid sequences, e.g., plasmids, comprising an AAV rep gene and/or an AAV cap gene are also used in preparing the AAV or other virus. Also disclosed herein are nucleic acid sequences, e.g., plasmids, comprising an adenovirus helper function gene. In some embodiments, the nucleic acids encoding the AAV rep, AAV cap, and/or adenovirus helper genes may be present in the same structure, e.g., a single plasmid, or they may be present in separate structures. In some embodiments, the one or more plasmids are cotransfected with the nucleic acid encoding the AAV vector into competent cells, and the cells are cultured to produce the recombinant virus. In some cases, the plasmids encoding AAV viral

genome and AAV rep and/or cap genes are transferred to cells permissible for infection with a helper virus of AAV (e.g., adenovirus, E1-deleted adenovirus or herpesvirus). In some embodiments, the rAAV genome is assembled into infectious viral particles with AAV capsid proteins in the cells after transfection. Techniques to produce rAAV particles, in which an AAV genome to be packaged, rep and cap genes, and helper virus functions are provided to a cell are known in the art and may include, e.g., electroporation. In some embodiments, production of rAAV involves the following components present within a single cell (denoted herein as a packaging cell): a rAAV vector, AAV rep and cap genes separate from (i.e., not in) the rAAV vector, and helper virus functions. Production of pseudotyped rAAV is disclosed in, for example, WO 01/83692 which is incorporated by reference herein in its entirety. In various embodiments, AAV capsid proteins may be modified to enhance delivery of the recombinant vector. Modifications to capsid proteins are generally known in the art. See, for example, US 2005/0053922 and US 2009/0202490, the disclosures of which are incorporated by reference herein in their entirety.

In various embodiments, general principles of viral vector production may be utilized to produce the vectors and virus, e.g., rAAV, disclosed herein. Carter, (1992) *Curr. Opinions Biotech.*, 1533-539; Muzyczka, (1992) *Curr. Topics Microbial. Immunol.*, 158:97-129. Various approaches are disclosed in Ratschin et al., (1984) *Mol. Cell. Biol.*, 4: 2072; Hennonat et al., (1984) *Proc. Natl. Acad. Sci. USA*, 81: 6466; Tratschin et al., (1985) *Mol. Cell. Biol.*, 5: 3251; McLaughlin et al., (1988) *J. Virol.*, 62: 1963; Lebkowski et al., (1988) *Mol. Cell. Biol.*, 7:349; Samulski et al. (1989) *J. Virol.*, 63:3822-3828; U.S. Pat. No. 5,173,414; WO 95/13365 and corresponding U.S. Pat. No. 5,658,776; WO 95/13392; WO 96/17947; PCT/US98/18600; WO 97/09441 (PCT/US96/14423); WO 97/08298 (PCT/US96/13872); WO 97/21825 (PCT/US96/20777); WO 97/06243 (PCT/FR96/01064); WO 99/11764; Perrin et al., (1995) *Vaccine*, 13: 1244-1250; Paul et al., (1993) *Hum. Gene Ther.*, 4: 609-615; Clark et al. (1996) *Gene Therapy*, 3: 1124-1132; U.S. Pat. No. 5,786,211; U.S. Pat. No. 5,871,982; and U.S. Pat. No. 6,258,595. The foregoing documents are hereby incorporated by reference in their entirety herein, with particular emphasis on those sections of the documents relating to rAAV production.

An exemplary method of generating a packaging cell is to create a cell line that stably expresses all the necessary components for AAV particle production. For example, a plasmid (or multiple plasmids) encoding a rAAV vector lacking AAV rep and cap genes, AAV rep and cap genes separate from the rAAV vector, and a selectable marker, such as a neomycin resistance gene, are integrated into the genome of a cell. AAV genomes have been introduced into bacterial plasmids by procedures such as GC tailing (Samulski et al., (1982) *Proc. Natl. Acad. Sci. USA*, 79: 2077-2081), addition of synthetic linkers containing restriction endonuclease cleavage sites (Laughlin et al., (1983) *Gene*, 23:65-73) or by direct, blunt-end ligation (Senapathy et al., (1984) *J. Biol. Chem.*, 259: 4661-4666). The packaging cell line is then infected with a helper virus such as adenovirus and/or a plasmid encoding a helper virus. The advantages of this method are that the cells are selectable and are suitable for large-scale production of

rAAV. Other examples of suitable methods employ adenovirus or baculovirus rather than plasmids to introduce rAAV vectors and/or rep and cap genes into packaging cells.

In some embodiments, a method of producing recombinant virus comprises providing a nucleic acid to be packaged. In some embodiments, the nucleic acid is a plasmid. In other embodiments, the nucleic acid comprises a transgene sequence interposed between a first AAV terminal repeat and a second AAV terminal repeat. In some embodiments, the transgene encodes human progranulin (hPGRN). In some embodiments, the method of producing recombinant virus comprises providing one or more additional nucleic acids. In some embodiments, the one or more additional nucleic acids comprises an AAV rep gene and/or an AAV cap gene. In some embodiments, the one or more additional nucleic acids comprises an AAV rep gene derived from an AAV serotype 1, AAV serotype 2, AAV serotype 3, AAV serotype 4, AAV serotype 5, AAV serotype 6, AAV serotype 7, AAV serotype 8, or AAV serotype 9. In some embodiments, the one or more additional nucleic acids comprises an AAV cap gene derived from an AAV serotype 1, AAV serotype 2, AAV serotype 3, AAV serotype 4, AAV serotype 5, AAV serotype 6, AAV serotype 7, AAV serotype 8, or AAV serotype 9. In some embodiments, the one or more additional nucleic acids comprises one or more of an adenovirus helper function gene.

In some embodiments, the nucleic acids are co-transfected into competent cells or packaging cells. Methods of co-transfection are known in the art, and include, but are not limited to, transfection by lipofectamine, electroporation, and polyethylenimine. Competent cells or packaging cells may be non-adherent cells cultured in suspension or adherent cells. In one embodiment any suitable packaging cell line may be used, such as HeLa cells, HEK 293 cells and PerC.6 cells (a cognate 293 line). In one embodiment, the packaging cells are human cells. In one embodiment, the packaging cells are HEK 293 cells. In one embodiment, the packaging cells are insect cells. In one embodiment, the packaging cells are Sf9 cells. In some embodiments, the method comprises culturing the transfected cells to produce recombinant virus. In some embodiments, the method comprises recovering the recombinant virus. Methods of recovering recombinant virus include, e.g., those disclosed in U.S. Patent No. 6,143,548 and U.S. Patent No. 9,408,904. In some embodiments, recombinant virus is secreted into cell culture media and purified from the media. In some embodiments, packaging cells are lysed, and the contents purified to recover the recombinant virus. In some embodiments, the virus is recovered from the packaging cell by filtration or centrifugation. In some embodiments, the virus is recovered from the packaging cell by chromatography.

In various embodiments, disclosed herein are cells comprising the nucleic acids disclosed herein, cells comprising the vectors disclosed herein, or cells comprising the viruses disclosed herein. The cells comprising the nucleic acids disclosed herein, cells comprising the vectors disclosed herein, or cells comprising the viruses disclosed herein, may be human cells. The cells comprising the nucleic acids disclosed herein, cells comprising the vectors disclosed herein, or cells comprising the viruses disclosed herein, may also be insect cells. In some embodiments, the cells comprising the nucleic acids disclosed

herein, cells comprising the vectors disclosed herein, or cells comprising the viruses disclosed herein are HEK293 cells. In some other embodiments, the cells comprising the nucleic acids disclosed herein, cells comprising the vectors disclosed herein, or cells comprising the viruses disclosed herein are Sf9 cells.

In some embodiments, the method of producing recombinant virus comprises transfecting an insect cell. In some embodiments, the method comprises transfecting an insect cell with a baculovirus comprising the nucleic acids as disclosed herein. In some embodiments, the method comprises transfecting an insect cell with baculovirus comprising a nucleic acid comprising a transgene sequence interposed between a first AAV terminal repeat and a second AAV terminal repeat. In some embodiments, the method comprises transfecting an insect cell with a baculovirus comprising one or more additional nucleic acids. In some embodiments, the one or more additional nucleic acids comprises an AAV rep gene and/or an AAV cap gene. In some embodiments, the one or more additional nucleic acids comprises an AAV rep gene derived from an AAV serotype 1, AAV serotype 2, AAV serotype 3, AAV serotype 4, AAV serotype 5, AAV serotype 6, AAV serotype 7, AAV serotype 8, or AAV serotype 9. In some embodiments, the one or more additional nucleic acids comprises an AAV cap gene derived from an AAV serotype 1, AAV serotype 2, AAV serotype 3, AAV serotype 4, AAV serotype 5, AAV serotype 6, AAV serotype 7, AAV serotype 8, or AAV serotype 9.c. In some embodiments, the one or more additional nucleic acids comprises one or more of an adenovirus helper function gene. In some embodiments, the insect cells are cultivated under conditions suitable to produce recombinant virus. In some embodiments, the virus is recovered from the insect cell. In some embodiments, the virus is recovered from the insect cell by filtration or centrifugation. In some embodiments, the virus is recovered from the insect cell by chromatography.

Pharmaceutical compositions

In various embodiments, pharmaceutical compositions are disclosed. In some embodiments, a pharmaceutical composition comprises one or more nucleic acids, vectors and/or viruses disclosed herein. In some embodiments, the pharmaceutical composition comprises a pharmaceutically acceptable carrier.

The nucleic acids, vectors, and/or recombinant virus according to the present disclosure (e.g., viral particles) can be formulated to prepare pharmaceutically useful compositions. Exemplary formulations include, for example, those disclosed in U.S. Patent No. 9,051,542 and U.S. Patent No. 6,703,237, which are incorporated by reference in their entirety. The compositions of the disclosure can be formulated for administration to a mammalian subject, e.g., a human. In some embodiments, delivery systems may be formulated for intramuscular, intradermal, mucosal, subcutaneous, intravenous, intrathecal, injectable depot type devices, or topical administration.

In some embodiments, when the delivery system is formulated as a solution or suspension, the delivery system is in an acceptable carrier, e.g., an aqueous carrier. A variety of aqueous carriers may be

used, e.g., water, buffered water, 0.8% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized and/or sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized. In some embodiments, the lyophilized preparation is combined with a sterile solution prior to administration.

In some embodiments, the compositions, e.g., pharmaceutical compositions, may contain pharmaceutically acceptable auxiliary substances to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc. In some embodiments, the pharmaceutical composition comprises a preservative. In some other embodiments, the pharmaceutical composition does not comprise a preservative.

Method of Use and Treatment

Without being bound by theory, the nucleic acids and other embodiments described herein are used in a method of conditionally expressing a molecule (e.g., protein) of interest, said method comprising: contacting an expression system, e.g. a cell comprising the nucleic acid molecule described herein, a vector described herein or a recombinant virus described herein, with a splice modulator, e.g., LMI070, wherein: a) in the presence of said splice modulator, expression of said protein of interest is increased, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, or 100 fold greater, relative to the level of expression of said protein of interest in the absence of said splice modulator; and b) in the absence of said splice modulator, expression of said protein of interest is substantially decreased, e.g., e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold less, relative to the level of expression of said protein of interest in the presence of the splice modulator.

In embodiments the nucleic acids and other embodiments described herein are used in a method of conditionally expressing a protein of interest, said method comprising: contacting an expression system, e.g. a cell comprising the nucleic acid molecule described herein, a vector described herein or a recombinant virus described herein, with a splice modulator, e.g., LMI070, wherein: a) in the absence of said splice modulator, expression of said protein of interest is increased, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, relative to the level of expression of said protein of interest in the presence of said splice modulator; and b) in the presence of said splice modulator, expression of said protein of interest is substantially decreased, e.g., e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold less, relative to the level of expression of said protein of interest in the absence of the splice modulator.

In embodiments, provided is a method of treating a subject in need of a gene therapy, said method comprising administering to said subject a nucleic acid molecule described herein, a vector described herein a recombinant virus described herein, or a pharmaceutical composition described herein. In embodiments, the method further comprises administering to the subject a splice modulator. In

embodiments, the splice modulator is administered periodically (e.g., for a time, separated by times of no administration). In embodiments, the method further comprises administering to the subject an amount of a splice modulator, e.g., LMI070, effective to cause at least a 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold increase or decrease in expression of the protein of interest, relative to the expression level of the protein of interest in the absence of the splice modulator.

Without being bound by theory, mutations in the gene encoding neuron-specific proteins such as progranulin may be implicated in neurodegenerative diseases. In some embodiments, the nucleic acids, vectors, and viruses disclosed herein may be administered to increase neuron-specific expression of a wild-type gene whose loss has been implicated in a neurodegenerative disease. For instance, administration may be used to increase levels of functional progranulin polypeptides. Co-treatment with a splice modulator allows for the expression level to be controlled (modulated). In some embodiments, administering a nucleic acid, vector, and/or virus disclosed herein may serve to treat, prevent, delay, slow, a disease, such as for example, frontotemporal dementia. In some embodiments, the nucleic acids, vectors, and viruses disclosed herein are used in conjunction with a splice modulator to modulate expression of the transgene.

As used herein, “frontotemporal dementia” (FTD) is an umbrella term for a diverse group of disorders that primarily affect the frontal and temporal lobes of the brain — the areas generally associated with personality, behavior and language. FTD is typically driven by degeneration of the frontotemporal lobar regions of the brain. In frontotemporal dementia, portions of these lobes shrink (atrophy). Signs and symptoms may vary, depending upon the portion of the brain affected. The most common signs and symptoms of frontotemporal dementia involve extreme changes in behavior and personality. These include increasingly inappropriate actions, loss of empathy and other interpersonal skills, lack of judgment and inhibition, apathy, repetitive compulsive behavior, a decline in personal hygiene, changes in eating habits, predominantly overeating, oral exploration and consumption of inedible objects, and a lack of awareness of thinking or behavioral changes. Rarer subtypes of FTD are characterized by problems with movement, similar to those associated with Parkinson's disease or amyotrophic lateral sclerosis. Since the discovery that several gene mutations can cause both FTD and amyotrophic lateral sclerosis (ALS), it is increasingly being recognized that FTD and ALS share neurodegenerative pathways and may be part of a common spectrum. Mutations in the progranulin gene (GRN) have recently been identified as a major cause of FTD, with majority of the mutations leading to loss of functional hPGRN polypeptide. Babykumari et al., (2017) *Brain*, 140(12): 3081-3104; Baker et al., (2006) *Nature*, 442: 916-19; Cruts et al., (2006) *Nature*, 442: 920-4 ; Gaweda-Walerych et al, (2018) *Neurobiol. Aging*, 72:186.e9-186.e12; Galimerti et al., (2018) *Expert Opin. Ther. Targets*, 22(7):579-585; Wauters et al., (2018) *Neurobiol. Aging*, 67:84-94; and Mendez, (2018) *Neuropsychiatr. Dis. Treat.*, 26:14:657-662. Methods for detecting mutations in PGRN include, e.g., those disclosed in WO2008/019187.

In various embodiments, the nucleic acids, vectors, and/or viruses disclosed herein may be used in methods of treating a disorder caused by one or more mutations in the gene encoding progranulin. In one embodiment, the term "treating" comprises the step of administering an effective dose, or effective multiple doses, of a composition comprising a nucleic acid, a vector, a recombinant virus, or a pharmaceutical composition as disclosed herein, to an animal (including a human being) in need thereof. If the dose is administered prior to development of a disorder/disease, the administration is prophylactic. If the dose is administered after the development of a disorder/disease, the administration is therapeutic. In embodiments, an effective dose is a dose that detectably alleviates (either eliminates or reduces) at least one symptom associated with the disorder/disease state being treated, that slows or prevents progression to a disorder/disease state, that slows or prevents progression of a disorder/disease state, that diminishes the extent of disease, that results in remission (partial or total) of disease, and/or that prolongs survival. The term encompasses but does not require complete treatment (i.e., curing) and/or prevention. In some embodiments, an effective dose comprises 1×10^{10} to 1×10^{15} vector genome per milliliter (vg/ml) of a virus as disclosed herein. In some embodiments, an effective dose comprises 1×10^6 to 1×10^{10} plaque forming units per milliliter (pfu/ml) of a virus as disclosed herein. In some embodiments, an effective dose comprises 1×10^6 to 1×10^9 transducing units per milliliter (TU/ml) of a virus as disclosed herein. Examples of disease states contemplated for treatment are set out herein.

In some embodiments, the mutations in the gene encoding progranulin are deletion mutations. In some embodiments, the mutations in the gene encoding progranulin are null mutations. In some embodiments, the mutations in the gene encoding progranulin are indels. In some embodiments, the mutations in the gene encoding progranulin are loss-of-function mutations. In some embodiments, the mutations in the gene encoding progranulin are knock-out mutations. In some embodiments, the mutations in the gene encoding progranulin results in loss of expression and/or function of the progranulin protein. In some embodiments, a patient in need of treatment with the nucleic acids, vectors, and/or viruses disclosed herein is identified by screening for a progranulin mutation prior to administration. In some embodiments, screening comprises obtaining a sample of cells or tissue from a subject and sequencing or genotyping one or more genetic loci in the sample to check for the presence of a progranulin mutation. In some embodiments, the screening is performed on genetic material from samples such as (but not limited to) saliva, blood, and/or skin cells.

In some embodiments, a method of treating comprises delivering to a subject in need thereof a therapeutically effective amount of a nucleic acid disclosed herein. In some embodiments, a method of treating comprises delivering to a subject in need thereof a therapeutically effective amount of a vector disclosed herein. In some embodiments, a method of treating comprises delivering to a subject in need thereof a therapeutically effective amount of a recombinant virus disclosed herein. In some embodiments, a method of treating comprises delivering to a subject in need thereof a therapeutically effective amount of a pharmaceutical composition disclosed herein. In some embodiments, the disorder is a neurodegenerative disorder. In some embodiments, the disorder is a frontotemporal dementia. In some

embodiments, the disorder is Alzheimer's disease. In some embodiments, the disorder is Parkinson's disease. In some embodiments, the disorder is amyotrophic lateral sclerosis (ALS).

In some embodiments, a nucleic acid, vector, recombinant virus, or pharmaceutical composition disclosed herein is used in treating a disorder caused by one or more mutations in the gene encoding progranulin, e.g., a mutation which results in loss of expression and/or function of the progranulin protein. In some embodiments, the disorder is a neurodegenerative disorder. In some embodiments, the disorder is a frontotemporal dementia. In some embodiments, the disorder is Alzheimer's disease. In some embodiments, the disorder is Parkinson's disease. In some embodiments, the disorder is amyotrophic lateral sclerosis (ALS).

In some embodiments, a nucleic acid, vector, recombinant virus, or pharmaceutical compositions disclosed herein is used in the manufacture of a medicament, for treating a subject in need thereof. In embodiments, the subject suffers from a disorder caused by one or more mutations in the gene encoding progranulin, e.g., a mutation which results in loss of expression and/or function of the progranulin protein.

In various embodiments, the nucleic acid, vector, recombinant virus, or pharmaceutical composition disclosed herein may be delivered to the subject in need thereof by an intravenous administration, direct brain administration (e.g., intrathecal, intracerebral, and/or intraventricular administration), intranasal administration, intra-aural administration, or intra-ocular route administration, or any combination thereof. In some embodiments, the nucleic acid, vector, recombinant virus, or pharmaceutical composition is delivered by intrathecal administration. In some embodiments, the nucleic acid, vector, recombinant virus, or pharmaceutical composition is delivered by an intracerebral or intraventricular route of administration. In some embodiments, the administered nucleic acid, vector, recombinant virus, or pharmaceutical composition is ultimately delivered to the brain, spinal cord, peripheral nervous system, and/or CNS, either directly or by transfer after administration to a separate tissue or fluid, e.g., blood.

Without being bound by theory, in some embodiments the methods disclosed herein may rescue cells that carry mutations on a gene coding for a polypeptide, e.g., progranulin, that result in a non-functioning polypeptide. In some embodiments, a method of expressing a molecule, for example a protein or ribonucleic acid (e.g., an siRNA), comprises delivering to a cell a nucleic acid, viral vector, virus, or pharmaceutical composition disclosed herein. In some embodiments, the cell is a neuronal cell. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is a human cell. In some embodiments, the neuronal cell is a neuron. In some embodiments, delivery is done in vitro. In some embodiments, delivery is done ex vivo. In some embodiments, the delivery is by systemic administration. In some embodiments, the delivery is local. In some embodiments, the delivery is by direct application to the target tissue. In some embodiments, the target tissue is the brain. In some embodiments, the delivery is by injection into the brain. In some embodiments, the delivery is by intrathecal administration. Without being bound by theory, the methods disclosed herein may reduce lipofuscin deposition, astrocyte and

microglia activation, and/or inflammation in the brain of a human or mouse with a mutation in the PGRN protein, thus providing potential benefits to subjects in need thereof.

In various embodiments, the nucleic acids, vectors, viruses, and pharmaceutical compositions disclosed herein may be used to treat a disorder, e.g., FTD. In some embodiments, a nucleic acid, vector, virus, and/or pharmaceutical composition disclosed herein may be used in the manufacture of a medicament for treating a disorder, e.g., FTD. In some embodiments, the disorder is caused by one or more mutations in the gene encoding progranulin. In some embodiments, the mutation in the progranulin gene results in a loss of expression of the progranulin protein. In some embodiments, the mutation in the progranulin gene results in loss of function of the progranulin protein. In some embodiments, the use comprises delivering to a subject in need thereof a therapeutically effective amount of a nucleic acid encoding hPGRN, e.g., in a vector, virus, and/or pharmaceutical composition disclosed herein.

Also provided herein is a kit comprising a nucleic acid molecule described herein, a vector described herein, a recombinant virus described herein, a cell described herein, or a pharmaceutical composition described herein; and a splice modulator.

The present disclosure is further illustrated by the following examples that should not be construed as limiting. The contents of all references, patents, and published patent applications cited throughout this application, as well as the Figures, are incorporated herein by reference in their entirety for all purposes.

EXAMPLES

The following examples are to be considered illustrative and not limiting on the scope of the disclosure disclosed above.

Example 1. Identification of splice-modulator binding sites from the human genome.

A normal Human Fibroblast line (HD1994) was treated with active (LMI070 and splice modulator 2) and inactive analogs (splice modulator 3) or DMSO for 24 hours. The following compound doses were used in both the NSC34 and HD1994 cell line 1:

- LMI070 was tested at a concentration of (100nM) and a high dose of (5uM)
- Splice modulator 2 was tested at 750nM
- Splice modulator 3 was tested at 5uM

DMSO treatment control was included for both cell lines. There were 3 biological replicates per group.

Total RNA was isolated using the Qiagen RNeasy Mini isolation kit. RNA-Seq libraries were prepared using the Illumina TruSeq RNA Sample Prep kit v2 and sequenced using the Illumina HiSeq2500 platform.

Each sample was sequenced on four different lanes belonging to the same flow cells to a length of 2x76 base-pairs (bp). The quality of the generated reads was assessed by running FastQC (version 0.10.1) on the FASTQ files provided by the sequencing lab (data release file for DM00012.txt). The average quality per base in Phred score was computed for each sample. The reads are of excellent quality (mean Phred score > 28 for all base positions). A similar quality trend that decreases to the 5' and 3' ends, was observed, as expected by Illumina chemistry.

A total of 847 million 76-base-pair (bp) paired-end reads were mapped to the Homo sapiens genome (hg19), the human RefSeq (Pruitt et al., 2007) transcripts (release 59, May 3, 2013) using TopHat (2.0.3)

TopHat (2.0.3) alignments against the human genome (hg19) were computed for each of the 15 replicates separately. In order to increase the ability to detect exons, three alignment files (bam files) were pooled for each of the five conditions (DMSO, LMI070 at 5uM, LMI070 at 100nM, splice modulator 2 at 750nM and splice modulator 3 at 5uM) before the transcript assembly by Cufflinks (2.1.1). After transcript assembly, the exon coordinates were extracted from the transcript gtf files. Exons on alternative chromosomes and on chromosome M were excluded and the strand information was ignored. That yielded 273866 putative exons. Exons that do not intersect any RefSeq exon (release 59, May 3, 2013) are considered as candidates for non annotated splice in events. That results in 19474 candidates. To gain further confidence, overlapping exons were merged in the full set of all RefSeq exons plus the initial 19474 candidates resulting in 229665 non overlapping exons. For this set of exons all possible exon-exon junctions within each RefSeq gene were considered. A junction database was created using R (2.15.2) scripts and bedtools (2.15.0). The first mate of each paired end read was then mapped against the database. Only non annotated exons supported by at least one junction alignment were retained. This excludes in particular candidates not attached to a RefSeq gene. That leaves 10898 final candidates. Sequences for these candidates were extracted from hg19 using bedtools. To assess variability separate Cufflinks assemblies were computed for each replicate and checked whether each candidate is seen in such an assembly. In addition the alignments against the junction database was used to determine the number of junctions that skip over a new exon. The information was used to estimate a splice-in fraction. Further the read coverage for the 10898 candidates was determined for each replicate using bedtools on the TopHat alignments (bam files) and then aggregated within each of the five conditions. The original fastq files were reprocessed with STAR (020201) and aligned against the human genome (hg38). The 10898 candidate exons were lifted over to hg38 using the UCSC genome browser tools – 7 candidates could not be lifted. The junctions detected by STAR were mapped to the remaining 10891 candidates and provide an alternative source of junction counts. The final 10 candidates (described in Table 1) for validation were selected from the 10898 putative not annotated exons found in the human SMA RNA Seq dataset as follows 1) STAR aggregated junction counts & TopHat exon coverage =0 for all samples for splice modulator 3 and DMSO (no leakiness); 2) LMI070 and splice modulator 2 STAR aggregated junction counts & TopHat exon coverage >60 (dynamic

range); 3) Exon length <100 bp (for feasibility); and 3' end AGA/GTAAG (to confirm presence of splice modulator binding site)

Example 2. Construction of minigene switch.

With the design concepts in mind (Fig. 1A and 1B), a specific minigene ON-switch was designed using the SNX7 gene sequences identified in Example 1. Fig. 2A shows a schematic diagram of the SNX7 locus identified in Example 1 containing a splice modulator (LMI070) exonic target binding site at chromosome: GRCh37:1:99204216:99204359:1

(AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGAATGAGCAGAAAATCATTTCAGGGCCTGTTCTC TATGTCCTTGCTATCCCTGTCTTCTGTAGCTATTCTGAAACCATCAACAAAGGAGCACACCATTCCAT CAGCAAAAGA (SEQ ID NO: 80)), as well as an intronic sequence downstream of exon 8 at chromosome:GRCh37:1:99203793:99203946:1

(CTTCCAGAGGAGATTGGAAAACCTGAAGATAAAGTGGAATGTGCTAATAATGCCCTGAAAGCAGATT GGGAGAGATGGAAACAAAATATGCAAAATGATATCAAGTTAGCATTACAGATATGGCTGAGGAGAA TATCCATTATTATGAACAG (SEQ ID NO: 99)), and 21,251 nucleotides upstream of exon 9 at chromosome:GRCh37:1:99225610:99225687:1

(TGCCTTGCTACGTGGGAGTCATTCCTTACATCACAGACCAACCTTCACTTGGAAGAAGCCTCTGAAG ATAAACCTTAA (SEQ ID NO: 100)). Using these sequence the non-naturally occurring SNX7 minigene was constructed (version 1) (Fig. 2) using exon 8 (called exon A), the 270 nucleotide intron located between exon 8 and the identified cryptic exon comprising the splice modulator binding site (AB), an exon comprising a splice modulator (e.g., LMI070) binding site at its 3' end (called exon B), and a 407 nucleotide intron fragment between the cryptic exon and exon 9 (shortened from 21,251 nt; BC), and exon 9 (called exon C). Additional modifications were made to the minigene to improve its performance, such as: 1) a Kozak consensus sequence and ATG codon (GCCACCATG) was inserted at position 65 in exon A; 2) All other ATG sequences in the minigene were replaced with TTG; 3) a TA at position 20 of exon A was replaced with AG to make GAAGAAGAA sequence (SEQ ID NO: 69); 4) 1 nt was removed from exon B to create frame shift (number of nucleotides = 3n-1) in ORF; 5) T was inserted at position 4 of exon C to create frame shift in ORF resulting in multiple stop codons; 6) TAC at position 9 of exon C was changed to TAA to create earlier termination codon; 7) CAG at position 34 of exon C was changed to ACC to mutate a potential cryptic splice site; 8) CTCT at position 60 of exon C was changed to TAGC to create a Nhe I restriction site; and 9) TAA at the end of exon C was removed to create continuous ORF. This sequence was then inserted into a scAAV vector using molecular cloning techniques. The scAAV was created by combining, AAV2 ITR containing a deletion of trs, followed by a JeT promoter, followed by the SNX7 minigene (see Fig. 2B), followed by a coding sequence for a furin cleavage site (RNRR (SEQ ID NO: 39)) added to the end of exon C, followed by coding sequence for a T2A peptide, followed by a transgene sequence (here, a coding sequence for EGFP without the first ATG); followed by a SV40 late

polyadenylation signal, followed by an AAV2 ITR (See Fig. 2C). Figure 3 shows the predicted mRNA products of the scAAV in the presence or absence of splice modulator.

Example 3. In vitro performance of ON-switch in HEK293 cells.

HEK293 cells were maintained in complete DMEM media and seeded in 24-well plate at 100,000 cells per well dense day before transfection. Each well was transfected with 2ug of pJSNX-GFP plasmid DNA using Lipofectamine2000 (Invitrogen) according to manufacturer's protocol. Transfection media was replaced with complete DMEM 4 hours later. Initial 1mM stock of LMI070 in DMSO was diluted in 1/1,000-1/500,000 in DMEM to achieve concentrations 2nM-1uM when added to the the cells in 24 hours later. Control cells received 0.1%DMSO. GFP expression was evaluated 48 hours post transfection using fluorescent microscope. No GFP expression was observed in control DMSO-treated cells (Fig. 4A). For quantitative analysis of GFP expression, cells were trypsinized and analyzed by FACS using SONY SH-800 flow cytometer. Mean fluorescence intensity was used for relative measurement of GFP expression. Control DMSO-treated cells showed no detectable GFP expression, while dose-dependent increase in GFP expression was observed in LMI070-treated samples (Fig. 4B). For RNA splicing analysis, total RNA was extracted from cells using Trizol (Invitrogen) according to manufacturer's protocol. cDNA was synthesized using Superscript III 1st strand supermix for qRT-PCR (Invitrogen). Inclusion of exon B was evaluated using qPCR by measuring amounts of exonB-exonC amplified by CAACAAAGGAGCACACCATTC (SEQ ID NO: 103) and GCGGTTGCGAGGTTTATCT (SEQ ID NO: 104) primers pair as compared to total transgene mRNA amplified by primers specific to exon C, GCGGTTGCGAGGTTTATCT (SEQ ID NO: 104) and CTCTTGCTAAGTGGGAGTCATT (SEQ ID NO: 105). Inclusion of exon B in 125nM LMI070-treated cells was found to be upregulated 75 times as compared to DMSO-treated cells (Fig. 4C). Amounts of constitutively spliced RNA (i.e. exonA-exonC) was measured using GTGCTAATAATGCCCTGAAAGC (SEQ ID NO: 106) and CCACTTAGCAAGAGCACTGT (SEQ ID NO: 107) primers pair. 1uM LMI070-treated cells demonstrated 60 times lower exonA-exonC splicing as compared to control DMSO-treated cells.

Example 4. Regulation of GFP expression by SNX7 switch in rat cortical neurons.

Primary rat neurons were prepared from dissected rat embryo cortices digested with papain and cultured in complete Neurobasal media in 24-well poly-D-Lysine plates (Corning) at density 150,000 cells/well for 7 days. Half of the media was replaced with fresh media day before transfection. Each well was transfected with 2ug of pJSNX-GFP plasmid DNA using Lipofectamine2000 (Invitrogen) according to manufacturer's protocol, except cells were washed three times with optiMEM before adding DNA-liposomes cocktails. Transfection media was replaced with conditioned media containing 50% fresh complete Neurobasal media 4 hours later. Next day, 1mM stock of LMI070 in DMSO was diluted in 1/1,000-1/500,000 in DMEM to achieve concentrations 2nM-1uM when added to the cells. Control cells received 0.1% DMSO. GFP expression was evaluated 6 days post transfection using fluorescent microscope. No GFP expression was observed in control DMSO-treated cells (Fig. 4A). For RNA splicing

analysis, total RNA was extracted from cells using Trizol (Invitrogen) according to manufacturer's protocol. cDNA was synthesized using Superscript III 1st strand supermix for qRT-PCR (Invitrogen). Inclusion of exon B was evaluated using qPCR by measuring amounts of exonB-exonC amplified by CAACAAAGGAGCACACCATTC (SEQ ID NO: 103) and GCGGTTGCGAGGTTTATCT (SEQ ID NO: 104) primers pair as compared to total transgene mRNA amplified by primers specific to exon C, GCGGTTGCGAGGTTTATCT (SEQ ID NO: 104) and CTCTTGCTAAGTGGGAGTCATT (SEQ ID NO: 105). Inclusion of exon B in 31nM LMI070-treated cells was found to be upregulated more than 100 times as compared to DMSO-treated cells (Fig. 4B). Amounts of constitutively spliced RNA (i.e. exonA-exonC) was measured using GTGCTAATAATGCCCTGAAAGC (SEQ ID NO: 106) and CCACTTAGCAAGAGCACTGT (SEQ ID NO: 107) primers pair. 500nM LMI070-treated cells demonstrated 30 times lower exonA-exonC splicing as compared to control DMSO-treated cells.

Example 5. Regulatable expression of human progranulin in rat cortical neurons by SNX7 switch.

Primary rat neurons were prepared from dissected rat embryo cortices digested with papain and cultured in complete Neurobasal media in 24-well poly-D-Lysine plates (Corning) at density 150,000 cells/well for 7 days. Half of the media was replaced with fresh media day before transfection. Each well was transfected with 2ug of pSyn-snx7-PGRN (Fig. 6A) or control pSyn-PGRN plasmids, which do not contain snx7 minigene, using Lipofectamine2000 (Invitrogen) according to manufacturer's protocol, except cells were washed three times with optiMEM before adding DNA-liposomes cocktails. Transfection media was replaced with conditioned media containing 50% fresh complete Neurobasal media 4 hours later. Next day, 1mM stock of LMI070 in DMSO was diluted in 1/10,000 in DMEM to achieve concentrations 100nm when added to the cells. Control cells received 0.01% DMSO. hPGRN expression was measured 6 days post transfection using TR-FRET assay. In pSyn-snx7-PGRN transfected cells, expression of hPGRN was induced by LMI070 more than 30 times comparing to DMSO-treated control (Fig. 6B). For RNA splicing analysis, total RNA was extracted from cells using Trizol (Invitrogen) according to manufacturer's protocol. cDNA was synthesized using Superscript III 1st strand supermix for qRT-PCR (Invitrogen). Inclusion of exon B was evaluated using qPCR by measuring amounts of exonB-exonC junction amplified by CAACAAAGGAGCACACCATTC (SEQ ID NO: 103) and GCGGTTGCGAGGTTTATCT (SEQ ID NO: 104) primers pair as compared to total transgene mRNA amplified by primers specific to exon C, GCGGTTGCGAGGTTTATCT (SEQ ID NO: 104) and CTCTTGCTAAGTGGGAGTCATT (SEQ ID NO: 105). Inclusion of exon B in LMI070-treated cells was found to be upregulated more than 150 times as compared to DMSO-treated cells for pSyn-snx7-PGRN transfected cells (Fig. 6C). Amounts of constitutively spliced RNA (i.e. exonA-exonC) was measured using GTGCTAATAATGCCCTGAAAGC (SEQ ID NO: 106) and CCACTTAGCAAGAGCACTGT (SEQ ID NO: 107) primers pair. LMI070-treated cells demonstrated 10 times lower exonA-exonC splicing in pSyn-snx7-PGRN transfected cells as compared to DMSO-treated cells.

Example 6. Modification of minigene to reduce size and eliminate peptide expression in the absence of LMI070

The SNX7 minigene was further modified to reduce the overall size and eliminate peptide expression in the absence of LMI070. In particular, exon A was shortened 109nt to 53 nt while the region adjacent to the 3' splice site was kept, the resulting exon A has the sequence of SEQ ID NO: 96. First intron was shortened from 150 nt to 120nt while splice sites and branch point were preserved. The resulting first intron has the sequence of SEQ ID NO: 97. The region containing the start codon in exon A of the first version of SNX7 minigene was deleted, and a start codon was constructed by changing TC to GG in exon B of the new version of SNX7 minigene. The resulting exon B has the sequence of SEQ ID NO: 98. By switching start codon to Exon B, protein expression occurs only in the presence of LMI070. The second intron was kept the same, as it was found that this sequence contains essential cis elements. The sequence of the modified SNX minigene (version 2) is shown in SEQ ID NO: 94. Fig. 9 shows schematic diagram of the new version (version 2) of minigene as compared to the previous version of SNX7 minigene (version 1).

The modified SNX7 minigene (version 2) was inserted into a scAAV vector using molecular cloning techniques. The sequence of the vector comprising the modified SNX7 minigene (version 2) is shown in SEQ ID NO: 95. HEK293 cells were maintained in complete DMEM media and seeded in 24-well plate at 100,000 cells per well density day before transfection. Each well was transfected with 2ug of plasmid DNA DL180 containing SNX7 switch version 1 or plasmid DL182 containing SNX7 version 2 using Lipofectamine2000 (Invitrogen) according to manufacturer's protocol. Transfection media was replaced with complete DMEM 4 hours later. Initial 1mM stock of LMI070 in DMSO was diluted in 1/1,000-1/500,000 in DMEM to achieve concentrations 100nm-1uM when added to the cells in 24 hours later. Control cells received 0.1%DMSO. GFP expression was evaluated 48 hours post transfection using fluorescent microscope. Fig.10 shows that the modified SNX7 minigene (version 2) is more sensitive than the previous version.

Example 7. Oral administration of LMI070 switches on transgene expression in mouse brain in time dependent manner

ssAAV9 viral vector encoding hPGRN under control of synapsin promoter with SNX7 switch (version 1) was produced in HEK293 cells and purified by iodixanol. 2e10vg of AAV vector in 2uL was injected ICV in C57Bl/6 neonatal mice at P0. At 4 weeks of age, 30 mg/kg of LMI070 or vehicle control was administered orally through gavage. 4-6 mice per group were taken down at specified time points (Fig. 7A). After transcardial perfusion with PBS posterior half of the left hemisphere was homogenized in Precellys tube. TR-FRET assay was used for measurement of human PGRN expressed from AAV vector. Results indicate rapid and transient induction of hPGRN expression in the brain after 24 hours post LMI070 administration. The transgenic protein expression returned to untreated levels after 4 days post LMI070 administration (Fig. 7B).

Example 8. LMI070 switch on transgene expression *in vivo* in dose dependent manner

ssAAV9 viral vector encoding hPGRN under control of synapsin promoter with SNX7 switch (version 1) was produced in HEK293 cells and purified by iodixanol. 2×10^{10} vg of AAV vector in 2 μ L was injected ICV in FVB neonatal mice at P0. At 4 weeks of age, 3, 10 or 30 mg/kg of LMI070 or vehicle control was administered orally through gavage. 6-7 mice per group were taken down at specified time points (Fig. 8A). After transcardial perfusion with PBS posterior half of the left hemisphere was homogenized in Precellys tube. TR-FRET assay was used for measurement of human PGRN expressed from AAV vector. Sample of human cortex was used as a control for physiological PGRN levels (~200pg/mg). Results indicate rapid (12 hours post LMI070 administration) accumulation of transgenic hPGRN in the brain, which starts to decline at 24 hour point. Transgene expression demonstrated dose response to LMI070 administration.

Throughout this specification and the claims that follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A nucleic acid molecule comprising a minigene linked to a transgene encoding a protein of interest, wherein the minigene comprises:

- i. A first exon;
- ii. A first intron;
- iii. A second exon;
- iv. A second intron; and
- v. A third exon;

wherein said second exon comprises a splice modulator binding sequence and wherein, in the presence of a splice modulator, said second exon is included in an mRNA product of the nucleic acid, and in the absence of said splice modulator, said second exon is not included in an mRNA product of the nucleic acid, wherein the third exon comprises a stop codon that is in frame in the mRNA product of the nucleic acid produced in the absence of the splice modulator and which is not in frame in the mRNA product of the nucleic acid produced in the presence of the splice modulator.

2. A nucleic acid molecule comprising a minigene linked to a transgene encoding a protein of interest, wherein the minigene comprises:

- i. A first exon;
- ii. A first intron;
- iii. A second exon;
- iv. A second intron; and
- v. A third exon;

wherein said second exon comprises a splice modulator binding sequence and wherein, in the presence of a splice modulator, said second exon is included in an mRNA product of the nucleic acid, and in the absence of said splice modulator, said second exon is not included in an mRNA product of the nucleic acid, wherein the second exon comprises a stop codon that is in frame in the mRNA product of the nucleic acid produced in the presence of the splice modulator.

3. A nucleic acid molecule comprising a minigene linked to a transgene encoding a protein of interest, wherein the minigene comprises:

- i. A first exon;
- ii. A first intron;
- iii. A second exon;
- iv. A second intron; and
- v. A third exon;

wherein said second exon comprises a splice modulator binding sequence and wherein, in the presence of a splice modulator, said second exon is included in an mRNA product of the nucleic acid, and in the absence of said splice modulator, said second exon is not included in an mRNA product of the nucleic acid, wherein the first and the third exons do not comprise a start codon, and wherein the second exon comprises a start codon.

4. The nucleic acid molecule of any one of claims 1-3, comprising a sequence encoding a protease cleavage site disposed between the minigene and the transgene; optionally wherein:

(a) said protease cleavage site is cleaved by a mammalian protease selected from group consisting of furin, PCSK1, PCSK5, PCSK6, PCSK7, cathepsin B, Granzyme B, Factor XA, Enterokinase, genenase, sortase, precision protease, thrombin, TEV protease, and elastase 1; and

(b) the protease cleavage site comprises a polypeptide having an cleavage motif selected from the group consisting of RX(K/R)R consensus motif, RXXX[KR]R consensus motif, RRX consensus motif, RNRR (SEQ ID NO: 39), I-E-P-D-X consensus motif (SEQ ID NO: 35), Glu/Asp-Gly-Arg, Asp-Asp-Asp-Asp-Lys (SEQ ID NO: 36), Pro-Gly-Ala-Ala-His-Tyr (SEQ ID NO: 37), LPXTG/A consensus motif, Leu-Glu-Val-Phe-Gln-Gly-Pro (SEQ ID NO: 38), Leu-Val-Pro-Arg-Gly-Ser (SEQ ID NO: 40), E-N-L-Y-F-Q-G (SEQ ID NO: 41), and [AGSV]-x (SEQ ID NO: 42).

5. The nucleic acid molecule of claim 4, wherein said cleavage site is cleaved by furin; optionally wherein the protease cleavage site cleaved by furin comprises RNRR (SEQ ID NO: 39) and the sequence encoding the protease cleave site comprises CGCAACCGCCGC (SEQ ID NO: 19).
6. The nucleic acid molecule of any one of claims 1-5, comprising a sequence encoding a self-cleaving peptide disposed between the minigene and the transgene, optionally wherein the self-cleaving peptide cleaves within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids of the N-terminus of the protein of interest, and optionally wherein the self-cleaving peptide comprises a T2A peptide comprising the

sequence of EGRGSLTCTGDVEENPGP (SEQ ID NO: 61) or (GSG)EGRGSLTCTGDVEENPGP (SEQ ID NO: 59).

7. The nucleic acid molecule of any one of claims 1-6, wherein the splice modulator binding sequence is located at the 3' terminus of the second exon.
8. The nucleic acid molecule of any one of claims 1-7, wherein the splice modulator binding sequence comprises AGA and the splice modulator is 5-(1H-Pyrazol-4-yl)-2-(6-((2,2,6,6-tetramethylpiperidin-4-yl)oxy)pyridazin-3-yl)phenol (LMI070).
9. The nucleic acid molecule of any one of claims 1-8, wherein the second exon comprises a sequence selected from:
 - i. CCTTGCTATCCCTGTCTTCTGTAGCTATTCTGAAACCATCAACAAAGGAGCACACCA
TTCCATCAGCAAAAGA (SEQ ID NO: 1);
 - ii. GTAATTAGCTGAGAAGGAAGATCTGAAGGTTTAACGAGAGAGGGCGAGAGATACAA
AATATCTGCTAGGAGA (SEQ ID NO: 2);
 - iii. GGATTGTTTGTATTCCTGCCAATGATTTGTGAGACAGTCTGTTCCCCACATCCTCGT
CAACAGA (SEQ ID NO: 3);
 - iv. CTTTCTGACATCTTAACGAGGCAATACAGAGAGACGAATTTTCATCAGTTTGTTCAG
GGAGACACATATAACAAAAGA (SEQ ID NO: 4);
 - v. ATCCATACATACTTAATGCTGAAATGTGAAGGGCTGAGAAAAAAGAAAAGA (SEQ ID
NO: 5);
 - vi. AATTGGAAACATCGAGGGAAAATGGGCTTTTTATTATTAACAAAACCTCAGTATTA
TCACTTAGAAACCTGAAATTGAACTCCAAAAGCCAAAGA (SEQ ID NO: 6);
 - vii. AAGAATGTTCTTTTGTGAAGAATGACTTAAGGAAGATTCATGATGACTGAGTGTGC
CCGTGTGGAACCTTTAGGACATAGATGCACTCCTACAGA (SEQ ID NO: 7);
 - viii. TTGTCCTTCACTCCGTACTCCAGTTGGCCAAGCATAGGTCGCATGCCAGGGTCAAG
GAGACTAAGGGAGA (SEQ ID NO: 8);
 - ix. GACATACAGACATGGCAGCCCCTAGCATGTGTATCCTAAGA (SEQ ID NO: 9);
 - x. ACATACAGACATGGCAGCCCCTAGCATGTGTATCCTAAGA (SEQ ID NO: 10);

- xii. A fragment or mutant of any of (i) to (xi) having at least 90%, at least 95% at least 96%, at least 97%, at least 98% or at least 99% identity thereto.
10. The nucleic acid molecule of any one of claims 1-9, wherein the second exon comprises a sequence derived from an exon of SNX7 or a cryptic exon of SNX7, and optionally wherein:
 - (a) the second exon comprises,
 - i. AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGAATGAGCAGAAAATCATTTTCAGG
GCCTGTTCTCTATGTCCTTGCTATCCCTGTCTTCTGTAGCTATTCTGAAACCATCAACA
AAAGGAGCACACCATTCCATCAGCAAAAGA (SEQ ID NO: 80) and
 - ii. a fragment of SEQ ID NO: 16; or
 - iii. a mutant sequence of SEQ ID NO: 16 or a fragment thereof having at least 90%, at least 95% at least 96%, at least 97%, at least 98% or at least 99% identity thereto; or
 - (b) the second exon comprises,
 - i. AGTTTGCAAAGGAAGGAAAGGAGCAGAGACTTGAATGAGCAGAAAATCATTTTCAGG
GCCTGTTCTCTATTGTCCTTGCTATCCTGTCTTCTGTAGCTATCTGAAACCATCAACA
AAGGAGCACACCATTCCATCAGCAAAAGA (SEQ ID NO: 16);
 - ii. a fragment of SEQ ID NO: 16; or
 - iii. a mutant sequence of SEQ ID NO: 16 or a fragment thereof having at least 90%, at least 95% at least 96%, at least 97%, at least 98% or at least 99% identity thereto;
11. The nucleic acid molecule of any one of claims 1-10, wherein the first exon comprises:
 - iii. Both (a) and (b).
12. The nucleic acid molecule of any one of claims 1-11, wherein the minigene has been modified to:

- i. Remove or mutate all but a single start codon, e.g., an ATG start codon, deposited within the first exon or the second exon;
 - ii. Remove or mutate all cryptic splice donor and splice acceptor sequences other than those at the termini of the first exon, the second exon and the third exon.
- 13. The nucleic acid molecule of any one of claims 1 and 3-12, wherein the minigene comprises, e.g., consists of, SEQ ID NO: 71 or SEQ ID NO: 94, or a sequence with at least 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identity thereto, or a functional fragment thereof; optionally wherein:
 - (a) the nucleic acid molecule further comprises a sequence encoding a furin cleavage site (e.g., SEQ ID NO: 19) and/or a sequence encoding a self-cleaving peptide (e.g. SEQ ID NO: 20), optionally wherein the minigene is disposed 5' to the sequence encoding the furin cleavage site (e.g., immediately 5' to the sequence encoding the furin cleavage site), the sequence encoding the furin cleavage site is disposed 5' to the sequence encoding the self-cleaving peptide (e.g., immediately 5' to the sequence encoding the self-cleaving peptide), and the sequence encoding the self-cleaving peptide is disposed 5' to the transgene (e.g., immediately 5' to the transgene);
 - (b) The nucleic acid molecule further comprises a promoter operably linked to the minigene and transgene, optionally wherein said promoter is disposed 5' to the minigene, and optionally wherein said promoter is a JeT promoter, a CBA promoter, a PGK promoter, or a synapsin promoter, or any promoter that does not comprise an intron;
 - (c) The nucleic acid molecule further comprises a post-transcriptional regulatory element comprising a PRE derived from hepatitis B (HPRE), bat (BPRE), ground squirrel (GSPRE), arctic squirrel (ASPRE), duck (DPRE), chimpanzee (CPRE) and wooly monkey (WMPRE) or woodchuck (WPRE), optionally wherein said post-transcriptional regulatory element is disposed 3' to the transgene;
 - (d) the nucleic acid molecule further comprises a polyadenylation signal (polyA), optionally wherein said polyA is disposed 3' to the transgene, and optionally wherein said poly A signal is an SV40 polyA, human growth hormone (HGH) polyA, or bovine growth hormone (BGH) polyA, a beta-globin polyA, an alpha-globin polyA, an ovalbumin polyA, a kappa-light chain polyA, and a synthetic polyA.
- 14. A vector comprising a nucleic acid according to any one of claims 1-13, wherein the viral vector is an adeno-associated viral (AAV) vector, chimeric AAV vector, adenoviral vector, retroviral vector, lentiviral vector, DNA viral vector, herpes simplex viral vector, baculoviral vector, or any mutant or derivative thereof.
- 15. A recombinant virus comprising the nucleic acid of any one of claims 1-13, or the vector of claim 14, optionally wherein the recombinant virus is an adeno-associated virus (AAV), chimeric AAV,

adenovirus, retrovirus, lentivirus, DNA virus, herpes simplex virus, baculovirus, or any mutant or derivative thereof.

16. A cell comprising the nucleic acid molecule of any one of claims 1-13, the vector of claim 14 or the recombinant virus of claim 15, optionally wherein the cell is a neuron or astrocyte.
17. A method of conditionally expressing a protein of interest, said method comprising: contacting an expression system (e.g. a cell, e.g., a cell of claim 16) comprising the nucleic acid molecule of any one of claims 1 and 3-13, the vector of claim 14 or the recombinant virus of claim 15, with a splice modulator, e.g., LMI070, wherein:
 - (a) in the presence of said splice modulator, expression of said protein of interest is increased, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, relative to the level of expression of said protein of interest in the absence of said splice modulator; and
 - (b) in the absence of said splice modulator, expression of said protein of interest is substantially decreased, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold less, relative to the level of expression of said protein of interest in the presence of the splice modulator.
18. A method of conditionally expressing a protein of interest, said method comprising: contacting an expression system (e.g. a cell, e.g., a cell of claim 16) comprising the nucleic acid molecule of any one of claims 2-13, the vector of claim 14 or the recombinant virus of claim 56, with a splice modulator, e.g., LMI070, wherein:
 - (a) in the absence of said splice modulator, expression of said protein of interest is increased, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold greater, relative to the level of expression of said protein of interest in the presence of said splice modulator; and
 - (b) in the presence of said splice modulator, expression of said protein of interest is substantially decreased, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold less, relative to the level of expression of said protein of interest in the absence of the splice modulator.
19. A method of treating a subject in need of a gene therapy, said method comprising
 - (a) administering to said subject the nucleic acid molecule of any one of claims 1-13, the vector of claim 14, the recombinant virus of claim 15, or the cell of claim 16; and
 - (b) administering to the subject an amount of a splice modulator, e.g., LMI070, effective to cause at least a 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold increase or decrease in expression of the

protein of interest, relative to the expression level of the protein of interest in the absence of the splice modulator.

20. Use of the nucleic acid molecule of any one of claims 1-13, the vector of claim 14, the recombinant virus of claim 15, or the cell of claim 16, in the manufacture of a medicament for treating a subject in need of a gene therapy, wherein treating the subject comprises:

- (a) administering to the subject the nucleic acid molecule, the vector, the recombinant virus, or the cell; and

- (b) administering to the subject an amount of a splice modulator, e.g., LMI070, effective to cause at least a 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50 or 100 fold increase or decrease in expression of the protein of interest, relative to the expression level of the protein of interest in the absence of the splice modulator.

Concepts of "ON switch" and "OFF-switch"

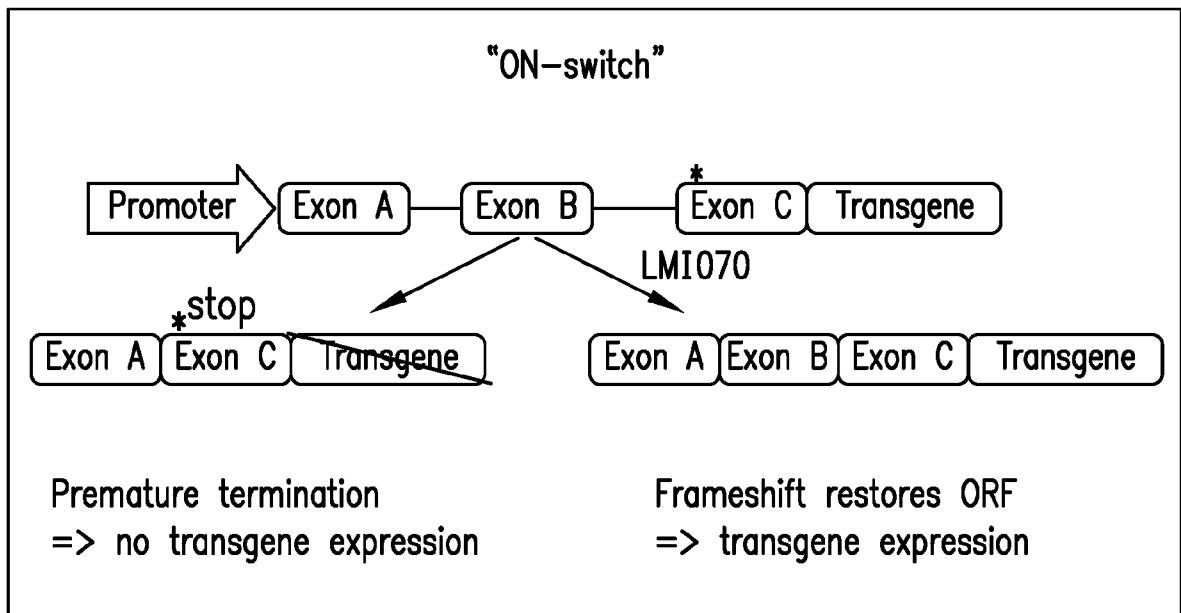


FIG.1A

Concepts of "ON switch" and "OFF-switch"

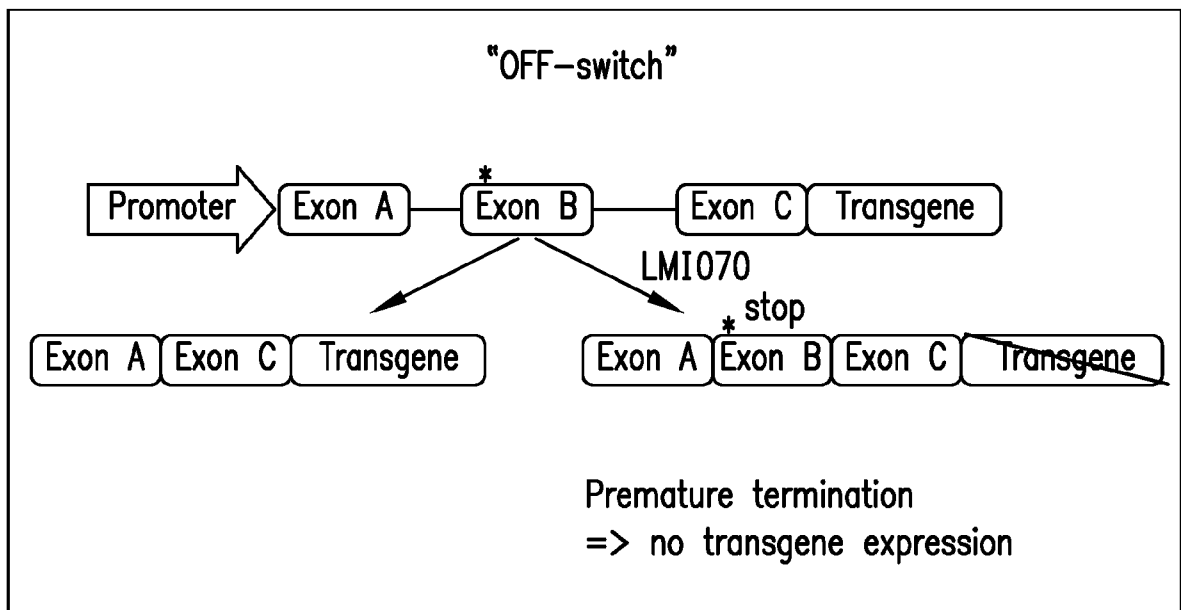


FIG.1B

Design of AAV vector with snx7 minigene-based switch (version 1)

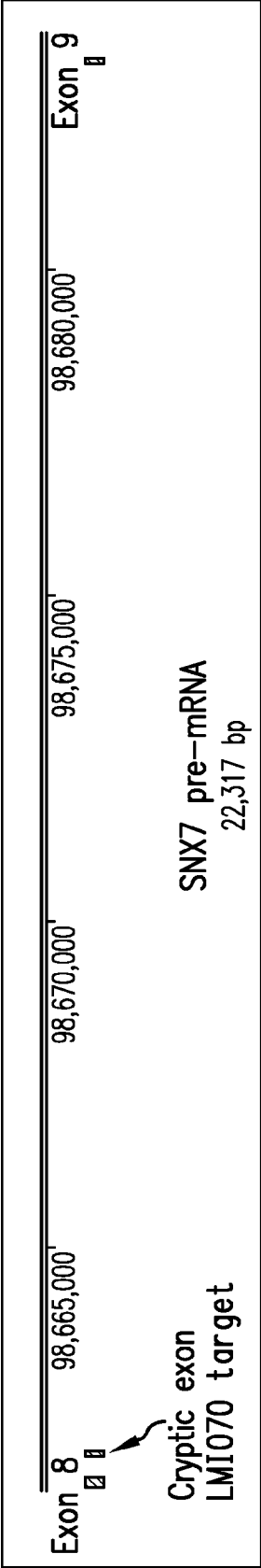


FIG. 2A

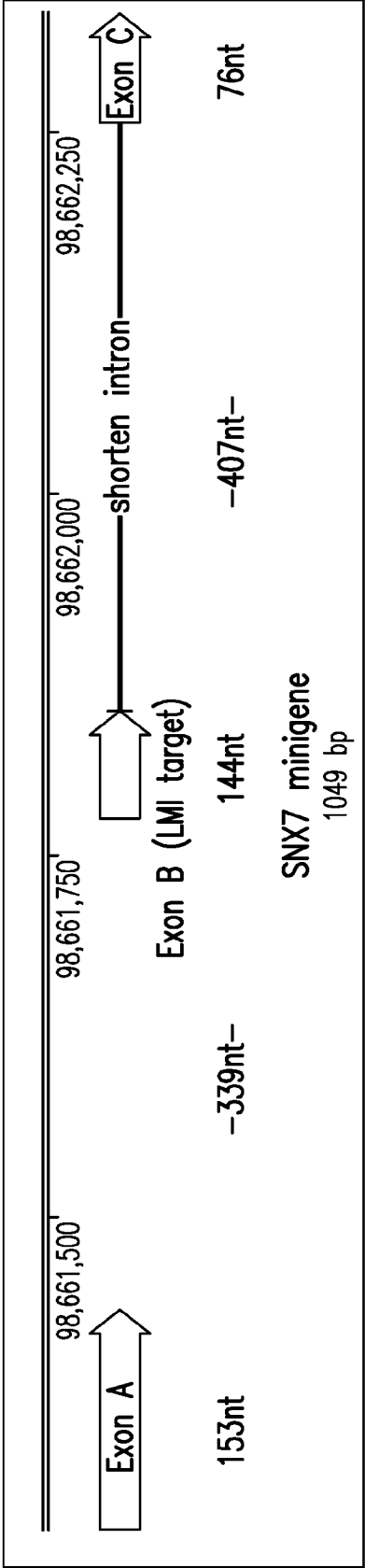


FIG. 2B

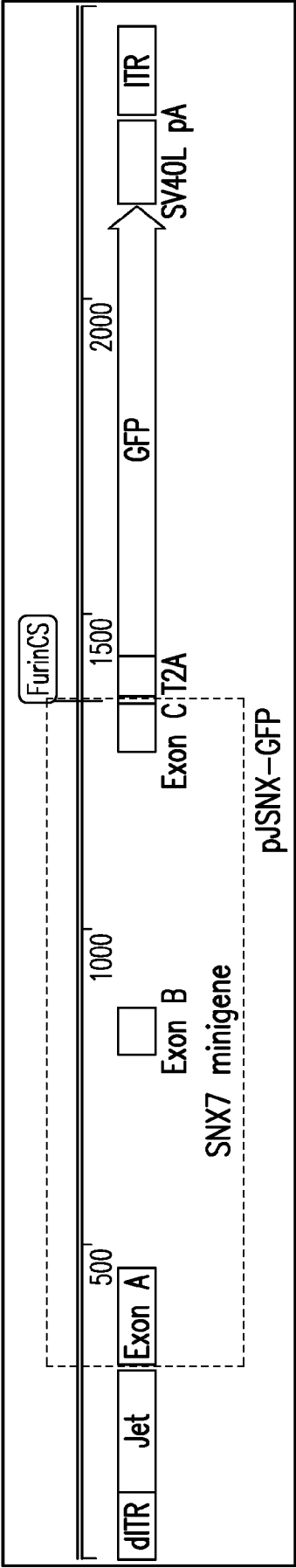


FIG.2C

Regulation of transgene expression by snx7 minigene-based ON-switch and OFF-switch

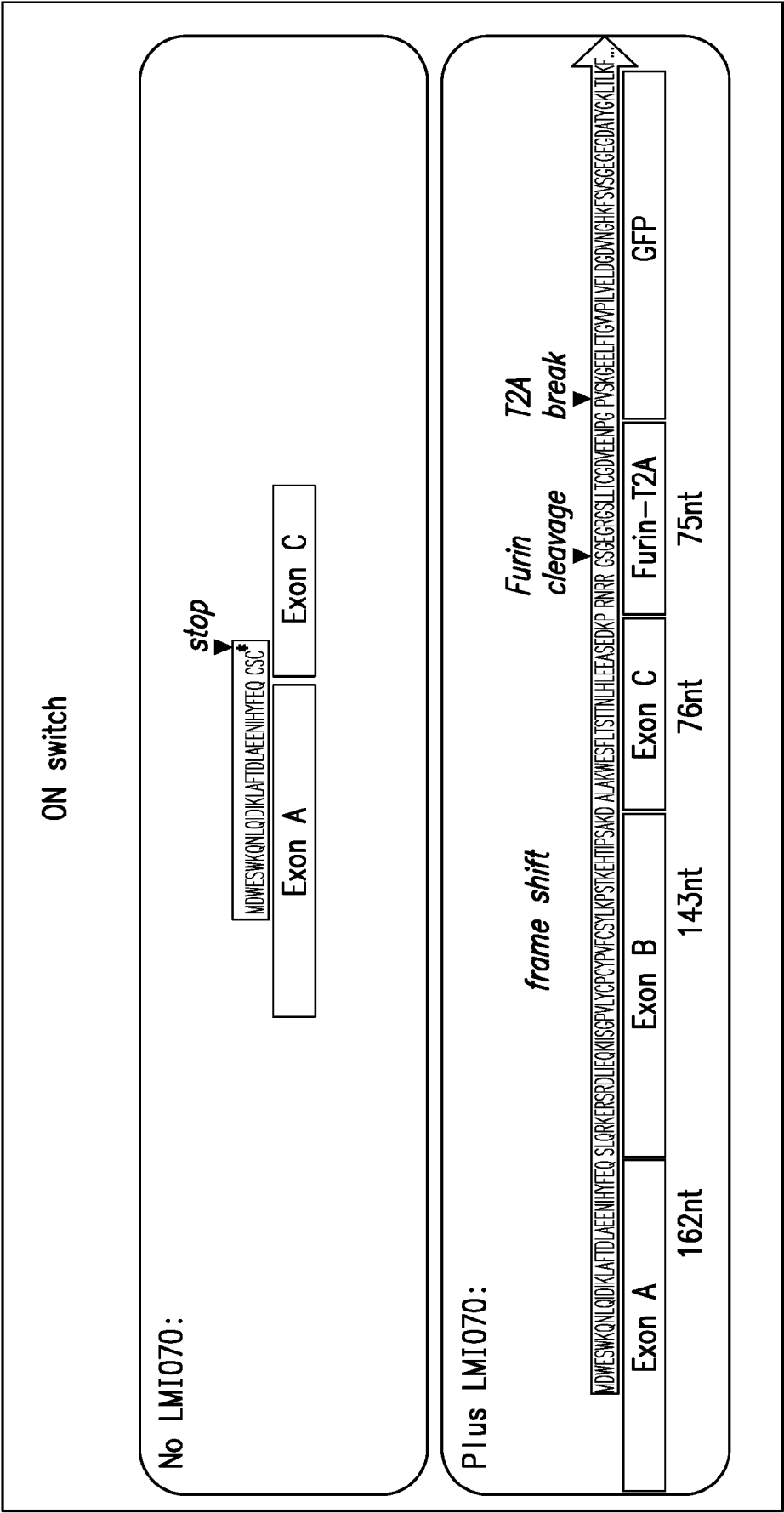


FIG.3A

Regulation of transgene expression by snx7 minigene-based ON-switch and OFF-switch

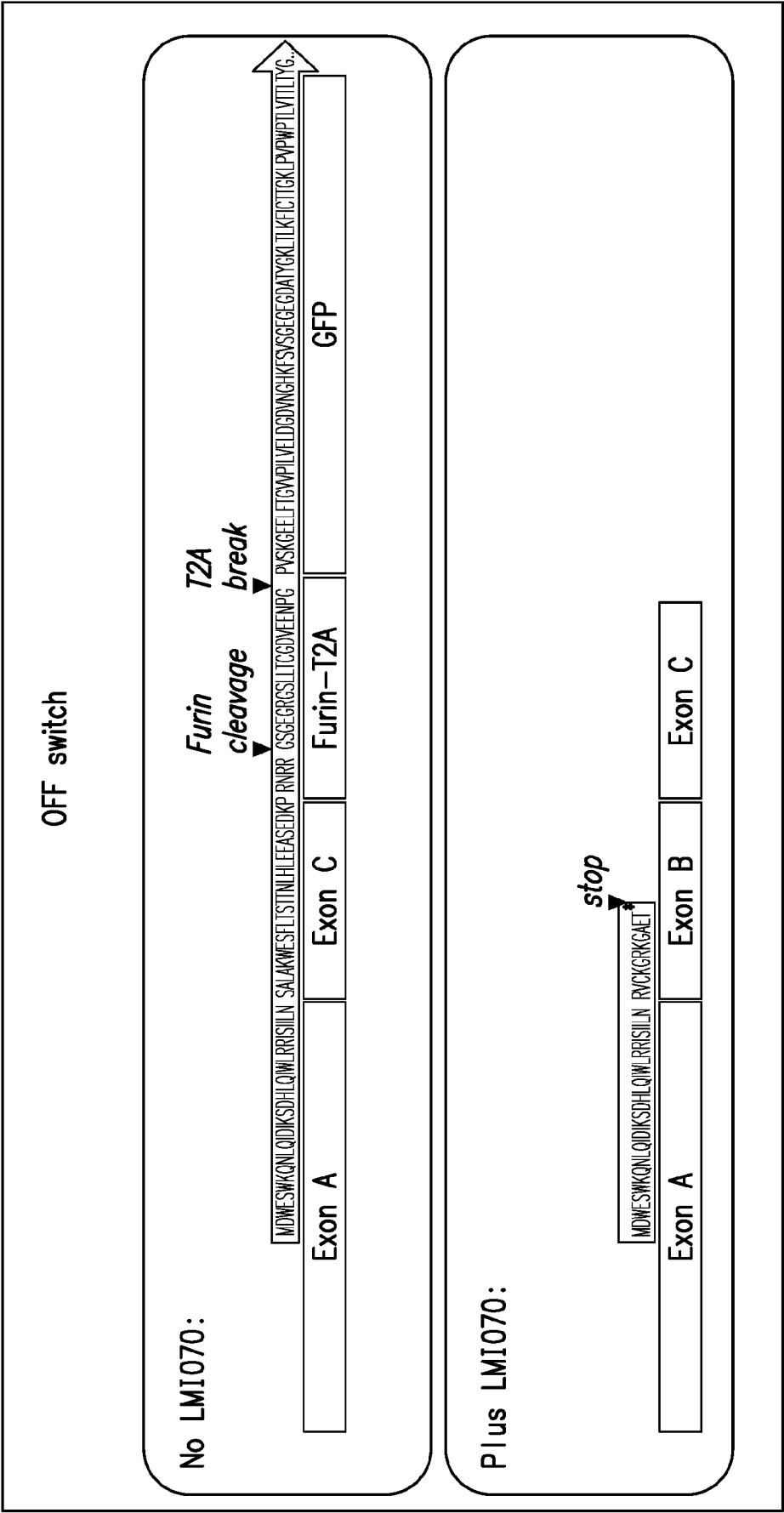


FIG.3B

In vitro testing in HEK293 cells shows GFP expression is regulated by snx7 switch (version 1)

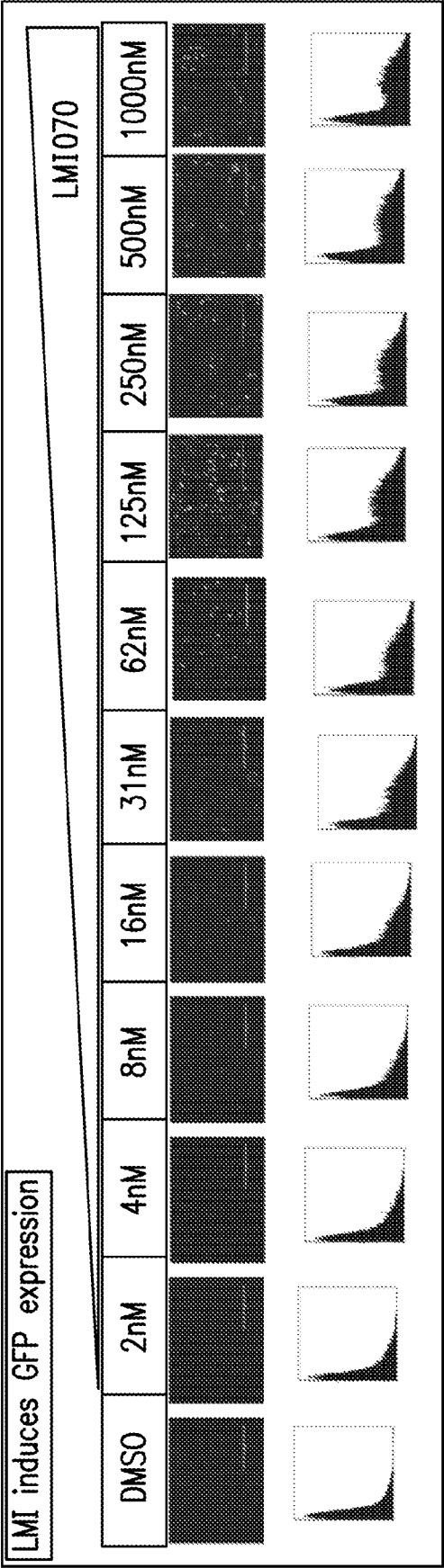


FIG.4A

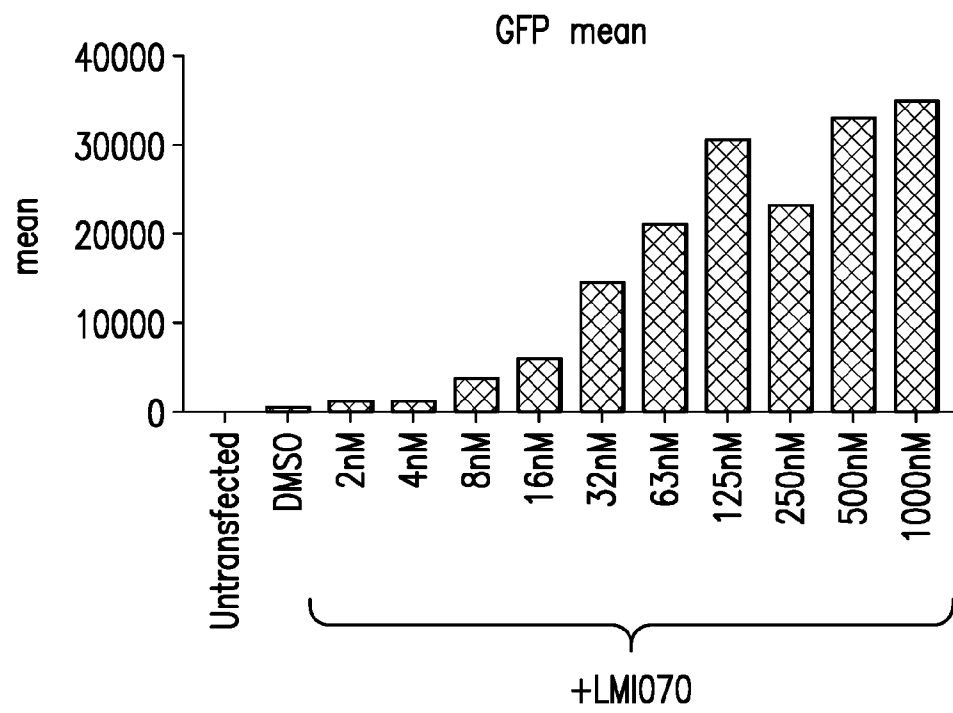


FIG.4B

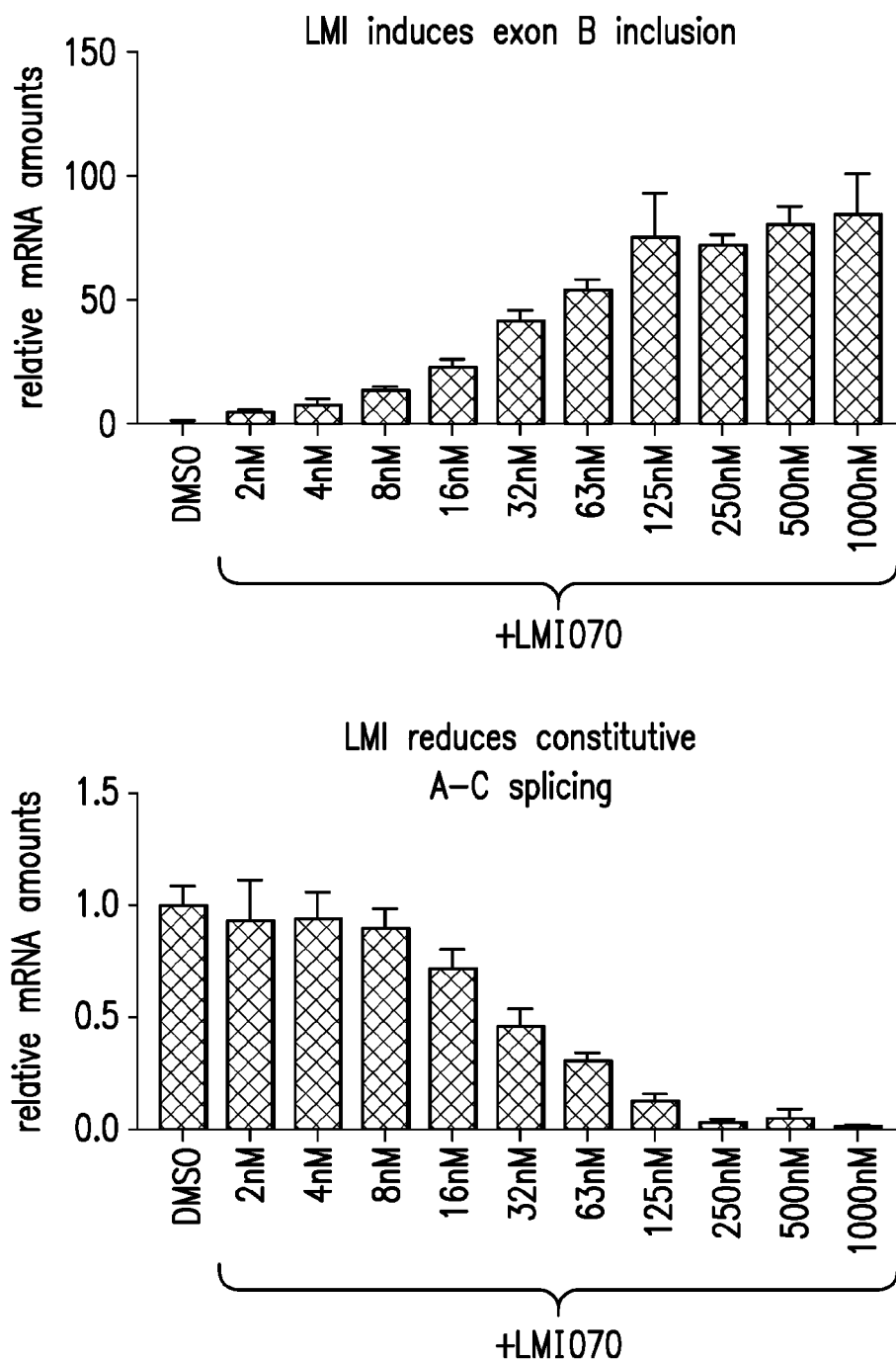


FIG.4C

In vitro testing in rat cortical neurons shows GFP expression is regulated by snx7 switch (version 1)

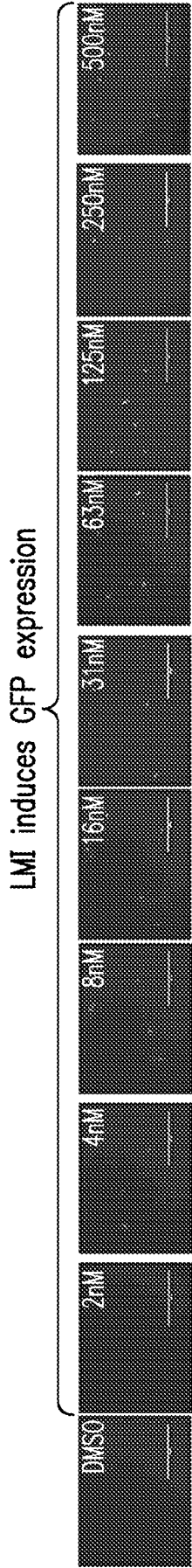


FIG.5A

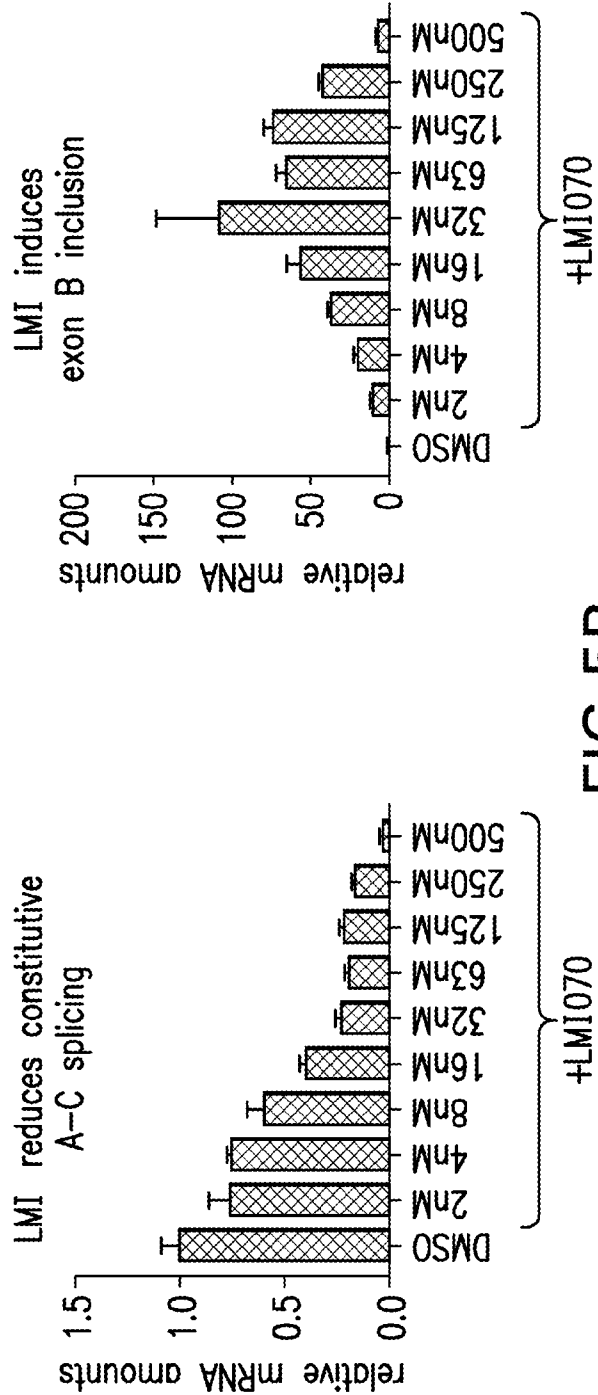


FIG.5B

Neuronal ssAAV vector with snx7 switch (version 1).
In vitro testing in rat cortical neurons shows hPGRN expression is regulated by snx7 switch

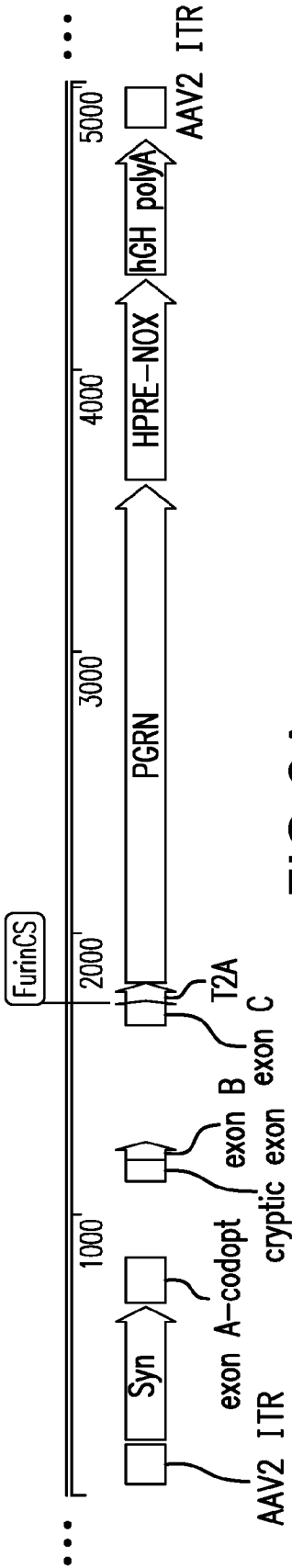


FIG. 6A

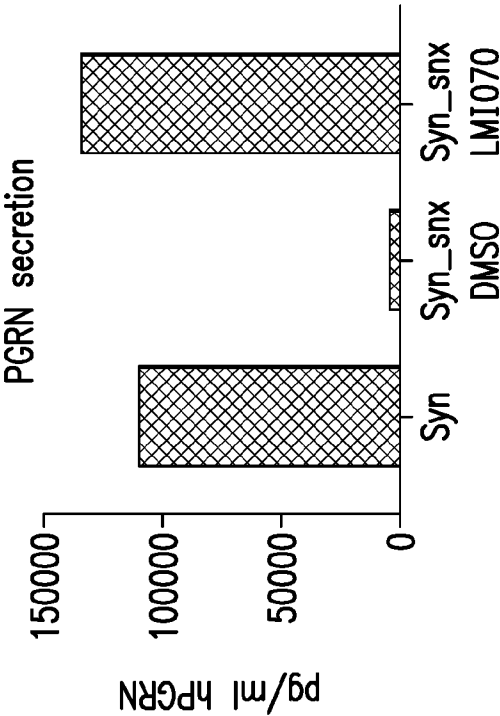


FIG. 6B

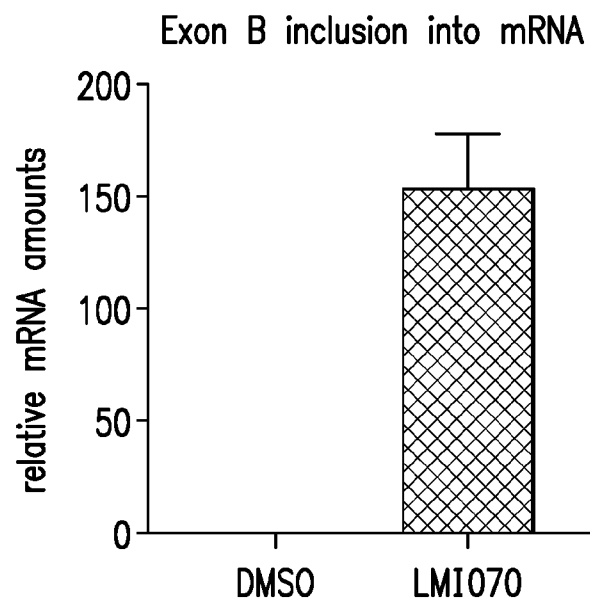
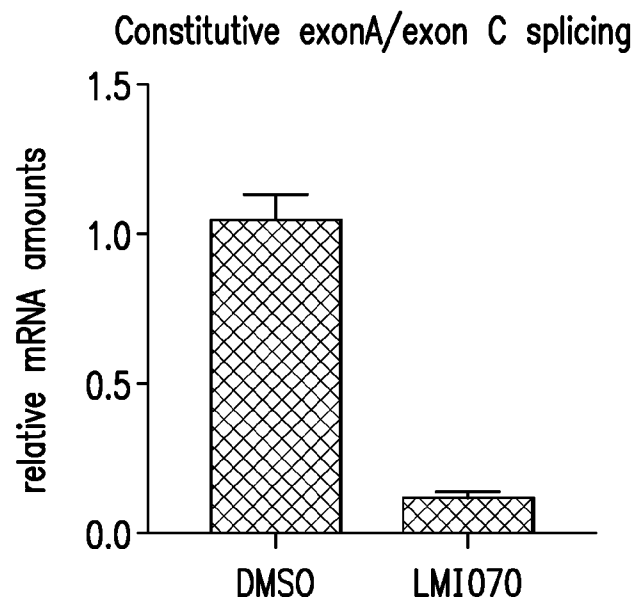


FIG.6C

Oral administration of LMI070 switches on transgene expression in mouse brain in time-dependent manner

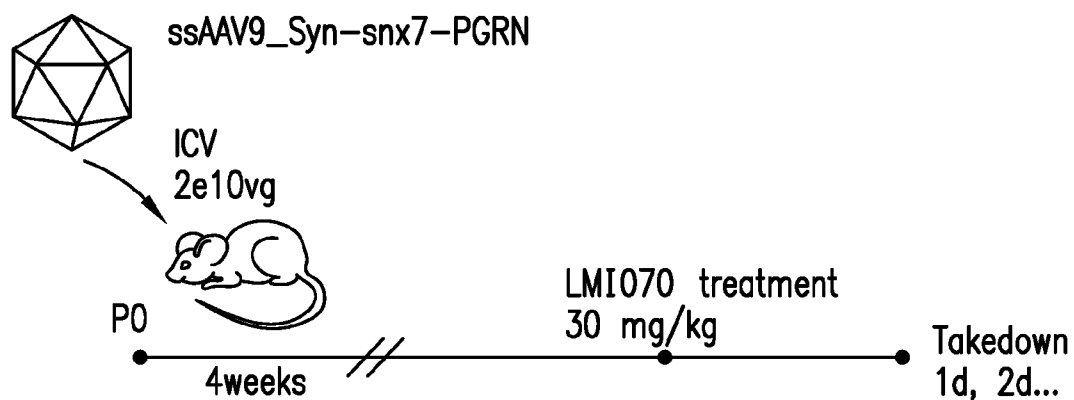


FIG.7A

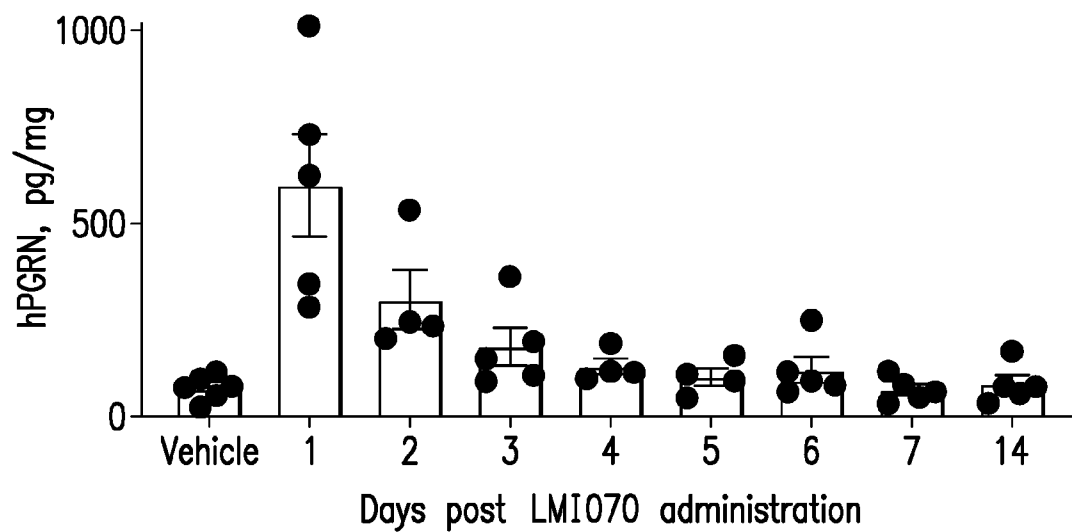


FIG.7B

LMI070 switch on transgene expression *in vivo*
in dose-dependent manner



FIG.8A

LMI070 switch on transgene expression *in vivo*
in dose-dependent manner

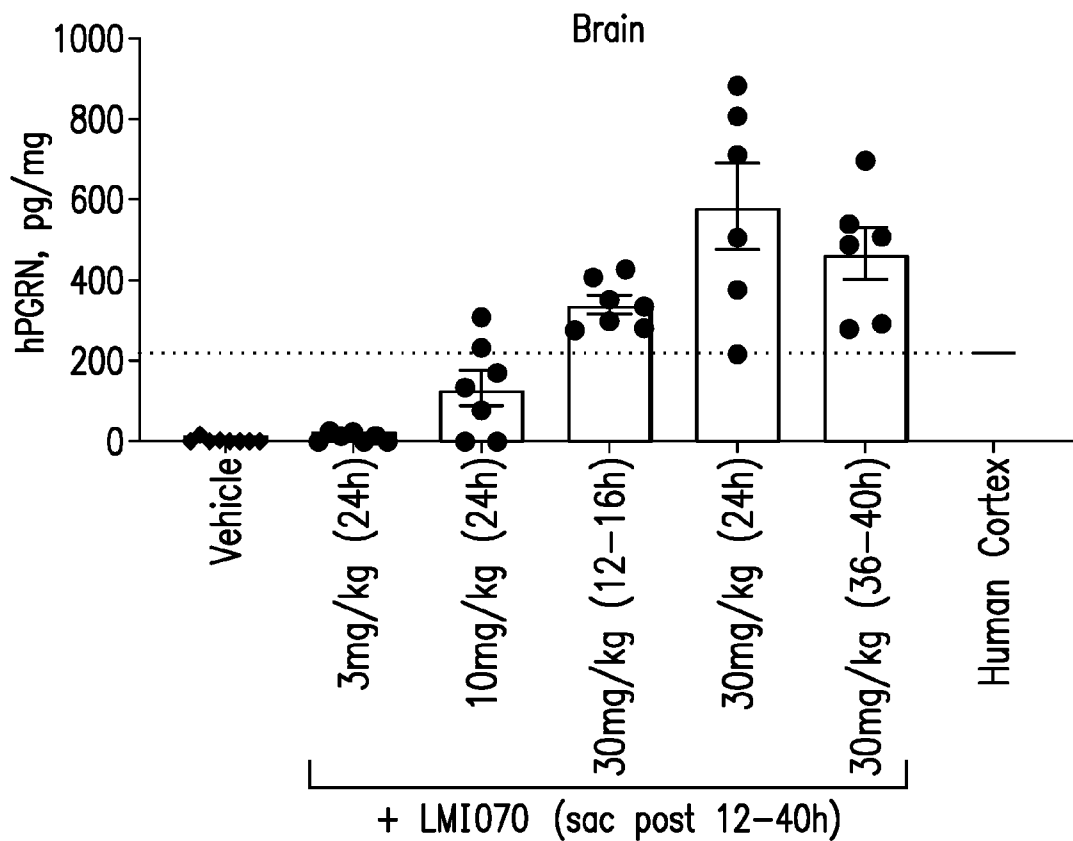


FIG.8B

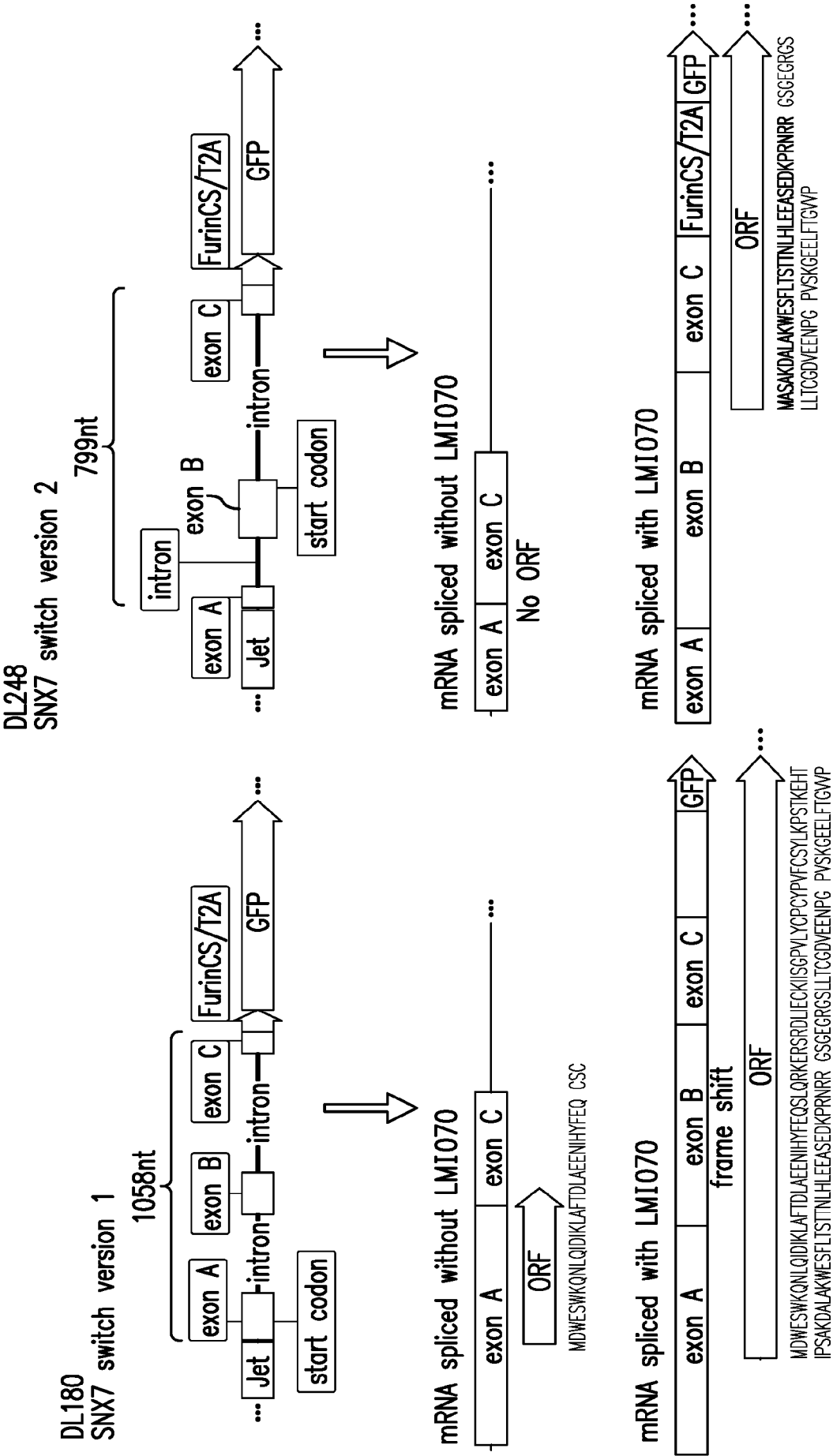
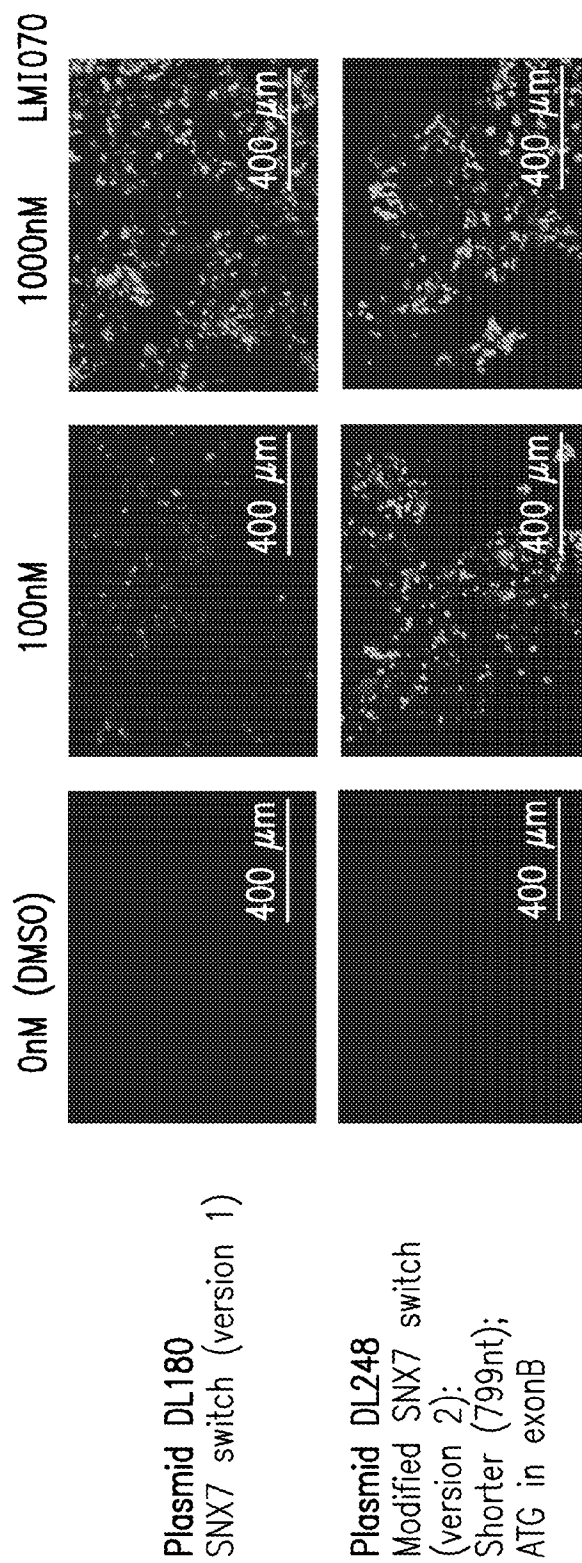


FIG. 9



SEQUENCE LISTING

<110> NOVARTIS AG

<120> REGULATABLE EXPRESSION SYSTEMS

<130> PAT058643-WO-PCT

<140>

<141>

<150> 63/013,283

<151> 2020-04-21

<150> 62/878,514

<151> 2019-07-25

<160> 113

<170> PatentIn version 3.5

<210> 1

<211> 73

<212> DNA

<213> Homo sapiens

<400> 1

ccttgctatc cctgtcttct gtagctattc tgaaccatc aacaaaggag cacaccattc 60

catcagcaaa aga 73

<210> 2

<211> 72

<212> DNA

<213> Unknown

<220>

<221> source

<223> /note="Description of Unknown:
Exonic target sequence of LMI070"

<400> 2

gtaattagct gagaaggaag atctgaaggt ttaacgagag agggcgagag atacaaaata 60

tctgctagga ga 72

<210> 3

<211> 64

<212> DNA

<213> Unknown

<220>

<221> source
 <223> /note="Description of Unknown:
 Exonic target sequence of LMI070"

<400> 3
 ggattgtttg tattcctgcc aatgatttgt gagacagtct gttccccaca tcctcgtcaa 60
 caga 64

<210> 4
 <211> 78
 <212> DNA
 <213> Unknown

<220>
 <221> source
 <223> /note="Description of Unknown:
 Exonic target sequence of LMI070"

<400> 4
 ctttctgaca tcttaacgag gcaatacaga gagacgaatt ttcacagtt tggtcaggga 60
 gacacatata acaaaaga 78

<210> 5
 <211> 51
 <212> DNA
 <213> Unknown

<220>
 <221> source
 <223> /note="Description of Unknown:
 Exonic target sequence of LMI070"

<400> 5
 atccatacat acttaatgct gaaatgtgaa gggctgagaa aaaagaaaag a 51

<210> 6
 <211> 97
 <212> DNA
 <213> Unknown

<220>
 <221> source
 <223> /note="Description of Unknown:
 Exonic target sequence of LMI070"

<400> 6
 aattggaaac atcgaggga aatgggcttt ttattattaa aacaaaacct cagtattatc 60
 acttagaaac ctgaaattga actccaaaag ccaaaga 97

<210> 7
<211> 95
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Exonic target sequence of LMI070"

<400> 7
aagaatgttc cttttgtgaa gaatgactta aggaagattc atgatgactg agtgtgcccg 60

tgtggaactt taggacatag atgcactcct acaga 95

<210> 8
<211> 70
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Exonic target sequence of LMI070"

<400> 8
ttgtccttca ctccgtactc cagttggcca agcatagggtc gcatgccagg gtcaaggaga 60

ctaagggaga 70

<210> 9
<211> 41
<212> DNA
<213> Unknown

<220>
<221> source
<223> /note="Description of Unknown:
Exonic target sequence of LMI070"

<400> 9
gacatacaga catggcagcc cctagcatgt gtatcctaag a 41

<210> 10
<211> 40
<212> DNA
<213> Unknown

<220>

<221> source
<223> /note="Description of Unknown:
Exonic target sequence of LMI070"

<400> 10
acatacagac atggcagccc ctagcatgtg taticctaaga 40

<210> 11
<211> 2438
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 11
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctc ggcgaccttt 60
ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtgggaatt cgggcggagt 120
tagggcggag ccaatcagcg tgcgccgttc cgaaagttgc cttttatggc tgggcggaga 180
atgggcggtg aacgccgatg attatataag gacgcgccgg gtgtggcaca gctagtccg 240
tcgcagccgg gatttggttc gcggttcttg tttgtggatc cctgtgatcg tcacttgaca 300
ccggtcttcc agaggagatt ggaaaacttg aagaagaagt ggattgtgct aatattgccc 360
tgaaagcagc caccatggat tgggagagtt ggaaacaaaa tttgcaaatt gatataag 420
tagcatttac agatttggtc gaggagaata tccattatct tgaacaggta attagtgttg 480
tttgatattg cttcatttta aagttatttg ctcatctact tttgggtccg ccatgttgga 540
aagagtgtat taaagaacaa gtgtcacatt ctattgcctc tctggtagct tggttttgtt 600
gaagtgtgca gttaccattt ggttttgttt atcctcagtt tgttgttttg gatttggtt 660
cttcaaaagc atttgatatt gctttctatt gattgtccta actactcctc tttcctctcc 720
cttctccatt tttgaagagt ttgcaaagga aggaaaggag cagagacttg attgagcaga 780
aaatcatttc agggcctgtt ctctattgtc cttgctatcc tgtcttctgt agctatctga 840
aaccatcaac aaaggagcac accattccat cagcaaaaga gtaacaacat ctttttttaa 900
gttcattttg tttttcagtt gattgtatct caattttttt acagctgact tttctcagag 960
aagttttttt tttattgtaa acatactttt tctagaaagt atattttaaa ataacatctt 1020
taaccttatc tctggctgaa ttattgaata tttgaaatta ttacattaac aaaattttgt 1080

cttacagcag	tggtcccca	ccttcttagc	agtagcatcc	ctcattaaga	attaaaattt	1140
gtagaaattg	acaaggattc	tgacaagctg	ttgggagaga	agaatagagc	agattgcagt	1200
aggaacagtt	gtgttagaat	ttattaatcc	tttaacactg	aaagtaaact	attgttgatt	1260
gcctcttggt	gtgtttccat	tattcagtg	tcttgctaag	tgggagtc	tccttacatc	1320
aaccaccaac	cttcacttgg	aagaagctag	cgaagataaa	cctcgcaacc	gccgcggcag	1380
cggcgaaggc	cgcggcagcc	tgctgacctg	cggcgatgtg	gaagaaaacc	cgggcccggg	1440
gagcaagggc	gaggagctgt	tcaccggggg	ggtgcccac	ctggtcgagc	tggacggcga	1500
cgtaaacggc	cacaagttca	gcgtgtccgg	cgagggcgag	ggcgatgcca	cctacggcaa	1560
gctgaccctg	aagttcatct	gcaccaccgg	caagctgccc	gtgccctggc	ccaccctcgt	1620
gaccaccctg	acctacggcg	tgcaagtgtt	cagccgctac	cccgaccaca	tgaagcagca	1680
cgacttcttc	aagtccgcca	tgcccgaagg	ctacgtccag	gagcgcacca	tcttcttcaa	1740
ggacgacggc	aactacaaga	cccgcgccga	ggtgaagttc	gagggcgaca	ccctggtgaa	1800
ccgcatcgag	ctgaagggca	tcgacttcaa	ggaggacggc	aacatcctgg	ggcacaagct	1860
ggagtacaac	tacaacagcc	acaacgtcta	tatcatggcc	gacaagcaga	agaacggcat	1920
caaggtgaac	ttcaagatcc	gccacaacat	cgaggacggc	agcgtgcagc	tcgccgacca	1980
ctaccagcag	aacaccccca	tcggcgacgg	ccccgtgctg	ctgcccgaca	accactacct	2040
gagcaccag	tccgccctga	gcaaagaccc	caacgagaag	cgcgatcaca	tggtcctgct	2100
ggagtctgtg	accgccgccg	ggatcactct	cggcatggac	gagctgtaca	agtaatgctt	2160
tatttgtgaa	atttgtgatg	ctattgcttt	atttgttaac	attataagct	gcaataaaca	2220
agttaacaac	aacaattgca	ttcatTTTTAT	gtttcagggt	cagggggagg	tgtgggagg	2280
TTTTTAAAT	cgatctgagg	aaccctagt	gatggagttg	gccactccct	ctctgcgcgc	2340
tcgctcgctc	actgaggccg	ggcgaccaa	ggtcgcccga	cggccgggct	ttgcccgggc	2400
ggcctcagtg	agcgagcgag	cgcgcagaga	gggagtg			2438

<210> 12

<211> 106

<212> DNA

<213> Artificial Sequence

<220>

<221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 12
 ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctc ggcgaccttt 60
 ggtcgcccgg cctcagttag cgagcgagcg cgcagagagg gagtgg 106

<210> 13
 <211> 194
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 13
 aattcgggcg gagttagggc ggagccaatc agcgtgcgcc gttccgaaag ttgcctttta 60
 tggctgggcg gagaatgggc ggtgaacgcc gatgattata taaggacgcg ccgggtgtgg 120
 cacagctagt tccgtcgag cggggatttg ggtcgcggtt cttgtttgtg gatccctgtg 180
 atcgtcactt gaca 194

<210> 14
 <211> 162
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 14
 cttccagagg agattggaaa acttgaagaa gaagtggatt gtgctaatat tgccctgaaa 60
 gcagccacca tggattggga gagttggaaa caaaatttgc aaattgatat caagttagca 120
 tttacagatt tggctgagga gaatatccat tattttgaac ag 162

<210> 15
 <211> 270
 <212> DNA
 <213> Artificial Sequence

<220>

<221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 15
 gtaattagtg ttgtttgata ttgcttcatt ttaaagttat ttgctcattt acttttggtc 60
 cgtccattgt tgaaagagtg tattaagaa caagtgtcac attctattgc ctctctggtg 120
 gcttggtttt gttgaagttg tcagttacca tttggttttg tttatcctca gtttgttggt 180
 ttggatttgg attcttcaaa agcatttgat attgctttct attgattgtc ctaactactc 240
 ctctttcctc tcccttctcc atttttgaag 270

<210> 16
 <211> 143
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 16
 agtttgcaaa ggaaggaaag gagcagagac ttgattgagc agaaaatcat ttcagggcct 60
 gttctctatt gtccttgcta tcctgtcttc tgtagctatc tgaaaccatc aacaaaggag 120
 cacaccattc catcagcaaa aga 143

<210> 17
 <211> 407
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 17
 gtaacaacat ctttttttaa gttcattttg tttttcagtt gattgtattt caattttttt 60
 acagctgact tttctcagag aagttttttt tttattgtaa acatactttt tctagaaagt 120
 atattttaaa ataacatctt taaccttata tctggctgaa ttattgaata ttgaaatta 180
 ttacattaac aaaattttgt cttacagcag tgggtcccaa ctttcttagc agtagcatcc 240
 ctcatthaaga attaaaattt gtagaaattg acaaggattc tgacaagctg ttgggagaga 300

agaatagagc agattgcagt aggaacagtt gtgttagaat ttattaatcc tttaacactg 360

aaagtaaact attgttgatt gcctcttggt gtgtttccat tattcag 407

<210> 18

<211> 76

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 18

tgctcttgct aagtgggagt cattccttac atcaaccacc aaccttcact tggaagaagc 60

tagcgaagat aaacct 76

<210> 19

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 19

cgcaaccgcc gc 12

<210> 20

<211> 63

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 20

ggcagcggcg aaggccgcgg cagcctgctg acctgcggcg atgtggaaga aaacccgggc 60

ccg 63

<210> 21

<211> 717

<212> DNA
<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 21

gtgagcaagg gcgaggagct gttcaccggg gtggtgcccc tcctggtcga gctggacggc	60
gacgtaaacg gccacaagtt cagcgtgtcc ggcgagggcg agggcgatgc cacctacggc	120
aagctgaccc tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg gcccaccctc	180
gtgaccaccc tgacctacgg cgtgcagtgc ttcagccgct accccgacca catgaagcag	240
cacgacttct tcaagtccgc catgcccga ggctacgtcc aggagcgcac catcttcttc	300
aaggacgacg gcaactacaa gaccgcgcc gaggtgaagt tcgagggcga caccctggtg	360
aaccgcatcg agctgaaggg catcgacttc aaggaggacg gcaacatcct ggggcacaag	420
ctggagtaca actacaacag ccacaacgtc tatatcatgg ccgacaagca gaagaacggc	480
atcaaggtga acttcaagat ccgccacaac atcgaggacg gcagcgtgca gctcgccgac	540
cactaccagc agaacacccc catcggcgac ggccccgtgc tgctgcccga caaccactac	600
ctgagcaccc agtccgccct gagcaaagac cccaacgaga agcgcgatca catggtcctg	660
ctggagttcg tgaccgccgc cgggatcact ctcggcatgg acgagctgta caagtaa	717

<210> 22

<211> 132

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 22

tgctttatth gtgaaatttg tgatgctatt gctttatthg taaccattat aagctgcaat	60
aaacaagtta acaacaacaa ttgcattcat tttatgtthc aggttcaggg ggaggtgtgg	120
gaggtththh aa	132

<210> 23

<211> 141

<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 23
aggaaccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg 60
ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc 120
gagcgcgcag agagggagtg g 141

<210> 24

<400> 24
000

<210> 25

<400> 25
000

<210> 26

<400> 26
000

<210> 27

<400> 27
000

<210> 28

<400> 28
000

<210> 29

<400> 29
000

<210> 30

<400> 30
000

<210> 31

<400> 31
000

<210> 32

<400> 32
000

<210> 33

<400> 33
000

<210> 34

<400> 34
000

<210> 35
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> MOD_RES
<222> (5)..(5)
<223> Any amino acid

<400> 35
Ile Glu Pro Asp Xaa
1 5

<210> 36
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 36
Asp Asp Asp Asp Lys
1 5

<210> 37
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 37
Pro Gly Ala Ala His Tyr
1 5

<210> 38
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 38
Leu Glu Val Phe Gln Gly Pro
1 5

<210> 39
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 39
Arg Asn Arg Arg
1

<210> 40
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 40
Leu Val Pro Arg Gly Ser
1 5

<210> 41
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 41
Glu Asn Leu Tyr Phe Gln Gly
1 5

<210> 42
<211> 2
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<220>
<221> VARIANT
<222> (1)..(1)
<223> /replace="Gly" or "Ser" or "Val"

<220>
<221> MOD_RES
<222> (2)..(2)
<223> Any amino acid

<220>
<221> SITE

<222> (1)..(2)
<223> /note="Variant residues given in the sequence have no preference with respect to those in the annotations for variant positions"

<400> 42
Ala Xaa
1

<210> 43
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 43
Arg Thr Lys Arg
1

<210> 44
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<400> 44
cgtactaaaa ga

12

<210> 45
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 45
Gly Thr Gly Ala Glu Asp Pro Arg Pro Ser Arg Lys Arg Arg Ser Leu
1 5 10 15

Gly Asp Val Gly
20

<210> 46
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 46
ggaaccggcg cggaagaccc ccggccctcc aggaagcgaa ggtccctcgg agacgtgggt 60

<210> 47
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 47
Gly Thr Gly Ala Glu Asp Pro Arg Pro Ser Arg Lys Arg Arg
1 5 10

<210> 48
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 48
ggaaccggcg cggaagaccc ccggccctcc aggaagcgaa gg 42

<210> 49
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 49

Leu Gln Trp Leu Glu Gln Gln Val Ala Lys Arg Arg Thr Lys Arg
1 5 10 15

<210> 50

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<400> 50

ctgcaatggc tggagcagca ggtggcgaag cggagaacta agcgg

45

<210> 51

<211> 18

<212> PRT

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 51

Gly Thr Gly Ala Glu Asp Pro Arg Pro Ser Arg Lys Arg Arg Ser Leu
1 5 10 15

Gly Gly

<210> 52

<211> 54

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<400> 52

ggcacaggtg ccgaggaccc tcggccaagc cgcaaaagga ggtcacttgg cggc

54

<210> 53
<211> 17
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 53
Gly Thr Gly Ala Glu Asp Pro Arg Pro Ser Arg Lys Arg Arg Ser Leu
1 5 10 15

Gly

<210> 54
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<400> 54
ggaaccggag cagaagatcc cagaccaagc cggaaaaggc ggtccctggg t 51

<210> 55
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 55
Ser Leu Asn Leu Thr Glu Ser His Asn Ser Arg Lys Lys Arg
1 5 10

<210> 56
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 56
agtctcaatt tgactgagtc acacaattcc aggaagaaaa gg

42

<210> 57
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<400> 57
Cys Lys Ile Asn Gly Tyr Pro Lys Arg Gly Arg Lys Arg Arg
1 5 10

<210> 58
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<400> 58
tgcaagatca acggctaccc taagaggggc agaaagcggc gg

42

<210> 59
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
peptide"

<220>
<221> VARIANT
<222> (1)..(3)
<223> /replace=" "

<220>
 <221> SITE
 <222> (1)..(21)
 <223> /note="Variant residues given in the sequence have no
 preference with respect to those in the annotations
 for variant positions"

 <400> 59
 Gly Ser Gly Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu
 1 5 10 15

Glu Asn Pro Gly Pro
 20

<210> 60

<400> 60
 000

<210> 61
 <211> 18
 <212> PRT
 <213> Thosea asigna virus 2A

<400> 61
 Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro
 1 5 10 15

Gly Pro

<210> 62
 <211> 53
 <212> DNA
 <213> Thosea asigna virus 2A

<400> 62
 gaaggccgcg gcagcctgct gacctgcggc gatgtggaag aaaacccggg ccc 53

<210> 63
 <211> 22
 <212> PRT
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 peptide"

<220>
<221> VARIANT
<222> (1)..(3)
<223> /replace=" "

<220>
<221> SITE
<222> (1)..(22)
<223> /note="Variant residues given in the sequence have no
preference with respect to those in the annotations
for variant positions"

<400> 63
Gly Ser Gly Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val
1 5 10 15

Glu Glu Asn Pro Gly Pro
20

<210> 64
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
oligonucleotide"

<220>
<221> variation
<222> (1)..(9)
<223> /replace=" "

<220>
<221> misc_feature
<222> (1)..(66)
<223> /note="Variant nucleotides given in the sequence have no
preference with respect to those in the annotations
for variant positions"

<400> 64
ggcagcggcg ccaccaactt cagcctgctg aagcaggccg gcgacgtgga ggagaacccc 60
ggcccc 66

<210> 65
<211> 23

<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<220>
<221> VARIANT
<222> (1)..(3)
<223> /replace=" "

<220>
<221> SITE
<222> (1)..(23)
<223> /note="Variant residues given in the sequence have no preference with respect to those in the annotations for variant positions"

<400> 65
Gly Ser Gly Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp
1 5 10 15

Val Glu Ser Asn Pro Gly Pro
20

<210> 66
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<220>
<221> variation
<222> (1)..(9)
<223> /replace=" "

<220>
<221> misc_feature
<222> (1)..(69)
<223> /note="Variant nucleotides given in the sequence have no preference with respect to those in the annotations for variant positions"

<400> 66

ggcagcggcc agtgcaccaa ctacgccctg ctgaagctgg ccggcgacgt ggagagcaac 60
cccggcccc 69

<210> 67
<211> 25
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<220>
<221> VARIANT
<222> (1)..(3)
<223> /replace=" "

<220>
<221> SITE
<222> (1)..(25)
<223> /note="Variant residues given in the sequence have no preference with respect to those in the annotations for variant positions"

<400> 67
Gly Ser Gly Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala
1 5 10 15

Gly Asp Val Glu Ser Asn Pro Gly Pro
20 25

<210> 68
<211> 75
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<220>
<221> variation
<222> (1)..(9)
<223> /replace=" "

<220>

<221> misc_feature
 <222> (1)..(75)
 <223> /note="Variant nucleotides given in the sequence have no preference with respect to those in the annotations for variant positions"

<400> 68
 ggagcggcg tgaagcagac cctgaacttc gacctgctga agctggccgg cgacgtggag 60
 agcaaccccg gcccc 75

<210> 69
 <211> 9
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<400> 69
 gaagaagaa 9

<210> 70
 <211> 6
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic oligonucleotide"

<400> 70
 gccacc 6

<210> 71
 <211> 1058
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 71
 cttccagagg agattggaaa acttgaagaa gaagtggatt gtgctaatat tgccctgaaa 60
 gcagccacca tggattggga gagttggaaa caaaatttgc aaattgatata caagttagca 120

tttacagatt tggctgagga gaatatccat tattttgaac aggtaattag tgttgtttga	180
tattgcttca ttttaaagtt atttgctcat ttacttttgg tccgtccatt gttgaaagag	240
tgtattaaag aacaagtgtc acattctatt gcctctctgg tagcttggtt ttgttgaagt	300
tgtcagttac catttggttt tgtttatcct cagtttggtt ttttggattt ggattcttca	360
aaagcatttg atattgcttt ctattgattg tcctaactac tcctctttcc tctcccttct	420
ccatttttga agagtttgca aaggaaggaa aggagcagag acttgattga gcagaaaatc	480
atttcagggc ctgttctcta ttgtccttgc tatcctgtct tctgtagcta tctgaaacca	540
tcaacaaagg agcacaccat tccatcagca aaagagtaac aacatctttt tttaagttca	600
ttttgttttt cagttgattg tatttcaatt tttttacagc tgacttttct cagagaagtt	660
ttttttttat tgtaaacata ctttttctag aaagtatat ttaaaataac atctttaacc	720
ttatctctgg ctgaattatt gaatatttga aattattaca ttaacaaaat tttgtcttac	780
agcagtggtc cccaaccttc ttagcagtag catccctcat taagaattaa aattttaga	840
aattgacaag gattctgaca agctgttggg agagaagaat agagcagatt gcagtaggaa	900
cagttgtgtt agaatttatt aatccttta cactgaaagt aaactattgt tgattgcctc	960
ttggtgtgtt tccattattc agtgctcttg ctaagtggga gtcattcctt acatcaacca	1020
ccaaccttca cttggaagaa gctagcgaag ataaacct	1058

<210> 72

<211> 720

<212> DNA

<213> Unknown

<220>

<221> source

<223> /note="Description of Unknown:
PRE sequence"

<400> 72

acaggcctat tgattggaaa gtatgtcaac gaattgtggg tcttttgggg tttgctgccc	60
cttttacgca atgtggatat cctgctttta tgcctttata tgcatgtata caagcaaaac	120
aggcttttac tttctcgcca acttacaagg cttttctaag taaacagtat ctgacccttt	180
accccgttgc tcggcaacgg cctggctctgt gccaaagtgt tgctgacgca acccccactg	240
gttggggctt ggccataggc catcagcgca tgcgtggaac ctttgtgtct cctctgccga	300

tccatactgc ggaactccta gccgcttggt ttgctcgcag caggtctgga gcgaaactca	360
tcgggactga caattctgtc gtgctctccc gcaagtatac atcgtttcca gggctgctag	420
gctgtgctgc caactggatc ctgcgcggga cgtcctttgt ttacgtcccg tcggcgctga	480
atccccgcgga cgaccctcc cggggccgct tggggctcta ccgcccgtt ctccgtctgc	540
cgtaccgacc gaccacgggg gcgacctctc tttacgcgga ctccccgtct gtgccttctc	600
atctgccgga ccgtgtgcac ttcgcttcac ctctgcacgt cgcatggaga ccaccgtgaa	660
cgcccaccgg aacctgccc aggtcttgca taagaggact cttggacttt cagcaatgtc	720

<210> 73

<211> 721

<212> DNA

<213> Unknown

<220>

<221> source

<223> /note="Description of Unknown:
PRE sequence"

<400> 73

aacaggccta ttgattggaa agtatgtcaa cgaattgtgg gtcttttggg gtttgctgcc	60
ccttttacgc aatgtggata tcctgcttta atgcctttat atgcatgtat acaagcaaaa	120
caggctttta ctttctcgcc aacttacaag gcctttctaa gtaaacagta tctgaccctt	180
taccccgttg ctcggcaacg gcctggtctg tgccaagtgt ttgtgacgc aacccccact	240
ggttggggct tggccatagg ccatcagcgc atgcgtggaa cctttgtgtc tcctctgccg	300
atccatactg cggaactcct agccgcttgt tttgctcgca gcaggtctgg agcgaaactc	360
atcgggactg acaattctgt cgtgctctcc cgcaagtata catcgtttcc agggctgcta	420
ggctgtgctg ccaactggat cctgcgcggg acgtcctttg tttacgtccc gtcggcgctg	480
aatccccgcg acgaccctc ccggggccgc ttggggctct accgcccgt tctccgtctg	540
ccgtaccgac cgaccacggg gcgcacctct ctttacgcgg actccccgtc tgtgccttct	600
catctgccgg accgtgtgca cttcgcttca cctctgcacg tcgcatggag accaccgtga	660
acgcccaccg gaacctgccc aaggtcttgc ataagaggac tcttggactt tcagcaatgt	720
c	721

<210> 74
<211> 2211
<212> DNA
<213> Adeno-associated virus

<400> 74
atggctgccg atggttatct tccagattgg ctcgaggaca accttagtga aggaattcgc 60
gagtgggtggg ctttgaaacc tggagcccct caacccaagg caaatcaaca acatcaagac 120
aacgctcgag gtcttgtgct tccgggttac aaataccttg gacccggcaa cggactcgac 180
aagggggagc cgggtcaacgc agcagacgcg gcggccctcg agcacgacaa ggcctacgac 240
cagcagctca aggccggaga caacccttac ctcaagtaca accacgccga cgccgagttc 300
caggagcggc tcaaagaaga tacgtctttt gggggcaacc tcgggcgagc agtcttccag 360
gccaaaaaga ggcttcttga acctcttggg ctggttgagg aagcggctaa gacggctcct 420
ggaaagaaga ggcctgtaga gcagtctcct caggaaccgg actcctccgc gggatttggc 480
aaatcgggtg cacagccgcg taaaaagaga ctcaatttcg gtcagactgg cgacacagag 540
tcagtcccag acctcaacc aatcggagaa cctcccgag cccctcagg tgtgggatct 600
cttacaatgg cttcaggtgg tggcgcacca gtggcagaca ataacgaagg tgccgatgga 660
gtgggtagtt cctcgggaaa ttggcattgc gattcccaat ggctggggga cagagtcac 720
accaccagca cccgaacctg ggccctgccc acctacaaca atcacctcta caagcaaadc 780
tccaacagca catctggagg atcttcaa at gacaacgcct acttcggcta cagcaccccc 840
tgggggtatt ttgacttcaa cagattccac tgccacttct caccacgtga ctggcagcga 900
ctcatcaaca acaactgggg attccggcct aagcgactca acttcaagct cttcaacatt 960
caggtcaaag aggttacgga caacaatgga gtcaagacca tcgccaataa ccttaccagc 1020
acggtccagg tcttcacgga ctcagactat cagctcccgt acgtgctcgg gtcggctcac 1080
gagggctgcc tcccgccgtt cccagcggac gttttcatga ttcctcagta cgggtatctg 1140
acgcttaatg atggaagcca ggccgtgggt cgttcgtcct tttactgcct ggaatatttc 1200
ccgtcgcaaa tgctaagaac gggtaacaac ttccagttca gctacgagtt tgagaacgta 1260
cctttccata gcagctacgc tcacagccaa agcctggacc gactaatgaa tccactcatc 1320
gaccaatact tgtactatct ctcaaagact attaacgggt ctggacagaa tcaacaaacg 1380
ctaaaattca gtgtggccgg acccagcaac atggctgtcc aggaagaaa ctacatacct 1440

ggacccagct accgacaaca acgtgtctca accactgtga ctcaaaacaa caacagcgaa	1500
tttgcttggc ctggagcttc ttcttgggct ctcaatggac gtaatagctt gatgaatcct	1560
ggacctgcta tggccagcca caaagaagga gaggaccgtt tctttccttt gtctggatct	1620
ttaatTTTTg gcaacaagg aactggaaga gacaacgtgg atgcggacaa agtcatgata	1680
accaacgaag aagaaattaa aactactaac ccggtagcaa cggagtccta tggacaagtg	1740
gccacaaacc accagagtgc ccaagcacag gcgcagaccg gctgggttca aaaccaagga	1800
atacttccgg gtatggtttg gcaggacaga gatgtgtacc tgcaaggacc catttgggcc	1860
aaaattcctc acacggacgg caactttcac ctttctccgc tgatgggagg gtttggaatg	1920
aagcaccgc ctctcagat cctcatcaaa aacacacctg tacctgcgga tcctccaacg	1980
gccttcaaca aggacaagct gaactctttc atcaccagct attctactgg ccaagtcagc	2040
gtggagatcg agtgggagct gcagaaggaa aacagcaagc gctggaaccc ggagatccag	2100
tacacttcca actattacaa gtctaataat gttgaatttg ctgttaatac tgaagggtga	2160
tatagtgaac cccgccccat tggcaccaga tacctgactc gtaatctgta a	2211

<210> 75

<211> 1800

<212> DNA

<213> Adeno-associated virus

<400> 75

acggctcctg gaaagaagag gcctgtagag cagtctcctc aggaaccgga ctctccgcg	60
ggtattggca aatcgggtgc acagcccgct aaaaagagac tcaatttcgg tcagactggc	120
gacacagagt cagtcccaga ccctcaacca atcggagaac ctcccgagc cccctcaggt	180
gtgggatctc ttacaatggc ttcaggtggt ggcgcaccag tggcagacaa taacgaaggt	240
gccgatggag tgggtagttc ctcgggaaat tggcattgcg attcccaatg gctgggggac	300
agagtcatca ccaccagcac ccgaacctgg gccctgcca cctacaacaa tcacctctac	360
aagcaaattc ccaacagcac atctggagga tcttcaaattg acaacgccta cttcggctac	420
agcaccctt ggggggtattt tgacttcaac agattccact gccatttctc accacgtgac	480
tggcagcgac tcatcaacaa caactgggga ttccggccta agcgactcaa cttcaagctc	540
ttcaacattc aggtcaaaga gggttacggac aacaatggag tcaagaccat cgccaataac	600
cttaccagca cggtccaggt cttcacggac tcagactatc agctcccgtg cgtgctcggg	660

tcggctcacg agggctgcct cccgccgttc ccagcggacg ttttcatgat tcctcagtac	720
gggtatctga cgcttaatga tggaagccag gccgtgggtc gttcgtcctt ttactgcctg	780
gaatatttcc cgtcgcaaat gctaagaacg ggtaacaact tccagttcag ctacgagttt	840
gagaacgtac ctttccatag cagctacgct cacagccaaa gcctggaccg actaatgaat	900
ccactcatcg accaatactt gtactatctc tcaaagacta ttaacggttc tggacagaat	960
caacaaacgc taaaattcag tgtggccgga cccagcaaca tggctgtcca gggaagaaac	1020
tacatacctg gaccagcta ccgacaacaa cgtgtctcaa ccactgtgac tcaaaacaac	1080
aacagcgaat ttgcttggcc tggagcttct tcttgggctc tcaatggacg taatagcttg	1140
atgaatcctg gacctgctat ggccagccac aaagaaggag aggaccgttt ctttcctttg	1200
tctggatctt taatTTTTGG caaacaagga actggaagag acaacgtgga tgcggacaaa	1260
gtcatgataa ccaacgaaga agaaattaaa actactaacc cggtagcaac ggagtcctat	1320
ggacaagtgg ccacaaacca ccagagtgcc caagcacagg cgagaccgg ctgggttcaa	1380
aaccaaggaa tacttccggg tatggttttg caggacagag atgtgtacct gcaaggacc	1440
atttgggcca aaattcctca cacggacggc aactttcacc cttctccgct gatgggaggg	1500
tttggaatga agcaccgcc tcctcagatc ctcatcaaaa acacacctgt acctgcggat	1560
cctccaacgg ctttcaacaa ggacaagctg aactctttca tcaccagta ttctactggc	1620
caagtcagcg tggagatcga gtgggagctg cagaaggaaa acagcaagcg ctggaacccg	1680
gagatccagt acatttcaa ctattacaag tctaataatg ttgaatttgc tgttaatact	1740
gaaggtgtat atagtgaacc ccgccccatt ggcaccagat acctgactcg taatctgtaa	1800

<210> 76

<211> 1605

<212> DNA

<213> Adeno-associated virus

<400> 76

atggcttcag gtggtggcgc accagtggca gacaataacg aaggtgccga tggagtgggt	60
agttcctcgg gaaattggca ttgcgattcc caatggctgg gggacagagt catcaccacc	120
agcaccgaa cctgggccct gccacctac aacaatcacc tctacaagca aatctccaac	180
agcacatctg gaggatcttc aaatgacaac gcctacttcg gctacagcac cccctggggg	240

tatthttgact tcaacagatt ccactgccac ttctcaccac gtgactggca gcgactcatc	300
aacaacaact ggggattccg gcctaagcga ctcaacttca agctcttcaa cattcaggtc	360
aaagagggtta cggacaacaa tggagtcaag accatcgcca ataaccttac cagcacggtc	420
cagggtcttca cggactcaga ctatcagctc ccgtacgtgc tcgggtcggc tcacgagggc	480
tgcctcccg cgttcccagc ggacgttttc atgattcctc agtacgggta tctgacgctt	540
aatgatggaa gccaggccgt gggtcgttcg tccttttact gcctggaata tttcccgtcg	600
caaatgctaa gaacgggtaa caacttccag tttagctacg agtttgagaa cgtacctttc	660
catagcagct acgctcacag ccaaagcctg gaccgactaa tgaatccact catcgaccaa	720
tacttgtact atctctcaaa gactattaac ggttctggac agaatacaaa aacgctaaaa	780
ttcagtggtg ccggaccag caacatggct gtccagggaa gaaactacat acctggaccc	840
agctaccgac aacaacgtgt ctcaaccact gtgactcaaa acaacaacag cgaatttgct	900
tggcctggag cttcttcttg ggctctcaat ggacgtaata gcttgatgaa tcctggacct	960
gctatggcca gccacaaaga aggagaggac cgtttctttc ctttgtctgg atctttaatt	1020
tttggcaaac aaggaactgg aagagacaac gtggatgcgg acaaagtcac gataaccaac	1080
gaagaagaaa ttaaaactac taaccggta gcaacggagt cctatggaca agtggccaca	1140
aaccaccaga gtgccaagc acaggcgcag accggctggg ttcaaaacca aggaatactt	1200
ccgggtatgg tttggcagga cagagatgtg tacctgcaag gaccatcttg ggccaaaatt	1260
cctcacacgg acggcaactt tcacccttct ccgctgatgg gagggtttgg aatgaagcac	1320
ccgcctctc agatcctcat caaaaacaca cctgtacctg cggatcctcc aacggccttc	1380
aacaaggaca agctgaactc tttcatcacc cagtattcta ctggccaagt cagcgtggag	1440
atcgagtggg agctgcagaa ggaaaacagc aagcgctgga acccgagat ccagtacact	1500
tccaactatt acaagtctaa taatgttgaa tttgctgtta atactgaagg tgtatatagt	1560
gaaccccgcc ccattggcac cagatacctg actcgtaatc tgtaa	1605

<210> 77

<211> 736

<212> PRT

<213> Adeno-associated virus

<400> 77

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser

1		5		10		15											
Glu	Gly	Ile	Arg	Glu	Trp	Trp	Ala	Leu	Lys	Pro	Gly	Ala	Pro	Gln	Pro		
		20						25					30				
Lys	Ala	Asn	Gln	Gln	His	Gln	Asp	Asn	Ala	Arg	Gly	Leu	Val	Leu	Pro		
		35					40					45					
Gly	Tyr	Lys	Tyr	Leu	Gly	Pro	Gly	Asn	Gly	Leu	Asp	Lys	Gly	Glu	Pro		
	50					55					60						
Val	Asn	Ala	Ala	Asp	Ala	Ala	Ala	Leu	Glu	His	Asp	Lys	Ala	Tyr	Asp		
65					70					75					80		
Gln	Gln	Leu	Lys	Ala	Gly	Asp	Asn	Pro	Tyr	Leu	Lys	Tyr	Asn	His	Ala		
				85					90					95			
Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly		
			100					105					110				
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro		
		115					120					125					
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg		
	130					135					140						
Pro	Val	Glu	Gln	Ser	Pro	Gln	Glu	Pro	Asp	Ser	Ser	Ala	Gly	Ile	Gly		
145					150					155					160		
Lys	Ser	Gly	Ala	Gln	Pro	Ala	Lys	Lys	Arg	Leu	Asn	Phe	Gly	Gln	Thr		
				165					170					175			
Gly	Asp	Thr	Glu	Ser	Val	Pro	Asp	Pro	Gln	Pro	Ile	Gly	Glu	Pro	Pro		
			180					185					190				
Ala	Ala	Pro	Ser	Gly	Val	Gly	Ser	Leu	Thr	Met	Ala	Ser	Gly	Gly	Gly		
		195					200					205					
Ala	Pro	Val	Ala	Asp	Asn	Asn	Glu	Gly	Ala	Asp	Gly	Val	Gly	Ser	Ser		
	210					215					220						

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser

435

440

445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
 450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
 465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
 485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
 500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
 515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
 530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
 545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
 565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
 580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
 595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
 610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
 625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
 645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 78

<211> 599

<212> PRT

<213> Adeno-associated virus

<400> 78

Thr Ala Pro Gly Lys Lys Arg Pro Val Glu Gln Ser Pro Gln Glu Pro
1 5 10 15

Asp Ser Ser Ala Gly Ile Gly Lys Ser Gly Ala Gln Pro Ala Lys Lys
20 25 30

Arg Leu Asn Phe Gly Gln Thr Gly Asp Thr Glu Ser Val Pro Asp Pro
35 40 45

Gln Pro Ile Gly Glu Pro Pro Ala Ala Pro Ser Gly Val Gly Ser Leu
50 55 60

Thr Met Ala Ser Gly Gly Gly Ala Pro Val Ala Asp Asn Asn Glu Gly
65 70 75 80

Ala Asp Gly Val Gly Ser Ser Ser Gly Asn Trp His Cys Asp Ser Gln
85 90 95

Trp Leu Gly Asp Arg Val Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu
100 105 110

Pro Thr Tyr Asn Asn His Leu Tyr Lys Gln Ile Ser Asn Ser Thr Ser
115 120 125

Gly Gly Ser Ser Asn Asp Asn Ala Tyr Phe Gly Tyr Ser Thr Pro Trp
130 135 140

Gly Tyr Phe Asp Phe Asn Arg Phe His Cys His Phe Ser Pro Arg Asp
145 150 155 160

Trp Gln Arg Leu Ile Asn Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu
165 170 175

Asn Phe Lys Leu Phe Asn Ile Gln Val Lys Glu Val Thr Asp Asn Asn
180 185 190

Gly Val Lys Thr Ile Ala Asn Asn Leu Thr Ser Thr Val Gln Val Phe
195 200 205

Thr Asp Ser Asp Tyr Gln Leu Pro Tyr Val Leu Gly Ser Ala His Glu
210 215 220

Gly Cys Leu Pro Pro Phe Pro Ala Asp Val Phe Met Ile Pro Gln Tyr
225 230 235 240

Gly Tyr Leu Thr Leu Asn Asp Gly Ser Gln Ala Val Gly Arg Ser Ser
245 250 255

Phe Tyr Cys Leu Glu Tyr Phe Pro Ser Gln Met Leu Arg Thr Gly Asn
260 265 270

Asn Phe Gln Phe Ser Tyr Glu Phe Glu Asn Val Pro Phe His Ser Ser
275 280 285

Tyr Ala His Ser Gln Ser Leu Asp Arg Leu Met Asn Pro Leu Ile Asp
290 295 300

Gln Tyr Leu Tyr Tyr Leu Ser Lys Thr Ile Asn Gly Ser Gly Gln Asn
305 310 315 320

Gln Gln Thr Leu Lys Phe Ser Val Ala Gly Pro Ser Asn Met Ala Val

325

330

335

Gln Gly Arg Asn Tyr Ile Pro Gly Pro Ser Tyr Arg Gln Gln Arg Val
 340 345 350

Ser Thr Thr Val Thr Gln Asn Asn Asn Ser Glu Phe Ala Trp Pro Gly
 355 360 365

Ala Ser Ser Trp Ala Leu Asn Gly Arg Asn Ser Leu Met Asn Pro Gly
 370 375 380

Pro Ala Met Ala Ser His Lys Glu Gly Glu Asp Arg Phe Phe Pro Leu
 385 390 395 400

Ser Gly Ser Leu Ile Phe Gly Lys Gln Gly Thr Gly Arg Asp Asn Val
 405 410 415

Asp Ala Asp Lys Val Met Ile Thr Asn Glu Glu Glu Ile Lys Thr Thr
 420 425 430

Asn Pro Val Ala Thr Glu Ser Tyr Gly Gln Val Ala Thr Asn His Gln
 435 440 445

Ser Ala Gln Ala Gln Ala Gln Thr Gly Trp Val Gln Asn Gln Gly Ile
 450 455 460

Leu Pro Gly Met Val Trp Gln Asp Arg Asp Val Tyr Leu Gln Gly Pro
 465 470 475 480

Ile Trp Ala Lys Ile Pro His Thr Asp Gly Asn Phe His Pro Ser Pro
 485 490 495

Leu Met Gly Gly Phe Gly Met Lys His Pro Pro Pro Gln Ile Leu Ile
 500 505 510

Lys Asn Thr Pro Val Pro Ala Asp Pro Pro Thr Ala Phe Asn Lys Asp
 515 520 525

Lys Leu Asn Ser Phe Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val
 530 535 540

Glu Ile Glu Trp Glu Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro
545 550 555 560

Glu Ile Gln Tyr Thr Ser Asn Tyr Tyr Lys Ser Asn Asn Val Glu Phe
565 570 575

Ala Val Asn Thr Glu Gly Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr
580 585 590

Arg Tyr Leu Thr Arg Asn Leu
595

<210> 79

<211> 534

<212> PRT

<213> Adeno-associated virus

<400> 79

Met Ala Ser Gly Gly Gly Ala Pro Val Ala Asp Asn Asn Glu Gly Ala
1 5 10 15

Asp Gly Val Gly Ser Ser Ser Gly Asn Trp His Cys Asp Ser Gln Trp
20 25 30

Leu Gly Asp Arg Val Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro
35 40 45

Thr Tyr Asn Asn His Leu Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly
50 55 60

Gly Ser Ser Asn Asp Asn Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly
65 70 75 80

Tyr Phe Asp Phe Asn Arg Phe His Cys His Phe Ser Pro Arg Asp Trp
85 90 95

Gln Arg Leu Ile Asn Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn
100 105 110

Phe Lys Leu Phe Asn Ile Gln Val Lys Glu Val Thr Asp Asn Asn Gly
115 120 125

Val Lys Thr Ile Ala Asn Asn Leu Thr Ser Thr Val Gln Val Phe Thr
130 135 140

Asp Ser Asp Tyr Gln Leu Pro Tyr Val Leu Gly Ser Ala His Glu Gly
145 150 155 160

Cys Leu Pro Pro Phe Pro Ala Asp Val Phe Met Ile Pro Gln Tyr Gly
165 170 175

Tyr Leu Thr Leu Asn Asp Gly Ser Gln Ala Val Gly Arg Ser Ser Phe
180 185 190

Tyr Cys Leu Glu Tyr Phe Pro Ser Gln Met Leu Arg Thr Gly Asn Asn
195 200 205

Phe Gln Phe Ser Tyr Glu Phe Glu Asn Val Pro Phe His Ser Ser Tyr
210 215 220

Ala His Ser Gln Ser Leu Asp Arg Leu Met Asn Pro Leu Ile Asp Gln
225 230 235 240

Tyr Leu Tyr Tyr Leu Ser Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln
245 250 255

Gln Thr Leu Lys Phe Ser Val Ala Gly Pro Ser Asn Met Ala Val Gln
260 265 270

Gly Arg Asn Tyr Ile Pro Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser
275 280 285

Thr Thr Val Thr Gln Asn Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala
290 295 300

Ser Ser Trp Ala Leu Asn Gly Arg Asn Ser Leu Met Asn Pro Gly Pro
305 310 315 320

Ala Met Ala Ser His Lys Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser
325 330 335

Gly Ser Leu Ile Phe Gly Lys Gln Gly Thr Gly Arg Asp Asn Val Asp

340

345

350

Ala Asp Lys Val Met Ile Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn
 355 360 365

Pro Val Ala Thr Glu Ser Tyr Gly Gln Val Ala Thr Asn His Gln Ser
 370 375 380

Ala Gln Ala Gln Ala Gln Thr Gly Trp Val Gln Asn Gln Gly Ile Leu
 385 390 395 400

Pro Gly Met Val Trp Gln Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile
 405 410 415

Trp Ala Lys Ile Pro His Thr Asp Gly Asn Phe His Pro Ser Pro Leu
 420 425 430

Met Gly Gly Phe Gly Met Lys His Pro Pro Pro Gln Ile Leu Ile Lys
 435 440 445

Asn Thr Pro Val Pro Ala Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys
 450 455 460

Leu Asn Ser Phe Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val Glu
 465 470 475 480

Ile Glu Trp Glu Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu
 485 490 495

Ile Gln Tyr Thr Ser Asn Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala
 500 505 510

Val Asn Thr Glu Gly Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg
 515 520 525

Tyr Leu Thr Arg Asn Leu
 530

<210> 80

<211> 144

<212> DNA

<213> Homo sapiens

<400> 80

agtttgcaaa ggaaggaaag gagcagagac ttgaatgagc agaaaatcat ttcagggcct 60

gttctctatg tccttgctat ccctgtcttc tgtagctatt ctgaaacat caacaaagga 120

gcacaccatt ccatcagcaa aaga 144

<210> 81

<211> 593

<212> PRT

<213> Homo sapiens

<400> 81

Met Trp Thr Leu Val Ser Trp Val Ala Leu Thr Ala Gly Leu Val Ala
1 5 10 15

Gly Thr Arg Cys Pro Asp Gly Gln Phe Cys Pro Val Ala Cys Cys Leu
20 25 30

Asp Pro Gly Gly Ala Ser Tyr Ser Cys Cys Arg Pro Leu Leu Asp Lys
35 40 45

Trp Pro Thr Thr Leu Ser Arg His Leu Gly Gly Pro Cys Gln Val Asp
50 55 60

Ala His Cys Ser Ala Gly His Ser Cys Ile Phe Thr Val Ser Gly Thr
65 70 75 80

Ser Ser Cys Cys Pro Phe Pro Glu Ala Val Ala Cys Gly Asp Gly His
85 90 95

His Cys Cys Pro Arg Gly Phe His Cys Ser Ala Asp Gly Arg Ser Cys
100 105 110

Phe Gln Arg Ser Gly Asn Asn Ser Val Gly Ala Ile Gln Cys Pro Asp
115 120 125

Ser Gln Phe Glu Cys Pro Asp Phe Ser Thr Cys Cys Val Met Val Asp
130 135 140

Gly Ser Trp Gly Cys Cys Pro Met Pro Gln Ala Ser Cys Cys Glu Asp
145 150 155 160

Arg Val His Cys Cys Pro His Gly Ala Phe Cys Asp Leu Val His Thr
165 170 175

Arg Cys Ile Thr Pro Thr Gly Thr His Pro Leu Ala Lys Lys Leu Pro
180 185 190

Ala Gln Arg Thr Asn Arg Ala Val Ala Leu Ser Ser Ser Val Met Cys
195 200 205

Pro Asp Ala Arg Ser Arg Cys Pro Asp Gly Ser Thr Cys Cys Glu Leu
210 215 220

Pro Ser Gly Lys Tyr Gly Cys Cys Pro Met Pro Asn Ala Thr Cys Cys
225 230 235 240

Ser Asp His Leu His Cys Cys Pro Gln Asp Thr Val Cys Asp Leu Ile
245 250 255

Gln Ser Lys Cys Leu Ser Lys Glu Asn Ala Thr Thr Asp Leu Leu Thr
260 265 270

Lys Leu Pro Ala His Thr Val Gly Asp Val Lys Cys Asp Met Glu Val
275 280 285

Ser Cys Pro Asp Gly Tyr Thr Cys Cys Arg Leu Gln Ser Gly Ala Trp
290 295 300

Gly Cys Cys Pro Phe Thr Gln Ala Val Cys Cys Glu Asp His Ile His
305 310 315 320

Cys Cys Pro Ala Gly Phe Thr Cys Asp Thr Gln Lys Gly Thr Cys Glu
325 330 335

Gln Gly Pro His Gln Val Pro Trp Met Glu Lys Ala Pro Ala His Leu
340 345 350

Ser Leu Pro Asp Pro Gln Ala Leu Lys Arg Asp Val Pro Cys Asp Asn
355 360 365

Val Ser Ser Cys Pro Ser Ser Asp Thr Cys Cys Gln Leu Thr Ser Gly
370 375 380

Glu Trp Gly Cys Cys Pro Ile Pro Glu Ala Val Cys Cys Ser Asp His
385 390 395 400

Gln His Cys Cys Pro Gln Gly Tyr Thr Cys Val Ala Glu Gly Gln Cys
405 410 415

Gln Arg Gly Ser Glu Ile Val Ala Gly Leu Glu Lys Met Pro Ala Arg
420 425 430

Arg Ala Ser Leu Ser His Pro Arg Asp Ile Gly Cys Asp Gln His Thr
435 440 445

Ser Cys Pro Val Gly Gln Thr Cys Cys Pro Ser Leu Gly Gly Ser Trp
450 455 460

Ala Cys Cys Gln Leu Pro His Ala Val Cys Cys Glu Asp Arg Gln His
465 470 475 480

Cys Cys Pro Ala Gly Tyr Thr Cys Asn Val Lys Ala Arg Ser Cys Glu
485 490 495

Lys Glu Val Val Ser Ala Gln Pro Ala Thr Phe Leu Ala Arg Ser Pro
500 505 510

His Val Gly Val Lys Asp Val Glu Cys Gly Glu Gly His Phe Cys His
515 520 525

Asp Asn Gln Thr Cys Cys Arg Asp Asn Arg Gln Gly Trp Ala Cys Cys
530 535 540

Pro Tyr Arg Gln Gly Val Cys Cys Ala Asp Arg Arg His Cys Cys Pro
545 550 555 560

Ala Gly Phe Arg Cys Ala Ala Arg Gly Thr Lys Cys Leu Arg Arg Glu
565 570 575

Ala Pro Arg Trp Asp Ala Pro Leu Arg Asp Pro Ala Leu Arg Gln Leu
580 585 590

Leu

<210> 82

<400> 82
000

<210> 83

<400> 83
000

<210> 84

<400> 84
000

<210> 85

<211> 141

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 85

cctgcaggca gctgcgcgct cgctcgctca ctgaggccgc ccgggcaaag cccgggcgtc 60

gggcgacctt tggtcgcccg gcctcagtga gcgagcgagc gcgcagagag ggagtggcca 120

actccatcac taggggttcc t 141

<210> 86

<211> 483

<212> DNA

<213> Homo sapiens

<400> 86

aaccgagtat ctgcagaggg ccctgcgtat gaggcgcaagt gggttttagg accaggatga 60

ggcgggggtgg gggtgcctac ctgacgaccg accccgaccc actggacaag cacccaaccc 120

ccattcccca aattgcgcat cccctatcag agaggggggag gggaaacagg atgcggcgag 180

gcgcgtgcgc actgccagct tcagcaccgc ggacagtgcc ttcgcccccg cctggcggcg	240
cgcgccaccg ccgcctcagc actgaaggcg cgctgacgtc actcgccgggt ccccccga	300
ctcccccttcc cggccacctt ggtcgcgtcc gcgccgccgc cggcccagcc ggaccgcacc	360
acgcgaggcg cgagataggg gggcacgggc gcgaccatct gcgctgcggc gccggcgact	420
cagcgctgcc tcagtctgcg gtgggcagcg gaggagtcgt gtcgtgcctg agagcgagc	480
tgt	483

<210> 87

<211> 1779

<212> DNA

<213> Homo sapiens

<400> 87

tggaccctgg tgagctgggt ggccttaaca gcagggtgg tggctggaac gcggtgcca	60
gatggtcagt tctgccctgt ggcctgctgc ctggaccccg gaggagccag ctacagctgc	120
tgccgtcccc ttctggacaa atggcccaca aactgagca ggcatctggg tggcccctgc	180
caggttgatg cccactgctc tgccggccac tcctgcatct ttaccgtctc agggacttcc	240
agttgctgcc cttcccaga ggccgtggca tgcggggatg gccatcactg ctgcccacgg	300
ggcttccact gcagtgcaga cgggcgatcc tgcttccaaa gatcaggtaa caactccgtg	360
ggtgccatcc agtgccctga tagtcagttc gaatgcccgg acttctccac gtgctgtgtt	420
atggtcgatg gtcctggggg gtgctgcccc atgccccagg cttcctgctg tgaagacagg	480
gtgactgct gtccgcacgg tgccttctgc gacctggttc acaccgctg catcacacc	540
acgggcaccc accccctggc aaagaagctc cctgcccaga ggactaacag ggagtggcc	600
ttgtccagct cggcatgtg tccggacgca cgggtcccgt gccctgatgg ttctacctgc	660
tgtgagctgc ccagtgggaa gtatggctgc tgcccaatgc ccaacgccac ctgctgctcc	720
gatcacctgc actgctgccc ccaagacact gtgtgtgacc tgatccagag taagtgcctc	780
tccaaggaga acgctaccac ggacctctc actaagctgc ctgcgcacac agtgggggat	840
gtgaaatgtg acatggaggt gagctgcccga gatggctata cctgctgccg tctacagtgc	900
ggggcctggg gctgctgccc ttttaccag gctgtgtgct gtgaggacca cataactgc	960
tgtcccgcg gggtttacgtg tgacacgcag aagggtacct gtgaacaggg gccccaccag	1020
gtgccctgga tggagaaggc cccagctcac ctcagcctgc cagaccaca agccttgaag	1080

agagatgtcc cctgtgataa tgtcagcagc tgtccctcct ccgatacctg ctgccaactc	1140
acgtctgggg agtggggctg ctgtccaatc ccagaggctg tctgctgctc ggaccaccag	1200
cactgctgcc cccagggcta cacgtgtgta gctgaggggc agtgtcagcg aggaagcgag	1260
atcgtggctg gactggagaa gatgcctgcc cgccgggctt ctttatcca cccagagac	1320
atcggctgtg accagcacac cagctgcccg gtggggcaga cctgctgccc gagcctgggt	1380
gggagctggg cctgctgcca gttgccccat gctgtgtgct gcgaggatcg ccagcactgc	1440
tgcccggtg gctacacctg caacgtgaag gctcgatcct gcgagaagga agtggctctt	1500
gcccagcctg ccaccttcct ggcccgtagc cctcacgtgg gtgtgaagga cgtggagtgt	1560
ggggaaggac acttctgcca tgataaccag acctgctgcc gagacaaccg acagggtgg	1620
gcctgctgtc cctaccgcca gggcgtctgt tgtgtgatc ggcgccactg ctgtcctgct	1680
ggcttccgct gcgcagccag gggtaccaag tgtttgcgca gggaggcccc gcgctgggac	1740
gcccctttga gggaccagc cttgagacag ctgctgtag	1779

<210> 88

<211> 721

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 88

aacaggccta ttgattggaa agtatgtcaa cgaattgtgg gtcttttggg gtttgctgcc	60
ccttttacgc aatgtggata tcctgcttta atgcctttat atgcatgtat acaagcaaaa	120
caggctttta ctttctcgcc aacttacaag gcctttctaa gtaaacagta tctgaccctt	180
taccccgttg ctcggaacg gcctggtctg tgccaagtgt ttgctgacgc aacccccact	240
ggttggggct tggccatagg ccatcagcgc atgcgtggaa cctttgtgtc tcctctgccg	300
atccatactg cggaactcct agccgcttgt tttgctcgca gcaggctctgg agcgaaactc	360
atcgggactg acaattctgt cgtgctctcc cgcaagtata catcgtttcc agggctgcta	420
ggctgtgctg ccaactggat cctgcgcggg acgtcctttg tttacgtccc gtcggcgctg	480
aatcccgcgg acgaccctc ccggggccgc ttggggctct accgcccgt tctccgtctg	540

ccgtaccgac cgaccacggg ggcacacctt ctttacgcgg actccccgtc tgtgccttct	600
catctgccgg accgtgtgca cttcgcttca cctctgcacg tcgcatggag accaccgtga	660
acgcccaccg gaacctgccc aaggtcttgc ataagaggac tcttggactt tcagcaatgt	720
c	721

<210> 89
 <211> 479
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 89	
acgggtggca tccctgtgac ccctccccag tgcctctcct ggccctggaa gttgccactc	60
cagtgccac cagccttgtc ctaataaaat taagttgcat cattttgtct gactaggtgt	120
ccttctataa tattatgggg tggagggggg tggatatggag caaggggcaa gttgggaaga	180
caacctgtag ggcctgcggg gtctattggg aaccaagctg gaggcagtg gcacaatctt	240
ggctcactgc aatctccgcc tcctgggttc aagcgattct cctgcctcag cctcccaggt	300
tgttgggatt ccaggcatgc atgaccaggc tcagctaatt tttgtttttt tggtagagac	360
ggggtttcac catattggcc aggctgggtc ccaactccta atctcaggtg atctaccac	420
cttggcctcc caaattgctg ggattacagg cgtgaaccac tgctcccttc cctgtcctt	479

<210> 90
 <211> 141
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 90	
aggaaccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg	60
ccgggcgacc aaaggtcgcc cgacggcgg gctttgccc ggcgccctca gtgagcgagc	120
gagcgcgag ctgcctgcag g	141

<210> 91
<211> 4958
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 91
cctgcaggca gctgcgcgct cgctcgctca ctgaggccgc ccgggcaaag cccgggcgtc 60
gggcgacctt tggtcgcccg gcctcagtga gcgagcgagc gcgcagagag ggagtggcca 120
actccatcac taggggttcc tgcggccgca cgcgtaaccg agtatctgca gagggccctg 180
cgtatgagtg caagtgggtt ttaggaccag gatgaggcgg ggtgggggtg cctacctgac 240
gaccgacccc gaccactgg acaagcacc aacccccatt ccccaaattg cgcattcccct 300
atcagagagg gggaggggaa acaggatgcg gcgaggcgcg tgcgactgc cagcttcagc 360
accgcgga gtccttcgc cccgccttg cggcgcgcgc caccgccgcc tcagcactga 420
aggcgcgctg acgtcactcg ccggtcccc gcaaactccc cttcccggcc accttggtcg 480
cgtccgcgcc gccgccggcc cagccggacc gcaccacgcg aggcgcgaga taggggggca 540
cgggcgcgac catctgcgct gcggcgccgg cgactcagcg ctgcctcagt ctgcggtggg 600
cagcggagga gtcgtgtcgt gcctgagagc gcagctgtga attccttcca gaggagattg 660
gaaaacttga agaagaagt gattgtgcta atattgccct gaaagcagcc accatggatt 720
gggagagttg gaaacaaat ttgcaaattg atatcaagtt agcatttaca gatttggtcg 780
aggagaatat ccattatatt gaacaggtaa ttagtgttgt ttgatattgc ttcattttaa 840
agttatttgc tcatttactt ttggtccgtc cattgttgaa agagtgtatt aaagaacaag 900
tgtcacattc tattgcctct ctggtagctt ggttttgttg aagttgtcag ttaccatttg 960
gttttgttta tcctcagttt gttgttttgg atttggattc ttcaaaagca tttgatattg 1020
ctttctattg attgtcctaa ctactcctct ttctctccc ttctccattt ttgaagagtt 1080
tgcaaaggaa ggaaaggagc agagacttga ttgagcagaa aatcatttca gggcctgttc 1140
tctattgtcc ttgctatcct gtcttctgta gctatctgaa accatcaaca aaggagcaca 1200
ccattccatc agcaaaagag taacaacatc tttttttaag ttcattttgt ttttcagttg 1260

attgtatttc aattttttta cagctgactt ttctcagaga agtttttttt ttattgtaaa	1320
catacttttt ctagaaagta tatttttaaaa taacatcttt aaccttatct ctggctgaat	1380
tattgaatat ttgaaattat tacattaaca aaattttgtc ttacagcagt ggtccccaac	1440
cttcttagca gtagcatccc tcattaagaa ttaaaatttg tagaaattga caaggattct	1500
gacaagctgt tgggagagaa gaatagagca gattgcagta ggaacagttg tgttagaatt	1560
tattaatcct ttaacactga aagtaaacta ttgttgattg cctcttggtg tgtttccatt	1620
attcagtgc cttgctaagt gggagtcatt ccttacatca accaccaacc ttcacttgga	1680
agaagctagc gaagataaac ctcgcaaccg ccgcggcagc ggcgaaggcc gcggcagcct	1740
gctgacctgc ggcatgttg aagaaaacc gggcccgtgg accctgggtga gctgggtggc	1800
cttaacagca gggctggtgg ctggaacgcg gtgccagat ggtcagttct gccctgtggc	1860
ctgctgcctg gaccccgag gagccagcta cagctgctgc cgtccccttc tggacaaatg	1920
gcccacaaca ctgagcaggc atctgggtgg cccctgccag gttgatgccc actgctctgc	1980
cggccactcc tgcactttta ccgtctcagg gacttccagt tgctgcccct tcccagaggc	2040
cgtggcatgc ggggatggcc atcactgctg cccacggggc ttccactgca gtgcagacgg	2100
gcgatcctgc ttccaaagat caggtaacaa ctccgtgggt gccatccagt gccctgatag	2160
tcagttcgaa tgcccggact tctccacgtg ctgtgttatg gtcgatggct cctgggggtg	2220
ctgccccatg ccccaggctt cctgctgtga agacagggtg cactgctgtc cgcacgggtg	2280
cttctgcgac ctggttcaca cccgctgcat cacaccacg ggcaccacc ccctggcaaa	2340
gaagctccct gccagagga ctaacagggc agtggccttg tccagctcgg tcatgtgtcc	2400
ggacgcacgg tcccgggtgc ctgatggttc tacctgctgt gagctgcca gtgggaagta	2460
tggctgctgc ccaatgcca acgccacctg ctgctccgat cacctgcact gctgccccca	2520
agacactgtg tgtgacctga tccagagtaa gtgcctctcc aaggagaacg ctaccacgga	2580
cctcctcact aagctgcctg cgcacacagt gggggatgtg aaatgtgaca tggaggtgag	2640
ctgcccagat ggctatacct gctgccgtct acagtcgggg gcctggggct gctgcccttt	2700
taccaggct gtgtgctgtg aggaccacat aactgctgt cccgcgggggt ttacgtgtga	2760
cacgcagaag ggtacctgtg aacaggggcc ccaccagggt ccctggatgg agaaggcccc	2820
agctcacctc agcctgccag acccacaagc cttgaagaga gatgtcccct gtgataatgt	2880

cagcagctgt	ccctcctccg	atacctgctg	ccaactcacg	tctggggagt	ggggctgctg	2940
tccaatccca	gaggctgtct	gctgctcgga	ccaccagcac	tgctgcccc	agggctacac	3000
gtgtgtagct	gaggggcagt	gtcagcgagg	aagcgagatc	gtggctggac	tggagaagat	3060
gcctgccccg	cgggcttcct	tatcccaccc	cagagacatc	ggctgtgacc	agcacaccag	3120
ctgcccggtg	gggcagacct	gctgcccag	cctgggtggg	agctgggcct	gctgccagtt	3180
gccccatgct	gtgtgctgcg	aggatcgcca	gcactgctgc	ccggctggct	acacctgcaa	3240
cgtgaaggct	cgatcctgcg	agaaggaagt	ggtctctgcc	cagcctgcca	ccttcctggc	3300
ccgtagccct	cacgtgggtg	tgaaggacgt	ggagtgtggg	gaaggacact	tctgccatga	3360
taaccagacc	tgctgccgag	acaaccgaca	gggctgggcc	tgctgtccct	accgccaggg	3420
cgtctgttgt	gctgatcggc	gccactgctg	tcctgctggc	ttccgctgcg	cagccagggg	3480
taccaagtgt	ttgcgcaggg	aggccccgcg	ctgggacgcc	cctttgaggg	accagcctt	3540
gagacagctg	ctgtaggtcg	actaaacagg	cctattgatt	ggaaagtatg	tcaacgaatt	3600
gtgggtcttt	tggggtttgc	tgcccccttt	acgcaatgtg	gatatcctgc	tttaatgcct	3660
ttatatgcat	gtatacaagc	aaaacaggct	tttactttct	cgccaactta	caaggccttt	3720
ctaagtaaac	agtatctgac	cctttacccc	gttgctcggc	aacggcctgg	tctgtgccaa	3780
gtgtttgctg	acgcaacccc	cactggtttg	ggcttgGCCA	taggccatca	gcgcatgcgt	3840
ggaacctttg	tgtctcctct	gccgatccat	actgcggaac	tcctagccgc	ttgttttgct	3900
cgcagcaggt	ctggagcgaa	actcatcggg	actgacaatt	ctgtcgtgct	ctcccgcaag	3960
tatacatcgt	ttccagggtc	gctaggctgt	gctgccaaact	ggatcctgcg	cgggacgtcc	4020
tttgtttacg	tcccgtcggc	gctgaatccc	gcggacgacc	cctcccgggg	ccgcttgggg	4080
ctctaccgcc	cgcttctccg	tctgccgtac	cgaccgacca	cggggcgcac	ctctctttac	4140
gcggactccc	cgtctgtgcc	ttctcatctg	ccggaccgtg	tgcacttcgc	ttcacctctg	4200
cacgtcgcac	ggagaccacc	gtgaacgccc	accggaacct	gcccagggtc	ttgcataaga	4260
ggactcttgg	actttcagca	atgtcaactc	gagagatcta	cgggtggcat	ccctgtgacc	4320
cctccccagt	gcctctcctg	gccctggaag	ttgccactcc	agtgcccacc	agccttgtcc	4380
taataaaaatt	aagttgcatc	attttgtctg	actaggtgtc	cttctataat	attatggggt	4440
ggaggggggt	ggtatggagc	aaggggcaag	ttgggaagac	aacctgtagg	gcctgcgggg	4500

tctattggga accaagctgg agtgcagtgg cacaatcttg gctcactgca atctccgcct	4560
cctgggttca agcgattctc ctgcctcagc ctcccagatt gttgggattc caggcatgca	4620
tgaccaggct cagctaattt ttgttttttt ggtagagacg gggtttcacc atattggcca	4680
ggctggtctc caactcctaa tctcaggtga tctaccacc ttggcctccc aaattgctgg	4740
gattacaggc gtgaaccact gtcctcttcc ctgtccttct gatattttag gtaaccacgt	4800
gcggaccgag cggccgcagg aaccctagtg gatggagttg gccactccct ctctgcgcgc	4860
tcgctcgtc actgaggccg ggcgacaaa ggctgcccga cggccgggct ttgcccgggc	4920
ggcctcagtg agcgagcgag cgcgcagctg cctgcagg	4958

<210> 92

<211> 9220

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 92

cctgcaggca gctgcgcgtc cgctcgtca ctgaggccgc ccgggcaaag cccgggcgtc	60
gggcgacctt tggtcgcccg gcctcagtga gcgagcgagc gcgcagagag ggagtggcca	120
actccatcac taggggttcc tgcggccgca cgcgtaaccg agtatctgca gagggccctg	180
cgtatgagtg caagtgggtt ttaggaccag gatgaggcgg ggtgggggtg cctacctgac	240
gaccgacccc gaccactgg acaagcacc aacccccatt ccccaaattg cgcacctcct	300
atcagagagg gggaggggaa acaggatgcg gcgaggcgcg tgcgactgc cagcttcagc	360
accgcggaca gtgccttcgc cccgccttg cggcgcgcgc caccgccgcc tcagcactga	420
aggcgcgctg acgtcactcg ccggtcccc gcaaactccc cttcccggcc accttggtcg	480
cgtccgcgcc gccgccggcc cagccggacc gcaccacgcg aggcgcgaga taggggggca	540
cgggcgcgac catctgcgtc gcggcgccgg cgactcagcg ctgcctcagt ctgcggtggg	600
cagcggagga gtcgtgtcgt gcctgagagc gcagctgtga attccttcca gaggagattg	660
gaaaacttga agaagaagtg gattgtgcta atattgccct gaaagcagcc accatggatt	720
gggagagttg gaaacaaaat ttgcaaattg atatcaagtt agcatttaca gatttggctg	780

aggagaatat ccattatntt gaacaggtaa ttagtgttgt ttgatattgc ttcattnttaa	840
agttatnttg tcattntactt ttggtccgtc cattgttgaa agagtgtatt aaagaacaag	900
tgtcacattc tattgcctct ctggtagctt ggtnttggtg aagttgtcag ttaccatttg	960
gnttgntta tcctcagntt gttgnttggt atttggttc ttcaaaagca tttgatattg	1020
ctntctattg attgtcctaa ctactcctct ttcctctccc ttctccattt ttgaagagtt	1080
tgcaaaggaa ggaaaggagc agagacttga ttgagcagaa aatcattntca gggcctgttc	1140
tctattgtcc ttgctatcct gtcttctgta gctatctgaa accatcaaca aaggagcaca	1200
ccattccatc agcaaaagag taacaacatc tntntntaag ttcattnttg tnttcagttg	1260
attgtatntc aatntntnta cagctgactt ttctcagaga agntntntnt ttattgtaaa	1320
catactntnt ctagaaagta tntntntaaa taacatctnt aaccttatct ctggctgaat	1380
tattgaatat ttgaaattat tacattaaca aaatnttgtc ttacagcagt ggtccccaac	1440
cttcttagca gtagcatccc tcattaagaa ttaaaatttg tagaaattga caaggattct	1500
gacaagctgt tgggagagaa gaatagagca gattgcagta ggaacagttg tgttagaatt	1560
tattaatcct ttaacactga aagtaacta ttgttgattg cctcttggtg tgtntccatt	1620
attcagtgc cttgctaagt gggagtcatt cttacatca accaccaacc ttcacttgga	1680
agaagctagc gaagataaac ctgcgaaccg ccgcggcagc ggcgaaggcc gcggcagcct	1740
gctgacctgc ggcgatgtgg aagaaaacc gggcccgtgg accctggtga gctgggtggc	1800
cttaacagca gggctggtgg ctggaacgcg gtgcccagat ggtcagttct gccctgtggc	1860
ctgctgcctg gaccccgag gagccagcta cagctgctgc cgtcccctc tggacaaatg	1920
gcccacaaca ctgagcaggc atctgggtgg cccctgccag gttgatgccc actgctctgc	1980
cggccactcc tgcatctnta ccgtctcagg gacttccagt tgctgcccct tcccagaggc	2040
cgtggcatgc ggggatggcc atcactgctg cccacggggc ttccactgca gtgcagacgg	2100
gcgatcctgc ttcaaagat caggtaacaa ctccgtgggt gccatccagt gccctgatag	2160
tcagttcgaa tgcccgact tctccacgtg ctgtgttatg gtcgatggct cctgggggtg	2220
ctgccccatg cccaggctt cctgctgtga agacagggtg cactgctgtc cgcacgggtc	2280
cttctgcgac ctggttcaca cccgctgcat cacaccacg ggcaccacc ccctggcaaa	2340
gaagctccct gccagagga ctaacagggc agtggccttg tccagctcgg tcatgtgtcc	2400

ggacgcacgg	tcccgggtgcc	ctgatggttc	tacctgctgt	gagctgcca	gtgggaagta	2460
tggctgctgc	ccaatgcca	acgccacctg	ctgctccgat	cacctgcact	gctgccccca	2520
agacactgtg	tgtgacctga	tccagagtaa	gtgcctctcc	aaggagaacg	ctaccacgga	2580
cctcctcact	aagctgcctg	cgcacacagt	gggggatgtg	aaatgtgaca	tggaggtgag	2640
ctgcccagat	ggctatacct	gctgccgtct	acagtcgggg	gcctggggct	gctgcccttt	2700
tacccaggct	gtgtgctgtg	aggaccacat	acactgctgt	cccgcgggg	ttacgtgtga	2760
cacgcagaag	ggtacctgtg	aacaggggcc	ccaccagggtg	ccctggatgg	agaaggcccc	2820
agctcacctc	agcctgccag	accacacaagc	cttgaagaga	gatgtcccct	gtgataatgt	2880
cagcagctgt	ccctcctccg	atacctgctg	ccaactcacg	tctggggagt	ggggctgctg	2940
tccaatcca	gaggctgtct	gctgctcgga	ccaccagcac	tgctgcccc	agggtacac	3000
gtgtgtagct	gaggggcagt	gtcagcgagg	aagcgagatc	gtggctggac	tggagaagat	3060
gcctgcccgc	cgggcttcct	tatcccaccc	cagagacatc	ggctgtgacc	agcacaccag	3120
ctgcccggtg	gggcagacct	gctgcccag	cctgggtggg	agctgggcct	gctgccagtt	3180
gccccatgct	gtgtgctgcg	aggatcgcca	gcactgctgc	ccggctggct	acacctgcaa	3240
cgtgaaggct	cgatcctgcg	agaaggaagt	ggtctctgcc	cagcctgcca	ccttcctggc	3300
ccgtagccct	cacgtgggtg	tgaaggacgt	ggagtgtggg	gaaggacact	tctgccatga	3360
taaccagacc	tgctgccgag	acaaccgaca	gggctgggcc	tgctgtccct	accgccaggg	3420
cgtctgttgt	gctgatcggc	gccactgctg	tcctgctggc	ttccgctgcg	cagccagggg	3480
taccaagtgt	ttgcgcaggg	aggccccgcg	ctgggacgcc	cctttgaggg	accagcctt	3540
gagacagctg	ctgtaggtcg	actaaacagg	cctattgatt	ggaaagtatg	tcaacgaatt	3600
gtgggtcttt	tggggtttgc	tgcccccttt	acgcaatgtg	gatatcctgc	tttaatgcct	3660
ttatatgcat	gtatacaagc	aaaacaggct	tttactttct	cgccaactta	caaggccttt	3720
ctaagtaa	agtatctgac	cctttacccc	gttgctcggc	aacggcctgg	tctgtgccaa	3780
gtgtttgctg	acgcaacccc	cactggttgg	ggcttgcca	taggccatca	gcgcatgcgt	3840
ggaacctttg	tgtctcctct	gccgatccat	actgcggaac	tcctagccgc	ttgttttgct	3900
cgcagcaggt	ctggagcgaa	actcatcggg	actgacaatt	ctgtcgtgct	ctcccgcaag	3960
tatacatcgt	ttccagggt	gctaggctgt	gctgccaact	ggatcctgcg	cgggacgtcc	4020

tttgtttacg tcccgtcggc gctgaatccc gcggacgacc cctcccgggg ccgcttgggg	4080
ctctaccgcc cgcttctccg tctgccgtac cgaccgacca cggggcgcac ctctctttac	4140
gcggactccc cgtctgtgcc ttctcatctg ccggaccgtg tgcacttcgc ttcacctctg	4200
cacgtcgcac ggagaccacc gtgaacgccc accggaacct gcccaaggct ttgcataaga	4260
ggactcttgg actttcagca atgtcaactc gagagatcta cgggtggcat ccctgtgacc	4320
cctccccagt gcctctcctg gccctggaag ttgccactcc agtgcccacc agccttgtcc	4380
taataaaatt aagttgcatc attttgtctg actaggtgtc cttctataat attatgggggt	4440
ggaggggggt ggtatggagc aaggggcaag ttgggaagac aacctgtagg gcctgcgggg	4500
tctattggga accaagctgg agtgcagtgg cacaatcttg gctcactgca atctccgcct	4560
cctgggttca agcgattctc ctgcctcagc ctcccagatt gttgggattc caggcatgca	4620
tgaccaggct cagctaattt ttgttttttt ggtagagacg gggtttcacc atattggcca	4680
ggctggtctc caactcctaa tctcaggtga tctaccacc ttggcctccc aaattgctgg	4740
gattacaggc gtgaaccact gctcccttcc ctgtccttct gatattttag gtaaccacgt	4800
gcggaccgag cggccgcagg aaccctagt gatggagttg gccactccct ctctgcgcgc	4860
tcgctcgctc actgaggccg ggcgaccaa ggctgcccga cggccgggct ttgcccgggc	4920
ggcctcagt agcgagcgag cgcgcagctg cctgcaggaa gctgtaagct ctaggagatc	4980
cgaaccagat aagtgaatc tagttccaaa ctattttgtc atttttaatt ttcgtattag	5040
cttacgacgc tacaccagc tcccatctat tttgtcactc ttccctaaat aatcctaaa	5100
aactccattt ccaccctcc cagttcccaa ctattttgtc cgcccacagc ggggcatttt	5160
tcttcctgtt atgtttttta tcaaacatcc tgccaactcc atgtgacaaa ccgtcatctt	5220
cggctacttt ttctctgtca cagaatgaaa atttttctgt catctcttcg ttattaatgt	5280
ttgtaattga ctgaatatca acgcttattt gcagcctgaa tggcgaatgg acgcgccctg	5340
tagcggcgca ttaagcgcgg cgggtgtggt ggttacgcgc agcgtgaccg ctacacttgc	5400
cagcgcccta gcgcccgtc ctttcgcttt cttcccttcc tttctcgcca cgttcgccgg	5460
ctttccccgt caagctctaa atcgggggct ccctttagggt ttccgattta gtgctttacg	5520
gcacctcgac cccaaaaaac ttgattagggt tgatggttca cgtagtgggc catcgccctg	5580
atagacggtt tttcgccctt tgacgttggg gtccacgttc tttaatagtg gactcttgtt	5640

ccaaactgga acaacactca accctatctc ggtctattct tttgatttat aagggatttt	5700
gccgatttcg gcctattgggt taaaaaatga gctgatttaa caaaaattta acgcgaattt	5760
taacaaaata ttaacgttta caatttcagg tggcactttt cggggaaatg tgcgcggaac	5820
ccctatttgt ttatTTTTtct aaatacattc aaatatgtat ccgctcatga gacaataacc	5880
ctgataaatg cttcaataat attgaaaaag gaagagtatg agtattcaac atttccgtgt	5940
cgcccttatt ccctTTTTtTg cggcattttg ctttcctgtt tttgctcacc cagaaacgct	6000
ggtgaaagta aaagatgctg aagatcagtt ggggtgcacga gtgggttaca tcgaactgga	6060
tctcaacagc ggtaagatcc ttgagagttt tcgccccgaa gaacgttttc caatgatgag	6120
cacttttaaa gttctgctat gtggcgcggt attatcccggt attgacgccg ggcaagagca	6180
actcggtcgc cgcatacact attctcagaa tgacttggtt gagtactcac cagtcacaga	6240
aaagcatctt acggatggca tgacagtaag agaattatgc agtgctgcca taaccatgag	6300
tgataacact gcggccaact tactttctgac aacgatcgga ggaccgaagg agctaaccgc	6360
TTTTTgcac aacatggggg atcatgtaac tcgccttgat cgttgggaac cggagctgaa	6420
tgaagccata ccaaacgacg agcgtgacac cacgatgcct gtagcaatgg caacaacgtt	6480
gcgcaaacta ttaactggcg aactacttac tctagcttcc cggcaacaat taatagactg	6540
gatggaggcg gataaagttg caggaccact tctgcgctcg gcccttccgg ctggctgggt	6600
tattgctgat aaatctggag ccggtgagcg tgggtctcgc ggtatcattg cagcactggg	6660
gccagatgggt aagccctccc gtatcgtagt tatctacacg acggggagtc aggcaactat	6720
ggatgaacga aatagacaga tcgctgagat aggtgcctca ctgattaagc attggtaact	6780
gtcagaccaa gtttactcat atatacttta gattgattta aaacttcatt tttaatTTaa	6840
aaggatctag gtgaagatcc tttttgataa tctcatgacc aaaatccctt aacgtgagtt	6900
ttcgttcac tgagcgtcag accccgtaga aaagatcaaa ggatcttctt gagatccttt	6960
TTTTctgcgc gtaatctgct gcttgcaaac aaaaaaacca ccgctaccag cggtggtttg	7020
tttgccggat caagagctac caactctttt tccgaaggta actggcttca gcagagcgca	7080
gataccaaat actgtccttc tagtgtagcc gtagttaggc caccacttca agaactctgt	7140
agcaccgcct acatacctcg ctctgctaata cctgttacca gtggctgctg ccagtggcga	7200
taagtcgtgt cttaccgggt tggactcaag acgatagtta ccggataagg cgcagcggtc	7260

gggctgaacg gggggttcgt gcacacagcc cagcttggag cgaacgacct acaccgaact	7320
gagataccta cagcgtgagc attgagaaaag cgccacgctt cccgaaggga gaaaggcgga	7380
caggatatccg gtaagcggca gggctcggaac aggagagcgc acgagggagc ttccaggggg	7440
aaacgcctgg tatctttata gtcctgtcgg gtttcgccac ctctgacttg agcgtcgatt	7500
tttgtgatgc tcgtcagggg ggcggagcct atggaaaaac gccagcaacg cggccttttt	7560
acggttcctg gccttttgct ggccttttgc tcacatgttc tttcctgcgt tatcccctga	7620
ttctgtggat aaccgtatta ccgcctttga gtgagctgat accgctcgcc gcagccgaac	7680
gaccgagcgc agcgagtcag tgagcgagga agcggaagag cgcctgatgc ggtattttct	7740
ccttacgcat ctgtgcggta tttcacaccg cagaccagcc gcgtaacctg gcaaaatcgg	7800
ttacggttga gtaataaatg gatgccctgc gtaagcgggt gtgggcggac aataaagtct	7860
taaactgaac aaaatagatc taaactatga caataaagtc ttaaactaga cagaatagtt	7920
gtaaactgaa atcagtccag ttatgctgtg aaaaagcata ctggactttt gttatggcta	7980
aagcaaactc ttcattttct gaagtgcaaa ttgcccgtcg tattaagag gggcgtggcc	8040
aagggcattg taaagactat attcgcggcg ttgtgacaat ttaccgaaca actccgcggc	8100
cgggaagccg atctcggctt gaacgaattg ttaggtggcg gtacttgggt cgatatcaaa	8160
gtgcatcact tcttcccgtg tgcccaactt tgtatagaga gccactgcgg gatcgtcacc	8220
gtaatctgct tgcacgtaga tcacataagc accaagcgcg ttggcctcat gcttgaggag	8280
attgatgagc gcggtggcaa tgccctgcct ccggtgctcg ccggagactg cgagatcata	8340
gatatagatc tctactacgc gctgctcaaa cctgggcaga acgtaagccg cgagagcgcc	8400
aacaaccgct tcttggtcga aggcagcaag cgcgatgaat gtcttactac ggagcaagtt	8460
cccgaggtaa tcggagtccg gctgatgttg ggagtaggtg gctacgtctc cgaactcacg	8520
accgaaaaga tcaagagcag ccgcgatgga tttgacttgg tcagggccga gcctacatgt	8580
gcgaatgatg ccataacttg agccacctaa ctttgtttta gggcgactgc cctgctgcgt	8640
aacatcgttg ctgctgcgta acatcgttgc tgctccataa catcaaacaat cgaccacagg	8700
cgtaacgcgc ttgctgcttg gatgcccag gcatagactg tacaaaaaaa cagtcataac	8760
aagccatgaa aaccgccact gcgccgttac caccgctgcg ttcgggtcaag gttctggacc	8820
agttgcgtga gcgcatacgc tacttgcatt acagttttacg aaccgaacag gcttatgtca	8880

actgggttcg tgccttcacg cgtttccacg gtgtgcgtca cccggcaacc ttgggcagca	8940
gcgaagtcga ggcatttctg tcctggctgg cgaacgagcg caaggtttcg gtctccacgc	9000
atcgtcaggc attggcggcc ttgctgttct tctacggcaa ggtgctgtgc acggatctgc	9060
cctggcttca ggagatcgga agacctcggc cgtcgcggcg cttgccggtg gtgctgaccc	9120
cggatgaagt ggttcgcacg ctcggttttc tggaaggcga gcatcgtttg ttcgcccagg	9180
actctagcta tagttctagt ggttggctac agcttgcatg	9220

<210> 93

<211> 5954

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 93

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc cccggcgctcg ggcgaccttt	60
ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cgggcggagt	120
tagggcggag ccaatcagcg tgcgccgttc cgaaagtgc cttttatggc tgggcggaga	180
atgggcgggtg aacgccgatg attatataag gacgcgccgg gtgtggcaca gctagttccg	240
tcgcagccgg gatttgggtc gcggttcttg tttgtggatc cctgtgatcg tcacttgaca	300
ccggtcttcc agaggagatt ggaaaacttg aagaagaagt ggattgtgct aatattgccc	360
tgaaagcagc caccatggat tgggagagtt ggaaacaaaa tttgcaaatt gatatcaagt	420
tagcatttac agatttggct gaggagaata tccattatit tgaacaggta attagtgttg	480
tttgatattg cttcatttta aagttatttg ctcatittact tttgggtccgt ccattgttga	540
aagagtgtat taaagaacaa gtgtcacatt ctattgcctc tctggtagct tggttttgtt	600
gaagtgttca gttaccattt ggttttgttt atcctcagtt tgttgttttg gatttggatt	660
cttcaaaagc atttgatatt gctttctatt gattgtccta actactcctc tttcctctcc	720
cttctccatt tttgaagagt ttgcaaagga aggaaaggag cagagacttg attgagcaga	780
aaatcatttc agggcctgtt ctctattgtc cttgctatcc tgtcttctgt agctatctga	840
aaccatcaac aaaggagcac accattccat cagcaaaaga gtaacaacat ctttttttaa	900

gttcattttg tttttcagtt gattgtattt caattttttt acagctgact tttctcagag	960
aagttttttt tttattgtaa acatactttt tctagaaagt atatttttaa ataacatctt	1020
taaccttatc tctggctgaa ttattgaata tttgaaatta ttacattaac aaaattttgt	1080
cttacagcag tggtcccaa ctttcttagc agtagcatcc ctcatthaaga attaaaattt	1140
gtagaaattg acaaggattc tgacaagctg ttgggagaga agaatagagc agattgcagt	1200
aggaacagtt gtgttagaat ttattaatcc ttttaactg aaagtaaact attgttgatt	1260
gcctcttggt gtgtttccat tattcagtg ctttgctaag tgggagtcac tccttacatc	1320
aaccaccaac cttcacttg aagaagctag cgaagataaa cctcgcaacc gccgcggcag	1380
cggcgaaggc cgcggcagcc tgctgacctg cggcgatgtg gaagaaaacc cgggcccggg	1440
gagcaagggc gaggagctgt tcaccggggg ggtgcccac ctggtcgagc tggacggcga	1500
cgtaaacggc cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa	1560
gctgaccctg aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt	1620
gaccaccctg acctacggcg tgcagtgtt cagccgtac ccgaccaca tgaagcagca	1680
cgacttcttc aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa	1740
ggacgacggc aactacaaga ccgcgccga ggtgaagttc gagggcgaca ccctggtgaa	1800
ccgcatcgag ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct	1860
ggagtacaac tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat	1920
caaggtgaac ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca	1980
ctaccagcag aacaccccca tcggcgacgg ccccgctgtg ctgcccgaca accactacct	2040
gagcaccag tccgccctga gcaaagacc caacgagaag cgcgatcaca tggctctgct	2100
ggagtctgtg accgccgccg ggatcactct cggcatggac gagctgtaca agtaatgctt	2160
tatttgtaga atttgatg ctattgctt atttgtaacc attataagct gcaataaaca	2220
agttaacaac aacaattgca ttcattttat gtttcagggt cagggggagg tgtgggagg	2280
tttttaaaat cgatctgagg aaccctagt gatggagttg gccactcccc ccccccccg	2340
gcgattctct tgtttgctcc agactctcag gcaatgacct gatagccttt gtagagacct	2400
ctcaaaaata gctaccctct ccggcatgaa tttatcagct agaacggttg aatatcatat	2460
tgatggtgat ttgactgtct ccggcctttc tcaccggtt gaatctttac ctacacatta	2520

ctcaggcatt gcatttaaaa tatatgaggg ttctaaaaat ttttatcctt gcgttgaaat	2580
aaaggcttct cccgcaaaag tattacaggg tcataatggt tttggtacaa ccgatttagc	2640
tttatgctct gaggctttat tgcttaattt tgctaattct ttgccttgcc tgtatgattt	2700
attggatggt ggaatcgcct gatgcggtat tttctcctta cgcatctgtg cggatattca	2760
caccgcatat ggtgcactct cagtacaatc tgctctgatg ccgcatagtt aagccagccc	2820
cgacacccgc caacactatg gtgcactctc agtacaatct gctctgatgc cgcatagtta	2880
agccagcccc gacacccgcc aacacccgct gacgcgccct gacgggcttg tctgctcccg	2940
gcatccgctt acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca	3000
ccgtcatcac cgaaacgcgc gagacgaaag ggcctcgtga tacgcctatt tttatagggt	3060
aatgtcatga taataatggt ttcttagacg tcagggtggca cttttcgggg aaatgtgcgc	3120
ggaacccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa	3180
taaccctgat aaatgcttca ataatttga aaaaggaaga gtatgagcca tattcaacgg	3240
gaaacgtcga ggccgcgatt aaattccaac atggatgctg atttatatgg gtataaatgg	3300
gctcgcgata atgtcgggca atcagggtgcg acaatctatc gcttgtatgg gaagcccgat	3360
gcgccagagt tgtttctgaa acatggcaaa ggtagcgttg ccaatgatgt tacagatgag	3420
atggtcagac taaactggct gacggaattt atgccacttc cgaccatcaa gcattttatc	3480
cgtactcctg atgatgcatg gttactcacc actgcgatcc ccggaaaaac agcgttccag	3540
gtattagaag aatatcctga ttcagggtgaa aatattgttg atgcgctggc agtgttcctg	3600
cgccggttgc actcgattcc tgtttgtaat tgtcctttta acagcgatcg cgtatttcgc	3660
ctcgctcagg cgcaatcacg aatgaataac ggtttggttg atgcgagtga tttgatgac	3720
gagcgtaatg gctggcctgt tgaacaagtc tggaaagaaa tgcataaact tttgccattc	3780
tcaccggatt cagtcgtcac tcatggtgat ttctcacttg ataaccttat tttgacgag	3840
gggaaattaa taggttgtat tgatgttgga cgagtcggaa tcgcagaccg ataccaggat	3900
cttgccatcc tatggaactg cctcggtgag ttttctcctt cattacagaa acggcttttt	3960
caaaaatatg gtattgataa tcctgatatg aataaattgc agtttcattt gatgctcgat	4020
gagtttttct aactgtcaga ccaagtttac tcatatatac tttagattga tttaaaactt	4080
catttttaat ttaaaaggat ctaggtgaag atcctttttg ataatctcat gacccaaatc	4140

ccttaacgtg agtttttcgtt ccactgagcg tcagacccccg tagaaaagat caaaggatct	4200
tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta	4260
ccagcgggtgg tttgtttgcc ggatcaagag ctaccaactc tttttccgaa ggtaactggc	4320
ttcagcagag cgcagatacc aaatactggtt cttctagtgt agccgtagtt aggccaccac	4380
ttcaagaact ctgtagcacc gcctacatac ctcgctctgc taatcctggt accagtggct	4440
gctgccagtg gcgataagtc gtgtcttacc gggttggact caagacgata gttaccggat	4500
aaggcgcagc ggtcgggctg aacgggggggt tcgtgcacac agcccagctt ggagcgaacg	4560
acctacaccg aactgagata cctacagcgt gagctatgag aaagcgccac gcttcccgaa	4620
gggagaaagg cggacaggta tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg	4680
gagcttccag ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg ccacctctga	4740
cttgagcgtc gatTTTTgtg atgctcgtca ggggggcgga gcctatggaa aaacgccagc	4800
aacgcggcct ttttacggtt cctggccttt tgctggcctt ttgctcacat gttctttcct	4860
gcgttatccc ctgattctgt ggataaccgt attaccgcct ttgagtgagc tgataccgct	4920
cgccgcagcc gaacgaccga gcgcagcgag tcagtgagcg aggaagcgga agagcgccca	4980
atacgcaaac cgcctctccc cgcgcgttgg ccgattcatt aatgcagctg gcgtaatagc	5040
gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatggcga	5100
ttccgttgca atggctggcg gtaatatgtt tctggatatt accagcaagg ccgatagttt	5160
gagttcttct actcaggcaa gtgatgttat tactaatcaa agaagtattg cgacaacggt	5220
taatttgcgt gatggacaga ctcttttact cggtggcctc actgattata aaaacacttc	5280
tcaggattct ggcgtaccgt tcctgtctaa aatcccttta atcggcctcc tgtttagctc	5340
ccgctctgat tctaacgagg aaagcacgtt atacgtgctc gtcaaagcaa ccatagtacg	5400
cgccctgtag cggcgcatta agcgcggcgg gtgtgggtgg tactgcgcagc gtgaccgcta	5460
cacttgccag cgccctagcg cccgctcctt tcgctttctt cccttccttt ctcgccacgt	5520
tcgccggctt tccccgtcaa gctctaaatc gggggctccc tttagggttc cgatttagtg	5580
ctttacggca cctcgacccc aaaaaacttg attagggtga tggttcacgt agtgggccat	5640
cgccctgata gacggttttt cgccctttga cgttggagtc cacgttcttt aatagtggac	5700
tcttgttcca aactggaaca acactcaacc ctatctcggg ctattctttt gatttataag	5760

ggatTTTgCC gatttcggcc tattggTTaa aaaatgagct gatttaacaa aaatttaacg	5820
cgaattTTta caaaatatta acgcttaca tTTaaatatt tgcttataca atcttcctgt	5880
TTTTggggct tttctgatta tcaaccgggg tacatatgat tgacatgcta gttttacgat	5940
taccgttcat cgcc	5954

<210> 94
 <211> 799
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 94	
gaagaagaag atatcaagtt agcatttaca gatttggctg aggagaagaa caggtaatta	60
gtgttgtttg atattgcttc atTTtaaagt tatttgctca tttagcattt gatattgctt	120
tctattgatt gtcctaacta ctctctttc ctctccctc tccatttttg aagagtttgc	180
aaaggaagga aaggagcaga gacttgattg agcagaaaat catttcaggg cctgttctct	240
attgtccttg ctatcctgtc ttctgtagct atctgaaacc atcaacaaag gagcacacca	300
tggcatcagc aaaagagtaa caacatcttt tTTaaagttc atTTtgTTTT tcagttgatt	360
gtatttcaat ttttttacag ctgacttttc tcagagaagt ttttttttta ttgtaaacat	420
actttttcta gaaagtatat tTTaaaataa catctTTaac cttatctctg gctgaattat	480
tgaatatttg aaattattac attaacaaaa ttttgtctta cagcagtggg cccaacctt	540
cttagcagta gcatccctca ttaagaatta aaattttag aaattgacaa ggattctgac	600
aagctgttgg gagagaagaa tagagcagat tgcagtagga acagttgtgt tagaatttat	660
taatccttta aactgaaag taaactattg ttgattgcct cttggtgtgt ttccattatt	720
cagtgtcttt gctaagtggg agtcattcct tacatcaacc accaaccttc acttggaaga	780
agctagcgaa gataaacct	799

<210> 95
 <211> 5446
 <212> DNA
 <213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 95

gccaatacg caaacgcct ctccccgcgc gttggccgat tcattaatgc agctgattct	60
aacgaggaaa gcacgttata cgtgctcgtc aaagcaacca tagtacgcgc cctgtagcgg	120
cgcattaagc gcggcgggtg tgggtggttac gcgcagcgtg accgctacac ttgccagcgc	180
cctagcgcgc gtccttttcg ctttcttccc ttcctttctc gccacgttcg ccggctttcc	240
ccgtcaagct ctaaatcggg ggctcccttt agggttccga tttagtgtt tacggcacct	300
cgaccccaa aaacttgatt agggatgatg ttcacgtagt gggccatcgc cctgatagac	360
ggtttttcgc cttttgacgt tggagtccac gttctttaat agtggactct tgttccaaac	420
tggaacaaca ctcaacccta tctcggctca ttcttttgat ttataaggga ttttgccgat	480
ttcggcctat tggttaaaaa atgagctgat ttaacaaaaa tttaacgcga attttaacaa	540
aatattaacg cttacaattt aaatatattgc ttatacaatc ttcctgtttt tggggctttt	600
ctgattatca accgggggtac atatgattga catgctagtt ttacgattac cgttcacgc	660
cctgcgcgct cgctcgtca ctgaggccgc ccgggcaaag cccgggcgtc gggcgacctt	720
tggtcgcccc gcctcagtga gcgagcgagc gcgcagagag ggagtggaat tcaattcggg	780
cggagttagg gcggagccaa tcagcgtgcg ccgttccgaa agttgccttt tatggctggg	840
cggagaatgg gcggtgaacg ccgatgatta tataaggacg cgccgggtgt ggcacagcta	900
gttccgtcgc agccgggatt tgggtcgcgg ttcttgtttg tggatccctg tgatcgtcac	960
ttgacaccgg tgaagaagaa gatataagat tagcatttac agatttggct gaggagaaga	1020
acaggtaatt agtggtggtt gatattgctt ctttttaaag ttatttgctc atttagcatt	1080
tgatattgct ttctattgat tgtcctaact actcctcttt cctctccctt ctccattttt	1140
gaagagtttg caaaggaagg aaaggagcag agacttgatt gagcagaaaa tcatttcagg	1200
gcctgttctc tattgtcctt gctatcctgt cttctgtagc tatctgaaac catcaacaaa	1260
ggagcacacc atggcatcag caaaagagta acaacatctt tttttaagtt cttttgttt	1320
ttcagttgat tgtatttcaa tttttttaca gctgactttt ctcagagaag tttttttttt	1380
attgtaaaca tactttttct agaaagtata ttttaaaata acatctttaa ccttatctct	1440

ggctgaatta ttgaatattt gaaattatta cattaacaaa attttgtctt acagcagtgg	1500
tccccaacct tcttagcagt agcatccctc attaagaatt aaaatttgta gaaattgaca	1560
aggattctga caagctgttg ggagagaaga atagagcaga ttgcagtagg aacagttgtg	1620
ttagaattta ttaatccttt aacactgaaa gtaaactatt gttgattgcc tcttggtgtg	1680
tttccattat tcagtgtctt tgctaagtgg gagtcattcc ttacatcaac caccaacctt	1740
cacttggaag aagctagcga agataaacct cgcaaccgcc gcggcagcgg cgaaggccgc	1800
ggcagcctgc tgacctgcgg cgatgtggaa gaaaaccgg gcccggtctt cacactcgaa	1860
gatttcgttg gggactggcg acagacagcc ggctacaacc tggaccaagt ccttgaacag	1920
ggaggtgtgt ccagtttgtt tcagaatctc ggggtgtccg taactccgat ccaaaggatt	1980
gtcctgagcg gtgaaaatgg gctgaagatc gacatccatg tcatcatccc gtatgaaggt	2040
ctgagcggcg accaaatggg ccagatcgaa aaaattttta aggtggtgta ccctgtggat	2100
gatcatcact ttaaggtgat cctgcactat ggcacactgg taatcgacgg ggttacgccg	2160
aacatgatcg actatttcgg acggccgtat gaaggcatcg ccgtgttcga cggcaaaaag	2220
atcactgtaa cagggaccct gtggaacggc aaaaaatta tcgacgagcg cctgatcaac	2280
cccgacggct ccctgctgtt ccgagtaacc atcaacggag tgaccggctg gcggctgtgc	2340
gaacgcattc tggcgtaacc tgcccgttg gcctcccaac gggccctcct cccctccttg	2400
caccaagctt atcgataccg tcgactagag ctgctgatc agcctcgact gtgccttcta	2460
gttgccagcc atctgttggt tgcccctccc ccgtgccttc cttgaccctg gaaggtgcca	2520
ctcccactgt cctttcctaa taaaatgagg aaattgcatc gcattgtctg agtaggtgtc	2580
attctattct ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata	2640
gcaggcatgc tggggagaga tcgatctgag gaaccctag tgatggagtt ggccactccc	2700
tctctgcgcg ctgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc	2760
tttgcccggg cggcctcagt gagcgagcga gcgcgcagag agggagtggc ccccccccc	2820
ccccccccgg cgattctctt gtttgctcca gactctcagg caatgacctg atagcctttg	2880
tagagacctc tcaaaaatag ctaccctctc cggcatgaat ttatcagcta gaacggttga	2940
atatcatatt gatggtgatt tgactgtctc cggcctttct caccgtttg aatctttacc	3000
tacacattac tcaggcattg catttaaaat atatgagggt tctaaaaatt tttatccttg	3060

cgttgaaata aaggcttctc ccgcaaaagt attacagggt cataatgttt ttggtacaac	3120
cgatttagct ttatgctctg aggctttatt gcttaatfff gctaattctt tgccttgcct	3180
gtatgattta ttggatgttg gaatcgctg atgcgggtatt ttctccttac gcatctgtgc	3240
ggatattcac accgcatatg gtgcactctc agtacaatct gctctgatgc cgcatagtta	3300
agccagcccc gacacccgcc aacacccgct gacgcgccct gacgggcttg tctgctcccg	3360
gcatccgctt acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca	3420
ccgtcatcac cgaaacgcgc gagacgaaag ggcctcgtga tacgcctatt tttataggtt	3480
aatgtcatga taataatggt ttcttagacg tcagggtggca cttttcgggg aaatgtgcgc	3540
ggaacccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa	3600
taaccctgat aaatgcttca ataatatga aaaaggaaga gtatgagtat tcaacatttc	3660
cgtgtcgcgc ttattccctt ttttgcggca ttttgccttc ctgtttttgc tcaccagaa	3720
acgctggatga aagtaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa	3780
ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg	3840
atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtattga cgccgggcaa	3900
gagcaactcg gtcgccgcat acactattct cagaatgact tggttgagta ctcaccagtc	3960
acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc	4020
atgagtata acactgcggc caacttactt ctgacaacga tcggaggacc gaaggagcta	4080
accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag	4140
ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgcctgtagc aatggcaaca	4200
acgttgcgca aactattaac tggcgaacta cttactctag cttcccggca acaattaata	4260
gactggatgg aggcgataa agttgcagga ccacttctgc gctcggccct tccggctggc	4320
tggtttattg ctgataaatc tggagccggt gagcgtgggt ctcgcggtat cattgcagca	4380
ctggggccag atggtaagcc ctcccgtatc gtagttatct acacgacggg gtagcaggca	4440
actatggatg aacgaaatag acagatcgct gagatagggt cctcactgat taagcattgg	4500
taactgtcag accaagttta ctcatatata ctttagattg atttaaaact tcatttttaa	4560
tttaaaagga tctaggtgaa gatccttttt gataatctca tgaccaaaat cccttaacgt	4620
gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat	4680

cctttttttc tgcgcgtaat ctgctgcttg caaacaacaaa aaccaccgct accagcggtg	4740
gtttgtttgc cggatcaaga gctaccaact ctttttccga aggtaactgg cttcagcaga	4800
gcgagatac caaatactgt ccttctagtg tagccgtagt taggccacca cttcaagaac	4860
tctgtagcac cgcctacata cctcgtcttg ctaatcctgt taccagtggc tgctgccagt	4920
ggcgataagt cgtgtcttac cgggttggac tcaagacgat agttaccgga taaggcgag	4980
cggtcgggct gaacggggggg ttcgtgcaca cagcccagct tggagcgaac gacctacacc	5040
gaactgagat acctacagcg tgagctatga gaaagcgcca cgcttcccga agggagaaaag	5100
gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag ggagcttcca	5160
gggggaaacg cctggtatct ttatagtcct gtcgggtttc gccacctctg acttgagcgt	5220
cgatTTTTgt gatgctcgtc aggggggagg agcctatgga aaaacgccag caacgcggcc	5280
tttttacggt tcctggcctt ttgctggcct tttgctcaca tgttctttcc tgcgttatcc	5340
cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc tcgccgcagc	5400
cgaacgaccg agcgagcga gtcagtgagc gaggaagcgg aagagc	5446

<210> 96
 <211> 53
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 oligonucleotide"

<400> 96	
gaagaagaag atatcaagtt agcatttaca gatttggctg aggagaagaa cag	53

<210> 97
 <211> 120
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> source
 <223> /note="Description of Artificial Sequence: Synthetic
 polynucleotide"

<400> 97	
gtaattagtg ttgtttgata ttgcttcatt ttaaagttat ttgctcattt agcatttgat	60

attgctttct attgattgtc ctaactactc ctctttcctc tcccttctcc atttttgaag 120

<210> 98

<211> 143

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 98

agtttgcaaa ggaaggaaag gagcagagac ttgattgagc agaaaatcat ttcagggcct 60

gttctctatt gtccttgcta tcctgtcttc tgtagctatc tgaaaccatc aacaaaggag 120

cacaccatgg catcagcaaa aga 143

<210> 99

<211> 153

<212> DNA

<213> Homo sapiens

<400> 99

cttcagagg agattggaaa acttgaagat aaagtggaat gtgctaataa tgccttgaag 60

gcagattggg agagatggaa acaaaatatg caaaatgata tcaagttagc atttacagat 120

atggctgagg agaatatcca ttattatgaa cag 153

<210> 100

<211> 78

<212> DNA

<213> Homo sapiens

<400> 100

tgccttgcta cgtgggagtc attccttaca tcacagacca accttcactt ggaagaagcc 60

tctgaagata aaccttaa 78

<210> 101

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic

oligonucleotide"

<220>
<221> misc_feature
<222> (1)..(18)
<223> /note="This sequence may encompass 1-6 'gaa' repeating units"

<400> 101
gaagaagaag aagaagaa 18

<210> 102
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic peptide"

<400> 102
Cys Cys Cys Cys Cys Cys Cys Cys Cys Cys Cys Cys
1 5 10

<210> 103
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 103
caacaaagga gcacaccatt c 21

<210> 104
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic primer"

<400> 104
gcggttgcga ggtttatct 19

<210> 105
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 105
ctcttgctaa gtgggagtca tt 22

<210> 106
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 106
gtgctaataa tgccctgaaa gc 22

<210> 107
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
primer"

<400> 107
ccacttagca agagcactgt 20

<210> 108
<211> 34
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 108

Met Asp Trp Glu Ser Trp Lys Gln Asn Leu Gln Ile Asp Ile Lys Leu
1 5 10 15

Ala Phe Thr Asp Leu Ala Glu Glu Asn Ile His Tyr Phe Glu Gln Cys
20 25 30

Ser Cys

<210> 109

<211> 149

<212> PRT

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 109

Met Asp Trp Glu Ser Trp Lys Gln Asn Leu Gln Ile Asp Ile Lys Ser
1 5 10 15

Asp His Leu Gln Ile Trp Leu Arg Arg Ile Ser Ile Ile Leu Asn Ser
20 25 30

Ala Leu Ala Lys Trp Glu Ser Phe Leu Thr Ser Thr Thr Asn Leu His
35 40 45

Leu Glu Glu Ala Ser Glu Asp Lys Pro Arg Asn Arg Arg Gly Ser Gly
50 55 60

Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro
65 70 75 80

Gly Pro Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile
85 90 95

Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser
100 105 110

Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe
115 120 125

Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr
130 135 140

Thr Leu Thr Tyr Gly
145

<210> 110
<211> 175
<212> PRT
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 110
Met Asp Trp Glu Ser Trp Lys Gln Asn Leu Gln Ile Asp Ile Lys Leu
1 5 10 15

Ala Phe Thr Asp Leu Ala Glu Glu Asn Ile His Tyr Phe Glu Gln Ser
20 25 30

Leu Gln Arg Lys Glu Arg Ser Arg Asp Leu Ile Glu Gln Lys Ile Ile
35 40 45

Ser Gly Pro Val Leu Tyr Cys Pro Cys Tyr Pro Val Phe Cys Ser Tyr
50 55 60

Leu Lys Pro Ser Thr Lys Glu His Thr Ile Pro Ser Ala Lys Asp Ala
65 70 75 80

Leu Ala Lys Trp Glu Ser Phe Leu Thr Ser Thr Thr Asn Leu His Leu
85 90 95

Glu Glu Ala Ser Glu Asp Lys Pro Arg Asn Arg Arg Gly Ser Gly Glu
100 105 110

Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro Gly
115 120 125

Pro Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
130 135 140

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
145 150 155 160

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe
165 170 175

<210> 111

<211> 42

<212> PRT

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 111

Met Asp Trp Glu Ser Trp Lys Gln Asn Leu Gln Ile Asp Ile Lys Ser
1 5 10 15

Asp His Leu Gln Ile Trp Leu Arg Arg Ile Ser Ile Ile Leu Asn Arg
20 25 30

Val Cys Lys Gly Arg Lys Gly Ala Glu Thr
35 40

<210> 112

<211> 142

<212> PRT

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 112

Met Asp Trp Glu Ser Trp Lys Gln Asn Leu Gln Ile Asp Ile Lys Leu
1 5 10 15

Ala Phe Thr Asp Leu Ala Glu Glu Asn Ile His Tyr Phe Glu Gln Ser
20 25 30

Leu Gln Arg Lys Glu Arg Ser Arg Asp Leu Ile Glu Gln Lys Ile Ile
35 40 45

Ser Gly Pro Val Leu Tyr Cys Pro Cys Tyr Pro Val Phe Cys Ser Tyr
50 55 60

Leu Lys Pro Ser Thr Lys Glu His Thr Ile Pro Ser Ala Lys Asp Ala
65 70 75 80

Leu Ala Lys Trp Glu Ser Phe Leu Thr Ser Thr Thr Asn Leu His Leu
85 90 95

Glu Glu Ala Ser Glu Asp Lys Pro Arg Asn Arg Arg Gly Ser Gly Glu
100 105 110

Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro Gly
115 120 125

Pro Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro
130 135 140

<210> 113

<211> 69

<212> PRT

<213> Artificial Sequence

<220>

<221> source

<223> /note="Description of Artificial Sequence: Synthetic
polypeptide"

<400> 113

Met Ala Ser Ala Lys Asp Ala Leu Ala Lys Trp Glu Ser Phe Leu Thr
1 5 10 15

Ser Thr Thr Asn Leu His Leu Glu Glu Ala Ser Glu Asp Lys Pro Arg
20 25 30

Asn Arg Arg Gly Ser Gly Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly
35 40 45

Asp Val Glu Glu Asn Pro Gly Pro Val Ser Lys Gly Glu Glu Leu Phe
50 55 60

Thr Gly Val Val Pro

