



(12) **DEMANDE DE BREVET CANADIEN**
CANADIAN PATENT APPLICATION

(13) **A1**

(86) **Date de dépôt PCT/PCT Filing Date:** 2022/10/05
(87) **Date publication PCT/PCT Publication Date:** 2023/04/13
(85) **Entrée phase nationale/National Entry:** 2024/03/27
(86) **N° demande PCT/PCT Application No.:** US 2022/077586
(87) **N° publication PCT/PCT Publication No.:** 2023/060112
(30) **Priorité/Priority:** 2021/10/05 (US63/252,520)

(51) **Cl.Int./Int.Cl.** **A61K 35/76** (2015.01),
C12N 15/63 (2006.01), **C12N 15/67** (2006.01)
(71) **Demandeur/Applicant:**
ALLEN INSTITUTE, US
(72) **Inventeurs/Inventors:**
DAIGLE, TANYA, US;
LEIN, EDWARD SEBASTIAN, US;
LEVI, BOAZ P., US;
TASIC, BOSILJKA, US;
TING, JONATHAN, US;
ZENG, HONGKUI, US;
MICH, JOHN K., US
(74) **Agent:** GOWLING WLG (CANADA) LLP

(54) **Titre : CONSTRUCTIONS D'EXPRESSION ARTIFICIELLES POUR MODULER L'EXPRESSION GENIQUE DANS LE CERVELET ET DANS UN TYPE DE CELLULE SECONDAIRE**
(54) **Title: ARTIFICIAL EXPRESSION CONSTRUCTS FOR MODULATING GENE EXPRESSION IN THE CEREBELLUM AND A SECONDARY CELL TYPE**

FIG. 1A



(57) **Abrégé/Abstract:**

Artificial expression constructs for modulating gene expression in targeted central nervous system cell types are described. The artificial expression constructs can be used to express synthetic genes or modify gene expression in a cerebellar cell type and a secondary cell type.



Date Submitted: 2024/03/27

CA App. No.: 3233342

Abstract:

Artificial expression constructs for modulating gene expression in targeted central nervous system cell types are described. The artificial expression constructs can be used to express synthetic genes or modify gene expression in a cerebellar cell type and a secondary cell type.

ARTIFICIAL EXPRESSION CONSTRUCTS FOR MODULATING GENE EXPRESSION IN THE CEREBELLUM AND A SECONDARY CELL TYPE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 63/252,520, filed October 5, 2021, which is incorporated herein by reference in its entirety as if fully set forth herein.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under MH114126 and MH120095 awarded by the National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO SEQUENCE LISTING

[0003] The Sequence Listing associated with this application is provided in XML format in lieu of a paper copy and is hereby incorporated by reference into the specification. The name of the XML file containing the Sequence Listing is A166-0031PCT.xml. The XML file is 291 KB, was created on October 4, 2022, and is being submitted electronically via Patent Center.

FIELD OF THE DISCLOSURE

[0004] The current disclosure provides artificial expression constructs for modulating gene expression in the cerebellum and a secondary cell type. Gene expression in the cerebellum can be modulated in Purkinje cells, granule cells, Bergmann glia, deep cerebellar nucleus (DCN) cells, molecular layer interneurons (MLI), mossy fiber (MF) cells, foliar white matter (WM), or oligodendrocytes. Gene expression in the secondary cell type can be modulated in astrocytes, oligodendrocytes, or neuronal cells (brain-wide); glutamatergic neurons in the thalamus; GABAergic neurons, glutamatergic neurons, astrocytes, pericytes, or specialized smooth-muscle cells (SMCs) in the neocortex; cholinergic interneurons or medium spiny neurons (MSN) in the striatum; or spinal motor neurons.

BACKGROUND OF THE DISCLOSURE

[0005] To fully understand the biology of the brain, different cell types need to be distinguished and defined and, to further study them, artificial expression constructs that can label and perturb them need to be identified. In mouse, recombinase driver lines have been used to great effect to label cell populations that share marker gene expression. However, the creation, maintenance, and use of such lines that label cell types with high specificity can be costly, frequently requiring triple transgenic crosses, which yield a low frequency of experimental animals. Furthermore, those tools require germline transgenic animals and thus are not applicable to humans.

[0006] Previously, artificial expression constructs that drive gene expression in astrocytes or oligodendrocytes (brain-wide), glutamatergic neurons in the thalamus or neocortex, GABAergic neurons in the neocortex, or cholinergic interneurons in the striatum have been described.

SUMMARY OF THE DISCLOSURE

[0007] The current disclosure provides artificial expression constructs that drive gene expression in the cerebellum and a secondary cell type. Gene expression in the cerebellum can be modulated in Purkinje cells, granule cells, Bergmann glia, deep cerebellar nucleus (DCN) cells, molecular layer interneurons (MLI), mossy fiber (MF) cells, foliar white matter (WM), or oligodendrocytes. Gene expression in the secondary cell type can be modulated in astrocytes, oligodendrocytes, or neuronal cells (brain-wide); glutamatergic neurons in the thalamus; GABAergic neurons (e.g., Pvalb (e.g., chandelier), SST (e.g., Chodl), LAMP5, VIP, medial ganglionic eminence (MGE))), glutamatergic neurons (e.g., L5 ET, L5 NP, L5 IT), astrocytes, pericytes, or specialized smooth-muscle cells (SMCs) in the neocortex; cholinergic interneurons or medium spiny neurons (MSN) in the striatum; or spinal motor neurons.

[0008] Particular embodiments of the artificial expression constructs utilize the following enhancers to drive gene expression within a cerebellar cell type and a secondary cell type as follows (enhancer / cerebellar cell type and secondary cell type):

eHGT_023h v1 / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_023m / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_082h / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_087h / Purkinje cells in the cerebellum and Sst interneurons in the neocortex;

eHGT_128h / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_181h / MLI cells in the cerebellum and Lamp5 interneurons in the neocortex;

eHGT_260h / Purkinje cells in the cerebellum and pan-glutamatergic neurons in the neocortex;

eHGT_023h v2 / Purkinje and MLI cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_356h / DCN cells in the cerebellum and Vip interneurons in the neocortex;

eHGT_359h / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_375m / Bergmann glia in the cerebellum and astrocytes brain wide;

eHGT_387m / Bergmann glia in the cerebellum and astrocytes brain wide;

eHGT_395h / oligodendrocytes, DCN, and foliar white matter (WM) in the cerebellum and oligodendrocytes brain wide;

eHGT_453m / DCN cells in the cerebellum and L5 ET neurons in the neocortex;

eHGT_470m / DCN cells in the cerebellum and Sst/Chodl neurons in the neocortex;
eHGT_479m / Purkinje and MLI cells in the cerebellum and chandelier cells in the neocortex;
eHGT_494m / Purkinje cells in the cerebellum and Vip and chandelier cells in the neocortex;
eHGT_467m / Purkinje cells in the cerebellum and Sst/Chodl neurons in the neocortex;
eHGT_483m / Granule and mossy fiber (MF) cells in the cerebellum and Vip interneurons in the neocortex;
eHGT_606h / Purkinje cells in the cerebellum and glutamatergic neurons in the thalamus;
eHGT_738m / MLI cells in the cerebellum and cholinergic interneurons in the striatum;
eHGT_796h / Purkinje cells in the cerebellum and pan-GABAergic neurons in the neocortex;
eHGT_710m / MLI cells in the cerebellum and chandelier cells in the neocortex;
eHGT_588m / Purkinje and MLI cells in the cerebellum and MGE cells in the neocortex;
eHGT_007m / MLI cells in the cerebellum and Vip interneurons in the neocortex;
eHGT_703m / MLI cells in the cerebellum and chandelier cells in the neocortex;
eHGT_589m / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;
eHGT_086h / MLI (basket) cells in the cerebellum and Pvalb interneurons in the neocortex;
eHGT_963m / MLI and DCN cells in the cerebellum and L5 NP neurons in the neocortex;
eHGT_534h / Purkinje cells in the cerebellum and pan-neuronal cells brain wide;
eHGT_830h / MLI cells in the cerebellum and glutamatergic neurons in the thalamus;
eHGT_540h / Purkinje cells in the cerebellum and pan-neuronal cells brain wide;
eHGT_882m / DCN cells in the cerebellum and MSN cells in the striatum;
eHGT_1137m / DCN cells in the cerebellum and spinal motor neurons;
eHGT_381h / Bergmann glia in the cerebellum and astrocytes in the neocortex;
MGT_E118 / DCN cells in the cerebellum and astrocytes in the neocortex;
MGT_E122 / Bergmann glia in the cerebellum and astrocytes in the neocortex;
MGT_E146 / MLI cells in the cerebellum and pericytes in the neocortex;
MGT_E16 / MLI cells in the cerebellum and L5 IT neurons in the neocortex;
MGT_E150 / Purkinje cells in the cerebellum and specialized smooth-muscle cells (SMCs) in the neocortex;
eHGT_1032h / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
eHGT_1027h / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
eHGT_1027m / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;

concatenated copies of an enhancer or concatenated copies of an enhancer core including 3xCore-eHGT_023h v3 (3xPVALBCore)/ Purkinje and MLI cells in the cerebellum and Pvalb interneurons in the neocortex;

3x(CoreB)eHGT_121h / MLI cells in the cerebellum and Pvalb interneurons in the neocortex;

3xcore3_eHGT_453m / DCN cells in the cerebellum;

3xcore2_eHGT_387m / Bergmann glia in the cerebellum and astrocytes in the neocortex;

3xcore2_eHGT_475m / Purkinje cells in the cerebellum and chandelier cells in the neocortex;

3xcore2_eHGT_351h / DCN cells in the cerebellum and pan-MSN cells in the striatum; and

3xcore_MGT_E116 / DCN cells in the cerebellum and L5 NP neurons in the neocortex.

[0009] In particular embodiments, the artificial enhancer elements include a core or concatenated core of an enhancer. Examples include a concatenated core of eHGT_023h v3, eHGT_121h, eHGT_453m, eHGT_387m, eHGT_475m, eHGT_351h, and/or MGT_E116. These artificial enhancer elements can provide higher levels and more rapid onset of transgene expression compared to a single full length original (native) enhancer.

[0010] In particular embodiments, the enhancer core includes the sequence as set forth in any one of SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 108, SEQ ID NO: 112, or SEQ ID NO: 121. In particular embodiments, these cores are concatenated and have 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core sequence. In particular embodiments, SEQ ID NO: 26 provides a three-copy concatemer of core-eHGT_023h v3 core. In particular embodiments, SEQ ID NO: 28 provides a three-copy concatemer of the (coreB)eHGT_121h. In particular embodiments, SEQ ID NO: 104 provides a three-copy concatemer of the core3_eHGT_453m. In particular embodiments, SEQ ID NO: 106 provides a three-copy concatemer of the core2_eHGT_387m. In particular embodiments, SEQ ID NO: 109 provides a three-copy concatemer of the core2_eHGT_475m. In particular embodiments, SEQ ID NO: 113 provides a three-copy concatemer of the core2_eHGT_351h. In particular embodiments, SEQ ID NO: 122 provides a three-copy concatemer of the core_MGT_E116. In particular embodiments, the enhancer includes an enhancer core and additional sequence.

[0011] Particular embodiments provide artificial expression constructs including the features and/or use of vectors described herein including vectors: CN1259, CN1280, CN1521, CN1528, CN1533, CN1621, CN1674, CN1778, CN1932, CN2043, CN2045, CN2085, CN2102, CN2157, CN2216, CN2251, CN2257, CN2258, CN2267, CN2316, CN2339, CN2431, CN2436, CN2643, CN2663, CN2717, CN2838, CN1415, CN2710, CN2839, CN2674, CN3301, CN3019, CN3569, CN2374, CN3584, CN3003, CN2379, CN3566, CN3453, HCT1, CN2146, AiP1347, AiP1351,

AiP1375, AiP1307, AiP1530, AiP1379, CN3916, CN3869, and/or CN3870.

BRIEF DESCRIPTION OF THE FIGURES

[0012] Some of the drawings submitted herein may be better understood in color. Applicant considers the color versions of the drawings as part of the original submission and reserves the right to present color images of the drawings in later proceedings.

[0013] FIGs. 1A-1C: CN2251 vector drives reporter transgene expression in primate deep cerebellar nucleus neurons. (1A) Epifluorescence image of macaque cerebellum with anti-NeuN antibody staining to reveal structural details and neurons. (1B) Epifluorescence image of anti-GFP signal in deep cerebellar nucleus neurons, the target of the stereotaxic intraparenchymal virus injection of AAV vector CN2251 containing enhancer eHGT_453m packaged with serotype PHP.eB. (1C) Zoom view of the boxed region highlighting anti-GFP signal in neurons in the deep cerebellar nuclei.

[0014] FIGs. 2A-2I. Vector: CN2045 and Enhancer: eHGT_359h. (2A-2E) Animal: mouse 200910-08. (2A) Native SYFP2 fluorescence montage of a sagittal section of a whole mouse brain, (2B) cerebellum, and (2C) higher magnified cerebellum showing targeted expression of SYFP2 in cells with Purkinje cell morphology. (2D) Native SYFP2 fluorescence of cerebellum in sagittal section showing labeling of Purkinje cells, and (2E) their overlap (arrows) with Pvalb mRNA expression. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN2045 virus packaged with the PHP.eB capsid. (2F, 2G) Animal: rat 585761. (2F) Native SYFP2 fluorescence montage of a sagittal section of a whole rat brain and the (2G) expanded view of the cerebellum showing targeted expression of SYFP2 in cells with Purkinje cell morphology. Virus was administered in one day postnatal rat pups by intracerebroventricular (ICV) injection of CN2045 virus packaged with the PHP.eB capsid. (2H, 2I) Animal: monkey Q21.26.022. (2H) Native SYFP2 fluorescence montage of a sagittal section of a macaque cerebellum and the (2I) expanded view of the cerebellum showing targeted expression of SYFP2 in cells with Purkinje cell morphology. Virus was administered by intraparenchymal injection of CN2045 virus packaged with the PHP.eB capsid.

[0015] FIGs. 3A-3I. Vector: CN1259 and Enhancer: eHGT_023h v1. (3A-3C) Animal: mouse 190329-06. (3A) Native SYFP2 fluorescence montage of a sagittal section of a whole mouse brain, (3B) cerebellum, and (3C) higher magnified cerebellum showing targeted expression of SYFP2 in cells with Purkinje cell morphology. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN1259 virus packaged with the PHP.eB capsid. (3D-3I) Mouse 190329-06 cerebellum transduced by CN1259 administered in adult mice by

intravascular (IV) injection (retro-orbital). (3D) Merged whole cerebellum or (3E) higher magnification Purkinje cells are shown. (3F-3I) Individual channels are shown (3F) including SYFP2 fluorescence, (3G) DAPI staining, (3H) Gad1 mFISH and (3I) Pvalb mFISH. Images are montages. Note all SYFP2 cells overlap with Pvalb and Gad1 mRNA, consistent with their identity as Purkinje cells.

[0016] FIGs. 4A-4I. Vector: CN1528, Enhancer: eHGT_082h, and Animal: mouse 190425-01. (4A) Native SYFP2 fluorescence montage of a sagittal section of a whole mouse brain, (4B) cerebellum, and (4C) higher magnified cerebellum showing targeted expression of SYFP2 in cells in the deep cerebellar nucleus. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN1528 virus packaged with the PHP.eB capsid. (4D-4I) Mouse cerebellum transduced by CN1528 administered in adult mice by intravascular (IV) injection (retro-orbital). (4D) Merged deep cerebellar nucleus or (4E) higher magnification image of deep cerebellar nuclear cells are shown. (4F-4I) Individual channels are shown (4F) including SYFP2 fluorescence, (4G) DAPI staining, (4H) Gad1 mFISH and (4I) Pvalb mFISH. Images are montages. Note all SYFP2 cells overlap with Pvalb, while only some express Gad1.

[0017] FIGs. 5A-5C. Vector: CN2102, Enhancer: eHGT_387m, and Animals: mouse 200113-05 (5A), 200910-05 (5B-5C). (5A) Native SYFP2 fluorescence montage of a sagittal section of a whole mouse brain, (5B) cerebellum, and (5C) higher magnified cerebellum showing targeted expression of SYFP2 in cells with Bergmann glial morphology. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN2102 virus packaged with the PHP.eB capsid.

[0018] FIGs. 6A, 6B. Vector: CN2717, Enhancer: eHGT_710m, and Animal: mouse C57BL6J-560070. (6A) Native SYFP2 fluorescence montage of a sagittal section of mouse cerebellum and (6B) expanded view of the cerebellum showing targeted expression of SYFP2 in cells with small interneuron morphology in the molecular layer. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN2717 virus packaged with the PHP.eB capsid.

[0019] FIGs. 7A-7B. Vector: CN2251, Enhancer: eHGT_453m, and Animal: mouse. (7A) Native SYFP2 fluorescence montage of a coronal section of mouse cerebellum and (7B), expanded view of the cerebellum showing targeted expression of SYFP2 in cells in the deep cerebellar nuclei. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN2251 virus packaged with the PHP.eB capsid.

[0020] FIG. 8. Sequences supporting the disclosure. Sequences include: eHGT_023h v1 (SEQ ID NO: 1); eHGT_023m (SEQ ID NO: 2); eHGT_359h (SEQ ID NO: 3); eHGT_023h v2 (SEQ ID NO: 4); eHGT_082h (SEQ ID NO: 5); eHGT_703m (SEQ ID NO: 6); eHGT_087h (SEQ ID NO: 7).

7); eHGT_128h (SEQ ID NO: 98); eHGT_181h (SEQ ID NO: 8); eHGT_260h (SEQ ID NO: 9); eHGT_356h (SEQ ID NO: 10); eHGT_375m (SEQ ID NO: 11); eHGT_387m (SEQ ID NO: 12); eHGT_395h (SEQ ID NO: 99); eHGT_007m (SEQ ID NO: 13); eHGT_453m (SEQ ID NO: 14); eHGT_470m (SEQ ID NO: 15); eHGT_479m (SEQ ID NO: 16); eHGT_494m (SEQ ID NO: 17); eHGT_467m (SEQ ID NO: 18); eHGT_483m (SEQ ID NO: 19); eHGT_606h (SEQ ID NO: 20); eHGT_738m (SEQ ID NO: 21); eHGT_796h (SEQ ID NO: 22); eHGT_710m (SEQ ID NO: 23); eHGT_588m (SEQ ID NO: 24); Core-eHGT_023h v3 (SEQ ID NO: 25); 3xCore-eHGT_023h v3 (SEQ ID NO: 26); (CoreB)eHGT_121h (SEQ ID NO: 27); 3x(CoreB)eHGT_121h (SEQ ID NO: 28); eHGT_589m (SEQ ID NO: 100); eHGT_086h (SEQ ID NO: 101); eHGT_963m (SEQ ID NO: 102); core3_eHGT_453m (SEQ ID NO: 103); 3xcore3_eHGT_453m (SEQ ID NO: 104); core2_eHGT_387m (SEQ ID NO: 105); 3xcore2_eHGT_387m (SEQ ID NO: 106); eHGT_534h (SEQ ID NO: 107); core2_eHGT_475m (SEQ ID NO: 108); 3xcore2_eHGT_475m (SEQ ID NO: 109); eHGT_830h (SEQ ID NO: 110); eHGT_540h (SEQ ID NO: 111); core2_eHGT_351h (SEQ ID NO: 112); 3xcore2_eHGT_351h (SEQ ID NO: 113); eHGT_882m (SEQ ID NO: 114); eHGT_1137m (SEQ ID NO: 115); eHGT_381h (SEQ ID NO: 116); MGT_E118 (SEQ ID NO: 117); MGT_E122 (SEQ ID NO: 118); MGT_E146 (SEQ ID NO: 119); MGT_E16 (SEQ ID NO: 120); core_MGT_E116 (SEQ ID NO: 121); 3xcore_MGT_E116 (SEQ ID NO: 122); MGT_E150 (SEQ ID NO: 123); eHGT_1032h (SEQ ID NO: 124); eHGT_1027h (SEQ ID NO: 125); eHGT_1027m (SEQ ID NO: 126); Beta-Globin Minimal Promoter (SEQ ID NO: 29); minCMV Promoter (SEQ ID NO: 30); Mutated minCMV Promoter (SEQ ID NO: 31); minRho Promoter (SEQ ID NO: 32); minRho* Promoter (SEQ ID NO: 33); Hsp68 minimal Promoter (SEQ ID NO: 34); SYFP2 (SEQ ID NO: 35); EGFP (SEQ ID NO: 36); Optimized Flp recombinase (SEQ ID NO: 37); 10aa (SEQ ID NO: 38); H2Bmod (SEQ ID NO: 39); hsA2 (Seq ID NO: 40); Improved Cre recombinase (SEQ ID NO: 41); SP10 insulator (SEQ ID NO: 42); 3xSP10ins (SEQ ID NO: 43); Flag encoding sequence (SEQ ID NO: 163); 3XFLAG (SEQ ID NO: 164); H2B (SEQ ID NO: 165); H2B* (SEQ ID NO: 166); WPRE3 (SEQ ID NO: 44); WPRE (SEQ ID NO: 45); BGHpA (SEQ ID NO: 46); HGHpA (SEQ ID NO: 47); P2A (SEQ ID NO: 48); T2A (SEQ ID NO: 49); E2A (SEQ ID NO: 50); F2A (SEQ ID NO: 51); Exemplary Plasmid Backbone 1 – Left ITR (SEQ ID NO: 52); Exemplary Plasmid Backbone 1 – Right ITR (SEQ ID NO: 53); Exemplary Plasmid Backbone 2 – Left ITR (SEQ ID NO: 54); Exemplary Plasmid Backbone 2 – Right ITR (SEQ ID NO: 55); PHP.eB capsid (SEQ ID NO: 56); AAV9 VP1 capsid protein (SEQ ID NO: 57); tet-Transactivator version 2 (SEQ ID NO: 58); GTPase HRas [Homo sapiens] (SEQ ID NO: 59); Substance P is position 58-68 of Protachykinin-1 [Homo sapiens] (SEQ ID NO: 60); Oxytocin is position 20-28 of Oxytocin-neurophysin 1 [Homo sapiens] (SEQ ID NO: 61); GCaMP6m (SEQ ID NO: 62); GCaMP6s (SEQ

ID NO: 63); GCaMP6f (SEQ ID NO: 64); CN1259 (SEQ ID NO: 65); CN1621 (SEQ ID NO: 127); CN2157 (SEQ ID NO: 128); CN2339 (SEQ ID NO: 66); CN1280 (SEQ ID NO: 67); CN2045 (SEQ ID NO: 68); CN1932 (SEQ ID NO: 69); CN1528 (SEQ ID NO: 70); CN2710 (SEQ ID NO: 71); CN1533 (SEQ ID NO: 72); CN1674 (SEQ ID NO: 73); CN1778 (SEQ ID NO: 74); CN2043 (SEQ ID NO: 75); CN2085 (SEQ ID NO: 76); CN2102 (SEQ ID NO: 77); CN1415 (SEQ ID NO: 78); CN2251 (SEQ ID NO: 79); CN2257 (SEQ ID NO: 80); CN2258 (SEQ ID NO: 81); CN2267 (SEQ ID NO: 82); CN2316 (SEQ ID NO: 83); CN2431 (SEQ ID NO: 84); CN2436 (SEQ ID NO: 85); CN2643 (SEQ ID NO: 86); CN2663 (SEQ ID NO: 87); CN2717 (SEQ ID NO: 88); CN2838 (SEQ ID NO: 89); CN1521 (SEQ ID NO: 90); CN2216 (SEQ ID NO: 91); CN2839 (SEQ ID NO: 129); CN2674 (SEQ ID NO: 130); CN3301 (SEQ ID NO: 131); CN3019 (SEQ ID NO: 132); CN3569 (SEQ ID NO: 133); CN2374 (SEQ ID NO: 134); CN3584 (SEQ ID NO: 135); CN3003 (SEQ ID NO: 136); CN2379 (SEQ ID NO: 137); CN3566 (SEQ ID NO: 138); CN3453 (SEQ ID NO: 139); HCT1 (SEQ ID NO: 140); CN2146 (SEQ ID NO: 141); AiP1347 (SEQ ID NO: 142); AiP1351 (SEQ ID NO: 143); AiP1375 (SEQ ID NO: 144); AiP1307 (SEQ ID NO: 145); AiP1530 (SEQ ID NO: 146); AiP1379 (SEQ ID NO: 147); CN3916 (SEQ ID NO: 148); CN3869 (SEQ ID NO: 149); and CN3870 (SEQ ID NO: 150).

DETAILED DESCRIPTION

[0021] To fully understand the biology of the brain, different cell types need to be distinguished and defined and, to further study them, artificial expression constructs that can label and perturb them need to be identified (Tasic, *Curr. Opin. Neurobiol.* 50, 242–249 (2018); Zeng & Sanes, *Nat. Rev. Neurosci.* 18, 530–546 (2017)). In mouse, recombinase driver lines have been used to great effect to label cell populations that share marker gene expression (Daigle *et al.*, *Cell* 174, 465–480.e22 (2018); Taniguchi, *et al.*, *Neuron* 71, 995–1013 (2011); Gong *et al.*, *J. Neurosci.* 27, 9817–9823 (2007)). However, the creation, maintenance, and use of such lines that label cell types with high specificity can be costly, frequently requiring triple transgenic crosses, which yield a low frequency of experimental animals. Furthermore, those tools require germline transgenic animals and thus are not applicable to humans.

[0022] The current disclosure provides artificial expression constructs that drive gene expression in the cerebellum and a secondary cell type. Gene expression in the cerebellum can be modulated in Purkinje cells, granule cells, Bergmann glia, deep cerebellar nucleus (DCN) cells, molecular layer interneurons (MLI), mossy fiber (MF) cells, foliar white matter (WM), or oligodendrocytes. Gene expression in the secondary cell type can be modulated in astrocytes, oligodendrocytes, or neuronal cells (brain-wide); glutamatergic neurons in the thalamus; GABAergic neurons (e.g.,

Pvalb (e.g., chandelier), SST (e.g., Chodl), LAMP5, VIP, medial ganglionic eminence (MGE)), glutamatergic neurons (e.g., L5 ET, L5 NP, L5 IT), astrocytes, pericytes, or specialized smooth-muscle cells (SMCs) in the neocortex; cholinergic interneurons or medium spiny neurons (MSN) in the striatum; or spinal motor neurons.

[0023] The current disclosure provides enhancers and associated artificial expression constructs that can drive gene expression in a cerebellar cell type and a secondary cell type as follows:

eHGT_023h v1 / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_023m / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_082h / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_087h / Purkinje cells in the cerebellum and Sst interneurons in the neocortex;

eHGT_128h / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_181h / MLI cells in the cerebellum and Lamp5 interneurons in the neocortex;

eHGT_260h / Purkinje cells in the cerebellum and pan-glutamatergic neurons in the neocortex;

eHGT_023h v2 / Purkinje and MLI cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_356h / DCN cells in the cerebellum and Vip interneurons in the neocortex;

eHGT_359h / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_375m / Bergmann glia in the cerebellum and astrocytes brain wide;

eHGT_387m / Bergmann glia in the cerebellum and astrocytes brain wide;

eHGT_395h / oligodendrocytes, DCN, and foliar white matter (WM) in the cerebellum and oligodendrocytes brain wide;

eHGT_453m / DCN cells in the cerebellum and L5 ET neurons in the neocortex;

eHGT_470m / DCN cells in the cerebellum and Sst/Chodl neurons in the neocortex;

eHGT_479m / Purkinje and MLI cells in the cerebellum and chandelier cells in the neocortex;

eHGT_494m / Purkinje cells in the cerebellum and Vip and chandelier cells in the neocortex;

eHGT_467m / Purkinje cells in the cerebellum and Sst/Chodl neurons in the neocortex;

eHGT_483m / Granule and mossy fiber (MF) cells in the cerebellum and Vip interneurons in the neocortex;

eHGT_606h / Purkinje cells in the cerebellum and glutamatergic neurons in the thalamus;

eHGT_738m / MLI cells in the cerebellum and cholinergic interneurons in the striatum;

eHGT_796h / Purkinje cells in the cerebellum and pan-GABAergic neurons in the neocortex;

eHGT_710m / MLI cells in the cerebellum and chandelier cells in the neocortex;

eHGT_588m / Purkinje and MLI cells in the cerebellum and MGE cells in the neocortex;
 eHGT_007m / MLI cells in the cerebellum and Vip interneurons in the neocortex;
 eHGT_703m / MLI cells in the cerebellum and chandelier cells in the neocortex;
 eHGT_589m / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_086h / MLI (basket) cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_963m / MLI and DCN cells in the cerebellum and L5 NP neurons in the neocortex;
 eHGT_534h / Purkinje cells in the cerebellum and pan-neuronal cells brain wide;
 eHGT_830h / MLI cells in the cerebellum and glutamatergic neurons in the thalamus;
 eHGT_540h / Purkinje cells in the cerebellum and pan-neuronal cells brain wide;
 eHGT_882m / DCN cells in the cerebellum and MSN cells in the striatum;
 eHGT_1137m / DCN cells in the cerebellum and spinal motor neurons;
 eHGT_381h / Bergmann glia in the cerebellum and astrocytes in the neocortex;
 MGT_E118 / DCN cells in the cerebellum and astrocytes in the neocortex;
 MGT_E122 / Bergmann glia in the cerebellum and astrocytes in the neocortex;
 MGT_E146 / MLI cells in the cerebellum and pericytes in the neocortex;
 MGT_E16 / MLI cells in the cerebellum and L5 IT neurons in the neocortex;
 MGT_E150 / Purkinje cells in the cerebellum and specialized smooth-muscle cells (SMCs) in the neocortex;
 eHGT_1032h / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
 eHGT_1027h / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
 eHGT_1027m / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
 3xCore-eHGT_023h v3 / Purkinje and MLI cells in the cerebellum and Pvalb interneurons in the neocortex;
 3x(CoreB)eHGT_121h / MLI cells in the cerebellum and Pvalb interneurons in the neocortex;
 3xcore3_eHGT_453m / DCN cells in the cerebellum;
 3xcore2_eHGT_387m / Bergmann glia in the cerebellum and astrocytes in the neocortex;
 3xcore2_eHGT_475m / Purkinje cells in the cerebellum and chandelier cells in the neocortex;
 3xcore2_eHGT_351h / DCN cells in the cerebellum and pan-MSN cells in the striatum; and
 3xcore_MGT_E116 / DCN cells in the cerebellum and L5 NP neurons in the neocortex.

[0024] In particular embodiments, the artificial enhancer elements include a core or concatenated core of an enhancer. Examples include a concatenated core of eHGT_023h v3, eHGT_121h,

eHGT_453m, eHGT_387m, eHGT_475m, eHGT_351h, and/or MGT_E116. These artificial enhancer elements can provide higher levels and more rapid onset of transgene expression compared to a single full length original (native) enhancer.

[0025] In particular embodiments, the enhancer core includes the sequence as set forth in any one of SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 108, SEQ ID NO: 112, or SEQ ID NO: 121. In particular embodiments, these cores are concatenated and have 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core sequence. In particular embodiments, SEQ ID NO: 26 provides a three-copy concatemer of Core-eHGT_023h v3 core. In particular embodiments, SEQ ID NO: 28 provides a three-copy concatemer of the (coreB)eHGT_121h. In particular embodiments, SEQ ID NO: 104 provides a three-copy concatemer of the core3_eHGT_453m. In particular embodiments, SEQ ID NO: 106 provides a three-copy concatemer of the core2_eHGT_387m. In particular embodiments, SEQ ID NO: 109 provides a three-copy concatemer of the core2_eHGT_475m. In particular embodiments, SEQ ID NO: 113 provides a three-copy concatemer of the core2_eHGT_351h. In particular embodiments, SEQ ID NO: 122 provides a three-copy concatemer of the core_MGT_E116. In particular embodiments, the enhancer includes an enhancer core and additional sequence.

[0026] Particular embodiments provide artificial expression constructs including the features and/or use of vectors described herein including vectors: CN1259, CN1280, CN1521, CN1528, CN1533, CN1621, CN1674, CN1778, CN1932, CN2043, CN2045, CN2085, CN2102, CN2157, CN2216, CN2251, CN2257, CN2258, CN2267, CN2316, CN2339, CN2431, CN2436, CN2643, CN2663, CN2717, CN2838, CN1415, CN2710, CN2839, CN2674, CN3301, CN3019, CN3569, CN2374, CN3584, CN3003, CN2379, CN3566, CN3453, HCT1, CN2146, AiP1347, AiP1351, AiP1375, AiP1307, AiP1530, AiP1379, CN3916, CN3869, and/or CN3870. The vectors can be used to drive gene expression in a cerebellar cell type and a secondary cell type as follows:

CN1259 / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;

CN1280 / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;

CN1521 / Purkinje and MLI cells in the cerebellum and Pvalb interneurons in the neocortex;

CN1528 / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;

CN1533 / Purkinje cells in the cerebellum and Sst interneurons in the neocortex;

CN1621 / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;

CN1674 / MLI cells in the cerebellum and Lamp5 interneurons in the neocortex;

CN1778 / Purkinje cells in the cerebellum and pan-glutamatergic neurons in the neocortex;

CN1932 / Purkinje and MLI cells in the cerebellum and Pvalb interneurons in the neocortex;

CN2043 / DCN cells in the cerebellum and Vip interneurons in the neocortex;
CN2045 / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;
CN2085 / Bergmann glia in the cerebellum and astrocytes brain wide;
CN2102 / Bergmann glia in the cerebellum and astrocytes brain wide;
CN2157 / oligodendrocytes, DCN, and foliar white matter (WM) in the cerebellum and oligodendrocytes brain wide;
CN2216 / MLI cells in the cerebellum and Pvalb interneurons in the neocortex;
CN2251 / DCN cells in the cerebellum and L5 ET neurons in the neocortex;
CN2257 / DCN cells in the cerebellum and Sst/Chodl neurons in the neocortex;
CN2258 / Purkinje and MLI cells in the cerebellum and chandelier cells in the neocortex;
CN2267 / Purkinje cells in the cerebellum and Vip and chandelier cells in the neocortex;
CN2316 / Purkinje cells in the cerebellum and Sst/Chodl neurons in the neocortex;
CN2339 / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;
CN2431 / Granule and mossy fiber (MF) cells in the cerebellum and Vip interneurons in the neocortex;
CN2436 / Purkinje cells in the cerebellum and glutamatergic neurons in the thalamus;
CN2643 / MLI cells in the cerebellum and cholinergic interneurons in the striatum;
CN2663 / Purkinje cells in the cerebellum and pan-GABAergic neurons in the neocortex;
CN2717 / MLI cells in the cerebellum and chandelier cells in the neocortex;
CN2838 / Purkinje and MLI cells in the cerebellum and MGE cells in the neocortex;
CN1415 / MLI cells in the cerebellum and Vip interneurons in the neocortex;
CN2710 / MLI cells in the cerebellum and chandelier cells in the neocortex;
CN2839 / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;
CN2674 / MLI (basket) cells in the cerebellum and Pvalb interneurons in the neocortex;
CN3301 / MLI and DCN cells in the cerebellum and L5 NP neurons in the neocortex;
CN3019 / DCN cells in the cerebellum;
CN3569 / Bergmann glia in the cerebellum and astrocytes in the neocortex;
CN2374 / Purkinje cells in the cerebellum and pan-neuronal cells brain wide;
CN3584 / Purkinje cells in the cerebellum and chandelier cells in the neocortex;
CN3003 / MLI cells in the cerebellum and glutamatergic neurons in the thalamus;
CN2379 / Purkinje cells in the cerebellum and pan-neuronal cells brain wide;

CN3566 / DCN cells in the cerebellum and pan-MSN cells in the striatum;
 CN3453 / DCN cells in the cerebellum and MSN cells in the striatum;
 HCT1 / DCN cells in the cerebellum and spinal motor neurons;
 CN2146 / Bergmann glia in the cerebellum and astrocytes in the neocortex;
 AiP1347 / DCN cells in the cerebellum and astrocytes in the neocortex;
 AiP1351 / Bergmann glia in the cerebellum and astrocytes in the neocortex;
 AiP1375 / MLI cells in the cerebellum and pericytes in the neocortex;
 AiP1307 / MLI cells in the cerebellum and L5 IT neurons in the neocortex;
 AiP1530 (3xcore_E116) / DCN cells in the cerebellum and L5 NP neurons in the neocortex;
 AiP1379 / Purkinje cells in the cerebellum and specialized smooth-muscle cells (SMCs) in the neocortex;
 CN3916 / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
 CN3869 / MLI cells in the cerebellum and pan-GABAergic neurons brain wide; and
 CN3870 / MLI cells in the cerebellum and pan-GABAergic neurons brain wide.

[0027] Aspects of the disclosure are now described with the following additional options and detail: (i) Artificial Expression Constructs & Vectors for Targeted Expression of Genes in Targeted Cell Types; (ii) Compositions for Administration (iii) Cell Lines Including Artificial Expression Constructs; (iv) Transgenic Animals; (v) Methods of Use; (vi) Kits and Commercial Packages; (vii) Exemplary Embodiments; and (viii) Closing Paragraphs.

[0028] (i) Artificial Expression Constructs & Vectors for Targeted Expression of Genes in Targeted Cell Types. Artificial expression constructs disclosed herein include (i) an enhancer sequence that leads to targeted expression of a coding sequence within a targeted central nervous system cell type, (ii) a coding sequence that is expressed, and (iii) a promoter. The artificial expression construct can also include other regulatory elements if necessary or beneficial.

[0029] In particular embodiments, an “enhancer” or an “enhancer element” is a cis-acting sequence that increases the level of transcription associated with a promoter and can function in either orientation relative to the promoter and the coding sequence that is to be transcribed and can be located upstream or downstream relative to the promoter or the coding sequence to be transcribed. There are art-recognized methods and techniques for measuring function(s) of enhancer element sequences. Particular examples of enhancer sequences utilized within artificial expression constructs disclosed herein include eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h

v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m.

[0030] In particular embodiments, a targeted central nervous system cell type enhancer is an enhancer that is uniquely or predominantly utilized by the targeted central nervous system cell type. A targeted central nervous system cell type enhancer enhances expression of a gene in the targeted central nervous system. In certain embodiments, a targeted central nervous system cell type enhancer is also a targeted central nervous system type enhancer that enhances expression of a gene in the targeted central nervous system and does not substantially direct expression of genes in other non-targeted cell types, thus having cell type specific transcriptional activity.

[0031] When a heterologous coding sequence operatively linked to an enhancer disclosed herein leads to expression in a targeted cell type, it leads to expression of the administered heterologous coding sequence in the intended cell type.

[0032] When a heterologous coding sequence is selectively expressed in selected cells, it leads to expression of the administered heterologous coding sequence in the intended cell type and is not substantially expressed in other cell types, as explained in additional detail below. In particular embodiments, not substantially expressed in other cell types is less than 50% expression in a reference cell type as compared to a targeted cell type; less than 40% expression in a reference cell type as compared to a targeted cell type; less than 30% expression in a reference cell type as compared to a targeted cell type; less than 20% expression in a reference cell type as compared to a targeted cell type; or less than 10% expression in a reference cell type as compared to a targeted cell type. In particular embodiments, a reference cell type refers to non-targeted cells. The non-targeted cells can be within the same anatomical structure as the targeted cells and/or can project to a common anatomical area. In particular embodiments, a reference cell type is within an anatomical structure that is adjacent to an anatomical structure that includes the targeted cell type. In particular embodiments, a reference cell type is a non-targeted cell with a different gene expression profile than the targeted cells.

[0033] In particular embodiments, the product of the coding sequence may be expressed at low levels in non-selected cell types, for example at less than 1% or 1%, 2%, 3%, 5%, 10%, 15% or

20% of the levels at which the product is expressed in selected cells. In particular embodiments, the targeted central nervous system cell type is the only cell type that expresses the right combination of transcription factors that bind an enhancer disclosed herein to drive gene expression. Thus, in particular embodiments, expression occurs exclusively within the targeted cell type.

[0034] In particular embodiments, targeted cell types (e.g., neuronal, and/or non-neuronal) can be identified based on transcriptional profiles, such as those described in Tasic *et al.*, Nature 563, 72-78 (2018) and Hodge *et al.*, Nature 573, 61-68 (2019). For reference, the following description of cell types and distinguishing features is also provided:

[0035] Neocortical GABAergic neuron Subclasses:

- All: Express GABA synthesis genes *Gad1/GAD1* and *Gad2/GAD2*.
- *Lamp5*, *Sncg*, *Serpinf1*, and *Vip* GABAergic neurons: Developmentally derived from neuronal progenitors from the caudal ganglionic eminence (CGE) or preoptic area (POA).
- *Sst* and *Pvalb* GABAergic neurons: Developmentally derived from neuronal progenitors in the medial ganglionic eminence (MGE).
- *Lamp5* GABAergic neurons: Found in many neocortical layers, especially upper (L1-L2/3), and have mainly neurogliaform and single bouquet morphology.
- *Lamp5_Lhx6* GABAergic neurons: A subset of *Lamp5* GABAergic neurons that co-express *Lamp5* and *Lhx6*.
- *Sncg* GABAergic neurons: Found in many neocortical layers, and have molecular overlaps with *Lamp5* and *Vip* cells, but inconsistent expression of *Lamp5* or *Vip*, with more consistent expression of *Sncg*.
- *Serpinf1* GABAergic neurons: Found in many neocortical layers, and have molecular overlaps with *Sncg* and *Vip* cells, but inconsistent expression of *Sncg* or *Vip*, with more consistent expression of *Serpinf1*.
- *Vip* GABAergic neurons: Found in many neocortical layers, but especially frequent in upper layers (L1-L4), and highly express the neurotransmitter vasoactive intestinal peptide (*Vip*).
- *Sst* GABAergic neurons: Found in many neocortical layers, but especially frequent in lower layers (L5-L6). They highly express the neurotransmitter somatostatin (*Sst*), and frequently block dendritic inputs to postsynaptic neurons. Included in this subclass are sleep-active *Sst Chodl* neurons (which also express *Nos1* and *Tacr1*) that are highly distinct from other *Sst* neurons but express some shared marker genes including *Sst*. In human, *SST* gene expression is often detected in layer 1 *LAMP5*⁺ GABAergic neuron

subtypes.

- *Pvalb* GABAergic neurons: Found in many neocortical layers, but especially frequent in lower layers (L5-L6). They highly express the calcium-binding protein parvalbumin (*Pvalb*), express neuropeptide Tac1, and frequently dampen the output of postsynaptic neurons. Most fast-spiking GABAergic neurons express *Pvalb* strongly. Included in this subclass are chandelier cells, which have distinct, chandelier-like morphology and express the markers *Cpne5* and *Vipr2* in mouse, and *NOG* and *UNC5B* in human.
- *Meis2*: A distinct subclass defined by a single type, only neocortical GABAergic neuron type that expresses *Meis2* gene, and does not express some other genes that are expressed by all other neocortical GABAergic neuron types (for example, *Thy1* and *Scn2b*). This type is found in L6b and subcortical white matter.
- White matter (WM)-interneurons: A subgroup of interstitial neurons that express GABA and other neuronal markers.

[0036] Neocortical glutamatergic neuron subclasses:

- All: Express glutamate transmitters Slc17a6 and/or Slc17a7. They all express Snap25 and lack expression of Gad1/Gad2.
- L2/3 IT glutamatergic neurons: Primarily reside in Layer 2/3 and have mainly intratelencephalic (cortico-cortical) projections.
- L4 IT glutamatergic neurons: Primarily reside in Layer 4 and mainly have either local or intratelencephalic (cortico-cortical) projections.
- L5 IT glutamatergic neurons: Primarily reside in Layer 5 and have mainly intratelencephalic (cortico-cortical) projections. Also called L5a.
- L5 PT glutamatergic neurons: Primarily reside in Layer 5 and have mainly cortico-subcortical (pyramidal tract or corticofugal) projections. Also called L5b or L5 CF (corticofugal) or L5 ET (extratelencephalic). This subclass includes cells that are located in the primary motor cortex and neighboring areas and are corticospinal projection neurons, which are associated with motor neuron/movement disorders, such as ALS. This subclass includes thick-tufted pyramidal neurons, including distinctive subtypes found only in specialized regions, e.g., Betz cells, Meynert cells, and von Economo cells.
- L5 NP glutamatergic neurons: Primarily reside in Layer 5 and have mainly nearby projections.
- L6 CT glutamatergic neurons: Primarily reside in Layer 6 and have mainly cortico-thalamic projections.
- L6 IT glutamatergic neurons: Primarily reside in Layer 6 and have mainly

intratelencephalic (cortico-cortical) projections.

- L6 IT Car3 glutamatergic neurons: Most densely present in claustrum and endopyriform nucleus, and sparsely throughout L6 in many cortical areas including the primary visual cortex. These cells have mainly intratelencephalic (cortico-cortical) projections. Additional marker genes for claustrum enriched neurons include *Gnb4* and *Ntng2*.
- L6b glutamatergic neurons: Primarily reside in the neocortical subplate (L6b), with local (near the cell body) projections and some cortico-cortical projections from VISp to anterior cingulate, and cortico-subcortical projections to the thalamus.
- CR neurons: A distinct subclass defined by a single type in L1, Cajal-Retzius cells express distinct molecular markers *Lhx5* and *Trp73*.

[0037] Striatal cell Classes and Subclasses:

- Medium spiny neurons, pan: include 95% of striatal neurons and known to express GABA synthesis genes *Gad1/GAD1* and *Gad2/GAD2*, as well as *Ppp1r1b/PPP1R1B*. Medium spiny neurons expressing *Drd3* are herein referred to as *Drd3+* medium spiny neurons.
- Medium spiny neurons, direct pathway-projecting: include nearly 50% of striatal neurons and are enriched for expression of *Drd1/DRD1*, *Pdyn/PDYN*, and *Slc35d3/SLC35D3*. The major axon projection from direct pathway medium spiny neurons is to the substantia nigra pars reticulata (SNr) or to the inner division of the globus pallidus (GPi).
- Medium spiny neurons, indirect pathway-projecting: include nearly 50% of striatal neurons and are enriched for expression of *Drd2/DRD2*, *Adora2a/ADORA2A*, *Gpr6/GPR6*, and *Penk/PENK*. The major axon projection from indirect pathway medium spiny neurons is to the external segment of the globus pallidus (GPe).
- Striatal interneuron-cholinergic: A rare interneuron population including 1% of striatal neurons. These local interneurons have large somata and aspiny dendrites and are known to express *Chat/CHAT* and release the neurotransmitter acetylcholine.

[0038] Cerebellum Subclasses:

- Purkinje cells: large GABAergic neurons that are the only projection neurons and the sole output from the cerebellum. Their cell bodies form a single layer, so called 'Purkinje cell layer', and they express parvalbumin. Other known marker genes include *Gad1*, *Pcp2*, and *Calb1*.
- Deep cerebellar nucleus (DCN) neurons: neurons located in the deep cerebellar nuclei structures. These include excitatory and inhibitory cells, some of which express the gene *Pvalb*. DCN excitatory cells can be identified by the lack of *Gad1* expression, and glutamatergic excitatory cells express *Slc17a6*, while glycinergic excitatory cells express

Slc6a5. Inhibitory DCN cells express Gad1, while glycinergic inhibitory cells also express Slc6a5.

- Molecular layer interneurons (MLI): control Purkinje cell activity via inhibitory synaptic transmission. These include stellate and basket cell types. Exemplary marker genes include Gad1, Pvalb, Kit, and Sorcs3. Stellate and basket cells are GABAergic interneurons marked by Pvalb expression within the molecular layer of the cerebellum.
- Bergmann glia cells: electrically nonexcitable cells chiefly responsible for glutamate uptake and extracellular K⁺ homeostasis. Exemplary marker genes include Gpr3711, Gdf10, Id4, and Npy.
- Granule cells: granule cells of the cerebellar cortex give rise to parallel fibers which are specialized axons that ascend to the molecular layer of the cerebellar cortex. Exemplary marker genes include Slc17a7, Calb2, Gabra6, and Fat2.

[0039] Cholinergic interneurons of the striatum are the main source of acetylcholine in the striatum and are important in basal ganglia physiology and pathophysiology.

[0040] Spinal motor neurons: Motor neurons are a specialized neuron located within the spinal cord and brain responsible for integrating signals from the central nervous system and the sensory systems to control voluntary and involuntary movements. Motor neurons in the spinal cord receive input from neurons in the cortex and relay information to the control muscles throughout the body.

[0041] Non-neuronal Subclasses:

- Astrocytes: Neuroectoderm-derived glial cells which express the marker Aqp4 and often GFAP, but do not express neuronal marker SNAP25. They can have a distinct star-shaped morphology and are involved in metabolic support of other cells in the brain. Multiple astrocyte morphologies are observed in mouse and human
- Oligodendrocytes: Neuroectoderm-derived glial cells, which express the marker Sox10. This category includes oligodendrocyte precursor cells (OPCs). Oligodendrocytes are the subclass that is primarily responsible for myelination of neurons.
- VLMCs: Vascular leptomeningeal cells (VLMCs) are part of the meninges that surround the outer layer of the cortex and express the marker genes Lum and Col1a1.
- Pericytes: Blood vessel-associated cells that express the marker genes Kcnj8 and Abcc9. Pericytes wrap around endothelial cells and are important for regulation of capillary blood flow and are involved in blood-brain barrier permeability.
- SMCs: Specialized smooth-muscle cells which are blood vessel-associated cells that express the marker gene Acta2. SMCs cover arterioles in the brain and are involved in blood-brain barrier permeability.

- Endothelial cells: Cells that line blood vessels of the brain. Endothelial cells express the markers Tek and PDGF-B.
- Microglia: hematopoietic-derived immune cells, which are brain-resident macrophages, and perivascular macrophages (PVMs) that may be transitionally associated with brain tissue or included as a byproduct of brain dissection methods. Microglia are known to express Cx3cr1, Tmem119, and PTPRC (CD45).

[0042] In particular embodiments, a coding sequence is a heterologous coding sequence that encodes an effector element. An effector element is a sequence that is expressed to achieve, and that in fact achieves, an intended effect. Examples of effector elements include reporter genes/proteins and functional genes/proteins.

[0043] Exemplary reporter genes/proteins include those expressed by Addgene ID#s 83894 (pAAV-hDlx-Flex-dTomato-Fishell_7), 83895 (pAAV-hDlx-Flex-GFP-Fishell_6), 83896 (pAAV-hDlx-GiDREADD-dTomato-Fishell-5), 83898 (pAAV-mDlx-ChR2-mCherry-Fishell-3), 83899 (pAAV-mDlx-GCaMP6f-Fishell-2), 83900 (pAAV-mDlx-GFP-Fishell-1), and 89897 (pcDNA3-FLAG-mTET2 (N500)). Exemplary reporter genes particularly can include those which encode an expressible fluorescent protein, or expressible biotin; blue fluorescent proteins (e.g. eBFP, eBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire); cyan fluorescent proteins (e.g. eCFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan, mTurquoise); green fluorescent proteins (e.g. GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green (mAzamigreen), CopGFP, AceGFP, avGFP, ZsGreen1, Oregon Green™ (Thermo Fisher Scientific)); Luciferase; orange fluorescent proteins (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato, dTomato); red fluorescent proteins (mKate, mKate2, mPlum, DsRed monomer, mCherry, mRuby, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred, Texas Red™ (Thermo Fisher Scientific)); far red fluorescent proteins (e.g., mPlum and mNeptune); yellow fluorescent proteins (e.g., YFP, eYFP, Citrine, SYFP2, Venus, YPet, PhiYFP, ZsYellow1); and tandem conjugates.

[0044] GFP is composed of 238 amino acids (26.9 kDa), originally isolated from the jellyfish *Aequorea victoria*/*Aequorea aequorea*/*Aequorea forskalea* that fluoresces green when exposed to blue light. The GFP from *A. victoria* has a major excitation peak at a wavelength of 395 nm and a minor one at 475 nm. Its emission peak is at 509 nm which is in the lower green portion of the visible spectrum. The GFP from the sea pansy (*Renilla reniformis*) has a single major excitation peak at 498 nm. Due to the potential for widespread usage and the evolving needs of researchers, many different mutants of GFP have been engineered. The first major improvement was a single

point mutation (S65T) reported in 1995 in Nature by Roger Tsien. This mutation dramatically improved the spectral characteristics of GFP, resulting in increased fluorescence, photostability and a shift of the major excitation peak to 488 nm with the peak emission kept at 509 nm. The addition of the 37°C folding efficiency (F64L) point mutant to this scaffold yielded enhanced GFP (EGFP). EGFP has an extinction coefficient (denoted ϵ), also known as its optical cross section of $9.13 \times 10^4 \text{ m}^2/\text{molecule}$, also quoted as $55,000 \text{ L}/(\text{mol} \cdot \text{cm})$. Superfolder GFP, a series of mutations that allow GFP to rapidly fold and mature even when fused to poorly folding peptides, was reported in 2006.

[0045] The "yellow fluorescent protein" (YFP) is a genetic mutant of green fluorescent protein, derived from *Aequorea victoria*. Its excitation peak is 514 nm and its emission peak is 527 nm.

[0046] Exemplary functional molecules include functioning ion transporters, cellular trafficking proteins, enzymes, transcription factors, neurotransmitters, calcium reporters, channelrhodopsins, guide RNA, nucleases, microRNA, or designer receptors exclusively activated by designer drugs (DREADDs).

[0047] Ion transporters are transmembrane proteins that mediate transport of ions across cell membranes. These transporters are pervasive throughout most cell types and important for regulating cellular excitability and homeostasis. Ion transporters participate in numerous cellular processes such as action potentials, synaptic transmission, hormone secretion, and muscle contraction. Many important biological processes in living cells involve the translocation of cations, such as calcium (Ca^{2+}), potassium (K^{+}), and sodium (Na^{+}) ions, through such ion channels. In particular embodiments, ion transporters include voltage gated sodium channels (e.g., *SCN1A*), potassium channels (e.g., *KCNQ2*), and calcium channels (e.g., *CACNA1C*).

[0048] Exemplary enzymes, transcription factors, receptors, membrane proteins, cellular trafficking proteins, signaling molecules, and neurotransmitters include enzymes such as lactase, lipase, helicase, alpha-glucosidase, amylase; transcription factors such as SP1, AP-1, Heat shock factor protein 1, C/EBP (CCAAT/enhancer binding protein), and Oct-1; receptors such as transforming growth factor receptor beta 1, platelet-derived growth factor receptor, epidermal growth factor receptor, vascular endothelial growth factor receptor, and interleukin 8 receptor alpha; membrane proteins, cellular trafficking proteins such as clathrin, dynamin, caveolin, Rab-4A, and Rab-11A; signaling molecules such as nerve growth factor (NGF), platelet-derived growth factor (PDGF), transforming growth factor β (TGF β), epidermal growth factor (EGF), GTPase and HRas; and neurotransmitters such as cocaine and amphetamine regulated transcript, substance P, oxytocin, and somatostatin.

[0049] In particular embodiments, functional molecules include reporters of cell function and

states such as calcium reporters. Intracellular calcium concentration is an important predictor of numerous cellular activities, which include neuronal activation, muscle cell contraction and second messenger signaling. A sensitive and convenient technique to monitor the intracellular calcium levels is through the genetically encoded calcium indicator (GECI). Among the GECIs, green fluorescent protein (GFP) based calcium sensors named GCaMPs are efficient and widely used tools. The GCaMPs are formed by fusion of M13 and calmodulin protein to N- and C-termini of circularly permuted GFP. Some GCaMPs yield distinct fluorescence emission spectra (Zhao *et al.*, Science, 2011, 333(6051): 1888-1891). Exemplary GECIs with green fluorescence include GCaMP3, GCaMP5G, GCaMP6s, GCaMP6m, GCaMP6f, jGCaMP7s, jGCaMP7c, jGCaMP7b, jGCaMP7f, jGCaMP8s, jGCaMP8m, and jGCaMP8f. Furthermore, GECIs with red fluorescence include jRGECO1a and jRGECO1b. AAV products containing GECIs are commercially available. For example, Vigene Biosciences provides AAV products including AAV8-CAG-GCaMP3 (Cat. No:BS4-CX3AAV8), AAV8-Syn-FLEX-GCaMP6s-WPRE (Cat. No:BS1-NXSAAV8), AAV8-Syn-FLEX-GCaMP6s-WPRE (Cat. No:BS1-NXSAAV8), AAV9-CAG-FLEX-GCaMP6m-WPRE (Cat. No:BS2-CXMAAV9), AAV9-Syn-FLEX-jGCaMP7s-WPRE (Cat. No:BS12-NXSAAV9), AAV9-CAG-FLEX-jGCaMP7f-WPRE (Cat. No:BS12-CXFAAV9), AAV9-Syn-FLEX-jGCaMP7b-WPRE (Cat. No:BS12-NXBAAV9), AAV9-Syn-FLEX-jGCaMP7c-WPRE (Cat. No:BS12-NXCAAV9), AAV9-Syn-FLEX-NES-jRGECO1a-WPRE (Cat. No:BS8-NXAAAV9), and AAV8-Syn-FLEX-NES-jRCaMP1b-WPRE (Cat. No:BS7-NXBAAV8).

[0050] In particular embodiments calcium reporters include the genetically encoded calcium indicators GECI, NTnC; Myosin light chain kinase, GFP, Calmodulin chimera; Calcium indicator TN-XXL; BRET-based auto-luminescent calcium indicator; and/or Calcium indicator protein OeNL(Ca²⁺)-18u).

[0051] In particular embodiments, functional molecules include modulators of neuronal activity like channelrhodopsins (e.g., channelrhodopsin-1, channelrhodopsin-2, and variants thereof). Channelrhodopsins are a subfamily of retinylidene proteins (rhodopsins) that function as light-gated ion channels. In addition to channelrhodopsin 1 (ChR1) and channelrhodopsin 2 (ChR2), several variants of channelrhodopsins have been developed. For example, Lin *et al.* (*Biophys J*, 2009, 96(5): 1803-14) describe making chimeras of the transmembrane domains of ChR1 and ChR2, combined with site-directed mutagenesis. Zhang *et al.* (*Nat Neurosci*, 2008, 11(6): 631-3) describe VChR1, which is a red-shifted channelrhodopsin variant. VChR1 has lower light sensitivity and poor membrane trafficking and expression. Other known channelrhodopsin variants include the ChR2 variant described in Nagel, *et al.*, *Proc Natl Acad Sci USA*, 2003, 100(24): 13940-5), ChR2/H134R (Nagel, G., *et al.*, *Curr Biol*, 2005, 15(24):

2279-84), and ChD/ChEF/ChIEF (Lin, J. Y., *et al.*, *Biophys J*, 2009, 96(5): 1803-14), which are activated by blue light (470 nm) but show no sensitivity to orange/red light. Additional variants are described in Lin, *Experimental Physiology*, 2010, 96.1: 19-25; Knopfel *et al.*, *The Journal of Neuroscience*, 2010, 30(45): 14998-15004; and Mardinly *et al.*, *Nat Neurosci.* 2018, 21(6):881-893).

[0052] In particular embodiments, functional molecules include DNA and RNA editing tools such as CRISPR/Cas (e.g., guide RNA and a nuclease, such as Cas, Cas9 or cpf1). Functional molecules can also include engineered Cpf1s such as those described in US 2018/0030425, US 2016/0208243, WO/2017/184768 and Zetsche *et al.* (2015) *Cell* 163: 759-771; single gRNA (see e.g., Jinek *et al.* (2012) *Science* 337:816-821; Jinek *et al.* (2013) *eLife* 2:e00471; Segal (2013) *eLife* 2:e00563) or editase, guide RNA molecules, microRNA, or homologous recombination donor cassettes.

[0053] In particular embodiments, functional molecules include a localizing cassette. In particular embodiments, localizing cassettes are used to localize a molecule (e.g., a vector, a protein, a sensor) to a specific subcellular compartment such as the soma, axon, or dendrite(s) of a neuron. In particular embodiments, localizing cassettes include a soma tag (e.g., soma (EE-RR)) to localize at the soma; an axon tag (e.g. derived from GAP43) or synaptophysin (sy) to localize at the axon; hydrophobic tails to localize at the plasma membrane; and hydrophobicity or alkyl chain to localize at the endoplasmic reticulum. In particular embodiments, localizing cassettes are fused to a sensor molecule such as a GECl. In particular embodiments, fusion proteins of a GECl and a localizing cassette includes soma-jGCaMP8s, axon-jRGECO1a, syGCaMP5G, and soma-jGCaMP7s.

[0054] In particular embodiments, functional molecules include tag cassettes. A tag cassette includes His tag (HHHHHH; SEQ ID NO: 151), Flag tag (DYKDDDDK; SEQ ID NO: 152), Xpress tag (DLYDDDDK; SEQ ID NO: 153), Avi tag (GLNDIFEAQKIEWHE; SEQ ID NO: 154), Calmodulin tag (KRRWKKNFIAVSAANRFKKISSSGAL; SEQ ID NO: 155), Polyglutamate tag, HA tag (YPYDVPDYA; SEQ ID NO: 156), Myc tag (EQKLISEEDL; SEQ ID NO: 157), Strep tag (which refers the original STREP® tag (WRHPQFGG; SEQ ID NO: 158), STREP® tag II (WSHPQFEK SEQ ID NO: 159 (IBA Institut für Bioanalytik, Germany); see, e.g., US 7,981,632), Softag 1 (SLAELLNAGLGGS; SEQ ID NO: 160), Softag 3 (TQDPSRVG; SEQ ID NO: 161), and V5 tag (GKPIPNPLLGLDST; SEQ ID NO: 162). In particular embodiments, a tag cassette includes a fusion of tag cassettes such as 3XFLAG. In particular embodiments, 3XFLAG includes the sequence set forth in SEQ ID NO: 164. In particular embodiments artificial expression constructs include 10aa (SEQ ID NO: 38)

[0055] Sequences are publicly-available. As examples, lactase (e.g., GenBank: EAX11622.1), lipase (e.g., GenBank: AAA60129.1), helicase (e.g., GenBank: AMD82207.1), amylase (e.g., GenBank: AAA51724.1), alpha-glucosidase (e.g., GenBank: ABI53718.1), transcription factor SP1 (e.g., UniProtKB/Swiss-Prot: P08047.3), transcription factor AP-1 (e.g., NP_002219.1), heat shock factor protein 1 (e.g., UniProtKB/Swiss-Prot: Q00613.1), CCAAT/enhancer-binding protein (C/EBP) beta isoform a (e.g., NP_005185.2), Oct-1 (e.g., UniProtKB/Swiss-Prot: P14859.2), TGF β (e.g., GenBank: CAF02096.2), glial cell line-derived neurotrophic factor (GDNF) (e.g., NP_001177397.1), platelet-derived growth factor receptor (e.g., GenBank: AAA60049.1), epidermal growth factor receptor (e.g., GenBank: CAA25240.1), vascular endothelial growth factor receptor (e.g., GenBank: AAC16449.2), interleukin 8 receptor alpha (e.g., GenBank: AAB59436.1), caveolin (e.g., GenBank: CAA79476.1), dynamin (e.g., GenBank: AAA88025.1), clathrin heavy chain 1 isoform 1 (e.g., NP_004850.1), clathrin heavy chain 2 isoform 1 (e.g., NP_009029.3), clathrin light chain A isoform a (e.g., NP_001824.1), clathrin light chain B isoform a (e.g., NP_001825.1), ras-related protein Rab-4A isoform 1 (e.g., NP_004569.2), ras-related protein Rab-11A (e.g., UniProtKB/Swiss-Prot: P62491.3), platelet-derived growth factor (e.g., GenBank: AAA60552.1), transforming growth factor-beta3 (e.g., GenBank: AAA61161.1), nerve growth factor (e.g., GenBank: CAA37703.1), EGF (e.g., GenBank: CAA34902.2), cocaine and amphetamine regulated transcript (Chain A) (e.g., PDB: 1HY9_A), protachykinin-1 (e.g., UniProtKB – P20366), oxytocin-neurophysin 1 (e.g., UniProtKB – P01178), somatostatin (e.g., GenBank: AAH32625.1), genetically-encoded green calcium indicator NTnC (chain A) [synthetic construct] (e.g., PDB: 5MWC_A), calcium indicator TN-XXL [synthetic construct], (e.g., GenBank: ACF93133.1), BRET-based auto-luminescent calcium indicator [synthetic construct] (e.g., GenBank ADF42668.1), calcium indicator protein OeNL(Ca²⁺)-18u [synthetic construct], ((e.g., GenBank BBB18812.1), myosin light chain kinase, Green fluorescent protein, Calmodulin chimera (Chain A) [synthetic construct] ((e.g., PDB: 3EKJ_A), channelopsin 1 (e.g., *UniProtKB* – F8UVI5), channelopsin 1 (e.g., GenBank: AER58217.1), channelrhodopsin-2 ((e.g., UniProtKB – B4Y105), channel rhodopsin 2 [synthetic construct] ((e.g., GenBank: ABO64386.1), CRISPR-associated protein (Cas) (e.g., GenBank: AKG27598.1), Cas9 [synthetic construct] (e.g., GenBank: AST09977.1), CRISPR-associated endonuclease Cpf1 (e.g., UniProtKB/Swiss-Prot: U2UMQ6.1), ribonuclease 4 or ribonuclease L (e.g., UniProtKB/Swiss-Prot: Q05823.2), deoxyribonuclease II beta (e.g., GenBank: AAF76893.1), sodium channel protein type 1 subunit alpha (e.g., UniProtKB – P35498), potassium voltage-gated channel subfamily KQT member 2 (e.g., UniProtKB – O43526), and voltage-dependent L-type calcium channel subunit alpha-1C (e.g., UniProtKB – Q13936).

[0056] Additional effector elements include Cre, iCre, dgCre, FlpO, and tTA2. iCre refers to a codon-improved Cre. dgCre refers to an enhanced GFP/Cre recombinase fusion gene with an N terminal fusion of the first 159 amino acids of the Escherichia coli K-12 strain chromosomal dihydrofolate reductase gene (DHFR or folA) harboring a G67S mutation and modified to also include the R12Y/Y100I destabilizing domain mutation. FlpO refers to a codon-optimized form of FLPe that greatly increases protein expression and FRT recombination efficiency in mouse cells. Like the Cre/LoxP system, the FLP/FRT system has been widely used for gene expression (and generating conditional knockout mice, mediated by the FLP/FRT system). tTA2 refers to tetracycline transactivator.

[0057] Exemplary expressible elements are expression products that do not include effector elements, for example, a non-functioning or defective protein. In particular embodiments, expressible elements can provide methods to study the effects of their functioning counterparts. In particular embodiments, expressible elements are non-functioning or defective based on an engineered mutation that renders them non-functioning. In these aspects, non-expressible elements are as similar in structure as possible to their functioning counterparts.

[0058] Exemplary self-cleaving peptides include the 2A peptides which lead to the production of two proteins from one mRNA. The 2A sequences are short (e.g., 20 amino acids), allowing more use in size-limited constructs. Particular examples include P2A, T2A, E2A, and F2A. In particular embodiments, the artificial expression constructs include an internal ribosome entry site (IRES) sequence. IRES allow ribosomes to initiate translation at a second internal site on a mRNA molecule, leading to production of two proteins from one mRNA.

[0059] Artificial expression constructs can encode nuclear localization proteins, such as Histone H1, Histone H2A, Histone H2B, Histone H3, Histone H4, histone-like protein HPhA, H2Bmod, or H2B*.

[0060] Coding sequences encoding molecules (e.g., RNA, proteins) described herein can be obtained from publicly available databases and publications. Coding sequences can further include various sequence polymorphisms, mutations, and/or sequence variants wherein such alterations do not affect the function of the encoded molecule. The term “encode” or “encoding” refers to a property of sequences of nucleic acids, such as a vector, a plasmid, a gene, cDNA, mRNA, to serve as templates for synthesis of other molecules such as proteins.

[0061] The term “gene” may include not only coding sequences but also regulatory regions such as promoters, enhancers, insulators, and/or post-regulatory elements, such as termination regions. The term further can include all introns and other DNA sequences spliced from the mRNA transcript, along with variants resulting from alternative splice sites. The sequences can also

include degenerate codons of a reference sequence or sequences that may be introduced to provide codon preference in a specific organism or cell type.

[0062] Promoters can include general promoters, tissue-specific promoters, cell-specific promoters, and/or promoters specific for the cytoplasm. Promoters may include strong promoters, weak promoters, constitutive expression promoters, and/or inducible promoters. Inducible promoters direct expression in response to certain conditions, signals or cellular events. For example, the promoter may be an inducible promoter that requires a particular ligand, small molecule, transcription factor or hormone protein in order to effect transcription from the promoter. Particular examples of promoters include minBglobin (also referred to as minBGprom), CMV, minCMV, minCMV* (minCMV* is minCMV with a SacI restriction site removed), minRho, minRho* (minRho* is minRho with a SacI restriction site removed), SV40 immediately early promoter, the Hsp68 minimal promoter (proHSP68), and the Rous Sarcoma Virus (RSV) long-terminal repeat (LTR) promoter. Minimal promoters have no activity to drive gene expression on their own but can be activated to drive gene expression when linked to a proximal enhancer element.

[0063] In particular embodiments, expression constructs are provided within vectors. The term vector refers to a nucleic acid molecule capable of transferring or transporting another nucleic acid molecule, such as an expression construct. The transferred nucleic acid is generally linked to, e.g., inserted into, the vector nucleic acid molecule. A vector may include sequences that direct autonomous replication in a cell or may include sequences that permit integration into host cell DNA. Useful vectors include, for example, plasmids (e.g., DNA plasmids or RNA plasmids), transposons, cosmids, bacterial artificial chromosomes, and viral vectors.

[0064] Viral vector is widely used to refer to a nucleic acid molecule that includes virus-derived components that facilitate transfer and expression of non-native nucleic acid molecules within a cell. The term adeno-associated viral vector refers to a viral vector or plasmid containing structural and functional genetic elements, or portions thereof, that are primarily derived from AAV. The term “retroviral vector” refers to a viral vector or plasmid containing structural and functional genetic elements, or portions thereof, that are primarily derived from a retrovirus. The term “lentiviral vector” refers to a viral vector or plasmid containing structural and functional genetic elements, or portions thereof, that are primarily derived from a lentivirus, and so on. The term “hybrid vector” refers to a vector including structural and/or functional genetic elements from more than one virus type.

[0065] Adenovirus vectors refer to those constructs containing adenovirus sequences sufficient to (a) support packaging of an artificial expression construct and (b) to express a coding sequence that has been cloned therein in a sense or antisense orientation. A recombinant Adenovirus vector

includes a genetically engineered form of an adenovirus. Knowledge of the genetic organization of adenovirus, a 36 kb, linear, double-stranded DNA virus, allows substitution of large pieces of adenoviral DNA with foreign sequences up to 7 kb. In contrast to retrovirus, the adenoviral infection of host cells does not result in chromosomal integration because adenoviral DNA can replicate in an episomal manner without potential genotoxicity. Also, adenoviruses are structurally stable, and no genome rearrangement has been detected after extensive amplification.

[0066] Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized genome, ease of manipulation, high titer, wide target-cell range, and high infectivity. Both ends of the viral genome contain 100-200 base pair inverted repeats (ITRs), which are cis elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome contain different transcription units that are divided by the onset of viral DNA replication. The E1 region (E1A and E1B) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication. These proteins are involved in DNA replication, late gene expression, and host cell shut-off. The products of the late genes, including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary transcript issued by the major late promoter (MLP). The MLP is particularly efficient during the late phase of infection, and all the mRNAs issued from this promoter possess a 5'-tripartite leader (TPL) sequence which makes them preferred mRNAs for translation.

[0067] Other than the requirement that an adenovirus vector be replication defective, or at least conditionally defective, the nature of the adenovirus vector is not believed to be crucial to the successful practice of particular embodiments disclosed herein. The adenovirus may be of any of the 42 different known serotypes or subgroups A-F. In particular embodiments, adenovirus type 5 of subgroup C is the preferred starting material in order to obtain a conditional replication-defective adenovirus vector for use in particular embodiments, since Adenovirus type 5 is a human adenovirus about which a great deal of biochemical and genetic information is known, and it has historically been used for most constructions employing adenovirus as a vector.

[0068] As indicated, the typical vector is replication defective and will not have an adenovirus E1 region. Thus, it will be most convenient to introduce the polynucleotide encoding the gene of interest at the position from which the E1-coding sequences have been removed. However, the position of insertion of the construct within the adenovirus sequences is not critical. The polynucleotide encoding the gene of interest may also be inserted in lieu of a deleted E3 region

in E3 replacement vectors or in the E4 region where a helper cell line or helper virus complements the E4 defect.

[0069] Adeno-Associated Virus (AAV) is a parvovirus, discovered as a contamination of adenoviral stocks. It is a ubiquitous virus (antibodies are present in 85% of the US human population) that has not been linked to any disease. It is also classified as a dependovirus, because its replication is dependent on the presence of a helper virus, such as adenovirus. Various serotypes have been isolated, of which AAV-2 is the best characterized. AAV has a single-stranded linear DNA that is encapsidated into capsid proteins VP1, VP2 and VP3 to form an icosahedral virion of 20 to 24 nm in diameter.

[0070] The AAV DNA is 4.7 kilobases long. It contains two open reading frames and is flanked by two ITRs. There are two major genes in the AAV genome: rep and cap. The rep gene codes for proteins responsible for viral replications, whereas cap codes for capsid protein VP1-3. Each ITR forms a T-shaped hairpin structure. These terminal repeats are the only essential cis components of the AAV for chromosomal integration. Therefore, the AAV can be used as a vector with all viral coding sequences removed and replaced by the cassette of genes for delivery. Three AAV viral promoters have been identified and named p5, p19, and p40, according to their map position. Transcription from p5 and p19 results in production of rep proteins, and transcription from p40 produces the capsid proteins.

[0071] AAVs stand out for use within the current disclosure because of their superb safety profile and because their capsids and genomes can be tailored to allow expression in targeted cell populations. scAAV refers to a self-complementary AAV. pAAV refers to a plasmid adeno-associated virus. rAAV refers to a recombinant adeno-associated virus.

[0072] Other viral vectors may also be employed. For example, vectors derived from viruses such as vaccinia virus, polioviruses and herpes viruses may be employed. They offer several attractive features for various mammalian cells.

[0073] Retroviruses are a common tool for gene delivery. "Retrovirus" refers to an RNA virus that reverse transcribes its genomic RNA into a linear double-stranded DNA copy and subsequently covalently integrates its genomic DNA into a host genome. Once the virus is integrated into the host genome, it is referred to as a "provirus." The provirus serves as a template for RNA polymerase II and directs the expression of RNA molecules which encode the structural proteins and enzymes needed to produce new viral particles.

[0074] Illustrative retroviruses suitable for use in particular embodiments, include: Moloney murine leukemia virus (M-MuLV), Moloney murine sarcoma virus (MoMSV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), gibbon ape leukemia virus

(GaLV), feline leukemia virus (FLV), spumavirus, Friend murine leukemia virus, Murine Stem Cell Virus (MSCV), Rous Sarcoma Virus (RSV), and lentivirus.

[0075] "Lentivirus" refers to a group (or genus) of complex retroviruses. Illustrative lentiviruses include: HIV (human immunodeficiency virus; including HIV type 1, and HIV type 2); visna-maedi virus (VMV); the caprine arthritis-encephalitis virus (CAEV); equine infectious anemia virus (EIAV); feline immunodeficiency virus (FIV); bovine immune deficiency virus (BIV); and simian immunodeficiency virus (SIV). In particular embodiments, HIV based vector backbones (i.e., HIV cis-acting sequence elements) can be used.

[0076] A safety enhancement for the use of some vectors can be provided by replacing the U3 region of the 5' LTR with a heterologous promoter to drive transcription of the viral genome during production of viral particles. Examples of heterologous promoters which can be used for this purpose include, for example, viral simian virus 40 (SV40) (e.g., early or late), cytomegalovirus (CMV) (e.g., immediate early), Moloney murine leukemia virus (MoMLV), Rous sarcoma virus (RSV), and herpes simplex virus (HSV) (thymidine kinase) promoters. Typical promoters are able to drive high levels of transcription in a Tat-independent manner. This replacement reduces the possibility of recombination to generate replication-competent virus because there is no complete U3 sequence in the virus production system. In particular embodiments, the heterologous promoter has additional advantages in controlling the manner in which the viral genome is transcribed. For example, the heterologous promoter can be inducible, such that transcription of all or part of the viral genome will occur only when the induction factors are present. Induction factors include one or more chemical compounds or the physiological conditions such as temperature or pH, in which the host cells are cultured.

[0077] In particular embodiments, viral vectors include a TAR element. The term "TAR" refers to the "trans-activation response" genetic element located in the R region of lentiviral LTRs. This element interacts with the lentiviral trans-activator (tat) genetic element to enhance viral replication. However, this element is not required in embodiments wherein the U3 region of the 5' LTR is replaced by a heterologous promoter.

[0078] The "R region" refers to the region within retroviral LTRs beginning at the start of the capping group (i.e., the start of transcription) and ending immediately prior to the start of the poly(A) tract. The R region is also defined as being flanked by the U3 and U5 regions. The R region plays a role during reverse transcription in permitting the transfer of nascent DNA from one end of the genome to the other.

[0079] In particular embodiments, expression of heterologous sequences in viral vectors is increased by incorporating posttranscriptional regulatory elements, efficient polyadenylation sites,

and optionally, transcription termination signals into the vectors. A variety of posttranscriptional regulatory elements can increase expression of a heterologous nucleic acid. Examples include the woodchuck hepatitis virus posttranscriptional regulatory element (WPRE; Zufferey *et al.*, 1999, *J. Virol.*, 73:2886); the posttranscriptional regulatory element present in hepatitis B virus (HPRE) (Smith *et al.*, *Nucleic Acids Res.* 26(21):4818-4827, 1998); and the like (Liu *et al.*, 1995, *Genes Dev.*, 9:1766). In particular embodiments, vectors include a posttranscriptional regulatory element such as a WPRE or HPRE. In particular embodiments, vectors lack or do not include a posttranscriptional regulatory element such as a WPRE or HPRE.

[0080] Elements directing the efficient termination and polyadenylation of a heterologous nucleic acid transcript can increase heterologous gene expression. Transcription termination signals are generally found downstream of the polyadenylation signal. In particular embodiments, vectors include a polyadenylation signal 3' of a polynucleotide encoding a molecule (e.g., protein) to be expressed. The term "poly(A) site" or "poly(A) sequence" denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript by RNA polymerase II. Polyadenylation sequences can promote mRNA stability by addition of a poly(A) tail to the 3' end of the coding sequence and thus, contribute to increased translational efficiency. Particular embodiments may utilize BGHpA, hGHpA, or SV40pA. In particular embodiments, a preferred embodiment of an expression construct includes a terminator element. These elements can serve to enhance transcript levels and to minimize read through from the construct into other plasmid sequences.

[0081] In particular embodiments, a viral vector further includes one or more insulator elements. Insulators elements may contribute to protecting viral vector-expressed sequences, e.g., effector elements or expressible elements, from integration site effects, which may be mediated by cis-acting elements present in genomic DNA and lead to deregulated expression of transferred sequences (*i.e.*, position effect; see, e.g., Burgess-Beusse *et al.*, *PNAS*, USA, 99:16433, 2002; and Zhan *et al.*, *Hum. Genet.*, 109:471, 2001). In particular embodiments, viral transfer vectors include one or more insulator elements at the 3' LTR and upon integration of the provirus into the host genome, the provirus includes the one or more insulators at both the 5' LTR and 3' LTR, by virtue of duplicating the 3' LTR. Suitable insulators for use in particular embodiments include the chicken β -globin insulator (see Chung *et al.*, *Cell* 74:505, 1993; Chung *et al.*, *PNAS USA* 94:575, 1997; and Bell *et al.*, *Cell* 98:387, 1999), SP10 insulator (Abhyankar *et al.*, *JBC* 282:36143, 2007), or other small CTCF recognition sequences that function as enhancer blocking insulators (Liu *et al.*, *Nature Biotechnology*, 33:198, 2015).

[0082] Beyond the foregoing description, a wide range of suitable expression vector types will be known to a person of ordinary skill in the art. These can include commercially available expression vectors designed for general recombinant procedures, for example plasmids that contain one or more reporter genes and regulatory elements required for expression of the reporter gene in cells. Numerous vectors are commercially available, e.g., from Invitrogen, Stratagene, Clontech, etc., and are described in numerous associated guides. In particular embodiments, suitable expression vectors include any plasmid, cosmid or phage construct that is capable of supporting expression of encoded genes in mammalian cell, such as pUC or Bluescript plasmid series.

[0083] Particular embodiments of vectors disclosed herein include:

Expression Construct Name	Features
CN1259	scAAV-eHGT_023h v1-minBglobin-SYFP2-WPRE3-BGHpA
CN1280	scAAV-eHGT_023m-minBglobin-SYFP2-WPRE3-BGHpA
CN1521	rAAV-hsA2-3xCore-eHGT_023h v3-minRho-SYFP2-WPRE3-BGHpA
CN1528	rAAV-hsA2-eHGT_082h-minRho-SYFP2-WPRE3-BGHpA
CN1533	rAAV-hsA2-eHGT_087h-minRho-SYFP2-WPRE3-BGHpA
CN1621	rAAV-hsA2-eHGT_128h-minRho-SYFP2-WPRE3-BGHpA
CN1674	rAAV-hsA2-eHGT_181h-minRho-SYFP2-WPRE3-BGHpA
CN1778	rAAV-hsA2-eHGT_260h-minRho-SYFP2-WPRE3-BGHpA
CN1932	rAAV-hsA2-eHGT_023h v2-minRho-SYFP2-WPRE3-BGHpA
CN2043	rAAV-3xSP10ins-eHGT_356h-minRho*-SYFP2-WPRE3-BGHpA
CN2045	rAAV-3xSP10ins-eHGT_359h-minRho*-SYFP2-WPRE3-BGHpA
CN2085	rAAV-eHGT_375m-minBglobin-SYFP2-WPRE3-BGHpA
CN2102	rAAV-eHGT_387m-minBglobin-SYFP2-WPRE3-BGHpA
CN2157	rAAV-eHGT_395h-minBglobin-SYFP2-WPRE3-BGHpA
CN2216	rAAV-3x(coreB)eHGT121h-minCMV*-SYFP2-WPRE3-BGHpA
CN2251	rAAV-eHGT_453m-minBglobin-SYFP2-WPRE3-BGHpA
CN2257	rAAV-eHGT_470m-minBglobin-SYFP2-WPRE3-BGHpA
CN2258	rAAV-eHGT_479m-minBglobin-SYFP2-WPRE3-BGHpA
CN2267	rAAV-eHGT_494m-minBglobin-SYFP2-WPRE3-BGHpA
CN2316	rAAV-eHGT_467m-minBglobin-SYFP2-WPRE3-BGHpA
CN2339	rAAV-eHGT_023h v1-minRho*-SYFP2-10aa-H2Bmod-WPRE3-BGHpA
CN2431	rAAV-eHGT_483m-minBglobin-SYFP2-WPRE3-BGHpA
CN2436	rAAV-eHGT_606h-minBglobin-SYFP2-WPRE3-BGHpA
CN2643	rAAV-eHGT_738m-minBglobin-SYFP2-WPRE3-BGHpA
CN2663	rAAV-eHGT_796h-minBglobin-SYFP2-WPRE3-BGHpA
CN2717	rAAV-eHGT_710m-minBglobin-SYFP2-WPRE3-BGHpA
CN2838	rAAV-eHGT_588m-minBglobin-SYFP2-WPRE3-BGHpA
CN1415	rAAV-eHGT_007m-minBglobin-SYFP2-WPRE3-BGHpA
CN2710	rAAV-eHGT_703m-minBglobin-SYFP2-WPRE3-BGHpA
CN2839	rAAV-eHGT_589m-minBglobin-SYFP2-WPRE3-BGHpA
CN2674	rAAV-eHGT_086h-minBglobin-SYFP2-WPRE3-BGHpA
CN3301	rAAV-eHGT_963m-minBG-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN3019	rAAV-3xcore3_eHGT_453m-minBglobin-SYFP2-WPRE3-BGHpA

CN3569	rAAV-3xcore2_eHGT_387m-minBglobin-SYFP2-WPRE3-BGHpA
CN2374	rAAV-eHGT_534h-minBglobin-SYFP2-WPRE3-BGHpA
CN3584	rAAV-3xcore2_eHGT_475m-minBglobin-SYFP2-WPRE3-BGHpA
CN3003	rAAV-eHGT_830h-minBglobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN2379	rAAV-eHGT_540h-minBglobin-SYFP2-WPRE3-BGHpA
CN3566	rAAV-3xSP10ins-3xcore2_eHGT_351h-minRho*-SYFP2-WPRE3-BGHpA
CN3453	rAAV-eHGT_882m-minBglobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
HCT1	rAAV-eHGT_1137m-minBglobin-SYFP2-WPRE3-bGHpA
CN2146	rAAV-eHGT_381h-minBglobin-SYFP2-WPRE3-BGHpA
AiP1347	pAAV-MGT_E118-minBglobin-SYFP2-WPRE3-bGHpA
AiP1351	pAAV-MGT_E122-minBglobin-SYFP2-WPRE3-bGHpA
AiP1375	pAAV-MGT_E146-minBglobin-SYFP2-WPRE3-bGHpA
AiP1307	pAAV-MGT_E16-minBglobin-SYFP2-WPRE3-bGHpA
AiP1530	pAAV-3xCore_MGT_E116-minBglobin-SYFP2-WPRE3-bGHpA
AiP1379	pAAV-MGT_E150-minBglobin-SYFP2-WPRE3-bGHpA
CN3916	rAAV-eHGT_1032h-minBglobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN3869	rAAV-eHGT_1027h-minBglobin-SYFP2-P2A-3XFLAG-10aa-H2B*-WPRE3-BGHpA
CN3870	rAAV-eHGT_1027m-minBglobin-SYFP2-P2A-3XFLAG-10aa-H2B*-WPRE3-BGHpA

[0084] Subcomponent sequences within the larger vector sequences can be readily identified by one of ordinary skill in the art and based on the contents of the current disclosure (see FIG. 8). Nucleotides between identifiable and enumerated subcomponents reflect restriction enzyme recognition sites used in assembly (cloning) of the constructs, and in some cases, additional nucleotides do not convey any identifiable function. These segments of complete vector sequences can be adjusted based on use of different cloning strategies and/or vectors. In general, short 6-nucleotide palindromic sequences reflect vector construction artifacts that are not important to vector function.

[0085] In particular embodiments vectors (e.g., AAV) with capsids that cross the blood-brain barrier (BBB) are selected. In particular embodiments, vectors are modified to include capsids that cross the BBB. Examples of AAV with viral capsids that cross the blood brain barrier include AAV9 (Gombash *et al.*, Front Mol Neurosci. 2014; 7:81), AAVrh.10 (Yang, *et al.*, Mol Ther. 2014; 22(7): 1299-1309), AAV1R6, AAV1R7 (Albright *et al.*, Mol Ther. 2018; 26(2): 510), rAAVrh.8 (Yang, *et al.*, supra), AAV-BR1 (Marchio *et al.*, EMBO Mol Med. 2016; 8(6): 592), AAV-PHP.S (Chan *et al.*, Nat Neurosci. 2017; 20(8): 1172), AAV-PHP.B (Deverman *et al.*, Nat Biotechnol. 2016; 34(2): 204), AAV-PPS (Chen *et al.*, Nat Med. 2009; 15: 1215), and PHP.eB. In particular embodiments, the PHP.eB capsid differs from AAV9 such that, using AAV9 as a reference, amino

acids starting at residue 586: S-AQ-A (SEQ ID NO: 92) are changed to S-DGTLAVPFK-A (SEQ ID NO: 93). In particular embodiments, PHP.eb refers to SEQ ID NO: 56.

[0086] AAV9 is a naturally occurring AAV serotype that, unlike many other naturally occurring serotypes, can cross the BBB following intravenous injection. It transduces large sections of the central nervous system (CNS), thus permitting minimally invasive treatments (Naso *et al.*, BioDrugs. 2017; 31(4): 317), for example, as described in relation to clinical trials for the treatment of spinal muscular atrophy (SMA) syndrome by AveXis (AVXS-101, NCT03505099) and the treatment of CLN3 gene-Related Neuronal Ceroid-Lipofuscinosis (NCT03770572).

[0087] AAVrh.10, was originally isolated from rhesus macaques and shows low seropositivity in humans when compared with other common serotypes used for gene delivery applications (Selot *et al.*, Front Pharmacol. 2017; 8: 441) and has been evaluated in clinical trials LYS-SAF302, LYSOGENE, and NCT03612869.

[0088] AAV1R6 and AAV1R7, two variants isolated from a library of chimeric AAV vectors (AAV1 capsid domains swapped into AAVrh.10), retain the ability to cross the BBB and transduce the CNS while showing significantly reduced hepatic and vascular endothelial transduction.

[0089] rAAVrh.8, also isolated from rhesus macaques, shows a global transduction of glial and neuronal cell types in regions of clinical importance following peripheral administration and also displays reduced peripheral tissue tropism compared to other vectors.

[0090] AAV-BR1 is an AAV2 variant displaying the NRGTEWD (SEQ ID NO: 94) epitope that was isolated during *in vivo* screening of a random AAV display peptide library. It shows high specificity accompanied by high transgene expression in the brain with minimal off-target affinity (including for the liver) (Körbelin *et al.*, EMBO Mol Med. 2016; 8(6): 609).

[0091] AAV-PHP.S (Addgene, Watertown, MA) is a variant of AAV9 generated with the CREATE method that encodes the 7-mer sequence QAVRTSL (SEQ ID NO: 95), transduces neurons in the enteric nervous system, and strongly transduces peripheral sensory afferents entering the spinal cord and brain stem.

[0092] AAV-PHP.B (Addgene, Watertown, MA) is a variant of AAV9 generated with the CREATE method that encodes the 7-mer sequence TLAVPFK (SEQ ID NO: 96). It transfers genes throughout the CNS with higher efficiency than AAV9 and transduces the majority of astrocytes and neurons across multiple CNS regions.

[0093] AAV-PPS, an AAV2 variant created by insertion of the DSPAHPS (SEQ ID NO: 97) epitope into the capsid of AAV2, shows a dramatically improved brain tropism relative to AAV2.

[0094] For additional information regarding capsids that cross the blood brain barrier, see Chan *et al.*, Nat. Neurosci. 2017 Aug; 20(8): 1172-1179.

[0095] (ii) Compositions for Administration. Artificial expression constructs and vectors of the present disclosure (referred to herein as physiologically active components) can be formulated with a carrier that is suitable for administration to a cell, tissue slice, animal (e.g., mouse, non-human primate), or human. Physiologically active components within compositions described herein can be prepared in neutral forms, as freebases, or as pharmacologically acceptable salts.

[0096] Pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

[0097] Carriers of physiologically active components can include solvents, dispersion media, vehicles, coatings, diluents, isotonic and absorption delaying agents, buffers, solutions, suspensions, colloids, and the like. The use of such carriers for physiologically active components is well known in the art. Except insofar as any conventional media or agent is incompatible with the physiologically active components, it can be used with compositions as described herein.

[0098] The phrase "pharmaceutically-acceptable carriers" refer to carriers that do not produce an allergic or similar untoward reaction when administered to a human, and in particular embodiments, when administered intravenously (e.g., at the retro-orbital plexus).

[0099] In particular embodiments, compositions can be formulated for intravenous, intraparenchymal, intraocular, intravitreal, parenteral, subcutaneous, intracerebro-ventricular, intramuscular, intrathecal, intraspinal, intraperitoneal, oral or nasal inhalation, or by direct injection in or application to one or more cells, tissues, or organs.

[0100] Compositions may include liposomes, lipids, lipid complexes, microspheres, microparticles, nanospheres, and/or nanoparticles.

[0101] The formation and use of liposomes is generally known to those of skill in the art. Liposomes have been developed with improved serum stability and circulation half-times (see, for instance, U.S. Pat. No. 5,741,516). Further, various methods of liposome and liposome like preparations as potential drug carriers have been described (see, for instance U.S. Pat. Nos. 5,567,434; 5,552,157; 5,565,213; 5,738,868; and 5,795,587).

[0102] The disclosure also provides for pharmaceutically acceptable nanocapsule formulations of the physiologically active components. Nanocapsules can generally entrap compounds in a stable and reproducible way (Quintanar-Guerrero *et al.*, *Drug Dev Ind Pharm* 24(12):1113-1128, 1998; Quintanar-Guerrero *et al.*, *Pharm Res.* 15(7):1056-1062, 1998; Quintanar-Guerrero *et al.*,

J. Microencapsul. 15(1):107-119, 1998; Douglas *et al.*, *Crit Rev Ther Drug Carrier Syst* 3(3):233-261, 1987). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles can be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkylcyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present disclosure. Such particles can be easily made, as described in Couvreur *et al.*, *J Pharm Sci* 69(2):199-202, 1980; Couvreur *et al.*, *Crit Rev Ther Drug Carrier Syst.* 5(1)1-20, 1988; zur Muhlen *et al.*, *Eur J Pharm Biopharm*, 45(2):149-155, 1998; Zambaux *et al.*, *J Control Release* 50(1-3):31-40, 1998; and U.S. Pat. No. 5,145,684.

[0103] Injectable compositions can include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U.S. Pat. No. 5,466,468). For delivery via injection, the form is sterile and fluid to the extent that it can be delivered by syringe. In particular embodiments, it is stable under the conditions of manufacture and storage, and optionally contains one or more preservative compounds against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (e.g., glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion, and/or by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and/or antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In various embodiments, the preparation will include an isotonic agent(s), for example, sugar(s) or sodium chloride. Prolonged absorption of the injectable compositions can be accomplished by including in the compositions of agents that delay absorption, for example, aluminum monostearate and gelatin. Injectable compositions can be suitably buffered, if necessary, and the liquid diluent first rendered isotonic with sufficient saline or glucose.

[0104] Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. As indicated, under ordinary conditions of storage and use, these preparations can contain a preservative to prevent the growth of microorganisms.

[0105] Sterile compositions can be prepared by incorporating the physiologically active component in an appropriate amount of a solvent with other optional ingredients (e.g., as enumerated above), followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized physiologically active components into a sterile vehicle that contains the basic dispersion medium and the required other ingredients (e.g., from those

enumerated above). In the case of sterile powders for the preparation of sterile injectable solutions, preferred methods of preparation can be vacuum-drying and freeze-drying techniques which yield a powder of the physiologically active components plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0106] Oral compositions may be in liquid form, for example, as solutions, syrups or suspensions, or may be presented as a drug product for reconstitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., almond oil, oily esters, or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g., pregelatinized maize starch, polyvinyl pyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). Tablets may be coated by methods well-known in the art.

[0107] Inhalable compositions can be delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

[0108] Compositions can also include microchip devices (U.S. Pat. No. 5,797,898), ophthalmic formulations (Bourlais *et al.*, *Prog Retin Eye Res*, 17(1):33-58, 1998), transdermal matrices (U.S. Pat. No. 5,770,219 and U.S. Pat. No. 5,783,208) and feedback-controlled delivery (U.S. Pat. No. 5,697,899).

[0109] Supplementary active ingredients can also be incorporated into the compositions.

[0110] Typically, compositions can include at least 0.1% of the physiologically active components or more, although the percentage of the physiologically active components may, of course, be varied and may conveniently be between 1 or 2% and 70% or 80% or more or 0.5-99% of the weight or volume of the total composition. Naturally, the amount of physiologically active components in each physiologically-useful composition may be prepared in such a way that a

suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of compositions and dosages may be desirable.

[0111] In particular embodiments, for administration to humans, compositions should meet sterility, pyrogenicity, and the general safety and purity standards as required by United States Food and Drug Administration (FDA) or other applicable regulatory agencies in other countries.

[0112] (iii) Cell Lines Including Artificial Expression Constructs. The present disclosure includes cells including an artificial expression construct described herein. A cell that has been transformed with an artificial expression construct can be used for many purposes, including in neuroanatomical studies, assessments of functioning and/or non-functioning proteins, and drug screens that assess the regulatory properties of enhancers.

[0113] A variety of host cell lines can be used, but in particular embodiments, the cell is a mammalian cell. In particular embodiments, the artificial express construct includes an enhancer and/or a vector sequence of eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m and/or CN1259, CN1280, CN1521, CN1528, CN1533, CN1621, CN1674, CN1778, CN1932, CN2043, CN2045, CN2085, CN2102, CN2157, CN2216, CN2251, CN2257, CN2258, CN2267, CN2316, CN2339, CN2431, CN2436, CN2643, CN2663, CN2717, CN2838, CN1415, CN2710, CN2839, CN2674, CN3301, CN3019, CN3569, CN2374, CN3584, CN3003, CN2379, CN3566, CN3453, HCT1, CN2146, AiP1347, AiP1351, AiP1375, AiP1307, AiP1530, AiP1379, CN3916, CN3869, and/or CN3870 and the cell line is a human, primate, or murine cell. Cell lines which can be utilized for transgenesis in the present disclosure also include primary cell lines derived from living tissue such as rat or mouse brains and organotypic cell cultures, including brain slices from animals such as rats, mice, non-human primates, or human neurosurgical tissue. The PC12 cell line (available from the American Type Culture Collection, ATCC, Manassas, VA) has been shown

to express a number of neuronal marker proteins in response to Neuronal Growth Factor (NGF). The PC12 cell line is considered to be a neuronal cell line and is applicable for use with this disclosure. JAR cells (available from ATCC) are a platelet derived cell-line that express some neuronal genes, such as the serotonin transporter gene, and may be used with embodiments described herein.

[0114] WO 91/13150 describes a variety of cell lines, including neuronal cell lines, and methods of producing them. Similarly, WO 97/39117 describes a neuronal cell line and methods of producing such cell lines. The neuronal cell lines disclosed in these patent applications are applicable for use in the present disclosure.

[0115] In particular embodiments, "neuronal" describes something that is of, related to, or includes, neuronal cells. Neuronal cells are defined by the presence of an axon and dendrites. The term "neuronal-specific" refers to something that is found, or an activity that occurs, in neuronal cells or cells derived from neuronal cells, but is not found in or occur in, or is not found substantially in or occur substantially in, non-neuronal cells or cells not derived from neuronal cells, for example glial cells such as astrocytes or oligodendrocytes.

[0116] In particular embodiments, non-neuronal cell lines may be used, including mouse embryonic stem cells. Cultured mouse embryonic stem cells can be used to analyze expression of genetic constructs using transient transfection with plasmid constructs. Mouse embryonic stem cells are pluripotent and undifferentiated. These cells can be maintained in this undifferentiated state by Leukemia Inhibitory Factor (LIF). Withdrawal of LIF induces differentiation of the embryonic stem cells. In culture, the stem cells form a variety of differentiated cell types. Differentiation is caused by the expression of tissue specific transcription factors, allowing the function of an enhancer sequence to be evaluated. (See for example Fiskerstrand *et al.*, *FEBS Lett* 458: 171-174, 1999).

[0117] Methods to differentiate stem cells into neuronal cells include replacing a stem cell culture media with a media including basic fibroblast growth factor (bFGF) heparin, an N2 supplement (e.g., transferrin, insulin, progesterone, putrescine, and selenite), laminin and polyornithine. A process to produce myelinating oligodendrocytes from stem cells is described in Hu, *et al.*, 2009, *Nat. Protoc.* 4:1614-22. Bibel, *et al.*, 2007, *Nat. Protoc.* 2:1034-43 describes a protocol to produce glutamatergic neurons from stem cells while Chatzi, *et al.*, 2009, *Exp. Neurol.* 217:407-16 describes a procedure to produce GABAergic neurons. This procedure includes exposing stem cells to all-trans-RA for three days. After subsequent culture in serum-free neuronal induction medium including Neurobasal medium supplemented with B27, bFGF and EGF, 95% GABA neurons develop

[0118] U.S. Publication No. 2012/0329714 describes use of prolactin to increase neural stem cell numbers while U.S. Publication No. 2012/0308530 describes a culture surface with amino groups that promotes neuronal differentiation into neurons, astrocytes and oligodendrocytes. Thus, the fate of neural stem cells can be controlled by a variety of extracellular factors. Commonly used factors include brain derived growth factor (BDNF; Shetty and Turner, 1998, *J. Neurobiol.* 35:395-425); fibroblast growth factor (bFGF; U.S. Pat. No.5,766,948; FGF-1, FGF-2); Neurotrophin-3 (NT-3) and Neurotrophin-4 (NT-4); Caldwell, *et al.*, 2001, *Nat. Biotechnol.* 1;19:475-9); ciliary neurotrophic factor (CNTF); BMP-2 (U.S. Pat. Nos. 5,948,428 and 6,001,654); isobutyl 3-methylxanthine; leukemia inhibitory growth factor (LIF; U.S. Patent No. 6,103,530); somatostatin; amphiregulin; neurotrophins (*e.g.*, cyclic adenosine monophosphate; epidermal growth factor (EGF); dexamethasone (glucocorticoid hormone); forskolin; GDNF family receptor ligands; potassium; retinoic acid (U.S. Patent No. 6,395,546); tetanus toxin; and transforming growth factor- α and TGF- β (U.S. Pat. Nos. 5,851,832 and 5,753,506).

[0119] In particular embodiments, yeast one-hybrid systems may also be used to identify compounds that inhibit specific protein/DNA interactions, such as transcription factors for eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m.

[0120] Transgenic animals are described below. Cell lines may also be derived from such transgenic animals. For example, primary tissue culture from transgenic mice (*e.g.*, also as described below) can provide cell lines with the artificial expression construct already integrated into the genome. (for an example see MacKenzie & Quinn, *Proc Natl Acad Sci USA* 96: 15251-15255, 1999).

[0121] (iv) Transgenic Animals. Another aspect of the disclosure includes transgenic animals, the genome of which contains an artificial expression construct including eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m,

eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m operatively linked to a heterologous coding sequence. In particular embodiments, the genome of a transgenic animal includes CN1259, CN1280, CN1521, CN1528, CN1533, CN1621, CN1674, CN1778, CN1932, CN2043, CN2045, CN2085, CN2102, CN2157, CN2216, CN2251, CN2257, CN2258, CN2267, CN2316, CN2339, CN2431, CN2436, CN2643, CN2663, CN2717, CN2838, CN1415, CN2710, CN2839, CN2674, CN3301, CN3019, CN3569, CN2374, CN3584, CN3003, CN2379, CN3566, CN3453, HCT1, CN2146, AiP1347, AiP1351, AiP1375, AiP1307, AiP1530, AiP1379, CN3916, CN3869, and/or CN3870. In particular embodiments, when a non-integrating vector is utilized, a transgenic animal includes an artificial expression construct including eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m and/or CN1259, CN1280, CN1521, CN1528, CN1533, CN1621, CN1674, CN1778, CN1932, CN2043, CN2045, CN2085, CN2102, CN2157, CN2216, CN2251, CN2257, CN2258, CN2267, CN2316, CN2339, CN2431, CN2436, CN2643, CN2663, CN2717, CN2838, CN1415, CN2710, CN2839, CN2674, CN3301, CN3019, CN3569, CN2374, CN3584, CN3003, CN2379, CN3566, CN3453, HCT1, CN2146, AiP1347, AiP1351, AiP1375, AiP1307, AiP1530, AiP1379, CN3916, CN3869, and/or CN3870 within one or more of its cells.

[0122] Detailed methods for producing transgenic animals are described in U.S. Pat. No. 4,736,866. Transgenic animals may be of any nonhuman species, but preferably include nonhuman primates (NHPs), sheep, horses, cattle, pigs, goats, dogs, cats, rabbits, chickens, and rodents such as guinea pigs, hamsters, gerbils, rats, mice, and ferrets.

[0123] In particular embodiments, construction of a transgenic animal results in an organism that has an engineered construct present in all cells in the same genomic integration site. Thus, cell

lines derived from such transgenic animals will be consistent in as much as the engineered construct will be in the same genomic integration site in all cells and hence will suffer the same position effect variegation. In contrast, introducing genes into cell lines or primary cell cultures can give rise to heterologous expression of the construct. A disadvantage of this approach is that the expression of the introduced DNA may be affected by the specific genetic background of the host animal.

[0124] As indicated above in relation to cell lines, the artificial expression constructs of this disclosure can be used to genetically modify mouse embryonic stem cells using techniques known in the art. Typically, the artificial expression construct is introduced into cultured murine embryonic stem cells. Transformed ES cells are then injected into a blastocyst from a host mother and the host embryo re-implanted into the mother. This results in a chimeric mouse whose tissues are composed of cells derived from both the embryonic stem cells present in the cultured cell line and the embryonic stem cells present in the host embryo. Usually the mice from which the cultured ES cells used for transgenesis are derived are chosen to have a different coat color from the host mouse into whose embryos the transformed cells are to be injected. Chimeric mice will then have a variegated coat color. As long as the germ-line tissue is derived, at least in part, from the genetically modified cells, then the chimeric mice crossed with an appropriate strain can produce offspring that will carry the transgene.

[0125] In addition to the methods of delivery described above, the following techniques are also contemplated as alternative methods of delivering artificial expression constructs to target cells or targeted tissues and organs of an animal, and in particular, to cells, organs, or tissues of a vertebrate mammal: sonophoresis (e.g., ultrasound, as described in U.S. Pat. No. 5,656,016); intraosseous injection (U.S. Pat. No. 5,779,708); microchip devices (U.S. Pat. No. 5,797,898); ophthalmic formulations (Bourlais *et al.*, *Prog Retin Eye Res*, 17(1):33-58, 1998); transdermal matrices (U.S. Pat. No. 5,770,219 and U.S. Pat. No. 5,783,208); feedback-controlled delivery (U.S. Pat. No. 5,697,899), and any other delivery method available and/or described elsewhere in the disclosure.

[0126] (v) Methods of Use. In particular embodiments, a composition including a physiologically active component described herein is administered to a subject to result in a physiological effect.

[0127] In particular embodiments, the disclosure includes the use of the artificial expression constructs described herein to modulate expression of a heterologous gene which is either partially or wholly encoded in a location downstream to that enhancer in an engineered sequence. Thus, there are provided herein methods of use of the disclosed artificial expression constructs in the research, study, and potential development of medicaments for preventing, treating or

ameliorating the symptoms of a disease, dysfunction, or disorder.

[0128] Particular embodiments include methods of administering to a subject an artificial expression construct that includes eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m and/or CN1259, CN1280, CN1521, CN1528, CN1533, CN1621, CN1674, CN1778, CN1932, CN2043, CN2045, CN2085, CN2102, CN2157, CN2216, CN2251, CN2257, CN2258, CN2267, CN2316, CN2339, CN2431, CN2436, CN2643, CN2663, CN2717, CN2838, CN1415, CN2710, CN2839, CN2674, CN3301, CN3019, CN3569, CN2374, CN3584, CN3003, CN2379, CN3566, CN3453, HCT1, CN2146, AiP1347, AiP1351, AiP1375, AiP1307, AiP1530, AiP1379, CN3916, CN3869, and/or CN3870 as described herein to drive expression of a gene in a targeted cell type. The subject can be an isolated cell, a network of cells, a tissue slice, an experimental animal, a veterinary animal, or a human.

[0129] As is well known in the medical arts, dosages for any one subject depends upon many factors, including the subject's size, surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Dosages for the compounds of the disclosure will vary, but, in particular embodiments, a dose could be from 10^5 to 10^{100} copies of an artificial expression construct of the disclosure. In particular embodiments, a patient receiving intravenous, intraparenchymal, intraspinal, retro-orbital, or intrathecal administration can be infused with from 10^6 to 10^{22} copies of the artificial expression construct.

[0130] An "effective amount" is the amount of a composition necessary to result in a desired physiological change in the subject. Effective amounts are often administered for research purposes. Effective amounts disclosed herein can cause a statistically-significant effect in an animal model, human study, *in vivo*, or *in vitro* assay.

[0131] The amount of expression constructs and time of administration of such compositions will be within the purview of the skilled artisan having benefit of the present teachings. It is likely, however, that the administration of effective amounts of the disclosed compositions may be

achieved by a single administration, such as for example, a single injection of sufficient numbers of infectious particles to provide an effect in the subject. Alternatively, in some circumstances, it may be desirable to provide multiple, or successive administrations of the artificial expression construct compositions or other genetic constructs, either over a relatively short, or a relatively prolonged period of time, as may be determined by the individual overseeing the administration of such compositions. For example, the number of infectious particles administered to a mammal may be 10^7 , 10^8 , 10^9 , 10^{10} , 10^{11} , 10^{12} , 10^{13} , or even higher, infectious particles/ml given either as a single dose or divided into two or more administrations as may be required to achieve an intended effect. In fact, in certain embodiments, it may be desirable to administer two or more different expression constructs in combination to achieve a desired effect.

[0132] In certain circumstances it will be desirable to deliver the artificial expression construct in suitably formulated compositions disclosed herein either by pipette, retro-orbital injection, subcutaneously, intraocularly, intravitreally, parenterally, subcutaneously, intravenously, intraparenchymally, intracerebro-ventricularly, intramuscularly, intrathecally, intraspinally, intraperitoneally, by oral or nasal inhalation, or by direct application or injection to one or more cells, tissues, or organs. The methods of administration may also include those modalities as described in U.S. Pat. No. 5,543,158; U.S. Pat. No. 5,641,515 and U.S. Pat. No. 5,399,363.

[0133] (vi) Kits and Commercial Packages. Kits and commercial packages contain an artificial expression construct described herein. The artificial expression construct can be isolated. In particular embodiments, the components of an expression product can be isolated from each other. In particular embodiments, the expression product can be within a vector, within a viral vector, within a cell, within a tissue slice or sample, and/or within a transgenic animal. Such kits may further include one or more reagents, restriction enzymes, peptides, therapeutics, pharmaceutical compounds, or means for delivery of the compositions such as syringes, injectables, and the like.

[0134] Embodiments of a kit or commercial package will also contain instructions regarding use of the included components, for example, in basic research, electrophysiological research, neuroanatomical research, and/or the research and/or treatment of a disorder, disease or condition.

[0135] The Exemplary Embodiments below are included to demonstrate particular embodiments of the disclosure. Those of ordinary skill in the art should recognize in light of the present disclosure that many changes can be made to the specific embodiments disclosed herein and still obtain a like or similar result without departing from the spirit and scope of the disclosure.

[0136] (vii) Exemplary Embodiments.

1. A concatenated core of an eHGT_023h v3, eHGT_121h, eHGT_453m, eHGT_387m, eHGT_475m, eHGT_351h, and/or MGT_E116 enhancer.
2. The concatenated core of embodiment 2, wherein the concatenated core includes the sequence as set forth in SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 108, SEQ ID NO: 112, and/or SEQ ID NO: 121 or a sequence having at least 90% sequence identity to the sequence as set forth in SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 108, SEQ ID NO: 112, and/or SEQ ID NO: 121.
3. The concatenated core of embodiments 2, wherein the concatenated core includes 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core of eHGT_023h v3, eHGT_121h, eHGT_453m, eHGT_387m, eHGT_475m, eHGT_351h, and/or MGT_E116.
4. The concatenated core of any of embodiments 2-3, including 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 108, SEQ ID NO: 112, and/or SEQ ID NO: 121 or a sequence having at least 90% sequence identity to the sequence as set forth in SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 108, SEQ ID NO: 112, and/or SEQ ID NO: 121.
5. The concatenated core of any of embodiments 2-4, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 25.
6. The concatenated core of any of embodiments 2-5, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 27.
7. The concatenated core of any of embodiments 2-6, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 103.
8. The concatenated core of any of embodiments 2-7, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 105.
9. The concatenated core of any of embodiments 2-8, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 108.
10. The concatenated core of any of embodiments 2-9, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 112.
11. The concatenated core of any of embodiments 2-10, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 121.
12. The concatenated core of embodiment 5, having 3 copies of the sequence as set forth in SEQ ID NO: 25.
13. The concatenated core of embodiment 6, having 3 copies of the sequence as set forth in SEQ ID NO: 27.
14. The concatenated core of embodiment 7, having 3 copies of the sequence as set forth in SEQ

ID NO: 103.

15. The concatenated core of embodiment 8, having 3 copies of the sequence as set forth in SEQ ID NO: 105.

16. The concatenated core of embodiment 9, having 3 copies of the sequence as set forth in SEQ ID NO: 108.

17. The concatenated core of embodiment 10, having 3 copies of the sequence as set forth in SEQ ID NO: 112.

18. The concatenated core of embodiment 11, having 3 copies of the sequence as set forth in SEQ ID NO: 121.

19. The concatenated core of embodiment 12, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 26.

20. The concatenated core of embodiment 13, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 28.

21. The concatenated core of embodiment 14, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 104.

22. The concatenated core of embodiment 15, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 106.

23. The concatenated core of embodiment 16, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 109.

24. The concatenated core of embodiment 17, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 113.

25. The concatenated core of embodiment 18, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 122.

26. An artificial expression construct including (i) an enhancer selected from eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m (ii) a promoter; and (iii) a heterologous coding sequence.

27. The artificial expression construct of embodiment 26, wherein the heterologous coding sequence encodes an effector element or an expressible element.
28. The artificial expression construct of embodiment 27, wherein the effector element includes a reporter protein or a functional molecule.
29. The artificial expression construct of embodiment 28, wherein the reporter protein includes a fluorescent protein.
30. The artificial expression construct of embodiment 28, wherein the functional molecule includes a functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or a designer receptor exclusively activated by designer drug (DREADD).
31. The artificial expression construct of embodiment 27, wherein the expressible element includes a non-functional molecule.
32. The artificial expression construct of embodiment 31, wherein the non-functional molecule includes a non-functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or DREADD.
33. The artificial expression construct of any of embodiments 26-32, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.
34. The artificial expression construct of embodiment 33, wherein the capsid includes PHP.eB, AAV-BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.
35. The artificial expression construct of any of embodiments 26-34, wherein the artificial expression construct includes or encodes a skipping element.
36. The artificial expression construct of embodiment 35, wherein the skipping element includes a 2A peptide or an internal ribosome entry site (IRES).
37. The artificial expression construct of embodiment 36, wherein the 2A peptide includes T2A, P2A, E2A, or F2A.
38. The artificial expression construct of any of embodiments 26-37, wherein the artificial expression construct includes or encodes a set of features selected from: eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m,

eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, eHGT_1027m, AAV, scAAV, rAAV, pAAV, minBglobin, CMV, minCMV, minCMV*, minRho, minRho*, fluorescent protein, 10aa, H2B, H2Bmod, H2B*, hsA2, Cre, iCre, dgCre, FlpO, tTA2, SP10, tag cassette nuclear localization protein, self-cleaving peptides, WPRE, WPRE3, hGHpA, and/or BGHpA.

39. The artificial expression construct of any of embodiments 26-38, wherein the artificial expression construct includes or encodes a set of features selected from:

eHGT_023h v1-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_023m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 hsA2-3xCore-eHGT_023h v3-minRho-[heterologous coding sequence]-WPRE3-BGHpA;
 hsA2-eHGT_082h-minRho-[heterologous coding sequence]-WPRE3-BGHpA;
 hsA2-eHGT_087h-minRho-[heterologous coding sequence]-WPRE3-BGHpA;
 hsA2-eHGT_128h-minRho-[heterologous coding sequence]-WPRE3-BGHpA;
 hsA2-eHGT_181h-minRho-[heterologous coding sequence]-WPRE3-BGHpA;
 hsA2-eHGT_260h-minRho-[heterologous coding sequence]-WPRE3-BGHpA;
 hsA2-eHGT_023h v2-minRho-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_356h-minRho*-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_359h-minRho*-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_375m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_387m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_395h-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 3x(coreB)eHGT121h-minCMV*-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_453m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_470m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_479m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_494m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_467m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_023h v1-minRho*-[heterologous coding sequence]-10aa-H2Bmod-WPRE3-BGHpA;

eHGT_023h v1-minRho*-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_483m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_606h-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_738m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_796h-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_710m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_588m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_007m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_703m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_589m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_086h-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_963m-minBG-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRES-BGHPA;
eHGT_963m-minBG-[heterologous coding sequence]-WPRES-BGHPA;
3xcore3_eHGT_453m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
3xcore2_eHGT_387m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_534h-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
3xcore2_eHGT_475m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_830h-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRES-BGHPA;
eHGT_830h-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_540h-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
3xcore2_eHGT_351h-minRho*-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_882m-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRES-BGHPA;
eHGT_882m-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_1137m-minBglobin-[heterologous coding sequence]-WPRES-bGHPA;
eHGT_381h-minBglobin-[heterologous coding sequence]-WPRES-BGHPA;
MGT_E118-minBglobin-[heterologous coding sequence]-WPRES-bGHPA;
MGT_E122-minBglobin-[heterologous coding sequence]-WPRES-bGHPA;
MGT_E146-minBglobin-[heterologous coding sequence]-WPRES-bGHPA;

MGT_E16-minBglobin-[heterologous coding sequence]-WPRES3-bGHPA;
 3xCore_MGT_E116-minBglobin-[heterologous coding sequence]-WPRES3-bGHPA;
 MGT_E150-minBglobin-[heterologous coding sequence]-WPRES3-bGHPA;
 eHGT_1032h-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRES3-BGHPA;
 eHGT_1032h-minBglobin-[heterologous coding sequence]-WPRES3-BGHPA;
 eHGT_1027h-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B*-WPRES3-BGHPA;
 eHGT_1027h-minBglobin-[heterologous coding sequence]-WPRES3-BGHPA;
 eHGT_1027m-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B*-WPRES3-BGHPA;
 eHGT_1027m-minBglobin-[heterologous coding sequence]-WPRES3-BGHPA;
 eHGT_023h v1-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_023m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-3xCore-eHGT_023h v3-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-eHGT_082h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-eHGT_087h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-eHGT_128h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-eHGT_181h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-eHGT_260h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-eHGT_023h v2-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_356h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_359h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_375m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_387m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_395h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

3x(coreB)eHGT121h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_453m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_470m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_479m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_494m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_467m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_023h v1-[minimal promoter]-[heterologous coding sequence]-10aa-H2Bmod-[post regulatory elements];

eHGT_483m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_606h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_738m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_796h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_710m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_588m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_007m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_703m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_589m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_086h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_963m-minBG-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-[post regulatory elements];

eHGT_963m-minBG-[heterologous coding sequence]-[post regulatory elements];

3xcore3_eHGT_453m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

3xcore2_eHGT_387m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_534h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

3xcore2_eHGT_475m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_830h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-[post regulatory elements];

eHGT_830h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

- eHGT_540h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- 3xcore2_eHGT_351h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- eHGT_882m-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-[post regulatory elements];
- eHGT_882m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- eHGT_1137m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- eHGT_381h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- MGT_E118-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- MGT_E122-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- MGT_E146-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- MGT_E16-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- 3xCore_MGT_E116-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- MGT_E150-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- eHGT_1032h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-[post regulatory elements];
- eHGT_1032h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- eHGT_1027h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B*-[post regulatory elements];
- eHGT_1027h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
- eHGT_1027m-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B*-[post regulatory elements]; or
- eHGT_1027m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements].
40. A vector including an artificial expression construct of any of embodiments 26-39.
41. The vector of embodiment 40, wherein the vector includes a viral vector.
42. The vector of embodiment 41, wherein the viral vector includes a recombinant adeno-associated viral (AAV) vector.
43. An adeno-associated viral (AAV) vector including at least one heterologous coding sequence, wherein the heterologous coding sequence is under the transcriptional control of a promoter and an enhancer selected from eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m,

eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and eHGT_1027m.

44. The AAV vector of embodiment 43, wherein the heterologous coding sequence encodes an effector element or an expressible element.

45. The AAV vector of embodiment 44, wherein the effector element includes a reporter protein or a functional molecule.

46. The AAV vector of embodiment 45, wherein the reporter protein includes a fluorescent protein.

47. The AAV vector of embodiment 45, wherein the functional molecule includes a functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or DREADD.

48. The AAV vector of embodiment 44, wherein the expressible element includes a non-functional molecule.

49. The AAV vector of embodiment 48, wherein the non-functional molecule includes a non-functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or DREADD.

50. A transgenic cell including an artificial expression construct or vector of any of the preceding embodiments.

51. The transgenic cell of embodiment 50, wherein the transgenic cell includes a cerebellar cell, GABAergic neuron, glutamatergic neuron, non-neuronal cell, cholinergic interneuron, medium spin neuron (MSN), and/or spinal motor neuron.

52. The transgenic cell of embodiment 51, wherein the cerebellar cell includes a Purkinje cell, molecular layer interneuron (MLI) cell, deep cerebellar nucleus (DCN) cell, Bergman glia, granule cell, mossy fiber (MF) cell, foliar white matter (VWM), and/or oligodendrocyte.

53. The transgenic cell of embodiments 51 or 52, wherein the GABAergic neuron includes a Pvalb neuron, Sst neuron, Lamp5 neuron, Vip neuron, medial ganglionic eminence (MGE) (Sst/Pvalb) cell, and/or pan-GABAergic neuron.

54. The transgenic cell of embodiment 53, wherein the Pvalb neuron includes a chandelier cell.
55. The transgenic cell of embodiment 53, wherein the Sst neuron includes an Sst Chodl cell.
56. The transgenic cell of any of embodiments 51-55, wherein the glutamatergic neuron includes a pan-glutamatergic neuron, layer 5 glutamatergic extraencephalic (L5 ET) neuron, L5 near-projecting (NP), L5 intratelencephalic (IT), and/or glutamatergic neuron in the thalamus
57. The transgenic cell of any of embodiments 51-56, wherein the non-neuronal cell includes an astrocyte, oligodendrocyte, pericyte, or specialized smooth-muscle cells (SMCs).
58. The transgenic cell of any of embodiments 50-57, wherein the transgenic cell is murine, human, or non-human primate.
59. A non-human transgenic animal including an artificial expression construct, vector, or transgenic cell of any of the preceding embodiments.
60. The non-human transgenic animal of embodiment 59, wherein the non-human transgenic animal is a mouse or a non-human primate.
61. An administrable composition including an artificial expression construct, vector, and/or a transgenic cell of any of the preceding embodiments.
62. A kit including an artificial expression construct, vector, a transgenic cell, non-human transgenic animal, and/or administrable compositions of any of the preceding embodiments.
63. A method for expressing a gene within a population of cerebellar cell types and a population of secondary cell types *in vivo* or *in vitro*, the method including providing the administrable composition of embodiment 61 in a sufficient dosage and for a sufficient time to a sample or subject including the population of cells thereby expressing the gene within the population of cells.
64. A method for expressing a heterologous gene within a population of cerebellar cell types and a population of secondary cell types *in vivo* or *in vitro*, the method including providing an administrable composition in a sufficient dosage and for a sufficient time to a sample or subject including the population of cells thereby expressing the gene within the population of cells wherein the composition includes an artificial expression construct including (i) an enhancer selected from eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16,

3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m; (ii) a promoter; and (iii) a heterologous coding sequence.

65. The method of embodiment 64, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.

66. The method of embodiment 65, wherein the capsid includes PHP.eB, AAV-BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.

67. The method of any of embodiments 64-66, wherein the artificial expression construct includes or encodes a skipping element.

68. The method of embodiment 67, wherein the skipping element includes a 2A peptide and/or an internal ribosome entry site (IRES).

69. The method of embodiment 68, wherein the 2A peptide includes T2A, P2A, E2A, or F2A.

70. The method of any of embodiments 64-69, wherein the artificial expression construct includes or encodes a set of features selected from: eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_087h, eHGT_128h, eHGT_181h, eHGT_260h, eHGT_023h v2, eHGT_356h, eHGT_359h, eHGT_375m, eHGT_387m, eHGT_395h, 3x(CoreB)eHGT_121h, eHGT_453m, eHGT_470m, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_483m, eHGT_606h, eHGT_738m, eHGT_796h, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_589m, eHGT_086h, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_387m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_830h, eHGT_540h, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, eHGT_381h, MGT_E118, MGT_E122, MGT_E146, MGT_E16, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, eHGT_1027m, hsA2, AAV, scAAV, rAAV, minBglobin, CMV, minCMV, minCMV*, minRho, minRho*, fluorescent protein (e.g., EGFP, SYFP, GFP), Cre, iCre, dgCre, FlpO, tTA2, SP10 (e.g., 3xSP10), 10aa, H2B, H2Bmod, H2B*, WPRE, WPRE3, hGHpA and/or BGHpA.

71. The method of any of embodiments 64-70, wherein the gene encodes an effector element or an expressible element

72. The method of embodiment 71, wherein the effector element includes a reporter protein or a functional molecule.

73. The method of embodiment 72, wherein the reporter protein includes a fluorescent protein.

74. The method of embodiment 72, wherein the functional molecule includes a functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or DREADD.

75. The method of embodiment 71, wherein the expressible element includes a non-functional molecule.
76. The method of embodiment 75, wherein the non-functional molecule includes a non-functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or DREADD.
77. The method of any of embodiments 64-76, wherein the providing includes pipetting.
78. The method of embodiment 77, wherein the pipetting is to a brain slice.
79. The method of embodiment 78, wherein the brain slice includes a cerebellar cell and/or a secondary cell type.
80. The method of embodiment 79, wherein the cerebellar cell includes a Purkinje cell, MLI cell, DCN cell, Bergman glia, granule cell, MF cell, foliar white matter (WM), and/or oligodendrocyte.
81. The method of embodiment 79, wherein the secondary cell type includes a GABAergic neuron, glutamatergic neuron, non-neuronal cell, cholinergic interneurons, MSN, or spinal motor neurons.
82. The method of embodiment 81, wherein the GABAergic neuron includes a Pvalb neuron, Sst neuron, Lamp5 neuron, Vip neuron, MGE (Sst/Pvalb) cell, and/or pan-GABAergic neuron.
83. The method of embodiment 82, wherein the Pvalb neuron includes a chandelier cell.
84. The method of embodiment 82, wherein the Sst neuron includes an Sst Chodl cell.
85. The method of embodiment 81, wherein the glutamatergic neuron includes a pan-glutamatergic neuron, L5 ET neuron, L5 NP, L5 IT, and/or glutamatergic neuron in the thalamus.
86. The method of embodiment 81, wherein the non-neuronal cell includes an astrocyte, oligodendrocyte, pericyte, or specialized SMCs.
87. The method of any of embodiments 79-86, wherein the cerebellar cell includes Purkinje cells and the enhancer is eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_087h, eHGT_260h, eHGT_023h v2, eHGT_359h, eHGT_479m, eHGT_494m, eHGT_467m, eHGT_606h, eHGT_796h, eHGT_588m, eHGT_534h, 3xcore2_eHGT_475m, eHGT_540h, or MGT_E150.
88. The method of any of embodiments 79-87, wherein the cerebellar cell includes MLI cells and the enhancer is 3xCore-eHGT_023h v3, eHGT_181h, eHGT_023h v2, 3x(CoreB)eHGT_121h, eHGT_479m, eHGT_738m, eHGT_710m, eHGT_588m, eHGT_007m, eHGT_703m, eHGT_086h, eHGT_963m, eHGT_830h, MGT_E146, MGT_E16, eHGT_1032h, eHGT_1027h, or eHGT_1027m.

89. The method embodiment 88, wherein the MLI cells are basket cells and the enhancer is eHGT_086h.
90. The method of any of embodiments 79-89, wherein the cerebellar cell includes DCN cells and the enhancer is eHGT_082h, eHGT_128h, eHGT_356h, eHGT_395h, eHGT_453m, eHGT_470m, eHGT_589m, eHGT_963m, 3xcore3_eHGT_453m, 3xcore2_eHGT_351h, eHGT_882m, eHGT_1137m, MGT_E118, or 3xcore_MGT_E116.
91. The method of any of embodiments 79-90, wherein the cerebellar cell includes Bergman glia and the enhancer is eHGT_375m, eHGT_387m, 3xcore2_eHGT_387m, eHGT_381h, or MGT_E122.
92. The method of any of embodiments 79-91, wherein the cerebellar cell includes granule and MF cells and the enhancer is eHGT_483m.
93. The method of any of embodiments 79-92, wherein the cerebellar cell includes oligodendrocytes, DNC, and/or foliar white matter (WM) and the enhancer is eHGT_395h.
94. The method of any of embodiments 79-93, wherein the secondary cell type is a Pvalb interneuron in the neocortex and the enhancer is eHGT_023h v1, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_082h, eHGT_128h, eHGT_023h v2, eHGT_359h, 3x(CoreB)eHGT_121h, eHGT_589m, or eHGT_086h.
95. The method of any of embodiments 79-94, wherein the secondary cell type is MGE cells in the neocortex and the enhancer is eHGT_588m.
96. The method of any of embodiments 79-95, wherein the secondary cell type is Sst/Chold neurons in the neocortex and the enhancer is eHGT_470m or eHGT_467m.
97. The method of any of embodiments 79-96, wherein the secondary cell type is Vip interneurons in the neocortex and the enhancer is eHGT_356h, eHGT_483m, eHGT_007m, or eHGT_494m.
98. The method of any of embodiments 79-97, wherein the secondary cell type is Sst interneurons in the neocortex and the enhancer is eHGT_087h.
99. The method of any of embodiments 79-98, wherein the secondary cell type is Lamp5 interneurons in the neocortex and the enhancer is eHGT_181h.
100. The method of any of embodiments 79-99, wherein the secondary cell type is pan-GABAergic neurons in the neocortex and the enhancer is eHGT_796h.
101. The method of any of embodiments 79-100, wherein the secondary cell type is L5 ET neurons in the neocortex and the enhancer is eHGT_453m.
102. The method of any of embodiments 79-101, wherein the secondary cell type is L5 NP neurons in the neocortex and the enhancer is 3xcore_MGT_E116.
103. The method of any of embodiments 79-102, wherein the secondary cell type is L5 IT

neurons in the neocortex and the enhancer is MGT_E16.

104. The method of any of embodiments 79-103, wherein the secondary cell type is pan-glutamatergic neurons in the neocortex and the enhancer is eHGT_260h.

105. The method of any of embodiments 79-104, wherein the secondary cell type is chandelier cells in the neocortex and the enhancer is eHGT_479m, eHGT_494m, 3xcore2_eHGT_475m, eHGT_710m, or eHGT_703m.

106. The method of any of embodiments 79-105, wherein the secondary cell type is astrocytes in the neocortex and the enhancer is 3xcore2_eHGT_387m, eHGT_381h, MGT_E118, or MGT_E122.

107. The method of any of embodiments 79-106, wherein the secondary cell type is pericytes in the neocortex and the enhancer is MGT_E146.

108. The method of any of embodiments 79-107, wherein the secondary cell type is specialized SMCs in the neocortex and the enhancer is MGT_E150.

109. The method of any of embodiments 79-108, wherein the secondary cell type is cholinergic interneurons in the striatum and the enhancer is eHGT_738m.

110. The method of any of embodiments 79-109, wherein the secondary cell type is MSN of the striatum and the enhancer is eHGT_882m.

111. The method of any of embodiments 79-110, wherein the secondary cell type is pan-MSN cells in the striatum and the enhancer is 3xcore2_eHGT_351h.

112. The method of any of embodiments 79-111, wherein the secondary cell type is glutamatergic neurons in the thalamus and the enhancer is eHGT_606h or eHGT_830h.

113. The method of any of embodiments 79-112, wherein the secondary cell type is astrocytes brain wide and the enhancer is eHGT_375m or eHGT_387m.

114. The method of any of embodiments 79-113, wherein the secondary cell type is oligodendrocyte brain wide and the enhancer is eHGT_395h.

115. The method of any of embodiments 79-114, wherein the secondary cell type is pan-GABAergic neurons brain wide and the enhancer is eHGT_1032h, eHGT_1027h, or eHGT_1027m.

116. The method of any of embodiments 79-115, wherein the secondary cell type is pan-neuronal cells brain wide and the enhancer is eHGT_534h or eHGT_540h.

117. The method of any of embodiments 79-116, wherein the secondary cell type is spinal motor neurons and the enhancer is eHGT_1137m.

118. The method of any of embodiments 78-117, wherein the brain slice is murine, human, or non-human primate.

119. The method of any of embodiments 77-118, wherein the providing includes administering to a living subject.

120. The method of embodiment 119, wherein the living subject is a human, non-human primate, or a mouse.

121. The method of embodiments 119 or 120, wherein the administering to a living subject is through injection.

122. The method of embodiment 121, wherein the injection includes intravenous injection, intraparenchymal injection into brain tissue, intracerebroventricular (ICV) injection, intra-cisterna magna (ICM) injection, or intrathecal injection.

123. An artificial expression construct including CN1259, CN1280, CN1521, CN1528, CN1533, CN1621, CN1674, CN1778, CN1932, CN2043, CN2045, CN2085, CN2102, CN2157, CN2216, CN2251, CN2257, CN2258, CN2267, CN2316, CN2339, CN2431, CN2436, CN2643, CN2663, CN2717, CN2838, CN1415, CN2710, CN2839, CN2674, CN3301, CN3019, CN3569, CN2374, CN3584, CN3003, CN2379, CN3566, CN3453, HCT1, CN2146, AiP1347, AiP1351, AiP1375, AiP1307, AiP1530, AiP1379, CN3916, CN3869, or CN3870.

[0137] (viii) Closing Paragraphs. Variants of the sequences disclosed and referenced herein are also included. Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological activity can be found using computer programs well known in the art, such as DNASTAR™ (Madison, Wisconsin) software. Preferably, amino acid changes in the protein variants disclosed herein are conservative amino acid changes, i.e., substitutions of similarly charged or uncharged amino acids. A conservative amino acid change involves substitution of one of a family of amino acids which are related in their side chains.

[0138] In a peptide or protein, suitable conservative substitutions of amino acids are known to those of skill in this art and generally can be made without altering a biological activity of a resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, e.g., Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Benjamin/Cummings Pub. Co., p. 224). Naturally occurring amino acids are generally divided into conservative substitution families as follows: Group 1: Alanine (Ala), Glycine (Gly), Serine (Ser), and Threonine (Thr); Group 2: (acidic): Aspartic acid (Asp), and Glutamic acid (Glu); Group 3: (acidic; also classified as polar, negatively charged residues and their amides): Asparagine (Asn), Glutamine (Gln), Asp, and Glu; Group 4: Gln and Asn; Group 5: (basic; also classified as polar, positively charged residues): Arginine (Arg), Lysine (Lys), and Histidine (His); Group 6 (large aliphatic, nonpolar residues): Isoleucine (Ile), Leucine (Leu), Methionine (Met), Valine (Val) and

Cysteine (Cys); Group 7 (uncharged polar): Tyrosine (Tyr), Gly, Asn, Gln, Cys, Ser, and Thr; Group 8 (large aromatic residues): Phenylalanine (Phe), Tryptophan (Trp), and Tyr; Group 9 (non-polar): Proline (Pro), Ala, Val, Leu, Ile, Phe, Met, and Trp; Group 11 (aliphatic): Gly, Ala, Val, Leu, and Ile; Group 10 (small aliphatic, nonpolar or slightly polar residues): Ala, Ser, Thr, Pro, and Gly; and Group 12 (sulfur-containing): Met and Cys. Additional information can be found in Creighton (1984) *Proteins*, W.H. Freeman and Company.

[0139] In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, *J. Mol. Biol.* 157(1), 105-32). Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are: Ile (+4.5); Val (+4.2); Leu (+3.8); Phe (+2.8); Cys (+2.5); Met (+1.9); Ala (+1.8); Gly (−0.4); Thr (−0.7); Ser (−0.8); Trp (−0.9); Tyr (−1.3); Pro (−1.6); His (−3.2); Glutamate (−3.5); Gln (−3.5); aspartate (−3.5); Asn (−3.5); Lys (−3.9); and Arg (−4.5).

[0140] It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, i.e., still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity.

[0141] As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: Arg (+3.0); Lys (+3.0); aspartate (+3.0 $\pm$ 1); glutamate (+3.0 $\pm$ 1); Ser (+0.3); Asn (+0.2); Gln (+0.2); Gly (0); Thr (−0.4); Pro (−0.5 $\pm$ 1); Ala (−0.5); His (−0.5); Cys (−1.0); Met (−1.3); Val (−1.5); Leu (−1.8); Ile (−1.8); Tyr (−2.3); Phe (−2.5); Trp (−3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

[0142] As outlined above, amino acid substitutions may be based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like.

[0143] As indicated elsewhere, variants of gene sequences can include codon optimized variants, sequence polymorphisms, splice variants, and/or mutations that do not affect the function of an encoded product to a statistically-significant degree.

[0144] Variants of the protein, nucleic acid, and gene sequences disclosed herein also include sequences with at least 70% sequence identity, 80% sequence identity, 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the protein, nucleic acid, or gene sequences disclosed herein.

[0145] "% sequence identity" refers to a relationship between two or more sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between protein, nucleic acid, or gene sequences as determined by the match between strings of such sequences. "Identity" (often referred to as "similarity") can be readily calculated by known methods, including those described in: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, NY (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, NY (1994); Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, NJ (1994); Sequence Analysis in Molecular Biology (Von Heijne, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Oxford University Press, NY (1992). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR, Inc., Madison, Wisconsin). Multiple alignment of the sequences can also be performed using the Clustal method of alignment (Higgins and Sharp CABIOS, 5, 151-153 (1989) with default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Relevant programs also include the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, Wisconsin); BLASTP, BLASTN, BLASTX (Altschul, *et al.*, J. Mol. Biol. 215:403-410 (1990); DNASTAR (DNASTAR, Inc., Madison, Wisconsin); and the FASTA program incorporating the Smith-Waterman algorithm (Pearson, Comput. Methods Genome Res., [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, N.Y. Within the context of this disclosure it will be understood that where sequence analysis software is used for analysis, the results of the analysis are based on the "default values" of the program referenced. As used herein "default values" will mean any set of values or parameters, which originally load with the software when first initialized.

[0146] Variants also include nucleic acid molecules that hybridizes under stringent hybridization conditions to a sequence disclosed herein and provide the same function as the reference sequence. Exemplary stringent hybridization conditions include an overnight incubation at 42 °C in a solution including 50% formamide, 5XSSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5XDenhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1XSSC at 50 °C. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, moderately high stringency conditions include an overnight incubation at 37°C in a solution including 6XSSPE (20XSSPE=3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50 °C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g., 5XSSC). Variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

[0147] The term concatenate is broadly used to describe linking together into a chain or series. It is used to describe the linking together of nucleotide or amino acid sequences into a single nucleotide or amino acid sequence, respectively. The term "concatamerize" should be interpreted to recite: "concatenate."

[0148] As will be understood by one of ordinary skill in the art, each embodiment disclosed herein can comprise, consist essentially of or consist of its particular stated element, step, ingredient or component. Thus, the terms "include" or "including" should be interpreted to recite: "comprise, consist of, or consist essentially of." The transition term "comprise" or "comprises" means has, but is not limited to, and allows for the inclusion of unspecified elements, steps, ingredients, or components, even in major amounts. The transitional phrase "consisting of" excludes any element, step, ingredient or component not specified. The transition phrase "consisting essentially of" limits the scope of the embodiment to the specified elements, steps, ingredients or components and to those that do not materially affect the embodiment. A material effect would cause a statistically significant reduction in targeted expression in the targeted cell population as determined by scRNA-Seq and the following enhancer / targeted cell population and cerebellum

subclass pairings:

eHGT_023h v1 / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_023m / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_082h / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_087h / Purkinje cells in the cerebellum and Sst interneurons in the neocortex;
 eHGT_128h / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_181h / MLI cells in the cerebellum and Lamp5 interneurons in the neocortex;
 eHGT_260h / Purkinje cells in the cerebellum and pan-glutamatergic neurons in the neocortex;
 eHGT_023h v2 / Purkinje and MLI cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_356h / DCN cells in the cerebellum and Vip interneurons in the neocortex;
 eHGT_359h / Purkinje cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_375m / Bergmann glia in the cerebellum and astrocytes brain wide;
 eHGT_387m / Bergmann glia in the cerebellum and astrocytes brain wide;
 eHGT_395h / oligodendrocytes, DCN, and foliar white matter (WM) in the cerebellum and oligodendrocytes brain wide;
 eHGT_453m / DCN cells in the cerebellum and L5 ET neurons in the neocortex;
 eHGT_470m / DCN cells in the cerebellum and Sst/Chodl neurons in the neocortex;
 eHGT_479m / Purkinje and MLI cells in the cerebellum and chandelier cells in the neocortex;
 eHGT_494m / Purkinje cells in the cerebellum and Vip and chandelier cells in the neocortex;
 eHGT_467m / Purkinje cells in the cerebellum and Sst/Chodl neurons in the neocortex;
 eHGT_483m / Granule and mossy fiber (MF) cells in the cerebellum and Vip interneurons in the neocortex;
 eHGT_606h / Purkinje cells in the cerebellum and glutamatergic neurons in the thalamus;
 eHGT_738m / MLI cells in the cerebellum and cholinergic interneurons in the striatum;
 eHGT_796h / Purkinje cells in the cerebellum and pan-GABAergic neurons in the neocortex;
 eHGT_710m / MLI cells in the cerebellum and chandelier cells in the neocortex;
 eHGT_588m / Purkinje and MLI cells in the cerebellum and MGE cells in the neocortex;
 eHGT_007m / MLI cells in the cerebellum and Vip interneurons in the neocortex;
 eHGT_703m / MLI cells in the cerebellum and chandelier cells in the neocortex;
 eHGT_589m / DCN cells in the cerebellum and Pvalb interneurons in the neocortex;
 eHGT_086h / MLI (basket) cells in the cerebellum and Pvalb interneurons in the neocortex;

eHGT_963m / MLI and DCN cells in the cerebellum and L5 NP neurons in the neocortex;
eHGT_534h / Purkinje cells in the cerebellum and pan-neuronal cells brain wide;
eHGT_830h / MLI cells in the cerebellum and glutamatergic neurons in the thalamus;
eHGT_540h / Purkinje cells in the cerebellum and pan-neuronal cells brain wide;
eHGT_882m / DCN cells in the cerebellum and MSN cells in the striatum;
eHGT_1137m / DCN cells in the cerebellum and spinal motor neurons;
eHGT_381h / Bergmann glia in the cerebellum and astrocytes in the neocortex;
MGT_E118 / DCN cells in the cerebellum and astrocytes in the neocortex;
MGT_E122 / Bergmann glia in the cerebellum and astrocytes in the neocortex;
MGT_E146 / MLI cells in the cerebellum and pericytes in the neocortex;
MGT_E16 / MLI cells in the cerebellum and L5 IT neurons in the neocortex;
MGT_E150 / Purkinje cells in the cerebellum and specialized smooth-muscle cells (SMCs) in the neocortex;
eHGT_1032h / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
eHGT_1027h / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
eHGT_1027m / MLI cells in the cerebellum and pan-GABAergic neurons brain wide;
concatenated copies of an enhancer or concatenated copies of an enhancer core including
3xCore-eHGT_023h v3 / Purkinje and MLI cells in the cerebellum and Pvalb interneurons in the neocortex;
3x(CoreB)eHGT_121h / MLI cells in the cerebellum and Pvalb interneurons in the neocortex;
3xcore3_eHGT_453m / DCN cells in the cerebellum;
3xcore2_eHGT_387m / Bergmann glia in the cerebellum and astrocytes in the neocortex;
3xcore2_eHGT_475m / Purkinje cells in the cerebellum and chandelier cells in the neocortex;
3xcore2_eHGT_351h / DCN cells in the cerebellum and pan-MSN cells in the striatum; and
3xcore_MGT_E116 / DCN cells in the cerebellum and L5 NP neurons in the neocortex.

[0149] In particular embodiments, artificial means not naturally occurring.

[0150] Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term “about.” Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and attached claims are approximations that may vary depending upon the desired properties sought to be

obtained by the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. When further clarity is required, the term “about” has the meaning reasonably ascribed to it by a person skilled in the art when used in conjunction with a stated numerical value or range, i.e. denoting somewhat more or somewhat less than the stated value or range, to within a range of $\pm 20\%$ of the stated value; $\pm 19\%$ of the stated value; $\pm 18\%$ of the stated value; $\pm 17\%$ of the stated value; $\pm 16\%$ of the stated value; $\pm 15\%$ of the stated value; $\pm 14\%$ of the stated value; $\pm 13\%$ of the stated value; $\pm 12\%$ of the stated value; $\pm 11\%$ of the stated value; $\pm 10\%$ of the stated value; $\pm 9\%$ of the stated value; $\pm 8\%$ of the stated value; $\pm 7\%$ of the stated value; $\pm 6\%$ of the stated value; $\pm 5\%$ of the stated value; $\pm 4\%$ of the stated value; $\pm 3\%$ of the stated value; $\pm 2\%$ of the stated value; or $\pm 1\%$ of the stated value.

[0151] Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

[0152] The terms “a,” “an,” “the” and similar referents used in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. Recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., “such as”) provided herein is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the invention.

[0153] Groupings of alternative elements or embodiments of the invention disclosed herein are not to be construed as limitations. Each group member may be referred to and claimed individually or in any combination with other members of the group or other elements found herein. It is anticipated that one or more members of a group may be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the

specification is deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

[0154] Certain embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Of course, variations on these described embodiments will become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventor expects skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

[0155] Furthermore, numerous references have been made to patents, printed publications, journal articles and other written text throughout this specification (referenced materials herein). Each of the referenced materials are individually incorporated herein by reference in their entirety for their referenced teaching.

[0156] In closing, it is to be understood that the embodiments of the invention disclosed herein are illustrative of the principles of the present invention. Other modifications that may be employed are within the scope of the invention. Thus, by way of example, but not of limitation, alternative configurations of the present invention may be utilized in accordance with the teachings herein. Accordingly, the present invention is not limited to that precisely as shown and described.

[0157] The particulars shown herein are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of various embodiments of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for the fundamental understanding of the invention, the description taken with the drawings and/or examples making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

[0158] Definitions and explanations used in the present disclosure are meant and intended to be controlling in any future construction unless clearly and unambiguously modified in the following examples or when application of the meaning renders any construction meaningless or essentially meaningless. In cases where the construction of the term would render it meaningless or essentially meaningless, the definition should be taken from Webster's Dictionary, 3rd Edition or

a dictionary known to those of ordinary skill in the art, such as the Oxford Dictionary of Biochemistry and Molecular Biology (Ed. Anthony Smith, Oxford University Press, Oxford, 2004).

CLAIMS

What is claimed is:

1. A method for expressing a heterologous gene within deep cerebellar nucleus (DCN) cells in the cerebellum and layer 5 glutamatergic extraencephalic (L5 ET) neurons in the neocortex *in vivo* or *in vitro*, the method comprising providing an administrable composition in a sufficient dosage and for a sufficient time to a sample or subject comprising the DCN cells in the cerebellum and L5 ET neurons in the neocortex thereby expressing the gene within the DCN cells in the cerebellum and L5 ET neurons in the neocortex wherein the composition comprises an artificial expression construct comprising (i) an eHGT_453m enhancer; (ii) a promoter; and (iii) a heterologous coding sequence.
2. The method of claim 1, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.
3. The method of claim 3, wherein the capsid comprises PHP.eB, AAV-BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.
4. A concatenated core of an eHGT_453m, eHGT_023h v3, eHGT_387m, and/or MGT_E116 enhancer.
5. The concatenated core of claim 4, wherein the concatenated core comprises the sequence as set forth in SEQ ID NO: 103, SEQ ID NO: 25, SEQ ID NO: 105, and/or SEQ ID NO: 121 or a sequence having at least 90% sequence identity to the sequence as set forth in SEQ ID NO: 103, SEQ ID NO: 25, SEQ ID NO: 105, and/or SEQ ID NO: 121.
6. The concatenated core of claim 4, wherein the concatenated core comprises 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core of eHGT_453m, eHGT_023h v3, eHGT_387m, and/or MGT_E116.
7. The concatenated core of claim 4, comprising 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 103, SEQ ID NO: 25, SEQ ID NO: 105, and/or SEQ ID NO: 121 or a sequence having at least 90% sequence identity to the sequence as set forth in SEQ ID NO: 103, SEQ ID NO: 25, SEQ ID NO: 105, and/or SEQ ID NO: 121.
8. The concatenated core of claim 4, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 103.
9. The concatenated core of claim 4, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 25.
10. The concatenated core of claim 4, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 105.
11. The concatenated core of claim 4, having 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 121.

12. The concatenated core of claim 8, having 3 copies of the sequence as set forth in SEQ ID NO: 103.
13. The concatenated core of claim 9, having 3 copies of the sequence as set forth in SEQ ID NO: 25.
14. The concatenated core of claim 10, having 3 copies of the sequence as set forth in SEQ ID NO: 105.
15. The concatenated core of claim 11, having 3 copies of the sequence as set forth in SEQ ID NO: 121.
16. The concatenated core of claim 12, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 104.
17. The concatenated core of claim 13, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 26.
18. The concatenated core of claim 14, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 106.
19. The concatenated core of claim 15, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 122.
20. An artificial expression construct comprising (i) an enhancer selected from 3xcore3_eHGT_453m, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_181h, eHGT_023h v2, eHGT_356h, eHGT_467m, eHGT_483m, eHGT_796h, eHGT_588m, eHGT_007m, eHGT_589m, eHGT_963m, 3xcore2_eHGT_387m, eHGT_882m, MGT_E122, MGT_E146, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m (ii) a promoter; and (iii) a heterologous coding sequence.
21. The artificial expression construct of claim 20, wherein the heterologous coding sequence encodes an effector element or an expressible element.
22. The artificial expression construct of claim 21, wherein the effector element comprises a reporter protein or a functional molecule.
23. The artificial expression construct of claim 22, wherein the reporter protein comprises a fluorescent protein.
24. The artificial expression construct of claim 22, wherein the functional molecule comprises a functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or a designer receptor exclusively activated by designer drug (DREADD).
25. The artificial expression construct of claim 19, wherein the expressible element comprises a

non-functional molecule.

26. The artificial expression construct of claim 25, wherein the non-functional molecule comprises a non-functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or DREADD.

27. The artificial expression construct of claim 20, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.

28. The artificial expression construct of claim 27, wherein the capsid comprises PHP.eB, AAV-BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.

29. The artificial expression construct of claim 20, wherein the artificial expression construct comprises or encodes a skipping element.

30. The artificial expression construct of claim 29, wherein the skipping element comprises a 2A peptide or an internal ribosome entry site (IRES).

31. The artificial expression construct of claim 30, wherein the 2A peptide comprises T2A, P2A, E2A, or F2A.

32. The artificial expression construct of claim 20, wherein the artificial expression construct comprises or encodes a set of features selected from: 3xcore3_eHGT_453m, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_181h, eHGT_023h v2, eHGT_356h, eHGT_467m, eHGT_483m, eHGT_796h, eHGT_588m, eHGT_007m, eHGT_589m, eHGT_963m, 3xcore2_eHGT_387m, eHGT_882m, MGT_E122, MGT_E146, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, eHGT_1027m, AAV, scAAV, rAAV, pAAV, minBglobin, CMV, minCMV, minCMV*, minRho, minRho*, fluorescent protein, 10aa, H2B, H2Bmod, H2B*, hsA2, Cre, iCre, dgCre, FlpO, tTA2, SP10, tag cassette nuclear localization protein, self-cleaving peptides, WPRE, WPRE3, hGHpA, and/or BGHpA.

33. The artificial expression construct of claim 20, wherein the artificial expression construct comprises or encodes a set of features selected from:

eHGT_023m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;

hsA2-3xCore-eHGT_023h v3-minRho-[heterologous coding sequence]-WPRE3-BGHpA;

hsA2-eHGT_181h-minRho-[heterologous coding sequence]-WPRE3-BGHpA;

hsA2-eHGT_023h v2-minRho-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_356h-minRho*-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_467m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_483m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_796h-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_588m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_007m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_589m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_963m-minBG-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA;
 eHGT_963m-minBG-[heterologous coding sequence]-WPRE3-BGHpA;
 3xcore3_eHGT_453m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 3xcore2_eHGT_387m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_882m-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA;
 eHGT_882m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 MGT_E122-minBglobin-[heterologous coding sequence]-WPRE3-bGHpA;
 MGT_E146-minBglobin-[heterologous coding sequence]-WPRE3-bGHpA;
 3xCore_MGT_E116-minBglobin-[heterologous coding sequence]-WPRE3-bGHpA;
 MGT_E150-minBglobin-[heterologous coding sequence]-WPRE3-bGHpA;
 eHGT_1032h-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA;
 eHGT_1032h-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_1027h-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B*-WPRE3-BGHpA;
 eHGT_1027h-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_1027m-minBglobin-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B*-WPRE3-BGHpA;
 eHGT_1027m-minBglobin-[heterologous coding sequence]-WPRE3-BGHpA;
 eHGT_023m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-3xCore-eHGT_023h v3-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-eHGT_181h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 hsA2-eHGT_023h v2-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];

eHGT_356h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_467m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_483m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_796h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_588m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_007m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_589m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_963m-minBG-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-[post regulatory elements];
 eHGT_963m-minBG-[heterologous coding sequence]-[post regulatory elements];
 3xcore3_eHGT_453m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 3xcore2_eHGT_387m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_882m-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-[post regulatory elements];
 eHGT_882m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 MGT_E122-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 MGT_E146-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 3xCore_MGT_E116-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 MGT_E150-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_1032h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-[post regulatory elements];
 eHGT_1032h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_1027h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B*-[post regulatory elements];
 eHGT_1027h-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements];
 eHGT_1027m-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B*-[post regulatory elements]; or
 eHGT_1027m-[minimal promoter]-[heterologous coding sequence]-[post regulatory elements].
 34. A vector comprising an artificial expression construct of claim 20.
 35. The vector of claim 34, wherein the vector comprises a viral vector.

36. The vector of claim 35, wherein the viral vector comprises a recombinant adeno-associated viral (AAV) vector.
37. A transgenic cell comprising an artificial expression construct of claim 20.
38. The transgenic cell of claim 37, wherein the transgenic cell comprises a cerebellar cell, GABAergic neuron, glutamatergic neuron, non-neuronal cell, cholinergic interneuron, medium spin neuron (MSN), and/or spinal motor neuron.
39. The transgenic cell of claim 38, wherein the cerebellar cell comprises a Purkinje cell, molecular layer interneuron (MLI) cell, deep cerebellar nucleus (DCN) cell, Bergman glia, granule cell, mossy fiber (MF) cell, foliar white matter (WM), and/or oligodendrocyte.
40. The transgenic cell of claim 38, wherein the GABAergic neuron comprises a Pvalb neuron, Sst neuron, Lamp5 neuron, Vip neuron, medial ganglionic eminence (MGE) (Sst/Pvalb) cell, and/or pan-GABAergic neuron.
41. The transgenic cell of claim 40, wherein the Pvalb neuron comprises a chandelier cell.
42. The transgenic cell of claim 40, wherein the Sst neuron comprises an Sst Chodl cell.
43. The transgenic cell of claim 38, wherein the glutamatergic neuron comprises a pan-glutamatergic neuron, layer 5 glutamatergic extraencephalic (L5 ET) neuron, L5 near-projecting (NP), L5 intratelencephalic (IT), and/or glutamatergic neuron in the thalamus
44. The transgenic cell of claim 38, wherein the non-neuronal cell comprises an astrocyte, oligodendrocyte, pericyte, or specialized smooth-muscle cells (SMCs).
45. The transgenic cell of claim 37, wherein the transgenic cell is murine, human, or non-human primate.
46. A non-human transgenic animal comprising an artificial expression construct of claim 20.
47. The non-human transgenic animal of claim 46, wherein the non-human transgenic animal is a mouse or a non-human primate.
48. An administrable composition comprising an artificial expression construct of claim 20.
49. A kit comprising an artificial expression construct of claim 18.
50. A method for expressing a gene within a population of cerebellar cell types and a population of secondary cell types *in vivo* or *in vitro*, the method comprising providing the administrable composition of claim 48 in a sufficient dosage and for a sufficient time to a sample or subject comprising the population of cells thereby expressing the gene within the population of cells.
51. A method for expressing a heterologous gene within a population of cerebellar cell types and a population of secondary cell types *in vivo* or *in vitro*, the method comprising providing an administrable composition in a sufficient dosage and for a sufficient time to a sample or subject comprising the population of cells thereby expressing the gene within the population of cells

wherein the composition comprises an artificial expression construct comprising (i) an enhancer selected from 3xcore3_eHGT_453m, eHGT_023m, 3xCore-eHGT_023h v3, eHGT_181h, eHGT_023h v2, eHGT_356h, eHGT_467m, eHGT_483m, eHGT_796h, eHGT_588m, eHGT_007m, eHGT_589m, eHGT_963m, 3xcore2_eHGT_387m, eHGT_882m, MGT_E122, MGT_E146, 3xcore_MGT_E116, MGT_E150, eHGT_1032h, eHGT_1027h, and/or eHGT_1027m; (ii) a promoter; and (iii) a heterologous coding sequence.

52. The method of claim 51, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.

53. The method of claim 52, wherein the capsid comprises PHP.eB, AAV-BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.

54. The method of claim 51, wherein the providing comprises pipetting.

55. The method of claim 54, wherein the pipetting is to a brain slice.

56. The method of claim 55, wherein the brain slice comprises a cerebellar cell and/or a secondary cell type.

57. The method of claim 56, wherein the cerebellar cell comprises a Purkinje cell, MLI cell, DCN cell, Bergman glia, granule cell, MF cell, foliar white matter (WM), and/or oligodendrocyte.

58. The method of claim 56, wherein the secondary cell type comprises a GABAergic neuron, glutamatergic neuron, non-neuronal cell, cholinergic interneurons, MSN, or spinal motor neurons.

59. The method of claim 58, wherein the GABAergic neuron comprises a Pvalb neuron, Sst neuron, Lamp5 neuron, Vip neuron, MGE (Sst/Pvalb) cell, and/or pan-GABAergic neuron.

60. The method of claim 59, wherein the Pvalb neuron comprises a chandelier cell.

61. The method of claim 59, wherein the Sst neuron comprises an Sst Chodl cell.

62. The method of claim 58, wherein the glutamatergic neuron comprises a pan-glutamatergic neuron, L5 ET neuron, L5 NP, L5 IT, and/or glutamatergic neuron in the thalamus

63. The method of claim 58, wherein the non-neuronal cell comprises an astrocyte, oligodendrocyte, pericyte, or specialized SMCs.

64. The method of claim 56, wherein the cerebellar cell comprises Purkinje cells and the enhancer is eHGT_023m, 3xCore-eHGT_023h v3, eHGT_023h v2, eHGT_467m, eHGT_796h, eHGT_588m, or MGT_E150.

65. The method of claim 56, wherein the cerebellar cell comprises MLI cells and the enhancer is 3xCore-eHGT_023h v3, eHGT_181h, eHGT_023h v2, eHGT_588m, eHGT_007m, eHGT_963m, MGT_E146, eHGT_1032h, eHGT_1027h, or eHGT_1027m.

66. The method claim 56, wherein the MLI cells are basket cells and/or stellate cells.

67. The method of claim 56, wherein the cerebellar cell comprises DCN cells and the enhancer is

3xcore3_eHGT_453m, eHGT_356h, eHGT_589m, eHGT_963m, eHGT_882m, or 3xcore_MGT_E116.

68. The method of claim 56, wherein the cerebellar cell comprises Bergman glia and the enhancer is 3xcore2_eHGT_387m, or MGT_E122.

69. The method of claim 56, wherein the cerebellar cell comprises granule and MF cells and the enhancer is eHGT_483m.

70. The method of claim 56, wherein the secondary cell type is a Pvalb interneuron in the neocortex and the enhancer is eHGT_023m, 3xCore-eHGT_023h v3, eHGT_023h v2, or eHGT_589m.

71. The method of claim 56, wherein the secondary cell type is MGE cells in the neocortex and the enhancer is eHGT_588m.

72. The method of claim 56, wherein the secondary cell type is Sst/Chold neurons in the neocortex and the enhancer is eHGT_467m.

73. The method of claim 56, wherein the secondary cell type is Vip interneurons in the neocortex and the enhancer is eHGT_356h, eHGT_483m, or eHGT_007m.

74. The method of claim 56, wherein the secondary cell type is Lamp5 interneurons in the neocortex and the enhancer is eHGT_181h.

75. The method of claim 56, wherein the secondary cell type is pan-GABAergic neurons in the neocortex and the enhancer is eHGT_796h.

76. The method of claim 56, wherein the secondary cell type is L5 NP neurons in the neocortex and the enhancer is 3xcore_MGT_E116.

77. The method of claim 56, wherein the secondary cell type is astrocytes in the neocortex and the enhancer is 3xcore2_eHGT_387m, or MGT_E122.

78. The method of claim 56, wherein the secondary cell type is pericytes in the neocortex and the enhancer is MGT_E146.

79. The method of claim 56, wherein the secondary cell type is specialized SMCs in the neocortex and the enhancer is MGT_E150.

80. The method of claim 56, wherein the secondary cell type is MSN of the striatum and the enhancer is eHGT_882m.

81. The method of claim 56, wherein the secondary cell type is pan-GABAergic neurons brain wide and the enhancer is eHGT_1032h, eHGT_1027h, or eHGT_1027m.

82. The method of claim 56, wherein the secondary cell type is pan-neuronal cells brain wide.

83. The method of claim 56, wherein the secondary cell type is spinal motor neurons.

84. The method of claim 55, wherein the brain slice is murine, human, or non-human primate.

85. The method of claim 54, wherein the providing comprises administering to a living subject.
86. The method of claim 86, wherein the living subject is a human, non-human primate, or a mouse.
87. The method of claim 86, wherein the administering to a living subject is through injection.
88. The method of claim 88, wherein the injection comprises intravenous injection, intraparenchymal injection into brain tissue, intracerebroventricular (ICV) injection, intra-cisterna magna (ICM) injection, or intrathecal injection.
89. An artificial expression construct consisting of or consisting essentially of CN1280, CN1521, CN1674, CN1932, CN2043, CN2316, CN2431, CN2663, CN2838, CN1415, CN2839, CN3301, CN3019, CN3569, CN3453, AiP1351, AiP1375, AiP1530, AiP1379, CN3916, CN3869, or CN3870.

FIG. 1A

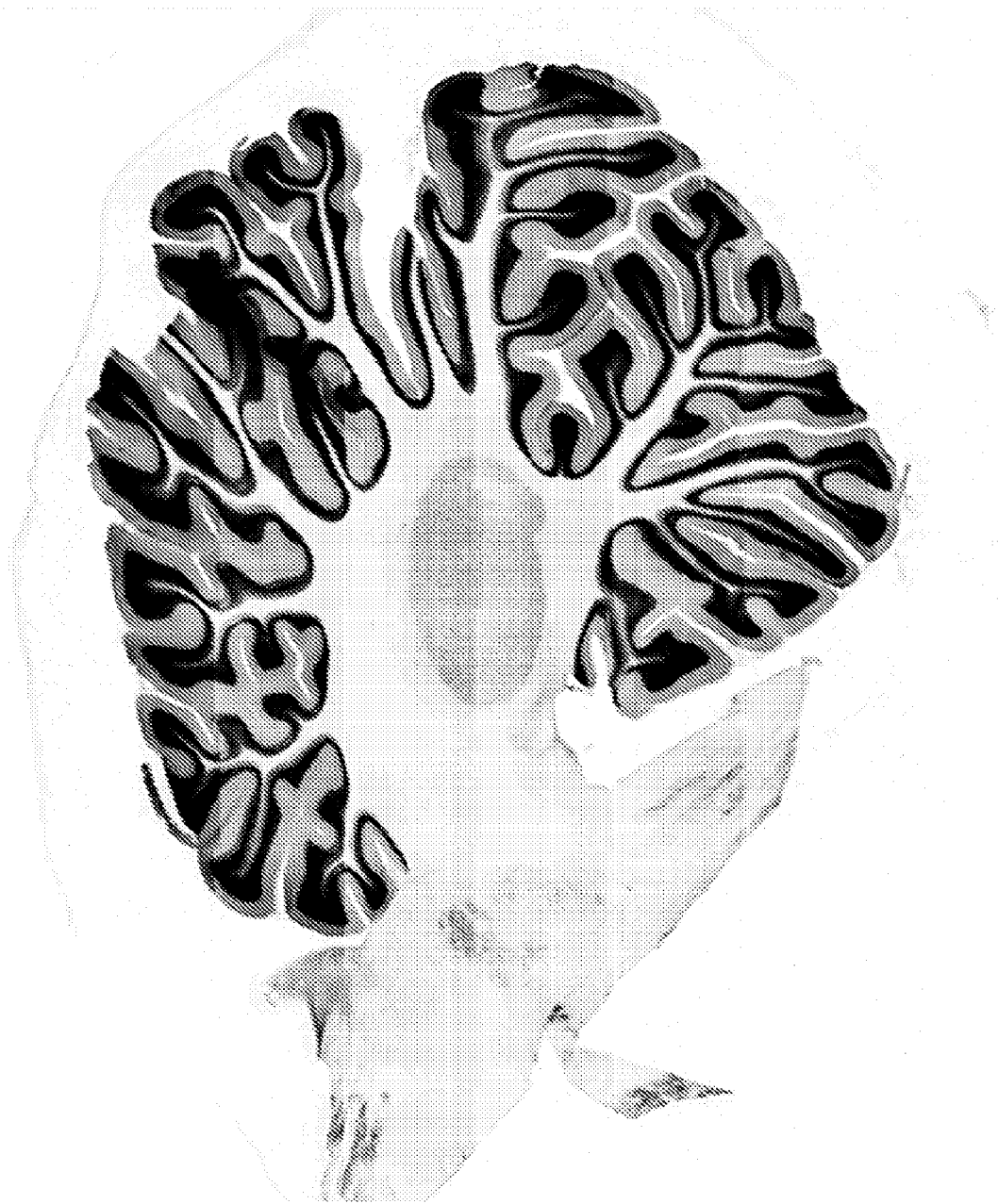


FIG. 1B

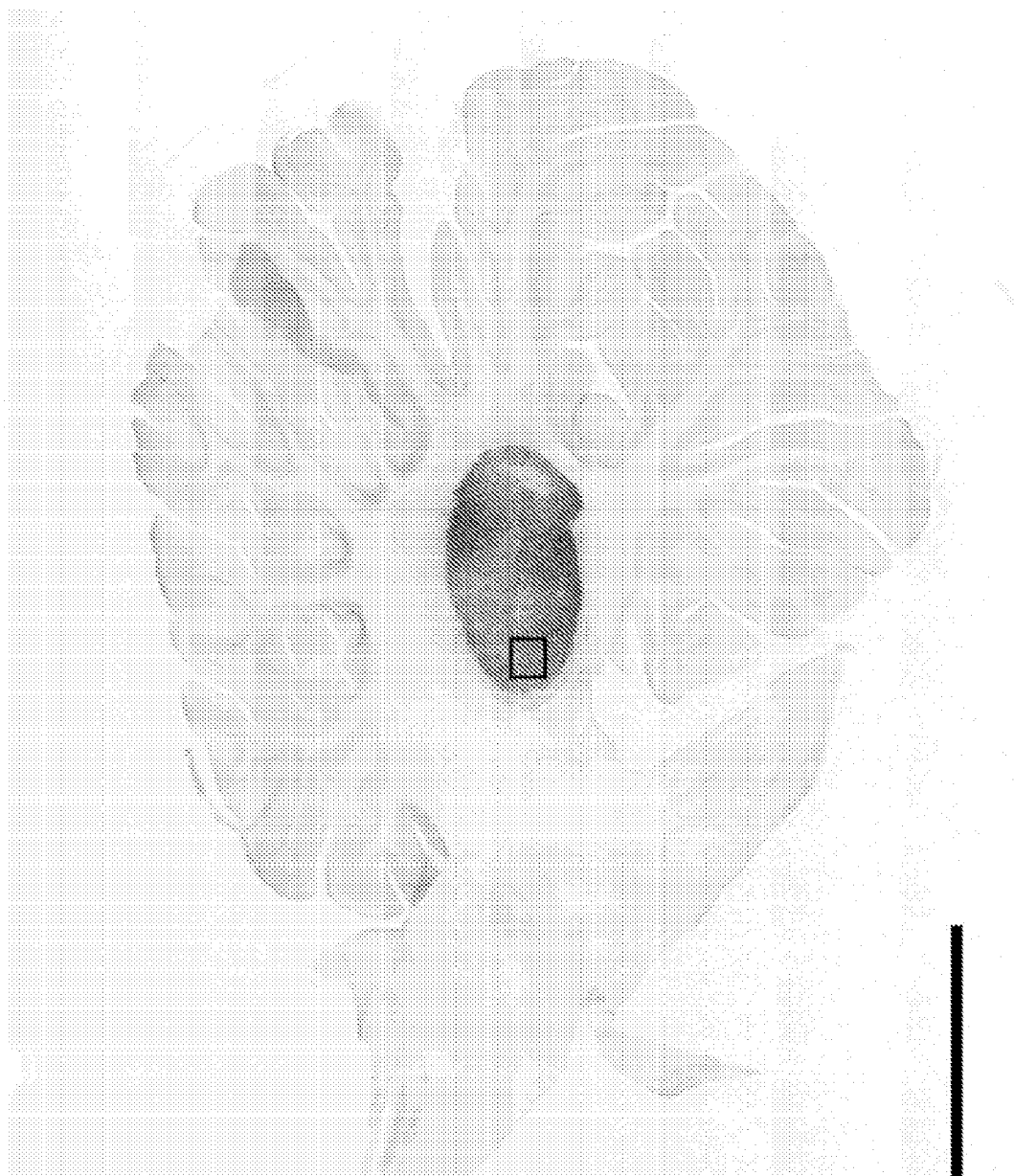


FIG. 1C

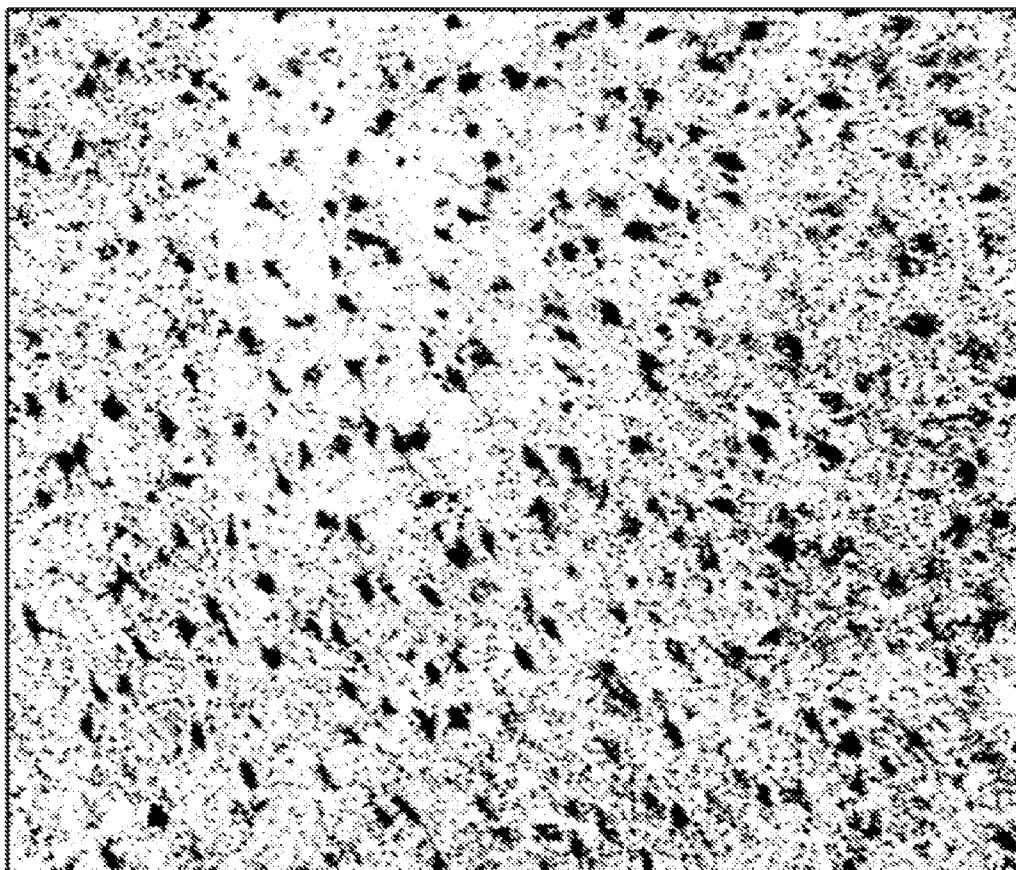
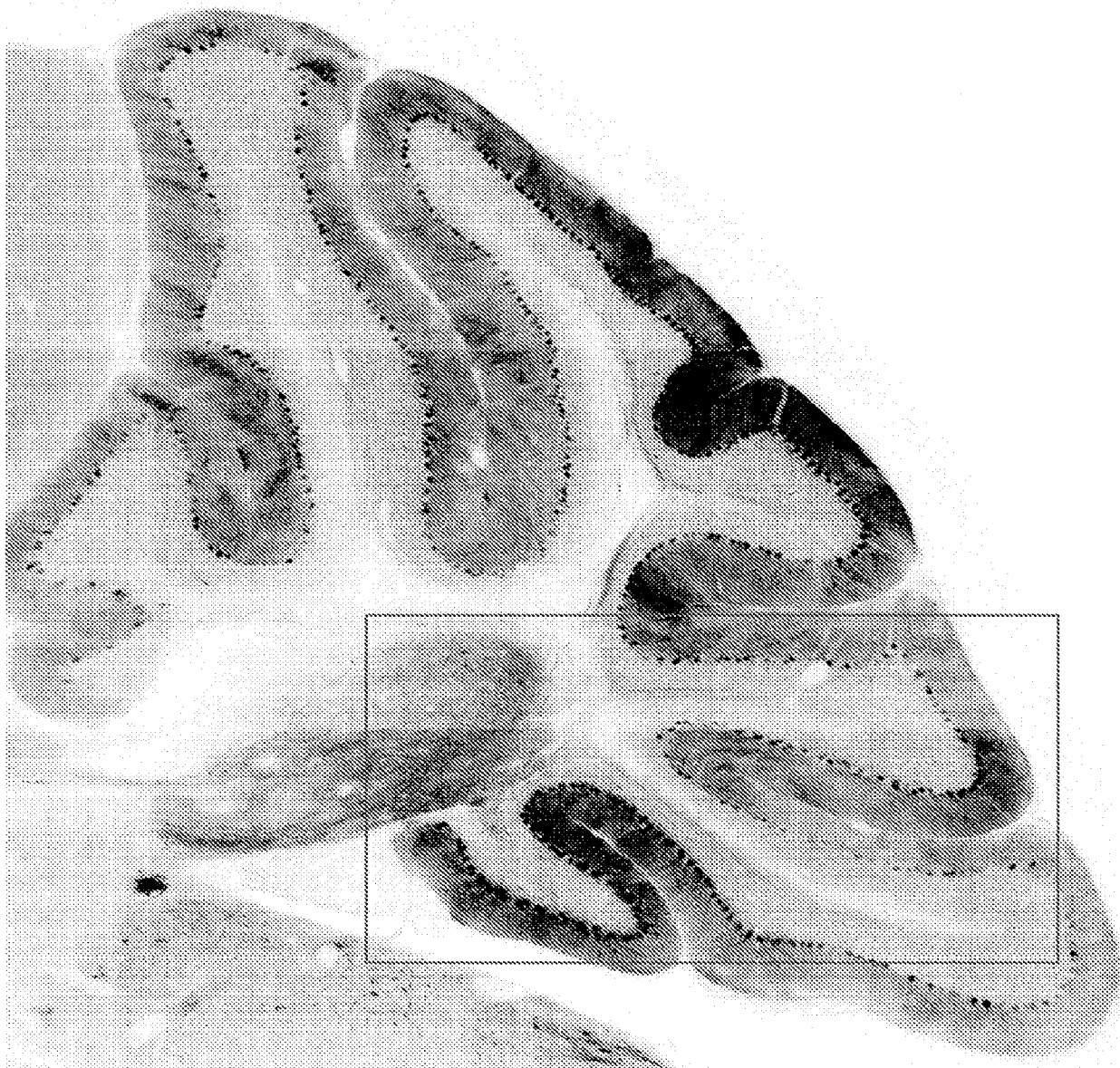


FIG. 2A



5/86

FIG. 2B



6/86

FIG. 2C



FIG. 2D

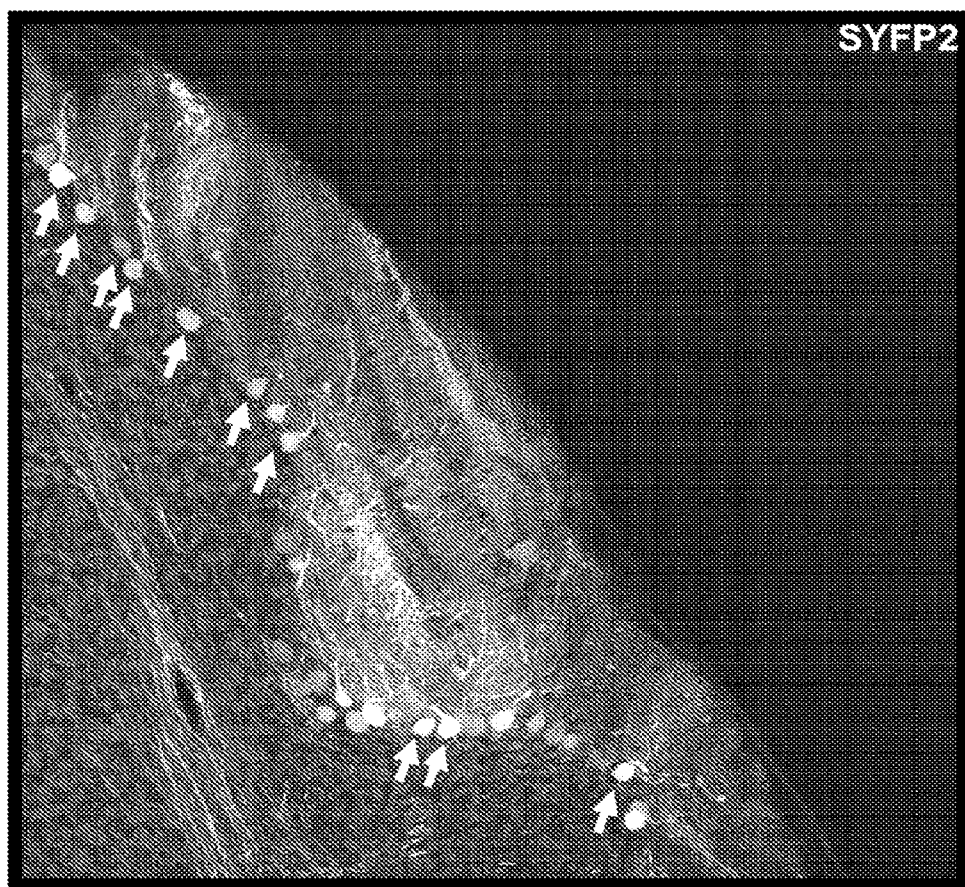


FIG. 2E

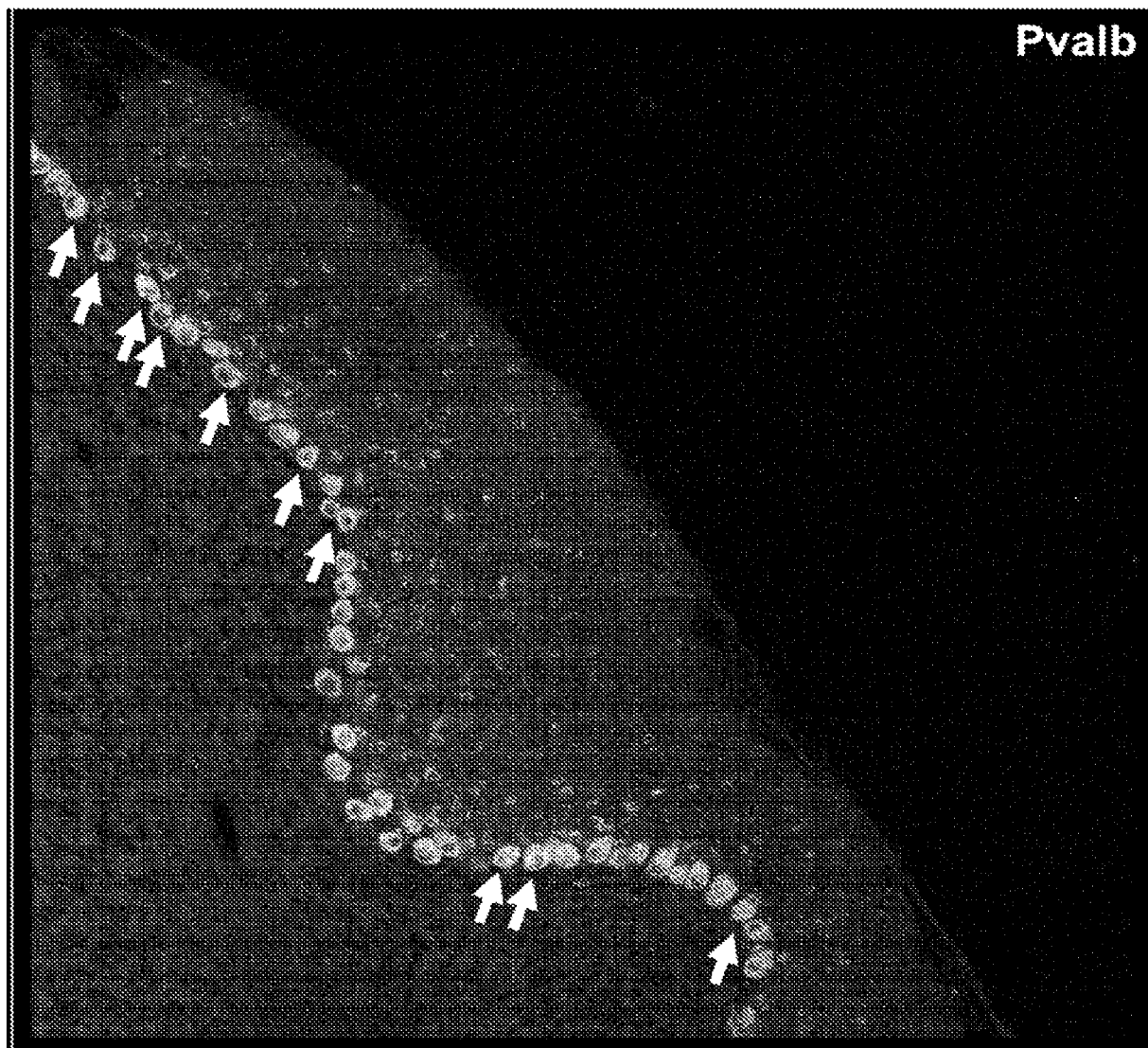
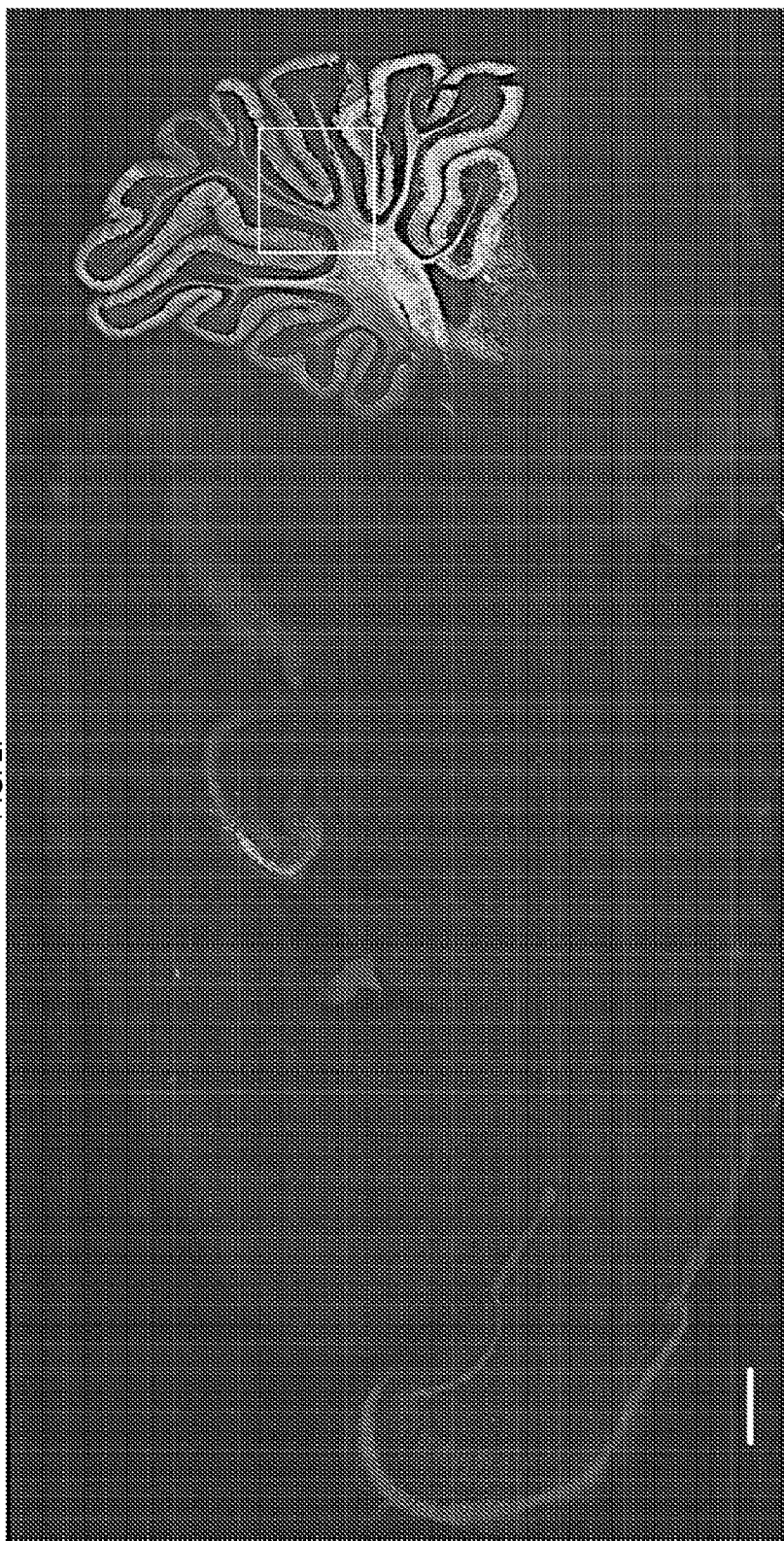


FIG. 2F



9/86

FIG. 2G

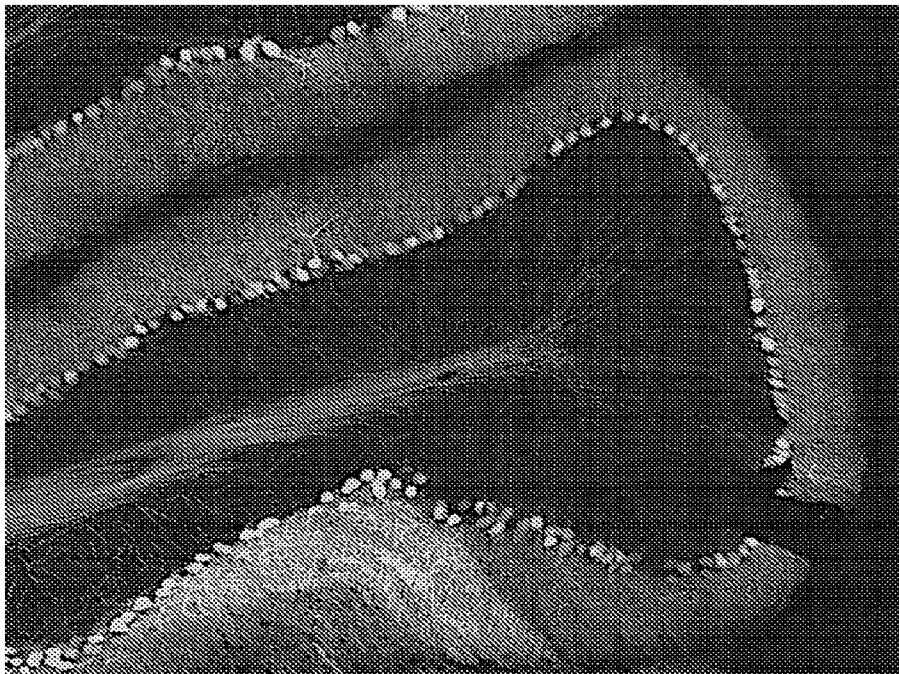


FIG. 2H

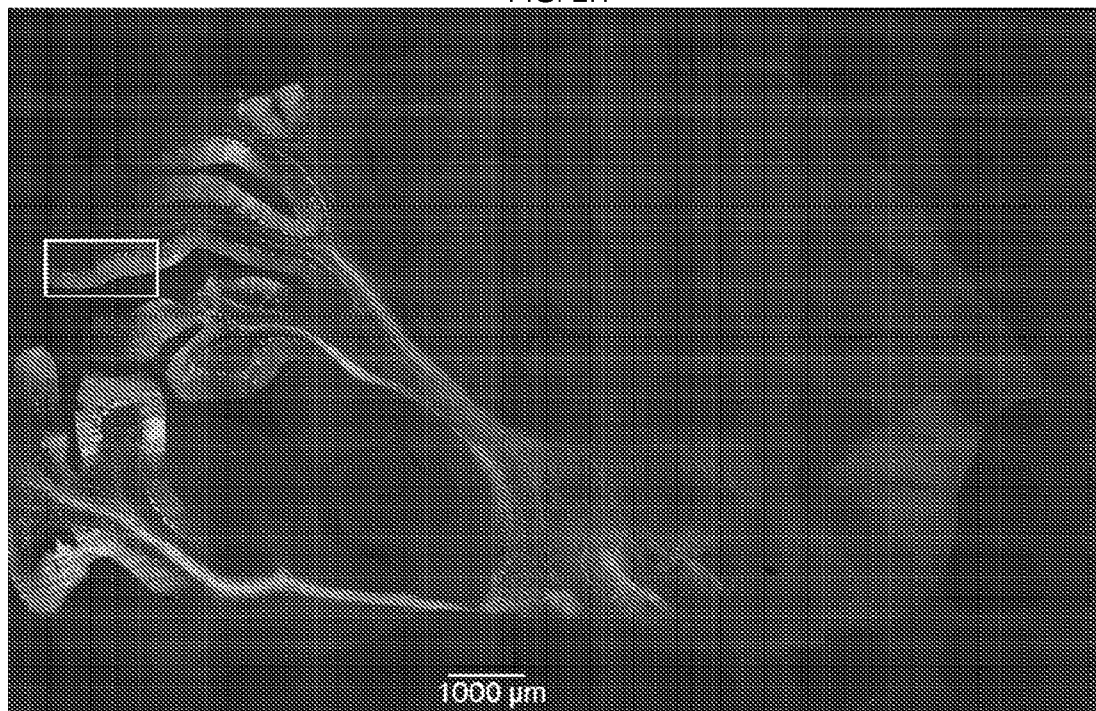


FIG. 2I

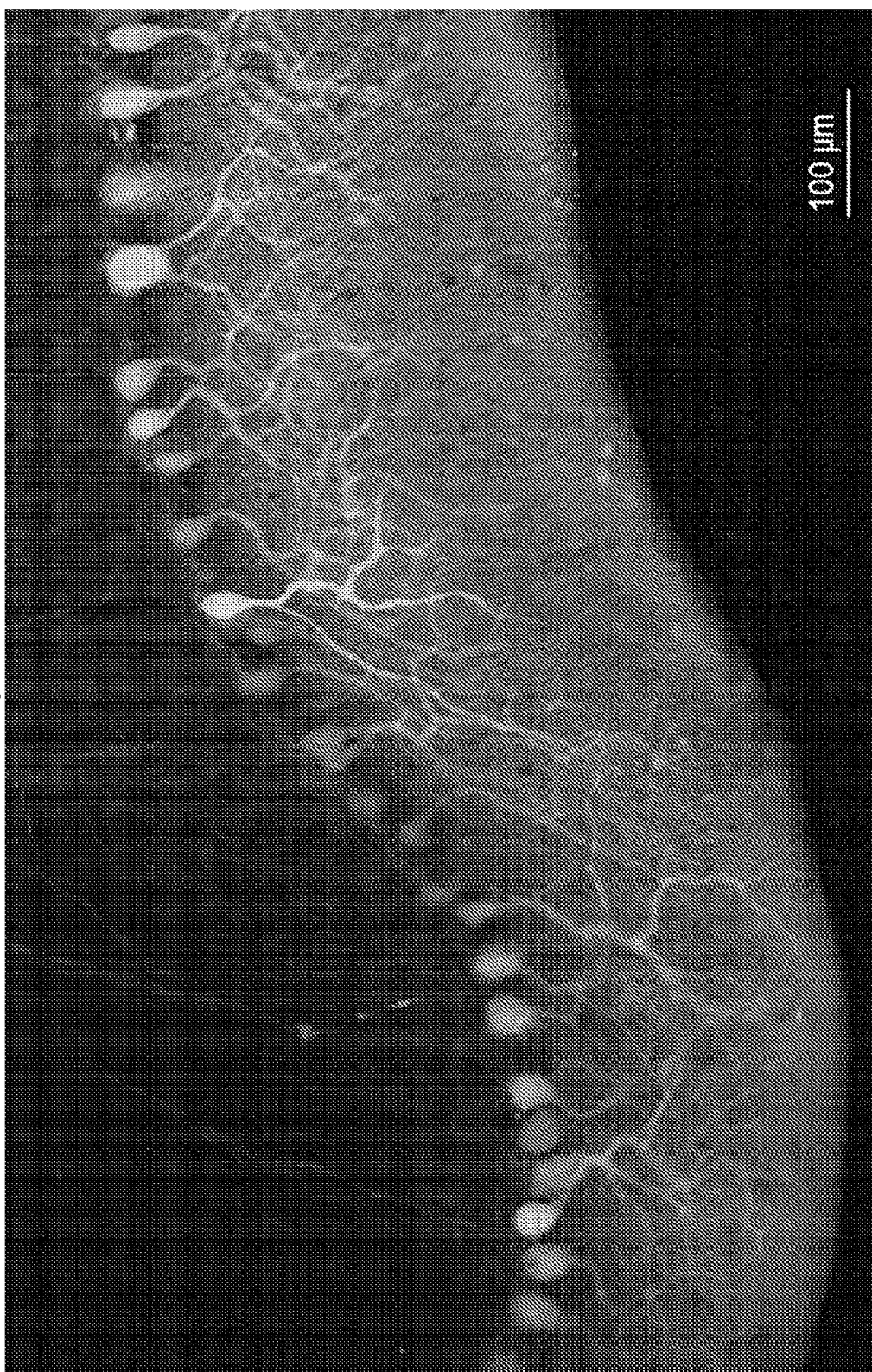
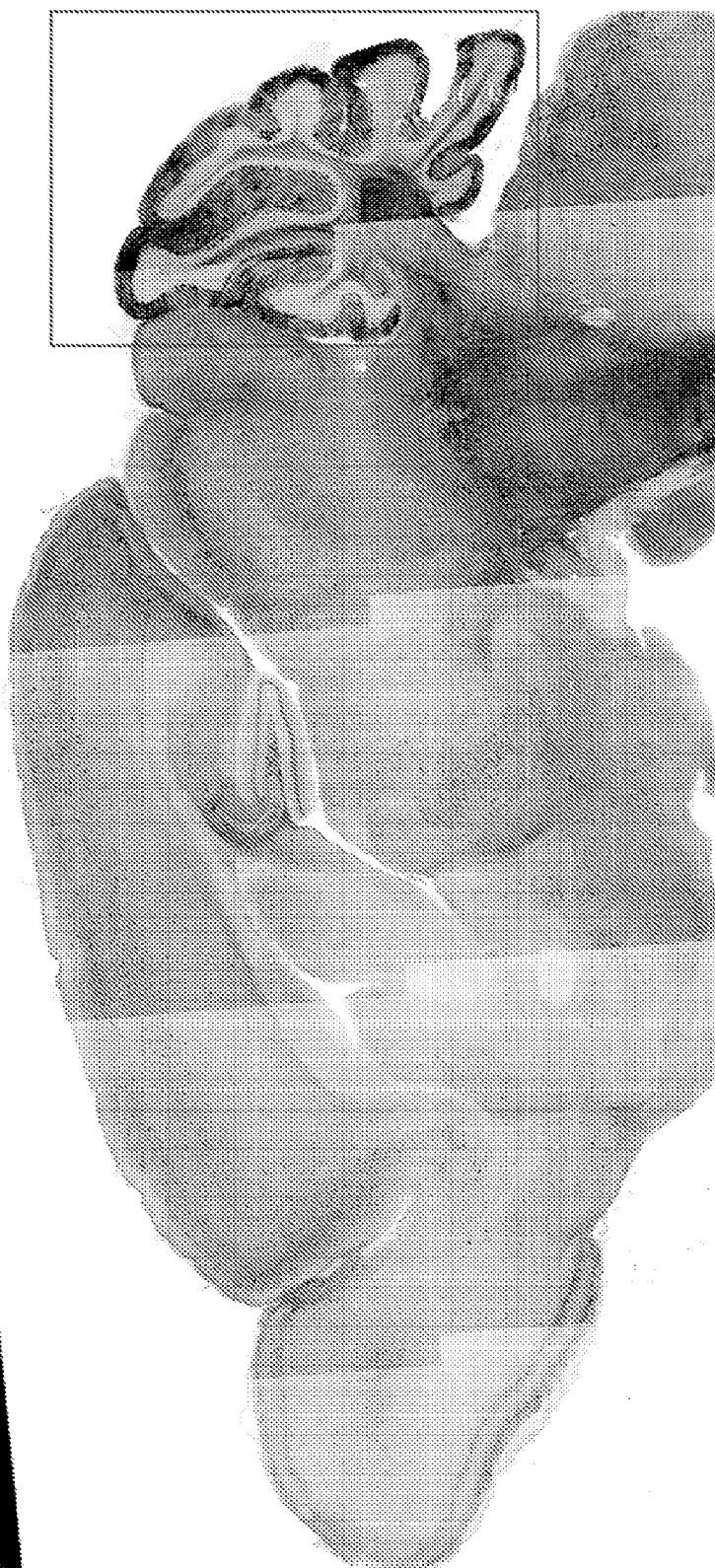


FIG. 3A



12/86

FIG. 3B

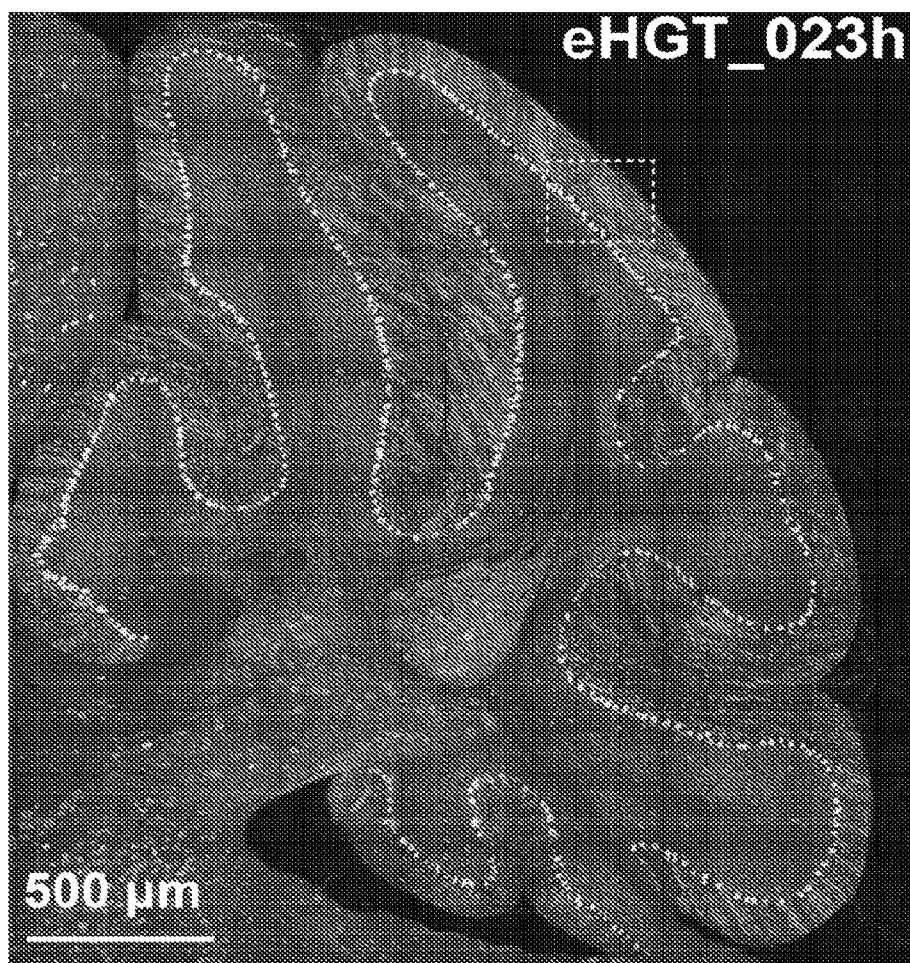


13/86

FIG. 3C



FIG. 3D



14/86

FIG. 3E

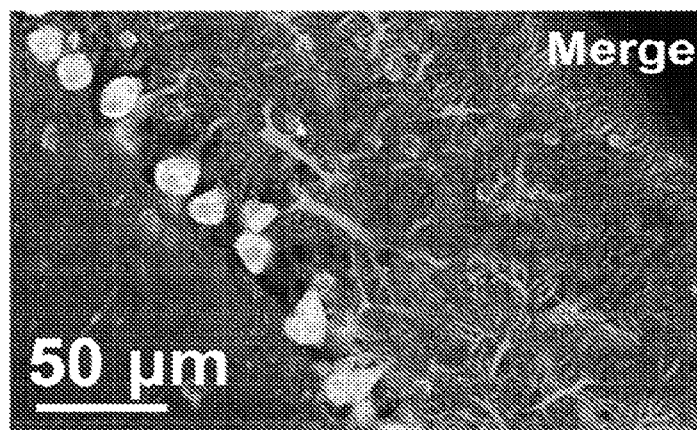
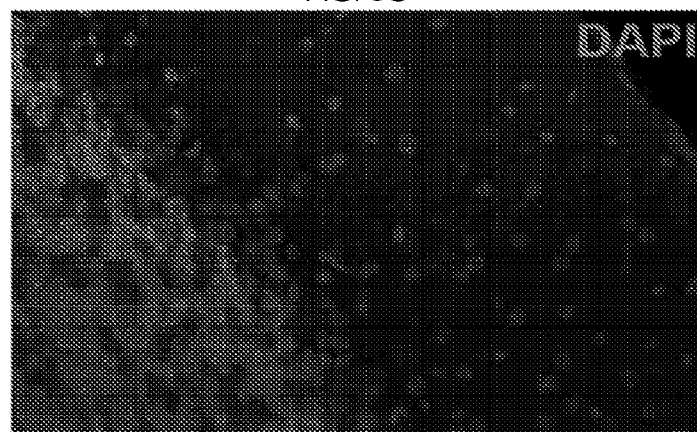


FIG. 3F



FIG. 3G



15/86

FIG. 3H

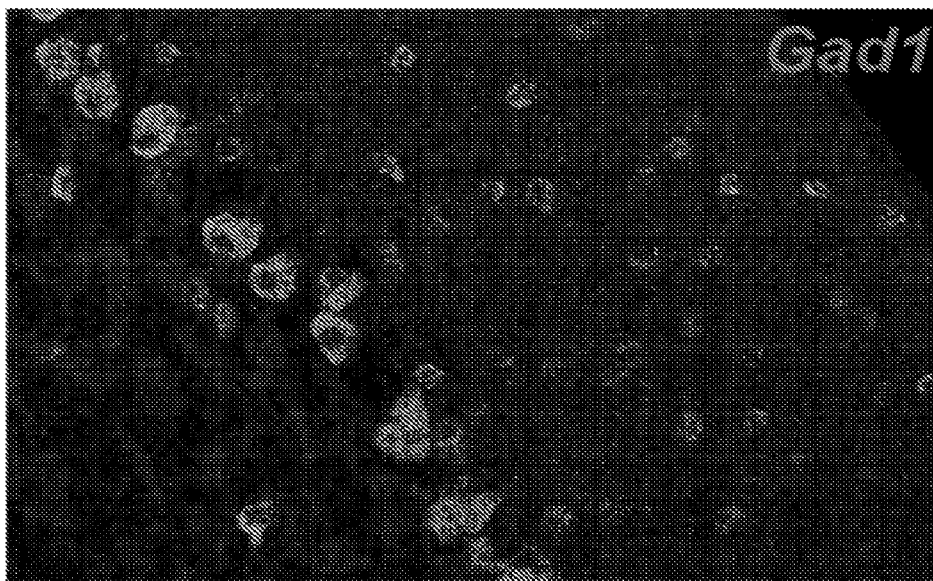


FIG. 3I

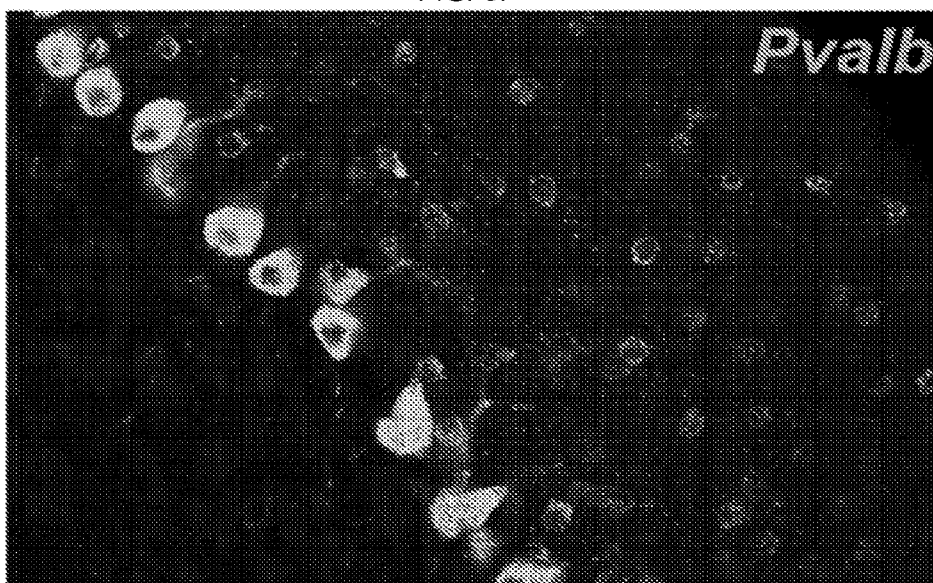


FIG. 4A



17/86

FIG. 4B

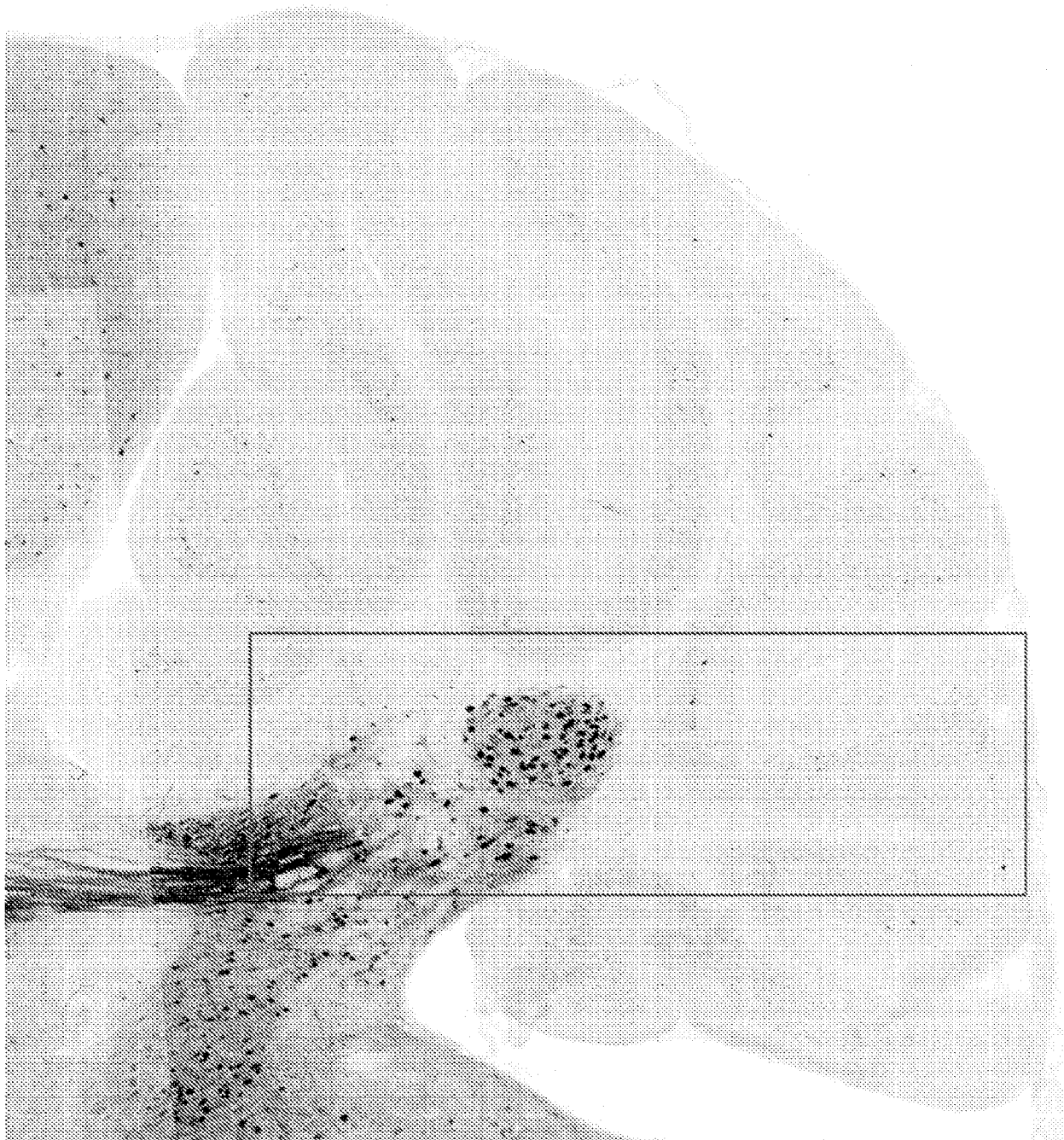
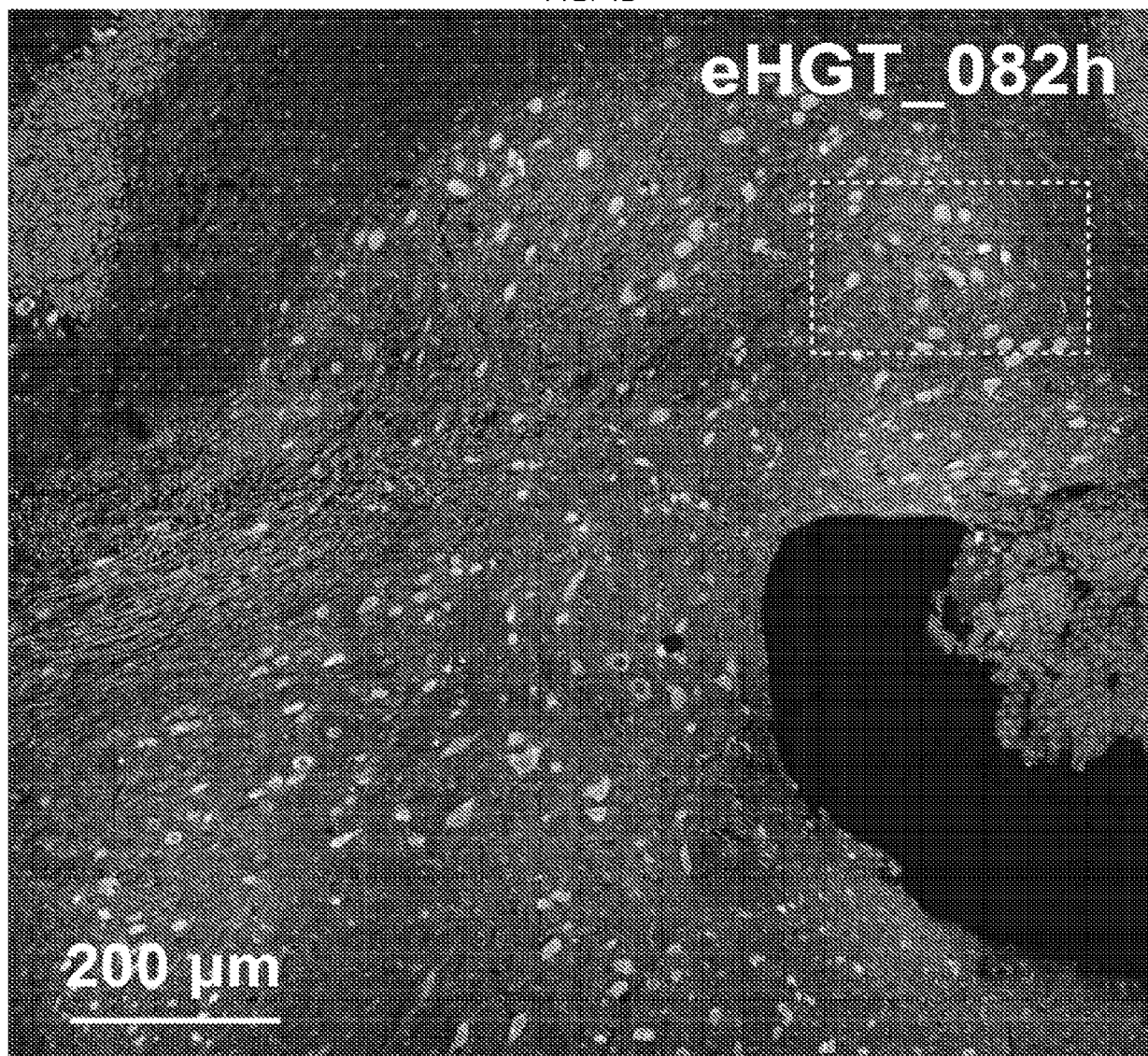


FIG. 4C



FIG. 4D



20/86

FIG. 4E

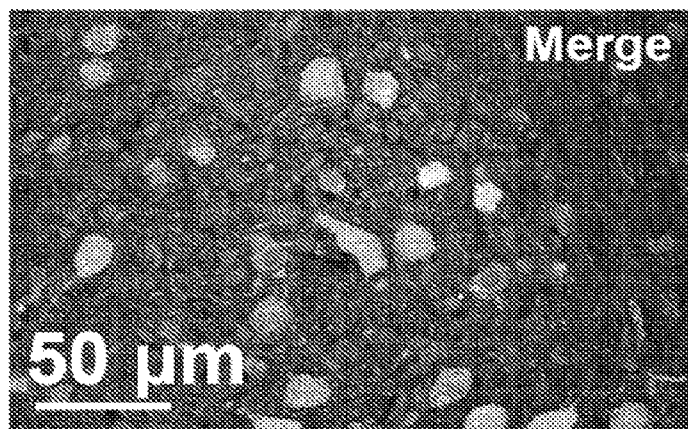


FIG. 4F

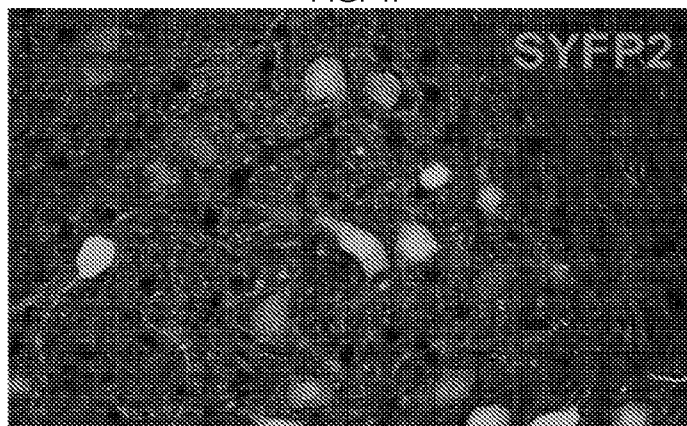
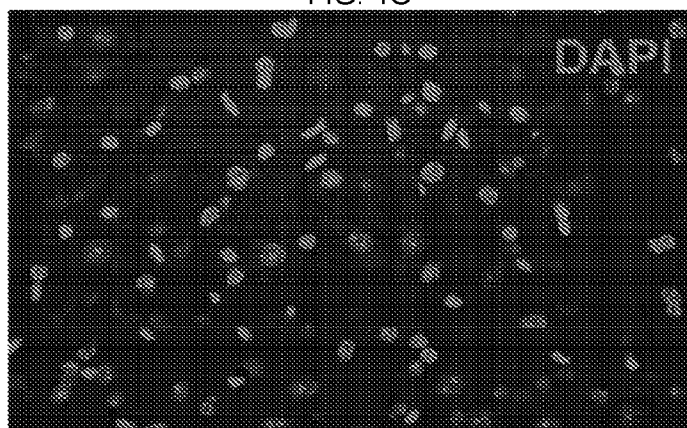


FIG. 4G



21/86

FIG. 4H

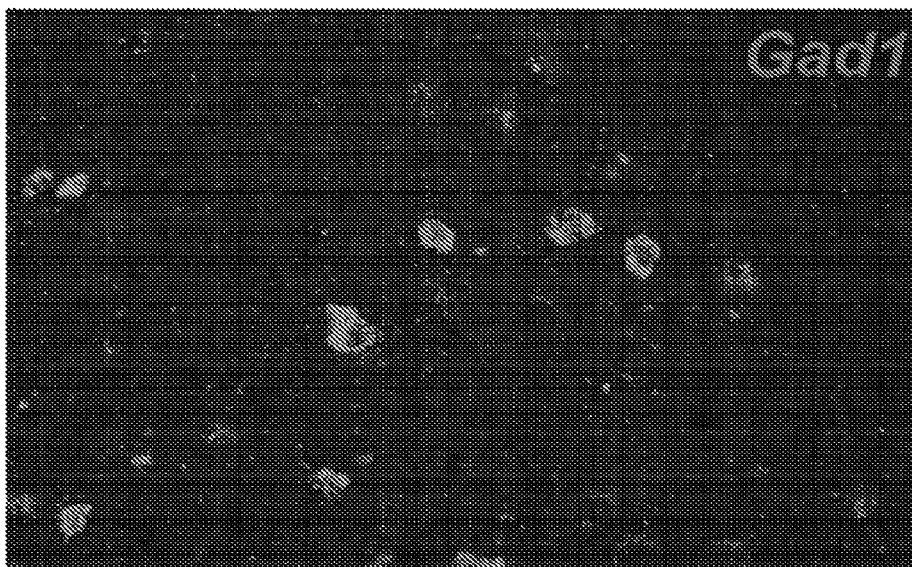


FIG. 4I

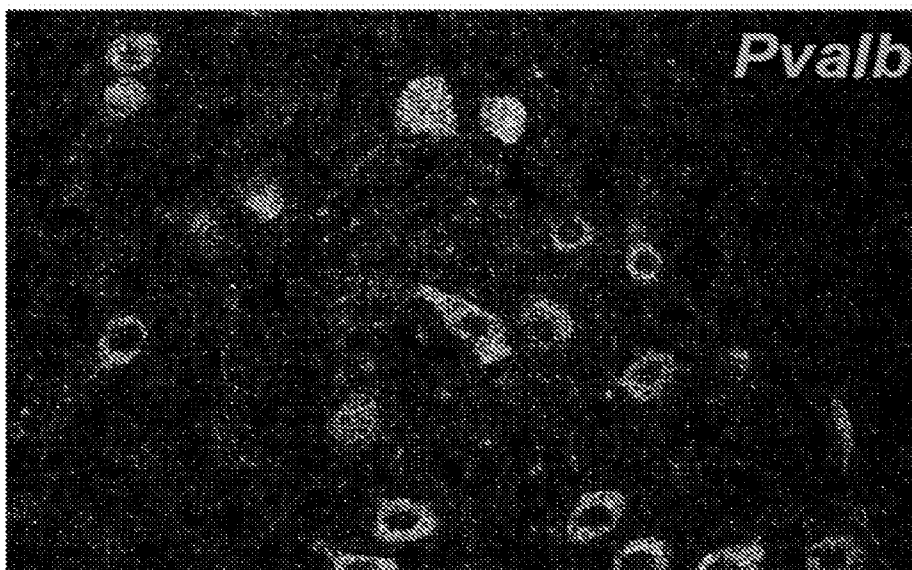


FIG. 5A

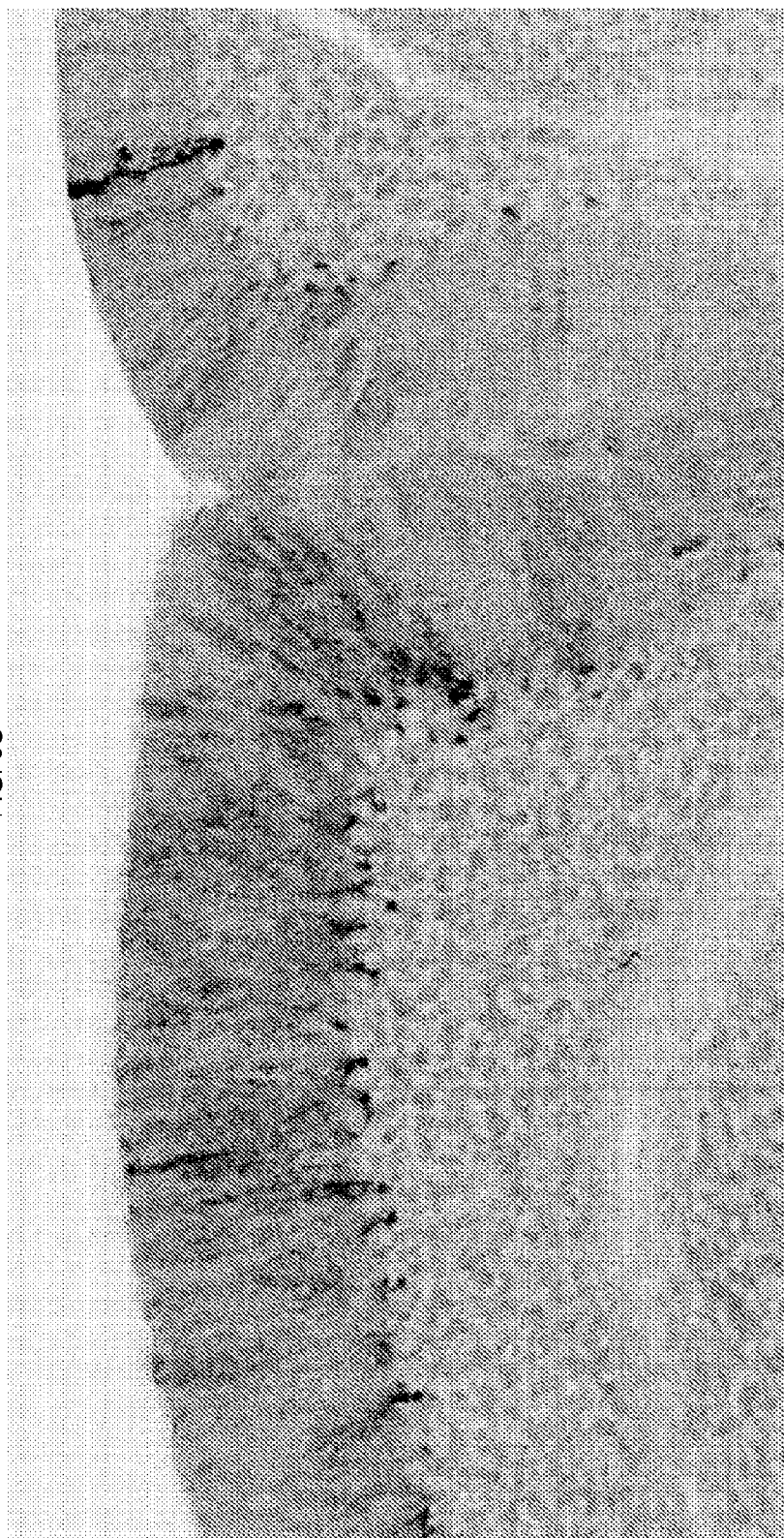


23/86

FIG. 5B



FIG. 5C



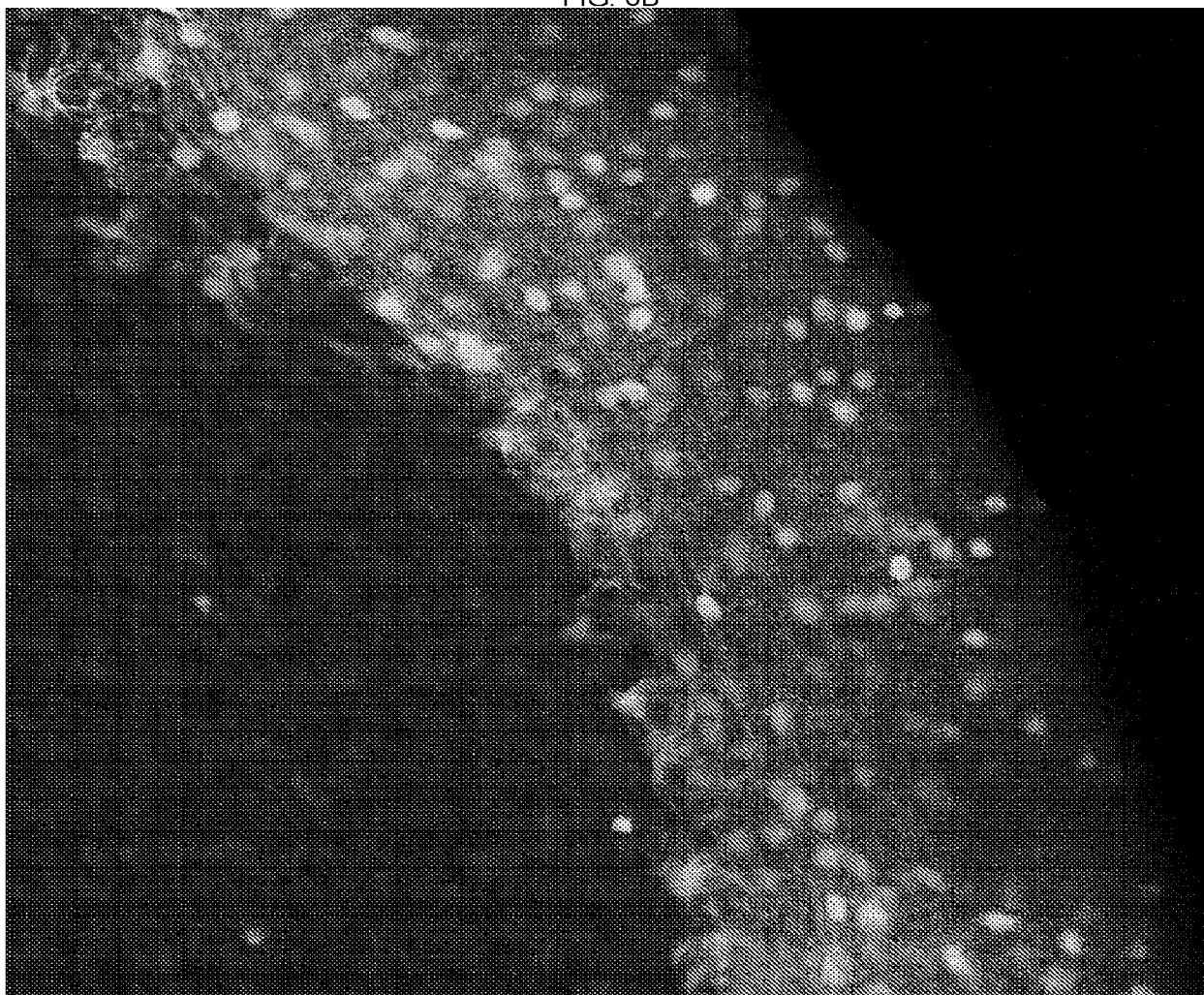
25/86

FIG. 6A



26/86

FIG. 6B



27/86

FIG. 7A



FIG. 7B

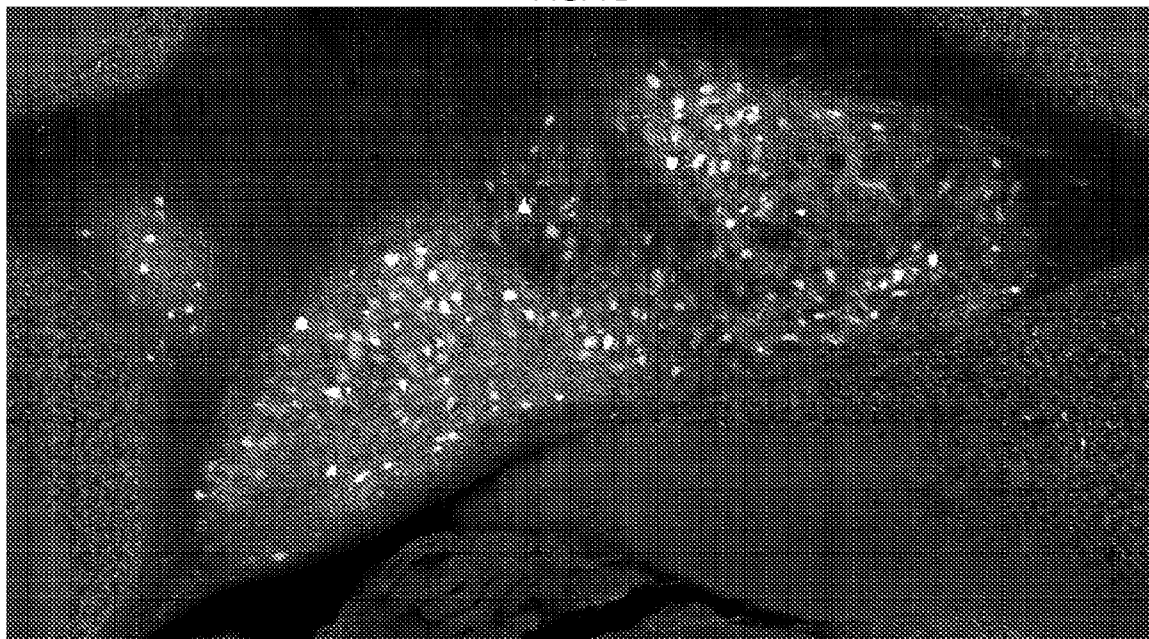


FIG. 8

eHGT_023h v1 (629 bp):

CCCTAAGCTGAGCTGAGACGTTTTAGTTATAAATGGCTTCTGTGCTTAGCAATCTGCTCTTT
TTATTCCCGTGTGGACTTTCCCTAGCTCTGGCCTTATGCTGCACTAGAAAAGATTTAGCAAG
GGGAGAGGAAGCAGCCTTCCCTACATAACTGGCCTCTTGTGAAAGGGAGCAGCTGCTTGG
TGGA AAAAGACATTCCCCTCCATGCATCCCTCTCCTTCTGCCTCTGGGGGTTGCAGCTTGA
GTCAGAACCAGGACCATTTAACTCCAACCTTTGAGGAAGAGACGCACCCTGGCCCCAGCC
ACGCCTGTTAGAATCTTCCCTAGCTGAGTGACACAGTGACACTCAGCCTCAGTTTCTCTGTA
AACAAAATGAAGATAAGAGAGCTGACGAGGATGAAATGGAATAACACACCCGTGCGGTGCC
TGGTACAGAGTGAGCCCCAGAACTGTTGACGCGGCCTCCTTGTGGCTGTCTGGCTTGACC
CGGAGTGACTCTGCCTCCCAGTTCTCCGGGATGGGAAGGTGATCCCTGTTTGAGATACCA
ATTTATAAGAAACCGAGCCCCGGGAGCTATTTAGAGGTGAGGTGATAAACCAGGAGGCCGG
CTCCTTCATCCCGGTCATCACGA (SEQ ID NO: 1)

eHGT_023m (603 bp):

GAAGCCGAGCTGAGCATTTGGAGTTTGAAATGTTTCCGTA CTTAGCAATCTGCTCCTCTATT
CCCGGGCGGACTTCCGCTAGCTCCGGCCTTATGCTGCACTAGATAAGATGGAGCAGGGA
GAGGACACGGCACTACTTATGTAACCGGCCTCTTGAAAAATGGAGCAGCGGTCAGGGCGG
AACAAAGACGTCCTCTCTACGCATCCCTCTCCTTTCCCTGCTAAGGCTGCAGCTGGAGTC
AGAGGCAGGGCTGTTCCAATCTGTCTTTGATCAGTAACGCAGCCAGCCTCCAGCCTCCGT
CAGCCTCCTCATGGCTGAGACCCGGCCTCAGTTTCCCCCACTTACATCCCGAGGATCAGA
GCCTGTGAGGATGAAATGGGATAAGGTAGCTGGAACCGTCTGGCAGAGAGCGAGTCCTCA
GGACTGTTGATGCCTGTGGCTGCCTGGCTTGACCCCAAGTGACCCCGCCTCCTCATCCTG
CAGCAGGAGAAAGTGATCCTGTTTGCGGCACCCATCTATAAGAAACTGAGCCCTGGAGCT
ATTTTGAGGCGAGGTGAGGTGATAAACCAGGAGCTGGGCCCTTGCTATGGTCATCAGAAC
G (SEQ ID NO: 2)

eHGT_359h (485 bp):

TCACAGCTCCTTGGTCTGAACTGGTACCCAGGATCAAGATGTCCCTCCCATGATTCTAATC
TCCAGCCCCCTTACACCATAGTCACTTCCAAACTCTGGATTTCTTAACATTGCAGTTTGCTT
CCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
CTATATCTCAGTGCAATTTCTACTTATCCACTCCCGACTATACCCTGCTTTCATCAGATTG
CAACTGACCTAATTTGTAAAGACTTCAGTGAAAAGAGCCTCATTAAGAGCCTTCATCTCTTT
TGCAGCCTCTCTTTCCCTCCCTCAAATGATCCAGGTCAGAAATGGCCCAATAAATAAATAAATA
GAAAATTACAAAGTCACTCAAGACTCTGCCAGGAGTAGAGATTTTAAGAATTAGTGACAGA
CATATACTGTCTAGGAGTAGGGTGGGAGTGGGTTGCATGGCTTAGGGACATAGA (SEQ ID
NO: 3)

eHGT_023h v2 (633 bp):

GCGTCCCTAAGCTGAGCTGAGACGTTTTAGTTATAAATGGCTTCTGTGCTTAGCAATCTGC
TCTTTTTATTCCCGTGTGGACTTTCCCTAGCTCTGGCCTTATGCTGCACTAGAAAAGATTTA
GCAAGGGGAGAGGAAGCAGCCTTCCCTACATAACTGGCCTCTTGTGAAAGGGAGCAGCTG
CTTGGTGGAAAAAGACATTCCCCTCCATGCATCCCTCTCCTTCTGCCTCTGGGGGTTGCAG
CTTGAGTCAGAACCAGGACCATTTAACTCCAACCTTTGAGGAAGAGACGCACCCTGGCCC
CAGCCACGCCTGTTAGAATCTTCCCTAGCTGAGTGACACAGTGACACTCAGCCTCAGTTTCT
CTGTAAACAAAATGAAGATAAGAGAGCCGACGAGGATGAAATGGAATAACACACCCGTGCG
GTGCCTGGTACAGAGTGAGCCCCAGAACTGTTGACGCGGCCTCCTTGTGGCTGTCTGGCT
TGACCCGGAGTGACTCTGCCTCCCAGTTCTCCGGGATGGGAAGGTGATCCCTGTTTGAGA
TACCAATTTATAAGAAACCGAGCCCCGGGAGCTATTTAGAGGTGAGGTGATAAACCAGGAG
GCCGGCTCCTTCATCCCGGTCATCACGA (SEQ ID NO: 4)

FIG. 8 cont'd

eHGT_082h (698 bp):

CCTTTGGCCGGCAGGAGCAGCCTGATTTTAGTGACTCGTTTCAGAAGGGAGGTGCAGCTGG
AAGGTTGTGTTTATTAAAGCCCAAGTGACTGGAGGTAAATGCAGGTGTGTTTCAGGAAATGT
AGCTGGTTACCTCTCAGTAGGCACAAGGCCAAAGTCAGCCGACTTGCAGATTGTAATGC
CGGTTCTCCACCCAGAGGTTATTTGAGTTCATCCAGCACTGCAGCTGAGCAGAATGCTGC
AGTCAGGGTGACACAATGCAAACTGGAGCCATGTCCTCAAGGGCATGCCAATCAAACCAA
CTTTAACCTTCGCTGGTTTGTGGCTGCTGACCTAGAGAGGAATTTGTTATATGGCTGAGGT
CCTGTCCTACCTAAGCCAGACTCCAGTGCAAAAAAAAAATCCTGTCAGATGCGCCTCTTCTT
CTCTGTCTCCATCTCCCTCTCTGCCTCCCGTTTCCCCTTTACTCTTTACTGTTAGTTGGCCT
GGGGTTCTGGGCTGAAGTGGAGTTTCTCCTGTCTCCGTGTGTTTGCTGCTCTGGTTTGC
TCTTCTTCCAGGGCCCCCTTGGTAAAAGAAGAGCTGCTGACAGGGTAGAGTGGAGGTGGGG
AAGTGACCACCAGACTGGGAGTGATGTAGGTTGTGGTCATCAATGATGTAGGTTGCTGTCA
CCAGGACACTCAAATCAGCTGTGACCAGAG (SEQ ID NO: 5)

eHGT_703m (322 bp):

TGGGCTTGATAGACAGACAGCTTCAGGTTTCCCGGGTTCCAGCAGACTTCCAGCCCCCTAC
CTGGCCTACTCTGTGATGACACAGAGCTGTAAAAGGCTGATGCAAGTTGGAAGAAGGGAA
TGAAACAGACTCTGTGGGGCAGATGGAGCTGCTGCCGGCCCCGTGGGACACAGGTCCCT
GTTCCACCTGCCTGTTTGCACCTGCTGATGTAATCACCCGGCCGCCCTTGGCTCAGGGCTG
GGCCTGCTCACCTGGTTTTCTAAGGGAGCTGCAGGTGGCGGGTGAGAGGGCAGGAGGGC
AGGGGATGACACAAGTGTCCAGG (SEQ ID NO: 6)

eHGT_087h (573 bp):

AATGGGGCCTCAGTTTCTCCCTTGGGATTAAGTGGGTACCAGGCAGGACAAAATAAGCCA
CCTGTGCCATTCTACCCTTTCCATCCTGCTGTTTTTCCACCCTGCAGCTGCTTTCTCAAAGA
ATCCCTGAATCTCCTCTCTAAATAGTGCAAGCTTGCTGTTGGGAACACCCACAGATGCACT
CTACAAATACACAACCTCACTTTGAGTAATCATATTGATTCGGAATTATCGGCCATTCAAATAC
CTCATTGAAGATGCCATCTGGGATGTTATTGCTCTTCAGAGGTAGATCATATTCTGAGAAAA
TCTGTCAAAATGTAGAAATAACTCATCAAAATGGCCATTGCAGAACTAGCTGCACTTCTTA
AGGAAGCAGACTCTGATGCTGCGGGTTTCATTGCCCTAAGTGAGATGCTGAAGGAAGAA
GGGCTGTTTCTTTTCTGTTTTCAAGTGTGATGAGAAATTTAAAGTTTCAGGAATGTGCAAGTT
CAGAGTGCTCCTGGACTCCTTTGCTCATTGAGAGACATCTCACTGAAAACAAACATTCTCTCA
AGGTTATCTCTGGTCCTGG (SEQ ID NO: 7)

eHGT_128h (531 bp):

CATCAAGGTCAGGGGCCCTTACCTCTCCTAGTATCATGAGGAGGATCTCAAAGCAGGGAAG
GGAAGCCAGCTGAGGGACCCCTGAGAGGAGATCAGGGGATAAATACTTTGCAACAGAATC
CCCATGGGAAACATGACAAAATTAGTAACAGGAATGGAGCCTCCCTCAGAGGATACATGAA
TCGTTGAGTTGAAAATAACCACAAGGTCATCAAATCCAGCTGCCACCAACTGCCTGAATATT
TTCAACAAATTTTTGAACAAGCAGTCCTCCAGCCTCCTCCAGCTGCATCCAGCAGCTCAGT
GCTCCACCTGACAGCCTCTTCTGATAGCGTGAAGGCTCAGAAAGCCCCTGTGGGGCCCAG
CACATCCTGGGGCAGGCCTCTAGTGGGCAGAGGGCAGCATACTGGGGTCTGGATCTAAC
ACACTCCTCCATCCAGGTGGTATATTTATAGCCCTGTGCTTTACAGAAAAGACTTCTGATGA
GATCAAGACTTTAGGCTGTTTATGTGGCTTCTTTCATGAGCCAAT (SEQ ID NO: 98)

eHGT_181h (263 bp):

TGTTGGGCTTTCTCTTGGGTTTATGACATGTGCTCCATTGTACTAATCCATTTCTCACAGTT
CCATCTCTCATAGCAATTCATGTAATTGCTGCAATTTTAAATTTTGGAACTATGTAGTGG
GGGAGGGGTGCAAGAAGATCTGGCAGTTTCTGCTACAGCTAGCTAATGGCTAATGGGGGT

30/86

FIG. 8 cont'd

GCCCCATAGATCTTTGTGGCAAAACAAAGAGTTAGACAGTGCAAAACAAGCTCTGGAAGGA
ACAGGGTCTGGCAGGAGA (SEQ ID NO: 8)

eHGT_260h (705 bp):

GCCTGTGGGACTCTGTGTGAGTATCCTCTCTGGAGCAATGCCTCTGTGCAATCTCTAGAAA
ACTCCCTATGTTAGTCTCGAACCACATGGATTGAAGGATTCTCCTGTAGCTAAGATTTTAA
AAGACCATGGCAGCACTGGGGAGCCCTGGAAGTCTCCCTCTTATCCTATACCATGTCTGG
GAGCTTCTCTTGGCTCCCTGCCAACCCTAGCCAAGCAGGAGGTCTTGCTTCCCTCTCCTTA
CTAGCTGTCGGTGCTTCCCTTCACTTCTGTGTTGAATCCCAAGTGTCTCTCCTAGACAATCTA
TTTTAAGTGTGAATATCTACTTGCTATTTTGGTTCCTCTCTGTGGAGGAGTGACTAGCTGC
ATCTAGTCAGCCATCTTGAATCTTCCCCCGCTTTTGTAAAGTGTTCCTTAATTTAGGTTTCCC
TGATGTTTCTTTATGATTCTGTTCAGGTTATGCTTCTGTGGCAGGAATATCACAGGAGCATT
GTTGTTCTCTTCTCATTGCATTAGATTAGGTGACACATGATTTGGATTTGTCCTATTATCGGT
GATGTAACTTTGATCATTTGATCAAGGTGATATCTGCCAGGTTTCTGCACTGTAAAGTTAC
TGTTTTCTCTTTTGAATTAATATATGTTTTGTGGGGAAATTTTCAAGACTGTTTAAAATATC
CCCGTTGTCATAAACTTTCACA (SEQ ID NO: 9)

eHGT_356h (564 bp):

AGGTGGAGTGAAGACTGAGAGGGGTGGTAGGGGGAGATGAGGCCGGGAGGAGGAGAGT
ACACCCTCAGGGATGGAGAGCCAACAAAGATTCTAAACGAGGAGTTCTGCAGTGAGACTG
GCATTTTGGAAACAATCGTCCTGGTTTTTGGTTGAGAGCACAACTTAGAGAACAGAGGCTAGA
GGTGGCAGGCTGGAGTGACAGCCTACTGCAGCTGTCCAGCCAATTAAGCCCTGTAGGAAG
AATGGCTGATGCCGCAGTCTATGTGATGCTGAACTGCCGTAGCCAGGCTGACCTGGCCCC
AATCCCGGGCTGTGACTGACCTTCAGATGCAGGCAGCAGAGAGAAGAGGACAAAAGTGCA
CCGCTGGGGAAAGGCAAGCAGCTGGCACCCAGGGCACCAAAGACCCGCACAGTGAGCAT
CAAGGCCAGCAAGGGGCACAAAGCCAGCTGTCACCTGTCTCTGGGCCAAGCCAAGGGGG
GCTGGCCTCCTGGCCACCTCCTTGGGGGAGCAGTACCCCCATTGTCATGACTTGTTTACA
GATTCCCAGACCTCCCAGTGCTATCAA (SEQ ID NO: 10)

eHGT_375m (322 bp):

TGCAGAGGGAAGATGAAGTCTTTGATGATAGCCAACTTAGTGCAATAAGTGGCTTTCTTT
GAGACATATTCAGATGGGAACGTCTTGCTTGCCAATTGCCATAGAAATCTTAACACACCAT
GAAGATTGTCCAAGCGCCAAGCCTTCCGTTCTGGACTAAATTACTTTGAAGTGGCTCAGGA
CGAGCAGTGGTCAATTTTAACTCTATAGACTGGACAGAAGAGGCTGGGAGTGGGAGATTG
TGCGTTTTTAAAGCAGAAAATAAGAAGGGGAACTTGTTTTATACACTCTATACAAGCTTCTG
CTCATTGTCAGATATCT (SEQ ID NO: 11)

eHGT_387m (492 bp):

TGGCTCTGGCAAAGGATGGAGATTTTTTTTTTTAACCTATTCTTTTGAGAATAAGGAGGGTCT
TTGTTTCTCTTCCCTGGGAACCGAGCAGCCCTTTCTCTGAGGGTAGAGTTGGGGCATGG
CTCAGCAGAACGACAGTCATCCTGGCTAGTAAGACTCAGAGGCTGGCCTTCAAAGGCTTG
AGCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGCCCTTGAATTGCTCCCCAGGC
GGGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGATTCTTAAAGAACTGGTATAA
TTTTGCACTGCATAGCAGACTCCCAAGACACACAGCCTTTTCCAGGAGGAGTTCTTATAGAG
GGGAAGGGGATAGGTTGCAATGGTCTGCTTAGCCTCCCGAAGGCTGCAGAAGCTGCTAAG
ACAGGGGTAAGTAAAGATAGACTCTGTGAGGTTGAACAGCAATCTTTATGACCTGACCCC
CTAGGCA (SEQ ID NO: 12)

31/86

FIG. 8 cont'd

eHGT_395h (624 bp):

CTGGCCTTGGTCTCTCCTGGGCTAGAACAGGCCCTTGTCTGCCTCCCTGAATCACTCTTG
AGGCTGGGCATGCACCTTCCCCACTCCAGATGTGCTGACAGCAGAAATGCCAACCCCTC
CCTTGAGGGCCAGTGAATTCAGCAGTGACAGGTCACTCTCTGAGCGAGACTGGTTCTCCC
TTCATCCGCTTGTAAAGGGAAAAACAGCCTCGGCCTCTGAGCCTGGCACCCAGGTTTG
GCCAGTCTTGTTGTTTGCCTAAGTCCAGTGCCCTTCAGTTCCCATCCCTCTTCTCTGACC
CCCAACTCCTGAGCCCCCTTGCTGTACCCCTCCCCACCCAGCTGCTGCATGTGCCAAC
CCCCACCCAGGAGAAAAACGAGGCTCTTGGAAGGGAATAAATGGAGGCTCTGTGCAGGCT
TGCAGCTGAAATGGCCCTGTGCGCGGCCACAGATGGGCCTCCTCTCCCCCTCCACATTTT
TGCATCAACAAAGCGTGACCTTGTGGATGACAGTCGCCCATGTTTTCCCATGACAAT
GCTTGCCTTGCCTTTCCTCCAGCGGGCTGTCTCAGGACATCTCTGTTCCCTGAAATTGG
GAAAGGGTGGTGGGGTGGAA (SEQ ID NO: 99)

eHGT_007m (508 bp):

CTGGTAGAGTGGCAGAGAGAGAGAGAGAAAAAGGAAACGCTAGAAAAGCTGAGCTGCAAG
CCCAGGCTGCCATTTGAAAGAGCAGTTGAGGAAGATCTGAATGAAAACTCTGCAAATTTA
TGCTAATTGTGTTCAAGTTCCATGAGTTTCATCCCATTTGAGTGGGAGCCAGGCCACCTGATT
GCTTCTCCTTTTGGTAAAGAACATCTCAGATGAACCAGCTGTCATGGACATGGTTCAGAAA
ATTGGCCCTAAAAAGAAACCTGGGCCATCTGTGTGTGCTGTTCTATTTCATAGGATGCCCTC
CAAATCACATATTTCTATTAGGAACCTCACAGCAAGGGCCTCTACCATGGGAGCAATAGAG
AACATTGGTCATCATCATCTATTGCACTCCCTTGGGGTGGGAGTGCTGTGATTTCTAGCAA
AGAGAACTCAGGGCTGAGGTAGATAACAAAGACAACCTTGAATGTCTAACGTGAGGAAGC
ATCCTTCTCTTCCTTTTGCTTTG (SEQ ID NO: 13)

eHGT_453m (721 bp):

TCACAAATCAGTAAGAATTACACTCCAAAAATAAAACAGGGAAGGGTATTAGTATGTGATAC
GCATGGATAAAGTAGAAGAGGCCAAAAATACAAAGAACCGTAAACTCACTAGTGATGGAAG
AAACGCTAATTAATAATGAGCTTTCCTGACCCTTTATCACATGGGCACATTTTCAAAGGA
CAAGACTCTGCTGAGGAGAACACAGAACAGGATCCTTCAGACAAGCTTCAGCAACCAATTT
TAGGATGTTCTTAATAGATCGATTCTTCTACTTTTTTTAAAAATAACATAGATCTGAAAAGC
TTTATAACTTTTGGTTTTTTTTAAGGACCATTATTAACAGCATCTTCCAGTTTTCTAGAGTAA
ATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCAATTTGTAAAAAGACTATAAA
AATAGAAGAGAAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGCCTTCAGC
TCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTCACACTGCCCTGCCTTTGT
GTGCCCTTGCCTGTCCCTCTGTGACGGTCACCTTTTCGGTTCTGGCTTGGAGCCTGGTCA
TCAGTGCCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTCGTTGTCTAGTGA
TGTGACCCTACCTTAGACTTCCCATCTCTTGAGGTGGCTA (SEQ ID NO: 14)

eHGT_470m (611 bp):

GGAGGGGCATAGTGCCACACTTCAGTCTTCATTCTTCTTGAGTTTCATGTGTTTAGCAAAT
GAAACTTTTATATTTCTGTCTTTGATAGGCTTACCAAATCCAAAGTGCTCCCCCTCCCCCA
ACTATTTTCAGGTTGAATATCTGTTCTCTGAAATGTTTGGTCTGAAAAGCATATCCAAATTTGG
AAGATCTGCATATATGTAATGAGACCTCTTGGGATGCATCTCAAGTCCAAACACAAAAGTCA
TTTATGTTTCATGCACACTTGACACACCTAGCCTGAAGCTAATTTGATACCATGTCTTTAGT
GCACCTGTGTTCTGACTGCAAACCACTATAAGGAGTCAGGTGTGGAATTTTCCATTTACAG
CATCGAGTCACTGTTCCAAATGTTTCAGCTTTTAGAGTGCTCTGTGATTTGTATTTTGGAT
CAGGAATGCTCAACCTGTATTTAATGCTTTGGCCTGAAACCTTCCCTCAGAAGGAAATGT
TTGATTGCAGACTGGCATTTTTGCCTCCTGGGATAGTGAGTTGTGTTTTGACTTTTTTTCATTT
ATTTATTTATTTTATATTCATGCCTGATTGAGGGAGGCTTTTAGACCATGGT (SEQ ID NO:
15)

FIG. 8 cont'd

eHGT_479m (808 bp):

NCCTGCGGCCGCACGCGTTCCCTAACCTAGAAAAACCCACCATCAACTATTGAGGTAATGT
TCCNTAGATGCAGGGTCATCAGTGGGCATTGTTACTTACTGGAAGAATGCTCTCTGGCGGA
TGGGGGAGGGAAGCAGGAAACTCTGCGTTAAGCAGGGCATAATCTCTTCAGGGACCTGC
TTCAGTTTCAATCCGACTCAAGTCGTGCTTTGCAGCTGGCCCCAGTTTTGGCTGCTCTGAG
AGTCAGTTGTTGGTTATGCACTGCCTGGGGTGAGGGCCAACCAGGGAGGGGGTTGGAGA
AGGATGCGAATAGTCAGAGCAGCTGGGAGCAAATGTATCCAGTTAAGGGGATCAAAACAG
AACGTCAGCAACCCCCGTAATGCCCCAAATCTTGCATAACTCTGTGTTATCTAGGCCATCT
CCCCCAACCCACCCCGCTTCTCATTTGGAACAAGGTCTTTCTATGTATGTAGCTAAAGTT
GACCTTGAACCTCATGATCTTCCCTGCCCAGCCTCTGAGTGCTAGGGTAAAATAGGCGTGCA
CCATCCTGCCGGGCTTCACCCACTCCCTTTTCAATTTCTGAGACACTTCCAGTTATTCTTCCTC
ATTCTTACTCCATTCACTCATACTGACCACTTCAAAATTCTGGAACCTTCCAGCATGCACTG
GCTGCAGATGAATTGTGCTTTTCTGCTCCCTCATCTGACTCTCCTTCTCGTCTGACTACTTA
GCTCTCATCTTCTCTACGGAGTCACCAGTCTTTTTTTTTTTTTTTTCAAGACAGAGTTTCTC
TGAATAGCCCTGGC (SEQ ID NO: 16)

eHGT_494m (654 bp):

AGCAAAGTTGAATGACATAGTTTTGATGAGATATTAGGTAGATTTTCAGCTCATTTAGAGAAA
AGAATATATTGTTCTGCCAAAGAATTCTGTGTCCGCTTTTCTGGCCTGGGTTCATTGTGTGT
CACGCTACAGAACTGGTATTCAAGAAAGGAGCCTGGATTTCTAATCATGTTGATATTTCCCA
TGAGCTTTTTTCTCAGCTCCTTGGACTCTTCACACAAAACATAAAAGGTCATTAATTATA
GGGTCCTTATGTAAATACTGGCCTGTTTTGTGATTCTATTCTGAGCAACGGTTATATACATC
TGCTCAAATAAATTAATTTTCTTGTCCTCAAATGACTCATTCCTGGAACCTAGAAGTTCCCATG
CCAGATCAAATAGAGAAAGGGAGAGAGGACACAGCACAATGCTGCGGATGGGAACCTTATT
CTGCCCTATACCAATTTATGAGTACTGGCACAAAATTTTAACGGGAAAAGCTCTAAGGGAAT
ATGTTTGTGTGTTTGTGTATGTGTGTAGGAGAGAGAAGGGGAGAGAGAGAATTATGAATCT
TGTCTTCTCATGTGTAATACAGCCATGTTTGAGATATGAATGTAATTCAGACATACTTGCCTT
TGAGTATTTTCTCTATATTTAATGAGCTGTTCTG (SEQ ID NO: 17)

eHGT_467m (605 bp):

TGTAGAAGGCCATGCTCACACTTTGGGCATTTTCCCTCAGCTCCTCCATAGCCTTAAGTTTC
AGATCCCAGCTTCACCTGCTGTTACCTATTTTATCTGCTTTCTTCTTAAAGCTGGAAAGCTC
TCTGTACAGCAGGCACAGAATTGCTGATGACTTCATGGCAGAGGCCAAAAATGGTAATTTT
CTTTCCTTTTGTTCAAAGCACTGTGCATGAGAATATAGTTCTAAGTTACTGACTCAGAGATG
ATAGAAGATTTTGAAGAAAATGATGACAAGCAGTGTTTTAAAGGTGAATTAGATCATAGAC
AGGTGTTTGGAGGAAAAAGCAGGGGAAGGGAATTTTCTAAATAGAGGTGCGGTCACCAAGA
CATATTGAAGCAGTACAAATGACATTAGAGAGAACAAAGGTGCAAACGAGTAAAAGGCCAGT
CCTGGGGGAACCTGAGTGCTTTCTAAGCAGATGATTATGAGAACAAGGAACAAATGTCCCTG
AAACCAATTAGGAAATGGGACTTGAGATGAGAAAACCTGTTCTGGGATACACAGTGATCTA
TTTTAGAGTCTGAATAATGTTGTCCTATCTAGAGTTGAAAATGATAGGTCCCT (SEQ ID NO:
18)

eHGT_483m (716 bp):

GTGGAGGTTGGGTGTTTCCATCACCTGGCTAACAGAGATCACAAATCTCCTTTTGTTCATCTT
AGGGATAAGGTTTGTCCATAGCTCGACTTGTCTGGATTTCTCTGTCCATCACATGCCTATGT
ACGCTCCCCACACCTCTCTCCTAGTGGCTCCTTGGTTTAAATGCCCCAGATCTGCAGCTC
ACAGAAGAACATTTGGCTACTCCCGTTTTTAATCTACTACTTCTGTAATGGGGCTCGGGAG
CTGGGCTGGGTCCCCCGAACCTGTGGGCAGCTGGCAGCCTGACCCTGGGCTGAGTACCA
GGCATGGGGAGGCCAGAAGGAAGGGGCAGATTGAGCCAGAGAGGTAAGACTGCAGAGAG
CCAAATAGGAGGGGAGGAGGCCAGAGCCTGAGGGTCCAGATGGGGCTGGCTGGGCACA
GGAGGGCATCTGATGCAATCCGCCAGGAGAGATGCCAGCCGGCCGGGTGCTGCAGAGGC

FIG. 8 cont'd

TCTGTTCTTGAAGAGTCTAGTGCCTGCTCTATTAATTCTGGCTCTGTGGGAATTGGGGAAG
ATTTCCCTATGAGGTCTCCTGAGCCCCAGCAACCTCCCCAAAGGCCTTAGCTCATAAAAGC
ACCTGTGTCTGAACAAGCAGGGGGTGGGGACATAGAACACCTGGGTTGGTGGCTGCCTG
GGGAAGGGCAAGCTGAAGTATCTTGTTCAGTATAAACACAGTACCTCAGCCCTGC (SEQ ID
NO: 19)

eHGT_606h (824 bp):

GGGAAGAATAGTGAGTTCTGGGGTGGAGGGTTTTGATTTTTTAAAGTGGAAGTCAAGAAA
TGCACCACGGAGAAGGTGGCATTGAGTGAAAACCTCAAAGTGGTTTGAGAGAAAGTGGTG
CAGATATGCGGACAGCGCTTCAGGCAGAGAAAATAACCTGGGATGTCTGAGAAAAGCAAA
GAAGCTGGGGCTACTGCAGTAAGAGCAAGAGAGTTAAGAATAGGTGTAGCCAGAGAGTTA
TCAGAATATAAGATTGCATAGGGCATTTTTAGGCCAATATAAGGACTGGCATTTTTAGGTCA
TGATGAGAACTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAGCATTTTGAGCAGAG
GAGTTGAACTGTCTGATTTAAAATTTTAAAGGTTCTGGCTACCTTGTGAGAAGAGACTGAA
GTAAGATATGCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAATAATCTAGGCAAGAGA
GATGGATCAGGGTGATAACAGTAACTTCATGAGAGCATTTCAGTTGCAAATATATTCCAGA
TTTGCTGACCAGATTTGATAAGGAATATAGAAAAAGAGGAAAGAACAAGAGATAACAGAGT
GTAGGCAGAAAAGAGAAAAAGGTTCAAAGACTGAGCCCTGCAGTATTTCAACATTAAGAGCT
TGAATTGCAGAATAGGAACCATCTCAAGGCTTTAAATACAAGTGCTGAACACTCTCAAATTT
ATATTTGTATCCCAGACTTCTCTAAAATTCAAGTTCTAAAATCCAGTTGGTTACTTTATGTCT
CCACTTGGAACAAGTCTTGAAGCAGT (SEQ ID NO: 20)

eHGT_738m (897 bp):

GAATGTATTGGATTATAACATGCACTAGTATTAAGTATAAAAGTATGTTCAAGAAAATTTAAA
GTTATTTTAAAGACTGATGTAAGGAAAGAAGGAGGAAGTTAGTCCATTAAATCACACTGAGC
AAGAGCTAACTTCTACAGAGTAGTCATTAAGAGGTAATTTTACAATCTAAATTTTTTTTTTTA
AAGGATGCTTTCCCTTTGTTGTGGTGTCAAGGTGAGAACTATGTAGATTTGATTCCGTAAAT
ATCTATGGTAAGGCAAGTGACAGGGTGAAGTCAAGTGTGAGCTGATGGAAATAGTTAGAACAT
AGAAACAGTCAATAACTATTCATTTTGTAAAAAATAATAATAAAATAGCTCCTTCAA
GTCCTGAAGCGTAATAAAATATTAGAAAACACAATGGAAAAATGAGGCACTCACAGAAGAA
ATACAGTTGGCTCTAAAGCATAGAAAGTCTCAATTATTTCAAATAATAGAAAGGTACTTTAA
GACAGTGAACCCGAGTGGTCTCTGGAACAGGCAGCTTGGTCTCAATCTGGCAGTCAG
TGCTAACTACACTGTGATCCTGGAAGATATGCTTTACTTCTCATTCAAAAACCTCAGAAT
TATAGTGCTGCCTGTGCCCTGATCTACCTGATGCCCGGCCCTACCTGCCTTTGGGGGTTG
CAGGTAAGGTGGTGTGGAATCAGGTGAGAAATGTCACAGCAAGCACCTCTGAAAGCTGCT
GCAAGCTCCTTTAATACGCTCACCCGTGTCCCATTTGGTCCCAAGAAAGCATCTCTTCTAAG
CAGCAGTTCCTGATTGGCTGGCATTGTCTGTGTGTCAGCAATCCTTGGTCTCTTAGGCAGTCC
TCAATTAGGTCCTCCTGTAGAAGGTGACTTTGCC (SEQ ID NO: 21)

eHGT_796h (551 bp):

TGCTCAGCCTCTATTTCCCGTTGAACTGGTTTTGGTAAGATATTTTACAGGCTAACTGTGTA
TTTCATCTAGCAGCCAGCAAGCGTTTGCACATCGTTATAAATTATGTCTTGTTATAAACTTGT
TCTAACAACAATTCACCTCTGAGCAGGCATTCCGCATGCTCATTACATAATATTTTTGTCTCT
GTGCAGCTGCACATCTGCACCGTAGTGTAGCTCCCTGTCCCCGGGTGACCGATTGGAAGT
AGGTTTCTCCATCCCCAATAGGCCTCACTGTGTCTCATCACACCAAAGTATTTTTAAAAGG
CTTAATGAAGACAGCATCCTCTGTATCATCCCCGGGAAGGCTGAGCAGGTTGCGCACCA
TAAATACACTTCAACGTTCTATCTGCCTAAGCCTTGAATCCCCTCCTCTATGATATAGCAA
GGTAAATGCCTGCATCCTGACTGCATTTAATGGACCACGGGCCCTGGGTTGGGGTTGCAG
CACCCAGTGAGCTATGAGAGCAGCCCCAGCTGCTCCGTCTGACAGATTGAAGCCAGGCT
(SEQ ID NO: 22)

FIG. 8 cont'd

eHGT_710m (382 bp):

AGTGACTTGGTGCTATGAGCCATATTTTGCTGTTGCTGTTGTTACTGGTAGTTTTTGTAAAT
CTGGGGCTAAACCTTGGGGTCTGGTATGCTGTCATTTACCAGTGAGCTATACCCTGGATAT
TATGATTTAGATGAATGTGAAATATCACCCAGACATACATATACTAAACACTTGGCCCTTG
GCCCATGATGCTAAATGGAGGAGATAGAAGCTTTTGGGGCACAGCCTAGTGGAAGGAAAT
GAGGTCAAATGACATGTACTCTGAAAGGAATATGGGTATTCTGGGCTTGCGTTATTCTCTCT
CTCCCTCTCTCTCCCTCTCTCTCCCTCTCTCTTTCTCCTTTTCTCTTTCTCTCCTCG
CCTTGTTTTCCA (SEQ ID NO: 23)

eHGT_588m (655 bp):

TGGCATTCTTGGATGTGTTAGGCCTCTTGCTGCTGCTCAATTTCTTTCCGTTTGACACATCT
CCAAACCATCATGTAGAATCTCATGGCTGAATTCATGCCCCGGCTCCATCTAATATTGTGTC
TTGACACCATCAAGGGCACCATCTAGGCTTTCCAACCCATTTATAAATGATACGCCACATC
CCACAAGGGCTTATGACTTTTGTAGGCATCAATAGTTCTGATTCCCTACGATAATGGTTCAT
TTTTCTTAGTAATAGAGAGCAGATCTTTTTCTTGCGCTAATTTTAAGCTTGTGATTGTAGAA
CTATATGAAGACGAGGCTCTTAATCATCAAGGCATTTTTTCCCCTAAACACTTAGGCTGATG
CAATTTTCTGAAATTTAGAAAACAAATTTTGCAATCACTTTTCTTCTTGATTGAAAAGAAGGT
AAAAGGACAGATCAAAAAGGCATAATTGCATGGTACATCTAGTCAACAAGTGTACAGGCAG
AGTGTCAAAAGTGTGCATCTTAACTGCCTCAAGACTTATTATGATGTCAATTTGCCTCCAG
AGGATCCATATGTATTAGGAGGGTGGCAGGATGCCCTAGTTCACAGCTTAACCTCTGCTTC
CTAGCCTAGATTCCAACCTCTGTTTCATCTACCCGA (SEQ ID NO: 24)

Core-eHGT_023h v3 used in 3xPVALBCore enhancer (260 bp):

GGTCCTGGTTCTGACTCAAGCTGCAACCCCCAGAGGCAGAAAGGAGAGGGATGCATGGAG
GGGAATGTCTTTTTCCACCAAGCAGCTGCTCCCTTTACAAGAGGCCAGTTATGTAAGGAA
GGCTGCTTCCTCTCCCCTTGCTAAATCTTTTCTAGTGCAGCATAAGGCCAGAGCTAGGGAA
AGTCCACACGGGAATAAAAAGAGCAGATTGCTAAGCACAGAAGCCATTTATAACTAAAACG
TCTCAGCTCAGCTTAGGG (SEQ ID NO: 25)

3xCore-eHGT_023h v3 (3xPVALBCore) (1000 bp):

GGTCCTGGTTCTGACTCAAGCTGCAACCCCCAGAGGCAGAAAGGAGAGGGATGCATGGAG
GGGAATGTCTTTTTCCACCAAGCAGCTGCTCCCTTTACAAGAGGCCAGTTATGTAAGGAA
GGCTGCTTCCTCTCCCCTTGCTAAATCTTTTCTAGTGCAGCATAAGGCCAGAGCTAGGGAA
AGTCCACACGGGAATAAAAAGAGCAGATTGCTAAGCACAGAAGCCATTTATAACTAAAACG
TCTCAGCTCAGCTTAGGGACGCGTCCTAGGTACGACTCCACCCCTCCAGGAACCCCTAG
TGAGGGAGTGGTGGGAGCAGAGCCCTCATCACACAGACTTGGTCCTGGTTCTGACTCAAG
CTGCAACCCCCAGAGGCAGAAAGGAGAGGGATGCATGGAGGGGAATGTCTTTTTCCACCAA
GCAGCTGCTCCCTTTACAAGAGGCCAGTTATGTAAGGAAGGCTGCTTCCTCTCCCCTTGCT
TAAATCTTTTCTAGTGCAGCATAAGGCCAGAGCTAGGGAAAGTCCACACGGGAATAAAAAG
AGCAGATTGCTAAGCACAGAAGCCATTTATAACTAAAACGTCTCAGCTCAGCTTAGGGACG
CGTCCTAGGTCACGACTCCACCCCTCCAGGAACCCCTAGTGAGGGAGTGGTGGTGGGTG
AGTAAGAGTAGGTGGCAACTGGTCCTGGTTCTGACTCAAGCTGCAACCCCCAGAGGCAGAA
AGGAGAGGGATGCATGGAGGGGAATGTCTTTTTCCACCAAGCAGCTGCTCCCTTTACAA
GAGGCCAGTTATGTAAGGAAGGCTGCTTCCTCTCCCCTTGCTAAATCTTTTCTAGTGCAGC
ATAAGGCCAGAGCTAGGGAAAGTCCACACGGGAATAAAAAGAGCAGATTGCTAAGCACAG
AAGCCATTTATAACTAAAACGTCTCAGCTCAGCTTAGGGACGCGTCCTAGGTACGACTCC
ACCCCTCCAGGAACCCCTAGTGAGGGAGTGGTGG (SEQ ID NO: 26)

35/86

FIG. 8 cont'd

(CoreB)eHGT_121h enhancer (150 bp):

ATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACA
TCTGCTCTGCATGAAGCCTGGGTGAGGCTGCCCCCTCTCATCTGTGACCTCCTGTGGTCGG
GAGGGACTTGCTCTGGATTATATAAAGGCT (SEQ ID NO: 27)

3x(CoreB)eHGT_121h (450 bp):

ATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACA
TCTGCTCTGCATGAAGCCTGGGTGAGGCTGCCCCCTCTCATCTGTGACCTCCTGTGGTCGG
GAGGGACTTGCTCTGGATTATATAAAGGCTATGAGAGGCTAGAGTTGGAGAGCACAGTTC
CTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTGAGGCTG
CCCCCTCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGGATTATATAAAGGCTA
TGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACAT
CTGCTCTGCATGAAGCCTGGGTGAGGCTGCCCCCTCTCATCTGTGACCTCCTGTGGTCGGG
AGGGACTTGCTCTGGATTATATAAAGGCT (SEQ ID NO: 28)

eHGT_589m (558 bp):

TGGAAGGAGATGCATTTGTATTGAAATTTTCTGTTTCATTTAAAGTTTGTTTTTTCTTTAAAT
ACTGGGATATATTTTCTTCTGTACAGCCTTGCCCTCCTTCCAATTTATGTTTCAGGATAAAGA
TTACATTTAGGCCATTTATAAATGTTAGAGTCAATCAACACGGTAAACATTTCGTATAGCAA
TTATTGCAAATTAACCACTTGATGTTAATCATTTCTGTTTGTGCTTTTTTTTTTCCCCTCTA
ATGTTGATTTCTGAGGCCTTTAAAGATGGTAACTTTATTATCCTTATAAGATGCCAACATTTG
AGTCATCAACATTGGTTAAAAATAAAGGGGGATCACCTTGAGCTGCAGACATGCAAAATCAA
CTGGAAAAACACCAGCTGGCCTTAATCTTTCACCATTCAGTTTTGTCTGCCTCCTGTCAG
CAGGCTAGCAGGTAGGCTTTTTATAGAAACAGGGACTTAGCAGATGTAGCTGCTGTTTCCT
TCTTGCAACACTGAAATTCCTTTTTAGCCCCCTTGACACACACAGACACCAACCGAA
(SEQ ID NO: 100)

eHGT_086h (452 bp):

ACAGCCTTTTCTCCTCTGACATTCCTGACCTCCCCGTGGCCTGCCTTCCCCTGTCTCCTT
GGCCGAGCCATCCTCTGTACCTTCACTGTTTCCACTAGGTTATTGAGGCCTTTTGGTCAT
GAGCTCCTGCCACCAAATCAGTGCCTGAGGCTGCTGCATAACAATCTCCTTCTTGATTAG
AGCATCTTAATTATGCTCCAGGAATTTGCTTAAACAGATTAATCATTCTGGAGGACAGATTT
AGAAGCTGTGCTAGCTGAGGGATCTGCTCTGAGTGTCCCCGGGGAGATTCTGGTGACAAC
CGCCTGTTGGGCCTTTGCCAAGGCCTTGGGTGCCTCGGAGACTCGGACACTGGCTGGAA
GCTCTACGCACCCCCGAAGGCTTCTCTTTGCTCCTCTGGGCCTGTAGAGCAGGCTCTGAA
ACTCTTGGATGTGAAACGTGGCAGGAGT (SEQ ID NO: 101)

eHGT_963m (431 bp):

ATCTTCAGGGTCACAGGCACATAAAGGTGGAAGGGGAGAACATATGCAGTCCATCTGTCT
GTCTGTCTTTGCACATAGGCCTCTTTACCAAGCCACCCTCGGCTCGATCTCCCTGCACGCA
ACCTACCCCACTTAGTAGATCTTCAACCCTGATGGCATCATCACGTTTCTGTCCCCTGAC
CTCTGCTCCGAGCTTACCCCACTTGTCTCCATTGCAGTGCCTACAGCGCTCAGCCCGTATG
TGCCTTGACACAGCCTCTGTGTGCCGTGCGCAGCATCAGTGTGCCGTGCGCAGCCTCTGTC
CACGTTTACAGTGGGTGTTACCAGCATTCCTCCTGCTTTATCATCCTCTGTGTTGATGCGG
AGGTCAACACATCCTTGGCTGCTATGAGAACAGAAGGACACAGTCAGTGGACTTTGTGGG
ACTTGGA (SEQ ID NO: 102)

FIG. 8 cont'd

core3_eHGT_453m (250 bp):

AAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGCCTTCAGCTCAGC
AGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTACACTGCCCTGCCTTTGTGTGCC
CTTGCCCTGTCCCTCTGTGACGGTCACCCCTTTCGGTTCTGGCTTGGAGCCTGGTCATCAGTG
CCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTCGTTGTCCTAGTGATGTGAC
CCTACG (SEQ ID NO: 103)

3xcore3_eHGT_453m (750 bp):

GAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGCCTTCAGCTCAG
CAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTACACTGCCCTGCCTTTGTGTGC
CCTTGCCCTGTCCCTCTGTGACGGTCACCCCTTTCGGTTCTGGCTTGGAGCCTGGTCATCAGT
GCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTCGTTGTCCTAGTGATGTGA
CCCTACGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGCCTTCA
GCTCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTACACTGCCCTGCCTTT
GTGTGCCCTTGCCTGTCCCTCTGTGACGGTCACCCCTTTCGGTTCTGGCTTGGAGCCTGGT
CATCAGTGCCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTCGTTGTCCTAGT
GATGTGACCCTACGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAG
GCCTTCAGCTCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTACACTGCCC
TGCCCTTTGTGTGCCCTTGCCTGTCCCTCTGTGACGGTCACCCCTTTCGGTTCTGGCTTGGAG
CCTGGTCATCAGTGCCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTCGTTGT
CCTAGTGATGTGACCCTAC (SEQ ID NO: 104)

core2_eHGT_387m (184 bp):

GCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGCCCTTGAATTGCTCCCCAGGCG
GGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGATTCTTAAAGAACTGGTATAAT
TTTGCACTGCATAGCAGACTCCCAAGACACACAGCCTTTTCCAGGAGGAGTTCCTTAGAGG
GG (SEQ ID NO: 105)

3xcore2_eHGT_387m (552 bp):

GCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGCCCTTGAATTGCTCCCCAGGCG
GGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGATTCTTAAAGAACTGGTATAAT
TTTGCACTGCATAGCAGACTCCCAAGACACACAGCCTTTTCCAGGAGGAGTTCCTTAGAGG
GGGCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGCCCTTGAATTGCTCCCCAGG
CGGGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGATTCTTAAAGAACTGGTATA
ATTTTGCACTGCATAGCAGACTCCCAAGACACACAGCCTTTTCCAGGAGGAGTTCCTTAGA
GGGGGCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGCCCTTGAATTGCTCCCCA
GGCGGGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGATTCTTAAAGAACTGGT
ATAATTTTGCACTGCATAGCAGACTCCCAAGACACACAGCCTTTTCCAGGAGGAGTTCCTT
AGAGGGG (SEQ ID NO: 106)

eHGT_534h (813 bp):

TCGGATTCATTTTGCATAATTTTGCTTTTTGGTTAAAAATGCTAACATACATATAAAATCAAG
AGTACCGATACTTGTTCAGGTTGCATTTAACCAGGAAAAATCTTGTTCTGTGAGATTGA
ACAGCTTTCCTCGCACTCACGCTGAAATGTGCAGTTTATTGCATTTGCTTTCTAAATGAGAAC
CACATTTAATGAAACCCTGACCATGTAATCAAGATAATGAGGGATTTTAAATTTGGGTGAGT
GACTCAGTTGTGGAGTACACACTTTGACTTGTGTTTCGGATGGGAGGTTGAGAAATTATCAC
TGCTTGGTGTACATGAGACACAAAGCAGAAAACAAAATCAGCATGCGCTGTGTCATATTT
GTACTGAGATAACACCGAGGAACAGCTGTGACTCTTTCAAATGACTCCATTTGTTTGATAAT
TGCTGGGTAAAAATAGTGCCACTATTGGCTATTTTCAGGAAGACAGATGGCGTTCTAGATTA
TGTTCTTTTCTTGTACCCAAAAAGAACAGGACCAAGTCTTCTGAAAATTACTGTCTGTGT

37/86

FIG. 8 cont'd

GTGTGAAATAGCATCGGGGAAGGGGACAGCGCATTTCAAGCTTTTGCCCCGCAGGTAGGT
AGGAAGCAGGGGCTGAGCCGTAAAGGAAACAGGGAATTCAATTATTCAGCCTGAGCTAG
GGTGGGAAATCAAGGCATCTGTCTGTTGCTTAGGAACAGAAATATCTTGTGCTTCAGCTAC
AGTTGGAGAAACCAGGTCATTTACTTTGAGAGTTACCTACTGGTTATCCAAACAGAGCTG
AAGGAGAAAGCACCT (SEQ ID NO: 107)

core2_eHGT_475m (226 bp):

TATAACAAATTATTTCTATTCTTTCTTCACTATTTCAGGAACATCTGTCCACCACTTACATG
ATCACTTATCTTGCTATTGTGTCAATTTTGATGAAAAAGAATTTTTTCTAAATATCTAAATACAA
GGCCCCATATTAACAGTGCTTTTTTAAATCCCCACAGATGTGGGAGATGACCCCTTTCCATC
CCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTT (SEQ ID NO: 108)

3xcore2_eHGT_475m (678 bp):

TATAACAAATTATTTCTATTCTTTCTTCACTATTTCAGGAACATCTGTCCACCACTTACATG
ATCACTTATCTTGCTATTGTGTCAATTTTGATGAAAAAGAATTTTTTCTAAATATCTAAATACAA
GGCCCCATATTAACAGTGCTTTTTTAAATCCCCACAGATGTGGGAGATGACCCCTTTCCATC
CCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTTTATAACAAATTATTTCTATTCTTT
CCTTCACTATTTCAGGAACATCTGTCCACCACTTACATGATCACTTATCTTGCTATTGTGTCA
TTTGATGAAAAAGAATTTTTTCTAAATATCTAAATACAAGGCCCCATATTAACAGTGCTTTTT
AAATCCCCACAGATGTGGGAGATGACCCCTTTCCATCCCTGAAGATTGTAATTGGGCCAGT
CTTTAGTACAGTTTTATAACAAATTATTTCTATTCTTTCTTCACTATTTCAGGAACATCTGTC
CACCCTTACATGATCACTTATCTTGCTATTGTGTCAATTTTGATGAAAAAGAATTTTTTCTAA
ATATCTAAATACAAGGCCCCATATTAACAGTGCTTTTTTAAATCCCCACAGATGTGGGAGATG
ACCCCTTTCCATCCCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTT (SEQ ID NO:
109)

eHGT_830h (719 bp):

TATTGTCTTTACCAGCCCCCGTGGGGACCCAGCTCATATCCACCTCCTCTGGGATGTCAT
CTCTGAAGATACCCTCCCTCAGGACTGTTGATTCAATTCAGCAAGCACATAAGGGGAACTGC
TTCATGCCAGCTGCTGCCCTAGCTCCAATTACAGTACTTATAAAAATACACTGTGATGATGT
GGTTCCTGCCTCTCCTCTCCCCAAAAACCTCACTCAGTACCAAGATCTTTAAAGGCTGGGG
TCATGCCTTATTCCATCTTGTATTCCCTGCATGGGACACTGGGCCCTGACACACAGCAGGA
CTCAGAAAAATATTTGCTGAATGAATAAATGGACAAAAGAAAAATGGGATTCATTTTATCTTCAT
GTCATGGGGTCAAATTCTCATTTTGTGGTATTTTGGGAGAACACTTCTGAAGAATGTCCAG
GTCATACTGTGTCTCTGTGACTGTAAATAGGGCTCTGCCCTAATTAAGTACTAGCTAGAGG
GTGAGAGAGCCCCAGGTGCGCACAGGGCAGTGTGTGGTTTGAAGCCGGTCCATGGAGTA
GCGTCCACGCAGGACAGCCTTGCTGGAAATCTGCGGATGCTAACGTGGTCCCACACTGTT
CTTCCCCAGCCTGAGATCCTGATTAGGAAGTAGCAGCTTCTTCTTTAAAACAACAAAACCTG
TCCTGTCTTCATAGGAAAAATGGGAATCCTAGCTGAAGACAGAAGGGG (SEQ ID NO: 110)

eHGT_540h (261 bp):

CCTCCACGGCCACTTCAGTGCCCTTTATCTTCCCCCTGGGGGTTACCCCAGACCCCCAGT
CAAGCAGCCGCTGCGGGGCCGCTCCCCGCGGGCCGCGACCTAGTTCCCTGCCGTTATTT
TTAGGGCGCGGGATGGCACCTGCCACGTGCCGGGCCCGCCCGGAGAGTGGGTGGGG
GGAGCTGCGTCTGCTCAAGGGAGGCGAGGCCCTGTCCCCGAAGCACCCCTTCATCCC
ACCCACGCAATCCCCAGTTTTTTGT (SEQ ID NO: 111)

core2_eHGT_351h (176 bp):

GCCCTTCTTTCTGTGAAGGGTGACACACCTCCTAGAAAGAAATCTGACTCCCCTCATTGTG
TTTTTAAATAACTCTCATTATTCCTCCTCTGAAGGTTATTTTCAAACCTTTATAACTCGATT

38/86

FIG. 8 cont'd

TTTTTTTAAATGCAAAGACAAGGAATCAAATCTGCCAACTTTCTATCAGAG (SEQ ID NO: 112)

3xcore2_eHGT_351h (528 bp):

GCCCTTCTTTCTGTGAAGGGTGACACACCTCCTAGAAAGAAATCTGACTCCCCTCATTGTG
TTTTTAAATAACTCTCATTATTCCTCCTCTGAAGGTTATTTTCAAACCTTTTATAACTCGATT
TTTTTTTAAATGCAAAGACAAGGAATCAAATCTGCCAACTTTCTATCAGAGGCCCTTCTTTCT
GTGAAGGGTGACACACCTCCTAGAAAGAAATCTGACTCCCCTCATTGTGTTTTTAAATAAC
TCTCATTATTCCTCCTCTGAAGGTTATTTTCAAACCTTTTATAACTCGATTTTTTTTTAAATGC
AAAGACAAGGAATCAAATCTGCCAACTTTCTATCAGAGGCCCTTCTTTCTGTGAAGGGTGA
CACACCTCCTAGAAAGAAATCTGACTCCCCTCATTGTGTTTTTAAATAACTCTCATTATTCC
TCCTCTGAAGGTTATTTTCAAACCTTTTATAACTCGATTTTTTTTTAAATGCAAAGACAAGGA
ATCAAATCTGCCAACTTTCTATCAGAG (SEQ ID NO: 113)

eHGT_882m (696 bp):

CCTTAATAGTCTTAATATAAGGTACGATTTTATTTATGTCTTTCTGACAAAGGGTTCAGACCC
CTGGAGGCCTCTGTTAACACAGTACTCTCAGATGTAAGATTTTTTAACACAAATTCCCAACCC
CTTGTTAGACTCCTGGTTTCTTCATCAGGCACAGCACAGAGAGAAGCTCTAGATTAACGAT
ATAGGCAGGAAAGTTGGGGGAGAAATGAAAACCTGTCACCCGTGTCATTATTTCCGCATTA
AGCCTGAAGGAGTAGGAGAGGTGTCACTCTCAGCGTCTCTGCGGTTGCCTGAGCTGCATC
AGTGTCTGCAGCTGCATGATACAGTCATCAGAGCTCTCTGAAGCACTCACCTCCCCCCCCC
CCACCTGTGCACCTCTCTCCGTGGAAACAAGGCAATGTTTGTTCAAGCTCAATGGAAAATA
GTAGCAAGCGTGATCTCACTTTGCTTTTTCTTTTTAAGGTAACAGCTGATCGCTGGCTTGAG
CACTTGCCAAAGAACGAGGTGCCACCCCGAGCTCTGAGCATTTCATTGAACCTTCATAGCA
ACGGTCCTAGGAGGTGGCAGCTATTATTATCTGCATTTTACAGGTAAAGTATCTGACGCTC
TGAGGTTACAGCATCTGCCCACTAGCAAGAAAGAACTGGAGTTCAAAAGGAAGTCAGGC
CATTGACACAAAACGGCCTGTCC (SEQ ID NO: 114)

eHGT_1137m (723 bp):

GGGAGTCTGGAGAGAGGGTGGGAGCAGCCATTCTGCAGCAGTGCCTTCTTGGGGTCATG
GGTCTGTAGGTGCTGCTGTGGAGGGAGAGATCAGCCTATTCTGGCTTCATTTCTGAGCTG
CAAACCTGCCTGGGTGTCTGGAGAAGCAGGTTGGCGTGGTGGTTAGCAGTGCCTGGGCGG
GGTTGCCCGCTCTTGATTTATGATTTCTTTGTCTCTGTGGAAGCACTTAAGTGCAGGCTTTA
GTTCCAATGACACTCAGGAGCCTCTGGATTCCAGCACTGGGGATGGGGGTGGGGTAGAAC
GTTCTCAGGCCTCACCAACCCCTCCCTGTGTGCTGCCTTTGGGAGAGTCCCAAGGCTTC
AGCATTACTTAATTAATTAGGCCTCTACTGCTACATAGGCTCAGATTCAAAAGAACAGAGTG
GCCCACGTGAGCCATTCCCGGAAAAGTCTGATGGCTGGAAGCCAGAGGACTATGTGTCTG
CCTTGCTGCCCTTGGCCAGCCCATCCTGAATGCCAGACTCGGACAATGGAGTAGGTACA
GAAGGGTAAAGACAGTGCTTTCTGTACCAGTAAGTGGGCCCTGATCTGCTCTCTACAGCTT
CCAGAGAAAGGGCCTGGCCAATGAGCGGCCTTTTGTAGTAGCAGATACCTCACATGCATTC
TGATAGAAAGCCTGGCCCCAGATCACTGTGACTTTAGCCCTCAGGTTTCTTTGCACTTC
(SEQ ID NO: 115)

eHGT_381h (232 bp):

GGGCTTTCTAAACTGGCCTGGAACACTGTTAGTTAATAAACATTCCCTTTCAGGGTTGTTCTC
TTTGCTTCTGATCTCATTGATGATGATTGTTGGACAAATAAGGAAGTAAACAGTGCAGAGT
TTTTTGGCTTCTGCTCTATTAAAAACATAGCAGAGAAGGACAAGAAGTTCCTGCCTGCTAAA
TAGGAAGATCTAGGGAATCAAAGAGTCTCTGTCAAGCTTCTTGTA (SEQ ID NO: 116)

39/86

FIG. 8 cont'd

MGT_E118 (1184 bp):

CCAATGTCCCAGGTGAAAATGAACACTTGCAGGTTTCTTCTTCCAGTCCCAGCCCTCTGGC
CCACACCCCACTGCTGACCACAGGAGCCTCTCACAGTGTCTCACTCAGGGCTTCAAAACTC
TCTCCCTGGCAGCTTCTTTCTATTCTGAAGTCTCCCAGTCCAGCACAAATTAGTCCCCTCTC
CTGGACTCCTACTGCGCCGCTTGGTACATCACAGAGCACTGGCTGAGAAAATCGCCCTT
TAGTCCCCGCAAATCACTCTCTTACCTGCTCCACTGACTAGGCCTGTGCCATTATTTATTAG
GGAATGAGCGCTAGCTGTTGAGGGTCAATTATTCCATTGACCAGGCATTACAATAATTCTT
GGTAATTAAGAGAGGTACTTATGTGAATCTGGATGCGTATTGAAAGAAACATTAGTCCTTTT
GTCAGCTTGGCAAGTCTATTGTTCTGAGCCAGGCCCAACCAATTAACATCTTTTGCCAATC
CCTGTCAGCAGGGGCTTGCAAGGAGGAGAGAAAGGGGGGGCCGGTCAGATGCTAATTTAG
ATACAATTGTGTCAGGGTGCATGGGAGGGGTAACCTCTGAAGAAGAGACTCCAGTTTAATGA
AGCAAACAGCTTGGGTCCCAGTGACCCTGCAGGGGTAAATTTTAAAGGCCCTCTGTGTTT
CTGTGCCAAATTGCAGAGGGCCAGCAACAAGAAGCCCTTCAAAAAAAAAAAAAAGAGGGAG
CAGGATAATCATCCACTTAATCTGCTCATTAAACAAATCTCTCAGGCGGACTTAACGTGAAT
TAGCACTTCTCTGAAAGGGGCAGCCTTGATGAGCAGCAGACAGCTCAGATCATTTGGAAG
GTCACAGTTTTTTTTTTTATTATTATATAAGCCCAGTTGTTATTAATGTATTCATATTTGAAGAA
ACTGAGCAGAGACCTGCTTTTCAATCTAGTGAGCAAAACAGTACGTTTTCTTAACCTGGA
AAACAAAACATAAATGCTAATCAGGTCCCAGCAGAGATCAAATCAAGCTAAATATAGCAGG
CACCTTTTGTGGTTTTTTTTTGTGTTTTGTTTTGTTTTGTTTTTGTGTTTTTTTAAATACAGAAATTA
AGGAAAGAAAAAAAACCATTGAATTAATTGTCTTCCTCTGTGCTTCGCACCTGACACCAGC
GAGGCTCACACAAGGGCCG (SEQ ID NO: 117)

MGT_E122 (1146 bp):

AAGGCTATAGTCCTTGGTGTGAAAGACTCATTCAATTTCTCTTATCTTCAATGTTAGGGA
CTGATTCAAAACAGCATAATCAAAACAAAGCAGAGCACAGCATAATTAATCATGTCTTCT
GAATAGGTAATTACAAAATAATTGGATATTGGTAATTTTCTTTTACCTGTTTCTGGCT
CCTTTCTCAGTTAATATGTAGTGCTATGTACAGATGAAGTTCAGATGCAATTATGTACAGGG
TCTATCTCACCACCAGGAAGCAGGACTTTTCTGTGTGATCTTCCATTTCTGCAAGGACTAC
AGAATTGAGGTACAAGTGCCTTTTGTGTTTCTCAAGTTGAGTCCTTCTGTTTTCTTGTCACT
GGCTTTTCATCCTCTGTCCCACTGGCTGTGGCTCAGTGTTTATTTATTGGCTGTTTAGTATGG
AGAAACATCCAGTCTTTCAACATTGACAACAAATCCAAATAATGAGCCTGCCCATCACGCTA
CACTGGGTATCATCTTTGCCTGTTAGGCTCTTGGCTGGTTTTTTGCTGTATTTGACTATCCT
CAAAGTTTGGGAGTTCTGACTGCGGGTGTGTTATTTTCCAATTCTTAAACACTTGCCAAAT
CTTTTGTGATGCATCTTTATACTCATGACAGCTTCTAGGTCAAATTTTAGTTTTGATCCTGG
ATCTTTAGAAAGGAAAGCTAAGGAACGTGACAGACATAGCCAGCTCCTTTTTTTTATGCTGGCAA
TGGATACTGATTAGCTGTGTGCTAGTTGTTGAAGTCATAGCAACAAGTCCTACTTACTATCT
TTTAGGCATCTCAATTAAGTTTCTTTGGTCTTTGGAAATATAACTACAATTAAGTTTCATCTC
CTGGAGAAAATGAGTATGTTTCAGGAGAACAACTTCATAGAGTTAATTAAGAACTCTTTTCTA
GAGTTCATTTTTAACTGGTAAGTTATTTGTACTCCTTCATTTTCTTTTGTGTTCCAGGTTTTA
TAATAGGCTTTTGTGATGGATTTAATCTTTTAAATTTCAAACGACTATTTGCCAAGATTCTGGAT
GCCTTTGCTGAGGCTACCAAGTTAAGTATATGCATCAGCAACCAGAGGAAAAACACTAACG
TTGGGCAGAGAAGCAATGAATACC (SEQ ID NO: 118)

MGT_E146 (689 bp):

ACTCAGCTGCTTTTCTGAAGGCAGCTTCAGCTTTTCCAAAGCCAGGGTGCTGCTGTGGTC
AGCTAGTGCTGCACTAAATATCCTCTTCCAAGGGGCTCTGATCCTGGCTCCAGCCCTCTG
GCCTCTGGGACTGGTTTTCTCTATTTCTTTCTTCTCCAAGGTTAGCTTTTTCAGACTCAAC
AACATAAATCAAGAGCCTCTAGGGAGGTGAACAGAAGTGGCCGGAGAGGAAGCACTGCTT
CAGTGGGGGAAGGGGAATGTCCTGGACTAGCAGGCCCCAGTAGGGGTGGAGTGTGGGG
GTATGGTGCATACTCAGATGTCTCCTTGAAGGAGTGGGTGGCAAGGATTCTTCATAAAC

40/86

FIG. 8 cont'd

TGATAAGCCGGGACACAGAGTCCCATCCCCCATCCCCACCCTGTTACACTGCTATAAACAG
AGAGGGCTGAAGGCAAAAGGACCCTGCTTGCTGCTGCAGCTGGTCCAGTCTAGGGGGCT
CTTCAGTAGAGGACAGCTCTCTTCTTGTCACGTGCTTTCTCTCTGAACAGAGCTGGTTGT
TTTGGGGCACTTGGGCTCCTGTGGACTAATGGGGTCAAGGACCAGGCCAGACACACACTA
AGGCCTGGAGTCCCTGGGGACTAGTGGTTCCAGTTCACCTAACCTCTGCCCTTCAAATCTG
GGCTCTGGCTCTCCCAATTGCTTCC (SEQ ID NO: 119)

MGT_E16 (599 bp):

AGGGTGCAGGAGAAATGTGACCTCAAAGTCTTGTTCTATAACTGTTGGACCTTAGGAGAGA
TCTGTGCTCAGCAATTTACCAGGACACCCCCACCCACATGTCTTGACCACTGTCTGGATA
ACTGGTATGCAGGACCACACTAGGCTTACTCACAGTGTAACCTCTCATAACCATCACTGGA
GCCCATCCTGCCTGGTAGACAAGGATTCAACCATGACTCATTGTACTTTAGTGGTGCCATG
CTTAGTCATCAGGTGCCCTGTGCTCTGACAGCCGAGGGTCAGAGCTGGAATCACACTCTT
GTTGTCTTTTAATCTCTCCCTCCCTTTCTTCTTCTTTCTTCACTCTGTTGTGATTGCTCATG
GAACAGATCCTAGCTGGTCTCCCTGGCAACCTACATGATTTGAGCCCAACAGATGGATAAT
GGGGACATCGACTTCCAATGTCATTCAACAGAATCATTGCCAAGGGAGTCTGATGAGCAG
GCAACTGAGATGACACCCTTATCAATATAGCTTCATTTTGGCAATCTGGAGTAGGTGTTTCA
AAAGGAGAGCCCCCACTGATGCCAGCAATACAGAACGTTTCATGGGCAAG (SEQ ID NO:
120)

core_MGT_E116 (219 bp):

CCTGGGCCTGAGAGAGACTCCGAGCAGCTGCCTGTGCAGCAGTGTGAGCCTGCCAGGTT
GTGCTTGGGGACACCAGTGACACCAGCTGCTCCTCATATCTCCACACGAGGCTATTTTAAG
AAACTCCGCACTTGATCAGGGAGCATGTGGGGTTAAGCCTTATGTTTTCCACTTTGGTTTTT
AGATCCTAAGCAGGTGCTGCAGTTAAAAGTTCAAAAT (SEQ ID NO: 121)

3xcore_MGT_E116 (598 bp):

TCACACCTGGACGACACCTGTGAAGGTCACTGTCTTCACTCACACCTGGACACCTGTGAA
GGTTACTGTCTCCACTCACACCTGAATGACACCTGTGAAAGTCAAGGTACACAAAAAAGG
CACTTTTGTGCTTATGCTTCTGTTTTAATCTGGGAACCCAGATGTTCTATCAGCATGCTTTAT
TCTGAGGGCCCCCGCACATCCTTCATGTATGCAAGCCTTCTCTTATGTAACCTTTTGGCTTTC
TGTGGCCAGATGGCTTCTCTTTCATTCATCCCTATGTGAACCAACTATGACTCCAAAATAT
CTGGCCTTTGTGTATCACATATACTCCAGCTTTCTGGTCTTTTTTAGGTTCTTGAAATATGC
CAGGCTGTTTCTCTTCTGTTGCTTCCAGTAGCAGTCTGGTATACCTGGGATATTTTTT
CTCTTATCCTCATTATATGTGTTCTTTCTAAATGTCACCTTCTCATGGAATCCACTTTACCTA
ATTTGTGTAGATTTTTCAGAAATTTATGCCCATGCCAAAGATCTTACCATAACTGGATCTAATC
ACCTTTATGAAGGTAAGAAAAGATCCCAAATTTATTTAG (SEQ ID NO: 122)

MGT_E150 (583 bp):

CGCTTTACCCAGCCGTATCTCATTTTTAGACATTTAAGTTACTTCCAACCTTTTCTGCTTATC
AGTTACGTCAGCATGGCTATTTTTATACCTAGATGGCCTGTGTTTCTGACTGCTTGGAAGTA
GACTTACTGGGTCAAAGGGGATGCCAGTGTGATTTTCAGAAAACGATGCCCGTTTCTGTTT
ACAGCTGGTTTCTATAAAGCATGGGAACGGCGCAGGCACTATTTATGTGCCCTGCTGAATT
CTCAGCTCCCTGCTTCATCTGACCTTCCGCCACTACTCAGAATCGGCTCTGTTCTTCTCCT
GGGCTCCCGATGGGAGCCAAGCTGGACTTGTGTGAGTGTATTAAGAAGTGCTCGTTATAG
CTCATTGCCCATGGGATGTCGACTGCTGCTGATTTCTCAGGGCTGGAGATGACACATTTTCT
TCACATTCTTCAAAGTCTTAAGGTTTCAGCAGGAGAGAAGGAAGGTCCGAGCCAAGGAGA

41/86

FIG. 8 cont'd

GCTGCGGCCATACCGACAGGGAGGGCACTCTCTTTCTCTAGCTCTCCCCCCCCCCCCCT
TGAAAAGGGGAGAAAAGCTCCCTCCCTGCCTCCT (SEQ ID NO: 123)

eHGT_1032h (300 bp):

CAGCTCACGTCTGAATCCCTCGGCGCCCCCTTTCTCTTCTCCTAGCCCCCTTCCTCACGTCC
CCTGCCTCCGGGTATTTCTCTCTCCAATCCCCCACCCTCGCACCGCCTGATTCCGAGGG
GCGGGAGCGCATTGGGCTGCGCACGGGTGGGGGCGCCGCGCCAGCTTCGCGTAGCTGC
TCTGACGCCGCTGCCGCCGCCGCCGCCGCCGCCCTCCGCAGCCCAGCTCGCGCCC
CGCGGCAGCTCCGCAGTGCAGTACCTAGCCACCACCGCCGCCGCCGCCGCTCCGCCAGACCTG
CTGC (SEQ ID NO: 124)

eHGT_1027h (466 bp):

GATTCTAGATTTTGAGAGACAGCTACGTCTTTGCTTCTCTGGTAGTTCCTATTATTAATTAT
TGGGGTTTTGACGGCGTTCTGGCCTCCGTTTTTGTGATTTCATTTTAGTGCTGCGTTTTCTC
CTGACTATGTTGATGATTTGCTGCTTATATAAGGGAAAGGATGTATTGAGGTGGACTGATG
CTTGGGGTTTTGAGATCGGCTACTGAGATTGGGTGGCGGTTGGTGTGTTTGTGATGTGC
ATGGTAGGAAGGTCTTTTCTTTTGGGTGGTAGGGGATGAGAGAATGGTGGGAGAACGGTA
GATGAGATGCTACAGCAAACGGGGATGTGGGGGGGAGACGGAGGTGAGGTTGGAAGTCAC
CGGTTTGCAGAATTTCTATTCTGCACCCCCCTTTCTCTCGCTCCTGGAAAAAAATCCATTA
TCTATAAAACACATTATTCTTCAATTGGCCTTTTAGAG (SEQ ID NO: 125)

eHGT_1027m (455 bp):

TTCGAGAGACTACTACGTCTTTGCTTCTCTGGTAGCTCCCTAATATTTATTACTGGGGGGAC
TGATGGGGTCTGGCCTCCGTTTTTCTTGGTTTCTTTTTAGTGCTGCGTTCCTATGTTGATGG
TTGCTGCTTATATAAGGAAGGGATGGATCGAGGTGGGTTGATGCTTGAGGTTTTGAGATCG
GCTACTGAGATTGGGTGGCGGTTGGTGTGTTTGTGATGTGCATGGTAGGAAGGTCTTTTC
TTTTGGATGGTGGGGGACAAGGGAATGGTCGGAGAATGGTAGATGAGATGCTCCAGTGAA
TGGGGACGTGAGGGGACACAGAGGTGAGGTTGGAAGTCACCGATTTGCAGAATTTCTGTC
CTGTATCCCCCTTTCTCTCGCTCCTTGAAAAAAATCCATTATCTCTAAAAACACATTAT
TCTTCAATTGGCCTTTTAGAGAAGAG (SEQ ID NO: 126)

Beta-Globin Minimal Promoter (pBGmin/minBGlobin/minBGprom):

GGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTG (SEQ ID
NO: 29)

minCMV Promoter:

GAGGTAGGCGTGACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGA
TCGCCTGG (SEQ ID NO: 30)

Mutated minCMV Promoter (Sacl RE site removed; CMV*):

GAGGTAGGCGTGACGGTGGGAGGCCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGA
TCGCCTGG (SEQ ID NO: 31)

minRho Promoter:

GATTACGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTG
GAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCC
CC (SEQ ID NO: 32)

FIG. 8 cont'd

minRho* Promoter:

GATTTCAGCCGGGAGCTTAGGGAGGGGAGGTCACCTTCATAAGGGCTTGGGGGGGGAGTTG
GAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCC
CC (SEQ ID NO: 33)

Hsp68 minimal Promoter (proHsp68):

CAGGAACATCCAAACTGAGCAGCCGGGGTCCCCCCCCACCCCCACCCCCGCCCCACGCGG
CAACTTTGAGCCTGTGCTGGGACAGAGCCTCTAGTTCCTAAATTAGTCCATGAGGTCAGAG
GCAGCACTGCCATTGTAACGCGATTGGAGAGGATCACGTACCGGACACGCCCCCAGGC
ATCTCCCTGGGTCTCCTAAACTTGGCGGGGAGAAGTTTTAGCCCTTAAGTTTTAGCCTTTAA
CCCCCATATTCAGAACTGTGCGAGTTGGCGAAACCCACAAATCACAACAACTGTACACA
ACACCGAGCTAGAGGTGATCTTTCTTGTCCATTCCACACAGGCCTTAGTAATGCGTCGCCA
TAGCAACAGTGTCACTAGTAGCACCAGCACTTCCCCACACCCTCCCCCTCAGGAATCCGTA
CTCTCCAGTGAACCCCAGAACTCTGGAGAGTTCTGGACAAGGGCGGAACCCACAACCTC
CGATTACTCAAGGGAGGGCGGGGAAGCTCCACCAGACGCGAAACTGCTGGAAGATTCTCTG
GCCCCAAGGCCTCCTCCGGCTCGCTGATTGGCCCAGCGGAGAGTGGGCGGGGCGCGGTG
AAGACTCCTTAAAGGCGCAGGGCGGGCGAGCAGGTCACCAGACGCTGACAGCTACTCAGA
ACCAAATCTGGTTCCATCCAGAGACAAGCGAAGACAAGAGAAGCAGAGCGAGCGGGCGCGT
TCCCGATCCTCGGCCAGGACCAGCCTTCCCCAGAGCATCCCTGCCGCGGAGCGCAACCT
TCCAGGAGCATCCCTGCCGCGGAGCGCAACTTTCCCCGGAGCATCCACGCCGCGGAGC
GCAGCCTTCCAGAAGCAGAGCGCGGGCGCC (SEQ ID NO: 34)

SYFP2:

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGGTGGTGCCCATCCTGGTTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCAT
CTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCACCCGGGACAAGCAGAA
AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGT
ACAAGTAA (SEQ ID NO: 35)

EGFP:

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGGTGGTGCCCATCCTGGTTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCA
CCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA
AGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCT
TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAA
GAACGGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGACGCT
CGCCGACCACTACCAGCAGAACACCCCCATCGGCCGACGGCCCCGTGCTGCTGCCCGACA
ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACA
TGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAA (SEQ ID NO: 36)

43/86

FIG. 8 cont'd

Optimized Flp recombinase (FlpO):

ATGGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTTTCGACATCCTGTGCAAGACCCCC
CCCAAGGTGCTGGTGCGGCAGTTTCGTGGAGAGATTTCGAGAGGCCAGCGGCCGAGAAGAT
CGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGATGATCACCCACAACGGCACCG
CCATCAAGAGGGCCACCTTCATGAGCTACAACACCATCATCAGCAACAGCCTGAGCTTCGA
CATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGACCCAGAAGGCCACCATCCTGGAGGC
CAGCCTGAAGAAGCTGATCCCCGCCTGGGAGTTCACCATCATCCCTTACAACGGCCAGAA
GCACCAGAGCGACATCACCGACATCGTGTCCAGCCTGCAGCTGCAGTTCGAGAGCAGCG
AGGAGGCCGACAAGGGCAACAGCCACAGCAAGAAGATGCTGAAGGCCCTGCTGTCCGAG
GGCGAGAGCATCTGGGAGATCACCGAGAAGATCCTGAACAGCTTCGAGTACACCAGCAGG
TTCACCAAGACCAAGACCCTGTACCAGTTCCTGTTCCCTGGCCACATTCATCAACTGCGGCA
GGTTCAGCGACATCAAGAACGTGGACCCCAAGAGCTTCAAGCTGGTGCAGAACAAGTACC
TGGGCGTGATCATTCAAGTGCCTGGTGACCGAGACCAAGACAAGCGTGTCCAGGCACATCT
ACTTTTTTCAGCGCCAGAGGCAGGATCGACCCCTGGTGTACCTGGACGAGTTCCTGAGGA
ACAGCGAGCCCGTGCTGAAGAGAGTGAACAGGACCGGCAACAGCAGCAGCAACAAGCAG
GAGTACCAGCTGCTGAAGGACAACCTGGTGCAGCTACAACAAGGCCCTGAAGAAGAAC
GCCCCCTACCCCATCTTCGCTATCAAGAACGGCCCTAAGAGCCACATCGGCAGGCACCTG
ATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAGCTGACAAACGTGGTGGGCAACTGG
AGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCACCTACACCCACCAGATCACCGCCAT
CCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTACGCCTACGACCCCATCAGCAAGGA
GATGATCGCCCTGAAGGACGAGACCAACCCCATCGAGGAGTGGCAGCACATCGAGCAGC
TGAAGGGCAGCGCCGAGGGCAGCATCAGATACCCCGCCTGGAACGGCATCATCAGCCAG
GAGGTGCTGGACTACCTGAGCAGCTACATCAACAGGCGGATCTGA (SEQ ID NO: 37)

10aa:

TCCGGACTCAGATCTGGAGGCTCCGGAGGC (SEQ ID NO: 38)

H2Bmod:

CCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAA
GGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCT
ATGTGTATAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGCCATGG
GCATCATGAACTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGTGAGGCTTCCCGCCT
GGCGCATTACAACAAGCGCTCGACCATCACCTCCAGGGAGATCCAGACGGCCGTGCGCC
TGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACC
AAGTACACCAGCGCTAAGTAATGAGTCGACGGCGCGCCCCCTGCAGGGAATTTCGATATC
(SEQ ID NO: 39)

hsA2:

TTAGAACAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTAGTTGCTGTTCTTCCAC
GTCAGAAGAGGACACAGACAAATTACCACCAGGTGGCGCTCAGAGTCTGCGGAGGCATCAC
AACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGACCTTTTACTACCTGACT
AGCT (SEQ ID NO: 40)

Improved Cre recombinase (iCre):

ATGGTGCCCAAGAAGAAGAGGAAAGTCTCCAACCTGCTGACTGTGCACCAAAACCTGCCT
GCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAGGAAGAACCTGATGGACATGTTTCAGG
GACAGGCAGGCCTTCTCTGAACACACCTGGAAGATGCTCCTGTCTGTGTGCAGATCCTGG
GCTGCCTGGTGCAAGCTGAACAACAGGAAATGGTTCCCTGCTGAACCTGAGGATGTGAGG
GACTACCTCCTGTACCTGCAAGCCAGAGGCCTGGCTGTGAAGACCATCCAACAGCACCTG
GGCCAGCTCAACATGCTGCACAGGAGATCTGGCCTGCCTCGCCCTTCTGACTCCAATGCT

44/86

FIG. 8 cont'd

GTGTCCCTGGTGATGAGGAGAATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAA
 GCAGGCCCTGGCCTTTGAACGCACTGACTTTGACCAAGTCAGATCCCTGATGGAGAACTC
 TGACAGATGCCAGGACATCAGGAACCTGGCCTTCCTGGGCATTGCCTACAACACCCTGCT
 GCGCATTGCCGAAATTGCCAGAATCAGAGTGAAGGACATCTCCCGCACCGATGGTGGGAG
 AATGCTGATCCACATTGGCAGGACCAAGACCCTGGTGTCCACAGCTGGTGTGGAGAAGGC
 CCTGTCCCTGGGGTTACCAAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGCTGA
 TGACCCCAACAACACTACCTGTTCTGCCGGGTGAGAAAGAATGGTGTGGCTGCCCTTCTGC
 CACCTCCCAACTGTCCACCCGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGAT
 CTATGGTGCCAAGGATGACTCTGGGCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAG
 AGTGGGTGCTGCCAGGGACATGGCCAGGGCTGGTGTGTCCATCCCTGAAATCATGCAGG
 CTGGTGGCTGGACCAATGTGAACATTGTGATGAACTACATCAGAAACCTGGACTCTGAGAC
 TGGGGCCATGGTGAAGGCTGCTCGAGGATGGGGACTAA (SEQ ID NO: 41)

SP10 insulator (SP10ins):

GAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAG (SEQ ID NO: 42)

3xSP10ins:

GAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGGAAGCTACCCC
 TAACACACTATTCTACACACAGAAAATGCTCTTCACTAGGAAGCTACCCCTAACACACTATT
 CTACACACAGAAAATGCTCTTCACTAG (SEQ ID NO: 43)

Flag encoding sequence:

GACTATAAAGACGATGACGATAAA (SEQ ID NO: 163)

3XFLAG:

GACTACAAAGACCATGACGGAGATTATAAAGATCATGACATCGATTACAAGGATGACGATG
 ACAAG (SEQ ID NO: 164)

H2B:

CCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAA
 GGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCT
 ATGTGTACAAGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCTGCAAGGCCATGG
 GCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTCCCGCCT
 GGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCGTGCGCC
 TGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACC
 AAGTACACCAGCGCTAAGTAA (SEQ ID NO: 165)

H2B*:

CCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAA
 GGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCT
 ATGTGTATAAGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCTGCAAGGCCATGG
 GCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGTGAGGCTTCCCGCCT
 GGCGCATTACAACAAGCGCTCGACCATCACCTCCAGGGAGATCCAGACGGCCGTGCGCC
 TGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACC
 AAGTACACCAGCGCTAAGTAA (SEQ ID NO: 166)

WPRES3:

ATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAGTATGTTGCTC
 CTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG
 GCTTTCAATTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAAGTCAATCGC

45/86

FIG. 8 cont'd

CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGG
(SEQ ID NO: 44)

WPRE:

GCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTA
TGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTCATTTTCTCCTCCTTGATAAATCCTGGTTGCTGTCTCTTTATGAGGAG
TTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCC
ACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTTCGCTTTCCCCCTCC
CTATTGCCACGGCGGAACCTCATCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCATCGTCCTTTCTTGGCTGC
TCGCCTATGTTGCCACCTGGATTCTGCGCGGGACGTCTTCTGCTACGTCCCTTCGGCCC
TCAATCCAGCGGACCTTCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTC
TTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCATCGATACC
G (SEQ ID NO: 45)

BGHpA:

CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGAC
CCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGT
CTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGA
TTGGGAAGACAATAGCAGGCATG (SEQ ID NO: 46)

HGHpA:

ACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTC
CAGTGCCCACCAGCCTTGTCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCC
TTCTATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACA
ACCTGTAGGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGG
CTCACTGCAATCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGT
TGGGATTCCAGGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTGGTAGAGACGGGG
TTTCACCATATTGGCCAGGCTGGTCTCCAACCTCCTAATCTCAGGTGATCTACCCACCTTGG
CCTCCCAAATTGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCTGTCTT (SEQ ID
NO: 47)

P2A:

GGCAGCGGCGCCACCAACTTCAGCCTGCTGAAGCAGGCCGCGACGTGGAGGAGAACCC
CGGCCCCGGAGCTAGCGGA (SEQ ID NO: 48)

T2A:

(GSG)EGRGSLTLCGDVEENPGP (SEQ ID NO: 49)

E2A:

(GSG)QCTNYALLKLAGDVESNPGPP (SEQ ID NO: 50)

F2A:

(GSG)VKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 51)

Exemplary Plasmid Backbone 1 – Left ITR:

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTCGGGCGACCTT
TGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCA
CTAGGGGTTCT (SEQ ID NO: 52)

46/86

FIG. 8 cont'd

Exemplary Plasmid Backbone 1 – Right ITR:

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCGGGCG
TCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTG
GCCAACTCCATCACTAGGGGGTTCCT (SEQ ID NO: 53)

Exemplary Plasmid Backbone 2 – Left ITR:

CATGTCCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCG
GGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGG
AGTGGCCAACTCCATCACTAGGGGGTTCCT (SEQ ID NO: 54)

Exemplary Plasmid Backbone 2 – Right ITR:

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGG
CCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGA
GCGAGCGCGCAGCTGCCTGCAGGGGCGCCTG (SEQ ID NO: 55)

PHP.eB capsid:

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLPGNGLDK
GEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKEDTSFGGNLGRAVFQAKK
RLEPLGLVVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDP
QPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRTVTTSTRT
WALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSRPRDWQRLINNNWG
FRPKRLNFKLFNIQVKEVTDNNGVKTIANNLTSTVQVFTDSYQLPYVLGSAHEGCLPPFPADV
FMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVFPFHSSYAHSSQLD
RLMNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
NNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK
VMITNEEEIKTTNPVATESYGQVATNHQSDGTLAVPFKAQAQTGWVQNGILPGMVWQDRDV
YLQGPWAKIPHTDGNFHPSPMLGGFGMKHPPQILIKNTVPADPPTAFNKDKLNSFITQYST
GQVSVEIEWELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL
(SEQ ID NO: 56)

AAV9 VP1 capsid protein:

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLPGNGLDK
GEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKEDTSFGGNLGRAVFQAKK
RLEPLGLVVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDP
QPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRTVTTSTRT
WALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSRPRDWQRLINNNWG
FRPKRLNFKLFNIQVKEVTDNNGVKTIANNLTSTVQVFTDSYQLPYVLGSAHEGCLPPFPADV
FMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVFPFHSSYAHSSQLD
RLMNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
NNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK
VMITNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGWVQNGILPGMVWQDRDVYLQGPW
AKIPHTDGNFHPSPMLGGFGMKHPPQILIKNTVPADPPTAFNKDKLNSFITQYSTGQVSVEIE
WELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL (SEQ ID NO:
57)

tet-Transactivator version 2 (tTA2):

ATGTCTAGACTGGACAAGAGCAAAGTCATAAACTCTGCTCTGGAATTACTCAATGAAGTCG
GTATCGAAGGCCTGACGACAAGGAACTCGCTCAAAAGCTGGGAGTTGAGCAGCCTACCC
TGTA CTGGCACGTGAAGAACAAAGCGGGCCCTGCTCGATGCCCTGGCAATCGAGATGCTGG
ACAGGCATCATACCCACTTCTGCCCCCTGGAAGGCGAGTCATGGCAAGACTTTCTGCGGA
ACAACGCCAAGTCATTCCGCTGTGCTCTCCTCTCACATCGCGACGGGGCTAAAGTGCATCT

47/86

FIG. 8 cont'd

CGGCACCCGCCCAACAGAGAAACAGTACGAAACCCTGGAAAATCAGCTCGCGTTCCTGTG
TCAGCAAGGCTTCTCCCTGGAGAACGCACTGTACGCTCTGTCCGCCGTGGGCCACTTTAC
ACTGGGCTGCGTATTGGAGGATCAGGAGCATCAAGTAGCAAAAGAGGAAAGAGAGACACC
TACCACCGATTCTATGCCCCCACTTCTGAGACAAGCAATTGAGCTGTTCCGACCATCAGGGA
GCCGAACCTGCCCTTCTTTTCGGCCTGGAACTAATCATATGTGGCCTGGAGAAACAGCTAA
AGTGCGAAAGCGGCGGGCCGCGGACGCCCTTGACGATTTTGA CTTAGACATGCTCCCAG
CCGATGCCCTTGACGACTTTGACCTTGATATGCTGCCTGCTGACGCTCTTGACGATTTTGA
CCTTGACATGCTCCCCGGGTAA (SEQ ID NO: 58)

GTPase HRas [Homo sapiens]:

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILD TAGQEEYS
AMRDQYMR TGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDLAARTVESR
QAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREIRQHKLRLNPPDESGPGCMSCKCVLS
(SEQ ID NO: 59)

Substance P is position 58-68 of Protachykinin-1 [Homo sapiens]:

RPKPQQFFGLM (SEQ ID NO: 60)

Oxytocin is position 20-28 of Oxytocin-neurophysin 1 [Homo sapiens]:

CYIQNCPLG (SEQ ID NO: 61)

GCaMP6m:

ATGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCC
ACAACATCGAGGACGGCGGCGTGCTGCTGCCCCGACAACCACTACCTGAGCGTGCACTCCAACTTTTCG
GGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCGTGCACTCCAACTTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTCAACGGGGTGGTGCCCTGCTGAGCTGAGCTGAGCGGCGAGCTA
AACGGCCACAAGTTCAAGCTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCCTCGTGAC
CACCCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCTTTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCTTGACAATGATGGCAAGAAAAGGGA
GCTACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGACGACAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 62)

GCaMP6s:

ATGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCCACATCCGCC

48/86

FIG. 8 cont'd

ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCGTGCAGTCCAACTTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCTTTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCAACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCTTGACAATGATGGCAAGAAAAATGA
AATACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 63)

GCaMP6f:

ATGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGGCAACTTCAAGATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCGTGCAGTCCAACTTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGAATTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCAACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCTTGACAATGATGGCAAGAAAAATGA
AATACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 64)

CN1259 (1956 bp between ITRs):

CCTAGGACGCGTCCCTAAGCTGAGCTGAGACGTTTTAGTTATAAATGGCTTCTGTGCTTAG
CAATCTGCTCTTTTTATTCCCGTGTGGACTTTCCCTAGCTCTGGCCTTATGCTGCACTAGAA
AAGATTTAGCAAGGGGAGAGGAAGCAGCCTTCCTTACATAACTGGCCTCTTGTGAAAGGG
AGCAGCTGCTTGGTGGAAAAAGACATTCCCCTCCATGCATCCCTCTCCTTCTGCCTCTGGG
GTTTGCAGCTTGAGTCAGAACCAGGACCATTTAACTCCAACCTTTGAGGAAGAGACGCAC

FIG. 8 cont'd

CCTGGCCCCAGCCACGCCTGTTAGAATCTTCCTAGCTGAGTGACACAGTGACACTCAGCC
 TCAGTTTCTCTGTAAACAAAATGAAGATAAGAGAGCTGACGAGGATGAAATGGAATAACAC
 ACCGTGCGGTGCCTGGTACAGAGTGAGCCCCAGAACTGTTGACGCGCCTCCTTGTGGCT
 GTCTGGCTTGACCCGGAGTGACTCTGCCTCCCAGTTCTCCGGGATGGGAAGGTGATCCCT
 GTTTGAGATACCAATTTATAAGAAACCGAGCCCCGGAGCTATTTAGAGGTGAGGTGATAAA
 CCAGGAGGCCGGCTCCTTCATCCCGGTCATCACGAGAGCTCGGGCTGGGCATAAAAGTCA
 GGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCT
 ACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCAACGGGGTGGTGCCCATCC
 TGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
 GGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCC
 CGTGCCCTGGCCACCTCGTGACCACCTGGGCTACGGCGTGCAAGTCTTCGCCCGCT
 ACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCC
 AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGT
 TCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
 GGCAACATCCTGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCACC
 GCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGAC
 GGCGGCGTGCAAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGT
 GCTGCTGCCCCGACAACCACTACCTGAGCTACCAAGTCCAAGCTGAGCAAAGACCCCAACGA
 GAAGCGCGATCACATGGTCCTGCTGGAGTTCTGTGACCGCCGCGGGGATCACTCTCGGCAT
 GGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTTCGATATCATAATCA
 ACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTAC
 GCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA
 TTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCGCCGCTG
 CTTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATCCGTGGCTCGAG
 CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGAC
 CTTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGT
 CTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGGAGGA
 TTGGGAAGACAATAGCAGGCATGACTAGT (SEQ ID NO: 65)

CN1621 (2170 bp between ITRs):

CGCGCCGCAACGCGTTTAGAACAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTA
 GTTGCTGTTCTTCCACGTCAGAAGAGGACAGACAAATTACCACCAGGTGGCGCTCAGAG
 TCTGCGGAGGCATCACAACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGA
 CTTTTTACTACCTGACTAGCTGTTTGTGTATTTTAGGTGTTTGTTCATCAAGGTCAGGGGC
 CTTACCTCTCCTAGTATCATGAGGAGGATCTCAAAGCAGGGAAGGGAAGCCAGCTGAGGG
 ACCCCTGAGAGGAGATCAGGGGATAAATACTTTGCAACAGAATCCCCATGGGAAACATGA
 CAAAATTAGTAACAGGAATGGAGCCTCCCTCAGAGGATACATGAATCGTTGAGTTGAAAT
 AACCACAAGGTCATCAAATCCAGCTGCCACCAACTGCCTGAATATTTTCAACAAATTTTGA
 ACAAGCAGTCCCTCCAGCCTCCTCCAGCTGCATCCAGCAGCTCAGTGCTCCACCTGACAGC
 CTCTTCTGATAGCGTGAAGGCTCAGAAAGCCCCTGTGGGGCCCAGCACATCCTGGGGCAG
 GCCTCTAGTGGGCAGAGGGCAGCATACTGGGGTCTGGATCTAACACACTCCTCCATCCAG
 GTGGTATATTTATAGCCCTGTGCTTTACAGAAAAGACTTCTGATGAGATCAAGACTTTAGGC
 TGTATATGTGGCTTCTTTTATGAGCCAACTGAATTCGATATCATAATCAACCATAGGTACCG
 AGCTCGGGATTACGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGG
 GGAGTTGGAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGA
 AGCATCCCCGGGTTGGATCCTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAG
 GGCGAGGAGCTGTTCAACGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAAA
 CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA
 CCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCTCGTGACC
 ACCCTGGGCTACGGCGTGCAAGTCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGCA
 CTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA

50/86

FIG. 8 cont'd

CGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACC
GCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG
GAGTACAACACTACAAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATC
AAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGCGCTGCAGCTCGCCGACCA
CTACCAGCAGAACACCCCATCGGCCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCT
GAGCTACCAGTCCAAGCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCT
GGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGA
CATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTT
GCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCG
TATGGCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCA
TCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCC
GTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCC
CCCGTGCTTCTTGAACCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGG
AAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG
ACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCC
GC (SEQ ID NO: 127)

CN2157 (1982 bp between ITRs):

GCGGCCGCACGCGTCTGGCCTTGGTCTCTCCTGGGCTAGAACAGGCCCTTGTCTGCCTC
CCTGAATCACTCTTGAGGCTGGGCATGCACCTTCCCCACTCCAGATGTGCTGACAGCAG
AAATGCCAACCCCTCCCTTGAGGGCCAGTGAATTCAGCAGTGACAGGTCACTCTCTGAGC
GAGACTGGTTCTCCCTTCATCCGCTTGTAAGGGGAAAAACAGCCTCGGCCTCTGAGCCTG
GCACCACCAGGTTTGGCCAGTCTTGTTGTTTGCCTAAGTCCCAGTGCCCTTCAGTTCCCAT
CCCTCTTCTCTGACCCCCAACTCCTGAGCCCCCTTGCTGTACCCCTCCCCACCCAGCT
GCTGCATGTGCCAACCCCCCACCAGGAGAAAACGAGGCTCTTGAAGGGGAATAAATGGA
GGCTCTGTGCAGGCTTGACGCTGAAATGGCCCTGTGCGCGGCCACAGATGGGCCTCCTC
TCCCCCTCCACATTTCTGCATCAACAAAGCGTGACCTTGTTTGGATGACAGTCGCCCCATG
TTTTCCCATGACAATGCTTGCCTTGCCTTGCCTCCAGCGGGCTGTCTCAGGACATCTC
TGTTCCCTGAAATTGGGAAAGGGTGGTGGGGTGGAAAGAGCTCGGGCTGGGCATAAAAGTC
AGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCAAGCTAGCGC
TACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATC
CTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAAGCTGTCCGGCGAGGGCGA
GGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGACGTGCTTCGCCCGC
TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA
GTTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAACAGCCACAACGTCTATATCA
CCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGG
ACGGCGGGCTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTTCGATATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTT
TACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTT
TCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCC
TGCTTGGCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCG
AGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG

FIG. 8 cont'd

GGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 128)

CN2339 (2450 bp between ITRs):

GCGGCCGCACGCGTCCCTAAGCTGAGCTGAGACGTTTTAGTTATAAATGGCTTCTGTGCTT
AGCAATCTGCTCTTTTTATTCCCGTGTGGACTTTCCCTAGCTCTGGCCTTATGCTGCACTAG
AAAAGATTTAGCAAGGGGAGAGGAAGCAGCCTTCCTTACATAACTGGCCTCTTGTGAAAGG
GAGCAGCTGCTTGGTGGAAAAAGACATTCCCCTCCATGCATCCCTCTCCTTCTGCCTCTGG
GGGTTGCAGCTTGAGTCAGAACCAGGACCATTTAACTCCAACCTTTGAGGAAGAGACGCA
CCCTGGCCCCAGCCACGCCTGTTAGAATCTTCCTAGCTGAGTGACACAGTGACACTCAGC
CTCAGTTTCTCTGTAAACAAAATGAAGATAAGAGAGCTGACGAGGATGAAATGGAATAACA
CACCGTGCGGTGCCTGGTACAGAGTGAGCCCCAGAAGTGTGACGCGGCCTCCTTGTGG
CTGTCTGGCTTGACCCGGAGTGACTCTGCCTCCCAGTTCTCCGGGATGGGAAGGTGATCC
CTGTTTGAGATACCAATTTATAAGAAACCGAGCCCCGGGAGCTATTTAGAGGTGAGGTGATA
AACCAGGAGGCCGGCTCCTTCATCCCGGTTCATCAGAGAGCTCGATTACGCCGGGAGCTT
AGGGAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTGGAGCCACGAGTCGTCCA
GCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCCTTCGAAGCT
AGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGGCGAGGAGCTGTTACCCGGGGTGGTGC
CCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAG
GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAA
GCTGCCCGTGCCCTGGCCACCTCGTGACCACCTGGGCTACGGCGTGCAGTGCTTCG
CCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCT
ACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG
GAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAATAACAAGCCACAACGTCTAT
ATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATC
GAGGACGGCGGCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGG
CCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCT
CGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTGGAGGCTCCGGAGGCCCCAGAG
CAGCAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAAGGCGCAG
AAGAAAGGCGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCTATGTGTAT
AAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCTGTCGAAGGCCATGGGCATCATG
AACTCGTTTTGTGAACGACATTTTTCGAGCGCATCGCAGGTGAGGCTTCCCGCCTGGCGCAT
TACAACAAGCGCTCGACCATCACCTCCAGGGAGATCCAGACGGCCGTGCGCCTGCTGCTG
CCTGGGGAGTTGGCCAAGCAGCGCGTGTCCGAGGGTACTAAGGCCATCACCAAGTACAC
CAGCGCTAAGTAATGAGTCGACGGCGCGCCCTGCAGGGAATTCGATATCATAATCAACC
TCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCT
ATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTT
CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCCTGCCTT
GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 66)

CN1280 (1930 bp between ITRs):

CCTAGGACGCGTGAAGCCGAGCTGAGCATTTGGAGTTTGAAATGTTTCCGTACTTAGCAAT
CTGCTCCTCTATTCCCGGGCGGACTTCCGCTAGCTCCGGCCTTATGCTGCACTAGATAAGA
TGGAGCAGGGAGAGGACACGGCACTACTTATGTAACCGGCCTCTTGAAAAATGGAGCAGC
GGTCAGGGCGGAACAAGACGTCTCTCTACGCATCCCTCTCCTTTCCCTGCTAAGGCT

52/86

FIG. 8 cont'd

GCAGCTGGAGTCAGAGGCAGGGCTGTTCCAATCTGTCTTTGATCAGTAACGCAGCCAGCC
TCCAGCCTCCGTCAGCCTCCTCATGGCTGAGACCCGGCCTCAGTTTCCCCCACTTACATC
CCGAGGATCAGAGCCTGTGAGGATGAAATGGGATAAGGTAGCTGGAACCGTCTGGCAGA
GAGCGAGTCCTCAGGACTGTTGATGCCTGTGGCTGCCTGGCTTGACCCCAAGTGACCCCG
CCTCCTCATCCTGCAGCAGGAGAAAGTGATCCTGTTTGCGGCACCCATCTATAAGAACTG
AGCCCTGGAGCTATTTTGAGGCGAGGTGAGGTGATAAACCAGGAGCTGGGCCCTTGCTAT
GGTCATCAGAACGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTA
CATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAG
CAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACG
TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGT
GACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGA
ACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACACAAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGC
ATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACAACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
ACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCGCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAG
TCGACGGCGCGCCGCGGCGCGAATTTCGATATCATAATCAACCTCTGGATTACAAAATTTG
TGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTT
TAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATAAAT
CCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAG
GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGCGACTGTGCCTTCTAGTTGCC
AGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATCTATT
CTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGC
ATGACTAGT (SEQ ID NO: 67)

CN2045 (2073 bp between ITRs):

GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTTCACAG
CTCCTTGGTCTGAACTGGTACCCAGGATCAAGATGTCCCTCCCATGATTCTAATCTCCAGC
CCCCTTACACCATAGTCACTTCCAACTCTGGATTTCTTAACTTGCAGTTTGCTTCCAAAA
AAAAATTAAAAAACCCTCATCTGGGCATTAGAAGTCATTCTGAAGAGGCTGTCCTATAT
CTCAGTGCAATTTCTACTTATCCACTCCCGACTATACCCTGCTTTCATCAGATTGCAACTG
ACCTAATTTGTAAAGACTTCAGTGAAAAGAGCCTCATTAAAGCCTTCATCTCTTTTGCAGC
CTCTCTTCTCCTCCTCAAATGATCCAGGTCAGAATGGCCCAATAAAAAATAGCTGGGAAAATT
ACAAAGTCACTCAAGACTCTGCCAGGAGTAGAGATTTTAAGAATTAGTGACAGACATATACT
GTCTAGGAGTAGGGTGGGAGTGGGTTGCATGGCTTAGGGACATAGAGAGCTCGATTACAGC
CGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTGGAGCCACG
AGTCGTCCAGCCGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCC
AGATCTTTTGAAGCTAGCGCTACCGGTGCCACCATGGTGAGCAAGGGCGAGGAGCTGTT
CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCA
GCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATC
TGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCTGGGCTACGG
CGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC
CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAA
GACCCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG

FIG. 8 cont'd

GCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACA
GCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGA
TCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACC
CCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAA
GCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGC
CGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATT
CTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCT
ATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCAC
GGCGGAACATCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCA
CTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTTCTA
ATAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGGTGGG
GTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCAGC
TGCGGACCGAGCGGCCGC (SEQ ID NO: 68)

CN1932 (2272 bp between ITRs):

GCGGCCGCAACCGCTTTAGAACAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTA
GTTGCTGTTCTTCCACGTCAGAAGAGGACAGACAAATTACCACCAGGTGGCGCTCAGAG
TCTGCGGAGGCATCACAACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGA
CCTTTTACTACCTGACTAGCTGTTTGTGTATTTTAGGTGTTTGTTCGCTCCCTAAGCTGAG
CTGAGACGTTTTAGTTATAAATGGCTTCTGTGCTTAGCAATCTGCTCTTTTATTCCCGTGT
GGACTTTCCCTAGCTCTGGCCTTATGCTGCACTAGAAAAGATTTAGCAAGGGGAGAGGAA
GCAGCCTTCCTTACATAACTGGCCTCTTGTGAAAGGGAGCAGCTGCTTGGTGGAAAAAGA
CATTCCCCTCCATGCATCCCTCTCCTTCTGCCTCTGGGGGTTGCAGCTTGAGTCAGAACCA
GGACCATTTAACTCCAACCTTTGAGGAAGAGACGCACCCTGGCCCCAGCCACGCCTGTTA
GAATCTTCCTAGCTGAGTGACACAGTGACACTCAGCCTCAGTTTCTCTGTAAACAAAATGAA
GATAAGAGAGCCGACGAGGATGAAATGGAATAACACACCGTGCGGTGCCTGGTACAGAGT
GAGCCCCAGAAGTGTGACGCGGCCCTCCTTGTGGCTGTCTGGCTTGACCCGAGTGACTC
TGCCTCCCAGTTCTCCGGGATGGGAAGGTGATCCCTGTTTGAGATACCAATTTATAAGAA
CCGAGCCCCGGGAGCTATTTAGAGGTGAGGTGATAAACAGGAGGCCGCTCCTTCATCCC
GGTCATCACGAGAATTCGATATCATAATCAACCATAGGTACCGAGCTCGGGATTACGCCGG
GAGCTTAGGGAGGGGAGGTCACCTTCATAAGGGCCTGGGGGGGAGTTGGAGCCACGAGT
CGTCCAGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCCCCGGGTTGGAT
CCTTCGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGGCGAGGAGCTGTTTAC
CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGC
ACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACCCTGGGCTACGGCGT
GCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT
GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC
CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCC
ACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCC
GCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTG
AGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTTCGTGACCGCCGC
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACATCATAATCAACCTCTGG
ATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTG
GATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCT
CCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCG
CTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGA

CN1528 (2359 bp between ITRs):

CN2710 (1680 bp between ITRs):

CA 03233342 2024-3-27

FIG. 8 cont'd

GACACAGGTCCCTGTTCCACCTGCCTGTTTGCACCTGCTGATGTAATCACCCGGCCGCCTT
GGCTCAGGGCTGGGCCTGCTCACCTGTTTTCTAAGGGAGCTGCAGGTGGCGGGTGAGA
GGGCAGGAGGGCAGGGGATGACACAAGTGTCCCAGGGAGCTCGGGCTGGGCATAAAAGT
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCG
CTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT
CCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCG
AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAAGTCTTCGCCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT
CCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA
GTTTCGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTATATCA
CCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGG
ACGGCGGGGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCGCGGAATTTCGATATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTT
TACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTT
TCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCGCCGCC
TGCTTGGCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCGTGGCTCG
AGAGATCTTCGACTGTGCCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 71)

CN1533 (2212 bp between ITRs):

GCGGCCGCAACGCGTTTAGAACAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTA
GTTGCTGTTCTCCACGTCAGAAGAGGCACAGACAAATTACCACCAGGTGGCGCTCAGAG
TCTGCGGAGGCATCACACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGA
CCTTTTACTACCTGACTAGCTGTTTGTGTATTTTAGGTGTTTGTTTAATGGGGCCTCAGTTT
CTCCCTTGGGATTAAGTGGGTACCAGGCAGGACAAAATAAGCCACCTGTGCCATTCTACCC
TTTCCATCCTGCTGTTTTTCCACCCTGCAGCTGCTTTCTCAAAGAATCCCTGAATCTCCTCT
CTAAATAGTGCAAGCTTGCTGTTGGGAACACCCACAGATGCACTCTACAAATACACAACCTC
ACTTTGAGTAATCATATTGATTCGGAATTATCGGCCATTCAAATACCTCATTGAAGATGCCA
TCTGGGATGTTATTGCTCTTCAGAGGTAGATCATATTCTGAGAAAATCTGTCAAATGTAGA
AATAACTCATCAAAATGGCCATTGCAGAATCTAGCTGCACTTCTTAAGGAAGCAGACTCTGA
TGCTGCGGGTTCATTGCCCTAAGTGAGATGCTGAAGGAAGAAGGGCTGTTTCTTTTCTG
TTTTCAAGTGTCATGAGAAATTTAAAGTTCAGGAATGTGCAAGTTCAGAGTGCTCCTGGACT
CCTTTGCTCATTGAGAGACATCTCACTGAAAACAAACATTCTCAAGGTTATCTCTGGTCCT
GGGAATTCGATATCATAATCAACCATAGGTACCGAGCTCGGGATTTCAGCCGGGAGCTTAG
GGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTGGAGCCACGAGTCGTCCAGC
CGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCCCCGGGTGGATCCTTCGAA
GCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGG
TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGC
GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGG
CAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAAGTGT
TCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA

56/86

FIG. 8 cont'd

AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCT
 ATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACA
 TCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC
 GGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGAC
 CCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCAC
 TCTCGGCATGGACGAGCTGTACAAGTAAGTCGACATCATAATCAACCTCTGGATTACAAAA
 TTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCT
 GCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTGAT
 AAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGG
 ACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCC
 TTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGT
 GCCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGT
 GTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGAC
 AATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 72)

CN1674 (1902 bp between

ITRs): GCGGCCGCAACGCGTTTAGAACAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGT
 GTTAGTTGCTGTTCTTCCACGTGAGAAGAGGCACAGACAAATTACCACCAGGTGGCGCTCA
 GAGTCTGCGGAGGCATCACAACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTG
 AGACCTTTTACTACCTGACTAGCTGTTTGTGATTTTAGGTGTTTGTGTTTGGGCTTTCTC
 TTGGGTTTATGACATGTGCTCCATTGTAATAATCCATTTCTCACAGTTCCATCTCTCATAGC
 AATTCATGTAATTGCTGCAATTTTAACATTTTGGAACTATGTAGTGGGGGAGGGGTGCAA
 GAAGATCTGGCAGTTTCTGCTACAGCTAGCTAATGGCTAATGGGGGTGCCCATAGATCTT
 TGTGGCAAAACAAAGAGTTAGACAGTGCAAAACAAGCTCTGGAAGGAACAGGGTCTGGCA
 GGAGAGAATTTCGATATCATAATCAACCATAGGTACCGAGCTCGGGATTACGCCGGGAGCT
 TAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTGGAGCCACGAGTCGTCC
 AGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCCCCGGGTGGATCCTTC
 GAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGG
 TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTACGCGTGTCC
 GGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATGTGACCAC
 CGGCAAGCTGCCCGTGCCTGGCCACCTCGTGACCACTGAGGCTACGGCTGAGGCTGAGT
 GCTTCGCCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCG
 AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
 CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
 TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAAC
 GTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCAC
 AACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGG
 CGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAA
 AGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGA
 TCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACATCATAATCAACCTCTGGATTAC
 AAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATAC
 GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTG
 TATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCT
 GGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTG
 CTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCTTGACCCTGGAAG
 GTGCCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
 GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
 ACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 73)

CN1778 (2344 bp between ITRs):

57/86

FIG. 8 cont'd

GCGGCCGCAACGCGTTTGAACAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTA
 GTTGCTGTTCTTCCACGTCAGAAGAGGCACAGACAAATTACCACCAGGTGGCGCTCAGAG
 TCTGCGGAGGCATCACAAACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGA
 CCTTTTACTACCTGACTAGCTGTTTGTGATTTTAGGTGTTTGTTCCTGTGGGACTCTGT
 GTGAGTATCCTCTCTGGAGCAATGCCTCTGTGCAATCTCTAGAAAACTCCCTATGTTAGTCT
 CGAACCCACATGGATTGAAGGATTCTCCTGTAGCTAAGATTTTAAAAGACCATGGCAGCAC
 TGGGGAGCCCTGGAAGTCTCCCTCTTATCCTATAACCATGTCTGGGAGCTTCTCTTGGCTCC
 CTGCCAACCCCTAGCCAAGCAGGAGGTCTTGCTTCCCTCTCCTTACTAGCTGTGCGGTGCTTC
 CTTTCACTTCTGTGTTGAATCCCAGTGTCTCTCCTAGACAATCTATTTTAAAGTGTGAATATCT
 ACTTGCTATTTTGGTTCCTCTCTGTGGAGGAGTGACTAGCTGCATCTAGTCAGCCATCTTG
 AATCTTCCCCCGCTTTTGTAAAGTGTTCCTTAATTTAGGTTTCCCTGATGTTTCTTTATGATT
 CTGTTTCAAGTTATGCTTCTGTGGCAGGAATATCACAGGAGCATTGTTGTTCTCTTCTCATTG
 CATTAGATTAGGTGACACATGATTTGGATTTGTCCTATTATCGGTGATGTTAACTTTGATCAT
 TTGATCAAGGTGATATCTGCCAGGTTTCTGCACTGTTAAGTTACTGTTTTCTCTTTTGCAATT
 AATATATGTTTTGTGGGGAAATTTTTCAAGACTGTTTAAAAATATCCCCGTTGTCAAAAATCT
 TCACAGAATTCGATATCATAATCAACCATAGGTACCGAGCTCGGGATTACAGCCGGGAGCTT
 AGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTGGAGCCACGAGTCGTCCA
 GCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCCCCGGGTGGATCCTTCG
 AAGCTAGCGCTACCGGTGCGCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGT
 GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAACCGGCCACAAGTTACAGCGTGTCCG
 GCGAGGGGCGAGGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACC
 GGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCTGGGCTACGGCGTGCAGTG
 CTTGCCCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGA
 AGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
 CGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACT
 TCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACG
 TCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACA
 ACATCGAGGACGGCGGCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGC
 GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAA
 GACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGAGCCGCCCGGGAT
 CACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACATCATAATCAACCTCTGGATTACA
 AAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATAC
 GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCTCCTTG
 TATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCCTGCCTTGCCCGCTGCT
 GGACAGGGGCTCGGCTGTTGGGCACTGACAATTCGCTGGCTCGAGAGATCTTCGACTGTG
 CTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCTTCTTGAACCTGGAAG
 GTGCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
 GTGTCAATCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
 ACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 74)

CN2043 (2152 bp between ITRs):

GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
 CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
 AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTAGGTGG
 AGTGAAGACTGAGAGGGGTGGTAGGGGGAGATGAGGCCGGGAGGAGGAGAGTACACCC
 TCAGGGATGGAGAGCCAACAAAGATTCTAAACGAGGAGTTCTGCAGTGAGACTGGCATT
 TGGAAACAATCGTCTTGTTTTGGTTGAGAGCACAACCTTAGAGAACAGAGGCTAGAGGTGGC
 AGGCTGGAGTGACAGCCTACTGCAGCTGTCCAGCCAATTAAGCCCTGTAGGAAGAATGGC
 TGATGCCGCACTATGTGATGCTGAACTGCGGTAGCCAGGCTGACCTGGCCCCAATCCC
 GGGCTGTGACTGACCTTCAGATGCAGGCAGCAGAGAGAAGAGGACAAAAGTGCACCGCT
 GGGGAAAGGCAAGCAGCTGGCACCCAGGGCACCAAAGACCCGCACAGTGAGCATCAAGG

58/86

FIG. 8 cont'd

CCAGCAAGGGGCACAAAGCCAGCTGTCACCTGTCCTCTGGGCCAAGCCAAGGGGGGCTGG
CCTCCTGGCCACCTCCTTGGGGGAGCAGTACCCCCATTGTCATGACTTGTTACAGATTCC
CAGACCTCCCAGTGCTATCAAGAGCTCGATTACGCCGGGAGCTTAGGGAGGGGAGGTCA
CTTCATAAGGGCTTGGGGGGGGAGTTGGAGCCACGAGTCGTCCAGCCGAGCCCCGTGT
GGCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCCAGATCTTTTGAAGCTAGCGCTACCG
GTCCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGT
CGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTG
CCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCC
CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCACCGCCG
ACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCG
GCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
CTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAG
CGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGA
CGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACC
TCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACCTATGTTGCTCCTTTTACGCT
ATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTT
CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCTGCCTT
GCCCCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCTT
GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ
ID NO: 75)

CN2085 (1680 bp between ITRs):

GCGGCCGCACGCGTTTGCAGAGGGAAGATGAAGTCTTTGATGATAGCCAACCTTAGTGTCAA
TAAGTGGCTTTCTTTGAGACATATTCAGATGGGAACGTCTTGCTTGCCAATTGCCATAGAAA
TCTTAACACACCATGAAGATTGTCCAAGCGCCAAGCCTTCCGTTCTGGACTAAATTACTTTG
AAGTGGCTCAGGACGAGCAGTGGTCAATTTTAACTCTATAGACTGGACAGAAGAGGCTGG
GAGTGGGAGATTGTGCGTTTTAAAGCAGAAAATAAGAAGGGGAACTTGTTTTATACACTC
TATACAAGCTTCTGCTCATTGTGAGATATCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCA
GAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTTGAAGCTAGCGCTACCGG
TCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCC
GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCACCGCCGA
CAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGG
CGTGACGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGCGGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCT
CTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACCTATGTTGCTCCTTTTACGCTA
TGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTT

59/86

FIG. 8 cont'd

CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTT
GCCCCGTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCCTGTCCTTTTCTAATAAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ
ID NO: 76)

CN2102 (1850 bp between ITRs):

GCGGCCGCACGCGTTGGCTCTGGCAAAGGATGGAGATTTTTTTTTTAACTATTCTTTTGA
GAATAAGGAGGGTCTTTGTTTCTCTTCCCTGGGAACCGAGCAGCCCCCTTCTCTGAGGGTA
GAGTTGGGGCATGGCTCAGCAGAACGACAGTCATCCTGGCTAGTAAGACTCAGAGGCTGG
CCTTCAAAGGCTTGAGCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGCCCTTGA
ATTGCTCCCCAGGCGGGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGATTCTTA
AAGAACTGGTATAATTTTGCCTGTCATAGCAGACTCCCAAGACACACAGCCTTTTCCAGG
AGGAGTTCTTAGAGGGGAAGGGGATAGGTTGCAATGGTCTGCTTAGCCTCCCGAAGGCT
GCAGAAGCTGCTAAGACAGGGGTAAGTGAAGATAGACTCTGTGAGGTTGAACAGCAATC
TTTATGACCTGACCCCTAGGCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCAT
CTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCCAC
CATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTCGAGCTGG
ACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC
TACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCC
CACCTCGTGACCACCCTGGGCTACGGCGTGCACTGCTTCGCCCGCTACCCCGACCAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCAT
CTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAG
AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCTACCAAGTCCAAGCTGAGCAAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCTCTGAGTTCTGAGTTCGTGACCGCCGCGGGTCACTCTCGGCATGACGAGCTGT
ACAAGTAAGTCGACGGCGCGCGCGCGCGCGAATTCGATATCATAATCAACCTCTGGATT
ACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGAT
ACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTTCTCCTCCT
TGATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTG
CTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTG
TGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCCTGTCCTTTTCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTA
GGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 77)

CN1415 (1860 bp between ITRs):

GCGGCCGCACGCGTCTGGTAGAGTGGCAGAGAGAGAGAGAGAGAAAAGGAAACGCTAGAAA
AGCTGAGCTGCAAGCCCAGGCTGCCATTTGAAAGAGCAGTTGAGGAAGATCTGAATGAAA
AACTCTGCAAATTTATGCTAATTGTGTTCAAGTTCCATGAGTTTCATCCCATGAGTGGGAGC
CAGGCCACCTGATTGCTTCTCCTTTTGGTAAAGAACATCTCAGATGAACCAGCTGTCATGG
ACATGGTTTCAGAAAATTGGCCCTAAAAAGAAACCTGGGCCATCTGTGTGTGCTGTTCTATT
CATAGGATGCCCTCCAAATCACATATTTCTATTAGGAACCTCACAGCAAGGGCCTCTACCA
TGGGAGCAATAGAGAACATTGGTCATCATCATCTATTGCACTCCCTTGGGGTGGGAGTGCT
GTGATTTCTAGCAAAGAGAACTCAGGGCTGAGGTAGATAACAAAGACAACCTTGGAAATGTCT
AACGTGAGGAAGCATCCTTCTCTTCTTTGCTTTGGAGCTCGGGCTGGGCATAAAAGTCA

60/86

FIG. 8 cont'd

GGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTTGAAGCTAGCGCT
 ACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCC
 TGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTACGCGTGTCCGGCGAGGGCGAG
 GCGGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCC
 CGTGCCCTGGCCCCACCTCGTGACCACCTGGGCTACGGCGTGCAGTGCTTCGCCCGCT
 ACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCC
 AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGT
 TCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
 GGCAACATCCTGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCACC
 GCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGAC
 GGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGT
 GCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGA
 GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGCGGGATCACTCTCGGCAT
 GGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGCGCGGAATTTCGATATCATAATCA
 ACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTAC
 GCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA
 TTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCCTG
 CTTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATCCGTGGCTCGAG
 AGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTT
 CTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTAATAAAATGAGGAAATTGCATC
 GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
 GGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID
 NO: 78)

CN2251 (2079 bp between ITRs):

GCGGCCGCACGCGTTCACAAATCAGTAAGAATTACACTCCAAAAATAAAACAGGGAAGGGT
 ATTAGTATGTGATACGCATGGATAAAGTAGAAGAGGCCAAAAATACAAAGAACCGTAAACT
 CACTAGTGATGGAAGAAACGCTAATTAATAATGAGCTTTCCTGACCCTTTATCATATGGG
 CACATTTTCAAAGGACAAGACTCTGCTGAGGAGAACACAGAACAGGATCCTTCAGACAAGC
 TTCAGCAACCAATTTTAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTAAAAATAACA
 TAGATCTGAAAAGCTTTTATAACTTTTTGTTTTTTTTAAGGACCATATTAACAGCATCTTCCA
 GTTTTTCTAGAGTAAATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTG
 TAAAAAGACTATAAAAAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGA
 CTAAGGCCCTTCAGCTCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTCACAC
 TGCCCTGCCTTTGTGTGCCCTTGCCCTGTCCCTCTGTGACGGTCACCCTTTTCGGTTCTGGCT
 TGGAGCCTGGTCATCAGTGCCCTGGCGGCCTGTGCCCTGGGCTGTGCTGAGCTCATTTGGTC
 GTTGTCTAGTGATGTGACCCTACCTTAGACTTCCCCATCTCTTGAGGTGGCTAGAGCTCG
 GGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGA
 TCTTTTGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTTAC
 CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
 TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGC
 ACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGT
 GCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT
 GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC
 CCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
 TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAAGCC
 ACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCC
 GCCACAACATCGAGGACGGCGGCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCC
 ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTG
 AGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC
 CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGC

61/86

FIG. 8 cont'd

GAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTA
 ACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATT
 GCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATATAAATCCTGGTTAGTTCTTGCCACGGC
 GGAACATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGCACTG
 ACAATTCGCTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
 GCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCTTAATA
 AAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG
 GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
 GGACCGAGCGGCCGC (SEQ ID NO: 79)

CN2257 (1969 bp between ITRs):

GCGGCCGCACGCGTGGAGGGGCGATAGTGTCCACACTTCAGTCTTCATTCTTCTTGAGTTTC
 ATGTGTTTAGCAAATGAAAACTTTTATATTTCTGTCTTTGATAGGCTTACCAAATCCAAAGTG
 CTCCCCCTCCCCCACTATTTTCAAGTTGAATATCTGTTCTCTGAAATGTTTGGTCTGAAAAG
 CATATCCAAATTTGGAAGATCTGCATATATGTAATGAGACCTCTTGGGATGCATCTCAAGTC
 CAAACACAAAAGTCATTTATGTTTCATGCACACTTGACACACCTAGCCTGAAGCTAATTTGA
 TACCATGTCTTTAGTGACCTGTGTTCTGACTGCAAACCACTATAAGGAGTCAGGTGTGGA
 ATTTTCCATTTACAGCATCGAGTCACTGTTCCAAAATGTTCAAGCTTTTAGAGTGCTCTGTGA
 TTTGTATTTTTGGATCAGGAATGCTCAACCTGTATTTTAATGCTTTGGCCTGAAACCTTCCC
 CTCAGAAGGAAATGTTTGATTGCAGACTGGCATTTTTGCTCTCTGGGATAGTGAGTTGTGT
 TTTGACTTTTTTTCATTTATTTATTTATTTATATTCATGCCTGATTGAGGGAGGCTTTTAGACC
 ATGGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTT
 CTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCACCATGGTGAGCAAGGGCGA
 GGAGCTGTTACACGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCC
 ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG
 AAGCTGATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACCCT
 GGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTT
 CAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
 CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
 AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGAGTAC
 AACTACAAGACGCCACAACGTCTATATCACCGCGCAGAAGCAAGCAAGCATCAAGGCC
 AACTTCAAGATCCGCCACAACATCGAGGACGGCGCGTGCAGCTCGCCGACCACTACCAG
 CAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTA
 CCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTCTGCTGGAGTT
 CGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCG
 CGCCGCGGGCGCGAATTTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATT
 GACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
 TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATATAAATCCTGGTTAG
 TTCTTGCCACGGCGGAACATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG
 CTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAG
 CCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTG
 TCCTTTCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTG
 GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
 AGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 80)

CN2258 (2166 bp between ITRs):

GCGGCCGCACGCGTNCCTGCGGCCGCACGCGTTCCCTAACCTAGAAAAACCCACCATCAA
 CTATTGAGGTAATGTTCCNTAGATGCAGGGTCATCAGTGGGCATTGTTACTTACTGGAAGA
 ATGCTCTCTGGCGGATGGGGGAGGGGAAGCAGGAAAACCTCTGCGTTAAGCAGGGGCATAATC
 TCTTCAGGGACCTGCTTCAGTTTCAATCCGACTCAAGTCGTGCTTTGCAGCTGGCCCCAGT
 TTTGGCTGCTCTGAGAGTCAGTTGTTGGTTATGCACTGCCTGGGGTGAGGGCCAACCAGG

FIG. 8 cont'd

GAGGGGGTTGGAGAAGGATGCGAATAGTCAGAGCAGCTGGGAGCAAATGTATCCAGTTAA
GGGGATCAAAACAGAACGTCAGCAACCCCCGTAATGCCCAAATCTTGCATAACTCTGTGT
TATCTAGGCCATCTCCCCCAACCCACCCCGCTTCTCATTTGGAACAAGGTCTTTCTATGTA
TGTAAGCTAAAGTTGACCTTGAACCTCATGATCTTCCCTGCCAGCCTCTGAGTGCTAGGGTA
AAATAGGCGTGACCATCCTGCCGGGCTTCACCCACTCCCTTTTCATTTCTGAGACACTTCC
AGTTATTCTTCTCATTCTTACTCCATTCACTCATACTGACCACTTCAAAATTCTGGAACCTT
CCAGCATGCACTGGCTGCAGATGAATTGTGCTTTTTCTGCTCCCTCATCTGACTCTCCTTCT
CGTCTGACTACTTAGCTCTCATCTTCTCTACGGAGTCACCAGTCTTTTTTTTTTTTTTTTCA
AGACAGAGTTTCTCTGAATAGCCCTGGCGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGA
GCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTC
GCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCC
CTGGCCACCCCTCGTGACCACCCTGGGCTACGGCGTGCAAGTCTTCCCGCTACCCCG
ACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCACCGCCGAC
AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCGACAACCACTACCTGAGCTACCAAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAAGTCGACGGCGCGCCGCGCGGAATTCGATATCATAATCAACCTCT
GGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG
TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTC
CTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCTGCCTTGCC
CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTC
GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCTTCTTGAAC
CTGGAAGGTGCCACTCCCACTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
TGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
81)

CN2267 (2012 bp between ITRs):

GCGGCCGCACGCGTAGCAAAGTTGAATGACATAGTTTTGATGAGATATTAGGTAGATTTCA
GCTCATTTAGAGAAAAGAATATATTGTTCTGCCAAAGAATTCTGTGTCCGCTTTTCTGGCCT
GGGTTCAATTGTGTGTACGCTACAGAACTGGTATTCAAGAAAGGAGCCTGGATTTCTAATC
ATGTTGATATTTCCCATGTAGCTTTTTTCTCAGCTCCTTGGACTCTTCACACAAAACATAA
AAGGTCATTAATTATAGGGTCCTTATGTAAATACTGGCCTGTTTTGTGATTCTATTCTGAGC
AACGGTTATATACATCTGCTCAAATAAATTAATTTTCTTGTCCCAAATGACTCATTCACTGGA
ACTAGAAGTTCCCATGCCAGATCAAATAGAGAAAGGGAGAGAGGACACAGCACAATGCTG
CGGATGGGAACCTTATTCTGCCCTATACCAATTTATGAGTACTGGCACAATAATTTAACGGGA
AAAGCTCTAAGGGAATATGTTTGTGTGTTTGTGTATGTGTGTAGGAGAGAGAAGGGGAGAG
AGAGAATTATGAATCTTGTCTTCTCATGTGTAATACAGCCATGTTTGAGATATGAATGTAATT
CAGACATACTTGCCTTTGAGTATTTTCTCTATATTTAATGAGCTGTTCTGGAGCTCGGGCTG
GGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTT
CGAAGCTAGCGCTACCGGTGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGG
GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC
CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCA
CCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCTGGGCTACGGCGTGACG
TGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCC

63/86

FIG. 8 cont'd

GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC
GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAA
CGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCA
CAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAAACACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGC
AAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCTGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCCGCCGCGAA
TTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACT
ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTTCAATTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAA
CTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAA
TTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCC
CTCCCCCGTGCCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAAT
GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGA
CCGAGCGGCCGC (SEQ ID NO: 82)

CN2316 (1963 bp between ITRs):

GCGGCCGCACGCGTTGTAGAAGGCCATGCTCACACTTTGGGCATTTTCCCTCAGCTCCTC
CATAGCCTTAAGTTTCAGATCCCAGCTTCACCTGCTGTTACCTATTTTATCTGCTTTCTTCT
TAAGCTGGAAAGCTCTCTGTACACAGCAGGCACAGAATTGCTGATGACTTCATGGCAGAGG
CAAAAATGGTAATTTTCTTTCTTTTGTTCAAAGCACTGTGCATGAGAATATAGTTCTAAGTT
ACTGACTCAGAGATGATAGAAGATTTTGAAGAAAATGATGACAAGCAGTGGTTTTAAAGGT
GAATTAGATCATAGACAGGTGTTTGGAGGAAAAAGCAGGGGAAGGGGAATTTCTAAATAGA
GGTCGGTCACCAAGACATATTGAAGCAGTACAAATGACATTAGAGAGAAACAAGGTGCAAAAC
GAGTAAAAGGCCAGTCCCTGGGGGAAGTGAAGTCTTCTAAGCAGATGATTATGAGAACAA
GGAACAAATGTCCCTGAAACCAATTAGGAAATGGGACTTGAGATGAGAAAACCTGTTCTGG
GATACACAGTGATCTATTTTAGAGTCTGAATAATGTTGTCTCTATCTAGAGTTGAAAATGATA
GGTCCCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTG
CTTCTGGGATCCAGATCTTTTGAAGCTAGCGCTACCGGTCCGCCACCATGGTGAGCAAGGG
CGAGGAGCTGTTTACCGGGGTGGTGGCCCATCTGCTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTTACGCTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC
CTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCAC
CCTGGGCTACGGCGTGCACTGCTTCGCCCGCTACCCCGACCATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGA
CGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA
TCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG
TACAACATAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAG
GCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTA
CCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGA
GCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGG
AGTTCTGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACG
GCGCGCCGCGGCGCGGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAG
ATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGC
CTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCTCCTTGATAAATCCTGGT
TAGTTCTTGCCACGGCGGAACATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTC
GGCTGTTGGGCACTGACAATTCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGC
CAGCCATCTGTTGTTTGCCCTCCCCCGTGCCCTTCTTGACCCTGGAAGGTGCCACTCCC
ACTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTAT

64/86

FIG. 8 cont'd

TCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG
CATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 83)

CN2431 (2074 bp between ITRs):

GCGGCCGCACGCGTGTGGAGGTTGGGTGTTTCCATCACCTGGCTAACAGAGATCACAATC
TCCTTTTGTTCATCTTAGGGATAAGGTTTGTCCATAGCTCGACTTGTCTGGATTTCTCTGTC
CATCACATGCCTATGTACGCTCCCCACACCTCTCTCCTAGTGGCTCCTTGGTTTAAAATGC
CCCAGATCTGCAGCTCACAGAAGAACATTTGGCTACTCCCGTTTTTAATCTACTACTTCTGT
AATGGGGCTCGGGAGCTGGGCTGGGTCCCCGAACCTGTGGGCAGCTGGCAGCCTGACC
CTGGGCTGAGTACCAGGCATGGGGAGGCCAGAAGGAAGGGGCAGATTGAGCCAGAGAG
GTAAGACTGCAGAGAGCCAAATAGGAGGGGAGGAGGCCAGAGCCTGAGGGTCCAGATGG
GGCTGGCTGGGCACAGGAGGGCATCTGATGCAATCCGCCAGGAGAGATGCCAGCCGGCC
GGGTGCTGCAGAGGCTCTGTTCTTGAAGAGTCTAGTGCCTGCTCTATTAATTCTGGCTCTG
TGGGAATTGGGGAAGATTTCCCTATGAGGTCTCCTGAGCCCCAGCAACCTCCCCAAAGGC
CTTAGCTCATAAAAGCACCTGTGTCTGAACAAGCAGGGGGTGGGGACATAGAACACCTGG
GTTGGTGGCTGCCTGGGGAAGGGCAAGCTGAAGTATCTTGTCAGTATAAACACAGTACCT
CAGCCCTGCGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATT
TGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCCACCATGGTGAGCAAG
GGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA
CCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACC
ACCCTGGGCTACGGCGTGCAAGTTCGCCCGCTACCCCGACCATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA
CGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACC
GCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG
GAGTACAACACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATC
AAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCA
CTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCT
GAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCT
GGAGTTCTGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTGA
CGGCGCGCCGCGGCGCAATTTCGATATCATATAACCTCTGGATTACAAAATTTGTGAA
AGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAAT
GCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTG
GTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGG
CTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTT
GCCAGCCATCTGTTGTTTGCCCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTC
CCTACTGTCCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTATTCT
ATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 84)

CN2436 (2182 bp between ITRs):

GCGGCCGCACGCGTGGGAAGAATAGTGAGTTCTGGGGTGGAGGGTTTTGATTTTTTAAAG
TGGAAGTCAAGAAATGCACCACGGAGAAGGTGGCATTGAGTGAAAACCTCAAAGTGGTTT
GAGAGAAAGTGGTGCAGATATGCGGACAGCGCTTCAGGCAGAGAAAATAACCTGGGATGT
CTGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGAGCAAGAGAGTTAAGAATAGGTG
TAGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCATTTTTAGGCCAATATAAGGACTG
GCATTTTTAGGTCATGATGAGAACTGGCATTACTCTGCCTAAAAATAGGAACCACTGGAG
CATTTTGAAGCAGAGGAGTTGAACTGTCTGATTTAAATTTTAAAGGTTCTGGCTACCTTGT
GAGAAGAGACTGAAGTAAGATATGCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAAT
AATCTAGGCAAGAGAGATGGATCAGGGTGATAACAGTAACTTCATGAGAGCATTTTCAGTT
GCAATATATTCCAGATTTGCTGACCAGATTTGATAAGGAATATAGAAAAAGAGGAAGAAG

65/86

FIG. 8 cont'd

AAGAGATAACAGAGTGTAGGCAGAAAAGAGAAAAGGTTCAAAGACTGAGCCCTGCAGTATT
TCAACATTAAGAGCTTGAATTGCAGAATAGGAACCATCTCAAGGCTTTAAATACAAGTGCTG
AACACTCTCAAATTTATATTTGTATCCCAGACTTCTCTAAAATTCAAGTTCTAAAATCCAGTT
GGTTACTTTTATGTCTCCACTTGGAACAAGTCTTGAAGCAGTGAGCTCGGGCTGGGCATAA
AAGTCAGGGCGAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTTGAAGCT
AGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGGCGAGGAGCTGTTACCCGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAG
GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAA
GCTGCCCGTGCCCTGGCCACCTCGTGACCACCTGGGCTACGGCGTGAGTGCTTCG
CCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCT
ACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG
GAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTAT
ATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATC
GAGGACGGCGGCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGG
CCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCT
CGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATC
ATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTC
CTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG
GCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCATCGC
CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCTGTTG
CTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCG
TGCTTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCCTAATAAAATGAGGAAAT
TGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAG
CAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGG
CCGC (SEQ ID NO: 85)

CN2643 (2255 bp between ITRs):

CGGGCCGCACGCGTGAATGTATTGGATTATAACATGCACTAGTATTAAGTATAAAAGTATGT
TCAAGAAAATTTAAAGTTATTTTAAAGACTGATGTAAGGAAAGAAGGAGGAAAGTTAGTCCAT
TAAATCACACTGAGCAAGAGCTAACTTCTACAGAGTAGTCATTAAGAGGTAATTTTACAATC
TAAATTTTTTTTTTTTAAAGGATGCTTTCCTTTGTTGTGGTGTCAAGGTGAGAACTATGTAG
ATTTGATTCCGTAAATATCTATGGTAAGGCAAGTGACAGGGTGAAGTCAAGTCAAGTCAAGT
GAAATAGTTAGAACATAGAAACAGTCAATAACTATTCTTTTGTAAAAAATAATAAT
AAAATAGCTCCTTCAAGTCCTGAAGCGTAATAAAATATTAGAAAACACAATGGAAAAATGAG
GCACTCACAGAAGAAATACAGTTGGCTCTAAAGCATAGAAAGTCTCAATTATTTCAAATAA
TAGAAAGGTACTTTAAGACAGTGAACCCGGAGTGGTCTCTGGGAACAGGCAGCTTGGTCC
TCAATCTGGCAGTCAGTGCTAACTACACTGTGATCCTGGAAAAGATATGCTTTACTTCCTCA
TTCAAAAACCTCAGAATTATAGTGCTGCCTGTGCCCTGATCTACCTGATGCCCCGCCCTACC
TGCTTTTGGGGGTTGCAGGTAAGGTGGTGTGGAATCAGGTGAGAAATGTACAGCAAGCA
CCTCTGAAAGCTGCTGCAAGCTCCTTTAATACGCTCACCCGTGTCCCATTTGGTCCCAAGAA
AGCATCTCTTCTAAGCAGCAGTTCCTGATTGGCTGGCATTGTCTGTGTGTCAGCAATCCTTGG
TCTCTTAGGCAGTCCTCAATTAGGTCTCCTGTAGAAGGTGACTTTGCCGAGCTCGGGCTG
GGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTT
CGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGG
GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTG
CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCA
CCGGCAAGCTGCCCGTGCCCTGGCCACCTCGTGACCACCTGGGCTACGGCGTGAG
TGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC

66/86

FIG. 8 cont'd

GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAA
CGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCA
CAACATCGAGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCG
GCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGC
AAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCTGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCCGCCGCGAA
TTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTA
ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTTCAATTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAA
CTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAA
TTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCC
CTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAAT
GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGA
CCGAGCGGCCGC (SEQ ID NO: 86)

CN2663 (1909 bp between ITRs):

GCGGCCGCACGCGTTGCTCAGCCTCTATTTCCCGTTGAACTGGTTTTGGTAAGATATTTTA
CAGGCTAACTGTGTATTTTCATCTAGCAGCCAGCAAGCGTTTGCACATCGTTATAAATTATGT
CTTGTTATAAACTTGTCTCTGTGCAGCTGCACATCTGCACCGTAGTGTAGCTCCCTGTCCCCGGG
TGACCGATTGGAAGTAGGTTTCTCCATCCCCAATAGGCCTCACTGTGTCTCATCACACCAA
AGTATTTTTAAAAAGGCTTAATGAAGACAGCATCCTCTGTATCATCCCCCGGGAAGGCTGA
GCAGGTTGCGCACCATAAATACACTTCAACGTTCTATCTGCCTAAGCCTTGAATCCCCTC
CTCTATGATATAGCAAGGTAAATGCCTGCATCCTGACTGCATTTAATGGACCACGGGCCCT
GGGTTGGGGTTGCAGCACCCAGTGAGCTATGAGAGCAGCCCCAGCTGCTCCGTCTGAC
AGATTGAAGCCAGGCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGC
TTACATTTGCTTCTGGGATCCAGATCTTTTGAAGCTAGCGCTACCGGTGCGCACCATGGTG
AGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGGA
CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCA
AGCTGACCCTGAAGCTGATCTGCACCAACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTC
GTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCATGAAGCA
GCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTT
CAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG
TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAAC
GGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGC
CGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC
ACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGT
AAGTCGACGGCGCGCCGCGGCGGAATTCGATATCATAATCAACCTCTGGATTACAAAAT
TTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTG
CTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATA
AATCCTGGTTAGTTCTTGCCACGGCGGAACATCGCCGCCTGCCTTGCCCGCTGCTGGA
CAGGGGCTCGGCTGTTGGGCACTGACAATTCGTGGCTCGAGAGATCTTCGACTGTGCCT
TCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCTTCTTGACCCTGGAAGGTG
CCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGT
CATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAA
TAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 87)

67/86

FIG. 8 cont'd

CN2717 (1740 bp between ITRs):

GCGGCCGCACGCGTAGTGACTTGGTGCTATGAGCCATATTTTGTCTGTTGCTGTTGTTACTG
GTAGTTTTTGTAAATCTGGGGCTAAAACCTTGGGGTCTGGTATGCTGTCAATTTACCAGTGAG
CTATACCCCTGGATATTATGATTTAGATGAATGTGAAATATCACCCCAGACATACATATACTAA
ACACTTGGCCCTTGGCCCATGATGCTAAATGGAGGAGATAGAAGCTTTTGGGGCACAGCC
TAGTGGAAGGAAATGAGGTCAAATGACATGTACTCTGAAAGGAATATGGGTATTCTGGGCT
TGCGTTATTCTCTCTCTCCCTCTCTCTCCCTCTCTCTCCCTCTCTCTTTTCTCCTTTT
CTCTTTCTCTCCTCGCCTTGTTTTCCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGC
CATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGC
CACCATGGTGAGCAAGGGCGAGGAGCTGTTCAACGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCACCGCCGACAAGC
AGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCTACCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGTCGACGGCGCGCCGCGGCGCGAATTCGATATCATAATCAACCTCTGGA
TTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGG
ATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTC
CTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCTGCCTTGCCCGC
TGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGAC
TGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCTTCTTGACCCTG
GAAGGTGCCACTCCCACTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
GAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 88)

CN2838 (2013 bp between ITRs):

GCGGCCGCACGCGTTGGCATTCTTGGATGTGTTAGGCCTCTTGCTGCTGCTCAATTTCTTT
CCGTTTGACACATCTCCAAACCATCATGTAGAATCTCATGGCTGAATTCATGCCCCGGCTC
CATCTAATATTGTGTCTTGACACCATCAAGGGCACCATCTAGGCTTTTCAACCCATTTATAA
ATGATACGCCACATCCCAACAAGGGCTTATGACTTTTGTAGGCATCAATAGTTCTGATTCCC
TACGATAATGGTTCAATTTTCTTAGTAATAGAGAGCAGATCTTTTTCTTGGGCTAATTTTAA
GCTTGTGATTGTAGAATATATGAAGACGAGGCTCTTAATCATCAAGGCATTTTTTCCCCTA
AACACTTAGGCTGATGCAATTTTCTGAAATTTAGAAAACAAATTTTGCAATCACTTTTCTTCT
TGATTGAAAAGAAGGTAAAAGGACAGATCAAAAAGGCATAATTGCATGGTACATCTAGTCA
ACAACGTACAGGCAGAGTGTCAAAAGTGTGCATCTTAACTGCCTCAAGACTTATTATGAT
GTCAATTTGCCTCCAGAGGATCCATATGTATTAGGAGGGTGGCAGGATGCCCTAGTTCACA
GCTTAACCTCTGCTTCTAGCCTAGATTCCAACCTCTGTTTCATCTACCCGAGAGCTCGGGC
TGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCT
TTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCG
GGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTG
TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCAC
CACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGC
AGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGC
CCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC

68/86

FIG. 8 cont'd

GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCAC
AACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGC
CACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCAT
CGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGA
GCAAAGACCCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTGTACCGCCGCC
GGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCCGCCGCG
AATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAA
CTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG
CTTCCCGTATGGCTTTTCAATTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCG
GAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGA
CAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTG
CCCCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAA
AATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGGTGGGGTG
GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 89)

CN1521 (2640 bp between ITRs):

GCGGCCGCAACGCGTTTAGAACAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTA
GTTGCTGTTCTTCCACGTCAGAAGAGGACAGACAAATTACCACCAGGTGGCGCTCAGAG
TCTGCGGAGGCATCACAACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGA
CCTTTTACTACCTGACTAGCTGTTTGTGTATTTTAGGTGTTTGTGTTTGGTCTGGTTCTGACT
CAAGCTGCAACCCCCAGAGGCAGAAGGAGAGGGATGCATGGAGGGGAATGTCTTTTCCA
CCAAGCAGCTGCTCCCTTTCACAAGAGGCCAGTTATGTAAGGAAGGCTGCTTCCTCTCCCC
TTGCTAAATCTTTTCTAGTGCAGCATAAGGCCAGAGCTAGGGAAAGTCCACACGGGAATAA
AAAGAGCAGATTGCTAAGCACAGAAGCCATTTATAACTAAAACGTCTCAGCTCAGCTTAGG
GACGCGTCTAGGTACGACTCCACCCCTCCAGGAACCCCTAGTGAGGGAGTGGTGGGA
GCAGAGCCCTCATCACACAGACTTGGTCCTGGTTCTGACTCAAGCTGCAACCCCCAGAGG
CAGAAGGAGAGGGATGCATGGAGGGGAATGTCTTTTCCACCAAGCAGCTGCTCCCTTTC
ACAAGAGGCCAGTTATGTAAGGAAGGCTGCTTCCTCTCCCTTGTCTAAATCTTTTCTAGTG
CAGCATAAGGCCAGAGCTAGGGAAAGTCCACACGGGAATAAAAAGAGCAGATTGCTAAGC
ACAGAAGCCATTTATAACTAAAACGTCTCAGCTCAGCTTAGGGACGCTCCTAGGTACGCA
CTCCACCCCTCCAGGAACCCCTAGTGAGGGAGTGGTGGTGGGTGAGTAAGAGTAGGTGG
CAACTGGTCCTGTTTCTGACTCAAGCTGCAACCCCCAGAGGCAGAAGGAGAGGGATGCAT
GGAGGGGAATGTCTTTTCCACCAAGCAGCTGCTCCCTTTCACAAGAGGCCAGTTATGTAA
GGAAGGCTGCTTCCTCTCCCTTGTCTAAATCTTTTCTAGTGACAGCATAAGGCCAGAGCTAG
GGAAGTCCACACGGGAATAAAAAGAGCAGATTGCTAAGCACAGAAGCCATTTATAACTAA
AACGTCTCAGCTCAGCTTAGGGACGCGTCTAGGTCACGACTCCACCCCTCCAGGAACCC
CTAGTGAGGGAGTGGTGGGAATTCGATATCATAATCAACCATAGGTACCGAGCTCGGGATT
CAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTGGAG
CCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCCCCG
GGTTGGATCCTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAG
CTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAA
GTTACGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGC
TGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGGC
TACGGCGTGCACTGCTTCGCCCGCTACCCCGACCATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAAC
TACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCT
GAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATA
CAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTT
CAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGA
ACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCACT

FIG. 8 cont'd

CCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA
CCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACATCATAATCA
ACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTAC
GCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA
TTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCGCCGCCTG
CCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAG
AGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCGTGCCTT
CCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTAATAAAATGAGGAAATTGCATC
GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID
NO: 90)

CN2216 (1818 bp between ITRs):

GCGGCCGCACGCGTATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGC
TTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTGAGGCTGCCCCCTCTCATCTGTGA
CCTCCTGTGGTCGGGAGGGACTTGCTCTGGATTATATAAAGGCTATGAGAGGCTAGAGTT
GGAGAGCACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAG
CCTGGGTGAGGCTGCCCCCTCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGG
ATTATATAAAGGCTATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGCT
TGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTGAGGCTGCCCCCTCTCATCTGTGAC
CTCCTGTGGTCGGGAGGGACTTGCTCTGGATTATATAAAGGCTGAGCTCAGAGGTAGGCG
TGTACGGTGGGAGGCCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGCCTGGGG
ATCCAGATCTTTTGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGC
TGTTACCGGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAG
TTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCT
GATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGGGCT
ACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGT
CCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACT
ACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTAC
AAGCCACACAGCTATATCACCGCCGACAGCAAGCAAGCAAGGCAAGGCAAGGCAAGTTC
AAGATCCGCCACAACATCGAGGACGGCGCGCTGCAGCTCGCCGACCCTACCGAGCAGAA
CACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGT
CCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA
CCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCG
CGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGG
TATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCA
TGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGC
CACGGCGGAACATCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGG
GCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTG
TTGTTTGCCCTCCCCCGTGCCTTCTTACCCTGGAAGGTGCCACTCCCACTGTCTTTT
CTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTG
CGGACCGAGCGGCCGC (SEQ ID NO: 91)

CN2839 (1916 bp between ITRs):

GCGGCCGCACGCGTTGGAAGGAGATGCATTTGTATTGAAATTTTCTGTTTCATTTAAAGTTT
GTTTTTTTCTTTAAATACTGGGATATATTTTCTTCTGTACAGCCTTGCCCTCCTTCCAATTTAT
GTTTCAAGGATAAAGATTACATTTAGGCCATTTATAAATGGTTAGAGTCAATCAACACGGTAA
CATTCGTATAGCAATTATTGCAAATTAACCACTTGATGTTAATCATTTCTGTTTGTGCTTTT

70/86

FIG. 8 cont'd

TTTTTTTCCCCTCTAATGTTGATTTCTGAGGCCTTTAAAGATGGTAACTTTATTATCCTTATAA
GATGCCAACATTTGAGTCATCAACATTGGTTAAAAATAAAGGGGGGATCACCTTGAGCTGCA
GACATGCAAATCAACTGGAAAAACACCAGCTGGCCTTAATCTTTCACCATTGCAGTTTTGTC
TGCCTCCTGTGCAGCAGGCTAGCAGGTAGGCTTTTTATAGAAACAGGGACTTAGCAGATGTA
GCTGCTGTTTCCTTCTTGTCAACACTGAAATTCTCCTTTTTAGCCCTCTGACACACACAGA
CACCAACCGAAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACA
TTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAA
GGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA
ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCAACCTCGTGAC
CACCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACG
ACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGG
ACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCT
GGAGTACAACACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCAT
CAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGGCTGCAGCTCGCCGACC
ACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTAC
CTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
CTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTC
GACGGCGCGCCGCGGGCCGCAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTG
AAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTA
ATGCCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATAAATCC
TGGTTAGTTCTTGCCACGGCGGAACATCGCCGCTGCCTTGCCCGCTGCTGGACAGGG
GCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACT
CCCAGTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATT
TATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCA
GGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 129)

CN2674 (1810 bp between ITRs):

GCGGCCGCACGCGTACAGCCTTTTCTCCTCTGACATTCTGACCTCCCCGTGGCCTGCC
TTCCCCTGTCTCCTTGCCGAGCCATCCTCTGTACCTTCACTGTTTCCACTAGGTTATTGA
GGCCTTTTGGTCATGAGCTCCTGCCACCAAATCAGTGCCTGAGGCTGCTGCATAACAATCT
CCTTCTTGATTAGAGCATCTTAATTATGCTCCAGGAATTTGCTTAAACAGATTAAATCATTCT
GGAGGACAGATTTAGAAGCTGTGCTAGCTGAGGGATCTGCTCTGAGTGTCCCCGGGGAGA
TTCTGGTGACAACCGCCTGTTGGGCCTTTGCCAAGGCCTTGGGTGCCTCGGAGACTCGGA
CACTGGCTGGAAGCTCTACGCACCCCCGAAGGCTTCTCTTTGCTCCTCTGGGCCTGTAGA
GCAGGCTCTGAAACTCTTGATGTGAAACGTGGCAGGAGTGAGCTCGGGCTGGGCATAAA
AGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTA
GCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGGTGGTGCC
CATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGG
GCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAG
CTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGC
CCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTA
CGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT
GAAGTTCGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGG
AGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATA
TCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCG
AGGACGGCGGGCTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC
CCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCC

71/86

FIG. 8 cont'd

AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCGCGCGAATTCGATATCA
TAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCC
TTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGG
CTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCGCC
GCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGC
TCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCGT
GCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATT
GCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGC
AAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGC
CGC (SEQ ID NO: 130)

CN3301 (2335 bp between ITRs):

GCGGCCGCACGCGTATCTTCAGGGTCCACAGGCACATAAAGGTGGAAGGGGAGAACATAT
GCAGTCCATCTGTCTGTCTGTCTTTGCACATAGGCCTCTTTACCAAGCCACCCTCGGCTCG
ATCTCCCTGCACGCAACCTACCCCACTTAGTAGATCTTCAACCCTGATGGCATCATCACGG
TTTCTGTCCCCTGACCTCTGCTCCGAGCTTACCCCACTTGTCTCCATTGCAGTGCCTACAG
CGCTCAGCCCGTATGTGCCTTGACAGCCTCTGTGTGCCGTGCGCAGCATCAGTGTGCCG
TGCGCAGCCTCTGTCCACGTTTACAGTGGGTGTTACCAGCATTCTCCTGCTTTATCATCC
TCTGTGTTGATGCGGAGGTCAACACATCCTTGGCTGCTATGAGAACAGAAGGACACAGTCA
GTGGACTTTGTGGGACTTGGAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCT
ATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCA
TGGTCTCCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGAC
GGCGATGTCAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTA
CGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCA
CCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATG
AAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
TTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GCACAAGCTGGAGTACAACACAGCCACAACGTCTATATCACCGCCGACAAGCAGAA
GAACGGCATCAAGGCCAATTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGC
TCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC
AACCACCTACCTGAGCTACCAGTCCAAGCTGAGCAAGAACCCCAACGAGAAGCGCGATCAC
ATGGTCTGCTGAGTTCTGAGCCGCGCCGGGATCACTCTCGGCATGGACGAGCTGTAC
AAGGGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGGGGATGTAGAGGAGAAT
CCAGGGCCTGGTGCTAGTGAGACTACAAGACCATGACGGAGATTATAAAGATCATGAC
ATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGATCTGGAGGCTCCGGAGGCCCA
GAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAAGGC
GCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCTATGT
GTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCTGTTCAAGGCCATGGGCAT
CATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTCCCGCCTGGC
GCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCGTGCGCCTGCT
GCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACCAAGT
ACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCGCGAATTCGATATCATAATCAACCTC
TGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTAT
GTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTC
TCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCGCCGCTGCCTTG
CCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATC
TTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTTG
ACCCTGGAAGGTGCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATT

72/86

FIG. 8 cont'd

GTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAG
GATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID
NO: 131)

CN3019 (2102 bp between ITRs):

GCGGCCGCACGCGTGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTA
AGGCCTTCAGCTCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTACACTGC
CCTGCCCTTTGTGTGCCCTTGCCTGTCCCTCTGTGACGGTCACCTTTTCGGTTCTGGCTTGG
AGCCTGGTCATCAGTGCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTCGTT
GTCCTAGTGATGTGACCCTACGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATC
AGACTAAGGCCTTCAGCTCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTCA
CACTGCCCTGCCTTTGTGTGCCCTTGCCTGTCCCTCTGTGACGGTCACCTTTTCGGTTCTG
GCTTGGAGCCTGGTCATCAGTGCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTG
GTCGTTGTCCTAGTGATGTGACCCTACGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACA
CGCATCAGACTAAGGCCTTCAGCTCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAA
GTGTCACACTGCCCTGCCTTTGTGTGCCCTTGCCTGTCCCTCTGTGACGGTCACCTTTTCG
GTTCTGGCTTGGAGCCTGGTCATCAGTGCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCT
CATTTGGTCGTTGTCCTAGTGATGTGACCCTACGAGCTCGGGCTGGGCATAAAAGTCAGG
GCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTAC
CGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTG
GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGGCGAGG
GCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCC
GTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGCAAGTGTCTCGCCCGCTA
CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA
GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTT
CGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCACCG
CCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACG
GCGGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG
CTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGCGGAATTCGATATCATAATCAA
CCTCTGGATTACAAATTTGTGAAAGATTGACTGGTATTCTTAAGTATGTTGCTCCTTTTACG
CTATGTGGATACGCTGCTTTAATGCTTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAT
TTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTATCGCCGCTGC
CTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGA
GATCTTCGACTGTGCCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCCTC
CTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTAATAAAATGAGGAAATTGCATCG
CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGG
GGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID
NO: 132)

CN3569 (1910 bp between ITRs):

GCGGCCGCACGCGTGCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGCCCTTGA
ATTGCTCCCCAGGCGGGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGATTCTTA
AAGAACTGGTATAATTTTGCAGTGCATAGCAGACTCCCAAGACACACAGCCTTTTCCAGG
AGGAGTTCTTAGAGGGGGCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGCCCT
TGAATTGCTCCCCAGGCGGGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGATT
TTAAAGAACTGGTATAATTTTGCAGTGCATAGCAGACTCCCAAGACACACAGCCTTTTCCA
GGAGGAGTTCTTAGAGGGGGCTCCTGGGTGACTGAGCAGTGGAGAAACAGAATCCTGC

73/86

FIG. 8 cont'd

CCTTGAATTGCTCCCCAGGCGGGCTTTATGCAGTCTGGGGAAGCAAGGGATGCCCTGTGA
TTCTTAAAGAACTGGTATAATTTTGCAGTGCATAGCAGACTCCCAAGACACACAGCCTTTT
CCAGGAGGAGTTCCCTTAGAGGGGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCA
TCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCA
CCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTG
GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCCGTACCCCGACCAC
ATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACACAACAGCCACAACGTCTATATCACCGCCGACAAGCAG
AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGT
ACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATT
ACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGAT
ACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTTCTCCTCCT
TGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCAGCCGCGCTGCCTTGCCCGCTG
CTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTG
TGCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCTTCTTGACCCTGGA
AGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTA
GGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 133)

CN2374 (2171 bp between ITRs):

GCGGCCGCACGCGTTCCGATTCAATTTGCATAATTTTGCTTTTTGGTTAAAAATGCTAACAT
ACATATAAAATCAAGAGTACCGATACTTGTTTCAGGTTGCATTTAACCAGGAAAAATCTTGT
GTTCTGTGAGATTGAACAGCTTCCTCGCACTCACGCTGAAATGTGCAGTTCAATTGCATTTG
CTTTCTAAATGAGAACCACATTTAATGAAACCTGACCATGTAATCAAGATAATGAGGGATT
TTTAATTTGGGTGAGTGACTCAGTTGTGGAGTACACACTTTGACTTGTGTTCCGATGGGAG
GTTGAGAAATTATCACTGCCTTGGTGATACATGAGACAAAGCAGAAAACAAAATCAGCAT
GCGCTGTGTCAATTTGTACTGAGATAACACCCGAGGAACAGCTGTGACTCTTTCAAATGAC
TCCATTTGTTTGATAATTGCTGGGTAAATAAGTGCCACTATTGGCTATTTTCAGGAAGACAG
ATGGCGTTCTAGATTATGTTTCTTTCTTGACCCAAAAAGAACAGGACCAAGTCTTTCTGA
AAATTACTGTCTGTGTGTGAAATAGCATCGGGGAAGGGGACAGCGCATTTCAAGCTTTT
GCCCCGCAGGTAGGTAGGAAGCAGGGGCTGAGCCGTAAAGGAAACAGGGAATTCAATTAT
TCCAGCCTGAGCTAGGGTGGGAAATCAAGGCATCTGTCTGTTGCTTAGGAACAGAAATATC
TTGTGCTTCAGCTACAGTTGGAGAAACCAGGTCATTTACTTTGAGAGTTACCTACTGGTTA
TCCAAACAGAGCTGAAGGAGAAAGCACCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCA
GAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGG
TCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCCGTACCCC
GACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACACAACAGCCACAACGTCTATATCACCGCCGA

74/86

FIG. 8 cont'd

CAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGG
CGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCTACCAAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCCGGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAAGTCGACGGCGCGCCGCGCCGCGAATTCGATATCATAATCAACCT
CTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTA
TGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTT
CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTT
GCCCCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ
ID NO: 134)

CN3584 (2036 bp between ITRs):

GCGGCCGCACGCGTTATAACAAATTATTTCTATTCTTTCTTCACTATTCAGGAACATCTG
TCCACCACTTACATGATCACTTATCTTGCTATTGTGTCATTTTGATGAAAAAGAATTTTTTCT
AAATATCTAAATACAAGGCCCCATATTAACAGTGCTTTTTAAATCCCCACAGATGTGGGAGA
TGACCCCTTTCCATCCCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTTTATAACAAA
TTATTTCTATTCTTTCTTCACTATTCAGGAACATCTGTCCACCACTTACATGATCACTTAT
CTTGCTATTGTGTCATTTTGATGAAAAAGAATTTTTTCTAAATATCTAAATACAAGGCCCCAT
ATTAACAGTGCTTTTTAAATCCCCACAGATGTGGGAGATGACCCCTTTCCATCCCTGAAGAT
TGTAATTGGGCCAGTCTTTAGTACAGTTTTATAACAAATTATTTCTATTCTTTCTTCACTAT
TCAGGAACATCTGTCCACCACTTACATGATCACTTATCTTGCTATTGTGTCATTTTGATGAAA
AAGAATTTTTTCTAAATATCTAAATACAAGGCCCCATATTAACAGTGCTTTTTAAATCCCCAC
AGATGTGGGAGATGACCCCTTTCCATCCCTGAAGATTGTAATTGGGCCAGTCTTTAGTACA
GTTTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTC
TGGGATCCAGATCTTTCAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAG
GAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCA
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGA
AGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCTG
GGCTACCGCGTGCAGTGCTTCGCCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTC
AAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA
GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACA
ACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAAGCGGCATCAAGGCCA
ACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTA
CCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTT
CGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCG
CGCCGCGGGCGCAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATT
GACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAG
TTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAG
CCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTG
TCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTG
GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
AGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 135)

75/86

FIG. 8 cont'd

CN3003 (2623 bp between ITRs):

GCGGCCGCACGCGTTATTGTCTTTACCAGCCCCCGTGGGGACCCAGCTCATATCCCACCT
CCTCTGGGATGTCATCTCTGAAGATACCCTCCCTCAGGACTGTTGATTCATTACAGCAAGCA
CATAAGGGGAAGTCTTCATGCCAGCTGCTGCCCTAGCTCCAATTACAGTACTTATAAAAA
TACACTGTGATGATGTGGTTCCTGCCTCTCCTCTCCCCAAAAACCTCACTCAGTACCAAGA
TCTTTAAAGGCTGGGGTCATGCCTTATTCCATCTTGATTCCCTGCATGGGACACTGGGCC
CTGACACACAGCAGGACTCAGAAAATATTTGCTGAATGAATAAATGGACAAAAGAAAATGG
GATTCATTTTATCTTCATGTCATGGGGTCAAATTCTCATTTTGTGGTATTTTGGGAGAACACT
TCTGAAGAATGTCCAGGTCATACTGTGTCTCTGTGACTGTAAATAGGGCTCTGCCCTAATTA
ACTACTAGCTAGAGGGTGAGAGAGCCCCAGGTGCGCACAGGGCAGTGTGTGGTTTGAAG
CCGGTCCATGGAGTAGCGTCCACGCAGGACAGCCTTGCTGGAAATCTGCGGATGCTAACG
TGGTCCCACACTGTTCTTCCCCAGCCTGAGATCCTGATTAGGAAGTAGCAGCTTCTTCCTT
TAAAACAACAAAAGTGTCTGTCTTCATAGGAAAATGGGAATCCTAGCTGAAGACAGAAGG
GGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCT
GGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCACCATGCTCTCAAGGGCGAGG
AGCTGTTACACGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGATGTCAACGGCCAC
AAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAA
GCTGATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGG
GCTACGGCGTGCAAGTGTTCGCCCCGTACCCCGACCACATGAAGCAGCACGACTTCTTCA
AGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAAGACGACGGCA
ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG
CTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
TACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAAC
TTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCA
GAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACC
AGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTCTGCTGGAGTTCTG
TGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGCTCTGGTGCTACC
AACTTCTCACTGTTGAAACAGGCAGGGGATGTAGAGGAGAATCCAGGGCCTGGTGCTAGT
GGAGACTACAAAGACCATGACGGAGATTATAAAGATCATGACATCGATTACAAGGATGACG
ATGACAAGTCCGGACTCAGATCTGGAGGCTCCGGAGGCCAGAGCCAGCGAAGTCTGCT
CCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAAGGCGCAGAAGAAGGCGGCAA
GAAGCCCAAGCGCAGCCGCAAGGAGAGCTATTCCATCTATGTGTACAAGGTTCTGAAGCA
GGTCCACCCTGACACCGGCATTTTCGTCCAAAGGCCATGGGCATCATGAATTCGTTTGTGAAC
GACATTTTCGAGCGCATCGCAGGAGAGGCTTCCCGCCTGGCGCATTACAACAAGCGCTCG
ACCATCACCTCCCGGGAGATCCAGACGGCCGTGCGCCTGCTGCTGCCTGGGGAGTTGGC
CAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACCAAGTACACCAGCGCTAAGTAATG
AGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAA
AGATTGACTGGTATTCTTAACCTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAAT
GCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTCCTCCTTGATAAATCCTG
GTTAGTTCTTGCCACGGCGGAAGTCAATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGG
CTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTT
GCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTC
CCACTGTCCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATCT
ATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 136)

CN2379 (1619 bp between ITRs):

GCGGCCGCACGCGTCTCCACGGCCACTTCAGTGCCCTTTATCTTCCCCCTGGGGGTTAC
CCCAGACCCCCAGTCAAGCAGCCGCTGCGGGGCGCGCTCCCCGCGGCCGCGACCTAGTT

76/86

FIG. 8 cont'd

CCCTGCCGTTATTTTTAGGGCGCGGGATGGCACCTGCCCACGTGCCGGGGCCCCGCCCGG
 AGAGTGGGTGGGGGGAGCTGCGTCTGCTCAAGGGAGGCGAGGCCCTGTCCCCCGAAGC
 ACCCCCTTCATCCCACCCACGCAATCCCCCAGTTTTTTGTGAGCTCGGGCTGGGCATAAAA
 GTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCAAGCTAG
 CGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCC
 ATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGG
 CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGC
 TGCCCGTGCCCTGGCCCAACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCC
 CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC
 GTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT
 GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGG
 AGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATA
 TCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCG
 AGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC
 CCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCC
 AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTC
 GGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGGCCGCAATTCGATATCA
 TAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCC
 TTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGG
 CTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCGCC
 GCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGC
 TCGAGAGATCTTCGACTGTGCCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGT
 GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATT
 GCATCGCATTGTCTGAGTAGGTGTCAATCTATTCTGGGGGGTGGGGTGGGGCAGGACAGC
 AAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGC
 CGC (SEQ ID NO: 137)

CN3566 (2116 bp between ITRs):

GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
 CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
 AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTGCCCTT
 CTTTCTGTGAAGGGTGACACACCTCCTAGAAAGAAATCTGACTCCCCTCATTGTGTTTTTAA
 AATACTCTCATTATTCCTCCTCTGAAGGTTATTTTCAAACACTCTTTATAACTCGATTGTTTTT
 AAATGCAAAGACAAGGAATCAAATCTGCCAACTTTCTATCAGAGGCCCTTCTTTCTGTGAAG
 GGTGACACACCTCCTAGAAAGAAATCTGACTCCCCTCATTGTGTTTTTAAATAACTCTCAT
 TATTCCTCCTCTGAAGGTTATTTTCAAACACTCTTTATAACTCGATTTTTTTTTTAAATGCAAAGAC
 AAGGAATCAAATCTGCCAACTTTCTATCAGAGGCCCTTCTTTCTGTGAAGGGTGACACACC
 TCCTAGAAAGAAATCTGACTCCCCTCATTGTGTTTTTAAATAACTCTCATTATTCCTCCTCT
 GAAGGTTATTTTCAAACACTCTTTATAACTCGATTTTTTTTTTAAATGCAAAGACAAGGAATCAAA
 TCTGCCAACTTTCTATCAGAGGAGCTCGATTTCAGCCGGGAGCTTAGGGAGGGGAGGTCAC
 TTCATAAAGGGCTTGGGGGGGGAGTTGGAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTG
 GCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCCAGATCTTTCGAAGCTAGCGCTACCGG
 TCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTC
 GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGGCGAGGGCG
 ATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCCTGTC
 CCTGGCCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCCGCTACCCC
 GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
 CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
 GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
 CATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCACCGCCGA

77/86

FIG. 8 cont'd

CAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGG
CGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCTACCAAGCTCAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCCGGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAAGTCGACGGCGCGCCGCGCCGCGAATTCGATATCATAATCAACCT
CTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTA
TGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTT
CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTT
GCCCCGTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ
ID NO: 138)

CN3453 (2600 bp between ITRs):

GCGGCCGCACGCGTCCTTAATAGTCTTAATATAAGGTACGATTTTATTTATGTCTTTCTGAC
AAAGGGTTTCAGACCCCTGGAGGCCTCTGTTAACACAGTACTCTCAGATGTAAGATTTTTAA
CACAAATTCCTAACCCCTTGTAGACTCCTGTTTTCTTCATCAGGCACAGCACAGAGAGAA
GCTCTAGATTAACGATATAGGCAGGAAAGTTGGGGGAGAAATGAAAACCTGTCACCCGTGT
CATTATTTCCGCATTAAGCCTGAAGGAGTAGGAGAGGTGTCACTCTCAGCGTCTCTGCGGT
TGCCTGAGCTGCATCAGTGTCTGCAGCTGCATGATACAGTCATCAGAGCTCTCTGAAGCAC
TCACCTCCCCCCCCCCCCACCTGTGCACCTCTCTCCGTGGAAACAAGGCAATGTTTGTTCAA
GCTCAATGGAAAATAGTAGCAAGCGTGATCTCACTTTGCTTTTTCTTTTAAGGTAACAGCT
GATCGCTGGCTTGAGCACTTGGCAAAGAACGAGGTGCCACCCCGAGCTCTGAGCATTTCA
TTGAACCTTCATAGCAACGGTCCTAGGAGGTGGCAGCTATTATTATCTGCATTTTACAGGTA
AAGTATCTGACGCTCTGAGGTTACAGCATCTGCCAACTAGCAAGAAAGAACTGGAGTTCA
AAAGGAAGTCAGGCCATTGACACAAAACGGCCTGTCCGAGCTCGGGCTGGGCATAAAAGT
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCG
CTACCGGTGCGCACCATGGTCTCCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCAT
CCTGGTCGAGCTGGACGGCGATGTCAACGGCCACAAGTTCAAGCTGATCTGCACCACCGGCAAGCTG
AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTG
CCCCGTCCTTGCCCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGCAGTGTCTCGCCCG
CTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT
CCAGGAGCGCACCATCTTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA
GTTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAGCCACAACGTCTATATCA
CCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGG
ACGGCGGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCTACCAAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGG
CATGGACGAGCTGTACAAGGGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGG
GGATGTAGAGGAGAATCCAGGGCCTGGTGCTAGTGGAGACTACAAAGACCATGACGGAGA
TTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGATCTGGA
GGCTCCGGAGGGCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAA
GGCGGTGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAG
AGCTATTCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCTGT
CCAAGGCCATGGGCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAG
AGGCTTCCCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGA
CGGCCGTGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACT

78/86

FIG. 8 cont'd

AAGGCCATCACCAAGTACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCCGCGAATTGCG
ATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGT
TGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCC
GTATGGCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAATC
ATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTC
CGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTC
CCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAG
GAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG
GACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCG
AGCGGCCGC (SEQ ID NO: 139)

HCT1 (2097 bp between ITRs):

GCGGCCGCACGCGTGGGAGTCTGGAGAGAGGGTGGGAGCAGCCATTCTGCAGCAGTGC
CTTCTTGGGGTCATGGGTCTGTAGGTGCTGCTGTGGAGGGAGAGATCAGCCTATTCTGGC
TTCATTTCTGAGCTGCAAACTGCCTGGGTGTCTGGAGAAGCAGGTTGGCGTGGTGGTTAG
CAGTGCGTGGGCGGGGTTGCCCGCTCTTGATTTATGATTTCTTTGTCTCTGTGGAAGCACT
TAAGTGCAAGGCTTTAGTTCCAATGACACTCAGGAGCCTCTGGATTCCAGCACTGGGGATG
GGGGTGGGGTAGAACGTTCTCAGGCCTACCAACCCCTCCCCTGTGTGCTGCCTTTGGGA
GAGTCCCAAGGCTTCAGCATTACTTAATTAATTAGGCCTCTACTGCTACATAGGCTCAGATT
CAAAAGAACAGAGTGGCCACGTCAGCCATTCCCGGAAAAGTCTGATGGCTGGAAGCCAG
AGGACTATGTGTCTGCCTTGCTGCCCTTGCCAGCCCATCCTGAATGCCAGACTCGGAC
AATGGAGTAGGTACAGAAGGGTAAAGACAGTGTCTTCTGTACCAGTAAGTGGGCCCTGAT
CTGCTCTCTACAGCTTCCAGAGAAAGGGCCTGGCCAATGAGCGGCCCTTTTGAGTAGCAGA
TACCTCACATGCATTCTGATAGAAAGCCTGGCCCCAGATCACTGTGACTTTAGCCCTCAGG
TTTCTTTTGACTTTCCTATTTCACTGAGTTTGAGCTCGGGCTGGGCATAAAAGTCAGGGCA
GAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGG
TCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCCGCTACCCC
GACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGCGGCACAAGCTGGAGTACAACCTACAAGCCACAACGTCTATATCACCGCCGA
CAAGCAGAAGAACGGCATCAAGGCCAATTCAAGATCCGCCACAACATCGAGGACGGCGG
CGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGGAATTCGATATCATAATCAACCT
CTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTA
TGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTT
CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTATCGCCGCCTGCCTT
GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCTT
GACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ
ID NO: 140)

79/86

FIG. 8 cont'd

CN2146 (1590 bp between ITRs):

GCGGCCGCACGCGTGGGCTTTCTAAACTGGCCTGGAACACTGTAGTTAATAAACATTCTT
TTCAGGGTTGTTCTCTTTGCTTCTGATCTCATTGATGATGATTTGTTGGACAAATAAGGAAG
TAAACAGTGCAGAGTTTTTTGGCTTCTGCTCTATTA AAAACATAGCAGAGAAGGACAAGAAG
TTCCTGCCTGCTAAATAGGAAGATCTAGGGAATCAAAGAGTCTCTGTCAAGCTTCTTGTA
GAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGG
GATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGA
GCTGTTCAACGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACA
AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
CTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGG
CTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAA
GTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAA
CTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
ACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAAGCGGCATCAAGGCCAACT
TCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCCTACCAGCAG
AACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCA
GTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT
GACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGC
CGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACT
GGTATTCTTAACATATGTTGCTCCTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTAT
CATGCTATTGCTTCCCGTATGGCTTTCATTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTT
GCCACGGCGGAACATCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGGCTCGGCTGTT
GGGCACTGACAATTCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCAT
CTGTTGTTTGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCTT
TTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGG
GTGGGGTGGGGCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATC
TCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 141)

AiP1347 (2543 bp between ITRs):

GCGGCCGCACGCGTCCAATGTCCCAGGTGAAAATGAACACTTGCAGGTTTCTTCTTCCAGT
CCCAGCCCTCTGGCCACACCCCACTGCTGACCACAGGAGCCTCTCACAGTGCTCACTCA
GGGCTTCAAACCTCTCTCCCTGGCAGCTTCTTTCTATTCTGAAGTCTCCAGTCCAGCACA
AATTAGTCCCCTCTCCTGACTCCTACTGCGCCGCTTGGTACATCACAGAGCACTGGCTG
AGAAAATCGCCCTTTAGTCCCGCAAATCACTCTTTACCTGCTCCACTGACTAGGCCTGT
GCCATTATTTATTAGGGAATGAGCGCTAGCTGTTGAGGGTCAATTATTCCATTGACCAGGC
ATTACAATAATTCCTGGTAATTAAGAGGTAATTATGTGAATCTGGATGCGTATTGAAAGA
AACATTAGTCCTTTTGTGAGCTTGGCAAGTCTATTGTTCTGAGCCAGGCCCAACCAATTAAC
ATCTTTTGCCAATCCCTGTGAGCAGGGGCTTGCAAGGAGGAGAGAAAGGGGGGCGGTC
AGATGCTAATTTAGATACAATTGTGTCAGGGTGCATGGGAGGGGTAACCTCTGAAGAAGAGA
CTCCAGTTTAATGAAGCAAACAGCTTGGGTCCCAGTGACCCTGCAGGGGTAAATTTTAAAG
GCCCTCTGTGTTCTGTGCCAAATTGCAGAGGGCCAGCAACAAGAAGCCCTTCAAAAAA
AAAAAAGAGGGAGCAGGATAATCATCCACTTAATCTGCTCATTAAACAATCTCTCAGGCG
GACTTAACGTGAATTAGCACTTCTCTGAAAGGGGCGAGCCTTGATGAGCAGCAGACAGCTCA
GATCATTTTGAAGGTACAGTTTTTTTTTTTATTATTATATAAGCCCAGTTGTTATTAATGTATT
CATATTTGAAGAACTGAGCAGAGACCTGCTTTCATTCAATCTAGTGAGCAAAACAGTACGT
TTTCCTAACTGGAAAACAAAACATAAATGCTAATCAGGTCCCAGCAGAGATCAAATCAAGCT
AAATATAGCAGGCACCCCTTTGTGGTTTTTTTTTTGTTTTGTTTTGTTTTTTTTTTTT
AATACAGAAATTAAGGAAAGAAAAAACCATTGAATTAATTGTCTTCTCTGTGCTTCGCA
CCTGACACCAGCGAGGCTCACACAAGGGCCGAGCTCGGGCTGGGCATAAAAGTCAGGG

FIG. 8 cont'd

CAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTTGAAGCTAGCGCTACC
 GGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACACGGGGTGGTGCCCATCCTGG
 TCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
 GATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTG
 CCCTGGCCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCC
 CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
 GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
 GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
 ACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCACCGCCG
 ACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCG
 GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
 CTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAG
 CGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGA
 CGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGCGCGGAATTTCGATATCATAATCAACC
 TCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAGTATGTTGCTCCTTTTACGCT
 ATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTT
 CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAAGTTCATCGCCGCTGCCTT
 GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
 CTTGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCGCGTGCCTTCCTT
 GACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCAT
 TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
 GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGCA (SEQ
 ID NO: 142)

AiP1351 (2504 bp between ITRs):

GCGGCCGCACGCGTAAGGCTATAGTCCTTGGTGTGAAAGACTCATTCAATTTCTCTCTTAT
 CTTCAATGTTAGGGACTGATTCAAAACAGCATAATCAAAACAAAGCAGAGCACAGCATAATT
 AAAATCATGTCTTCTGAATAGGTAATTACAAAATAATTGGATATTGGTAATTATTTTCTTTT
 ACCTGTTCTTGGCTCCTTTCTCAGTTAATATGTAGTGCTATGTACAGATGAAGTTCAGATGC
 AATTATGTACAGGGTCTATCTCACCACCAGGAAGCAGGACTTTTCTGTGTGATCTTCCATTT
 CCTGCAAGGACTACAGAATTGAGGTACAAGTGCCCTTTTGTTCCTCAAGTTGAGTCCTTCT
 GTTTTCTTTGTCACTGGCTTTTCATCCTCTGTCCCCTGGCTGTGGCTCAGTGTTTATTTATT
 GGCTGTTTAGTATGGAGAAACATCCAGTCTTTCAACATTGACAACAAATCCAAATAATGAGC
 CTGCCCATCAGCTACACTGGGTATCATCTTTGCCTGTTAGGCTCTTTGGCTGTTTTTTGCT
 GTATTTGACTATCCTCAAAGTTTTGGGAGTTCTGACTGCGGGTTGTTATTTTCCAATTCATTAT
 AACACTTGCCAAATCTTTTGTATGCATCTTTATACTCATGACAGCTTCTAGGTCAAATTTTA
 GTTTTGATCACTGGATCTTTAGAAGGAAAGCTAAGGAACTGACAGACATAGCCAGCTCCTT
 TTTTATGCTGGCAATGGATACTGATTAGCTGTGTGCTAGTTGTTGAAGTCATAGCAACAAG
 TCCTACTTACTATCTTTTAGGCATCTCAATTAAGTTTCTTTGGTCTTTGGAAATATAACTACA
 ATTAAGTTTCATCTCCTGGAGAAAATGAGTATGTTTCAGGAGAACAACTTCATAGAGTTAATT
 AAGAACTCTTTTCTAGAGTTCAATTTTAACTGGTAAGTTATTTGACTCCTTCATTTTCTTTT
 TGTTTCCAGGTTTTATAATAGGCTTTTGATGGATTAACTTTTAAATTTCAAAGTACTATTTG
 CCAAGATTCTGGATGCCTTTGCTGAGGCTACCAAGTTAAGTATATGCATCAGCAACCAGAG
 GAAAAACACTAACGTTGGGCAGAGAAGCAATGAATACCGAGCTCGGGCTGGGCATAAAAG
 TCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGC
 GCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACACGGGGTGGTGCCCA
 TCCTGGTGCAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGC
 GAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCT
 GCCCGTGCCCTGGCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCC
 GCTACCCCGACCATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG

FIG. 8 cont'd

TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAG
GACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATC
ACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAG
GACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCCGACGGCCC
CGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAA
CGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCG
GCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATA
ATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTT
TTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCT
TTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGC
CTGCCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTC
GAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCTAATAAAATGAGGAAATTGC
ATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA
GGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCG
C (SEQ ID NO: 143)

AiP1375 (2047 bp between ITRs):

GCGGCCGCACGCGTACTCAGCTGCTTTTCTGAAGGCAGCTTCAGCTTTTCCAAAGCCAG
GGTGCTGCTGTGGTCAGCTAGTGCTGCACTAAATATCCTCTTCCAAGGGGCTCTGATCCTG
GCTCCAGCCCCCTCTGGCCTCTGGGACTGGTTTTCTCTATTTCTTTCTTCTCCAAGGTTA
GCTTTTCAGACTCAACAACATAAATCAAGAGCCTCTAGGGAGGTGAACAGAAGTGGCCGGA
GAGGAAGCACTGCTTCAGTGGGGGAAGGGGAATGTCTTGACTAGCAGGCCCCAGTAGG
GGTGGAGTGTGGGGGTATGGTGCATACTCAGATGTCTCCTTGAAGGAGTGGGTGGCAA
GGATTCTTCCATAAACTGATAAGCCGGGACACAGAGTCCCATCCCCATCCCCACCTGTT
ACACTGCTATAAACAGAGAGGGCTGAAGGCCAAAAGGACCCTGCTTGCTGCTGCAGCTGGT
CCAGTCTAGGGGGCTCTTCAGTAGAGGACAGCTCTCTTCTTGTCAGTGCTTTCTCTCTG
AACAGAGCTGGTTGTTTTGGGGCACTTGGGCTCCTGTGGACTAATGGGGTCAAGGACCAG
GCCAGACACACACTAAGGCCTGGAGTCCCTGGGGACTAGTGGTTCCAGTTCACTAACCTC
TGCCCTTCAAATCTGGGCTCTGGCTCTCCCAATTGCTTCCGAGCTCGGGCTGGGCATAAAA
GTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAG
CGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCC
ATCCTGGTGCAGCTGACGGCGACGTAACGGCCACAAGTTACGCTGTCCGGCGAGGG
CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGC
TGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGCAGTGCTTCGCC
CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC
GTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT
GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGG
AGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATA
TCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCG
AGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC
CCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCC
AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCA
TAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCC
TTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGG
CTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCC
GCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGC
TCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCGT

82/86

FIG. 8 cont'd

GCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTCCTAATAAAATGAGGAAATT
GCATCGCATTGTCTGAGTAGGTGTCTATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGC
AAGGGGGAGGATTGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGC
CGC (SEQ ID NO: 144)

AiP1307 (1957 bp between ITRs):

GCGGCCGCACGCGTAGGGTGCCAGGAGAAATGTGACCTCAAAGTCTTGTTCTATAACTGTT
GGACCTTAGGAGAGATCTGTGCTCAGCAATTTACCAGGACACCCCCACCCACATGTCTTG
ACCACTGTCTGGATAACTGGTATGCAGGACCACACTAGGCTTACTCACAGTGTAACCTCTC
ATAACCATCACTGGAGCCCATCCTGCCTGGTAGACAAGGATTCAACCATGACTCATTGTAC
TTTAGTGGTGCCATGCTTAGTCATCAGGTGCCCTGTGCTCTGACAGCCGAGGGTCAGAGC
TGGAATCACACTCTTGTTGTCTTTAATCTCTCCCTCCCTTTCTTCTTCTTCTTCACTCTG
TTGTGATTGCTCATGGAACAGATCCTAGCTGGTCTCCCTGGCAACCTACATGATTTGAGCC
CAACAGATGGATAATGGGGACATCGACTTCCAATGTCTATTCAACAGAATCATTGCCAAGGG
AGTCTGATGAGCAGGCAACTGAGATGACACCCTTATCAATATAGCTTCATTTTGCAATCTG
GAGTAGGTGTTTCAAAGGAGAGCCCCCACTGATGCCAGCAATACAGAACGTTTCATGGGC
AAGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCT
GGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCCACCATGGTGAGCAAGGGCGAG
GAGCTGTTACCGGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCA
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGA
AGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTG
GGCTACGGCGTGCAAGTGCTTCGCCCCTACCCCGACCATGAAGCAGCACGACTTCTTC
AAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGCATCGA
GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACA
ACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAAGCGGCATCAAGGCCA
ACTTCAAGATCCGCCACAACATCGAGGACGGCGGGCGTGAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTA
CCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTT
CGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCG
CGCCGCGGGCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATT
GACTGGTATTCTTAAGTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGATCATGCTATTGCTTCCCGTATGGCTTTTCATTTTCTCCTTCTTGTATAAATCCTGGTTAG
TTCTTGGCAGCGGGAACCTCATCGCCGCTGCTTGCCTTGCCTGCTGGCAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAG
CCATCTGTTGTTTGCCCTCCCCCGTGCTTCTTGACCCTGGAAGGTGCCACTCCCACTG
TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCTATTCTATTCTG
GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
AGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 145)

AiP1530 (3xcore_E116) (2046 bp between ITRs):

GCGGCCGCACGCGTTGTAAGAGTAATAATCGCCGGCGCCTGGGCCTGAGAGAGACTCCG
AGCAGCTGCCTGTGCAGCAGTGTGAGCCTGCCAGGTTGTGCTTGGGGACACCAGTGACA
CCAGCTGCTCCTCATATCTCCACACGAGGCTATTTTAAGAACTCCGCACTTGATCAGGGA
GCATGTGGGGTTAAGCCTTATGTTTTCCACTTTGGTTTTTAGATCCTAAGCAGGTGCTGCA
GTTAAAAGTTCAAAATCCTGGGCCTGAGAGAGACTCCGAGCAGCTGCCTGTGCAGCAGTG
TGAGCCTGCCAGGTTGTGCTTGGGGACACCAGTGACACCAGCTGCTCCTCATATCTCCAC
ACGAGGCTATTTTAAGAACTCCGCACTTGATCAGGGAGCATGTGGGGTTAAGCCTTATGT
TTTCCACTTTGTTTTTATAGATCCTAAGCAGGTGCTGCAGTTAAAAGTTCAAAATCCTGGGCC
TGAGAGAGACTCCGAGCAGCTGCCTGTGCAGCAGTGTGAGCCTGCCAGGTTGTGCTTGG

FIG. 8 cont'd

GGACACCAGTGACACCAGCTGCTCCTCATATCTCCACACGAGGCTATTTTAAGAACTCCG
CACTTGATCAGGGAGCATGTGGGGTTAAGCCTTATGTTTTCCACTTTGGTTTTAGATCCTA
AGCAGGTGCTGCAGTTAAAAGTTCAAAATCGCCGGCGGAGCTCGGGCTGGGCATAAAAGT
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTTGAAGCTAGCG
CTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT
CCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCGGCGAGGGCG
AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTG
CCCGTGCCCTGGCCCAACCTCGTGACCACCTGGGCTACGGCGTGCAGTGCTTCGCCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT
CCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA
GTTGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTATATCA
CCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGG
ACGGCGGGGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCGCGGAATTGATATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTT
TACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTT
TCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCGCCGCC
TGCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCGTGGCTCG
AGAGATCTTCGACTGTGCCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 146)

AiP1379 (1941 bp between ITRs):

GCGGCCGCACGCGTCGCTTTACCCAGCCGTATCTCATTTTTAGACATTTAAGTTACTTCCA
ACTTTTCCTGCTTATCAGTTACGTCAGCATGGCTATTTTTATACCTAGATGGCCTGTGTTCC
TGACTGCTTGGAAGTAGACTTACTGGGTCAAAGGGGATGCCAGTGTGATTTTCAGAAAACG
ATGCCCGTTTCTGTTACAGCTGTTTTCTATAAAGCATGGGAACGGCGCAGGCACTATTTA
TGTGCCCTGCTGAATTCTCAGCTCCCTGCTTCACTGACCTTCCGCCACTACGAAATCG
GCTCTGTTCTTCTCCTGGGCTCCCGATGGGAGCCAAAGCTGGACTTGTGTCAGTGATTAAG
AAGTGCTCGTTATAGCTCATTGCCCATGGGATGTGCACTGCTGCTGATTCTCAGGGCTGG
AGATGACACATTTTCTCACATTCTTCAAAGTCTTAAGGTTTCAGCAGGAGAGAAGGAAGGT
CCGAGCCAAGGAGAGCTGCGGCCATACCGACAGGGAGGGCACTCTCTCTTTCTCTAGCTC
TCCCCCCCCCCCCCTTGAAGAGGGGAGAAAAGCTCCCTCCCTGCCTCCTGAGCTCGGGCTG
GGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTT
CGAAGCTAGCGCTACCGGTGCGCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGG
GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC
CGGCGAGGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCA
CCGGCAAGCTGCCCCGTGCCCTGGCCCAACCTCGTGACCACCTGGGCTACGGCGTGCGAG
TGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC
GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAA
CGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCA
CAACATCGAGGACGGCGGGCTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGC

84/86

FIG. 8 cont'd

AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGCGCGAA
TTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAC
ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAA
CTCATCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAA
TTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCC
CTCCCCCGTGCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAAT
GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGA
CCGAGCGGCCGC (SEQ ID NO: 147)

CN3916 (2204 bp between ITRs):

GCGGCCGCACGCGTCAGCTCACGTCTGAATCCCTCGGCGCCCCCTTTCTCTTCTCCTAGC
CCCTTCCTCACGTCCCCTGCCTCCGGGTATTTCTCTCTCCAATCCCCACCCCCGCACC
GCCTGATTCCGAGGGGGCGGGAGCGCATTGGGCTGCGCACGGGTGGGGGCGCCGCGCCA
GCTTCGCGTAGCTGCTCTGACGCCGCTGCCGCCGCCGCCGCCGCCGCCGCCCTCCGCA
GCCCAGCTCGCGCCCCGCGGCAGCTCCGCAGTGCCTAGCCACCACCGCCGCCGCCGC
CGCTCCGCCAGACCTGCTGCGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCT
ATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCA
TGGTCTCCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGAC
GGCGATGTCAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTA
CGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCA
CCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCCGCTACCCCGACCATG
AAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
TTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCACCGCCGACAAGCAGAA
GAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGC
TCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC
AACCCTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCAC
ATGGTCTGCTGGAGTTCTGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTAC
AAGGGCTCTGGTCTACCAACTTCTCACTGTTGAAACAGGCGAGGGGATGTAGAGGAGAAT
CCAGGGCCTGGTGCTAGTGGAGACTACAAGACCATGACGGAGATTATAAAGATCATGAC
ATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGATCTGGAGGCTCCGGAGGCCCA
GAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAAGGC
GCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCTATGT
GTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCTGTTCAAGGCCATGGGCAT
CATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTCCCGCCTGGC
GCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCGTGCGCCTGCT
GCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACCAAGT
ACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTC
TGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTAT
GTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTTC
TCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTG
CCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATC
TTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCTTCTCTTG
ACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATT
GTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAG

FIG. 8 cont'd

GATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 148)

CN3869 (2370 bp between ITRs):

GCGGCCGCACGCGTGATTCTAGATTTTGAGAGACAGCTACGTCTTTGCTTCTCTGGTAGTT
CCCTATTATTAATTATTGGGGTTTTGACGGCGTTCTGGCCTCCGTTTTTTGTTGATTTTCATTTT
AGTGCTGCGTTTTCTCCTGACTATGTTGATGATTTGCTGCTTATATAAGGGAAAGGATGTATT
GAGGTGGACTGATGCTTGGGGTTTTGAGATCGGCTACTGAGATTGGGTGGCGGTTGGTGT
TTTTGTGCGATGTGCATGGTAGGAAGGTCTTTTCTTTTGGGTGGTAGGGGATGAGAGAATGG
TGGGAGAACGGTAGATGAGATGCTACAGCAAACGGGGATGTGGGGGGAGACGGAGGTGA
GGTTGGAAGTCACCGGTTTGCAGAATTTCTATTCTGCACCCCCCTTTCTCTCGCTCCTGGA
AAAAAATCCATTATCTATAAAAACACATTATTCTTCAATTGGCCTTTTAGAGGAGCTCGGG
CTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATC
TTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTCTCCAAGGGCGAGGAGCTGTTACCGG
GGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGATGTCAACGGCCACAAGTTCAGCGTG
TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCAC
CACCGGCAAGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGC
AGTGCTTCGCCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGC
CCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAAGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCAC
AACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGC
CACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCAT
CGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGA
GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCC
GGGATCACTCTCGGCATGGACGAGCTGTACAAGGGCTCTGGTGCTACCAACTTCTCACTG
TTGAAACAGGCAGGGGATGTAGAGGAGAATCCAGGGCCTGGTGCTAGTGAGACTACAAA
GACCATGACGGAGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTCCG
GACTCAGATCTGGAGGCTCCGGAGGGCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAA
AAGGGCTCCAAGAAGGCGGTGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCG
CAGCCGCAAGGAGAGCTATTCCATCTATGTGTATAAGGTTCTGAAGCAGGTCCACCCTGAC
ACCGGCATTTTCGTCCAAGGCCATGGGCATCATGAACCTCGTTTGTGAACGACATTTTCGAGC
GCATCGCAGGTGAGGCTTCCCGCTTGGCGCATTACAACAAGCGCTCGACCATCACCTCCA
GGGAGATCCAGACGGCCGTGCGCCTGCTGCTGCTGGGGAGTTGGCCAAGCACGCCGTG
TCCGAGGGTACTAAGGCCATCACCAAGTACACCAGCGCTAAGTAATGAGGCGCGCCGCG
GCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTA
TTCTTAATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATG
CTATTGCTTCCCGTATGGCTTTTCATTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCA
CGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGC
ACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTT
GTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCAC
GTGCGGACCGAGCGGCCGC (SEQ ID NO: 149)

CN3870 (2359 bp between ITRs):

GCGGCCGCACGCGTTTTGAGAGACTACTACGTCTTTGCTTCTCTGGTAGCTCCCTAATATT
TATTACTGGGGGGACTGATGGGGTCTGGCCTCCGTTTTTCTTGTTTTCTTTTGTGCTGC
GTTCTATGTTGATGGTTGCTGCTTATATAAGGAAGGGATGGATCGAGGTGGGTGATGCT
TGAGGTTTTGAGATCGGCTACTGAGATTGGGTGGCGGTTGGTGTGTTTGTGCGATGTGCAT

FIG. 8 cont'd

GGTAGGAAGGTCTTTTCTTTTGGATGGTGGGGGACAAGGGAATGGTCGGAGAATGGTAGA
TGAGATGCTCCAGTGAATGGGGACGTGAGGGGACACAGAGGTGAGGTTGGAAGTCACCG
ATTTGCAGAATTTCTGTCTGTATCCCCCTTTCTCTCGCTCCTTGAAAAAAAATCCATTA
TCTCTAAAAACACATTATTCTTCAATTGGCCTTTTAGAGAAGAGGAGCTCGGGCTGGGCAT
AAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAG
CTAGCGCTACCGGTGCGCCACCATGGTCTCCAAGGGCGAGGAGCTGTTACCGGGGTGGT
GCCCATCCTGGTCGAGCTGGACGGCGATGTCAACGGCCACAAGTTCAGCGTGTCCGGCG
AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGC
AAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGCAAGTCTT
CGCCCGCTACCCCGACCATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGG
CTACGTCCAGGAGCGCACCATCTTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAACAGCCACAACGTCTA
TATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACAT
CGAGGACGGCGGCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACG
GCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAAGTCCAAGCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGGATCACT
CTCGGCATGGACGAGCTGTACAAGGGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAG
GCAGGGGATGTAGAGGAGAATCCAGGGCCTGGTGCTAGTGGAGACTACAAAGACCATGA
CGGAGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGA
TCTGGAGGCTCCGGAGGGCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTC
CAAGAAGGCGGTGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCA
AGGAGAGCTATTCCATCTATGTGTATAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCAT
TTCGTCCAAGGCCATGGGCATCATGAACTCGTTTGTGAACGACATTTTCGAGCGCATCGCA
GGTGAGGCTTCCCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCAGGGAGATC
CAGACGGCCGTGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGG
GTAATAAGGCCATCACCAAGTACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCGCGCA
ATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAC
TATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCT
TCCCGTATGGCTTTTCATTTTCTCCTCCTTGATATAAATCCTGGTTAGTTCTTGCCACGGCGGA
ACTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACA
ATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCC
CCTCCCCCGTGCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTCTCTAATAAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGG
GCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGG
ACCGAGCGGCCGC (SEQ ID NO: 150)

FIG. 1A

