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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 May 2022 (19.05.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/103859 A1

(51) International Patent Classification:
C12N 15/86 (2006.01) *C12N 15/861* (2006.01)

(21) International Application Number:
PCT/US2021/058812

(22) International Filing Date:
10 November 2021 (10.11.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/112,102 10 November 2020 (10.11.2020) US

(71) Applicant: **ALLEN INSTITUTE** [US/US]; 615 Westlake Avenue North, Seattle, Washington 98109 (US).

(72) Inventors: **DAIGLE, Tanya**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **GRAYBUCK, Lucas T.**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **KALMBACH, Brian Edward**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **LEIN, Edward Sebastian**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **LEVI, Boaz P.**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **MICH, John K.**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **SEDENO CORTES, Adriana Estela**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **TASIC, Bosiljka**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **TING, Jonathan**; 615 Westlake Avenue North, Seattle, Washington 98109 (US). **ZENG, Hongkui**; 615 Westlake Avenue North, Seattle, Washington 98109 (US).

(74) Agent: **WINGER, C. Rachal** et al.; Lee & Hayes PC, 601 W. Riverside Avenue, Suite 1400, Spokane, Washington 99201 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,

(54) **Title:** ARTIFICIAL EXPRESSION CONSTRUCTS FOR MODULATING GENE EXPRESSION IN CHANDELIER CELLS

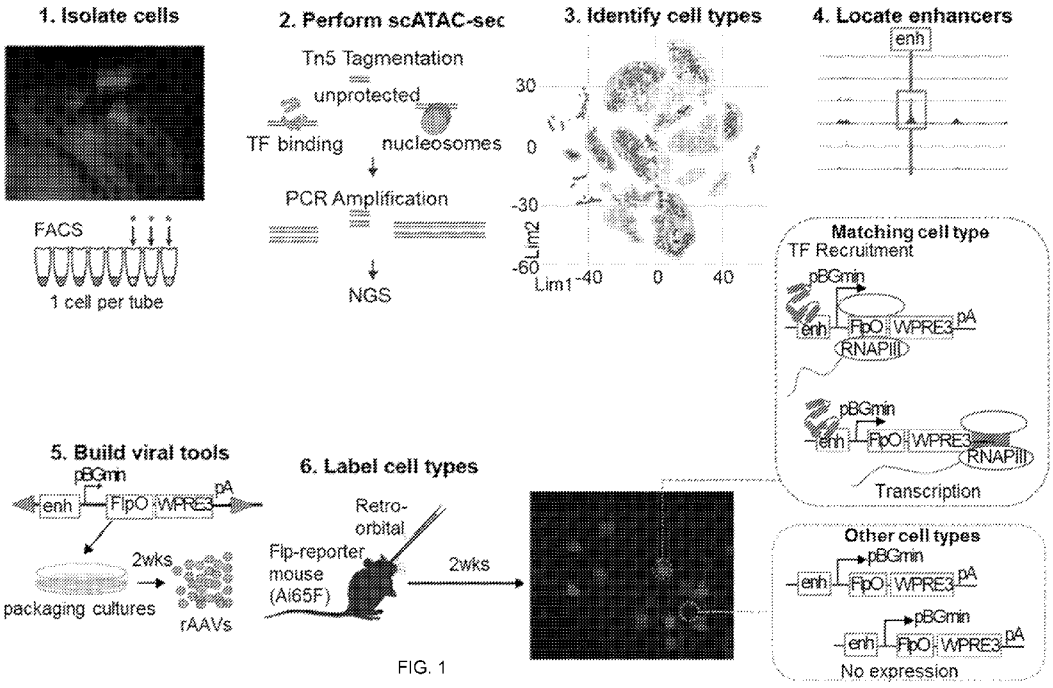


FIG. 1

(57) **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 chandelier cells. Chandelier cells are a subtype of GABAergic interneurons that that have been implicated in disorders such as epilepsy and schizophrenia.

[Continued on next page]

WO 2022/103859 A1

SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

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

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

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

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

ARTIFICIAL EXPRESSION CONSTRUCTS FOR MODULATING GENE EXPRESSION IN CHANDELIER CELLS

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority to U.S. Provisional Patent Application No. 63/112,102 filed November 10, 2020 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 DA036909 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 text format in lieu of a paper copy and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is A166-0026PCT_ST25.txt. The text file is 184 KB, was created on November 10, 2021, and is being submitted electronically via EFS-Web

FIELD OF THE DISCLOSURE

[0004] The current disclosure provides artificial expression constructs for modulating gene expression in targeted central nervous system cell types. The artificial expression constructs can be used to express synthetic genes or modify gene expression in chandelier cells.

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 double or 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] Chandelier cells are a subtype of GABAergic interneurons that were discovered in the 1970s and that have been implicated in disorders such as epilepsy and schizophrenia. While significant information has been learned about this cell type, many questions regarding their function and brain-wide effects remain.

SUMMARY OF THE DISCLOSURE

[0007] The current disclosure provides artificial expression constructs that drive gene expression in chandelier cells. Particular embodiments of the artificial expression constructs utilize the enhancers eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, and eHGT_1024m to drive gene expression within chandelier cells. Additional embodiments utilize the enhancer eHGT_503m to drive gene expression in chandelier cells and VIP cells. Additional embodiments utilize enhancer eHGT_710m to drive gene expression in chandelier cells, glutamatergic neurons in the thalamus, and/or molecular layer GABAergic interneurons in the cerebellum.

[0008] Particular embodiments provide artificial expression constructs including the features of vectors described herein including vectors: CN1917, CN2047, CN2048, CN2049, CN2427, CN2320, CN2321, CN2719, CN2707, CN2713, CN2717, AiP1104, AiP1089, AiP1105, AiP1090, AiP1106, AiP1091, AiP1092, CN2581, CN2782, CN3407, CN3408, CN3409, CN2580, CN2825, CN3270, CN3316, CN3271, CN3793, CN3794, CN3795, CN3790, CN3751, and CN3752.

BRIEF DESCRIPTION OF THE FIGURES

[0009] 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.

[0010] FIG. 1: Overview of enhancer discovery for viral tools. To build cell type-specific labeling tools, cells from adult mouse cortex were isolated and a single cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) was performed. Samples were clustered and compared to single cell RNA sequencing (scRNA-seq) datasets to identify the clusters. Single cells matching the same transcriptomic types were then pooled and the genome was searched for type-specific putative enhancers. These regions were cloned upstream of a minimal promoter in an AAV genomic backbone, which was used to generate self-complementary adeno-associated viral vectors (scAAVs), recombinant adeno-associated viral vectors (rAAVs), plasmid adeno-associated virus vectors (pAAVs). These viral tools were delivered retro-orbitally to label specific cortical populations. In cells with a matching cell type, enhancers recruit their cognate transcription factors to drive cell type-specific expression. In other cells, viral genomes are present, but transcripts are not expressed.

[0011] FIGs. 2A,2B. CN2320 (eHGT_475m) in mouse neocortex. (2A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the neocortex 28 days after intracerebroventricular injection of 8.0E10 viral genome copies of AAV vector #CN2320 (eHGT_475m) in 2-day old mouse pup. Scale bar: 100 microns. (2B) Higher magnification view showing a cell body and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0012] FIG. 3. CN2320 (eHGT_475m) in VISp of 554420 and 584422 animals. Mapping of single cell transcriptomic profiles of SYPF2+ cells sorted from the VISp region of the mouse neocortex after retro-orbital injection of CN2320 virus packaged with the PHP.eB capsid. Number of cells mapped to the final leaf are shown on the bar plot below the dendrogram. This data shows eHGT_475m enhancer-driven reporter expression is enriched in chandelier cells (Pvalb Vipr2 cells) when mouse VISp is evaluated. Transcriptomic cell types are shown on the bottom and read (from left to right): 169 L2/3 IT VISp Rrad, 168 L2/3 IT VISp Adamts2, 167 L2/3 IT VISp Agmat, 164 L4 IT VISp Rspo1, 163 L5 IT VISp Hsd11b1 Endou, 162 L5 IT VISp Whrn Tox2, 160 L5 IT VISp Batf3, 158 L5 IT VISp Col6a1 Fezf2, 157 L5 IT VISp Col27a1, 154 L6 IT VISp Penk Col27a1, 153 L6 IT VISp Penk Fst, L6 IT VISp Col23a1 Adamts2, 149 L6 IT VISp Col18a1, 146 L6 IT VISp Car3, 144 L5 PT VISp Chrna6, 143 L5 PT VISp Lgr5, 142 L5 PT VISp C1ql2 Ptgfr, 141 L5 PT VISp C1ql2 Cdh13, 140 L5 PT VISp Krt80, 134 L5 NP VISp Trhr Cpne7, 133 L5 NP VISp Trhr Met, 131 L6 CT Nxph2 Sla, 130 L6 CT VISp Krt80 Sla, 128 L6 CT VISp Nxph2 Wls, 127 L6 CT VISp Ctxn3 Brinp3, 126 L6 CT VISp Ctxn3 Sla, 122 L6 CT VISp Gpr139, 120 L6b Col8a1 Rprm, 119 L6b VISp Mup5, 118 L6b Col8a1 Rxfp1, 115 L6b P2ry12, 114 L6b VISp Crh, 110 Lamp5 Krt73, 109 Lamp5 Fam19a1 Pax6, 108 Lamp5 Fam19a1 Tmem182, 106 Lamp5 Ntn1 Npy2r, 105 Lamp5 Plch2 Dock5, 101 Lamp5 Lsp1, 100 Lamp5 Lhx6, 97 Sncg Slc17a8, 96 Sncg Vip Nptx2, 95 Sncg Gpr50, 93 Sncg Vip Itih5, 90 Serpinf1 Clrn1, 89 Serpinf1 Agp5 Vip, 85 Vip Igfbp6 Car10, 84 Vip Igfbp6 Pltp, 82 Vip Lmo1 Fam159b, 81 Vip Lmo1 Myl1, 79 Vip Igfbp4 Mab21l1, 78 Vip Arhgap36 Hmcn1, 77 Vip Gpc3 Slc18a3, 74 Vip Ptprt Pkp2, 73 Vip Rspo4 Rxfp1 Chat, 71 Vip Lect Oxtr, 70 Vip Rspo1 Itga4, 67 Vip Chat Htr1f, 66 Vip Pygm C1ql1, 61 Vip Crispld2 Htr2c, 60 Vip Crispld2 Kcne4, 58 Vip Col15a1 Pde1a, 54 Sst Chodl, 53 Sst Mme Fam114a1, 52 Sst Tac1 Htr1d, 50 Sst Tac1 Tacr3, 49 Sst Calb2 Necab1, 48 Sst Calb2 Pdlim5, 46 Sst Nr2f2 Necab1, 45 Sst Myh8 Etv1, 44 Sst Chrna2 Glra3, 42 Sst Myh8 Fibin, 40 Sst Chrna2 Ptgdr, 37 Sst Tac2 Myh4, 36 Sst Hpse Sema3c, 34 Sst Crhr2 Efemp1, 33 Sst Crh 4930553C11Rik, 31 Sst Esm1, 29 Sst Tac2 Tacstd2, 28 Sst Rxfp1 Eya1, 27 Sst Rxfp1 Prdm8, 23 Sst Nts, 21 Pvalb Gabrg1, 20 Pvalb Th Sst, 18 Pvalb Calb1 Sst, 17 Pvalb Akr1c18 Ntf3, 16 Pvalb Sema3e Kank4, 14 Pvalb Gpr149 Islr, 11 Pvalb Reln Itm2a, 10 Pvalb Reln Tac1, 9 Pvalb Tpbg, 4 Pvalb Vipr2, 1 Meis2 Adamts19, 170 Astro Aqp4, 171 OPC Pdgfra Grm5, 173 Oligo Serpinb1a, 174 Oligo Synpr,

175 VLMC Osr1 Cd74, 176 VLMC Osr1 Mc5r, 177 VLMC Spp1 Col15a1, 178 Peri Kcnj8, 179 SMC Acta2, 180 Endo Ctla2a, and 181 Microglia Siglech.

[0013] FIGs. 4A, 4B. CN2320 (eHGT_475m) in rat neocortex. (4A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the neocortex 18 days after intracerebroventricular injection of 1.5E11 viral genome copies of AAV vector #CN2320 (eHGT_475m) in 1-day old rat pup. Scale bar: 100 microns. (4B) Higher magnification view showing a cell body and signature chandelier cell axon cartridges. Scale bar: 25 microns.

[0014] FIGs. 5A, 5B. CN2320 (eHGT_475m) in macaque frontal cortex. (5A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the superior frontal cortex region 64 days after stereotaxic injection of 8.0E10 viral genome copies of AAV vector #CN2320 (eHGT_475m) in adult macaque in vivo. Scale bar: 100 microns. (5B) Higher magnification view showing a cell body and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0015] FIGs. 6A-6C. CN2320 (eHGT_475m) in macaque frontal cortex. (6A) Mapping of single cell transcriptomic profiles of SYPF2+ cells recorded using the Patch-seq technique (patch-clamp electrophysiology recording with nucleus extraction and post-hoc single nucleus RNA sequencing) from the frontal cortex region of the macaque cortex after in vivo stereotaxic injection of CN2320 virus packaged with the PHP.eB capsid. Experiments were conducted for two injected macaque monkeys and pooled. Number of cells mapped to the final leaf are shown on the bar plot below the dendrogram. Transcriptomic cell types of the human middle temporal gyrus are shown on the bottom. This data shows eHGT_475m enhancer-driven reporter expression is highly specific to primate chandelier cells (PVALB SCUBE3 type) in neocortex. (6B) Exemplary Patch-seq recorded neuron showing individual cell RNA-seq mapping with high confidence to the chandelier cell type PVALB SCUBE3, as well as (6C) the signature firing pattern in response to current injection step (top) and the cell morphology from biocytin fill and reaction (bottom). Scale bar in FIG. 6C (bottom): 20 microns. The text at the bottom of FIGs. 6A and 6B read (from left to right): 4 Inh L1-2 PAX6 CH12, 5 Inh L1-2 PAX6 TNFAIP8L3, 7 Inh L1-2 SST NMBR (ADARB2+), 10 Inh L1-4 LAMP5 LCP2 (rosehip), 11 Inh L1-2 LAMP5 DBP, 12 Inh L2-6 LAMP5 CA1 (Igtp), 18 Inh L1 SST CHRNA4 (ADARB2+), 19 Inh L1-2 GAD1 MC4R (ADARB2+), 20 Inh L1-2 SST BAGE2 (ADARB2+), 22 Inh L1-3 PAX6 SYT6 (Sncg), 24 Inh L1-2 VIP TSPAN12, 25 Inh L1-4 VIP CHRNA6, 27 Inh L1-3 VIP ADAMTSL1, 28 Inh L1-4 VIP PENK, 33 Inh L2-6 VIP QPCT, 34 Inh L3-6 VIP HS3ST3A1, 35 Inh L1-2 VIP PCDH20, 36 Inh L2-5 VIP SERPINF1, 37 Inh L2-5 VIP TYR, 41 Inh L1-3 VIP CHRM2, 42 Inh L2-4 VIP CBLN1, 43 Inh L1-3 VIP CCDC184, 44 Inh L1-3 VIP GGH, 47 Inh L1-2 VIP LHB, 48 Inh L2-3 VIPCASC6, 50 Inh L2-4 VIP SPAG17, 51 Inh L1-4 VIP OPRM1, 54 Inh L3-6 SST NPY (Chodl), 59 Inh L3-6 SST HPGD, 63 Inh L4-6 SST B3GAT2,

64 Inh L5-6 SST KLHDC8A, 66 Inh L5-6 SST NPM1P10, 67 Inh L4-6 SST GXYLT2, 68 Inh L4-5 SST STK32A, 71 Inh L1-3 SST CALB1, 72 Inh L3-5 SST ADGRG6, 73 Inh L2-4 SST FRZB, 74 Inh L5-6 SST TH, 76 Inh L5-6 GAD1 GLP1R (LHX6+), 78 Inh L5-6 PVALB LGR5, 81 Inh L4-5 PVALB MEPE, 83 Inh L2-4 PVALB WFDC2, 84 Inh L4-6 PVALB SULF1, 85 Inh L5-6 SST MIR548F2, 86 Inh L2-5 PVALB SCUBE3 (chandelier), 93 Exc L2 LAMP5 LTK, 94 Exc L2-4 LINC00507 GLP2R, 95 Exc L2-3 LINC00507 FREM3, 96 Exc L5-6 THEMIS C1QL3, 99 Exc L3-4 RORB CARM1P1, 101 Exc L3-5 RORB ESR1, 103 Exc L3-5 RORB COL22A1, 105 Exc L3-5 RORB FILIP1L, 106 Exc L3-5 RORB TWIST2, 109 Exc L4-5 RORB FOLH1B, 111 Exc L4-6 RORB SEMA3E, 112 Exc L4-5 RORB DAPK2, 114 Exc L5-6 RORB TTC12, 115 Exc L4-6 RORB C1R, 116 Exc L4-5 FEZF2 SCN4B (PT), 118 Exc L5-6 THEMIS DCSTAMP, 120 Exc L5-6 THEMIS CRABP1, 121 Exc L5-6 THEMIS FGF10, 123 Exc L4-6 FEZF2 IL26 (NP), 125 Exc L5-6 FEZF2 ABO, 129 Exc L6 FEZF2 SCUBE1, 130 Exc L5-6 SLC17A7 IL15, 131 Exc L6 FEZF2 OR2T8, and 132 Exc L5-6 FEZF2 EFTUD1P1.

[0016] FIGs. 7A, 7B. CN2321 (eHGT_476m) in mouse neocortex. (7A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the neocortex 42 days after retro-orbital delivery of 5.48E11 viral genome copies of AAV vector #CN2321 (eHGT_476m). Scale bar: 100 microns. (7B) Higher magnification view showing a sparse cell bodies and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0017] FIG. 8. CN2321 (eHGT_476m) in 569787 animal VISp. Mapping of single cell transcriptomic profiles of SYPF2+ cells sorted from the VISp region of the mouse neocortex after retro-orbital injection of CN2321 virus packaged with the PHP.eB capsid. Number of cells mapped to the final leaf are shown on the bar plot below the dendrogram. Transcriptomic cell types are shown on the bottom. This data shows eHGT_476m enhancer-driven reporter expression is enriched in chandelier cells (Pvalb Vipr2 cells) and Pvalb Tpbg cells when mouse VISp is evaluated. The text at the bottom of the figure from left to right reads the same as that described in FIG. 3.

[0018] FIGs. 9A, 9B. CN2321 (eHGT_476m) in macaque frontal cortex. (9A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the superior frontal cortex region 43 days after stereotaxic injection of 9.59E10 viral genome copies of AAV vector #CN2321 (eHGT_476m) in adult macaque in vivo. Scale bar: 200 microns. (9B) Higher magnification view showing a cell body and signature chandelier cell axon cartridges. Scale bar: 100 microns.

[0019] FIGs. 10A, 10B. CN2782 (3xCore-eHGT_476m) in mouse neocortex. (10A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the neocortex 42 days after retro-orbital delivery of 4.48E11 viral genome copies of AAV vector #CN2782

(3xCore-eHGT_476m). Scale bar: 100 microns. (10B) Higher magnification view showing a sparse cell bodies and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0020] FIGs. 11A, 11B. CN2580 (eHGT_476m) in Ai14 mouse neocortex. (11A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the neocortex 29 days after retro-orbital delivery of 1.0E11 viral genome copies of AAV vector #CN2580 (eHGT_476m) into Ai14 Cre-dependent reporter mouse line. Scale bar: 100 microns. (11B) Higher magnification view showing a sparse cell bodies and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0021] FIGs. 12A, 12B. CN2717 (eHGT_710m) in mouse neocortex. (12A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the neocortex 40 days after retro-orbital delivery of 6.0E11 viral genome copies of AAV vector #CN2717 (eHGT_710m). Scale bar: 200 microns. (12B) Higher magnification view showing a sparse cell bodies and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0022] FIGs. 13A, 13B. CN2717 (eHGT_710m) in macaque frontal cortex. (13A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the superior frontal cortex region 43 days after stereotaxic injection of 3.46E11 viral genome copies of AAV vector #CN2717 (eHGT_710m) in adult macaque in vivo. Scale bar: 200 microns. (13B) Higher magnification view showing a cell body and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0023] FIG. 14. Sequences supporting the disclosure, including eHGT_297m (SEQ ID NO: 1); eHGT_303m (SEQ ID NO: 2); eHGT_307m (SEQ ID NO: 3); eHGT_308m (SEQ ID NO: 4); eHGT_472m (SEQ ID NO: 5); eHGT_475m (SEQ ID NO: 6); eHGT_476m (SEQ ID NO: 7); eHGT_476m core (SEQ ID NO: 8); 3x_eHGT_476m core (SEQ ID NO: 9); eHGT_503m (SEQ ID NO: 10); eHGT_571m (SEQ ID NO: 11); eHGT_706m (SEQ ID NO: 12); eHGT_710m (SEQ ID NO: 13); eHGT_296m (SEQ ID NO: 14); eHGT_299m (SEQ ID NO: 15); eHGT_300m (SEQ ID NO: 16); eHGT_306m (SEQ ID NO: 17); eHGT_309m (SEQ ID NO: 18); eHGT_310m (SEQ ID NO: 19); eHGT_890m (SEQ ID NO: 20); eHGT_891m (SEQ ID NO: 21); eHGT_892m (SEQ ID NO: 22); eHGT_476m (SEQ ID NO: 23); eHGT_1022m (SEQ ID NO: 24); eHGT_1023m (SEQ ID NO: 25); eHGT_1024m (SEQ ID NO: 26); Beta-Globin Minimal Promoter (SEQ ID NO: 27); minCMV Promoter (SEQ ID NO: 28); Mutated minCMV Promoter (SEQ ID NO: 29); minRho Promoter (SEQ ID NO: 30); minRho* Promoter (SEQ ID NO: 31); Hsp68 minimal Promoter (SEQ ID NO: 32); SYFP2 (SEQ ID NO: 33); EGFP (SEQ ID NO: 34); Optimized Flp recombinase (SEQ ID NO: 35); Improved Cre recombinase (iCre) (SEQ ID NO: 36); iCre(R297T) (SEQ ID NO: 37); CreN-inteinN (SEQ ID NO: 38); inteinC-CreC (SEQ ID NO: 39); SP10 insulator (SP10ins) (SEQ ID NO: 40); 3xSP10ins (SEQ ID NO: 41); 4X2C (SEQ ID NO: 42); miR128 Recognition Sequence (SEQ ID NO: 136); miR221 Recognition Sequence (SEQ ID NO: 137); 3XFLAG (SEQ ID NO: 43);

10 aa (SEQ ID NO: 44); H2B (SEQ ID NOs: 135 and 45); WPRE3 (SEQ ID NO: 46); WPRE (SEQ ID NO: 47); BGHpA (SEQ ID NO: 48); HGHpA (SEQ ID NO: 49); P2A (SEQ ID NOs: 50 and 51); T2A (SEQ ID NO: 52); E2A (SEQ ID NO: 53); F2A (SEQ ID NO: 54); Exemplary Plasmid Backbone 1 – Left ITR (SEQ ID NO: 55); Exemplary Plasmid Backbone 1 – Right ITR (SEQ ID NO: 56); Exemplary Plasmid Backbone 2 – Left ITR (SEQ ID NO: 57); Exemplary Plasmid Backbone 2 – Right ITR (SEQ ID NO: 58); PHP.eB capsid (SEQ ID NO: 59); AAV9 VP1 capsid protein (SEQ ID NO: 60); tet-Transactivator version 2 (tTA2) (SEQ ID NO: 61); GTPase HRas [Homo sapiens] (SEQ ID NO: 62); Substance P is position 58-68 of Protachykinin-1 [Homo sapiens] (SEQ ID NO: 63); Oxytocin is position 20-28 of Oxytocin-neurophysin 1 [Homo sapiens] (SEQ ID NO: 64); GCaMP6m (SEQ ID NO: 65); GCaMP6s (SEQ ID NO: 66); GCaMP6f (SEQ ID NO: 67); CN1917 (SEQ ID NO: 68); CN2047 (SEQ ID NO: 69); CN2048 (SEQ ID NO: 70); CN2049 (SEQ ID NO: 71); CN2427 (SEQ ID NO: 72); CN2320 (SEQ ID NO: 73); CN2321 (SEQ ID NO: 74); CN2719 (SEQ ID NO: 75); CN2707 (SEQ ID NO: 76); CN2713 (SEQ ID NO: 77); CN2717 (SEQ ID NO: 78); AiP1104 (SEQ ID NO: 79); AiP1089 (SEQ ID NO: 80); AiP1105 (SEQ ID NO: 81); AiP1090 (SEQ ID NO: 82); AiP1106 (SEQ ID NO: 83); AiP1091 (SEQ ID NO: 84); AiP1092 (SEQ ID NO: 85); CN2581 (SEQ ID NO: 86); CN2782 (SEQ ID NO: 87); CN3407 (SEQ ID NO: 88); CN3408 (SEQ ID NO: 89); CN3409 (SEQ ID NO: 90); CN2580 (SEQ ID NO: 91); CN2825 (SEQ ID NO: 92); CN3270 (SEQ ID NO: 93); CN3316 (SEQ ID NO: 94); CN3271 (SEQ ID NO: 95); CN3793 (SEQ ID NO: 96); CN3794 (SEQ ID NO: 97); CN3795 (SEQ ID NO: 98); CN3790 (SEQ ID NO: 99); CN3751 (SEQ ID NO: 100); and CN3752 (SEQ ID NO: 101).

DETAILED DESCRIPTION

[0024] 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 double or 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 or non-human primates.

[0025] Chandelier cells are a subtype of GABAergic interneurons that were discovered in the 1970s and that have been implicated in disorders such as epilepsy and schizophrenia. Chandelier

cells (also referred to as axo-axonic cells) selectively innervate pyramidal neurons at the axon initial segment where action potentials are generated. This innervation occurs in cortical layers 2, 3, 5a, and 5b. Chandelier cells have a distinct morphology resembling candlesticks on a chandelier due to axonal arborization patterns. These cells express the markers *Cpne5* and *Vipr2* in mouse, and *NOG* in human. Putative conserved cross-species marker genes for mouse, monkey, and human chandelier cells include UNC5B and C1QL1. For more information regarding chandelier cells, see Wang, et al., Chandelier Cells in Functional and Dysfunctional Neural Circuits, Frontiers in Neural Circuits, doi:10.3389/fncir.2016.00044. While significant information has been learned about this cell type, many questions regarding their function and brain-wide effects remain.

[0026] The current disclosure provides artificial expression constructs that drive gene expression in chandelier cells. Particular embodiments of the artificial expression constructs utilize the enhancers eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, and eHGT_1024m to drive gene expression within chandelier cells. Additional embodiments utilize the enhancer eHGT_503m to drive gene expression in chandelier cells and VIP cells. Additional embodiments utilize enhancer eHGT_710m to drive gene expression in chandelier cells, glutamatergic neurons in the thalamus, and molecular layer GABAergic interneurons in the cerebellum.

[0027] Particular embodiments provide artificial expression constructs including the features of vectors described herein including vectors: CN1917, CN2047, CN2048, CN2049, CN2427, CN2320, CN2321, CN2719, CN2707, CN2713, CN2717, AiP1104, AiP1089, AiP1105, AiP1090, AiP1106, AiP1091, AiP1092, CN2581, CN2782, CN3407, CN3408, CN3409, CN2580, CN2825, CN3270, CN3316, CN3271, CN3793, CN3794, CN3795, CN3790, CN3751, and CN3752.

[0028] 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. These headings are provided for organization purposes only and do not limit the scope or interpretation of the disclosure.

[0029] (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.

[0030] 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_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, and eHGT_503m.

[0031] 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 selective 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.

[0032] 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. 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 an intended cell type; less than 40% expression in a reference cell type as compared to an intended cell type; less than 30% expression in a reference cell type as compared to an intended cell type; less than 20% expression in a reference cell type as compared to an intended cell type; or less than 10% expression in a reference cell type as compared to an intended 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-targeted 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 targeted 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.
- *Sst* and *Pvalb* GABAergic neurons: Developmentally derived from neuronal progenitors in the medial ganglionic eminence (MGE).
- *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 *C1ql1*, *Cpne5, Unc5b* and *Vipr2* in mouse, and *NOG, C1QL1,* and *UNC5B* in human. Notably, *Pvalb* expression is not always detected in all chandelier cells in the adult brain.
- 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.

[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. Included in this subclass are L6 IT Car3 cells, which are highly similar to intracortical-projecting cells in the claustrum.
- 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] Cerebellar 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.

[0038] Deep cerebellar nucleus neurons: neurons located in the deep cerebellar nuclei structures. These include glutamatergic and GABAergic cells that express the gene Pvalb.

[0039] 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 biproduct of brain dissection methods. Microglia are known to express Cx3cr1, Tmem119, and PTPRC (CD45).

[0040] 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.

[0041] 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, AmCyanl, Midoriishi-Cyan, mTurquoise); green fluorescent proteins (*e.g.* GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green (mAzamigreen), CopGFP, AceGFP, avGFP, ZsGreenl, 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, HcRedl, 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, ZsYellowl); and tandem conjugates.

[0042] 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 9.13X10-21 m^2/molecule, also quoted as 55,000 L/(mol•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.

[0043] 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.

[0044] 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).

[0045] 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 (Ca2+), 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)).

[0046] 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 (CCAA-T/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 (TGFβ), epidermal growth factor (EGF), GTPase and HRas; and neurotransmitters such as cocaine and amphetamine regulated transcript, substance P, oxytocin, and somatostatin.

[0047] 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 permutated 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, and jGCaMP7f. 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).

[0048] 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).

[0049] 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 and Knopfel et al., *The Journal of*

Neuroscience, 2010, 30(45): 14998-15004).

[0050] In particular embodiments, functional molecules include DNA and RNA editing tools such 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.

[0051] 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), TGFβ (e.g., GenBank: CAF02096.2), 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., *UniProt*KB - 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).

[0052] In particular embodiments, a functional molecule includes intein-mediated protein splicing elements. Intein-mediated protein splicing is a process whereby an intein catalyzes its removal from a protein precursor, permitting synthesis of a mature, active protein. When a pair of split inteins are involved in the splicing process, the mature and active protein is formed from two separate protein precursors. An intein is an in-frame intervening sequence in a protein precursor. The intein disrupts the coding region of a gene, until it catalyzes its own excision from the protein precursor through a post-translational protein splicing process to yield the free intein and a mature protein. A “split intein” is made up of two distinct polypeptides or proteins, referred to as the “N-terminal” or inteinN and the “C-terminal” or inteinC because of their homology to the N-terminal and C-terminal regions of non-split inteins, respectively. Together inteinN and inteinC polypeptides, when operably linked to foreign polypeptides, possess all necessary functionality to complete a trans-protein splicing reaction, whereby the two foreign “extein” fragments are ligated together by formation of a peptide bond. DNA sequences encoding inteinN and inteinC may be separated by many kilobases of nucleotides in a genome or on different chromosomes. In particular embodiments, the extein is a split Cre gene. In particular embodiments, intein-mediated protein splicing elements include CreN-inteinN and inteinC-CreC.

[0053] Additional effector elements include Cre, iCre, dgCre, FlpO, and tTA2. iCre refers to a codon-improved Cre. iCre includes the sequence as set forth in SEQ ID NOs: 36 and 37. 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.

[0054] 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.

[0055] In particular embodiments, artificial expression constructs include tag cassettes. Exemplary tag cassettes encode a His tag (HHHHHH; SEQ ID NO: 123), Flag tag (DYKDDDDK; SEQ ID NO: 124), Xpress tag (DLYDDDDK; SEQ ID NO: 125), Avi tag (GLNDIFEAQKIEWHE; SEQ ID NO: 126), Calmodulin tag (KRRWKKNFIAVSAANRFKKISSSGAL; SEQ ID NO: 127), Polyglutamate tag, HA tag (YPYDVPDYA; SEQ ID NO: 128), Myc tag (EQKLISEEDL; SEQ ID NO: 129), Strep tag (which refers the original STREP® tag (WRHPQFGG; SEQ ID NO: 130), STREP® tag II (WSHPQFEK SEQ ID NO: 131(IBA Institut fur Bioanalytik, Germany); *see*, *e.g.,* US 7,981,632), Softag 1 (SLAELLNAGLGGS; SEQ ID NO: 132), Softag 3 (TQDPSRVG; SEQ ID NO: 133), and V5 tag (GKPIPNPLLGLDST; SEQ ID NO: 134). In particular embodiments, artificial expression constructs include three tandem FLAG tag epitopes including 3XFLAG (SEQ ID NO: 43). In particular embodiments, artificial expression constructs include miRNA-guided neuron tags (mAGNETs). Exemplary mAGNETs include 4X2C, 4X3C, and 8X2C. In particular embodiments, artificial expression constructs include 4X2C (SEQ ID NO: 42). In particular embodiments artificial expression constructs includes 10 aa (SEQ ID NO: 44).

[0056] 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 (e.g. IRES2) allow ribosomes to initiate translation at a second internal site on a mRNA molecule, leading to production of two proteins from one mRNA.

[0057] 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.

[0058] 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.

[0059] 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, 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.

[0060] 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.

[0061] 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.

[0062] 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.

[0063] 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.

[0064] 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.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 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.

[0069] 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.

[0070] 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.

[0071] 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.

[0072] "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.

[0073] 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.

[0074] 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.

[0075] 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.

[0076] 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.

[0077] 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.

[0078] 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).

[0079] Artificial expression constructs can encode linker segments between other components.

A particular example of a protein linker sequence includes SGLRSGGSGG (SEQ ID NO: 102). Examples of Gly-Ser linkers include sets of glycine and serine repeats such as from one to ten repeats of $(Gly_xSer_y)_n$, wherein x and y are independently an integer from 0 to 10 provided that x and y are not both 0 and wherein n is an integer of 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10). Particular examples include $(Gly_4Ser)_n$ (SEQ ID NO: 103), $(Gly_3Ser)_n(Gly_4Ser)_n$ (SEQ ID NO: 104), $(Gly_3Ser)_n(Gly_2Ser)_n$ (SEQ ID NO: 105), and $(Gly_3Ser)_n(Gly_4Ser)_1$ (SEQ ID NO: 106). In particular embodiments, the linker is $(Gly_4Ser)_4$ (SEQ ID NO: 107), $(Gly_4Ser)_3$ (SEQ ID NO: 108), $(Gly_4Ser)_2$ (SEQ ID NO: 109), $(Gly_4Ser)_1$ (SEQ ID NO: 110), $(Gly_3Ser)_2$ (SEQ ID NO: 111), $(Gly_3Ser)_1$ (SEQ ID NO: 112), $(Gly_2Ser)_2$ (SEQ ID NO: 113) or $(Gly_2Ser)_1$, GGSGGGSGGSG (SEQ ID NO: 114), GGSGGGSGSG (SEQ ID NO: 115), and GGSGGGSG (SEQ ID NO: 116).

[0080] Artificial expression constructs can encode nuclear localization proteins, such as Histone H1, Histone H2A, Histone H2B, Histone H3, Histone H4, histone-like protein HPhA.

[0081] 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.

[0082] Particular embodiments of vectors disclosed herein include:

Expression Construct Name	Expression Construct Features
CN1917	rAAV-3xSP10ins-eHGT_297m-minRho*-SYFP2-WPRE3-BGHpA
CN2047	rAAV-3xSP10ins-eHGT_303m-minRho*-SYFP2-WPRE3-BGHpA
CN2048	rAAV-3xSP10ins-eHGT_307m-minRho*-SYFP2-WPRE3-BGHpA
CN2049	rAAV-3xSP10ins-eHGT_308m-minRho*-SYFP2-WPRE3-BGHpA
CN2427	rAAV-eHGT_472m-minBglobin-SYFP2-WPRE3-BGHpA
CN2320	rAAV-eHGT_475m-minBglobin-SYFP2-WPRE3-BGHpA
CN2321	rAAV-eHGT_476m-minBglobin-SYFP2-WPRE3-BGHpA
CN2719	rAAV-eHGT_503m-minBglobin-SYFP2-WPRE3-BGHpA
CN2707	rAAV-eHGT_571m-minBglobin-SYFP2-WPRE3-BGHpA
CN2713	rAAV-eHGT_706m-minBglobin-SYFP2-WPRE3-BGHpA
CN2717	rAAV-eHGT_710m-minBglobin-SYFP2-WPRE3-BGHpA
AiP1104	rAAV-eHGT_296m-minBglobin-FlpO-WPRE-HGHpA

AiP1089	rAAV-eHGT_297m-minBglobin-FlpO-WPRE-HGHpA
AiP1105	rAAV-eHGT_299m-minBglobin-FlpO-WPRE-HGHpA
AiP1090	rAAV-eHGT_300m-minBglobin-FlpO-WPRE-HGHpA
AiP1106	rAAV-eHGT_306m-minBglobin-FlpO-WPRE-HGHpA
AiP1091	rAAV-eHGT_309m-minBglobin-FlpO-WPRE-HGHpA
AiP1092	rAAV-eHGT_310m-minBglobin-FlpO-WPRE-HGHpA
CN2581	rAAV-eHGT_476m-minRho*-SYFP2-10aa-H2B-WPRE3-BGHpA
CN2782	pAAV-eHGT_3x_eHGT_476m-minBGlobin-SYFP2-WPRE3-BGHpA
CN3407	pAAV-eHGT_890m-minBGlobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN3408	pAAV-eHGT_891m-minBGlobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN3409	pAAV-eHGT_892m-minBGlobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN2580	rAAV-eHGT_476m-minBGlobin-iCre-4X2C-WPRE3-BGHpA
CN2825	rAAV-eHGT_476m-minBGlobin-FlpO-4X2C-WPRE3-BGHpA
CN3270	rAAV-eHGT_475m-minBglobin-SYFP2-4X2C-WPRE3-BGHpA
CN3316	rAAV-eHGT_710m-minBglobin-SYFP2-4X2C-WPRE3-BGHpA
CN3271	pAAV-eHGT_3x_eHGT_476m- minBGlobin -SYFP2-4X2C-WPRE3-BGHpA
CN3793	pAAV-eHGT_1022m-minBGlobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN3794	pAAV-eHGT_1023m- minBGlobin -SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN3795	pAAV-eHGT_1024m- minBGlobin -SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN3790	rAAV-eHGT_475m- minBGlobin -iCre(R297T)-BGHpA
CN3751	rAAV-eHGT_475m- minBGlobin -CreN-inteinN-WPRE3-BGHpA
CN3752	rAAV-eHGT_476m- minBGlobin -inteinC-CreC-WPRE3-BGHpA

[0083] 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. 3). 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.

[0084] 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: 117) are changed to S-DGTLAVPFK-A (SEQ ID NO: 118). In particular embodiments, PHP.eb refers to SEQ ID NO: 59.

[0085] 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).

[0086] 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.

[0087] 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.

[0088] 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.

[0089] AAV-BR1 is an AAV2 variant displaying the NRGTEWD (SEQ ID NO: 119) 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).

[0090] 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: 120), transduces neurons in the enteric nervous system, and strongly transduces peripheral sensory afferents entering the spinal cord and brain stem.

[0091] 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: 121). It transfers genes

throughout the CNS with higher efficiency than AAV9 and transduces the majority of astrocytes and neurons across multiple CNS regions.

[0092] AAV-PPS, an AAV2 variant crated by insertion of the DSPAHPS (SEQ ID NO: 122) epitope into the capsid of AAV2, shows a dramatically improved brain tropism relative to AAV2.

[0093] For additional information regarding capsids that cross the blood brain barrier, see Chan et al., Nat. Neurosci. 2017 Aug: 20(8): 1172-1179.

[0094] (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, rat), or human. Physiologically active components within compositions described herein can be prepared in neutral forms, as freebases, or as pharmacologically acceptable salts.

[0095] 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.

[0096] 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.

[0097] 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).

[0098] 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.

[0099] Compositions may include liposomes, lipids, lipid complexes, microspheres, microparticles, nanospheres, and/or nanoparticles.

[0100] 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).

[0101] 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 polyalkyl-cyanoacrylate 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.

[0102] 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.

[0103] 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.

[0104] 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.

[0105] 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.

[0106] 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.

[0107] 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).

[0108] Supplementary active ingredients can also be incorporated into the compositions.

[0109] 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.

[0110] 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.

[0111] (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.

[0112] 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_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, eHGT_503m, CN1917, CN2047, CN2048, CN2049, CN2427, CN2320, CN2321, CN2719, CN2707, CN2713, CN2717, AiP1104, AiP1089, AiP1105, AiP1090, AiP1106, AiP1091, AiP1092, CN2581, CN2782, CN3407, CN3408, CN3409, CN2580, CN2825, CN3270, CN3316, CN3271, CN3793, CN3794, CN3795, CN3790, CN3751, and/or CN3752, 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 or mice. 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.

[0113] 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.

[0114] 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.

[0115] 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).

[0116] 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

[0117] 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).

[0118] 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_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, and/or eHGT_503m.

[0119] 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).

[0120] (iv) Transgenic Animals. Another aspect of the disclosure includes transgenic animals, the genome of which contains an artificial expression construct including eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, and/or eHGT_503m operatively linked to a heterologous coding sequence. In particular embodiments, the genome of

a transgenic animal includes CN1917, CN2047, CN2048, CN2049, CN2427, CN2320, CN2321, CN2719, CN2707, CN2713, CN2717, AiP1104, AiP1089, AiP1105, AiP1090, AiP1106, AiP1091, AiP1092, CN2581, CN2782, CN3407, CN3408, CN3409, CN2580, CN2825, CN3270, CN3316, CN3271, CN3793, CN3794, CN3795, CN3790, CN3751, and/or CN3752.

[0121] In particular embodiments, when a non-integrating vector is utilized, a transgenic animal includes an artificial expression construct including eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, eHGT_503m, CN1917, CN2047, CN2048, CN2049, CN2427, CN2320, CN2321, CN2719, CN2707, CN2713, CN2717, AiP1104, AiP1089, AiP1105, AiP1090, AiP1106, AiP1091, AiP1092, CN2581, CN2782, CN3407, CN3408, CN3409, CN2580, CN2825, CN3270, CN3316, CN3271, CN3793, CN3794, CN3795, CN3790, CN3751, and/or CN3752within 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_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, eHGT_503m, CN1917, CN2047, CN2048, CN2049, CN2427, CN2320, CN2321, CN2719, CN2707, CN2713, CN2717, AiP1104, AiP1089, AiP1105, AiP1090, AiP1106, AiP1091, AiP1092, CN2581, CN2782, CN3407, CN3408, CN3409, CN2580, CN2825, CN3270, CN3316, CN3271, CN3793, CN3794, CN3795, CN3790, CN3751, and/or CN3752as 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 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 concatemer including a core of eHGT_476m.

2. The concatemer of embodiment 1, wherein the core of eHGT_476m includes the sequence as set forth in SEQ ID NO: 8.

3. The concatemer of embodiment 1 or 2, wherein the concatemer includes 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core of eHGT_476m.

4. The concatemer of any of embodiments 1-3, wherein the concatemer includes the sequence as set forth in SEQ ID NO: 9.

5. An artificial expression construct including (i) an enhancer including eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, and/or eHGT_503m; (ii) a promoter; and (iii) a heterologous encoding sequence.

6. The artificial expression construct of embodiment 5, wherein the heterologous encoding sequence encodes an effector element or an expressible element.

7. The artificial expression construct of embodiment 6, wherein the effector element includes a reporter protein or a functional molecule.

8. The artificial expression construct of embodiment 7, wherein the reporter protein includes a fluorescent protein.

9. The artificial expression construct of embodiment 7 or 8, 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).

10. The artificial expression construct of embodiment 6, wherein the expressible element includes a non-functional molecule.

11. The artificial expression construct of embodiment 10, 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 a DREADD.

12. The artificial expression construct of any of embodiments 5-11, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.

13. The artificial expression construct of embodiment 8, wherein the capsid includes PHP.eB, AAV-BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.

14. The artificial expression construct of any of embodiments 5-13, wherein the artificial expression construct includes or encodes a skipping element.

15. The artificial expression construct of embodiment 14, wherein the skipping element includes a 2A peptide and/or an internal ribosome entry site (IRES).

16. The artificial expression construct of embodiment 15, wherein the 2A peptide includes T2A, P2A, E2A, or F2A.

17. The artificial expression construct of any of embodiments 5-16, wherein the artificial expression construct encodes a linker.

18. The artificial expression construct of embodiment 17, wherein the linker has the sequence: set forth in SEQ ID NO: 102.

19. The artificial expression construct of embodiment 17, wherein the linker includes a Gly-Ser linker.

20. The artificial expression construct of any of embodiments 5-19, wherein the artificial expression construct encodes a nuclear localization protein.

21. The artificial expression construct of embodiment 20, wherein the nuclear localization protein includes Histone H1, Histone H2A, Histone H2B, Histone H3, Histone H4, and/or histone-like protein HPhA

22. The artificial expression construct of any of embodiments 5-21, wherein the artificial expression construct includes or encodes a set of features including: hsA2, eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, eHGT_503m, AAV, scAAV, rAAV, pAAV, minBglobin, CMV, minCMV, minRho, minRho*, fluorescent protein (e.g., EGFP, SYFP, GFP), Cre, iCre, dgCre, CreN-inteinN, inteinC-CreC, FlpO, tTA2, 3XFLAG, Histone H1, Histone H2A, Histone H2B, Histone H3, Histone H4, histone-like protein HPhA, a linker, SP10 insulator (e.g., 3xSP10ins), 10 amino acids (10 aa), 4X2C, P2A, WPRE, WPRE3, HGHpA, and/or BGHpA.

23. The artificial expression construct of any of embodiments 5-22, wherein the artificial expression construct includes or encodes a set of features including:

3xSP10ins-eHGT_297m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_303m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_307m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_308m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_472m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_475m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_476m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_503m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_571m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_706m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_710m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_296m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_297m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_299m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_300m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_306m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_309m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_310m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_476m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_3x_eHGT_476m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_890m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_891m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_892m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_1022m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_1023m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA ;

eHGT_1024m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_475m-minBGlobin-[heterologous encoding sequence]-BGHpA;

3xSP10ins-eHGT_297m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_303m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_307m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_308m-[minimal promoter]*-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_472m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_475m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_476m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_503m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_571m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_706m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_710m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_296m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_297m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_299m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_300m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_306m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_309m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_310m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_3x_eHGT_476m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_890m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_891m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_892m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_1022m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_1023m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA ;

eHGT_1024m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA; or

eHGT_475m-[minimal promoter]-[heterologous encoding sequence]-BGHpA.

24. The artificial expression construct of embodiment 23, wherein the heterologous encoding sequence encodes a reporter protein.

25. The artificial expression construct of embodiment 24, wherein the heterologous encoding sequence further encodes a linker.

26. The artificial expression construct of embodiment 24 or 25, wherein the heterologous encoding sequence further encodes a nuclear localization protein.

27. A vector including an artificial expression construct of any of embodiments 5-26.

28. The vector of embodiment 27, wherein the vector includes a viral vector.

29. The vector of embodiment 27 or 28, wherein the viral vector includes a recombinant adeno-associated viral (AAV) vector.

30. An adeno-associated viral (AAV) vector including at least one heterologous encoding sequence, wherein the heterologous encoding sequence is under control of a promoter and an enhancer including eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, and/or eHGT_503m.

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

32. The transgenic cell of embodiment 31, wherein the transgenic cell is a chandelier cell and optionally a vasoactive intestinal peptide (Vip) cell, a glutamatergic neuron in the thalamus, or a molecular layer GABAergic interneuron in the cerebellum.

33. A non-human transgenic animal including an artificial expression construct, vector, or transgenic cell of any of the preceding embodiments.

34. The non-human transgenic animal of embodiment 33, wherein the non-human transgenic animal is a mouse, a rat, or a non-human primate.

35. An administrable composition including an artificial expression construct, vector, or transgenic cell of any of the preceding embodiments.

36. A kit including an artificial expression construct, vector, transgenic cell, transgenic animal, and/or administrable compositions of any of the preceding embodiments.

37. A method for expressing a heterologous gene within a targeted population of cells in vivo or in vitro, the method including providing the administrable composition of embodiment 35 in a sufficient dosage and for a sufficient time to a sample or subject including the targeted population of cells thereby expressing the gene within the targeted population of cells.

38. The method of embodiment 37, wherein the heterologous gene encodes an effector element or an expressible element.

39. The method of embodiment 38, wherein the effector element includes a reporter protein or a functional molecule.

40. The method of embodiment 39, wherein the reporter protein includes a fluorescent protein.

41. The method of embodiment 39, 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 DREADD.

42. The method of embodiment 39, wherein the expressible element includes a non-functional molecule.

43. The method of embodiment 42, 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.

44. The method of any of embodiments 37-43, wherein the targeted population of cells includes a chandelier cell.

45. The method of any of embodiments 37-44, wherein the targeted population of cells includes a chandelier cell and optionally an additional cell type.

46. The method of embodiment 45, wherein the additional cell type includes a VIP cell.

47. The method of embodiment 45, wherein the additional cell type includes a glutamatergic neuron in the thalamus and a molecular layer GABAergic interneuron in the cerebellum.

48. The method of any of embodiments 37-44, wherein the targeted population of cells includes a selective population of cells including chandelier cells.

49. The method of any of embodiments 37 - 48, wherein the providing includes pipetting.

50. The method of embodiment 49, wherein the pipetting is to a brain slice.

51. The method of embodiment 50, wherein the brain slice includes a chandelier cell.

52. The method of embodiments 50 or 51, wherein the brain slice includes a VIP cell.

53. The method of any embodiments 50-52, wherein the brain slice includes a glutamatergic neuron in the thalamus.

54. The method of any of embodiments 50-53, wherein the brain slice includes a molecular layer GABAergic interneuron in the cerebellum.

55. The method of any of embodiments 50-54, wherein the brain slice is murine, human, or non-human primate.

56. The method of any of embodiments 37-55, wherein the providing includes administering to a living subject.

57. The method of embodiment 56, wherein the living subject is a human, a non-human primate, a rat, or a mouse.

58. The method of embodiments 56 or 57, wherein the administering to a living subject is through injection.

59. The method of embodiment 58, wherein the injection includes intravenous injection, intraparenchymal injection into brain tissue, intracerebroventricular (ICV) injection, intra-cisterna magna (ICM) injection, or intrathecal injection.

60. An artificial expression construct including CN1917, CN2047, CN2048, CN2049, CN2427, CN2320, CN2321, CN2719, CN2707, CN2713, CN2717, AiP1104, AiP1089, AiP1105, AiP1090, AiP1106, AiP1091, AiP1092, CN2581, CN2782, CN3407, CN3408, CN3409, CN2580, CN2825, CN3270, CN3316, CN3271, CN3793, CN3794, CN3795, CN3790, CN3751, or CN3752.

[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±1); glutamate (+3.0±1); Ser (+0.3); Asn (+0.2); Gln (+0.2); Gly (0); Thr (−0.4); Pro (−0.5±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, 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 NaH_2PO_4; 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] 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 chandelier cells using enhancer eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, or a statistically significant reduction in expression in chandelier cells and VIP cells using enhancer eHGT_503m.

[0148] In particular embodiments, artificial means not naturally occurring.

[0149] 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 ±20% of the stated value; ±19% of the stated value; ±18% of the stated value; ±17% of the stated value; ±16% of the stated value; ±15% of the stated value; ±14% of the stated value; ±13% of the stated value; ±12% of the stated value; ±11% of the stated value; ±10% of the stated value; ±9% of the stated value; ±8% of the stated value; ±7% of the stated value; ±6% of the stated value; ±5% of the stated value; ±4% of the stated value; ±3% of the stated value; ±2% of the stated value; or ±1% of the stated value.

[0150] 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.

[0151] 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.

[0152] 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.

[0153] 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.

[0154] 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.

[0155] 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.

[0156] 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.

[0157] 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. An artificial expression construct comprising (i) an enhancer selected from eHGT_475m or eHGT_476m; (ii) a promoter; and (iii) a heterologous encoding sequence.

2. The artificial expression construct of claim 1, wherein the enhancer comprises eHGT_475m and the promoter is minBglobin.

3. The artificial expression construct of claim 1, wherein the enhancer comprises eHGT_475m and the heterologous encoding sequence encodes a fluorescent protein.

4. The artificial expression construct of claim 3, wherein the fluorescent protein comprises SYFP2.

5. The artificial expression construct of claim 1, wherein the enhancer comprises eHGT_476m and the promoter is minBglobin.

6. The artificial expression construct of claim 1, wherein the enhancer comprises eHGT_476m and the heterologous encoding sequence encodes a fluorescent protein.

7. The artificial expression construct of claim 6, wherein the fluorescent protein comprises SYFP2.

8. A concatemer comprising a core of eHGT_476m.

9. The concatemer of claim 8, wherein the core of eHGT_476m has the sequence as set forth in SEQ ID NO: 8.

10. The concatemer of claim 9, wherein the concatemer has 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core of eHGT_476m.

11. The concatemer of claim 8, wherein the concatemer has the sequence as set forth in SEQ ID NO: 9.

12. An artificial expression construct comprising (i) an enhancer comprising the sequence set forth for eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, and/or eHGT_503m; (ii) a promoter; and (iii) a heterologous encoding sequence.

13. The artificial expression construct of claim 12, wherein the heterologous encoding sequence encodes an effector element or an expressible element.

14. The artificial expression construct of claim 13, wherein the effector element comprises a reporter protein or a functional molecule.

15. The artificial expression construct of claim 14, wherein the reporter protein comprises a fluorescent protein.

16. The artificial expression construct of claim 14, 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).

17. The artificial expression construct of claim 13, wherein the expressible element comprises a non-functional molecule.

18. The artificial expression construct of claim 17, 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 a DREADD.

19. The artificial expression construct of claim 12, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.

20. The artificial expression construct of claim 19, wherein the capsid comprises PHP.eB, AAV-BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.

21. The artificial expression construct of claim 12, wherein the artificial expression construct comprises or encodes a skipping element.

22. The artificial expression construct of claim 21, wherein the skipping element comprises a 2A peptide and/or an internal ribosome entry site (IRES).

23. The artificial expression construct of claim 22, wherein the 2A peptide comprises T2A, P2A, E2A, or F2A.

24. The artificial expression construct of claim 12, wherein the artificial expression construct encodes a linker.

25. The artificial expression construct of claim 24, wherein the linker has the sequence set forth in SEQ ID NO: 102.

26. The artificial expression construct of claim 24, wherein the linker comprises a Gly-Ser linker.

27. The artificial expression construct of claim 12, wherein the artificial expression construct encodes a nuclear localization protein.

28. The artificial expression construct of claim 27, wherein the nuclear localization protein comprises Histone H1, Histone H2A, Histone H2B, Histone H3, Histone H4, and/or histone-like

protein HPhA.

29. The artificial expression construct of claim 12, wherein the artificial expression construct comprises or encodes a set of features comprising: eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, eHGT_503m, hsA2, AAV, scAAV, rAAV, pAAV, minBglobin, CMV, minCMV, minRho, minRho*, fluorescent protein, Cre, iCre, dgCre, CreN-inteinN, inteinC-CreC, FlpO, tTA2, 3XFLAG, Histone H1, Histone H2A, Histone H2B, Histone H3, Histone H4, histone-like protein HPhA, a linker, SP10 insulator, 10 amino acid (10 aa), 4X2C, P2A, WPRE, WPRE3, HGHpA, and/or BGHpA.

30. The artificial expression construct of claim 12, wherein the artificial expression construct comprises a set of features comprising:

eHGT_475m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_476m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_476m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_3x_eHGT_476m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_475m- minBGlobin-[heterologous encoding sequence]-BGHpA;

3xSP10ins-eHGT_297m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_303m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_307m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_308m-minRho*-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_472m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_503m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_571m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_706m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_710m-minBglobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_296m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_297m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_299m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_300m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_306m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_309m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_310m-minBglobin-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_890m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_891m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_892m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_1022m-minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_1023m- minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA ;

eHGT_1024m- minBGlobin-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_475m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_476m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_3x_eHGT_476m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_475m-[minimal promoter]-[heterologous encoding sequence]-BGHpA;

3xSP10ins-eHGT_297m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_303m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_307m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_308m-[minimal promoter]*-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_472m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_503m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_571m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_706m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_710m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;

eHGT_296m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_297m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_299m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_300m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_306m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_309m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;

eHGT_310m-[minimal promoter]-[heterologous encoding sequence]-WPRE-HGHpA;
eHGT_890m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;
eHGT_891m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;
eHGT_892m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;
eHGT_1022m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA;
eHGT_1023m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA; or
eHGT_1024m-[minimal promoter]-[heterologous encoding sequence]-WPRE3-BGHpA.

31. The artificial expression construct of claim 30, wherein the heterologous encoding sequence encodes a reporter protein.

32. The artificial expression construct of claim 31, wherein the heterologous encoding sequence further encodes a linker.

33. The artificial expression construct of claim 31, wherein the heterologous encoding sequence further encodes a nuclear localization protein.

34. A vector comprising an artificial expression construct of claim 12.

35. The vector of claim 34, wherein the vector comprises a viral vector.

36. The vector of claim 34, wherein the viral vector comprises a recombinant adeno-associated viral (AAV) vector.

37. An adeno-associated viral (AAV) vector comprising at least one heterologous encoding sequence, wherein the heterologous encoding sequence is under control of a promoter and an enhancer comprising eHGT_475m, eHGT_476m, a core of eHGT_476m, a concatemer of the core of eHGT_476m, eHGT_297m, eHGT_303m, eHGT_307m, eHGT_308m, eHGT_472m, eHGT_571m, eHGT_706m, eHGT_710m, eHGT_296m, eHGT_299m, eHGT_300m, eHGT_306m, eHGT_309m, eHGT_310m, eHGT_890m, eHGT_891m, eHGT_892m, eHGT_1022m, eHGT_1023m, eHGT_1024m, and/or eHGT_503m.

38. A transgenic cell comprising an artificial expression construct of claim 12.

39. The transgenic cell of claim 38, wherein the transgenic cell is a chandelier cell.

40. The transgenic cell of claim 38, wherein the transgenic cell is a vasoactive intestinal peptide (VIP) cell.

41. The transgenic cell of claim 38, wherein the transgenic cell is a glutamatergic neuron in the thalamus.

42. The transgenic cell of claim 38, wherein the transgenic cell is a molecular layer GABAergic interneuron in the cerebellum.

43. A non-human transgenic animal comprising an artificial expression of claim 12.

44. The non-human transgenic animal of claim 43, wherein the non-human transgenic animal is a mouse, a rat, or a non-human primate.

45. An administrable composition comprising an artificial expression construct of claim 12.

46. A kit comprising an artificial expression construct of claim 12.

47. A method for expressing a heterologous gene within a targeted population of cells in vivo or in vitro, the method comprising providing the administrable composition of claim 45 in a sufficient dosage and for a sufficient time to a sample or subject comprising the targeted population of cells thereby expressing the gene within the targeted population of cells.

48. The method of claim 47, wherein the heterologous gene encodes an effector element or an expressible element.

49. The method of claim 48, wherein the effector element comprises a reporter protein or a functional molecule.

50. The method of claim 49, wherein the reporter protein comprises a fluorescent protein.

51. The method of claim 49, 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 DREADD.

52. The method of claim 48, wherein the expressible element comprises a non-functional molecule.

53. The method of claim 52, 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.

54. The method of claim 47, wherein the population of cells comprises a chandelier cell.

55. The method of claim 47, wherein the population of cells comprises a chandelier cell and optionally an additional cell type.

56. The method of claim 55, wherein the additional cell type comprises a VIP cell.

57. The method of claim 55, wherein the additional cell type comprises a glutamatergic neuron in the thalamus and a molecular layer GABAergic interneuron in the cerebellum.

58. The method of claim 47, wherein the targeted population of cells comprises a selective population of cells comprises chandelier cells.

59. The method of claim 47, wherein the providing comprises pipetting.

60. The method of claim 59, wherein the pipetting is to a brain slice.

61. The method of claim 60, wherein the brain slice comprises a chandelier cell.

62. The method of claim 60, wherein the brain slice comprises a VIP cell.

63. The method of claim 60, wherein the brain slice comprises a glutamatergic neuron in the thalamus.

64. The method of claim 60, wherein the brain slice comprises a molecular layer GABAergic interneuron in the cerebellum.

65. The method of claim 60, wherein the brain slice is murine, human, or non-human primate.

66. The method of claim 47, wherein the providing comprises administering to a living subject.

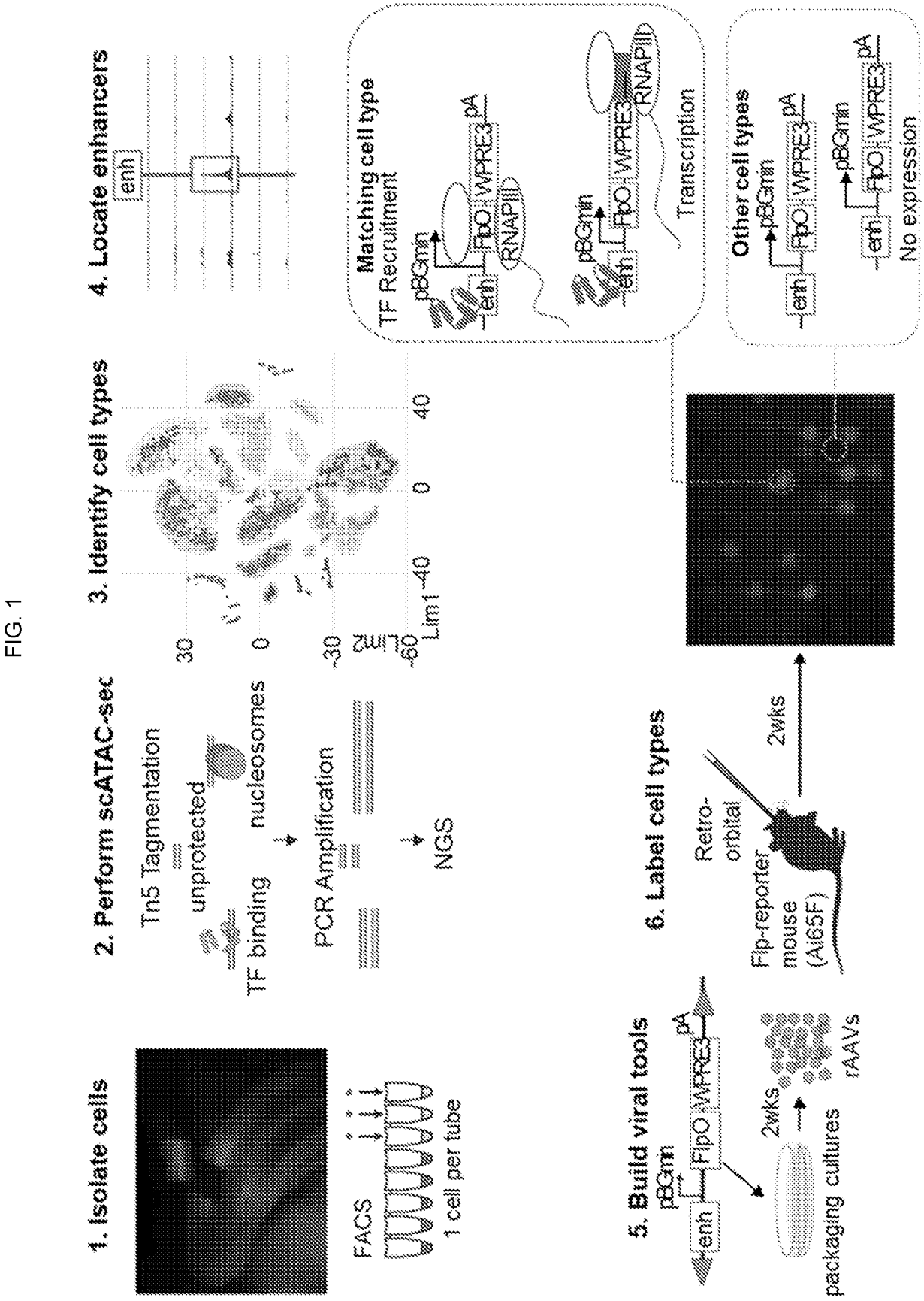
67. The method of claim 66, wherein the living subject is a human, a non-human primate, a rat, or a mouse.

68. The method of claim 66, wherein the administering to a living subject is through injection.

69. The method of claim 68, wherein the injection comprises intravenous injection, intraparenchymal injection into brain tissue, intracerebroventricular (ICV) injection, intra-cisterna magna (ICM) injection, or intrathecal injection.

70. An artificial expression construct comprising CN2320, CN2321, CN2581, CN2782, CN2580, CN2825, CN3270, CN3271, CN3790, CN3751, CN3752, CN1917, CN2047, CN2048, CN2049, CN2427, CN2719, CN2707, CN2713, CN2717, AiP1104, AiP1089, AiP1105, AiP1090, AiP1106, AiP1091, AiP1092, CN3407, CN3408, CN3409, CN3316, CN3793, CN3794, or CN3795.

FIG. 1


1. Isolate cells
FACS
1 cell per tube
2. Perform scATAC-seq
Tn5 Tagmentation
unprotected
nucleosomes
TF binding
PCR Amplification
NGS
3. Identify cell types
30
0
-30
-60
-40
0
40
Lim1
4. Locate enhancers
enh
Matching cell type
TF Recruitment
pBGmin
enh
FlpO
WPRE3
pA
RNAPII
Transcription
Other cell types
No expression
5. Build viral tools
6. Label cell types
Retro-orbital
Flp-reporter mouse (Ai65F)
2wks
packaging cultures
rAAVs

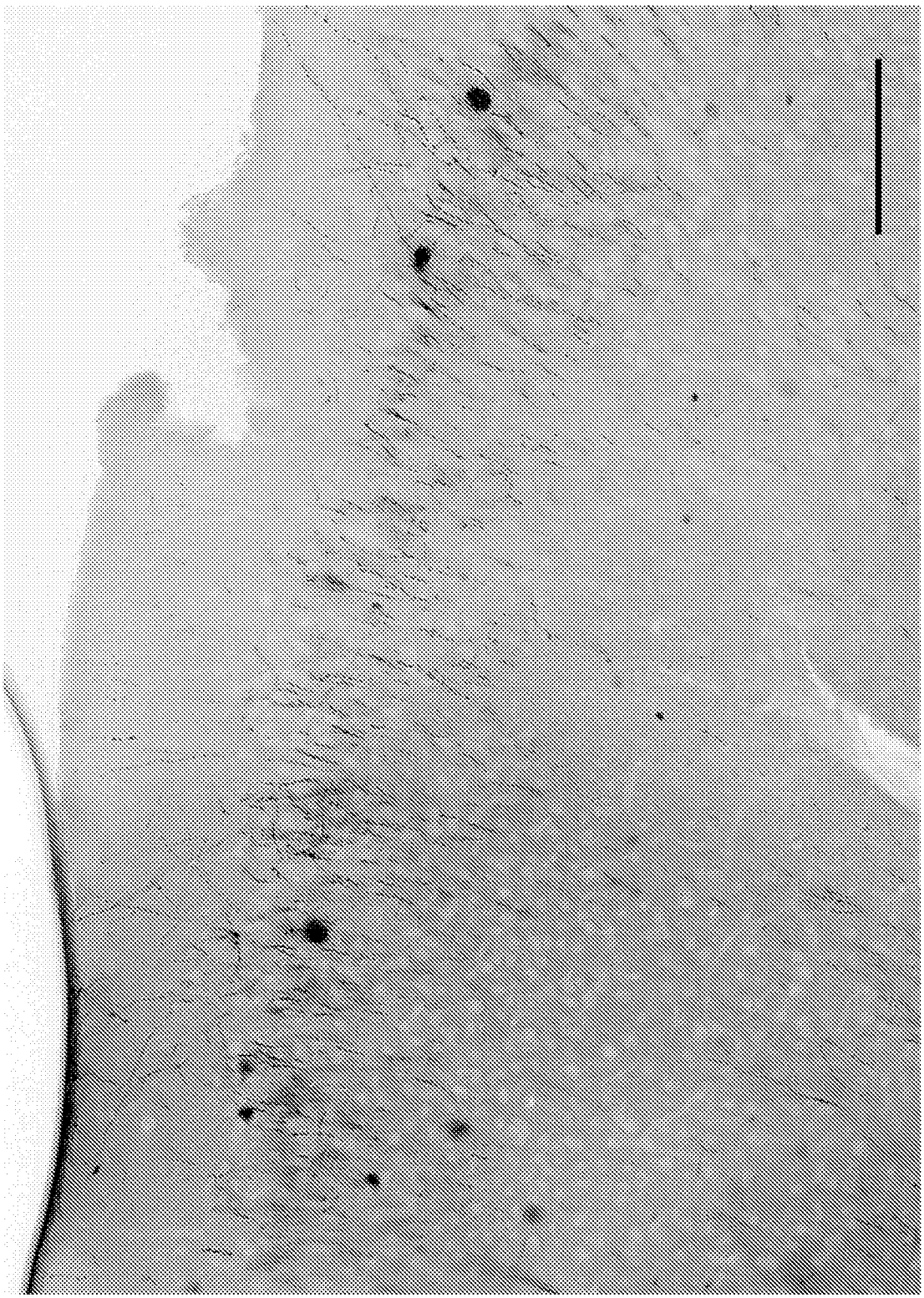


FIG. 2A

FIG. 2B

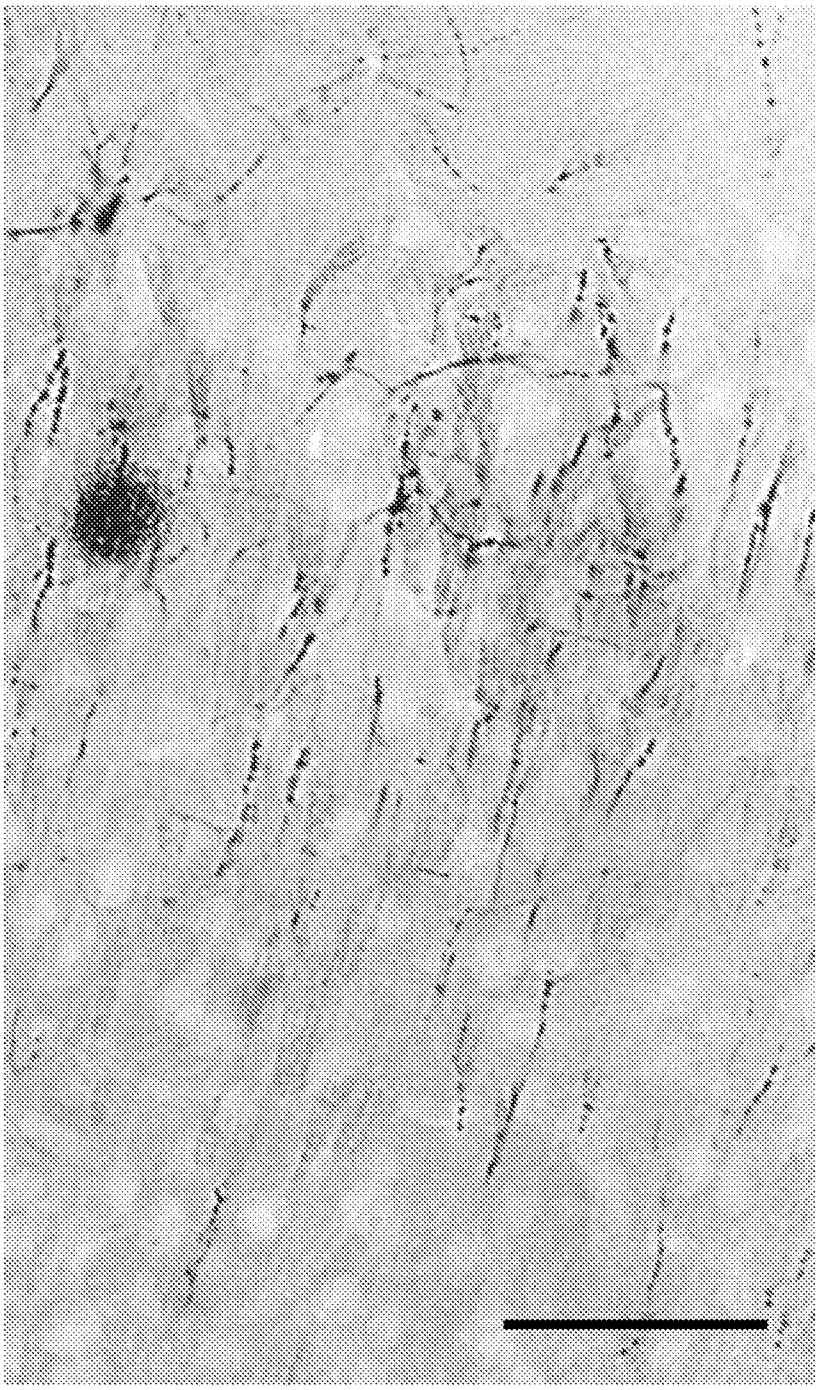
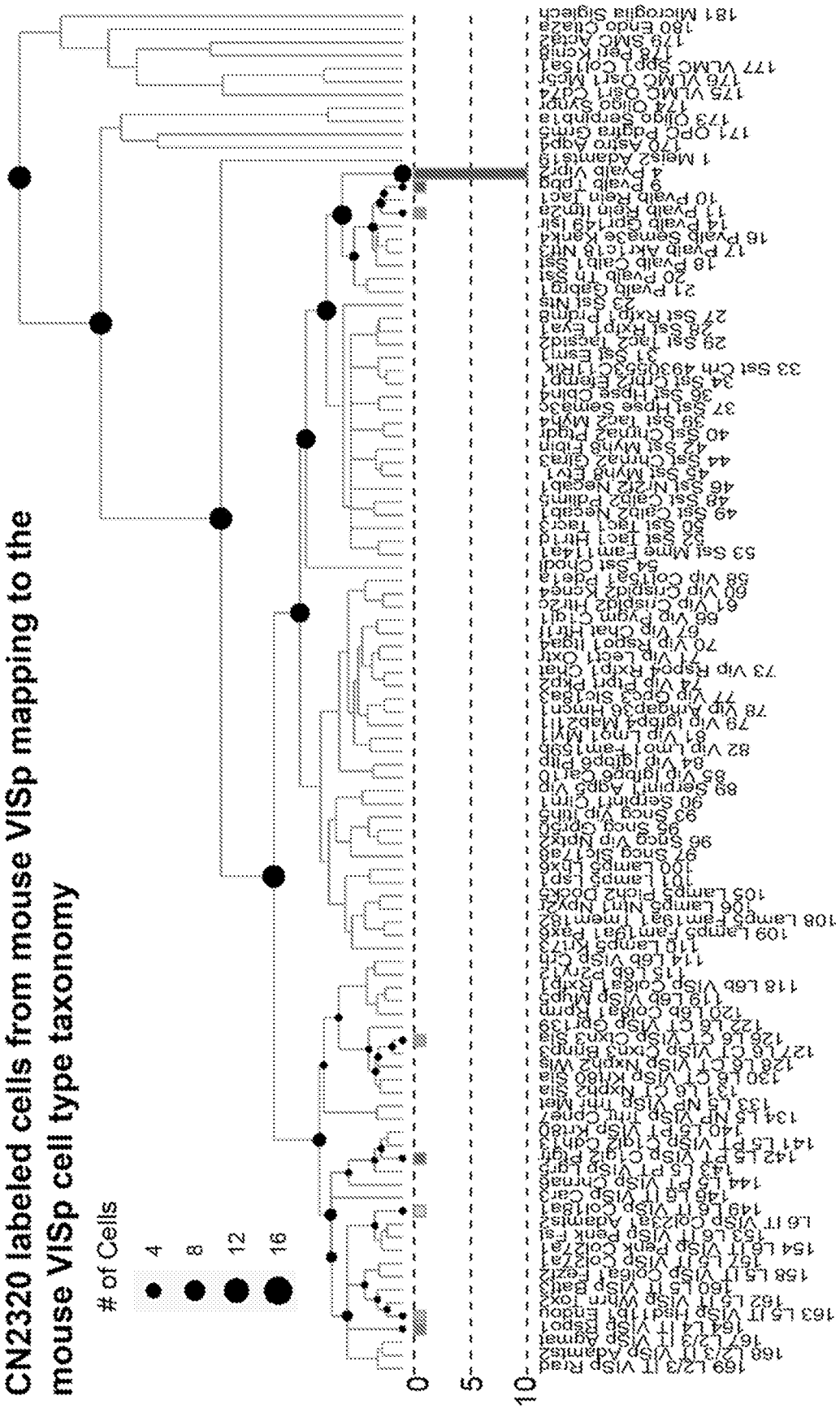


FIG. 3


CN2320 labeled cells from mouse VISp mapping to the mouse VISp cell type taxonomy
# of Cells
4
8
12
16
0
5
10
1 Meis2 Adamts19
4 Pvalb Vipr2
9 Pvalb Tpbg
10 Pvalb Rein Tac1
11 Pvalb Rein Ifitm2
14 Pvalb Gpr149 Isl1
16 Pvalb Sema3e Kank1
17 Pvalb Akr1c18 Nnat
18 Pvalb Calb1 Sst
20 Pvalb Tac1 Sst
21 Pvalb Gabbr1
23 Sst Nts
27 Sst Rxfp1 Eya1
29 Sst Tac2 Tacstd2
31 Sst Esm1
33 Sst Chodl
34 Sst Chrna2 Ptgdr
36 Sst Hpse Cbln4
37 Sst Hpse Sema3c
40 Sst Chrna2 Ptgdr
42 Sst Myh8 Etv1
44 Sst Chrna2 Glra3
45 Sst Myh8 Etv1
46 Sst Nr2f2 Ne2ab1
48 Sst Calb2 Pdlim5
49 Sst Calb2 Necab1
50 Sst Tac1 Tac2
52 Sst Tac1 Htr1d
53 Sst Mme Fam114a1
54 Sst Chodl
58 Vip Col15a1 Pde1a
60 Vip Crispld2 Kcnip4
61 Vip Crispld2 Htr2c
66 Vip Pygm C1ql1
67 Vip Chat Htr3a
70 Vip Rspo1 Itga4
71 Vip Lect1 Oxtr
73 Vip Rspo4 Rxfp1 Chat
74 Vip Plpp4 Pkp2
77 Vip Gpc3 Slc24a3
78 Vip Arhgap36 Hs3st1
79 Vip Igfbp4 Mab21l1
82 Vip Lmo1 Tam1 Sstr2
84 Vip Igfbp5 Plp1
85 Vip Igfbp5 Car10
89 Serpinf1 Aqp5 Vip
90 Serpinf1 Clnp1
93 Sncg Vip Itih5
95 Sncg Cpne5
96 Sncg Vip Nptx2
97 Sncg Slc17a8
100 Lamp5 Lhx6
101 Lamp5 Lsp1
103 Lamp5 Plch2 Dock5
106 Lamp5 Nrn1 Npy2r
108 Lamp5 Fam19a1 Tmem182
109 Lamp5 Fam19a1 Pax6
110 Lamp5 Krt73
114 Lsc Visp Cux2
115 Lsc Pvalb Vipr2
118 Lsc Visp Col23a1 Pvalb
119 Lsc Visp Rspo4 Rxfp1
120 Lsc Col23a1 Tpbpa
122 Lsc Ct Visp Gpr139
126 Lsc Ct Visp Cux2 Slit3
127 Lsc Ct Visp Cux2 Rgs12
128 Lsc Ct Visp Npy2r Wfs1
130 Lsc Ct Visp Kcnip2 Sla
131 Lsc Ct Nxph2 Sla
133 Lsc Np Visp Tmt Met
134 Lsc Np Visp Tmt Cpne7
140 Lsc Pt Visp Krt80
141 Lsc Pt Visp Col19a1 Cdh13
142 Lsc Pt Visp Col19a1 Pdgfc
143 Lsc Pt Visp Lgr5
144 Lsc Pt Visp Chrnb3
146 Lsc It Visp Calb1
149 Lsc It Visp Col12a1 Adamts2
Lsc It Visp Col23a1 Adamts2
154 Lsc It Visp Penk Col12a1
163 Lsc It Visp Penk Fst
158 Lsc It Visp Col6a1 Fezf2
160 Lsc It Visp Wnt7b
162 Lsc It Visp Wnt7b Tox
163 Lsc It Visp Hs3st1 Batf3
164 Lsc It Visp Adamts2
167 Lsc It Visp Col23a1
168 Lsc/3 It Visp Rrad
169 Lsc/3 It Visp Adamts2
181 Microglia Siglech
180 Endo Cldn5
175 SMC Acta2
176 Peri Kcnj8
177 VLMC Spp1 Col13a1
176 VLMC Osr1 Mgp
175 VLMC Osr1 Cd74
174 Oligo Syngr1
173 Oligo Serpinb1a
171 OPC Pdgfra Gpr17
170 Astro Agt

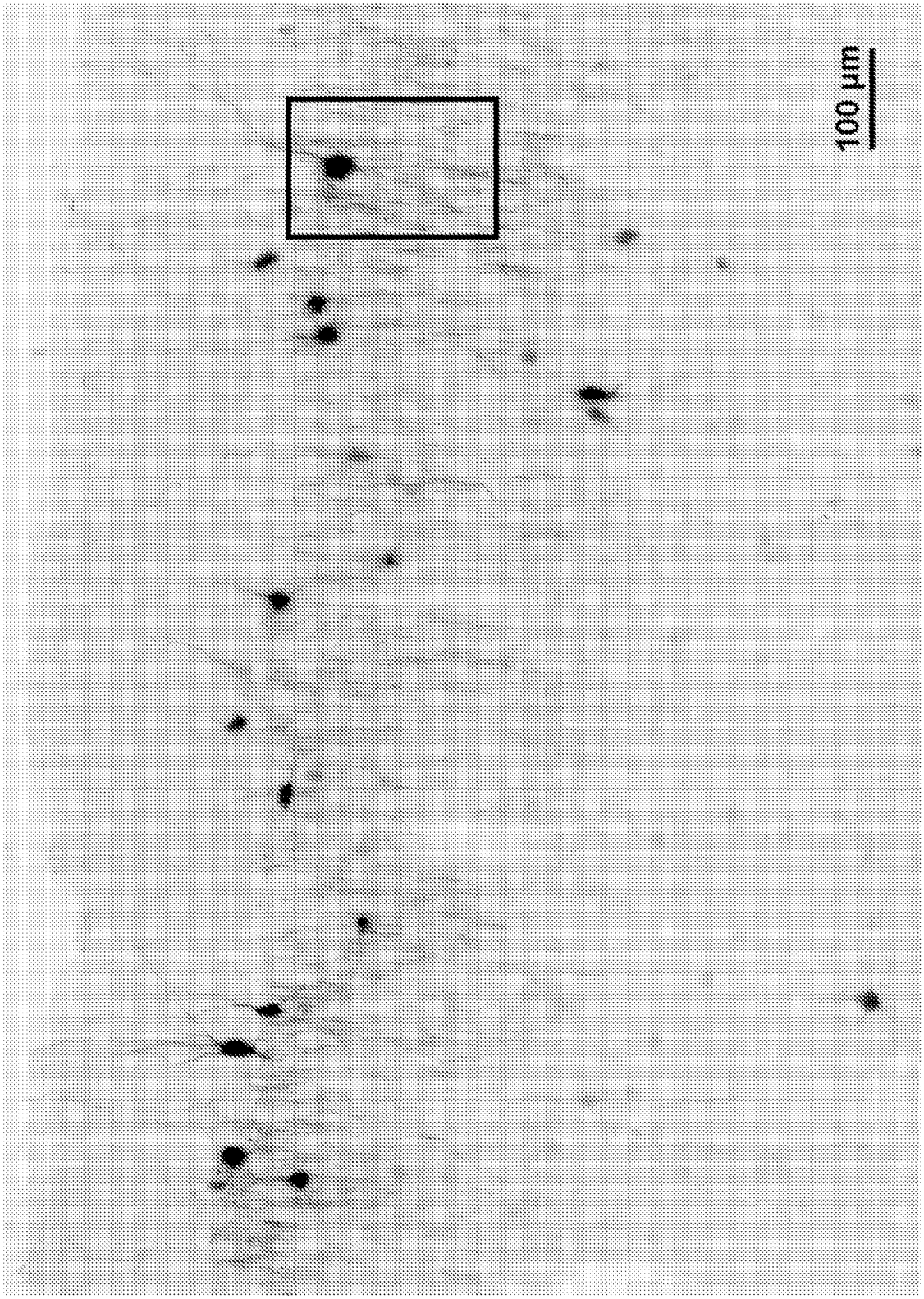
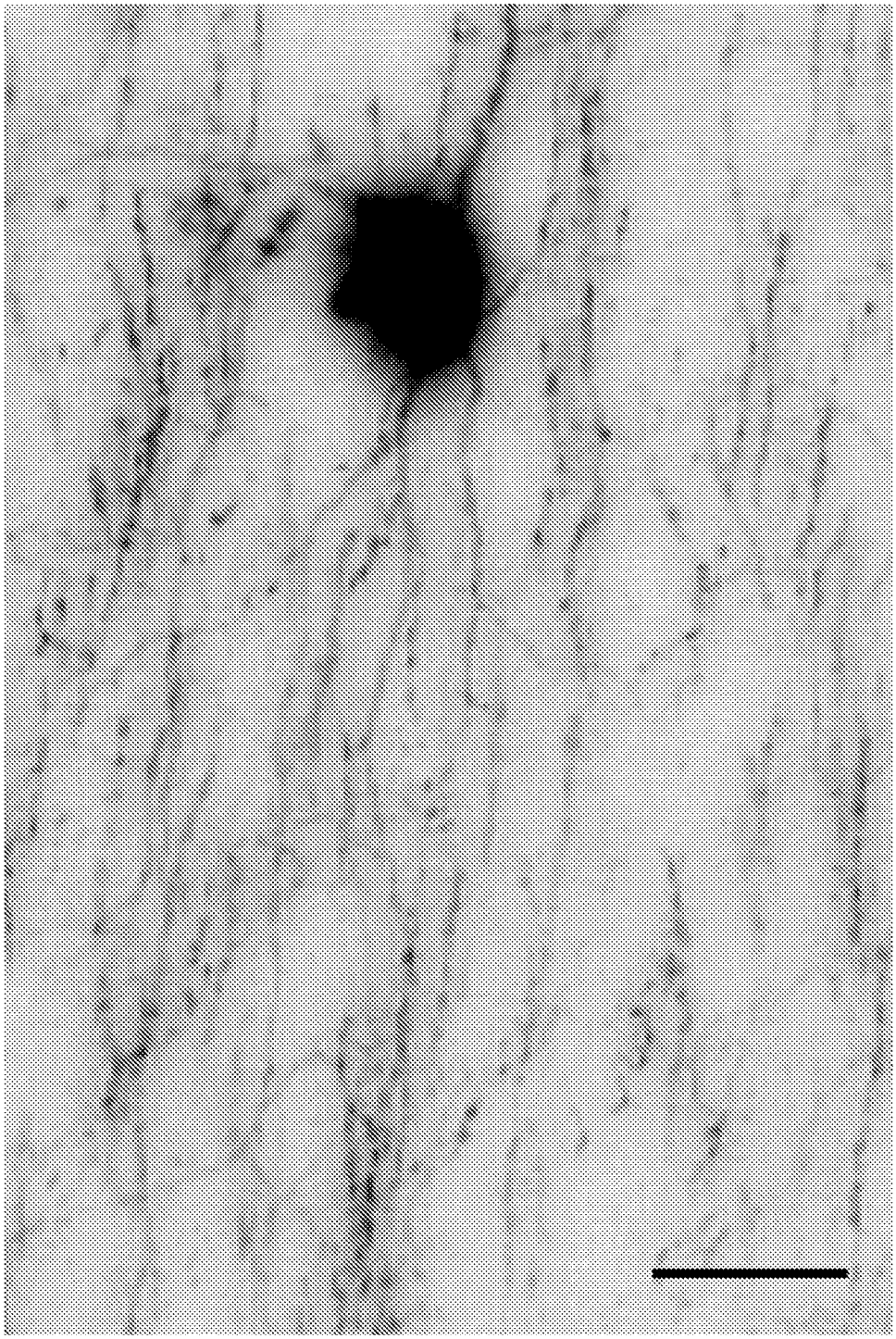


FIG. 4A

FIG. 4B



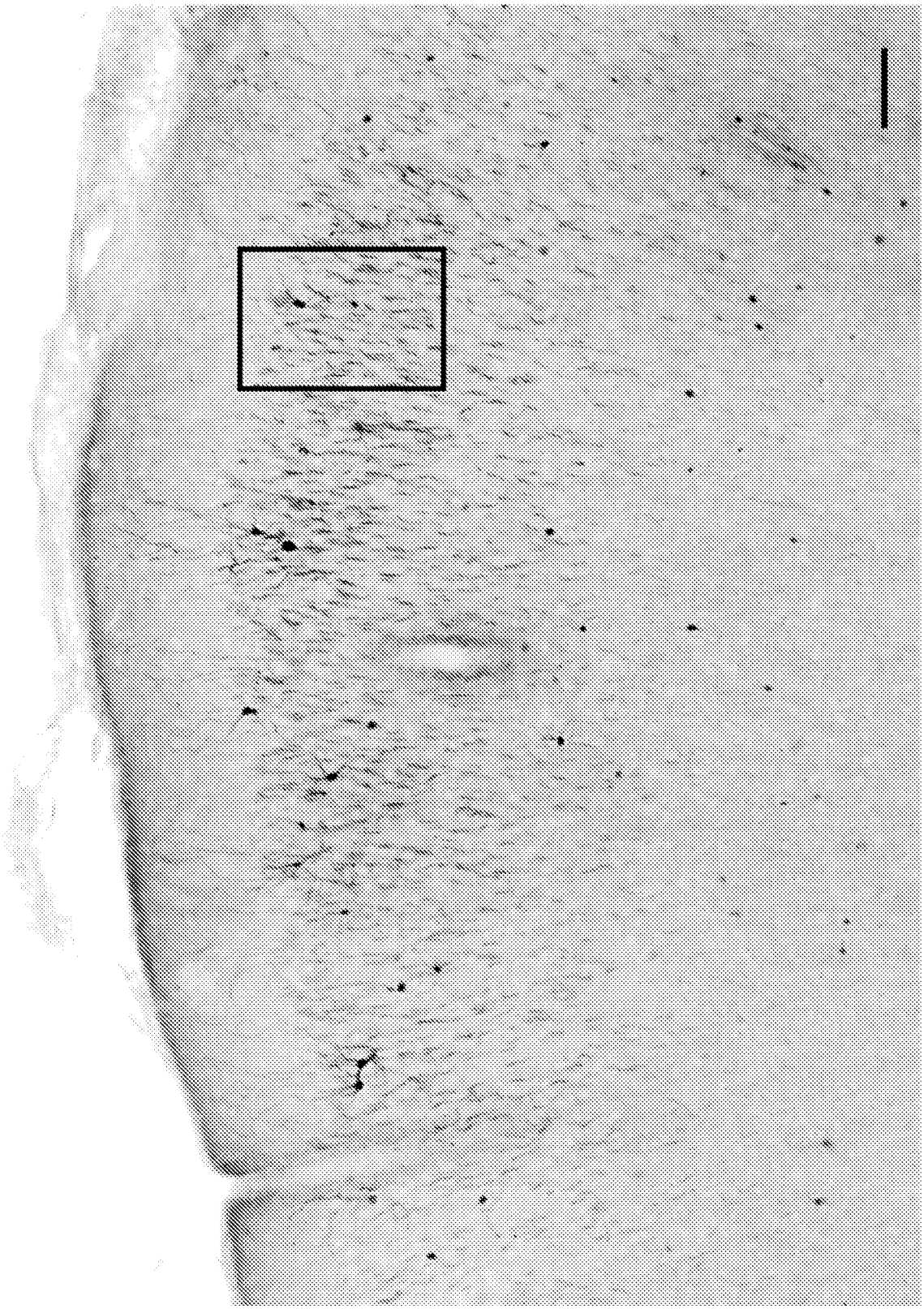


FIG. 5A

FIG. 5B

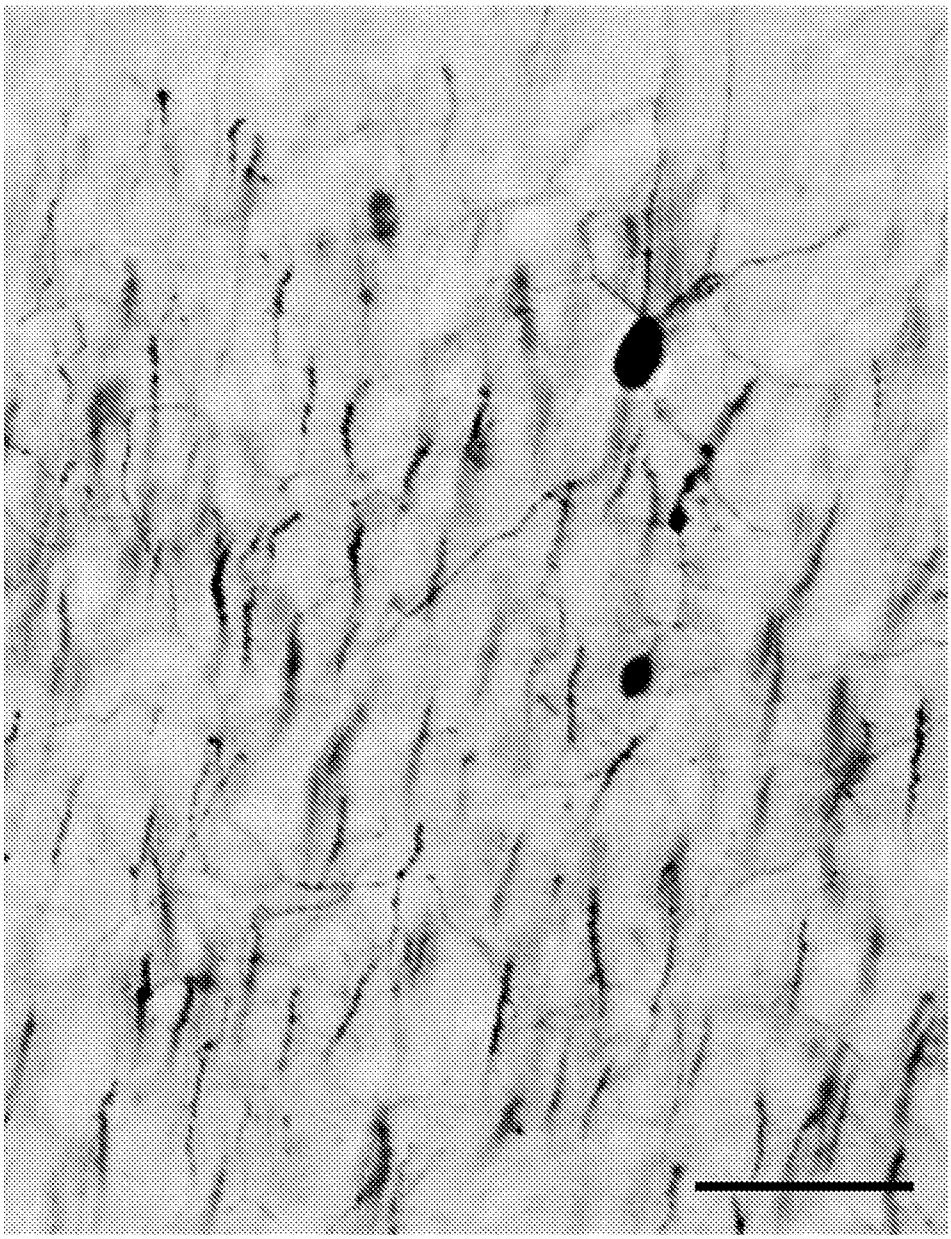
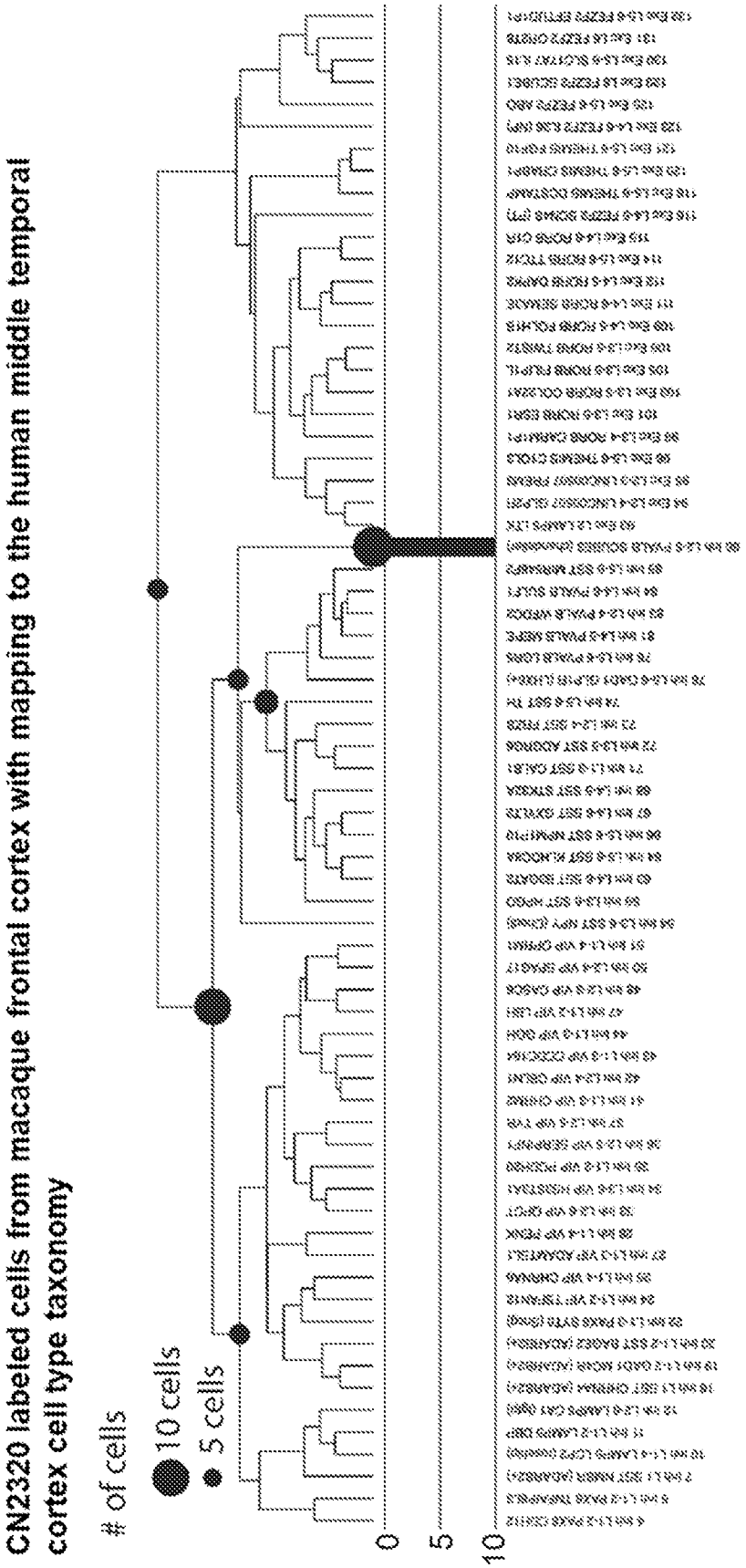


FIG. 6A


CN2320 labeled cells from macaque frontal cortex with mapping to the human middle temporal cortex cell type taxonomy
# of cells
10 cells
5 cells
0
5
10

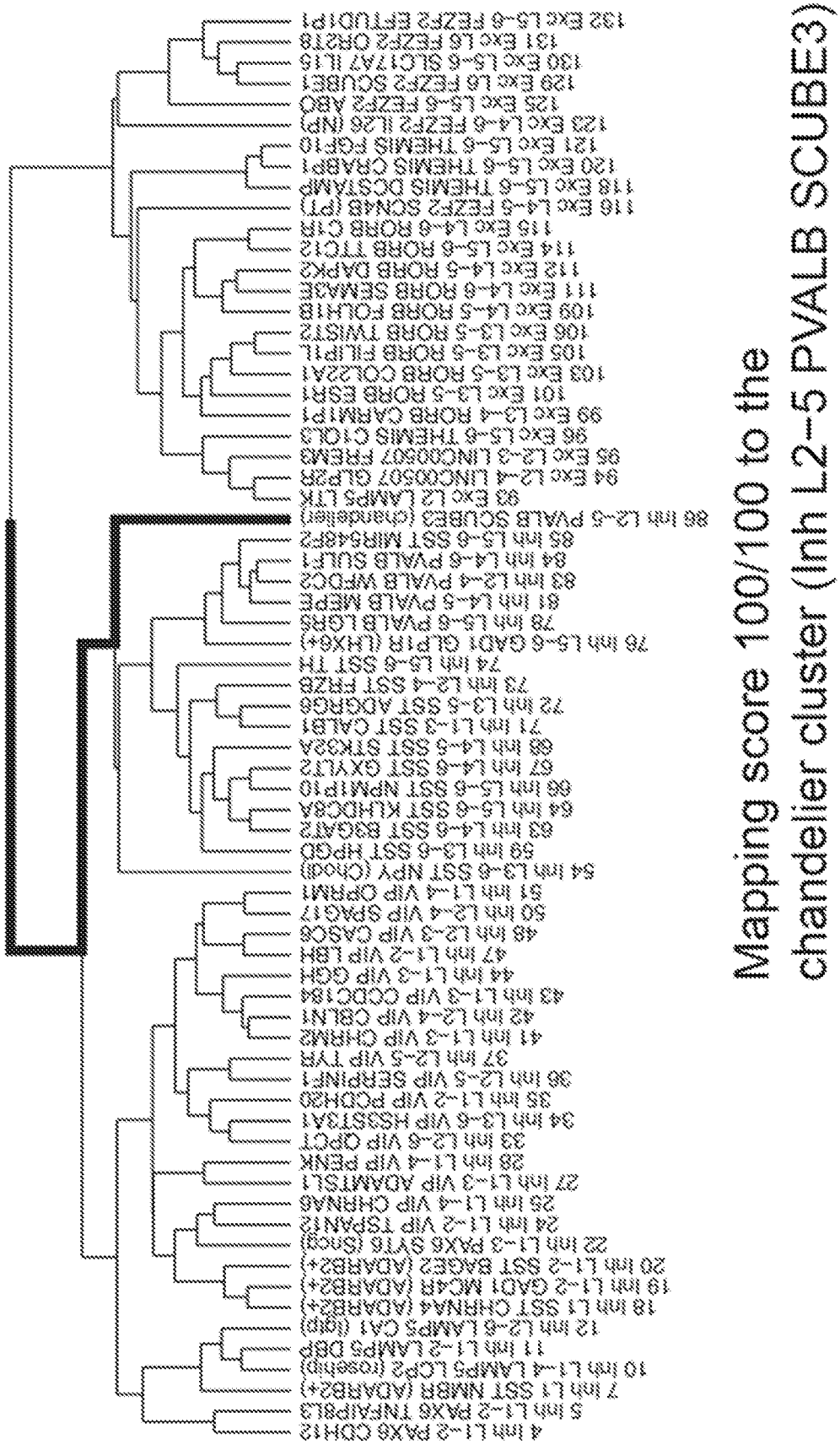
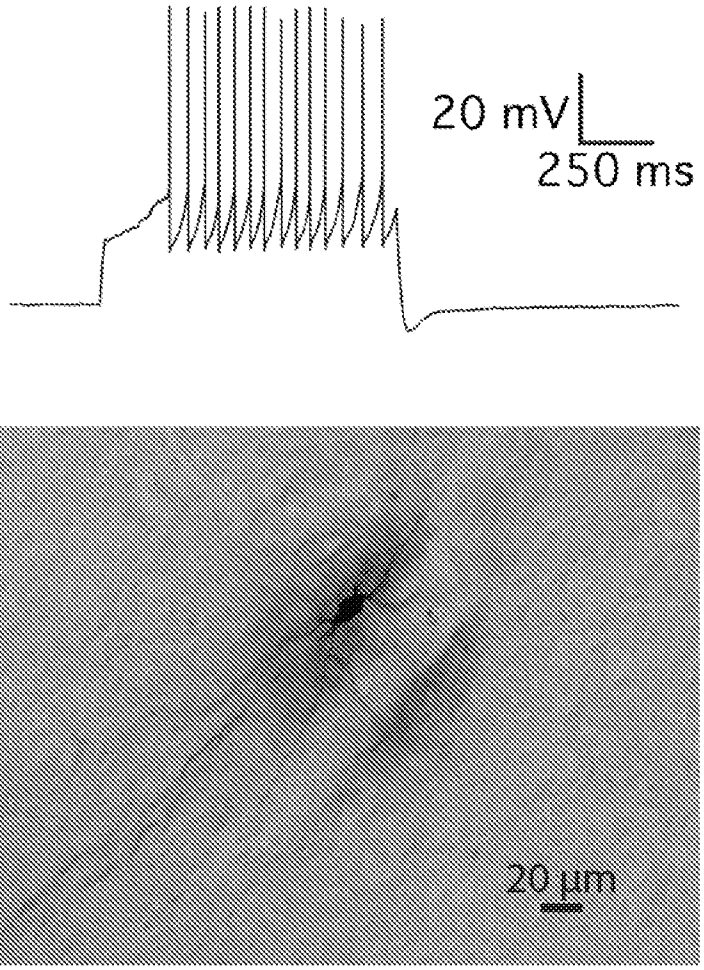

Mapping score 100/100 to the
chandelier cluster (Inh L2–5 PVALB SCUBE3)
1 Inh L1–2 PAX6 CDH12
2 Inh L1–2 PAX6 TNFAIP8L3
3 Inh L1 SST NMBR (ADARB2+)
4 Inh L1–4 LAMP5 LCP2 (rosehip)
5 Inh L1–2 LAMP5 DBP
6 Inh L2–6 LAMP5 CA1 (ngf)
7 Inh L1 SST CHRNA4 (ADARB2+)
8 Inh L1–2 GAD1 MC4R (ADARB2+)
9 Inh L1–2 SST BAGE2 (ADARB2+)
10 Inh L1–3 PAX6 SYT6 (Sncg)
11 Inh L1–2 VIP TSPAN12
12 Inh L1–4 VIP CHRNA6
13 Inh L1–3 VIP ADAMTSL1
14 Inh L1–4 VIP PENK
15 Inh L2–6 VIP QPCT
16 Inh L3–6 VIP HS3ST3A1
17 Inh L1–2 VIP PCDH20
18 Inh L2–5 VIP SERPINF1
19 Inh L2–5 VIP TYR
20 Inh L1–2 VIP CHRNA2
21 Inh L2–4 VIP CBLN1
22 Inh L1–3 VIP CDC14B
23 Inh L1–3 VIP GGH
24 Inh L1–2 VIP LBH
25 Inh L2–3 VIP CASC6
26 Inh L2–4 VIP SPAG17
27 Inh L1–4 VIP CPNM1
28 Inh L3–6 SST NPY (Chodl)
29 Inh L3–6 SST HPGD
30 Inh L4–6 SST B3GAT2
31 Inh L5–6 SST KLHDC8A
32 Inh L5–6 SST NPM1P10
33 Inh L4–6 SST GXYLT2
34 Inh L4–5 SST STK32A
35 Inh L1–3 SST CALB1
36 Inh L3–5 SST ADGRG6
37 Inh L2–4 SST FRZB
38 Inh L5–6 SST TH
39 Inh L5–6 GAD1 GLP1R (LHX6+)
40 Inh L5–6 PVALB LGR5
41 Inh L4–5 PVALB MEPE
42 Inh L2–4 PVALB WFDC2
43 Inh L4–6 PVALB SULF1
44 Inh L5–6 SST MIR548F2
45 Inh L2–5 PVALB SCUBE3 (chandelier)
46 Exc L2 LAMP5 LTK
47 Exc L2–4 LINC00507 GLP2R
48 Exc L2–3 LINC00507 FREM3
49 Exc L5–6 THEMIS C1QL3
50 Exc L3–4 RORB CARM1P1
51 Exc L3–5 RORB ESR1
52 Exc L3–5 RORB COL22A1
53 Exc L3–5 RORB FILIP1L
54 Exc L3–5 RORB TWIST2
55 Exc L4–5 RORB FOLH1B
56 Exc L4–6 RORB SEMA3E
57 Exc L4–6 RORB DAPK2
58 Exc L5–6 RORB TTC12
59 Exc L4–6 RORB C1R
60 Exc L4–5 FEZF2 SCN4B (PT)
61 Exc L5–6 THEMIS DCSTAMP
62 Exc L5–6 THEMIS CRABP1
63 Exc L5–6 THEMIS FGF10
64 Exc L4–6 FEZF2 IL26 (NP)
65 Exc L5–6 FEZF2 ABO
66 Exc L6 FEZF2 SCUBE1
67 Exc L5–6 SLC17A7 IL16
68 Exc L6 FEZF2 CR2T8
69 Exc L5–6 FEZF2 EFTUD1P1

FIG. 6B

FIG. 6C



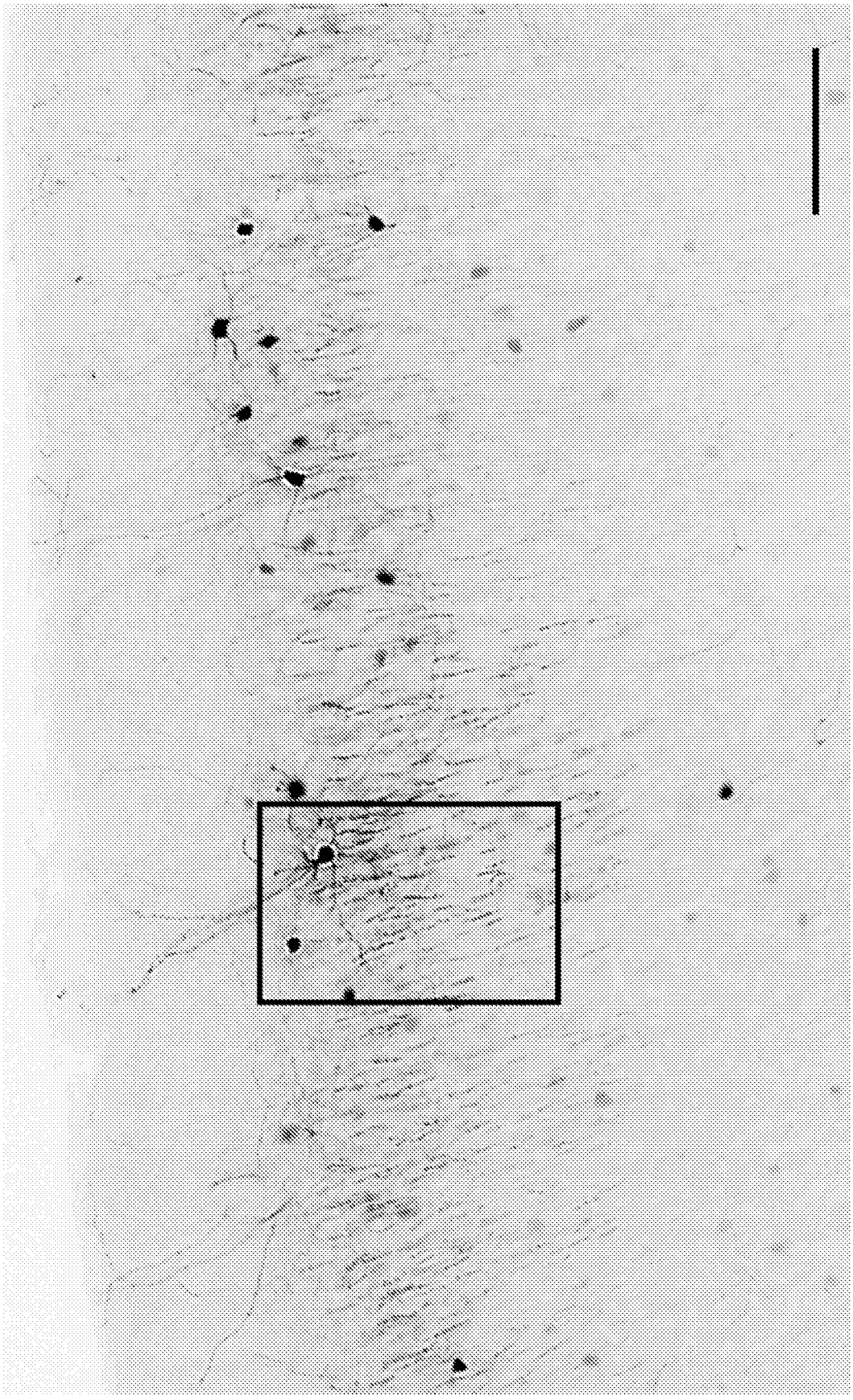


FIG. 7A

FIG. 7B

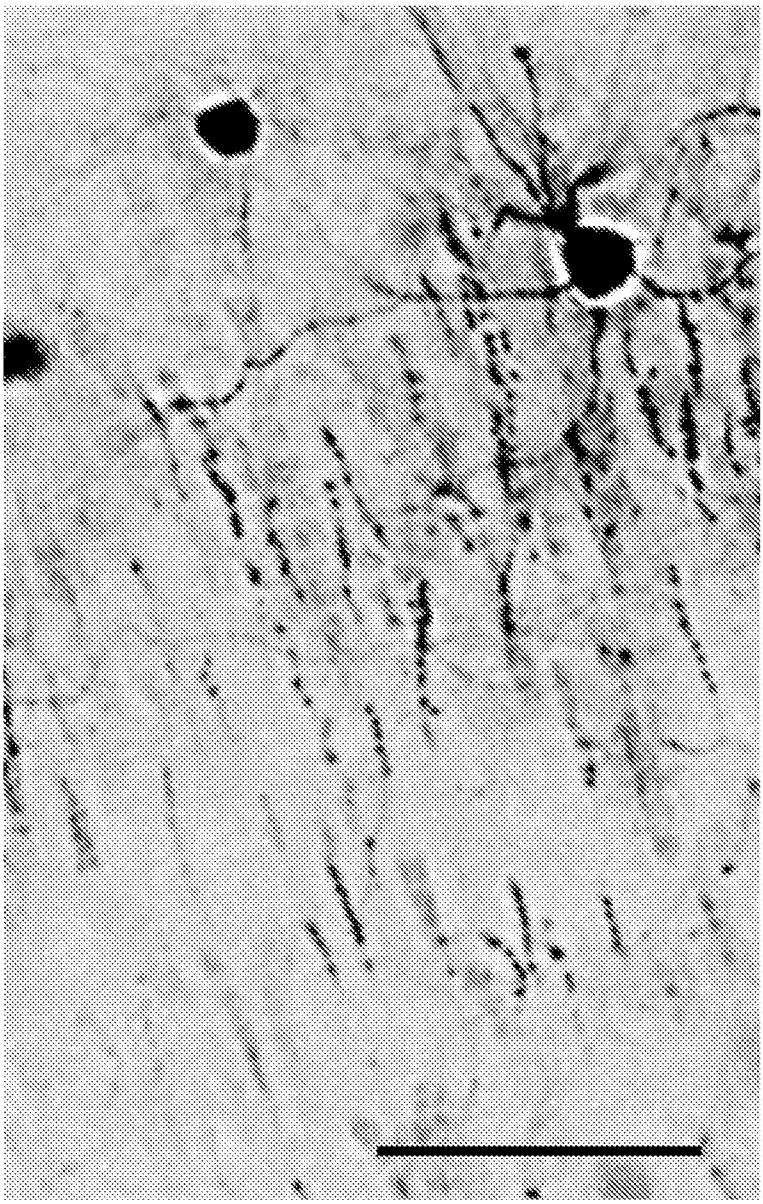
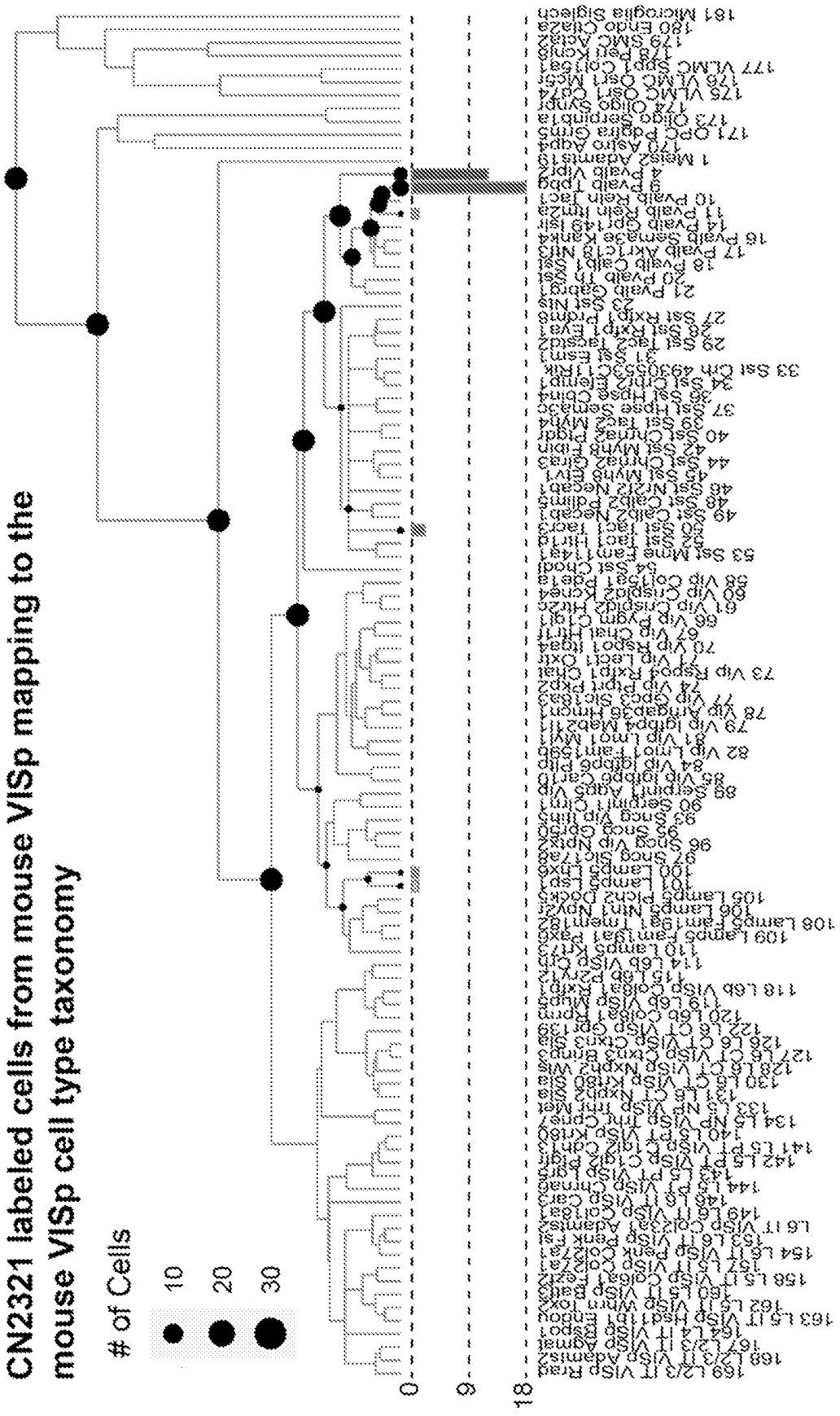


FIG. 8


CN2321 labeled cells from mouse VISp mapping to the mouse VISp cell type taxonomy
# of Cells
10
20
30
0
9
18

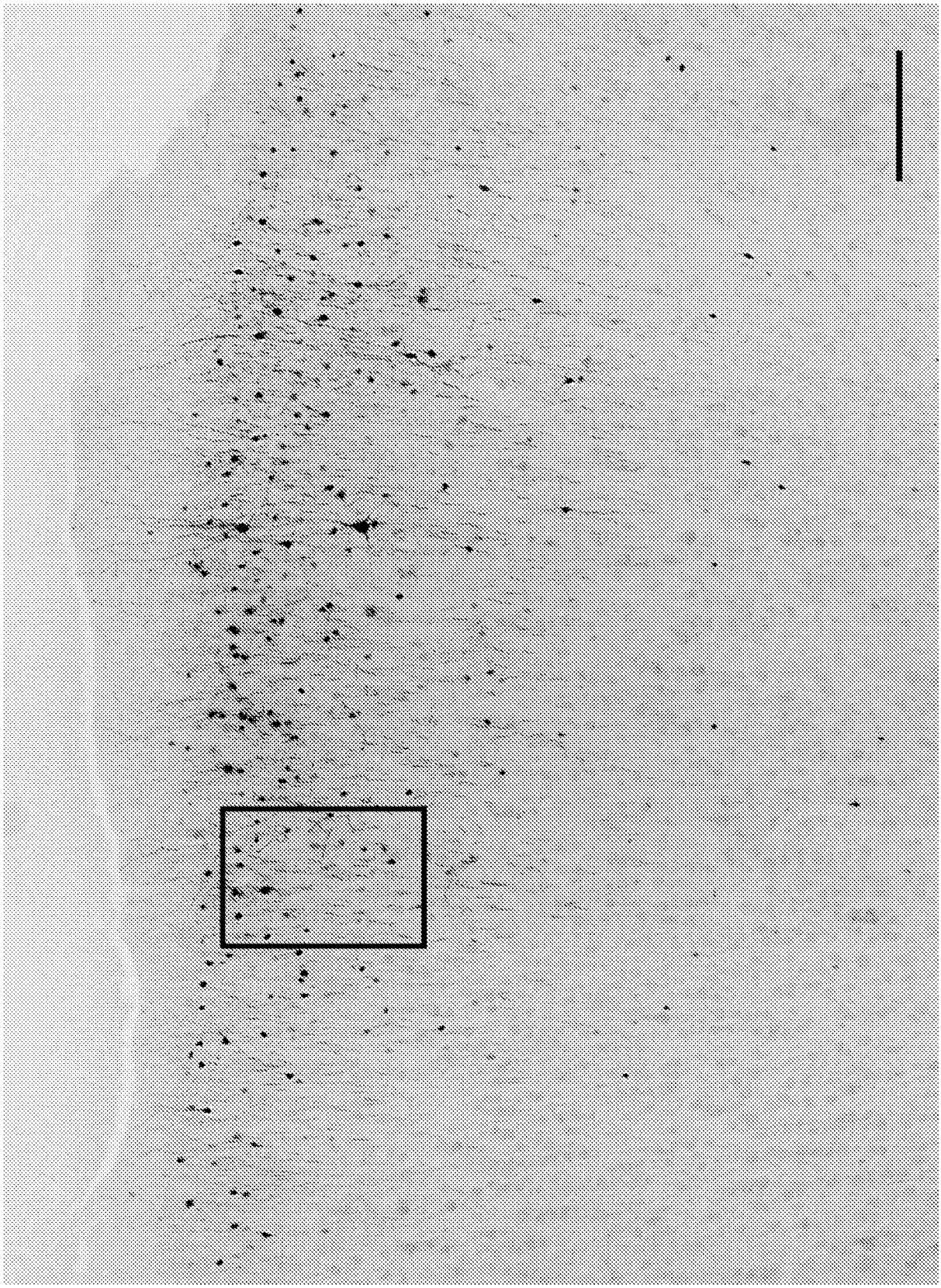
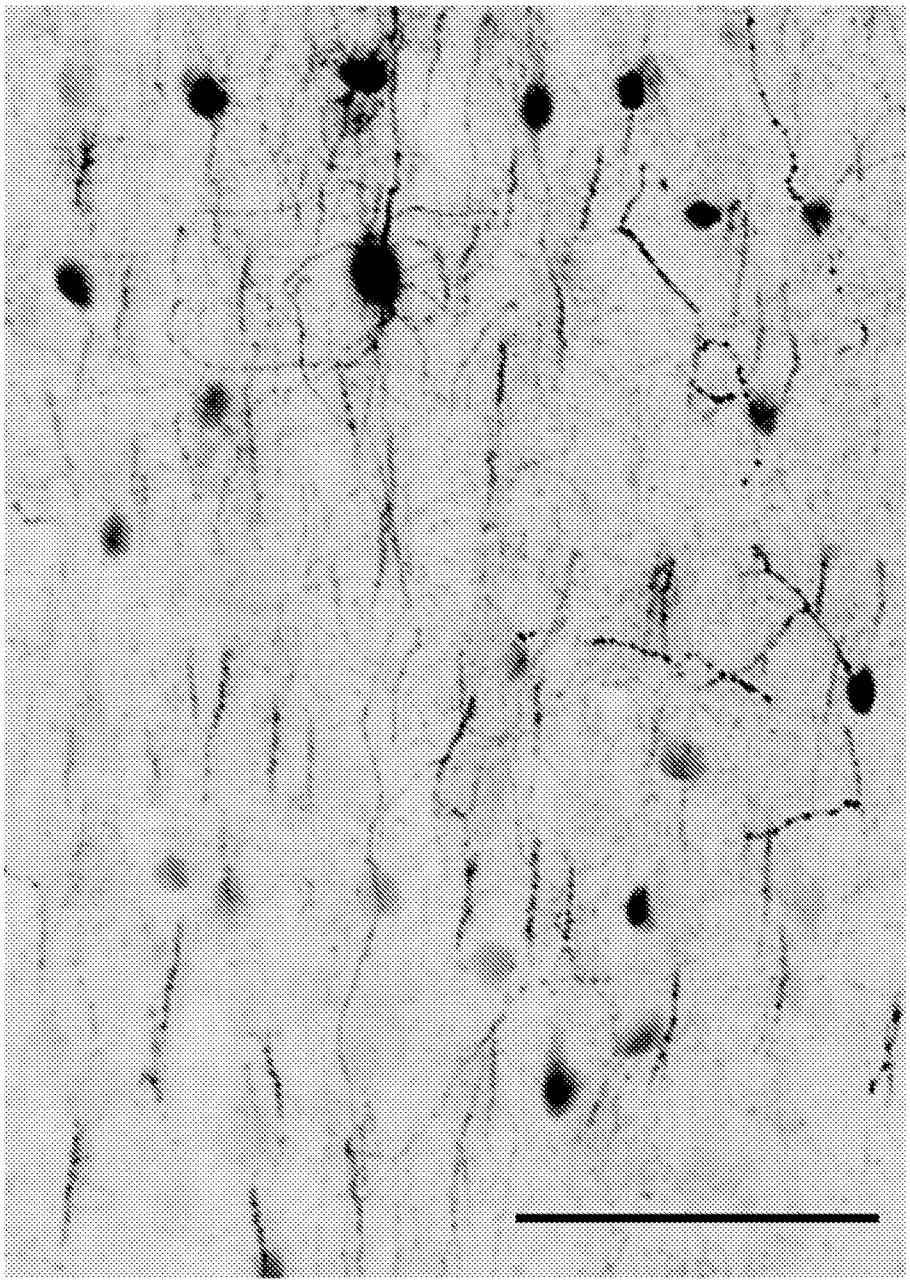


FIG. 9A

FIG. 9B



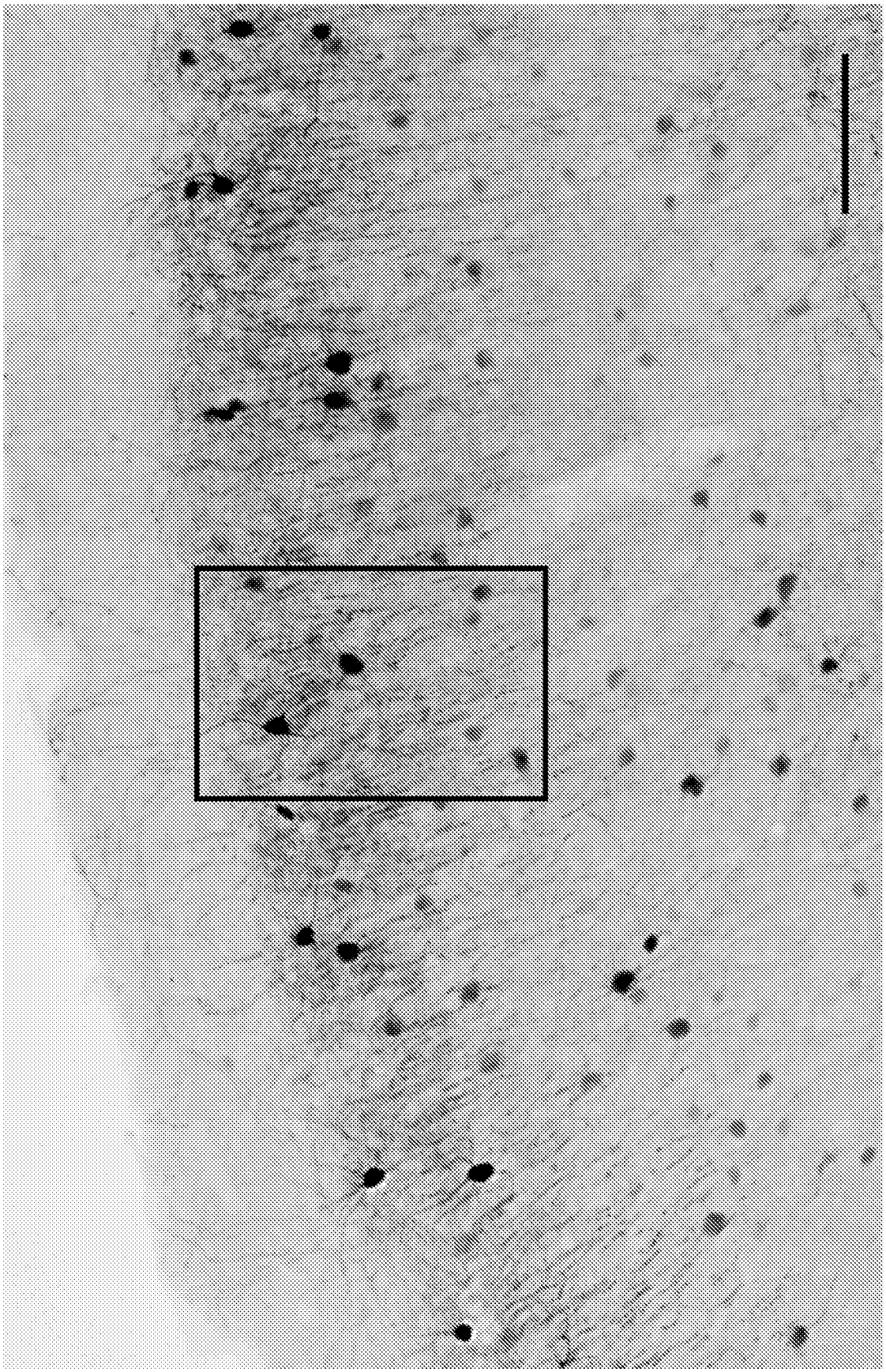
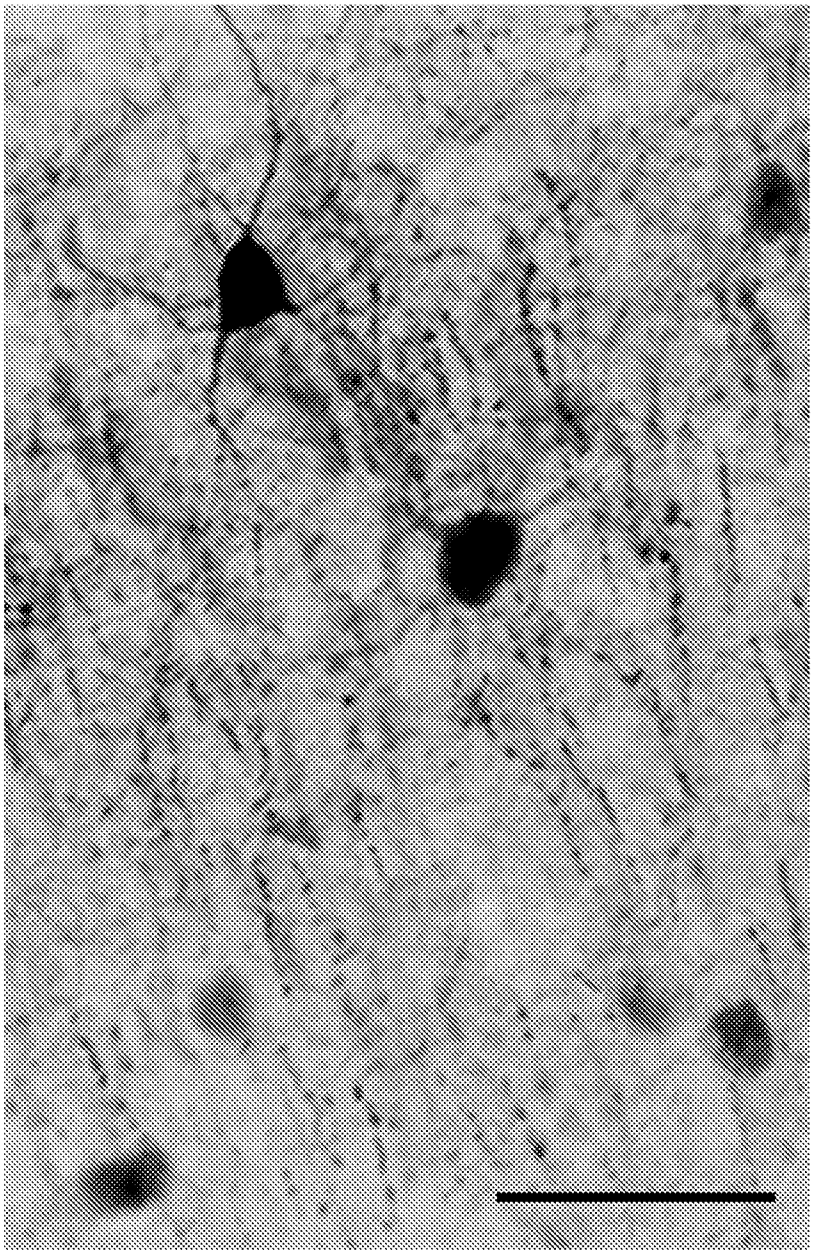


FIG. 10A

FIG. 10B



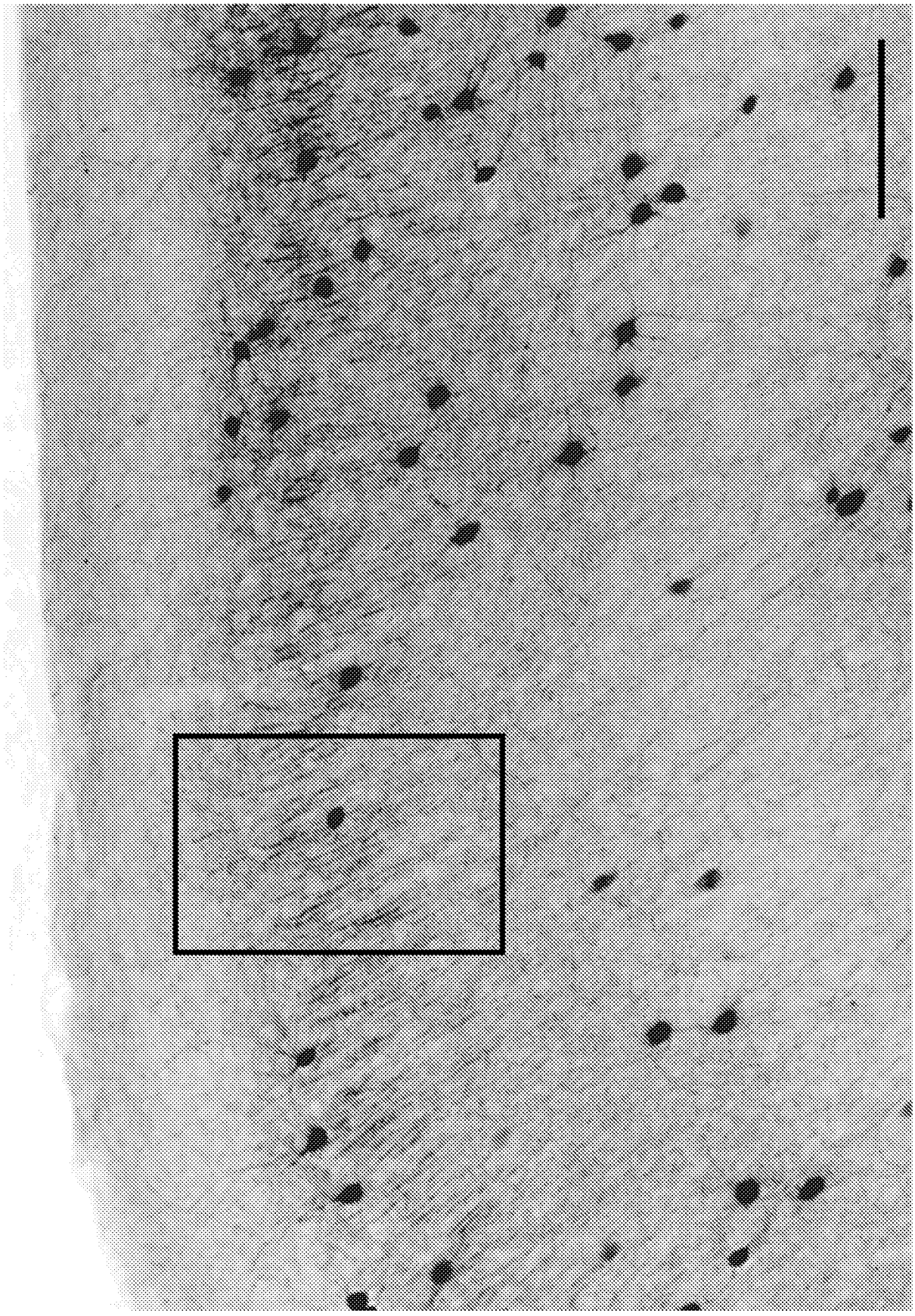
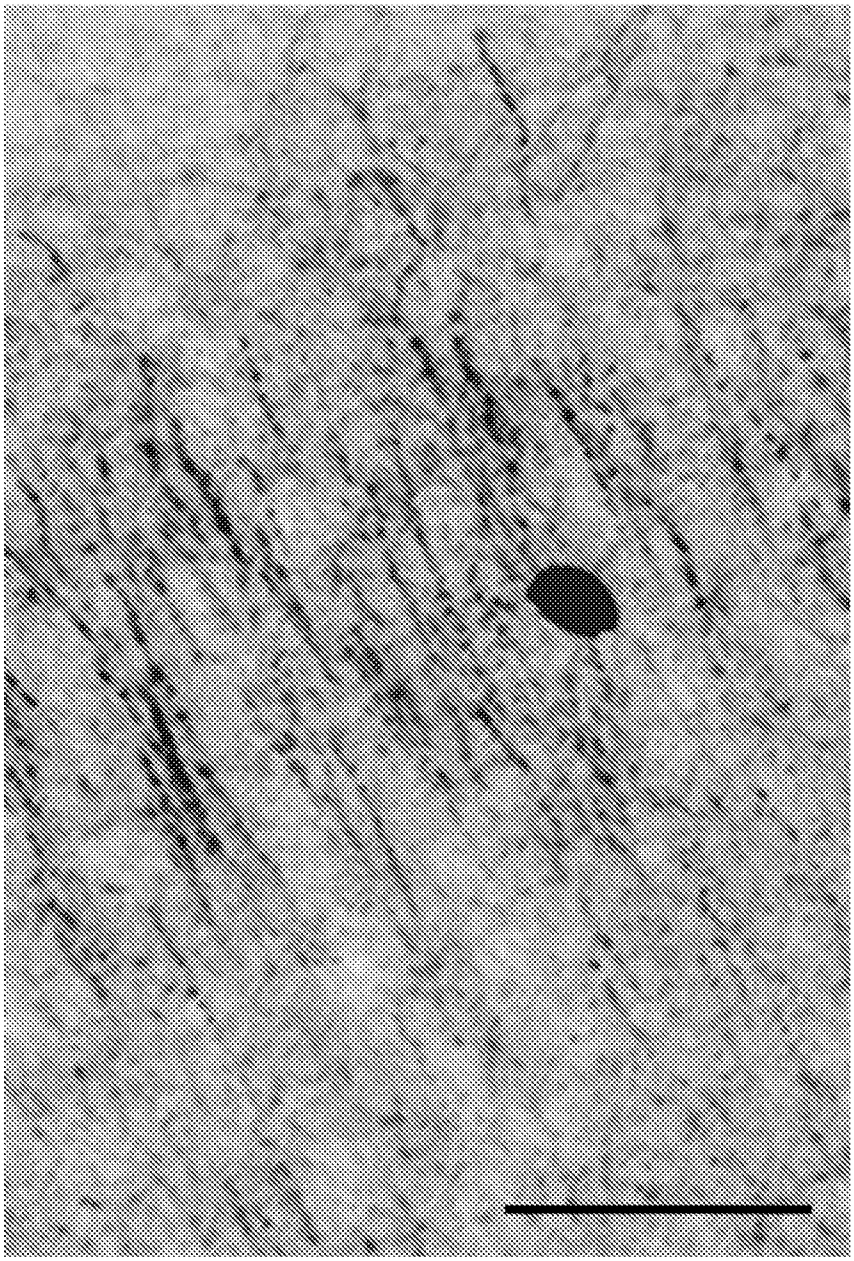


FIG. 11A

FIG. 11B



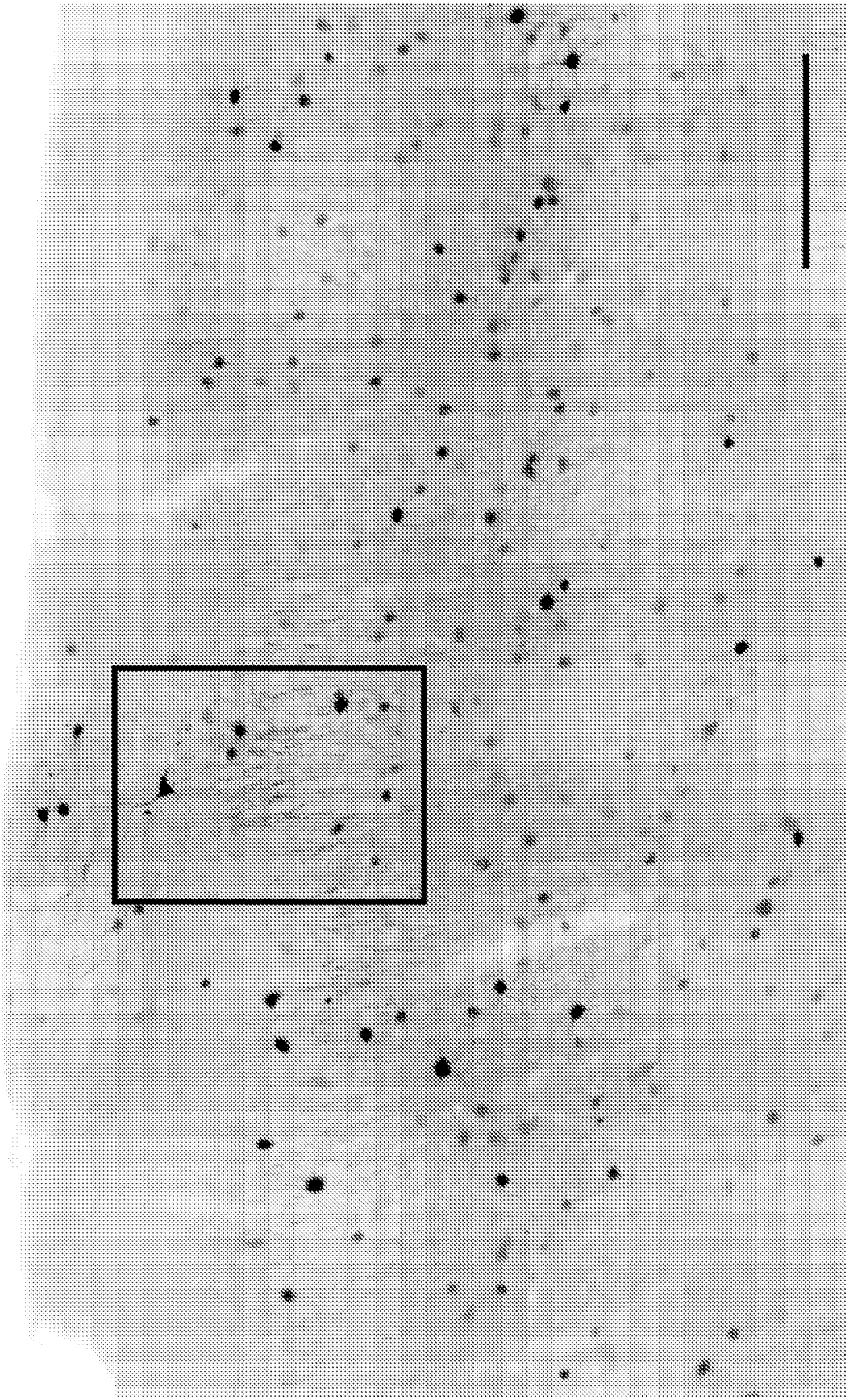
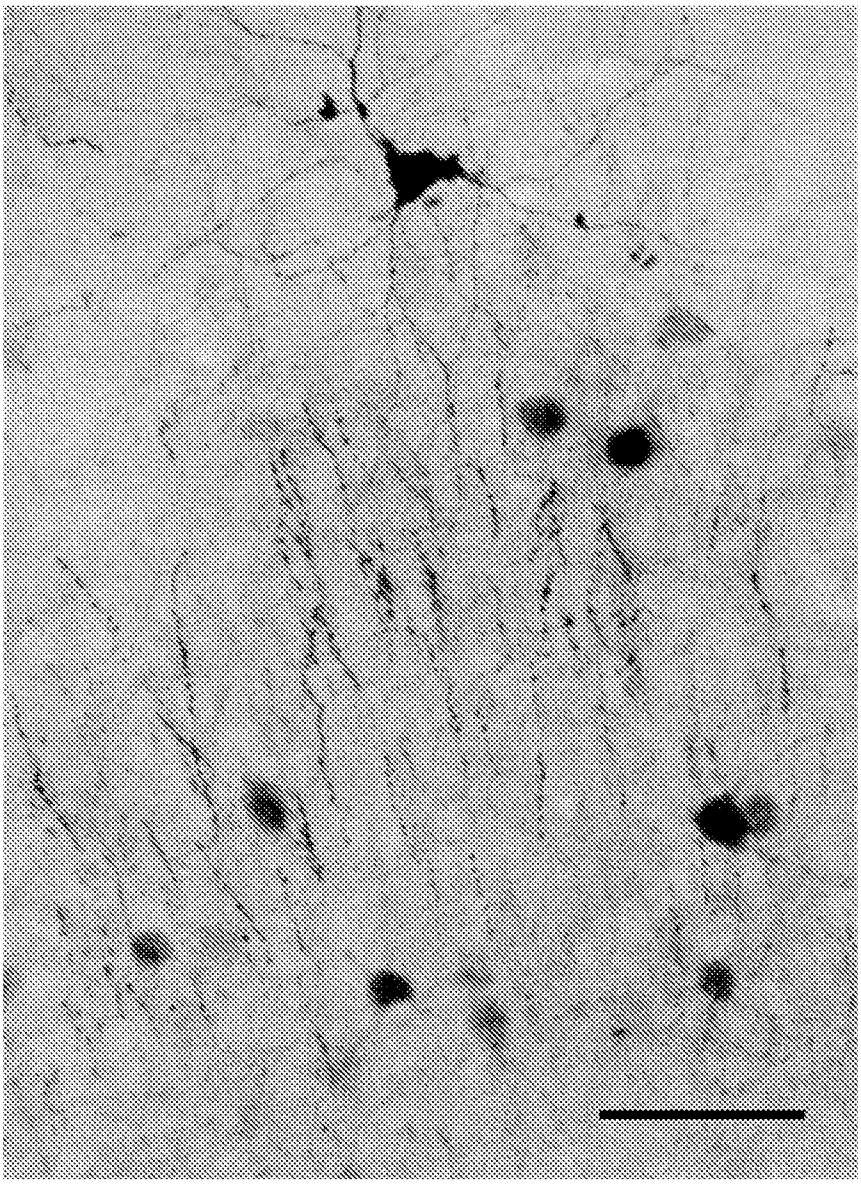


FIG. 12A

FIG. 12B



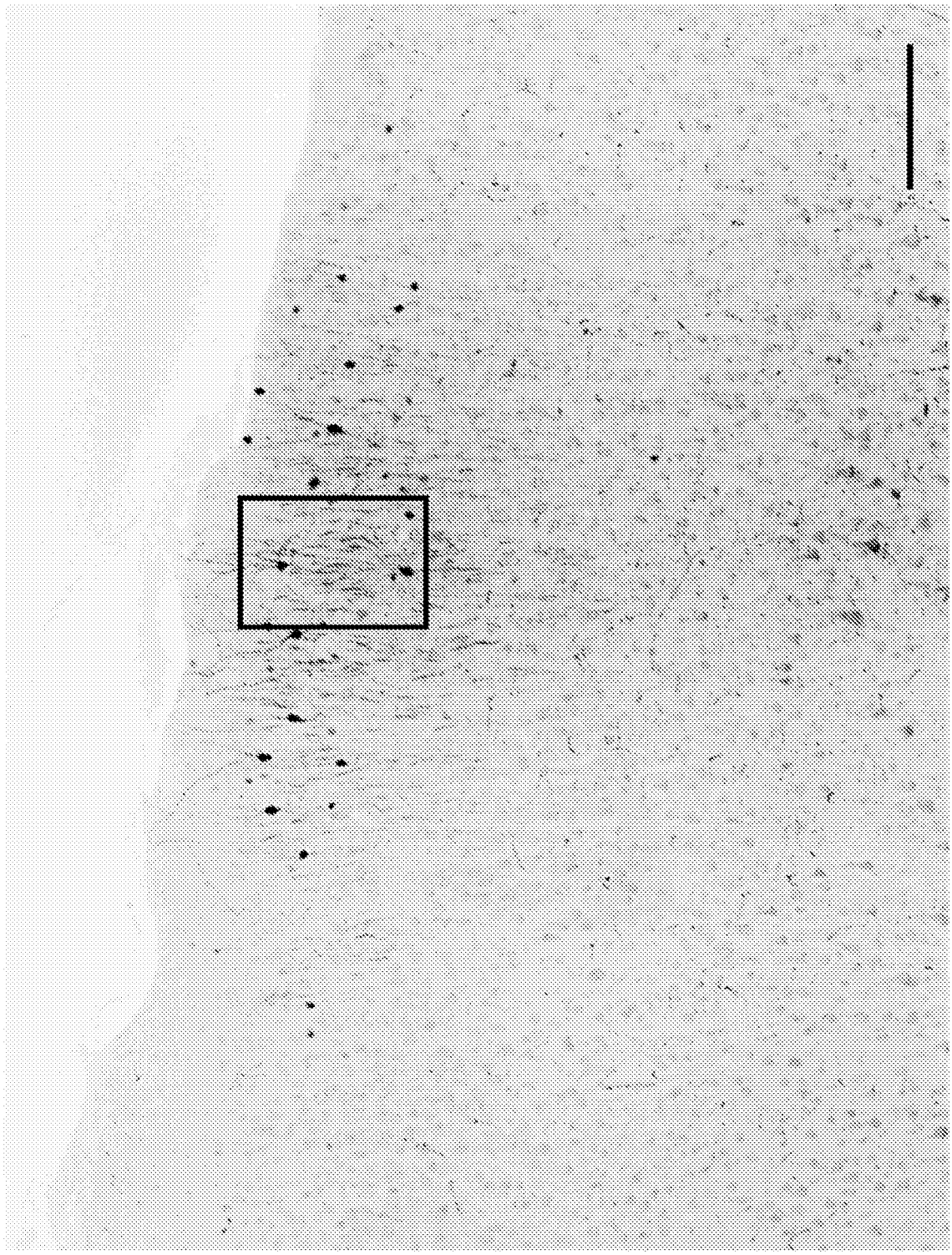


FIG. 13A

FIG. 13B

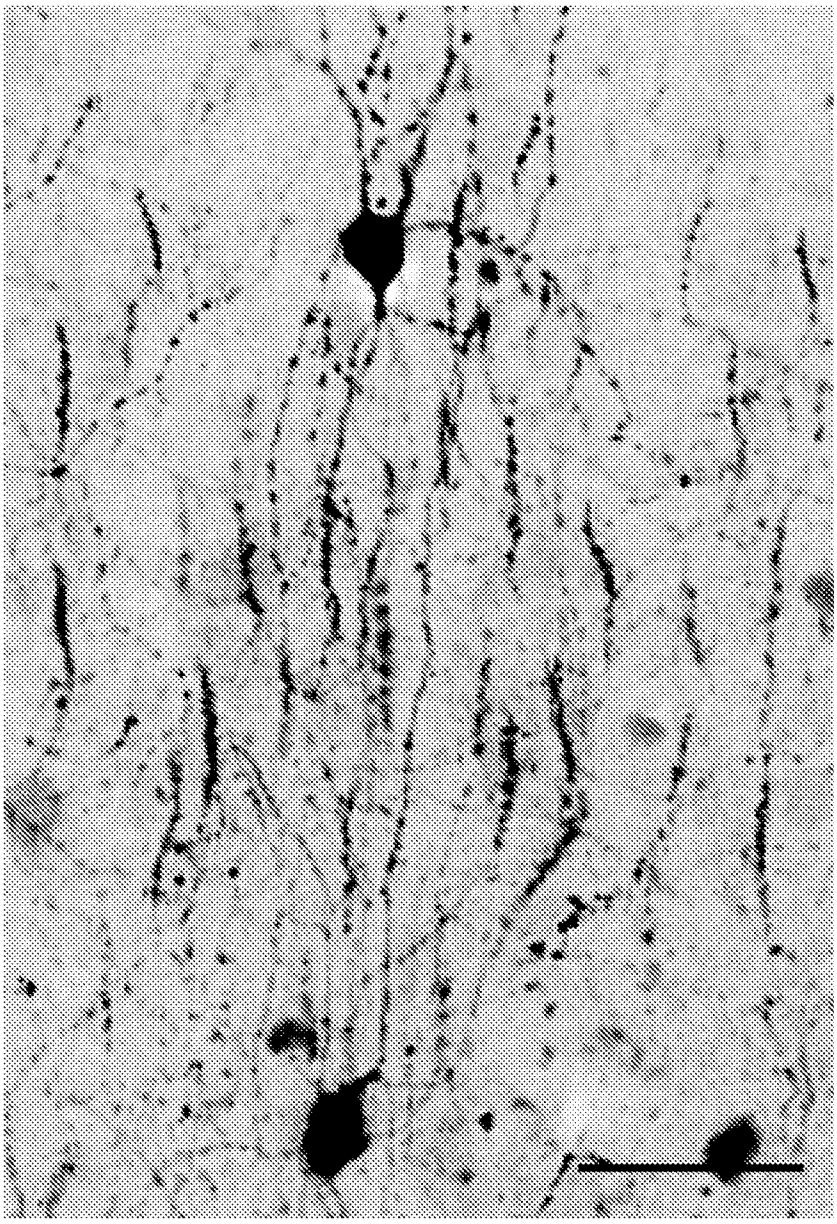


FIG. 14

eHGT_297m (907 bp):
ACTATTCCAGCCACGAGGTATAAACACTGGGAAGGAAAGTCCTGGCTCTGTATTGTCCACA
AAGACCCGAAGCTGCAGCAAAGTTGGCAAAAGAGAAACAAAAAGAGCAGAGAAGGCTCAG
CTTTCAACAGCTAGGCTCACCCAAACATCAAGAGGTGGACAAATATTTACAGTGTGAACCT
TAACCCAAAGAACGGCAGTGCGGTCATGCTGCACAGAGTCAACTTCAGAACCAAGACTGT
GACCAGGGCTGGAGCAAGAGACACTTCACGCTAATAAATAGGCCACTTAATCAAGAAGCT
GTCACAGTCCTAAATATGTATGCACCGAGCATTAGCACTTCCAAGTAGAGTGGAACAGCTA
AAGATAGAGGCAAGAAGCAAGCAGACAAATCTTCGGTGATTGTTGGAAATCACAGCATTTC
TCTCAGCAATTGTCAGGACACAGAAAATCAGCGAGAAGACAGAAGAGTCTCACAATAGTCC
CCATCAACTTGACCTAATTGACATTTATGGAGCTTGGCATCCAACAGCCGTGGAGCGCATG
CGCTCTTTAAGGCAGAATACAGACCATCAAACAAAACCAGGGAGACCAAAGTCACAGAAAA
TATGCTCTGTGAACATGACATAATAAAGTGGAAATCGATAACAGAGAGATCGCTGCAAAAT
CCCCCAAGTGATTGGTAATTAAATGCTCTACTCCTGAATGAATGATGGGCGAGAAAGGAAA
GCCACGGGGGAAAGCAGATTTCTGCGTTGAAAGAGCATGGAGACAGACTTCGTCAAGATG
AGAGAGCACGTGGGGCTGGAGGGATGGCTCAGCACTTCAGAGGCACTCACGCTCTTCCAT
AGGACCTAGGTTCACTTCTCAGCACACACATGGCAACTCACACCTGTGATGCAGAGAA
(SEQ ID NO: 1)

eHGT_303m (392 bp):
TCCTTCTCAGAACCTAGGTAAGGTAAGTTCCTTTCAAGGCTGGCTATATAAATAATCATCTC
AGTTAGGATGCTTGGTGGGACCAAAGAACCAAAACTGCCGAGCAGGCATGATCTGACTTG
GAGTGGTTCCAGGACCTTCCTGTGAATGCTGGAGTCATTCAGTGTAGAGCTCTCCTCTGTG
ACTGGGTCAAGGTTGCCCCACTGTAAACCCAGGGAAGCTAGCCCAGCCTTCCTCTCAGGG
AATGTGTATGCTTCCCTTACACCTGACCCTGGCACAGACCTGGTGGTTGTTTTTCAGAAGC
ATCAGTGTCTTTGCCTTAGGCATTTGTCCTCAAAGGGCAGCGACACTGTCTACTGACTGCT
TTGTACAGGGTAACTGCTTAACTAATT (SEQ ID NO: 2)

eHGT_307m (339 bp):
AAATGGAGACTGCCAAGGGCTGAAACGGGGTGCGGGAACCAGGGACCGAGCCCCCCCCC
TCCCCACATGAGAATCTGTCACATTGCTGCTCCAGTGGCCTGAAAGACCAGCACAGCCCC
AGCTGGAGCCTCTCCCCTCTGGATCTTGTCAATGTGGCTTTGCTTTGCTGCTTGGGCAGCC
GGGAGTGGTGACAAGCAGGGAGAGAGCGCCCAAGGCATCTGGCTGTGCCACTCCAGCCT
GACTGCCAGCTCACCCATCAGTGCCCATCTCATCATCGAGAGGGACCCAGATGAGACCGG
GGATCAGCACTGTCCTTACCTTGAAGGGACGTGTCAGGAA (SEQ ID NO: 3)

eHGT_308m (408 bp):
TCTGCTTTCTCTTCCCTTGGCCTCCTGCTGGGATAGAAGGGGTTGGTGTGAGTGTGTATGG
TGGGGTGCTGTGAGATTAATTAGCAGCCGTGCCAGGCAGCAGGCGGTGGGGTGCAGAGT
AGGCTGGCTTTCCCTGCTATAGATCCATGCTCTCTGGGAGAGGCACTAGCCGGCTGCTTT
GGGCTCTGGCTCAGCTATTTTAGGAATATTCTTAACCCTTCCAGAACCGCTGCCATTGCCA
GATCTCTCTCCCAGAACACAGGCCAGCTCCAGATTGCCCCTCCTTTCTGCCCCCGCCCTG
CACCCCACCTAGCCTCTGCTCTTCCTCCCTACAAGTTGAGAAGGTCAAGGTTTGACTTTTA
CCAAAGAAAACTCCTGGCTCCTGATCCCACTCTCTGTGCTTTACCT (SEQ ID NO: 4)

eHGT_472m (713 bp):
GGGATGCGATGCAGTGGCATGGTAAAAATGCCGTACCTAATGATGCAGTTTCAGCCATCAT
GACGTATGGATGGAATGAAAAGATGACAAAGCTACTAGCGGATCTGTTCACAAAGGAAAAG
AGGCTCAGTTTCCAACAGTCTCTGGCATTTCATCTTTGGAAAACTTGTCAAAGCTAAGATGA
TTTGTGAGAGTCCTAACTCATTTTCCTAAATATGAAGTCGCTATTATGGATAAAGAAAGGTT

FIG. 14 cont'd

ACAATGGAGCATCCATTTCCACTCTAGTCATTCTGCTTGATTGCACCAATTAGCCCTGCATT
CACTGTGCATTTTAACTTCATAATGGTCTATTATTTGGACCACTGACCATAATGATGCAATTC
TTTCCTTGACGAAAATAAATGCTTAGTGATAAATAAGTAATCATAATTAAAGCTTTCATAGTA
CTTATAGTACCACAATTGGTATTTAGCCCATTGATTATAATTTAAACACATTTTAATTAATATG
TAACTATATTACTTCCCTATTTTTCTTTTCTTTCCTCCAAAATGCCTCATGTCTCCTTCTATCA
ACACCCCCCATACCCCCCAGACACTCCTTGTCAAATTGATGGCCTTTTAAAATTATTATTAT
TGTAACATGAAGAAATAGTTGAACAAATATATAAATACAGTACAATGAGTCTGTTTATGGTG
GCTCCTATGGGTATGATTTGGGGGCCT (SEQ ID NO: 5)

eHGT_475m (627 bp):
TATCTTAGAGTGGGAAGATTTGAGAAGTGCCATGGTTAATATGACTGACTTTTTATTCTTATT
TCTTTTAATTTCATGGTTCTAAATCCGAATTTAATCATAGTACCCAGAAAAGCAGAGGTGTA
GAGGTTCACAGTGGGAGTTGTAATCTAGCCCTATTCATTTTGACCTCAAAACCCAAATTATT
TATAACAAATTATTTCCTATTCTTTCCTTCACTATTCAGGAACATCTGTCCACCACTTACATG
ATCACTTATCTTGCTATTGTGTCATTTTGATGAAAAAGAATTTTTTCTAAATATCTAAATACAA
GGCCCCATATTAACAGTGCTTTTTAAATCCCCACAGATGTGGGAGATGACCCCTTTCCATC
CCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTTGTTCCAATAAAGAGATACAATTTT
ATTCATTAATTTGTGTATTCATTTAGCAAATCACTTTAGAGTCTTATTATATCAGGATTTTGG
GGTCTATTTTAGTATATCTTTTTGTATTTCTTGGAACCTCTCCAATTATTCTAGACTCTTTCAA
AGGTTGGTGATCAATATTAGACATTATTATGAAAAGAATCTTACTTGCTAAAAGGGTTAGAT
G (SEQ ID NO: 6)

eHGT_476m (486 bp):
TATGATGTGCCAGGCTTGGGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAAC
TTCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCC
GCATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCT
TCACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATT
ACTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAA
ACTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTT
ATTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGA
ATGCAAAAAAATATTGAATATGAAAAGAAAAACATTTGCAGTAAAGCTTGTTCT (SEQ ID NO:
7)

eHGT_476m core (439 bp):
AGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAGCTTGAAATATGGGAG
AGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGAGAACAATTCTCCCTGC
TCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTTGAACAAGCTAGTTAAC
AACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAACCACCAGTTACAAAATG
AGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATACTTTTAGCAGGCTGATT
TCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAGTTGATTGGTTGGACTTG
TGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAAAAAATATTGAATATGAA
AAG (SEQ ID NO: 8)

3x_eHGT_476m core (1317 bp):
AGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAGCTTGAAATATGGGAG
AGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGAGAACAATTCTCCCTGC
TCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTTGAACAAGCTAGTTAAC
AACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAACCACCAGTTACAAAATG

FIG. 14 cont'd

AGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATACTTTTAGCAGGCTGATT
TCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAGTTGATTGGTTGGACTTG
TGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAAAAAATATTGAATATGAA
AAGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAGCTTGAAATATGG
GAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGAGAACAATTCTCC
CTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTTGAACAAGCTAG
TTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAACCACCAGTTACAA
AATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATACTTTTAGCAGGC
TGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAGTTGATTGGTTGG
ACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAAAAAATATTGAAT
ATGAAAAGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAGCTTGAAA
TATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGAGAACAAT
TCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTTGAACAA
GCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAACCACCAG
TTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATACTTTTAG
CAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAGTTGATTG
GTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAAAAAATAT
TGAATATGAAAAG (SEQ ID NO: 9)

eHGT_503m (988 bp):
AAAGCTACATCTCTGGGCTCCATTATTAAAGCTGCTTTCCTTCCTTTCTCTCTCTCTCGCTC
TCTCTCTCTCTCTCTTTCTTTCTTTCTTTTTCCTGAGATGGGTCTCATGTAGCCCAGGCTGG
CCTCAAGCTTGCCACATAGCCAAGGAGCCAAGGATGGCATTGAACTCCTAGATCCTCCAG
CCTCTGCCTCCTGTTAGGATTAGAGGTGAGCTACAATCCCAAGGGCCTTTAGTGTAACAGT
CAAAAGCTACTGGGAGTCAGGCAATTGGCTCAGTATAACCCTGATTCTCCTTTTGTGCCCA
GGCACGTTGGTCAGGAGTCTGTCTTCAGCCTGTCATGGCAGCAGCTCAGCTTCAGTGACC
AATCTATACTCACTCACAGGAGACTCTGAAATCCCAGATTCTGTGCTATAAAGTCCCCGCTC
GAGTGAGTCGTGACTGCTCCAAACAGCCTGGGCAGCTGCGAACCCTCATGGCATCTAGGT
GACCCTGTTCATCCTACAGCTGTTCTCACTGAGGGGAGGGGAGCTTTTGAGTGAGCCAGT
CAAAACTCTGTGCTCGGTGATCCTGTGAGGCTCGGAACGGTGGCACCCGAAGCCATGGGT
GCACACACAAACAGGGCTCTAATCGGTGGGATCACAATCCATGAACAAGCATGAGACCTC
CCTTCTTCTCACACACACACACACACACACACACACACACACACACACTCACTCACATACATGAG
CTGGTTTCCACAACTGTGGGGTTAGCCTGGAAGGTGTCTGTCCTATATAGTACTCCAGTAC
CAGTGTTGCAGACTCTAGGCCCAGGAAAAGGTTTTTGATGTTTGCTGGTGTTTTTGATGAC
CTTCATCGTGGGTAGAGCAGGCTGCGCTGGTTCATAAAGAGAAGACAAGACCTAGAGTGC
TGTGACCCTTTAAGGCATCATGGTGTGATGACCCCCAACCATAAAGTTATTTTCGCTGCTG
CTTTGTAACTGTAAGT (SEQ ID NO: 10)

eHGT_571m (428 bp):
TGAATCTTCATGGAGAAAACATCATGAATAGAGAATGAAGAAGTAAAGATGAACAAGTGAG
AATTCAGATCATCAGAGGGTTTTCTAGGATTTCTGTAAATTTCTCTGTGTTTTGGAATGCAAT
AGGAATGCCAGCCCAAAGCCATATAGGTCACAGCTGCCCAAGAACAGTTACCAATACAGTA
TAAATAACGTCCTAAACTTAGAATATTGTGAAATTCTTTTTATAACTCAGCCATTTAATTCTTG
AGTGTAAATTTTGTAGATGGAGTCATGTTGTAATGTTAGACACATTTGCTAGTGATGTGACA
ACATAATATTCCCATGAACTGATGTCAAATGTTGTATTGTACTTTGACCAGGTATATAAGGTT
TATTATCTTCTTGACCTTGGAGTAATTTCAGTCCCAACTTAAATCCCTAGTGGT (SEQ ID NO:
11)

FIG. 14 cont'd

eHGT_706m (352 bp):
TATGTAGATACAGGTCATAGAACTTGCCCTGGGGAATGGCTCCATTTGGTACCAACAGGCT
GACCCCTAGGGAGGAAGGAAGGCTATCAGCAAGAGGAGGAGGAGGTAGCAGAGATGAGA
AAGATGGGGTAGACTCTGGCTCCAACCTAGGGAAGGGAAAGACTCTAGACTCGGGGGTAT
GGGGGTGGATAGATACAGGGAGCACACAGGCTACTTGGCCTGGTCTGCCCATGAATACAG
GGGGCCTCTAACATTGCTGGGGTAGGAGGGTCAGAATGCTCCAGTGCTAGCCCTCATGCT
GGCTCAGGACAGGACTCTGAAAAGCCACCAGCTGCCACTTTCACAAGCTGAG (SEQ ID
NO: 12)

eHGT_710m (382 bp):
AGTGACTTGGTGCTATGAGCCATATTTTGCTGTTGCTGTTGTTACTGGTAGTTTTTGTAATT
CTGGGGCTAAAACTTGGGGTCTGGTATGCTGTCATTTACCAGTGAGCTATACCCTGGATAT
TATGATTTAGATGAATGTGAAATATCACCCCAGACATACATATACTAAACACTTGGCCCTTG
GCCCATGATGCTAAATGGAGGAGATAGAAGCTTTTGGGGCACAGCCTAGTGGAAGGAAAT
GAGGTCAAATGACATGTACTCTGAAAGGAATATGGGTATTCTGGGCTTGCGTTATTCTCTCT
CTCCCTCTCTCTCCCTCTCTCTCCCTCTCCCTCTCTCTTTCTCCTTTTCTCTTTCTCTCCTCG
CCTTGTTTTCCA (SEQ ID NO: 13)

eHGT_296m (912 bp):
TCCAGCTACCACCAGCCTGTCCAAAAGGGGACACCAAAGGAGAGGAGGAAGTCTGAGAGA
CACCTCTCTGCACCATGGCCATCTTAGTAGTCAGACCCAGAACAAAACTCTCTGATGAGTG
TGCGGAGCGTCGCTTCTGGTCTTTGCTCAAACTATTCACTGAAGATTTAAAGCAATCCCGT
GAGTGATACATTTGGGTGAATTTGTTCTCTAGAAGGTATCACAGAAATCTGGTCACTGGGC
CACCCGAGACATCCTGATAGGCCCTCTGGTAACCCATCACATGCTGCAGACTGACTCTGG
GGGCCTAGAACCCAGATCAGAAGCAACCTTGACCCCGGCCCACCCGCCACGGAAGCACC
ATCATCTCTCTGATTAAAAACCTCGATCACGGACCCGGGGGCGTGCCCGGAAGAGCTAAG
ATAATCAGCGTCAGCACTTTGCCTTCGCCGTCCAAGACTGCAGACGGCCTTCATTTGACCT
GATTCGTGGTGTTAATGACAGCAGAGCAATTTTGAGAGGCAGCTTGCTCTCGGCATCTATA
AGGAGAGGAAAAGCACTGAGGGCTGGGGACCAAGCTCCTTGCAGAGGCGGCAGCTGCAG
TCACCCTCCCCCTCCACCCCTGCCCCTCCCCTCCCCTCCAGAGGCACTTTGAGTAAGTGC
TGCCCTCCGATCTGCCCTGATACGATGGGAGAAAGCTGATGTGAGGGCTGGAGCCAGAGT
GTGCAAGGGGACAGTGTGTGCATGTGCGTGTGTCGGGGAGAGGTACCCGTGCTATACCT
GAGAACATTGCTGGGTGAACACAGCCTTGGACCTGGAAGAGCGCATAGCTTACTTAGAGG
CATGGGCTGCACATGAGCTGCCCATTTACCTGCTCATTTAGAAGCTACTATGAAGGCTGGT
GAGATGGCT (SEQ ID NO: 14)

eHGT_299m (481 bp):
ACCAGAAGTTCAGTGAGCAGAAGATGGGCTAAAATGAAAAGGGTACTGTCTTGAACTGAAG
ATGGAATCCTGCAGCTTCATTCTGGCCAAAAGAAGATCTATTCCCAGGAGGAGGGTAAAGG
CTTTGTTCTTAAGAGATGCTGAGGCTGGCCCTGTGAATCTGATGTCAAGATGTCCCTTGTC
ACTCTGCAGAAGCGTATGTCTCTTGCATTTCCTTCTTATTTCCTTGGGTGAAATTGCTGTGG
CATTGTGTCACTCATCCTAATGGGTCATGTCTAACATCTGCGTGCTTACAAATCAGGCATGC
TCATTTCTGGGCTTATGGAGCTTGTATAACACCAGGACAGGCAAGACATGTTGCCCACTCA
GGAAGAATAGAAGCTGGGCACAGCTGGAGTGCAAAGTAGGTCAGTTCAGAGAGCAAAGG
GAGTTGATGGAGCAATGAGTTGTTAGTGGGAAAGTTCTAACCAACTGTCCCAGA (SEQ ID
NO: 15)

FIG. 14 cont'd

eHGT_300m (990 bp):
TGTCTTAGACTGAGTTGCTGTAACAAAAATCTGGGATGAGGTCATTTCTAAAGCACAGCAAT
TTATTTCCCACAGTCTGAAGGCTGCAGATTCCAAGATCACTGGCAAGATCAGTTGTAAAGG
CTTCGTCCTCCAGAGGCGGGGATGCTGCATACTCCCTGGGCAGAGAGACAGGAAAACTCC
GTAACTGCATGTGTCCTTCCTGATGCCTCTTCTATATAGGCCTGGATCCCAATCACCCTGT
GAACCTCTCCCTGCTTCGTGGCCCCCACCTCTTAACACTACCACATTGGCAACTCCTGAAA
TTTGAAGGGGACACACTGAACCATGGCACAACAGCTTTCTGACTGATGCAGTAACCCAATG
GCAGTGCAGAAGGGGCCAGCTAAAAGCCCAAATGGTTAGCTCAAAATTCGCTGTCTCTTCC
GAGTGTCTGAACCCTTAGTCCTGGTATGTAAAGACATCAGAACATTTCCCCTTGTGTCCATC
AGATTTCTGTCTAGTGAAACGATGACACTGTAACCTCCAAGATCTCACACGAAATGATCTTT
TCTCCTTTGTGGAAGGAAACCAGCATTTAGCTCATCTCTCCTTCGTAGCAGCTCAGAATGT
CCACAGTGACCCAGTTACCATAGCTAAAGGCTTCCTTTTCAAAACACAGAGCAGAGGCAGC
CAATTCAGTATGTGCTGCTGCCATCCTCTGATTCTTTCCTGCTTCCATAGACACCAACTCTA
TTGTAACTAAGCCTTATACATTGTGTCTTCCTCCTTTACATTAGCTTGTGCTGGGGTGGTTC
ATGAGGCCCGCTGAGTAGTTTCAGTGACAGCCTATCCCTCTGCCAGTGCTGCTTTGAGCC
ATCTTATTGGTGAGGCTGTAAGAGAAGCCTGAAGTCACAGGGTAAAGCTATGTTGAAGGCA
GCCCCAGAACCAAGTTTCCCTATTTCTATCTCCTTACGCTGTTTGAGCCTCAGGGGTAGAT
CAGGTGCCTGT (SEQ ID NO: 16)

eHGT_306m (860 bp):
AGATGAGTCTGCAGCTGGGTACAGTGACCTTTCAGTCCATGTTTATTTGGAAATGACCTTTA
AGCAGCAATCAGCAAGAATAAAATGCTTCAAAGGAATACATTAGATAATGCAGAAGTCCCC
CAGAGGTAAGTTACACCCAAGGCACCCAGCTGACAATGAAAGTGGCCCTGCCCTGGGAAG
CCAAGGACTAGGCCACCCATTAGACTAACAAGTGAACACAGGGTCCCCAGAGTTTGGTCT
AATAGACAATGGGGAGTCTGAAGACAGGGTGACCTGGGCAAGACACAAGAGCAGTTCCAA
AATTAAACCTCTGTCGTAATGAAGGATGCCTAGTTGTGCTTTTTCCATCCTAGGATGGGGAA
TCCTCAAGGGCAGGGCACAGCTGTGCCAGGGGAACTGTACGGGCTCCATCCTGCCTCCC
TCCCATGGGGTGAGCTGATAGTCTTCCTCATACTGAGCTCTTGTCTCTGCTGTGTGCTGGG
GAGTCTGAAATGCTAGAGAAACTAAGCCTTCCCACTCAAAGACAGAGAAAGAGCTGGCCC
ATGGCTCCGTGCCCTCTCCTCTCTCTGTGCGTGTCTTTAACTCTGTATGTTCTATTTTCCCC
CTCCTCGTCCCCTGCTTCGCGCTCACAGAGTCACTCCTAGTAGCACCAAAGAGAGATGCTT
GGCAGTTCACTAACCCCTTGAGCTGAAATAGAAATAAATATCCCAAAAGAGAAATCAGAAAA
GCAGGGTGTCGCGCTGGAGAAGAGGCAGGAAGATCAGAAATACAAGGTCATCTGTGGCTA
CACATCTAGTCCAAGTCCCAGCCTGGGCTATGTGAGATGGAGGGGAATCGCTCAGAAACA
AGGCTGTACA (SEQ ID NO: 17)

eHGT_309m (778 bp):
TTCACCCACCTGACACTTGGGTTAGACCTGAATGTCGTTTCTTTAACTCACACTGCTCATCC
CACTGGCCTTTGCTGTGCTTCTCTGTGCCTCCTCAGAGATACATGAAACTGTCCCATCCCC
CTAACGATGCTGGATGGATGGCTCCAACAGCTCACTGCTCTCACCTTGACACAAAGTCCTA
GCGTCTGCATCTGTGAGACAAGTTGGAATTTATATATTTCCAGTGGAGATTAATAATTCATT
AGATGCTGAAGTAGAAAAACAAAGTACCGATTAATCAAGGCTCTGCTGAGGCCTGCTTTGC
AGCCACCAGTCTGTGGGGATTGGCAGTGCTTTTACACTGGAAGTAGGTCAGGACCACAGA
AAAGCAGCTCTCATGCACTAGCATCTGTTCGCACTAATCACTGTACACAGCTTTGGGTCTTA
CTATAGTTTTTATTAGTTATCCCAGCTGGGATTTATGTCTCAGGAATAAAGAGCCAAGAATG
GGAGGAGTTACCCTCGAAAGATCCAGGTCATGTGGTGCAGGGCAGGGAATATGGCTGACT
CAATCTCTTTGCCCATAGAGCCTCAGAGTATCAGATCTTAGCACTCTAAGGAGGGAGACTC
AGAGGGTACAAGTCTTAGAAGTCTCCCTAGGGCTTGGTGCCCAGCAAATATATGCTGTTTG

FIG. 14 cont'd

TGACTTCCCTAATACCAGGTACAGGCCAACACAAAGGACCTGTCCAAGGGAAACTCACGG
CTCAGACCTGATCTATTTACAGGTTGAGTTTGGGTGAAGCCAAGA (SEQ ID NO: 18)

eHGT_310m (549 bp):
ATCCCTGGAGATGAGGAGTCCTCTCTGGCAGGGTCCCCTCACTCTAGAGCAGCCCCTATC
CCAGGCCCCCTAGGAGTCTCTAATTAAAGGGCCGGCACGCCCCTCTGGGACTCATTAGGC
CCGCTGTGCAGAGAACATTTAATCATTGCTCAGAGCATCGATTGGAAAATCAATTTCTTTGT
CTCTTCGCACGAGGCGCGCTGGAGAAGTGGGGGGAGTGCTGACCTCCTTCTGCTGCCGT
GTAAAGCGCTGCACATTTAATCAGGGAACAGAAATCAATTAGCCACTTACGAGGTTGGCTT
TAGTTACCGAGTCGGCAAGGCCCGCGCCACAGCTCAGCCGCTGACAGTAGCGAATCTCCT
CCTCTCGGCCCTGCTGCATGGCTCTGTCTCCCTCCCTGTATCTCTCTGGCTTCCTTCTTTC
CCAGAGTGCTCTGGGTTCTCACCATCTTGGCAGATCCTCACAGAACTCCAAACAAGTCCCG
AGAAGCCTTCCTAATGCCCAGTCTCCTCGGCCACCTTCTTGTTCTCAGCTCTAGACGTTTC
AAGA (SEQ ID NO: 19)

eHGT_890m (575 bp):
TGAGCTTCAACCAAATCAGGCATTGATGGATTTTATAGTTTGATTAACAAAGATAATAGCAA
ACCCCAGATTTAGTTTAAACATAAAAAGTATTAAGGTTGTATCCTGCTTGTATAGCATATGCA
AATGACCTCGTTTCTGCTACTGCATTTGGAAATGTAGCAGAAGAAAAAAAAAAGGCACTTCA
ATTGCAGCTCTCATCAGTTATTCACTGTATCCAGGCCTCTCAATTGTGTTCTTTTCTTTAATG
CAATAGCAAGCAGCAATCACCCAGCTGTGCTTGGTAGAGTGAACTATATACACATCTATATT
GAGATTTCATACACACATAACATAAAAGCGAGAGAAAAAGCCTCAAGAATGTTTGGCCCATT
GCAAATCACACAAAAGGACTAATGAATCTCTCTCCAAATGGATCTGTAGTGACCATCTGTAA
GCCTTGATTGATTCATATTCCATAACGGTATCAGCATCCAGGAAGTGATTACTTCAAGGTGC
AACACAACTTCCCCTATGAAAGCTCAGTCTCTTTAATCATACCTAGTCAGTATCTGTCACGG
GGATAAACTAAGGCA (SEQ ID NO: 20)

eHGT_891m (611 bp):
ACATTTGCAGTAAAGCTTGTTCTTTTTCTTGAAGTATATTTTAAGATTTTGAGTTCTACTATCA
TTAAAGACAGATAATTAATAGTTTATTTTTATTTACTTTTGTTAGTAGTGACTTGGTGCTATGA
GCCATATTTTGCTGTTGCTGTTGTTACTGGTAGTTTTTGTAATTCTGGGGCTAAAACTTGGG
GTCTGGTATGCTGTCATTTACCAGTGAGCTATACCCTGGATATTATGATTTAGATGAATGTG
AAATATCACCCCAGACATACATATACTAAACACTTGGCCCTTGGCCCATGATGCTAAATGGA
GGAGATAGAAGCTTTTGGGGCACAGCCTAGTGGAAGGAAATGAGGTCAAATGACATGTAC
TCTGAAAGGAATATGGGTATTCTGGGCTTGCGTTATTCTCTCTCTCCCTCTCTCTCCCTCTC
TCTCCCTCTCCCTCTCTCTTTCTCCTTTTCTCTTTCTCTCCTCGCCTTGTTTTCCAGCTGCCA
GAAGGTAGGCCTCTTCTCTGCTGAATATCTGTGTCATGTTATGCACCAACACAGTACTAACT
GTCATGTTATACCTAGTGGCCAGGTAACCATGGACCAAAATGGCAGAGCA (SEQ ID NO:
21)

eHGT_892m (661 bp):
TGCAAAATAAAGATTTCTTGGGATACAGAGAAAAAAACAAATCTGACAGGAGAGGAAGAAG
CACCCGGTGGGCTATAACGGTGCAATTCAGCTGATTATATGTTACAAGTAACAAGGACGAG
AAAAAATGTTATTTCTTTGAAAATAAAACTAACCAGGCCATACATATTTAACAGGACTGCATG
AGAGAAGAAGAAGCCAGCTGCAGGAGTGACTGTGGGGGGGAGGGGGAACTTGACAAAAA
AAGCAAAATGGCAGTCCTGCTTCCAAAGTCCTCAAGGTCACAGTTATTTGGGCATTCTTGC
GGGCACTGCTTATACAAGAATGTGCTTTCAGTCAAGGCTTTCTAATAGATTCTCAAAATTTG
GGACAAATGTTATTTTTGTATCTGTAGAAATGTACTGATTCAGAAAGATCTTTGAGCAATACA

FIG. 14 cont'd

GATGTTAAAACATTTAAGTCACAAAATGGGTCTATTTAATCAATGCGACTAGTTTGGAACATT
ATTCAAACTGCCAGAAATACAATGTAAATGAAACCTCAGGCCAATATTTTGGAGCCCTAAAA
GATTTGATGGCTAATTTTATCGTAGACACTAATTATAAATAGGAGACCCCAGGATGGGACTA
GAAAACCAAGCCAGCTTTTTAATTTACCCCTCCAGGACTTTGCT (SEQ ID NO: 22)

eHGT_476m (1316 bp):
AGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAGCTTGAAATATGGGAG
AGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGAGAACAATTCTCCCTGC
TCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTTGAACAAGCTAGTTAAC
AACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAACCACCAGTTACAAAATG
AGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATACTTTTAGCAGGCTGATT
TCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAGTTGATTGGTTGGACTTG
TGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAAAAAATATTGAATATGAA
AAGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAGCTTGAAATATGG
GAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGAGAACAATTCTCC
CTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTTGAACAAGCTAG
TTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAACCACCAGTTACAA
AATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATACTTTTAGCAGGC
TGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAGTTGATTGGTTGG
ACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAAAAAATATTGAAT
ATGAAAAGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAGCTTGAAA
TATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGAGAACAAT
TCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTTGAACAA
GCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAACCACCAG
TTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATACTTTTAG
CAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAGTTGATTG
GTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAAAAAATAT
TGAATATGAAAA (SEQ ID NO: 23)

eHGT_1022m (505 bp):
CAGTTTCCAGCGTGGTTGTTGATGAGGCTCAGAGAAAAGACTCTAAAGTTATGATGGGAAA
TTACCATGCCATTCATCATCATACACATTCACCTCACACTTTCTGAGTCTCCTATACAAAGTC
AGTTCTCTGCCAAGGGCATGGAAGAGCGAGGAACAGGATGTTAGGAAGGGCTGACAGCG
CTGTTTTAGCCTGACAGGCAGATTTACAACAGGAGAATGAATGTACCACTTGTATAAGAAG
GCCATGCGGCACTGCTAATGCACAAGTTGGCAGTACATCAACATCTCTATCGTCCTCATAT
TCATGAAGCAGAGAACGGAAATGGCACACTGCTTGTACCGGCGAATAACCAAAGTGAACG
CCCTACGGCTGCCATTCACTGTGTCCTTCCAAAAGCATTTTTCTACTGAGCTCTTCCCAGA
GATTTAGGGTTTGCTTAGACAGGTCTTATGACGCCACGTGATAGGTCATTCTTCTGTTCTGA
GGAGCTTGGAGAAGATC (SEQ ID NO: 24)

eHGT_1023m (439 bp):
GTAGAACATACTTATTAACACATTCGTACATAAAATAAAATTCTACTCTCCCGACCTTTTCCT
CACCATCTTGCTTTTCAACGTATGGCGTTAGACCTAACAGCGAGTCCACTTCTTCCCCTTTC
ATTCTGTAGCAAGAACACACGGCTCACTGTAACAGGGACTTGGCTGTGGGTTGCAGACTG
GCTTCCTGCTGCCTCCACTTGAGCCCCACACAGCTGTGGCTTTGTGTTTACAACCCTCCAG
GCTGCCATTCATTCGGTGCTGTGGGCTCATGTACTGGAAGACAGCTTCCATCACAACCTTC
CCGTCCCAGCAGGAGAACTCCCTTGCTTCCTTGGGGAACATTTGCTTGCTCCTGCTGCTTG

FIG. 14 cont'd

GCTCTTCCCACTTTTGCCTCACTCTGGAGTTTCTCTCTCCCGTTTTGAATTCTAGTAGTAAA
CACATGGCC (SEQ ID NO: 25)

eHGT_1024m (507 bp):
CCAGTGAGATCTTCCACCAGCAGAACTCATGGACACAAACTAGACAGCTCACTTCTTGCCT
GTATTCCAGGAGTGGCTTTTTCTCTACTCCTGTACTGATGCCAGTCATTCAGAGTGCACTCA
AGACACTTGACCCACATCAGTTAAGAGAATGAAAATCAAGCTCTGAAAGCCATTAGCTTCTA
TTGCACACCCAGAAAACAGGCTCATCAAACACCTTCTTATGGTAATGCCTTTGATCAAAAGG
AGGGTTAATTCAACAAATGGTTTGCACCGTGACCCCATCAAAGCCTGAGCACCAGTGTCCT
CATTTCCTTTCCCCTGGTGTATAATGAGTTGTTAGTCTGGCTCACCTTGTCATCCCCATCAT
ACTGCCATAATCCACATCTCTAAAGAGTGGATTACAACAGTCCCGTCTGTGACACTCAGGA
CTGGCATCAAGGTTCCCAAGCTCTAGTCTATTGTGACATTGATACAAATAGGGCTCAGAGT
CTCACTGATCACACC (SEQ ID NO: 26)

Beta-Globin Minimal Promoter (pBGmin/minBGlobin/minBGprom):
GGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTG (SEQ ID
NO: 27)

minCMV Promoter:
GAGGTAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGA
TCGCCTGG (SEQ ID NO: 28)

Mutated minCMV Promoter (SacI RE site removed):
GAGGTAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGA
TCGCCTGG (SEQ ID NO: 29)

minRho Promoter:
GATTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTG
GAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCC
CC (SEQ ID NO: 30)

minRho* Promoter:
GATTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTG
GAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCC
CC (SEQ ID NO: 31)

Hsp68 minimal Promoter (proHsp68):
CAGGAACATCCAAACTGAGCAGCCGGGGTCCCCCCCACCCCCCACCCCGCCCCACGCGG
CAACTTTGAGCCTGTGCTGGGACAGAGCCTCTAGTTCCTAAATTAGTCCATGAGGTCAGAG
GCAGCACTGCCATTGTAACGCGATTGGAGAGGATCACGTCACCGGACACGCCCCCAGGC
ATCTCCCTGGGTCTCCTAAACTTGGCGGGGAGAAGTTTTAGCCCTTAAGTTTTAGCCTTTAA
CCCCCATATTCAGAACTGTGCGAGTTGGCGAAACCCCACAAATCACAACAAACTGTACACA
ACACCGAGCTAGAGGTGATCTTTCTTGTCCATTCCACACAGGCCTTAGTAATGCGTCGCCA
TAGCAACAGTGTCACTAGTAGCACCAGCACTTCCCCACACCCTCCCCCTCAGGAATCCGTA
CTCTCCAGTGAACCCCAGAAACCTCTGGAGAGTTCTGGACAAGGGCGGAACCCACAACTC
CGATTACTCAAGGGAGGCGGGGAAGCTCCACCAGACGCGAAACTGCTGGAAGATTCCTG
GCCCCAAGGCCTCCTCCGGCTCGCTGATTGGCCCAGCGGAGAGTGGGCGGGGCCGGTG
AAGACTCCTTAAAGGCGCAGGGCGGCGAGCAGGTCACCAGACGCTGACAGCTACTCAGA

FIG. 14 cont'd

ACCAAATCTGGTTCCATCCAGAGACAAGCGAAGACAAGAGAAGCAGAGCGAGCGGCGCGT
TCCCGATCCTCGGCCAGGACCAGCCTTCCCCAGAGCATCCCTGCCGCGGAGCGCAACCT
TCCCAGGAGCATCCCTGCCGCGGAGCGCAACTTTCCCCGGAGCATCCACGCCGCGGAGC
GCAGCCTTCCAGAAGCAGAGCGCGGCGCC (SEQ ID NO: 32)

SYFP2:
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCAT
CTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAG
AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
ACAAGTAA (SEQ ID NO: 33)

EGFP:
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCA
CCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA
AGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCT
TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAA
GAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCT
CGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACA
ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACA
TGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAA (SEQ ID NO: 34)

Optimized Flp recombinase (FlpO):
ATGGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTTCGACATCCTGTGCAAGACCCCC
CCCAAGGTGCTGGTGCGGCAGTTCGTGGAGAGATTCGAGAGGCCCAGCGGCGAGAAGAT
CGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGATGATCACCCACAACGGCACCG
CCATCAAGAGGGCCACCTTCATGAGCTACAACACCATCATCAGCAACAGCCTGAGCTTCGA
CATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGACCCAGAAGGCCACCATCCTGGAGGC
CAGCCTGAAGAAGCTGATCCCCGCCTGGGAGTTCACCATCATCCCTTACAACGGCCAGAA
GCACCAGAGCGACATCACCGACATCGTGTCCAGCCTGCAGCTGCAGTTCGAGAGCAGCG
AGGAGGCCGACAAGGGCAACAGCCACAGCAAGAAGATGCTGAAGGCCCTGCTGTCCGAG
GGCGAGAGCATCTGGGAGATCACCGAGAAGATCCTGAACAGCTTCGAGTACACCAGCAGG
TTCACCAAGACCAAGACCCTGTACCAGTTCCTGTTCCTGGCCACATTCATCAACTGCGGCA
GGTTCAGCGACATCAAGAACGTGGACCCCAAGAGCTTCAAGCTGGTGCAGAACAAGTACC
TGGGCGTGATCATTCAGTGCCTGGTGACCGAGACCAAGACAAGCGTGTCCAGGCACATCT

FIG. 14 cont'd

ACTTTTTCAGCGCCAGAGGCAGGATCGACCCCCTGGTGTACCTGGACGAGTTCCTGAGGA
ACAGCGAGCCCGTGCTGAAGAGAGTGAACAGGACCGGCAACAGCAGCAGCAACAAGCAG
GAGTACCAGCTGCTGAAGGACAACCTGGTGCGCAGCTACAACAAGGCCCTGAAGAAGAAC
GCCCCCTACCCCATCTTCGCTATCAAGAACGGCCCTAAGAGCCACATCGGCAGGCACCTG
ATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAGCTGACAAACGTGGTGGGCAACTGG
AGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCACCTACACCCACCAGATCACCGCCAT
CCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTACGCCTACGACCCCATCAGCAAGGA
GATGATCGCCCTGAAGGACGAGACCAACCCCATCGAGGAGTGGCAGCACATCGAGCAGC
TGAAGGGCAGCGCCGAGGGCAGCATCAGATACCCCGCCTGGAACGGCATCATCAGCCAG
GAGGTGCTGGACTACCTGAGCAGCTACATCAACAGGCGGATCTGA (SEQ ID NO: 35)

Improved Cre recombinase (iCre):
ATGGTGCCCAAGAAGAAGAGGAAAGTCTCCAACCTGCTGACTGTGCACCAAAACCTGCCT
GCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAGGAAGAACCTGATGGACATGTTCAGG
GACAGGCAGGCCTTCTCTGAACACACCTGGAAGATGCTCCTGTCTGTGTGCAGATCCTGG
GCTGCCTGGTGCAAGCTGAACAACAGGAAATGGTTCCCTGCTGAACCTGAGGATGTGAGG
GACTACCTCCTGTACCTGCAAGCCAGAGGCCTGGCTGTGAAGACCATCCAACAGCACCTG
GGCCAGCTCAACATGCTGCACAGGAGATCTGGCCTGCCTCGCCCTTCTGACTCCAATGCT
GTGTCCCTGGTGATGAGGAGAATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAA
GCAGGCCCTGGCCTTTGAACGCACTGACTTTGACCAAGTCAGATCCCTGATGGAGAACTC
TGACAGATGCCAGGACATCAGGAACCTGGCCTTCCTGGGCATTGCCTACAACACCCTGCT
GCGCATTGCCGAAATTGCCAGAATCAGAGTGAAGGACATCTCCCGCACCGATGGTGGGAG
AATGCTGATCCACATTGGCAGGACCAAGACCCTGGTGTCCACAGCTGGTGTGGAGAAGGC
CCTGTCCCTGGGGGTTACCAAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGCTGA
TGACCCCAACAACTACCTGTTCTGCCGGGTCAGAAAGAATGGTGTGGCTGCCCCTTCTGC
CACCTCCCAACTGTCCACCCGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGAT
CTATGGTGCCAAGGATGACTCTGGGCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAG
AGTGGGTGCTGCCAGGGACATGGCCAGGGCTGGTGTGTCCATCCCTGAAATCATGCAGG
CTGGTGGCTGGACCAATGTGAACATTGTGATGAACTACATCAGAAACCTGGACTCTGAGAC
TGGGGCCATGGTGAGGCTGCTCGAGGATGGGGACTAA (SEQ ID NO: 36)

iCre(R297T):
ATGGTGCCCAAGAAGAAGAGGAAAGTCTCCAACCTGCTGACTGTGCACCAAAACCTGCCT
GCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAGGAAGAACCTGATGGACATGTTCAGG
GACAGGCAGGCCTTCTCTGAACACACCTGGAAGATGCTCCTGTCTGTGTGCAGATCCTGG
GCTGCCTGGTGCAAGCTGAACAACAGGAAATGGTTCCCTGCTGAACCTGAGGATGTGAGG
GACTACCTCCTGTACCTGCAAGCCAGAGGCCTGGCTGTGAAGACCATCCAACAGCACCTG
GGCCAGCTCAACATGCTGCACAGGAGATCTGGCCTGCCTCGCCCTTCTGACTCCAATGCT
GTGTCCCTGGTGATGAGGAGAATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAA
GCAGGCCCTGGCCTTTGAACGCACTGACTTTGACCAAGTCAGATCCCTGATGGAGAACTC
TGACAGATGCCAGGACATCAGGAACCTGGCCTTCCTGGGCATTGCCTACAACACCCTGCT
GCGCATTGCCGAAATTGCCAGAATCAGAGTGAAGGACATCTCCCGCACCGATGGTGGGAG
AATGCTGATCCACATTGGCAGGACCAAGACCCTGGTGTCCACAGCTGGTGTGGAGAAGGC
CCTGTCCCTGGGGGTTACCAAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGCTGA
TGACCCCAACAACTACCTGTTCTGCCGGGTCAGAAAGAATGGTGTGGCTGCCCCTTCTGC
CACCTCCCAACTGTCCACCCGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGAT
CTATGGTGCCAAGGATGACTCTGGGCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAG
AGTGGGTGCTGCCACCGACATGGCCAGGGCTGGTGTGTCCATCCCTGAAATCATGCAGGC

FIG. 14 cont'd

TGGTGGCTGGACCAATGTGAACATTGTGATGAACTACATCAGAAACCTGGACTCTGAGACT
GGGGCCATGGTGAGGCTGCTCGAAGATGGGGACTGA (SEQ ID NO: 37)

CreN-inteinN:
ATGACGAGTGATGAGGTTCGCAAGAACCTGATGGACATGTTCAGGGATCGCCAGGCGTTT
TCTGAGCATACCTGGAAAATGCTTCTGTCCGTTTGCCGGTCGTGGGCGGCATGGTGCAAG
TTGAATAAATTTGCGGAATATTGCCTCAGTTTTGGCACCGAAATTTTAACCGTTGAGTACGG
CCCATTGCCCATTGGCAAAATTGTGAGTGAAGAAATTAATTGTTCTGTGTACAGTGTTGATC
CAGAAGGGAGAGTTTACACCCAGGCGATCGCCCAATGGCATGACCGGGGAGAGCAGGAA
GTATTGGAATATGAATTGGAAGATGGTTCAGTAATCCGAGCTACCTCTGACCACCGCTTTTT
AACCACCGATTATCAACTGTTGGCGATCGAAGAAATTTTTGCTAGGCAACTGGACTTGTTG
ACTTTAGAAAATATTAAGCAAACTGAAGAAGCTCTTGACAACCATCGTCTTCCCTTTCCATT
ACTTGACGCTGGGACAATTAAATAA (SEQ ID NO: 38)

inteinC-CreC:
ATGGTTAAAGTTATCGGTCGTCGTTCCCTCGGAGTGCAAAGAATATTTGATATTGGTCTTCC
CCAAGACCATAATTTTCTGCTAGCCAATGGGGCGATCGCCGCCAATTGTTTTAACAAATCC
AACCGGAAATGGTTTCCCGCAGAACCTGAAGATGTTCGCGATTATCTTCTATATCTTCAGG
CGCGCGGTCTGGCAGTAAAAACTATCCAGCAACATTTGGGCCAGCTAAACATGCTTCATCG
TCGGTCCGGGCTGCCACGACCAAGTGACAGCAATGCTGTTTCACTGGTTATGCGGCGGAT
CCGAAAAGAAAACGTTGATGCCGGTGAACGTGCAAAACAGGCTCTAGCGTTCGAACGCAC
TGATTTCGACCAGGTTCGTTCACTCATGGAAAATAGCGATCGCTGCCAGGATATACGTAAT
CTGGCATTTCTGGGGATTGCTTATAACACCCTGTTACGTATAGCCGAAATTGCCAGGATCA
GGGTTAAAGATATCTCACGTACTGACGGTGGGAGAATGTTAATCCATATTGGCAGAACGAA
AACGCTGGTTAGCACCGCAGGTGTAGAGAAGGCACTTAGCCTGGGGGTAACTAAACTGGT
CGAGCGATGGATTTCCGTCTCTGGTGTAGCTGATGATCCGAATAACTACCTGTTTTGCCGG
GTCAGAAAAAATGGTGTTGCCGCGCCATCTGCCACCAGCCAGCTATCAACTCGCGCCCTG
GAAGGGATTTTTGAAGCAACTCATCGATTGATTTACGGCGCTAAGGATGACTCTGGTCAGA
GATACCTGGCCTGGTCTGGACACAGTGCCCGTGTCGGAGCCGCGCGAGATATGGCCCGC
GCTGGAGTTTCAATACCGGAGATCATGCAAGCTGGTGGCTGGACCAATGTAAATATTGTCA
TGAACTATATCCGTAACCTGGATAGTGAAACAGGGGCAATGGTGCGCCTGCTGGAAGATG
GCGATTAG (SEQ ID NO: 39)

SP10 insulator (SP10ins):
GAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAG (SEQ ID NO: 40)

3xSP10ins:
GAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGGAAGCTACCCC
TAACACACTATTCTACACACAGAAAATGCTCTTCACTAGGAAGCTACCCCTAACACACTATT
CTACACACAGAAAATGCTCTTCACTAG (SEQ ID NO: 41)

FIG. 14 cont'd

4X2C:
GCGGCCTTAAAGAGACCGGTTCACTGTGACAGTAAAAGAGACCGGTTCACTGTGAGAATG
AAAGAGACCGGTTCACTGTGATCGGAAAAGAGACCGGTTCACTGTGAGCGGCCTT**GAAAC**
CCAGCAGACAATGTAGCTCAGTA**GAAACCCAGCAGACAATGTAGCT**GAATG**GAAACCCA**
GCAGACAATGTAGCTTCGGA**GAAACCCAGCAGACAATGTAGCT**GTCGAC (SEQ ID NO:
42)
In this sequence, the miR128 recognition sequences are underlined and the miR221 recognition sequences are in bold.

miR128 Recognition Sequence:
AAAGAGACCGGTTCACTGTGA (SEQ ID NO: 136)

miR221 Recognition Sequence:
GAAACCCAGCAGACAATGTAGCT (SEQ ID NO: 137)

3XFLAG:
GACTACAAAGACCATGACGGAGATTATAAAGATCATGACATCGATTACAAGGATGACGATG
ACAAG (SEQ ID NO: 43)

10 aa:
TCCGGACTCAGATCTGGAGGCTCCGGAGGC (SEQ ID NO: 44)

H2B:
CCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAA
GGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCT
ATGTGTATAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGCCATGG
GCATCATGAACTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGTGAGGCTTCCCGCCT
GGCGCATTACAACAAGCGCTCGACCATCACCTCCAGGGAGATCCAGACGGCCGTGCGCC
TGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACC
AAGTACACCAGCGCTAAGTAATGA (SEQ ID NO: 135)

CCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAA
GGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCT
ATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGCCATGG
GCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTCCCGCCT
GGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCGTGCGCC
TGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACC
AAGTACACCAGCGCTAAGTAA (SEQ ID NO: 45)

WPRE3:
ATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTC
CTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG
GCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGC
CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGG
(SEQ ID NO: 46)

FIG. 14 cont'd

WPRE:
GCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTA
TGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAG
TTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCC
ACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCC
CTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGC
TCGCCTATGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCC
TCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTC
TTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCATCGATACC
G (SEQ ID NO: 47)

BGHpA:
CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGAC
CCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGT
CTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGA
TTGGGAAGACAATAGCAGGCATG (SEQ ID NO: 48)

HGHpA:
ACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTC
CAGTGCCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCC
TTCTATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACA
ACCTGTAGGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGG
CTCACTGCAATCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGT
TGGGATTCCAGGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGG
TTTCACCATATTGGCCAGGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGG
CCTCCCAAATTGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTT (SEQ ID
NO: 49)

P2A:
GGCAGCGGCGCCACCAACTTCAGCCTGCTGAAGCAGGCCGGCGACGTGGAGGAGAACCC
CGGCCCCGGAGCTAGCGGA (SEQ ID NO: 50)

GGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGGGGATGTAGAGGAGAATCCA
GGGCCTGGTGCTAGTGGA (SEQ ID NO: 51)

T2A:
(GSG)EGRGSLLTCGDVEENPGP (SEQ ID NO: 52)

E2A:
(GSG)QCTNYALLKLAGDVESNPGPP (SEQ ID NO: 53)

F2A:
(GSG)VKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 54)

FIG. 14 cont'd

Exemplary Plasmid Backbone 1 – Left ITR:
CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTCGGGCGACCTT
TGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCA
CTAGGGGTTCCT (SEQ ID NO: 55)

Exemplary Plasmid Backbone 1 – Right ITR:
CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCG
TCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTG
GCCAACTCCATCACTAGGGGTTCCT (SEQ ID NO: 56)

Exemplary Plasmid Backbone 2 – Left ITR:
CATGTCCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCG
GGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGG
AGTGGCCAACTCCATCACTAGGGGTTCCT (SEQ ID NO: 57)

Exemplary Plasmid Backbone 2 – Right ITR:
AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGG
CCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGA
GCGAGCGCGCAGCTGCCTGCAGGGGCGCCTG (SEQ ID NO: 58)

PHP.eB capsid:
MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDK
GEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQERLKEDTSFGGNLGRAVFQAKK
RLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDP
QPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVITTSTRT
WALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWG
FRPKRLNFKLFNIQVKEVTDNNGVKTIANNLTSTVQVFTDSDYQLPYVLGSAHEGCLPPFPADV
FMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLD
RLMNPLIDQYLYYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
NNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK
VMITNEEEIKTTNPVATESYGQVATNHQSDGTLAVPFKAQAQTGWVQNQGILPGMVWQDRDV
YLQGPIWAKIPHTDGNFHPSPLMGGFGMKHPPPQILIKNTPVPADPPTAFNKDKLNSFITQYST
GQVSVEIEWELQKENSKRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL
(SEQ ID NO: 59)

AAV9 VP1 capsid protein:
MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDK
GEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQERLKEDTSFGGNLGRAVFQAKK
RLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDP
QPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVITTSTRT
WALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWG
FRPKRLNFKLFNIQVKEVTDNNGVKTIANNLTSTVQVFTDSDYQLPYVLGSAHEGCLPPFPADV
FMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLD
RLMNPLIDQYLYYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
NNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK
VMITNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGWVQNQGILPGMVWQDRDVYLQGPIW
AKIPHTDGNFHPSPLMGGFGMKHPPPQILIKNTPVPADPPTAFNKDKLNSFITQYSTGQVSVEIE

FIG. 14 cont'd

WELQKENSKRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL (SEQ ID NO: 60)

tet-Transactivator version 2 (tTA2):
ATGTCTAGACTGGACAAGAGCAAAGTCATAAACTCTGCTCTGGAATTACTCAATGAAGTCG
GTATCGAAGGCCTGACGACAAGGAAACTCGCTCAAAAGCTGGGAGTTGAGCAGCCTACCC
TGTACTGGCACGTGAAGAACAAGCGGGCCCTGCTCGATGCCCTGGCAATCGAGATGCTGG
ACAGGCATCATACCCACTTCTGCCCCCTGGAAGGCGAGTCATGGCAAGACTTTCTGCGGA
ACAACGCCAAGTCATTCCGCTGTGCTCTCCTCTCACATCGCGACGGGGCTAAAGTGCATCT
CGGCACCCGCCCAACAGAGAAACAGTACGAAACCCTGGAAAATCAGCTCGCGTTCCTGTG
TCAGCAAGGCTTCTCCCTGGAGAACGCACTGTACGCTCTGTCCGCCGTGGGCCACTTTAC
ACTGGGCTGCGTATTGGAGGATCAGGAGCATCAAGTAGCAAAAGAGGAAAGAGAGACACC
TACCACCGATTCTATGCCCCCACTTCTGAGACAAGCAATTGAGCTGTTCGACCATCAGGGA
GCCGAACCTGCCTTCCTTTTCGGCCTGGAACTAATCATATGTGGCCTGGAGAAACAGCTAA
AGTGCGAAAGCGGCGGGCCGGCCGACGCCCTTGACGATTTTGACTTAGACATGCTCCCAG
CCGATGCCCTTGACGACTTTGACCTTGATATGCTGCCTGCTGACGCTCTTGACGATTTTGA
CCTTGACATGCTCCCCGGGTAA (SEQ ID NO: 61)

GTPase HRas [Homo sapiens]:
MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYS
AMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDLAARTVESR
QAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREIRQHKLRKLNPPDESGPGCMSCKCVLS
(SEQ ID NO: 62)

Substance P is position 58-68 of Protachykinin-1 [Homo sapiens]:
RPKPQQFFGLM (SEQ ID NO: 63)

Oxytocin is position 20-28 of Oxytocin-neurophysin 1 [Homo sapiens]:
CYIQNCPLG (SEQ ID NO: 64)

GCaMP6m:
ATGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCTTTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC

FIG. 14 cont'd

GGTCTCTGGGGCAGAACCCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCCTGACAATGATGGCAAGAAAAGGGA
GCTACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 65)

GCaMP6s:
ATGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCCACATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCTTTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCCTGACAATGATGGCAAGAAAAATGA
AATACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 66)

GCaMP6f:
ATGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG

FIG. 14 cont'd

AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGAATTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCCTGACAATGATGGCAAGAAAAATGA
AATACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 67)

CN1917 (length between ITRs: 2495 bp):
GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTACTATTC
CAGCCACGAGGTATAAACACTGGGAAGGAAAGTCCTGGCTCTGTATTGTCCACAAAGACC
CGAAGCTGCAGCAAAGTTGGCAAAAGAGAAACAAAAAGAGCAGAGAAGGCTCAGCTTTCA
ACAGCTAGGCTCACCCAAACATCAAGAGGTGGACAAATATTTACAGTGTGAACCTTAACCC
AAAGAACGGCAGTGCGGTCATGCTGCACAGAGTCAACTTCAGAACCAAGACTGTGACCAG
GGCTGGAGCAAGAGACACTTCACGCTAATAAATAGGCCACTTAATCAAGAAGCTGTCACAG
TCCTAAATATGTATGCACCGAGCATTAGCACTTCCAAGTAGAGTGGAACAGCTAAAGATAG
AGGCAAGAAGCAAGCAGACAAATCTTCGGTGATTGTTGGAAATCACAGCATTTCTCTCAGC
AATTGTCAGGACACAGAAAATCAGCGAGAAGACAGAAGAGTCTCACAATAGTCCCCATCAA
CTTGACCTAATTGACATTTATGGAGCTTGGCATCCAACAGCCGTGGAGCGCATGCGCTCTT
TAAGGCAGAATACAGACCATCAAACAAAACCAGGGAGACCAAAGTCACAGAAAATATGCTC
TGTGAACATGACATAATAAAGTGGAAATCGATAACAGAGAGATCGCTGCAAAATCCCCCAA
GTGATTGGTAATTAAATGCTCTACTCCTGAATGAATGATGGGCGAGAAAGGAAAGCCACGG
GGGAAAGCAGATTTCTGCGTTGAAAGAGCATGGAGACAGACTTCGTCAAGATGAGAGAGC
ACGTGGGGCTGGAGGGATGGCTCAGCACTTCAGAGGCACTCACGCTCTTCCATAGGACCT
AGGTTCACTTCTCAGCACACACATGGCAACTCACACCTGTGATGCAGAGAAGAGCTCGATT
CAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTGGAGC
CACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCCCCGG
ATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGC
TGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAG
TTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCT
GATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCT
ACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGT
CCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACT
ACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTAC
AACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTC
AAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAA
CACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGT
CCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA
CCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCG
CGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGG
TATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCA
TGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGC
CACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGG
GCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTG

FIG. 14 cont'd

TTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTC
CTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTC
ACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 68)

CN2047 (length between ITRs: 1986 bp):
GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTTCCTTCT
CAGAACCTAGGTAAGGTAAGTTCCTTTCAAGGCTGGCTATATAAATAATCATCTCAGTTAGG
ATGCTTGGTGGGACCAAAGAACCAAAACTGCCGAGCAGGCATGATCTGACTTGGAGTGGT
TCCAGGACCTTCCTGTGAATGCTGGAGTCATTCAGTGTAGAGCTCTCCTCTGTGACTGGGT
CAAGGTTGCCCCACTGTAAACCCAGGGAAGCTAGCCCAGCCTTCCTCTCAGGGAATGTGT
ATGCTTCCCTTACACCTGACCCTGGCACAGACCTGGTGGTTGTTTTTCAGAAGCATCAGTG
TCTTTGCCTTAGGCATTTGTCCTCAAAGGGCAGCGACACTGTCTACTGACTGCTTTGTACA
GGGTAACTGCTTAACTAATTCTTAAGGAGCTCGATTCAGCCGGGAGCTTAGGGAGGGGAG
GTCACTTCATAAGGGCTTGGGGGGGGAGTTGGAGCCACGAGTCGTCCAGCCGGAGCCCC
GTGTGGCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCCAGATCTTTCGAAGCTAGCGCT
ACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCC
TGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCC
CGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCT
ACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCC
AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGT
TCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
GGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACC
GCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGAC
GGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGT
GCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCAT
GGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCA
ACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTAC
GCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA
TTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTG
CCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAG
AGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTT
CCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATC
GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 69)

CN2048 (length between ITRs: 1927 bp):
GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTAAATGGA
GACTGCCAAGGGCTGAAACGGGGTGCGGGAACCAGGGACCGAGCCCCCCCCCTCCCCA
CATGAGAATCTGTCACATTGCTGCTCCAGTGGCCTGAAAGACCAGCACAGCCCCAGCTGG
AGCCTCTCCCCTCTGGATCTTGTCAATGTGGCTTTGCTTTGCTGCTTGGGCAGCCGGGAGT

FIG. 14 cont'd

GGTGACAAGCAGGGAGAGAGCGCCCAAGGCATCTGGCTGTGCCACTCCAGCCTGACTGC
CAGCTCACCCATCAGTGCCCATCTCATCATCGAGAGGGACCCAGATGAGACCGGGGATCA
GCACTGTCCTTACCTTGAAGGGACGTGTCAGGAAGAGCTCGATTCAGCCGGGAGCTTAGG
GAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTGGAGCCACGAGTCGTCCAGCC
GGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCCAGATCTTTCGAAG
CTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT
GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCG
AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGC
AAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTT
CGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGG
CTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTA
TATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACAT
CGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACG
GCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACT
CTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATA
TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGC
TCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTA
TGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATC
GCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGT
GGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCC
CGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAA
ATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGAC
AGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGC
GGCCGC (SEQ ID NO: 70)

CN2049 (length between ITRs: 1996 bp):
GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTTCTGCTT
TCTCTTCCCTTGGCCTCCTGCTGGGATAGAAGGGGTTGGTGTGAGTGTGTATGGTGGGGT
GCTGTGAGATTAATTAGCAGCCGTGCCAGGCAGCAGGCGGTGGGGTGCAGAGTAGGCTG
GCTTTCCCTGCTATAGATCCATGCTCTCTGGGAGAGGCACTAGCCGGCTGCTTTGGGCTC
TGGCTCAGCTATTTTAGGAATATTCTTAACCCTTCCAGAACCGCTGCCATTGCCAGATCTCT
CTCCCAGAACACAGGCCAGCTCCAGATTGCCCCTCCTTTCTGCCCCCGCCCTGCACCCCA
CCTAGCCTCTGCTCTTCCTCCCTACAAGTTGAGAAGGTCAAGGTTTGACTTTTACCAAAGAA
AACTCCTGGCTCCTGATCCCACTCTCTGTGCTTTACCTGAGCTCGATTCAGCCGGGAGCTT
AGGGAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTGGAGCCACGAGTCGTCCA
GCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCCAGATCTTTC
GAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGG
TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCC
GGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCAC
CGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGT
GCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG
AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC

FIG. 14 cont'd

TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAAC
GTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCAC
AACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGG
CGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
TCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATT
CGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT
GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTC
CCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAAC
TCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATT
CCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCT
CCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA
GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCA
GGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACC
GAGCGGCCGC (SEQ ID NO: 71)

CN2427 (length between ITRs: 2071 bp):
GCGGCCGCACGCGTGGGATGCGATGCAGTGGCATGGTAAAAATGCCGTACCTAATGATGC
AGTTTCAGCCATCATGACGTATGGATGGAATGAAAAGATGACAAAGCTACTAGCGGATCTG
TTCACAAAGGAAAAGAGGCTCAGTTTCCAACAGTCTCTGGCATTTCATCTTTGGAAAACTTG
TCAAAGCTAAGATGATTTGTGAGAGTCCTAACTCATTTTCCTAAATATGAAGTCGCTATTAT
GGATAAAGAAAGGTTACAATGGAGCATCCATTTCCACTCTAGTCATTCTGCTTGATTGCACC
AATTAGCCCTGCATTCACTGTGCATTTTAACTTCATAATGGTCTATTATTTGGACCACTGAC
CATAATGATGCAATTCTTTCCTTGACGAAAATAAATGCTTAGTGATAAATAAGTAATCATAAT
TAAAGCTTTCATAGTACTTATAGTACCACAATTGGTATTTAGCCCATTGATTATAATTTAAAC
ACATTTTAATTAATATGTAACTATATTACTTCCCTATTTTTCTTTTCTTTCCTCCAAAATGCCT
CATGTCTCCTTCTATCAACACCCCCCATACCCCCCAGACACTCCTTGTCAAATTGATGGCC
TTTTAAAATTATTATTATTGTAACATGAAGAAATAGTTGAACAAATATATAAATACAGTACAAT
GAGTCTGTTTATGGTGGCTCCTATGGGTATGATTTGGGGGCCTGAGCTCGGGCTGGGCAT
AAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAG
CTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT
GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCG
AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGC
AAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTT
CGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGG
CTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTA
TATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACAT
CGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACG
GCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACT
CTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATA
TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGC
TCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTA
TGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATC
GCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGT
GGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCC

FIG. 14 cont'd

CGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAA
ATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGAC
AGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGC
GGCCGC (SEQ ID NO: 72)

CN2320 (length between ITRs: 1985 bp):
GCGGCCGCACGCGTTATCTTAGAGTGGGAAGATTTGAGAAGTGCCATGGTTAATATGACT
GACTTTTTATTCTTATTTCTTTTAATTTCATGGTTCTAAATCCGAATTTAATCATAGTACCCAG
AAAAGCAGAGGTGTAGAGGTTCACAGTGGGAGTTGTAATCTAGCCCTATTCATTTTGACCT
CAAAACCCAAATTATTTATAACAAATTATTTCCTATTCTTTCCTTCACTATTCAGGAACATCTG
TCCACCACTTACATGATCACTTATCTTGCTATTGTGTCATTTTGATGAAAAAGAATTTTTTCT
AAATATCTAAATACAAGGCCCCATATTAACAGTGCTTTTTAAATCCCCACAGATGTGGGAGA
TGACCCCTTTCCATCCCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTTGTTCCAATA
AAGAGATACAATTTTATTCATTAATTTGTGTATTCATTTAGCAAATCACTTTAGAGTCTTATTA
TATCAGGATTTTGGGGTCTATTTTAGTATATCTTTTTGTATTTCTTGGAACCTCTCCAATTATT
CTAGACTCTTTCAAAGGTTGGTGATCAATATTAGACATTATTATGAAAAGAATCTTACTTGCT
AAAAGGGTTAGATGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT
ACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAG
CAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACG
TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT
GACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGA
ACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGC
ATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
ACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAG
TCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTG
TGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTT
TAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAAT
CCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAG
GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCT
AGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCC
ACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCA
TTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA
GCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 73)

CN2321 (length between ITRs: 1844 bp):
GCGGCCGCACGCGTTATGATGTGCCAGGCTTGGGAGAAACACCACAAGCAAAGCCAAAAT
AGGTGGCCTAGAACTTCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCA
GCTGCCGGGTGCCGCATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGAT
GACAGCTGGCTCTTCACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATA
ATTCATATTTAATTACTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAAT
TTGCTATTGACAAACTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATG
AAAAATGAAGCTTATTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAA

FIG. 14 cont'd

AAGCCAAACTTGAATGCAAAAAAATATTGAATATGAAAAGAAAAACATTTGCAGTAAAGCTT
GTTCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTT
CTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGA
GGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCC
ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG
AAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCT
GGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTT
CAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCC
AACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTA
CCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTT
CGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCG
CGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATT
GACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAG
TTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAG
CCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTG
TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTG
GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
AGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 74)

CN2719 (length between ITRs: 2346 bp):
GCGGCCGCACGCGTAAAGCTACATCTCTGGGCTCCATTATTAAAGCTGCTTTCCTTCCTTT
CTCTCTCTCTCGCTCTCTCTCTCTCTCTCTTTCTTTCTTTCTTTTTCCTGAGATGGGTCTCAT
GTAGCCCAGGCTGGCCTCAAGCTTGCCACATAGCCAAGGAGCCAAGGATGGCATTGAACT
CCTAGATCCTCCAGCCTCTGCCTCCTGTTAGGATTAGAGGTGAGCTACAATCCCAAGGGC
CTTTAGTGTAACAGTCAAAAGCTACTGGGAGTCAGGCAATTGGCTCAGTATAACCCTGATT
CTCCTTTTGTGCCCAGGCACGTTGGTCAGGAGTCTGTCTTCAGCCTGTCATGGCAGCAGC
TCAGCTTCAGTGACCAATCTATACTCACTCACAGGAGACTCTGAAATCCCAGATTCTGTGCT
ATAAAGTCCCCGCTCGAGTGAGTCGTGACTGCTCCAAACAGCCTGGGCAGCTGCGAACCC
TCATGGCATCTAGGTGACCCTGTTCATCCTACAGCTGTTCTCACTGAGGGGAGGGGAGCT
TTTGAGTGAGCCAGTCAAAACTCTGTGCTCGGTGATCCTGTGAGGCTCGGAACGGTGGCA
CCCGAAGCCATGGGTGCACACACAAACAGGGCTCTAATCGGTGGGATCACAATCCATGAA
CAAGCATGAGACCTCCCTTCTTCTCACACACACACACACACACACACACACACACACACTC
ACTCACATACATGAGCTGGTTTCCACAACTGTGGGGTTAGCCTGGAAGGTGTCTGTCCTAT
ATAGTACTCCAGTACCAGTGTTGCAGACTCTAGGCCCAGGAAAAGGTTTTTGATGTTTGCT
GGTGTTTTTGATGACCTTCATCGTGGGTAGAGCAGGCTGCGCTGGTTCATAAAGAGAAGAC
AAGACCTAGAGTGCTGTGACCCTTTAAGGCATCATGGTGTGATGACCCCCAACCATAAAGT
TATTTTCGCTGCTGCTTTGTAACTGTAAGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAG
AGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGT
CGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCC
CTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCG

FIG. 14 cont'd

ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGAC
AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCT
GGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG
TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTC
CTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCC
CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTC
GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC
CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
TGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
75)

CN2707 (length between ITRs: 1786 bp):
GCGGCCGCACGCGTTGAATCTTCATGGAGAAAACATCATGAATAGAGAATGAAGAAGTAAA
GATGAACAAGTGAGAATTCAGATCATCAGAGGGTTTTCTAGGATTTCTGTAAATTTCTCTGT
GTTTTGGAATGCAATAGGAATGCCAGCCCAAAGCCATATAGGTCACAGCTGCCCAAGAACA
GTTACCAATACAGTATAAATAACGTCCTAAACTTAGAATATTGTGAAATTCTTTTTATAACTC
AGCCATTTAATTCTTGAGTGTAAATTTTGTAGATGGAGTCATGTTGTAATGTTAGACACATTT
GCTAGTGATGTGACAACATAATATTCCCATGAACTGATGTCAAATGTTGTATTGTACTTTGA
CCAGGTATATAAGGTTTATTATCTTCTTGACCTTGGAGTAATTTCAGTCCCAACTTAAATCCC
TAGTGGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTG
CTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGG
CGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC
CTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCAC
CCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGA
CGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA
TCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG
TACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAG
GCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTA
CCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGA
GCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGG
AGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACG
GCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAG
ATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGC
CTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGT
TAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTC
GGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGC
CAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCC
ACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTAT

FIG. 14 cont'd

TCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG
CATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 76)

CN2713 (length between ITRs: 1710 bp):
GCGGCCGCACGCGTTATGTAGATACAGGTCATAGAACTTGCCCTGGGGAATGGCTCCATT
TGGTACCAACAGGCTGACCCCTAGGGAGGAAGGAAGGCTATCAGCAAGAGGAGGAGGAG
GTAGCAGAGATGAGAAAGATGGGGTAGACTCTGGCTCCAACCTAGGGAAGGGAAAGACTC
TAGACTCGGGGGTATGGGGGTGGATAGATACAGGGAGCACACAGGCTACTTGGCCTGGT
CTGCCCATGAATACAGGGGGCCTCTAACATTGCTGGGGTAGGAGGGTCAGAATGCTCCAG
TGCTAGCCCTCATGCTGGCTCAGGACAGGACTCTGAAAAGCCACCAGCTGCCACTTTCAC
AAGCTGAGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTT
GCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGG
GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC
CCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCA
CCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGAC
TTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAA
GGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG
AGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG
GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGAC
GGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAA
GATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATG
CCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGG
TTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCT
CGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCC
CACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTA
TTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG
GCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 77)

CN2717 (length between ITRs: 1740 bp):
GCGGCCGCACGCGTAGTGACTTGGTGCTATGAGCCATATTTTGCTGTTGCTGTTGTTACTG
GTAGTTTTTGTAATTCTGGGGCTAAAACTTGGGGTCTGGTATGCTGTCATTTACCAGTGAG
CTATACCCTGGATATTATGATTTAGATGAATGTGAAATATCACCCCAGACATACATATACTAA
ACACTTGGCCCTTGGCCCATGATGCTAAATGGAGGAGATAGAAGCTTTTGGGGCACAGCC
TAGTGGAAGGAAATGAGGTCAAATGACATGTACTCTGAAAGGAATATGGGTATTCTGGGCT
TGCGTTATTCTCTCTCTCCCTCTCTCTCCCTCTCTCTCCCTCTCCCTCTCTCTTTCTCCTTTT
CTCTTTCTCTCCTCGCCTTGTTTTCCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGC
CATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGC
CACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA

FIG. 14 cont'd

CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGC
AGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGA
TTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGG
ATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTC
CTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGC
TGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGAC
TGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTG
GAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
GAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 78)

AiP1104 (length between ITRs: 3455 bp):
GCGGCCGCACGCGTTCCAGCTACCACCAGCCTGTCCAAAAGGGGACACCAAAGGAGAGG
AGGAAGTCTGAGAGACACCTCTCTGCACCATGGCCATCTTAGTAGTCAGACCCAGAACAAA
ACTCTCTGATGAGTGTGCGGAGCGTCGCTTCTGGTCTTTGCTCAAACTATTCACTGAAGAT
TTAAAGCAATCCCGTGAGTGATACATTTGGGTGAATTTGTTCTCTAGAAGGTATCACAGAAA
TCTGGTCACTGGGCCACCCGAGACATCCTGATAGGCCCTCTGGTAACCCATCACATGCTG
CAGACTGACTCTGGGGGCCTAGAACCCAGATCAGAAGCAACCTTGACCCCGGCCCACCC
GCCACGGAAGCACCATCATCTCTCTGATTAAAAACCTCGATCACGGACCCGGGGGCGTGC
CCGGAAGAGCTAAGATAATCAGCGTCAGCACTTTGCCTTCGCCGTCCAAGACTGCAGACG
GCCTTCATTTGACCTGATTCGTGGTGTTAATGACAGCAGAGCAATTTTGAGAGGCAGCTTG
CTCTCGGCATCTATAAGGAGAGGAAAAGCACTGAGGGCTGGGGACCAAGCTCCTTGCAGA
GGCGGCAGCTGCAGTCACCCTCCCCCTCCACCCCTGCCCCTCCCCTCCCCTCCAGAGGC
ACTTTGAGTAAGTGCTGCCCTCCGATCTGCCCTGATACGATGGGAGAAAGCTGATGTGAG
GGCTGGAGCCAGAGTGTGCAAGGGGACAGTGTGTGCATGTGCGTGTGTCGGGGAGAGGT
ACCCGTGCTATACCTGAGAACATTGCTGGGTGAACACAGCCTTGGACCTGGAAGAGCGCA
TAGCTTACTTAGAGGCATGGGCTGCACATGAGCTGCCCATTTACCTGCTCATTTAGAAGCT
ACTATGAAGGCTGGTGAGATGGCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCC
ATCTATTGCTTACATTTGCTTCTGGCGTGGCCACCATGGCTCCTAAGAAGAAGAGGAAGGT
GATGAGCCAGTTCGACATCCTGTGCAAGACCCCCCCCAAGGTGCTGGTGCGGCAGTTCGT
GGAGAGATTCGAGAGGCCCAGCGGCGAGAAGATCGCCAGCTGTGCCGCCGAGCTGACCT
ACCTGTGCTGGATGATCACCCACAACGGCACCGCCATCAAGAGGGCCACCTTCATGAGCT
ACAACACCATCATCAGCAACAGCCTGAGCTTCGACATCGTGAACAAGAGCCTGCAGTTCAA
GTACAAGACCCAGAAGGCCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCCT
GGGAGTTCACCATCATCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACCGACATCG
TGTCCAGCCTGCAGCTGCAGTTCGAGAGCAGCGAGGAGGCCGACAAGGGCAACAGCCAC
AGCAAGAAGATGCTGAAGGCCCTGCTGTCCGAGGGCGAGAGCATCTGGGAGATCACCGA
GAAGATCCTGAACAGCTTCGAGTACACCAGCAGGTTCACCAAGACCAAGACCCTGTACCA
GTTCCTGTTCCTGGCCACATTCATCAACTGCGGCAGGTTCAGCGACATCAAGAACGTGGA
CCCCAAGAGCTTCAAGCTGGTGCAGAACAAGTACCTGGGCGTGATCATTCAGTGCCTGGT
GACCGAGACCAAGACAAGCGTGTCCAGGCACATCTACTTTTTCAGCGCCAGAGGCAGGAT
CGACCCCCTGGTGTACCTGGACGAGTTCCTGAGGAACAGCGAGCCCGTGCTGAAGAGAG

FIG. 14 cont'd

TGAACAGGACCGGCAACAGCAGCAGCAACAAGCAGGAGTACCAGCTGCTGAAGGACAAC
CTGGTGCGCAGCTACAACAAGGCCCTGAAGAAGAACGCCCCCTACCCCATCTTCGCTATC
AAGAACGGCCCTAAGAGCCACATCGGCAGGCACCTGATGACCAGCTTTCTGAGCATGAAG
GGCCTGACCGAGCTGACAAACGTGGTGGGCAACTGGAGCGACAAGAGGGCCTCCGCCGT
GGCCAGGACCACCTACACCCACCAGATCACCGCCATCCCCGACCACTACTTCGCCCTGGT
GTCCAGGTACTACGCCTACGACCCCATCAGCAAGGAGATGATCGCCCTGAAGGACGAGAC
CAACCCCATCGAGGAGTGGCAGCACATCGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCA
TCAGATACCCCGCCTGGAACGGCATCATCAGCCAGGAGGTGCTGGACTACCTGAGCAGCT
ACATCAACAGGCGGATCTGAGAATTCGATATCAAGCTTATCGATAATCAACCTCTGGATTAC
AAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATAC
GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTG
TATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCG
TGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTC
AGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCG
CCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGT
TGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTATGTTGCCACCTGGATTCTGCG
CGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGG
CCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGAT
CTCCCTTTGGGCCGCCTCCCCGCATCGATACCGAGCGCTGCTCGAGAGATCTACGGGTG
GCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCC
CACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAA
TATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTA
GGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTG
CAATCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATT
CCAGGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACC
ATATTGGCCAGGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCA
AATTGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTTGTAGGT
AACCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 79)

AiP1089 (length between ITRs: 3450 bp):
GCGGCCGCACGCGTACTATTCCAGCCACGAGGTATAAACACTGGGAAGGAAAGTCCTGGC
TCTGTATTGTCCACAAAGACCCGAAGCTGCAGCAAAGTTGGCAAAAGAGAAACAAAAAGAG
CAGAGAAGGCTCAGCTTTCAACAGCTAGGCTCACCCAAACATCAAGAGGTGGACAAATATT
TACAGTGTGAACCTTAACCCAAAGAACGGCAGTGCGGTCATGCTGCACAGAGTCAACTTCA
GAACCAAGACTGTGACCAGGGCTGGAGCAAGAGACACTTCACGCTAATAAATAGGCCACT
TAATCAAGAAGCTGTCACAGTCCTAAATATGTATGCACCGAGCATTAGCACTTCCAAGTAGA
GTGGAACAGCTAAAGATAGAGGCAAGAAGCAAGCAGACAAATCTTCGGTGATTGTTGGAAA
TCACAGCATTTCTCTCAGCAATTGTCAGGACACAGAAAATCAGCGAGAAGACAGAAGAGTC
TCACAATAGTCCCCATCAACTTGACCTAATTGACATTTATGGAGCTTGGCATCCAACAGCC
GTGGAGCGCATGCGCTCTTTAAGGCAGAATACAGACCATCAAACAAAACCAGGGAGACCA
AAGTCACAGAAAATATGCTCTGTGAACATGACATAATAAAGTGGAAATCGATAACAGAGAG
ATCGCTGCAAAATCCCCCAAGTGATTGGTAATTAAATGCTCTACTCCTGAATGAATGATGG
GCGAGAAAGGAAAGCCACGGGGGAAAGCAGATTTCTGCGTTGAAAGAGCATGGAGACAG
ACTTCGTCAAGATGAGAGAGCACGTGGGGCTGGAGGGATGGCTCAGCACTTCAGAGGCA
CTCACGCTCTTCCATAGGACCTAGGTTCACTTCTCAGCACACACATGGCAACTCACACCTG
TGATGCAGAGAAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTAC
ATTTGCTTCTGGCGTGGCCACCATGGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTT
CGACATCCTGTGCAAGACCCCCCCCAAGGTGCTGGTGCGGCAGTTCGTGGAGAGATTCGA

FIG. 14 cont'd

GAGGCCCAGCGGCGAGAAGATCGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGA
TGATCACCCACAACGGCACCGCCATCAAGAGGGCCACCTTCATGAGCTACAACACCATCA
TCAGCAACAGCCTGAGCTTCGACATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGACCC
AGAAGGCCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCCTGGGAGTTCACC
ATCATCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACCGACATCGTGTCCAGCCTG
CAGCTGCAGTTCGAGAGCAGCGAGGAGGCCGACAAGGGCAACAGCCACAGCAAGAAGAT
GCTGAAGGCCCTGCTGTCCGAGGGCGAGAGCATCTGGGAGATCACCGAGAAGATCCTGA
ACAGCTTCGAGTACACCAGCAGGTTCACCAAGACCAAGACCCTGTACCAGTTCCTGTTCCT
GGCCACATTCATCAACTGCGGCAGGTTCAGCGACATCAAGAACGTGGACCCCAAGAGCTT
CAAGCTGGTGCAGAACAAGTACCTGGGCGTGATCATTCAGTGCCTGGTGACCGAGACCAA
GACAAGCGTGTCCAGGCACATCTACTTTTTCAGCGCCAGAGGCAGGATCGACCCCCTGGT
GTACCTGGACGAGTTCCTGAGGAACAGCGAGCCCGTGCTGAAGAGAGTGAACAGGACCG
GCAACAGCAGCAGCAACAAGCAGGAGTACCAGCTGCTGAAGGACAACCTGGTGCGCAGC
TACAACAAGGCCCTGAAGAAGAACGCCCCCTACCCCATCTTCGCTATCAAGAACGGCCCT
AAGAGCCACATCGGCAGGCACCTGATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAG
CTGACAAACGTGGTGGGCAACTGGAGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCAC
CTACACCCACCAGATCACCGCCATCCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTA
CGCCTACGACCCCATCAGCAAGGAGATGATCGCCCTGAAGGACGAGACCAACCCCATCGA
GGAGTGGCAGCACATCGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCATCAGATACCCCG
CCTGGAACGGCATCATCAGCCAGGAGGTGCTGGACTACCTGAGCAGCTACATCAACAGGC
GGATCTGAGAATTCGATATCAAGCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAA
AGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAAT
GCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTG
GTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCAC
TGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCC
GGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCC
CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAA
ATCATCGTCCTTTCCTTGGCTGCTCGCCTATGTTGCCACCTGGATTCTGCGCGGGACGTCC
TTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCG
GCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGG
GCCGCCTCCCCGCATCGATACCGAGCGCTGCTCGAGAGATCTACGGGTGGCATCCCTGT
GACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCCCACCAGCCTT
GTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAATATTATGGGG
TGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGGCCTGCGG
GGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGC
CTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCATG
CATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCCA
GGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGGG
ATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTTGTAGGTAACCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 80)

AiP1105 (length between ITRs: 3024 bp):
GCGGCCGCACGCGTACCAGAAGTTCAGTGAGCAGAAGATGGGCTAAAATGAAAAGGGTAC
TGTCTTGAACTGAAGATGGAATCCTGCAGCTTCATTCTGGCCAAAAGAAGATCTATTCCCA
GGAGGAGGGTAAAGGCTTTGTTCTTAAGAGATGCTGAGGCTGGCCCTGTGAATCTGATGT
CAAGATGTCCCTTGTCACTCTGCAGAAGCGTATGTCTCTTGCATTTCCTTCTTATTTCCTTG
GGTGAAATTGCTGTGGCATTGTGTCACTCATCCTAATGGGTCATGTCTAACATCTGCGTGC
TTACAAATCAGGCATGCTCATTTCTGGGCTTATGGAGCTTGTATAACACCAGGACAGGCAA

FIG. 14 cont'd

GACATGTTGCCCACTCAGGAAGAATAGAAGCTGGGCACAGCTGGAGTGCAAAGTAGGTCA
GTTCAGAGAGCAAAGGGAGTTGATGGAGCAATGAGTTGTTAGTGGGAAAGTTCTAACCAAC
TGTCCCAGAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTT
GCTTCTGGCGTGGCCACCATGGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTTCGAC
ATCCTGTGCAAGACCCCCCCCAAGGTGCTGGTGCGGCAGTTCGTGGAGAGATTCGAGAG
GCCCAGCGGCGAGAAGATCGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGATGA
TCACCCACAACGGCACCGCCATCAAGAGGGCCACCTTCATGAGCTACAACACCATCATCA
GCAACAGCCTGAGCTTCGACATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGACCCAGA
AGGCCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCCTGGGAGTTCACCATCA
TCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACCGACATCGTGTCCAGCCTGCAGC
TGCAGTTCGAGAGCAGCGAGGAGGCCGACAAGGGCAACAGCCACAGCAAGAAGATGCTG
AAGGCCCTGCTGTCCGAGGGCGAGAGCATCTGGGAGATCACCGAGAAGATCCTGAACAG
CTTCGAGTACACCAGCAGGTTCACCAAGACCAAGACCCTGTACCAGTTCCTGTTCCTGGCC
ACATTCATCAACTGCGGCAGGTTCAGCGACATCAAGAACGTGGACCCCAAGAGCTTCAAG
CTGGTGCAGAACAAGTACCTGGGCGTGATCATTCAGTGCCTGGTGACCGAGACCAAGACA
AGCGTGTCCAGGCACATCTACTTTTTCAGCGCCAGAGGCAGGATCGACCCCCTGGTGTAC
CTGGACGAGTTCCTGAGGAACAGCGAGCCCGTGCTGAAGAGAGTGAACAGGACCGGCAA
CAGCAGCAGCAACAAGCAGGAGTACCAGCTGCTGAAGGACAACCTGGTGCGCAGCTACAA
CAAGGCCCTGAAGAAGAACGCCCCCTACCCCATCTTCGCTATCAAGAACGGCCCTAAGAG
CCACATCGGCAGGCACCTGATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAGCTGAC
AAACGTGGTGGGCAACTGGAGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCACCTACA
CCCACCAGATCACCGCCATCCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTACGCCT
ACGACCCCATCAGCAAGGAGATGATCGCCCTGAAGGACGAGACCAACCCCATCGAGGAGT
GGCAGCACATCGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCATCAGATACCCCGCCTGG
AACGGCATCATCAGCCAGGAGGTGCTGGACTACCTGAGCAGCTACATCAACAGGCGGATC
TGAGAATTCGATATCAAGCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATT
GACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGC
TGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGT
TTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGA
CTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCT
GCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCAT
CGTCCTTTCCTTGGCTGCTCGCCTATGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTG
CTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCT
GCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGC
CTCCCCGCATCGATACCGAGCGCTGCTCGAGAGATCTACGGGTGGCATCCCTGTGACCCC
TCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCCCACCAGCCTTGTCCTA
ATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAATATTATGGGGTGGAG
GGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGGCCTGCGGGGTCT
ATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGCCTCCT
GGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCATGCATGA
CCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCCAGGCT
GGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGGGATTA
CAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTTGTAGGTAACCACGTGCGGAC
CGAGCGGCCGC (SEQ ID NO: 81)

FIG. 14 cont'd

AiP1090 (length between ITRs: 3533 bp):

GCGGCCGCACGCGTTGTCTTAGACTGAGTTGCTGTAACAAAAATCTGGGATGAGGTCATTT
CTAAAGCACAGCAATTTATTTCCCACAGTCTGAAGGCTGCAGATTCCAAGATCACTGGCAA
GATCAGTTGTAAAGGCTTCGTCCTCCAGAGGCGGGGATGCTGCATACTCCCTGGGCAGAG
AGACAGGAAAACTCCGTAACTGCATGTGTCCTTCCTGATGCCTCTTCTATATAGGCCTGGA
TCCCAATCACCCTGTGAACCTCTCCCTGCTTCGTGGCCCCCACCTCTTAACACTACCACAT
TGGCAACTCCTGAAATTTGAAGGGGACACACTGAACCATGGCACAACAGCTTTCTGACTGA
TGCAGTAACCCAATGGCAGTGCAGAAGGGGCCAGCTAAAAGCCCAAATGGTTAGCTCAAA
ATTCGCTGTCTCTTCCGAGTGTCTGAACCCTTAGTCCTGGTATGTAAAGACATCAGAACATT
TCCCCTTGTGTCCATCAGATTTCTGTCTAGTGAAACGATGACACTGTAACCTCCAAGATCTC
ACACGAAATGATCTTTTCTCCTTTGTGGAAGGAAACCAGCATTTAGCTCATCTCTCCTTCGT
AGCAGCTCAGAATGTCCACAGTGACCCAGTTACCATAGCTAAAGGCTTCCTTTTCAAAACA
CAGAGCAGAGGCAGCCAATTCAGTATGTGCTGCTGCCATCCTCTGATTCTTTCCTGCTTCC
ATAGACACCAACTCTATTGTAACTAAGCCTTATACATTGTGTCTTCCTCCTTTACATTAGCTT
GTGCTGGGGTGGTTCATGAGGCCCGCTGAGTAGTTTCAGTGACAGCCTATCCCTCTGCCA
GTGCTGCTTTGAGCCATCTTATTGGTGAGGCTGTAAGAGAAGCCTGAAGTCACAGGGTAAA
GCTATGTTGAAGGCAGCCCCAGAACCAAGTTTCCCTATTTCTATCTCCTTACGCTGTTTGAG
CCTCAGGGGTAGATCAGGTGCCTGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGC
CATCTATTGCTTACATTTGCTTCTGGCGTGGCCACCATGGCTCCTAAGAAGAAGAGGAAGG
TGATGAGCCAGTTCGACATCCTGTGCAAGACCCCCCCCAAGGTGCTGGTGCGGCAGTTCG
TGGAGAGATTCGAGAGGCCCAGCGGCGAGAAGATCGCCAGCTGTGCCGCCGAGCTGACC
TACCTGTGCTGGATGATCACCCACAACGGCACCGCCATCAAGAGGGCCACCTTCATGAGC
TACAACACCATCATCAGCAACAGCCTGAGCTTCGACATCGTGAACAAGAGCCTGCAGTTCA
AGTACAAGACCCAGAAGGCCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCCT
GGGAGTTCACCATCATCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACCGACATCG
TGTCCAGCCTGCAGCTGCAGTTCGAGAGCAGCGAGGAGGCCGACAAGGGCAACAGCCAC
AGCAAGAAGATGCTGAAGGCCCTGCTGTCCGAGGGCGAGAGCATCTGGGAGATCACCGA
GAAGATCCTGAACAGCTTCGAGTACACCAGCAGGTTCACCAAGACCAAGACCCTGTACCA
GTTCCTGTTCCTGGCCACATTCATCAACTGCGGCAGGTTCAGCGACATCAAGAACGTGGA
CCCCAAGAGCTTCAAGCTGGTGCAGAACAAGTACCTGGGCGTGATCATTCAGTGCCTGGT
GACCGAGACCAAGACAAGCGTGTCCAGGCACATCTACTTTTTCAGCGCCAGAGGCAGGAT
CGACCCCCTGGTGTACCTGGACGAGTTCCTGAGGAACAGCGAGCCCGTGCTGAAGAGAG
TGAACAGGACCGGCAACAGCAGCAGCAACAAGCAGGAGTACCAGCTGCTGAAGGACAAC
CTGGTGCGCAGCTACAACAAGGCCCTGAAGAAGAACGCCCCCTACCCCATCTTCGCTATC
AAGAACGGCCCTAAGAGCCACATCGGCAGGCACCTGATGACCAGCTTTCTGAGCATGAAG
GGCCTGACCGAGCTGACAAACGTGGTGGGCAACTGGAGCGACAAGAGGGCCTCCGCCGT
GGCCAGGACCACCTACACCCACCAGATCACCGCCATCCCCGACCACTACTTCGCCCTGGT
GTCCAGGTACTACGCCTACGACCCCATCAGCAAGGAGATGATCGCCCTGAAGGACGAGAC
CAACCCCATCGAGGAGTGGCAGCACATCGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCA
TCAGATACCCCGCCTGGAACGGCATCATCAGCCAGGAGGTGCTGGACTACCTGAGCAGCT
ACATCAACAGGCGGATCTGAGAATTCGATATCAAGCTTATCGATAATCAACCTCTGGATTAC
AAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATAC
GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTG
TATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCG
TGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTC
AGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCG
CCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGT
TGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTATGTTGCCACCTGGATTCTGCG

FIG. 14 cont'd

CGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGG
CCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGAT
CTCCCTTTGGGCCGCCTCCCCGCATCGATACCGAGCGCTGCTCGAGAGATCTACGGGTG
GCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCC
CACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAA
TATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTA
GGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTG
CAATCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATT
CCAGGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACC
ATATTGGCCAGGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCA
AATTGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTTGTAGGT
AACCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 82)

AiP1106 (length between ITRs: 3409 bp):
GCGGCCGCACGCGTAGATGAGTCTGCAGCTGGGTACAGTGACCTTTCAGTCCATGTTTAT
TTGGAAATGACCTTTAAGCAGCAATCAGCAAGAATAAAATGCTTCAAAGGAATACATTAGAT
AATGCAGAAGTCCCCCAGAGGTAAGTTACACCCAAGGCACCCAGCTGACAATGAAAGTGG
CCCTGCCCTGGGAAGCCAAGGACTAGGCCACCCATTAGACTAACAAGTGAACACAGGGTC
CCCAGAGTTTGGTCTAATAGACAATGGGGAGTCTGAAGACAGGGTGACCTGGGCAAGACA
CAAGAGCAGTTCCAAAATTAAACCTCTGTCGTAATGAAGGATGCCTAGTTGTGCTTTTTCCA
TCCTAGGATGGGGAATCCTCAAGGGCAGGGCACAGCTGTGCCAGGGGAACTGTACGGGC
TCCATCCTGCCTCCCTCCCATGGGGTGAGCTGATAGTCTTCCTCATACTGAGCTCTTGTCT
CTGCTGTGTGCTGGGGAGTCTGAAATGCTAGAGAAACTAAGCCTTCCCACTCAAAGACAGA
GAAAGAGCTGGCCCATGGCTCCGTGCCCTCTCCTCTCTCTGTGCGTGTCTTTAACTCTGTA
TGTTCTATTTTCCCCCTCCTCGTCCCCTGCTTCGCGCTCACAGAGTCACTCCTAGTAGCAC
CAAAGAGAGATGCTTGGCAGTTCACTAACCCCTTGAGCTGAAATAGAAATAAATATCCCAA
AAGAGAAATCAGAAAAGCAGGGTGTCGCGCTGGAGAAGAGGCAGGAAGATCAGAAATACA
AGGTCATCTGTGGCTACACATCTAGTCCAAGTCCCAGCCTGGGCTATGTGAGATGGAGGG
GAATCGCTCAGAAACAAGGCTGTACACTTAAGGAGCTCGGGCTGGGCATAAAAGTCAGGG
CAGAGCCATCTATTGCTTACATTTGCTTCTGGCGTGGCCACCATGGCTCCTAAGAAGAAGA
GGAAGGTGATGAGCCAGTTCGACATCCTGTGCAAGACCCCCCCCCAAGGTGCTGGTGCGG
CAGTTCGTGGAGAGATTCGAGAGGCCCAGCGGCGAGAAGATCGCCAGCTGTGCCGCCGA
GCTGACCTACCTGTGCTGGATGATCACCCACAACGGCACCGCCATCAAGAGGGCCACCTT
CATGAGCTACAACACCATCATCAGCAACAGCCTGAGCTTCGACATCGTGAACAAGAGCCTG
CAGTTCAAGTACAAGACCCAGAAGGCCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATC
CCCGCCTGGGAGTTCACCATCATCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACC
GACATCGTGTCCAGCCTGCAGCTGCAGTTCGAGAGCAGCGAGGAGGCCGACAAGGGCAA
CAGCCACAGCAAGAAGATGCTGAAGGCCCTGCTGTCCGAGGGCGAGAGCATCTGGGAGA
TCACCGAGAAGATCCTGAACAGCTTCGAGTACACCAGCAGGTTCACCAAGACCAAGACCC
TGTACCAGTTCCTGTTCCTGGCCACATTCATCAACTGCGGCAGGTTCAGCGACATCAAGAA
CGTGGACCCCAAGAGCTTCAAGCTGGTGCAGAACAAGTACCTGGGCGTGATCATTCAGTG
CCTGGTGACCGAGACCAAGACAAGCGTGTCCAGGCACATCTACTTTTTCAGCGCCAGAGG
CAGGATCGACCCCCTGGTGTACCTGGACGAGTTCCTGAGGAACAGCGAGCCCGTGCTGA
AGAGAGTGAACAGGACCGGCAACAGCAGCAGCAACAAGCAGGAGTACCAGCTGCTGAAG
GACAACCTGGTGCGCAGCTACAACAAGGCCCTGAAGAAGAACGCCCCCTACCCCATCTTC
GCTATCAAGAACGGCCCTAAGAGCCACATCGGCAGGCACCTGATGACCAGCTTTCTGAGC
ATGAAGGGCCTGACCGAGCTGACAAACGTGGTGGGCAACTGGAGCGACAAGAGGGCCTC
CGCCGTGGCCAGGACCACCTACACCCACCAGATCACCGCCATCCCCGACCACTACTTCGC

FIG. 14 cont'd

CCTGGTGTCCAGGTACTACGCCTACGACCCCATCAGCAAGGAGATGATCGCCCTGAAGGA
CGAGACCAACCCCATCGAGGAGTGGCAGCACATCGAGCAGCTGAAGGGCAGCGCCGAGG
GCAGCATCAGATACCCCGCCTGGAACGGCATCATCAGCCAGGAGGTGCTGGACTACCTGA
GCAGCTACATCAACAGGCGGATCTGAGAATTCGATATCAAGCTTATCGATAATCAACCTCT
GGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG
TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTC
CTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAA
CGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACC
ACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCA
TCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCC
GTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTATGTTGCCACCTGGA
TTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTT
CCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACG
AGTCGGATCTCCCTTTGGGCCGCCTCCCCGCATCGATACCGAGCGCTGCTCGAGAGATCT
ACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTC
CAGTGCCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCC
TTCTATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACA
ACCTGTAGGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGG
CTCACTGCAATCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGT
TGGGATTCCAGGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGG
TTTCACCATATTGGCCAGGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGG
CCTCCCAAATTGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTT
TGTAGGTAACCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 83)

AiP1091 (length between ITRs: 3320 bp):
GCGGCCGCACGCGTTTCACCCACCTGACACTTGGGTTAGACCTGAATGTCGTTTCTTTAAC
TCACACTGCTCATCCCACTGGCCTTTGCTGTGCTTCTCTGTGCCTCCTCAGAGATACATGA
AACTGTCCCATCCCCCTAACGATGCTGGATGGATGGCTCCAACAGCTCACTGCTCTCACCT
TGACACAAAGTCCTAGCGTCTGCATCTGTGAGACAAGTTGGAATTTATATATTTCCAGTGGA
GATTAATAATTCATTAGATGCTGAAGTAGAAAAACAAAGTACCGATTAATCAAGGCTCTGCT
GAGGCCTGCTTTGCAGCCACCAGTCTGTGGGGATTGGCAGTGCTTTTACACTGGAAGTAG
GTCAGGACCACAGAAAAGCAGCTCTCATGCACTAGCATCTGTTCGCACTAATCACTGTACA
CAGCTTTGGGTCTTACTATAGTTTTTATTAGTTATCCCAGCTGGGATTTATGTCTCAGGAAT
AAAGAGCCAAGAATGGGAGGAGTTACCCTCGAAAGATCCAGGTCATGTGGTGCAGGGCAG
GGAATATGGCTGACTCAATCTCTTTGCCCATAGAGCCTCAGAGTATCAGATCTTAGCACTC
TAAGGAGGGAGACTCAGAGGGTACAAGTCTTAGAAGTCTCCCTAGGGCTTGGTGCCCAGC
AAATATATGCTGTTTGTGACTTCCCTAATACCAGGTACAGGCCAACACAAAGGACCTGTCC
AAGGGAAACTCACGGCTCAGACCTGATCTATTTACAGGTTGAGTTTGGGTGAAGCCAAGAG
AGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGC
GTGGCCACCATGGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTTCGACATCCTGTGC
AAGACCCCCCCCCAAGGTGCTGGTGCGGCAGTTCGTGGAGAGATTCGAGAGGCCCAGCGG
CGAGAAGATCGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGATGATCACCCACAA
CGGCACCGCCATCAAGAGGGCCACCTTCATGAGCTACAACACCATCATCAGCAACAGCCT
GAGCTTCGACATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGACCCAGAAGGCCACCAT
CCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCCTGGGAGTTCACCATCATCCCTTACAA
CGGCCAGAAGCACCAGAGCGACATCACCGACATCGTGTCCAGCCTGCAGCTGCAGTTCGA
GAGCAGCGAGGAGGCCGACAAGGGCAACAGCCACAGCAAGAAGATGCTGAAGGCCCTGC
TGTCCGAGGGCGAGAGCATCTGGGAGATCACCGAGAAGATCCTGAACAGCTTCGAGTACA

FIG. 14 cont'd

CCAGCAGGTTCACCAAGACCAAGACCCTGTACCAGTTCCTGTTCCTGGCCACATTCATCAA
CTGCGGCAGGTTCAGCGACATCAAGAACGTGGACCCCAAGAGCTTCAAGCTGGTGCAGAA
CAAGTACCTGGGCGTGATCATTCAGTGCCTGGTGACCGAGACCAAGACAAGCGTGTCCAG
GCACATCTACTTTTTCAGCGCCAGAGGCAGGATCGACCCCCTGGTGTACCTGGACGAGTT
CCTGAGGAACAGCGAGCCCGTGCTGAAGAGAGTGAACAGGACCGGCAACAGCAGCAGCA
ACAAGCAGGAGTACCAGCTGCTGAAGGACAACCTGGTGCGCAGCTACAACAAGGCCCTGA
AGAAGAACGCCCCCTACCCCATCTTCGCTATCAAGAACGGCCCTAAGAGCCACATCGGCA
GGCACCTGATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAGCTGACAAACGTGGTGG
GCAACTGGAGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCACCTACACCCACCAGATC
ACCGCCATCCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTACGCCTACGACCCCATC
AGCAAGGAGATGATCGCCCTGAAGGACGAGACCAACCCCATCGAGGAGTGGCAGCACAT
CGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCATCAGATACCCCGCCTGGAACGGCATCA
TCAGCCAGGAGGTGCTGGACTACCTGAGCAGCTACATCAACAGGCGGATCTGAGAATTCG
ATATCAAGCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTC
TTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTA
TTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATG
AGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAA
CCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCC
CCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGG
GCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTT
GGCTGCTCGCCTATGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTT
CGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTC
CGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCATC
GATACCGAGCGCTGCTCGAGAGATCTACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCC
TCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCCCACCAGCCTTGTCCTAATAAAATTAAGT
TGCATCATTTTGTCTGACTAGGTGTCCTTCTATAATATTATGGGGTGGAGGGGGGTGGTAT
GGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGGCCTGCGGGGTCTATTGGGAACCA
AGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAATCTCCGCCTCCTGGGTTCAAGCG
ATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCAGGCATGCATGACCAGGCTCAGC
TAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCCAGGCTGGTCTCCAACTC
CTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGGGATTACAGGCGTGAACC
ACTGCTCCCTTCCCTGTCCTTCTGATTTTGTAGGTAACCACGTGCGGACCGAGCGGCCG
(SEQ ID NO: 84)

AiP1092 (length between ITRs: 3092 bp):
GCGGCCGCACGCGTATCCCTGGAGATGAGGAGTCCTCTCTGGCAGGGTCCCCTCACTCTA
GAGCAGCCCCTATCCCAGGCCCCCTAGGAGTCTCTAATTAAAGGGCCGGCACGCCCCTCT
GGGACTCATTAGGCCCGCTGTGCAGAGAACATTTAATCATTGCTCAGAGCATCGATTGGAA
AATCAATTTCTTTGTCTCTTCGCACGAGGCGCGCTGGAGAAGTGGGGGGAGTGCTGACCT
CCTTCTGCTGCCGTGTAAAGCGCTGCACATTTAATCAGGGAACAGAAATCAATTAGCCACT
TACGAGGTTGGCTTTAGTTACCGAGTCGGCAAGGCCCGCGCCACAGCTCAGCCGCTGACA
GTAGCGAATCTCCTCCTCTCGGCCCTGCTGCATGGCTCTGTCTCCCTCCCTGTATCTCTCT
GGCTTCCTTCTTTCCCAGAGTGCTCTGGGTTCTCACCATCTTGGCAGATCCTCACAGAACT
CCAAACAAGTCCCGAGAAGCCTTCCTAATGCCCAGTCTCCTCGGCCACCTTCTTGTTCTCA
GCTCTAGACGTTTCAAGAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATT
GCTTACATTTGCTTCTGGCGTGGCCACCATGGCTCCTAAGAAGAAGAGGAAGGTGATGAG
CCAGTTCGACATCCTGTGCAAGACCCCCCCCAAGGTGCTGGTGCGGCAGTTCGTGGAGA
GATTCGAGAGGCCCAGCGGCGAGAAGATCGCCAGCTGTGCCGCCGAGCTGACCTACCTG

FIG. 14 cont'd

TGCTGGATGATCACCCACAACGGCACCGCCATCAAGAGGGCCACCTTCATGAGCTACAAC
ACCATCATCAGCAACAGCCTGAGCTTCGACATCGTGAACAAGAGCCTGCAGTTCAAGTACA
AGACCCAGAAGGCCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCCTGGGAG
TTCACCATCATCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACCGACATCGTGTCCA
GCCTGCAGCTGCAGTTCGAGAGCAGCGAGGAGGCCGACAAGGGCAACAGCCACAGCAAG
AAGATGCTGAAGGCCCTGCTGTCCGAGGGCGAGAGCATCTGGGAGATCACCGAGAAGAT
CCTGAACAGCTTCGAGTACACCAGCAGGTTCACCAAGACCAAGACCCTGTACCAGTTCCT
GTTCCTGGCCACATTCATCAACTGCGGCAGGTTCAGCGACATCAAGAACGTGGACCCCAA
GAGCTTCAAGCTGGTGCAGAACAAGTACCTGGGCGTGATCATTCAGTGCCTGGTGACCGA
GACCAAGACAAGCGTGTCCAGGCACATCTACTTTTTCAGCGCCAGAGGCAGGATCGACCC
CCTGGTGTACCTGGACGAGTTCCTGAGGAACAGCGAGCCCGTGCTGAAGAGAGTGAACA
GGACCGGCAACAGCAGCAGCAACAAGCAGGAGTACCAGCTGCTGAAGGACAACCTGGTG
CGCAGCTACAACAAGGCCCTGAAGAAGAACGCCCCCTACCCCATCTTCGCTATCAAGAAC
GGCCCTAAGAGCCACATCGGCAGGCACCTGATGACCAGCTTTCTGAGCATGAAGGGCCTG
ACCGAGCTGACAAACGTGGTGGGCAACTGGAGCGACAAGAGGGCCTCCGCCGTGGCCAG
GACCACCTACACCCACCAGATCACCGCCATCCCCGACCACTACTTCGCCCTGGTGTCCAG
GTACTACGCCTACGACCCCATCAGCAAGGAGATGATCGCCCTGAAGGACGAGACCAACCC
CATCGAGGAGTGGCAGCACATCGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCATCAGAT
ACCCCGCCTGGAACGGCATCATCAGCCAGGAGGTGCTGGACTACCTGAGCAGCTACATCA
ACAGGCGGATCTGAGAATTCGATATCAAGCTTATCGATAATCAACCTCTGGATTACAAAATT
TGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGC
TTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAA
ATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGT
GTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCT
CCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTG
CCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTC
GGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTATGTTGCCACCTGGATTCTGCGCGG
GACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCT
GCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTC
CCTTTGGGCCGCCTCCCCGCATCGATACCGAGCGCTGCTCGAGAGATCTACGGGTGGCAT
CCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCCCACC
AGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAATATT
ATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGGC
CTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGGCTCACTGCAAT
CTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTTGGGATTCCA
GGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATAT
TGGCCAGGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAATT
GCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTTCTGATTTTGTAGGTAACC
ACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 85)

CN2581 (length between ITRs: 2307 bp):
GCGGCCGCACGCGTTATGATGTGCCAGGCTTGGGAGAAACACCACAAGCAAAGCCAAAAT
AGGTGGCCTAGAACTTCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCA
GCTGCCGGGTGCCGCATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGAT
GACAGCTGGCTCTTCACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATA
ATTCATATTTAATTACTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAAT
TTGCTATTGACAAACTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATG
AAAAATGAAGCTTATTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAA

FIG. 14 cont'd

AAGCCAAACTTGAATGCAAAAAAATATTGAATATGAAAAGAAAAACATTTGCAGTAAAGCTT
GTTCTGAGCTCGATTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCTTGGG
GGGGGAGTTGGAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTC
AGAAGCATCCCCGGATCCTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGG
GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC
CCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCA
CCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGAC
TTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAA
GGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG
AGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG
GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACT
CAGATCTGGAGGCTCCGGAGGCCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGG
GCTCCAAGAAGGCGGTGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGC
CGCAAGGAGAGCTATTCCATCTATGTGTATAAGGTTCTGAAGCAGGTCCACCCTGACACCG
GCATTTCGTCCAAGGCCATGGGCATCATGAACTCGTTTGTGAACGACATTTTCGAGCGCAT
CGCAGGTGAGGCTTCCCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCAGGGA
GATCCAGACGGCCGTGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCG
AGGGTACTAAGGCCATCACCAAGTACACCAGCGCTAAGTAATGAGTCGACGGCGCGCCCC
TGCAGGGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGT
ATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCAT
GCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCC
ACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGG
CACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGT
TGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCC
TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 86)

CN2782 (length between ITRs: 2675 bp):
GCGGCCGCACGCGTAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAG
CTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGA
GAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTT
GAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAAC
CACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATA
CTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAG
TTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAA
AAAATATTGAATATGAAAAGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACT
TCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCG
CATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTT
CACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTA
CTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAA
CTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTA
TTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAA

FIG. 14 cont'd

TGCAAAAAAATATTGAATATGAAAAGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCT
AGAACTTCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGG
TGCCGCATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGG
CTCTTCACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTT
AATTACTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGA
CAAACTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAG
CTTATTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACT
TGAATGCAAAAAAATATTGAATATGAAAAGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAG
AGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGT
CGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCC
CTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCG
ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGAC
AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCT
GGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG
TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTC
CTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCC
CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTC
GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC
CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
TGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
87)

CN3407 (length between ITRs: 2479 bp):
GCGGCCGCACGCGTTGAGCTTCAACCAAATCAGGCATTGATGGATTTTATAGTTTGATTAA
CAAAGATAATAGCAAACCCCAGATTTAGTTTAAACATAAAAAGTATTAAGGTTGTATCCTGC
TTGTATAGCATATGCAAATGACCTCGTTTCTGCTACTGCATTTGGAAATGTAGCAGAAGAAA
AAAAAAAGGCACTTCAATTGCAGCTCTCATCAGTTATTCACTGTATCCAGGCCTCTCAATTG
TGTTCTTTTCTTTAATGCAATAGCAAGCAGCAATCACCCAGCTGTGCTTGGTAGAGTGAACT
ATATACACATCTATATTGAGATTTCATACACACATAACATAAAAGCGAGAGAAAAAGCCTCA
AGAATGTTTGGCCCATTGCAAATCACACAAAAGGACTAATGAATCTCTCTCCAAATGGATCT
GTAGTGACCATCTGTAAGCCTTGATTGATTCATATTCCATAACGGTATCAGCATCCAGGAAG
TGATTACTTCAAGGTGCAACACAACTTCCCCTATGAAAGCTCAGTCTCTTTAATCATACCTA
GTCAGTATCTGTCACGGGGATAAACTAAGGCAGAGCTCGGGCTGGGCATAAAAGTCAGGG
CAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACC
GGTCGCCACCATGGTCTCCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGATGTCAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTG
CCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCC

FIG. 14 cont'd

CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCG
ACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCG
GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
CTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAG
CGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGA
CGAGCTGTACAAGGGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGGGGATGT
AGAGGAGAATCCAGGGCCTGGTGCTAGTGGAGACTACAAAGACCATGACGGAGATTATAA
AGATCATGACATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGATCTGGAGGCTC
CGGAGGCCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGG
TGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTAT
TCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGG
CCATGGGCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTC
CCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCG
TGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCC
ATCACCAAGTACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCCGCGAATTCGATATCAT
AATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCT
TTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGC
TTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCG
CCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCT
CGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG
CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTG
CATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCA
AGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCC
GC (SEQ ID NO: 88)

CN3408 (length between ITRs: 2515 bp):
GCGGCCGCACGCGTACATTTGCAGTAAAGCTTGTTCTTTTTCTTGAAGTATATTTTAAGATT
TTGAGTTCTACTATCATTAAAGACAGATAATTAATAGTTTATTTTTATTTACTTTTGTTAGTAG
TGACTTGGTGCTATGAGCCATATTTTGCTGTTGCTGTTGTTACTGGTAGTTTTTGTAATTCT
GGGGCTAAAACTTGGGGTCTGGTATGCTGTCATTTACCAGTGAGCTATACCCTGGATATTA
TGATTTAGATGAATGTGAAATATCACCCCAGACATACATATACTAAACACTTGGCCCTTGGC
CCATGATGCTAAATGGAGGAGATAGAAGCTTTTGGGGCACAGCCTAGTGGAAGGAAATGA
GGTCAAATGACATGTACTCTGAAAGGAATATGGGTATTCTGGGCTTGCGTTATTCTCTCTCT
CCCTCTCTCTCCCTCTCTCTCCCTCTCCCTCTCTCTTTCTCCTTTTCTCTTTCTCTCCTCGCC
TTGTTTTCCAGCTGCCAGAAGGTAGGCCTCTTCTCTGCTGAATATCTGTGTCATGTTATGCA
CCAACACAGTACTAACTGTCATGTTATACCTAGTGGCCAGGTAACCATGGACCAAAATGGC
AGAGCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGC
TTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTCTCCAAGGGC
GAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGATGTCAACGG
CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACC
CTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTC
TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAAGACGAC
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT

FIG. 14 cont'd

ACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGG
CCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACC
AGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGC
TACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAG
TTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGCTCTGGTGCT
ACCAACTTCTCACTGTTGAAACAGGCAGGGGATGTAGAGGAGAATCCAGGGCCTGGTGCT
AGTGGAGACTACAAAGACCATGACGGAGATTATAAAGATCATGACATCGATTACAAGGATG
ACGATGACAAGTCCGGACTCAGATCTGGAGGCTCCGGAGGCCCAGAGCCAGCGAAGTCT
GCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAAGGCGCAGAAGAAAGGCGG
CAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCTATGTGTACAAGGTTCTGAA
GCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGCCATGGGCATCATGAATTCGTTTGT
GAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTCCCGCCTGGCGCATTACAACAAGCG
CTCGACCATCACCTCCCGGGAGATCCAGACGGCCGTGCGCCTGCTGCTGCCTGGGGAGT
TGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACCAAGTACACCAGCGCTAAGT
AATGAGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTG
TGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTT
TAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAAT
CCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAG
GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCT
AGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCC
ACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCA
TTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA
GCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 89)

CN3409 (length between ITRs: 2565 bp):
GCGGCCGCACGCGTTGCAAAATAAAGATTTCTTGGGATACAGAGAAAAAAACAAATCTGAC
AGGAGAGGAAGAAGCACCCGGTGGGCTATAACGGTGCAATTCAGCTGATTATATGTTACAA
GTAACAAGGACGAGAAAAAATGTTATTTCTTTGAAAATAAAACTAACCAGGCCATACATATT
TAACAGGACTGCATGAGAGAAGAAGAAGCCAGCTGCAGGAGTGACTGTGGGGGGGAGGG
GGAACTTGACAAAAAAAGCAAAATGGCAGTCCTGCTTCCAAAGTCCTCAAGGTCACAGTTA
TTTGGGCATTCTTGCGGGCACTGCTTATACAAGAATGTGCTTTCAGTCAAGGCTTTCTAATA
GATTCTCAAAATTTGGGACAAATGTTATTTTTGTATCTGTAGAAATGTACTGATTCAGAAAGA
TCTTTGAGCAATACAGATGTTAAAACATTTAAGTCACAAAATGGGTCTATTTAATCAATGCGA
CTAGTTTGGAACATTATTCAAACTGCCAGAAATACAATGTAAATGAAACCTCAGGCCAATAT
TTTGGAGCCCTAAAAGATTTGATGGCTAATTTTATCGTAGACACTAATTATAAATAGGAGAC
CCCAGGATGGGACTAGAAAACCAAGCCAGCTTTTTAATTTACCCCTCCAGGACTTTGCTGA
GCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGAT
CCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTCTCCAAGGGCGAGGAGCTG
TTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGATGTCAACGGCCACAAGTT
CAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGA
TCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTAC
GGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCC
GCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAAGACGACGGCAACTAC
AAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA
GGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAA
CAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAA
GATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACA
CCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCC

FIG. 14 cont'd

AAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACC
GCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGCTCTGGTGCTACCAACTTC
TCACTGTTGAAACAGGCAGGGGATGTAGAGGAGAATCCAGGGCCTGGTGCTAGTGGAGAC
TACAAAGACCATGACGGAGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACA
AGTCCGGACTCAGATCTGGAGGCTCCGGAGGCCCAGAGCCAGCGAAGTCTGCTCCCGCC
CCGAAAAAGGGCTCCAAGAAGGCGGTGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCG
CAAGCGCAGCCGCAAGGAGAGCTATTCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCA
CCCTGACACCGGCATTTCGTCCAAGGCCATGGGCATCATGAATTCGTTTGTGAACGACATT
TTCGAGCGCATCGCAGGAGAGGCTTCCCGCCTGGCGCATTACAACAAGCGCTCGACCATC
ACCTCCCGGGAGATCCAGACGGCCGTGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCA
CGCCGTGTCCGAGGGTACTAAGGCCATCACCAAGTACACCAGCGCTAAGTAATGAGGCGC
GCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTG
ACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTT
GTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGT
TCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGC
TGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGC
CATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGT
CCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGG
GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGA
GATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 90)

CN2580 (length between ITRs: 2378 bp):
GCGGCCGCACGCGTTATGATGTGCCAGGCTTGGGAGAAACACCACAAGCAAAGCCAAAAT
AGGTGGCCTAGAACTTCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCA
GCTGCCGGGTGCCGCATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGAT
GACAGCTGGCTCTTCACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATA
ATTCATATTTAATTACTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAAT
TTGCTATTGACAAACTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATG
AAAAATGAAGCTTATTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAA
AAGCCAAACTTGAATGCAAAAAAATATTGAATATGAAAAGAAAAACATTTGCAGTAAAGCTT
GTTCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTT
CTGGGATCCAGATCTTTCGAAGCTAGCACCATGGTGCCCAAGAAGAAGAGGAAAGTCTCC
AACCTGCTGACTGTGCACCAAAACCTGCCTGCCCTCCCTGTGGATGCCACCTCTGATGAA
GTCAGGAAGAACCTGATGGACATGTTCAGGGACAGGCAGGCCTTCTCTGAACACACCTGG
AAGATGCTCCTGTCTGTGTGCAGATCCTGGGCTGCCTGGTGCAAGCTGAACAACAGGAAA
TGGTTCCCTGCTGAACCTGAGGATGTGAGGGACTACCTCCTGTACCTGCAAGCCAGAGGC
CTGGCTGTGAAGACCATCCAACAGCACCTGGGCCAGCTCAACATGCTGCACAGGAGATCT
GGCCTGCCTCGCCCTTCTGACTCCAATGCTGTGTCCCTGGTGATGAGGAGAATCAGAAAG
GAGAATGTGGATGCTGGGGAGAGAGCCAAGCAGGCCCTGGCCTTTGAACGCACTGACTTT
GACCAAGTCAGATCCCTGATGGAGAACTCTGACAGATGCCAGGACATCAGGAACCTGGCC
TTCCTGGGCATTGCCTACAACACCCTGCTGCGCATTGCCGAAATTGCCAGAATCAGAGTGA
AGGACATCTCCCGCACCGATGGTGGGAGAATGCTGATCCACATTGGCAGGACCAAGACCC
TGGTGTCCACAGCTGGTGTGGAGAAGGCCCTGTCCCTGGGGGTTACCAAGCTGGTGGAG
AGATGGATCTCTGTGTCTGGTGTGGCTGATGACCCCAACAACTACCTGTTCTGCCGGGTCA
GAAAGAATGGTGTGGCTGCCCCTTCTGCCACCTCCCAACTGTCCACCCGGGCCCTGGAAG
GGATCTTTGAGGCCACCCACCGCCTGATCTATGGTGCCAAGGATGACTCTGGGCAGAGAT
ACCTGGCCTGGTCTGGCCACTCTGCCAGAGTGGGTGCTGCCAGGGACATGGCCAGGGCT
GGTGTGTCCATCCCTGAAATCATGCAGGCTGGTGGCTGGACCAATGTGAACATTGTGATG

FIG. 14 cont'd

AACTACATCAGAAACCTGGACTCTGAGACTGGGGCCATGGTGAGGCTGCTCGAGGATGGG
GACTAATGAGGCGCGCCGCGGCCTTAAAGAGACCGGTTCACTGTGACAGTAAAAGAGACC
GGTTCACTGTGAGAATGAAAGAGACCGGTTCACTGTGATCGGAAAAGAGACCGGTTCACT
GTGAGCGGCCTTGAAACCCAGCAGACAATGTAGCTCAGTAGAAACCCAGCAGACAATGTA
GCTGAATGGAAACCCAGCAGACAATGTAGCTTCGGAGAAACCCAGCAGACAATGTAGCTG
TCGACGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTA
TTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATG
CTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCA
CGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGC
ACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTT
GTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG
GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCG
GACCGAGCGGCCGC (SEQ ID NO: 91)

CN2825 (length between ITRs: 2625 bp):
GCGGCCGCACGCGTTATGATGTGCCAGGCTTGGGAGAAACACCACAAGCAAAGCCAAAAT
AGGTGGCCTAGAACTTCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCA
GCTGCCGGGTGCCGCATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGAT
GACAGCTGGCTCTTCACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATA
ATTCATATTTAATTACTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAAT
TTGCTATTGACAAACTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATG
AAAAATGAAGCTTATTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAA
AAGCCAAACTTGAATGCAAAAAAATATTGAATATGAAAAGAAAAACATTTGCAGTAAAGCTT
GTTCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTT
CTGGGATCCAGATCTTTCGAAGCTAGCCACCATGGCTCCTAAGAAGAAGAGGAAGGTGAT
GAGCCAGTTCGACATCCTGTGCAAGACCCCCCCCCAAGGTGCTGGTGCGGCAGTTCGTGG
AGAGATTCGAGAGGCCCAGCGGCGAGAAGATCGCCAGCTGTGCCGCCGAGCTGACCTAC
CTGTGCTGGATGATCACCCACAACGGCACCGCCATCAAGAGGGCCACCTTCATGAGCTAC
AACACCATCATCAGCAACAGCCTGAGCTTCGACATCGTGAACAAGAGCCTGCAGTTCAAGT
ACAAGACCCAGAAGGCCACCATCCTGGAGGCCAGCCTGAAGAAGCTGATCCCCGCCTGG
GAGTTCACCATCATCCCTTACAACGGCCAGAAGCACCAGAGCGACATCACCGACATCGTG
TCCAGCCTGCAGCTGCAGTTCGAGAGCAGCGAGGAGGCCGACAAGGGCAACAGCCACAG
CAAGAAGATGCTGAAGGCCCTGCTGTCCGAGGGCGAGAGCATCTGGGAGATCACCGAGA
AGATCCTGAACAGCTTCGAGTACACCAGCAGGTTCACCAAGACCAAGACCCTGTACCAGTT
CCTGTTCCTGGCCACATTCATCAACTGCGGCAGGTTCAGCGACATCAAGAACGTGGACCC
CAAGAGCTTCAAGCTGGTGCAGAACAAGTACCTGGGCGTGATCATTCAGTGCCTGGTGAC
CGAGACCAAGACAAGCGTGTCCAGGCACATCTACTTTTTCAGCGCCAGAGGCAGGATCGA
CCCCCTGGTGTACCTGGACGAGTTCCTGAGGAACAGCGAGCCCGTGCTGAAGAGAGTGA
ACAGGACCGGCAACAGCAGCAGCAACAAGCAGGAGTACCAGCTGCTGAAGGACAACCTG
GTGCGCAGCTACAACAAGGCCCTGAAGAAGAACGCCCCCTACCCCATCTTCGCTATCAAG
AACGGCCCTAAGAGCCACATCGGCAGGCACCTGATGACCAGCTTTCTGAGCATGAAGGGC
CTGACCGAGCTGACAAACGTGGTGGGCAACTGGAGCGACAAGAGGGCCTCCGCCGTGGC
CAGGACCACCTACACCCACCAGATCACCGCCATCCCCGACCACTACTTCGCCCTGGTGTC
CAGGTACTACGCCTACGACCCCATCAGCAAGGAGATGATCGCCCTGAAGGACGAGACCAA
CCCCATCGAGGAGTGGCAGCACATCGAGCAGCTGAAGGGCAGCGCCGAGGGCAGCATCA
GATACCCCGCCTGGAACGGCATCATCAGCCAGGAGGTGCTGGACTACCTGAGCAGCTACA
TCAACAGGCGGATCTGAGTCGACGGCGCGCCGCGGCCTTAAAGAGACCGGTTCACTGTG

FIG. 14 cont'd

ACAGTAAAAGAGACCGGTTCACTGTGAGAATGAAAGAGACCGGTTCACTGTGATCGGAAAA
GAGACCGGTTCACTGTGAGCGGCCTTGAAACCCAGCAGACAATGTAGCTCAGTAGAAACC
CAGCAGACAATGTAGCTGAATGGAAACCCAGCAGACAATGTAGCTTCGGAGAAACCCAGC
AGACAATGTAGCTGTCGACGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGA
AAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAAT
GCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTG
GTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGG
CTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTT
GCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTC
CCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCT
ATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 92)

CN3270 (length between ITRs: 2205 bp):
GCGGCCGCACGCGTTATCTTAGAGTGGGAAGATTTGAGAAGTGCCATGGTTAATATGACT
GACTTTTTATTCTTATTTCTTTTAATTTCATGGTTCTAAATCCGAATTTAATCATAGTACCCAG
AAAAGCAGAGGTGTAGAGGTTCACAGTGGGAGTTGTAATCTAGCCCTATTCATTTTGACCT
CAAAACCCAAATTATTTATAACAAATTATTTCCTATTCTTTCCTTCACTATTCAGGAACATCTG
TCCACCACTTACATGATCACTTATCTTGCTATTGTGTCATTTTGATGAAAAAGAATTTTTTCT
AAATATCTAAATACAAGGCCCCATATTAACAGTGCTTTTTAAATCCCCACAGATGTGGGAGA
TGACCCCTTTCCATCCCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTTGTTCCAATA
AAGAGATACAATTTTATTCATTAATTTGTGTATTCATTTAGCAAATCACTTTAGAGTCTTATTA
TATCAGGATTTTGGGGTCTATTTTAGTATATCTTTTTGTATTTCTTGGAACCTCTCCAATTATT
CTAGACTCTTTCAAAGGTTGGTGATCAATATTAGACATTATTATGAAAAGAATCTTACTTGCT
AAAAGGGTTAGATGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT
ACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAG
CAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACG
TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT
GACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGA
ACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGC
ATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
ACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAG
TCGACGGCGCGCCGCGGCCTTAAAGAGACCGGTTCACTGTGACAGTAAAAGAGACCGGTT
CACTGTGAGAATGAAAGAGACCGGTTCACTGTGATCGGAAAAGAGACCGGTTCACTGTGA
GCGGCCTTGAAACCCAGCAGACAATGTAGCTCAGTAGAAACCCAGCAGACAATGTAGCTG
AATGGAAACCCAGCAGACAATGTAGCTTCGGAGAAACCCAGCAGACAATGTAGCTGTCGA
CGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTT
AACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATT
GCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGC
GGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTG
ACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATA

FIG. 14 cont'd

AAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG
GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 93)

CN3316 (length between ITRs: 1960 bp):
GCGGCCGCACGCGTAGTGACTTGGTGCTATGAGCCATATTTTGCTGTTGCTGTTGTTACTG
GTAGTTTTTGTAATTCTGGGGCTAAAACTTGGGGTCTGGTATGCTGTCATTTACCAGTGAG
CTATACCCTGGATATTATGATTTAGATGAATGTGAAATATCACCCCAGACATACATATACTAA
ACACTTGGCCCTTGGCCCATGATGCTAAATGGAGGAGATAGAAGCTTTTGGGGCACAGCC
TAGTGGAAGGAAATGAGGTCAAATGACATGTACTCTGAAAGGAATATGGGTATTCTGGGCT
TGCGTTATTCTCTCTCTCCCTCTCTCTCCCTCTCTCTCCCTCTCCCTCTCTCTTTCTCCTTTT
CTCTTTCTCTCCTCGCCTTGTTTTCCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGC
CATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGC
CACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGC
AGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGTCGACGGCGCGCCGCGGCCTTAAAGAGACCGGTTCACTGTGACAGTAAA
AGAGACCGGTTCACTGTGAGAATGAAAGAGACCGGTTCACTGTGATCGGAAAAGAGACCG
GTTCACTGTGAGCGGCCTTGAAACCCAGCAGACAATGTAGCTCAGTAGAAACCCAGCAGA
CAATGTAGCTGAATGGAAACCCAGCAGACAATGTAGCTTCGGAGAAACCCAGCAGACAAT
GTAGCTGTCGACGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTG
ACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTT
GTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGT
TCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGC
TGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGC
CATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGT
CCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGG
GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGA
GATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 94)

CN3271 (length between ITRs: 2895 bp):
GCGGCCGCACGCGTAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACTTCCAG
CTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCGCATGA
GAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTTCACTTT
GAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTACTTAAC
CACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAACTCATA
CTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTATTTTAG
TTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAATGCAAA
AAAATATTGAATATGAAAAGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCTAGAACT

FIG. 14 cont'd

TCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGGTGCCG
CATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGGCTCTT
CACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTTAATTA
CTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGACAAA
CTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAGCTTA
TTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACTTGAA
TGCAAAAAAATATTGAATATGAAAAGAGAAACACCACAAGCAAAGCCAAAATAGGTGGCCT
AGAACTTCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCAGCTGCCGGG
TGCCGCATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGATGACAGCTGG
CTCTTCACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATAATTCATATTT
AATTACTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAATTTGCTATTGA
CAAACTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATGAAAAATGAAG
CTTATTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAAAAGCCAAACT
TGAATGCAAAAAAATATTGAATATGAAAAGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAG
AGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGT
CGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCC
CTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCG
ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGAC
AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCTTAAAGAGACCGGTTCACTGTGACAGT
AAAAGAGACCGGTTCACTGTGAGAATGAAAGAGACCGGTTCACTGTGATCGGAAAAGAGA
CCGGTTCACTGTGAGCGGCCTTGAAACCCAGCAGACAATGTAGCTCAGTAGAAACCCAGC
AGACAATGTAGCTGAATGGAAACCCAGCAGACAATGTAGCTTCGGAGAAACCCAGCAGAC
AATGTAGCTGTCGACGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGA
TTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCC
TTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTT
AGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTC
GGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGC
CAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCC
ACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTAT
TCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGG
CATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 95)

CN3793 (length between ITRs: 2409 bp):
GCGGCCGCACGCGTCAGTTTCCAGCGTGGTTGTTGATGAGGCTCAGAGAAAAGACTCTAA
AGTTATGATGGGAAATTACCATGCCATTCATCATCATACACATTCACCTCACACTTTCTGAG
TCTCCTATACAAAGTCAGTTCTCTGCCAAGGGCATGGAAGAGCGAGGAACAGGATGTTAG
GAAGGGCTGACAGCGCTGTTTTAGCCTGACAGGCAGATTTACAACAGGAGAATGAATGTA
CCACTTGTATAAGAAGGCCATGCGGCACTGCTAATGCACAAGTTGGCAGTACATCAACATC
TCTATCGTCCTCATATTCATGAAGCAGAGAACGGAAATGGCACACTGCTTGTACCGGCGAA

FIG. 14 cont'd

TAACCAAAGTGAACGCCCTACGGCTGCCATTCACTGTGTCCTTCCAAAAGCATTTTTCTACT
GAGCTCTTCCCAGAGATTTAGGGTTTGCTTAGACAGGTCTTATGACGCCACGTGATAGGTC
ATTCTTCTGTTCTGAGGAGCTTGGAGAAGATCGAGCTCGGGCTGGGCATAAAAGTCAGGG
CAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACC
GGTCGCCACCATGGTCTCCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGATGTCAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTG
CCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCC
CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCG
ACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCG
GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
CTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAG
CGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGA
CGAGCTGTACAAGGGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGGGGATGT
AGAGGAGAATCCAGGGCCTGGTGCTAGTGGAGACTACAAAGACCATGACGGAGATTATAA
AGATCATGACATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGATCTGGAGGCTC
CGGAGGCCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGG
TGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTAT
TCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGG
CCATGGGCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTC
CCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCG
TGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCC
ATCACCAAGTACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCCGCGAATTCGATATCAT
AATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCT
TTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGC
TTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCG
CCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCT
CGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG
CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTG
CATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCA
AGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCC
GC (SEQ ID NO: 96)

CN3794 (length between ITRs: 2343 bp):
GCGGCCGCACGCGTGTAGAACATACTTATTAACACATTCGTACATAAAATAAAATTCTACTC
TCCCGACCTTTTCCTCACCATCTTGCTTTTCAACGTATGGCGTTAGACCTAACAGCGAGTC
CACTTCTTCCCCTTTCATTCTGTAGCAAGAACACACGGCTCACTGTAACAGGGACTTGGCT
GTGGGTTGCAGACTGGCTTCCTGCTGCCTCCACTTGAGCCCCACACAGCTGTGGCTTTGT
GTTTACAACCCTCCAGGCTGCCATTCATTCGGTGCTGTGGGCTCATGTACTGGAAGACAGC
TTCCATCACAACCTTCCCGTCCCAGCAGGAGAACTCCCTTGCTTCCTTGGGGAACATTTGC
TTGCTCCTGCTGCTTGGCTCTTCCCACTTTTGCCTCACTCTGGAGTTTCTCTCTCCCGTTTT
GAATTCTAGTAGTAAACACATGGCCGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCC
ATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCC
ACCATGGTCTCCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCT
GGACGGCGATGTCAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCA

FIG. 14 cont'd

CCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGG
CCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCA
CATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCAC
CATCTTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGA
CACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCT
GGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCA
GAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGC
AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC
GACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGAT
CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGGGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGGGGATGTAGAGGA
GAATCCAGGGCCTGGTGCTAGTGGAGACTACAAAGACCATGACGGAGATTATAAAGATCAT
GACATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGATCTGGAGGCTCCGGAGGC
CCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAA
GGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCT
ATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGCCATGG
GCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTCCCGCCT
GGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCGTGCGCC
TGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACC
AAGTACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAA
CCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACG
CTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCAT
TTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGC
CTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGA
GATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTC
CTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCG
CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGG
GGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 97)

CN3795 (length between ITRs: 2411 bp):
GCGGCCGCACGCGTCCAGTGAGATCTTCCACCAGCAGAACTCATGGACACAAACTAGACA
GCTCACTTCTTGCCTGTATTCCAGGAGTGGCTTTTTCTCTACTCCTGTACTGATGCCAGTCA
TTCAGAGTGCACTCAAGACACTTGACCCACATCAGTTAAGAGAATGAAAATCAAGCTCTGA
AAGCCATTAGCTTCTATTGCACACCCAGAAAACAGGCTCATCAAACACCTTCTTATGGTAAT
GCCTTTGATCAAAAGGAGGGTTAATTCAACAAATGGTTTGCACCGTGACCCCATCAAAGCC
TGAGCACCAGTGTCCTCATTTCCTTTCCCCTGGTGTATAATGAGTTGTTAGTCTGGCTCACC
TTGTCATCCCCATCATACTGCCATAATCCACATCTCTAAAGAGTGGATTACAACAGTCCCGT
CTGTGACACTCAGGACTGGCATCAAGGTTCCCAAGCTCTAGTCTATTGTGACATTGATACA
AATAGGGCTCAGAGTCTCACTGATCACACCGAGCTCGGGCTGGGCATAAAAGTCAGGGCA
GAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGG
TCGCCACCATGGTCTCCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGATGTCAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCC
GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA

FIG. 14 cont'd

CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGA
CAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGG
CGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGGGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGGGGATGTA
GAGGAGAATCCAGGGCCTGGTGCTAGTGGAGACTACAAAGACCATGACGGAGATTATAAA
GATCATGACATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGATCTGGAGGCTCC
GGAGGCCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGT
GACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATT
CCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTCGTCCAAGGC
CATGGGCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTCC
CGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCGT
GCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCA
TCACCAAGTACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCCGCGAATTCGATATCATA
ATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTT
TTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCT
TTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGC
CTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTC
GAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGC
ATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA
GGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCG
C (SEQ ID NO: 98)

CN3790 (length between ITRs: 2043 bp):
GCGGCCGCACGCGTTATCTTAGAGTGGGAAGATTTGAGAAGTGCCATGGTTAATATGACT
GACTTTTTATTCTTATTTCTTTTAATTTCATGGTTCTAAATCCGAATTTAATCATAGTACCCAG
AAAAGCAGAGGTGTAGAGGTTCACAGTGGGAGTTGTAATCTAGCCCTATTCATTTTGACCT
CAAAACCCAAATTATTTATAACAAATTATTTCCTATTCTTTCCTTCACTATTCAGGAACATCTG
TCCACCACTTACATGATCACTTATCTTGCTATTGTGTCATTTTGATGAAAAAGAATTTTTTCT
AAATATCTAAATACAAGGCCCCATATTAACAGTGCTTTTTAAATCCCCACAGATGTGGGAGA
TGACCCCTTTCCATCCCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTTGTTCCAATA
AAGAGATACAATTTTATTCATTAATTTGTGTATTCATTTAGCAAATCACTTTAGAGTCTTATTA
TATCAGGATTTTGGGGTCTATTTTAGTATATCTTTTTGTATTTCTTGGAACCTCTCCAATTATT
CTAGACTCTTTCAAAGGTTGGTGATCAATATTAGACATTATTATGAAAAGAATCTTACTTGCT
AAAAGGGTTAGATGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT
ACATTTGCTTCTGGGATCCAGATCTTTCCAAGCTAGCCACCATGGTGCCCAAGAAGAAGAG
GAAAGTCTCCAACCTGCTGACTGTGCACCAAAACCTGCCTGCCCTCCCTGTGGATGCCAC
CTCTGATGAAGTCAGGAAGAACCTGATGGACATGTTCAGGGACAGGCAGGCCTTCTCTGA
ACACACCTGGAAGATGCTCCTGTCTGTGTGCAGATCCTGGGCTGCCTGGTGCAAGCTGAA
CAACAGGAAATGGTTCCCTGCTGAACCTGAGGATGTGAGGGACTACCTCCTGTACCTGCA
AGCCAGAGGCCTGGCTGTGAAGACCATCCAACAGCACCTGGGCCAGCTCAACATGCTGCA
CAGGAGATCTGGCCTGCCTCGCCCTTCTGACTCCAATGCTGTGTCCCTGGTGATGAGGAG
AATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAAGCAGGCCCTGGCCTTTGAAC
GCACTGACTTTGACCAAGTCAGATCCCTGATGGAGAACTCTGACAGATGCCAGGACATCA
GGAACCTGGCCTTCCTGGGCATTGCCTACAACACCCTGCTGCGCATTGCCGAAATTGCCA
GAATCAGAGTGAAGGACATCTCCCGCACCGATGGTGGGAGAATGCTGATCCACATTGGCA

FIG. 14 cont'd

GGACCAAGACCCTGGTGTCCACAGCTGGTGTGGAGAAGGCCCTGTCCCTGGGGGTTACC
AAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGCTGATGACCCCAACAACTACCTG
TTCTGCCGGGTCAGAAAGAATGGTGTGGCTGCCCCTTCTGCCACCTCCCAACTGTCCACC
CGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGATCTATGGTGCCAAGGATGAC
TCTGGGCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAGAGTGGGTGCTGCCACCGA
CATGGCCAGGGCTGGTGTGTCCATCCCTGAAATCATGCAGGCTGGTGGCTGGACCAATGT
GAACATTGTGATGAACTACATCAGAAACCTGGACTCTGAGACTGGGGCCATGGTGAGGCT
GCTCGAAGATGGGGACTGAGGCGCGCCGAATTCAAGCTTCTCGAGAGATCTTCGACTGTG
CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAG
GTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 99)

CN3751 (length between ITRs: 1747 bp):
GCGGCCGCACGCGTTATCTTAGAGTGGGAAGATTTGAGAAGTGCCATGGTTAATATGACT
GACTTTTTATTCTTATTTCTTTTAATTTCATGGTTCTAAATCCGAATTTAATCATAGTACCCAG
AAAAGCAGAGGTGTAGAGGTTCACAGTGGGAGTTGTAATCTAGCCCTATTCATTTTGACCT
CAAAACCCAAATTATTTATAACAAATTATTTCCTATTCTTTCCTTCACTATTCAGGAACATCTG
TCCACCACTTACATGATCACTTATCTTGCTATTGTGTCATTTTGATGAAAAAGAATTTTTTCT
AAATATCTAAATACAAGGCCCCATATTAACAGTGCTTTTTAAATCCCCACAGATGTGGGAGA
TGACCCCTTTCCATCCCTGAAGATTGTAATTGGGCCAGTCTTTAGTACAGTTTGTTCCAATA
AAGAGATACAATTTTATTCATTAATTTGTGTATTCATTTAGCAAATCACTTTAGAGTCTTATTA
TATCAGGATTTTGGGGTCTATTTTAGTATATCTTTTTGTATTTCTTGGAACCTCTCCAATTATT
CTAGACTCTTTCAAAGGTTGGTGATCAATATTAGACATTATTATGAAAAGAATCTTACTTGCT
AAAAGGGTTAGATGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT
ACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCAATTCGCCACCATGACGAGTGATG
AGGTTCGCAAGAACCTGATGGACATGTTCAGGGATCGCCAGGCGTTTTCTGAGCATACCT
GGAAAATGCTTCTGTCCGTTTGCCGGTCGTGGGCGGCATGGTGCAAGTTGAATAAATTTGC
GGAATATTGCCTCAGTTTTGGCACCGAAATTTTAACCGTTGAGTACGGCCCATTGCCCATT
GGCAAAATTGTGAGTGAAGAAATTAATTGTTCTGTGTACAGTGTTGATCCAGAAGGGAGAG
TTTACACCCAGGCGATCGCCCAATGGCATGACCGGGGAGAGCAGGAAGTATTGGAATATG
AATTGGAAGATGGTTCAGTAATCCGAGCTACCTCTGACCACCGCTTTTTAACCACCGATTAT
CAACTGTTGGCGATCGAAGAAATTTTTGCTAGGCAACTGGACTTGTTGACTTTAGAAAATAT
TAAGCAAACTGAAGAAGCTCTTGACAACCATCGTCTTCCCTTTCCATTACTTGACGCTGGG
ACAATTAAATAACTCGAATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTG
GTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATC
ATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTG
CCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTG
GGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCT
GTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTT
CCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGG
TGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCT
CACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 100)

CN3752 (length between ITRs: 2071 bp):
GCGGCCGCACGCGTTATGATGTGCCAGGCTTGGGAGAAACACCACAAGCAAAGCCAAAAT
AGGTGGCCTAGAACTTCCAGCTTGAAATATGGGAGAGAATGAGGGAGGCACTGTAGAGCA
GCTGCCGGGTGCCGCATGAGAACAATTCTCCCTGCTCATAATTAATCCTACCTATTTCTGAT

FIG. 14 cont'd

GACAGCTGGCTCTTCACTTTGAACAAGCTAGTTAACAACTTTCTTCTCACATTGAGCAAATA
ATTCATATTTAATTACTTAACCACCAGTTACAAAATGAGAATCATCAAGGAATCACAATTAAT
TTGCTATTGACAAACTCATACTTTTAGCAGGCTGATTTCTACTTTATACTTAGATTGGTAATG
AAAAATGAAGCTTATTTTAGTTGATTGGTTGGACTTGTGTATGAATATTATCTATTATTTGAA
AAGCCAAACTTGAATGCAAAAAAATATTGAATATGAAAAGAAAAACATTTGCAGTAAAGCTT
GTTCTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTT
CTGGGATCCAGATCTTTCGAAGCTAGCAATTCGCCACCATGGTTAAAGTTATCGGTCGTCG
TTCCCTCGGAGTGCAAAGAATATTTGATATTGGTCTTCCCCAAGACCATAATTTTCTGCTAG
CCAATGGGGCGATCGCCGCCAATTGTTTTAACAAATCCAACCGGAAATGGTTTCCCGCAGA
ACCTGAAGATGTTCGCGATTATCTTCTATATCTTCAGGCGCGCGGTCTGGCAGTAAAAACT
ATCCAGCAACATTTGGGCCAGCTAAACATGCTTCATCGTCGGTCCGGGCTGCCACGACCA
AGTGACAGCAATGCTGTTTCACTGGTTATGCGGCGGATCCGAAAAGAAAACGTTGATGCC
GGTGAACGTGCAAAACAGGCTCTAGCGTTCGAACGCACTGATTTCGACCAGGTTCGTTCA
CTCATGGAAAATAGCGATCGCTGCCAGGATATACGTAATCTGGCATTTCTGGGGATTGCTT
ATAACACCCTGTTACGTATAGCCGAAATTGCCAGGATCAGGGTTAAAGATATCTCACGTAC
TGACGGTGGGAGAATGTTAATCCATATTGGCAGAACGAAAACGCTGGTTAGCACCGCAGG
TGTAGAGAAGGCACTTAGCCTGGGGGTAACTAAACTGGTCGAGCGATGGATTTCCGTCTC
TGGTGTAGCTGATGATCCGAATAACTACCTGTTTTGCCGGGTCAGAAAAAATGGTGTTGCC
GCGCCATCTGCCACCAGCCAGCTATCAACTCGCGCCCTGGAAGGGATTTTTGAAGCAACT
CATCGATTGATTTACGGCGCTAAGGATGACTCTGGTCAGAGATACCTGGCCTGGTCTGGA
CACAGTGCCCGTGTCGGAGCCGCGCGAGATATGGCCCGCGCTGGAGTTTCAATACCGGA
GATCATGCAAGCTGGTGGCTGGACCAATGTAAATATTGTCATGAACTATATCCGTAACCTG
GATAGTGAAACAGGGGCAATGGTGCGCCTGCTGGAAGATGGCGATTAGCTCGAATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTT
TACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTT
TCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCC
TGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCG
AGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 101)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/58812

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 15/86, C12N 15/861 (2022.01)

CPC - A01K 67/0275, A01K 2217/15, A01K 2227/105, A01K 2217/206, C12N 2750/14143, C12N 2830/008

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

B. FIELDS SEARCHED

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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2020/168279 A2 (ALLEN INSTITUTE) 20 August 2020 (20.08.2020) para [0005], [0032], [0063], [0072], [0080], [0081]	1-4, 12-15, 29-31, 34-37, 45-46, 70
A	GenBank submission AC123673.8, 23 May 2006 [online]. [Retrieved on 14 March 2022]. Retrieved from the internet <URL: https://cipweb.cardinal-ip.com/pctsrs/PCTSRS_DATA/PCT-US%2021-58812/20220311_075435_pct-us21-58812-6.rge.txt> nt 156408-157034	1-4, 12-15, 29-31, 34-37, 45-46, 70
A	GenBank submission FR242253/c, 12 September 2009 [online]. [Retrieved on 16 March 2022]. Retrieved from the internet <URL: https://www.ncbi.nlm.nih.gov/nuccore/258063011> nt 736-940	1-4, 12-15, 29-31, 34-37, 45-46, 70

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
16 March 2022	MAR 30 2022
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Kari Rodriquez Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/58812

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/58812

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

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

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

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

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

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

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

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

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

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-4, 12-15, 29-31, 34-37, 45-46 and 70, limted to SEQ ID NO: 6

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/58812

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

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

Group I+: Claims 1-46 and 70 directed to a composition comprising an artificial expression construct comprising (i) an enhancer, (ii) a promoter; and (iii) a heterologous encoding sequence, the expression construct further comprised in a viral vector or a cell. The expression construct will be searched to the extent that the enhancer is eHGT_475m (SEQ ID NO: 6), the promoter is minBglobin, the heterologous encoding sequence encodes a SYFP2 fluorescent protein, WPRE3 and BGHpA 3' elements, and the expression construct is comprises in a rAAV viral vector It is believed that claims 1-4, 12-15, 29-31, 34-37, 45-46 and 70 encompass this first named invention, and thus these claims will be searched without fee to the extent that they encompass said artificial expression construct. Additional artificial expression constructs will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected Additional artificial expression constructs. Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be an artificial expression construct wherein the enhancer is eHGT_476m (SEQ ID NO: 7), the promoter is minBglobin, the heterologous encoding sequence encodes a functional ion channel, WPRE3 and BGHpA 3' elements, and a vector, and the expression construct is comprised in a transgenic chandelier cell (claims 1, 5, 12-14, 16, 29-30, 34 and 38-39).

Group II: Claims 47-69, drawn to a method for expressing a heterologous gene within a targeted population of cells in vivo or in vitro by administrating a composition comprising an expression construct, the construct comprising an enhancer, a promoter and a heterologous encoding sequence, with a sufficient dosage and for a sufficient time to a sample or subject to express the gene within the targeted population of cells.

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

Special Technical Features

No technical features are shared between the nucleic acid sequences of the artificial expression construct of Group I+ and, accordingly, Group I+ lacks unity a priori.

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

Group I+ requires an isolated artificial expression construct composition comprised in a vector, not required by Group II.

Group II requires a method for expressing a heterologous gene within a targeted population of cells in vivo or in vitro, not required by Group I+.

Common Technical Features

The feature shared by the inventions of Group I+ and group II is an artificial expression construct comprising an enhancer, a promoter, a heterologous gene, and WPRE3 and BGHpA 3' elements in a vector, optionally an rAAV vector, the expression construct optionally comprised in a cell tissue, or animal.

However, these shared technical features do not represent a contribution over prior art, because the shared technical features are taught by WO 2020/168279 A2 to Allen Institute (hereinafter 'Allen'). Allen teaches said artificial expression construct comprising an enhancer, a promoter, a heterologous gene, and WPRE3 and BGHpA 3' elements in a vector, optionally an rAAV vector (para "[0032] The current disclosure provides artificial expression constructs that selectively drive gene expression", claim 1"An artificial expression construct comprising (i) an enhancer selected from eHGT_140h, ... and eHGT_359; (ii) a promoter; and (iii) a heterologous encoding sequence.", para [0063] Promoters can include general promoters, tissue-specific promoters, cell-specific promoters ... Particular examples of promoters include minBglobin, CMV, minCMV", para [0080] "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 ... vectors include a posttranscriptional regulatory element such as a WPRE", para [0081] "Elements directing the efficient termination and polyadenylation of a heterologous nucleic acid transcript can increase heterologous gene expression ... Particular embodiments may utilize BGHpA", para [0072] 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 selected cell populations. ... rAAV refers to a recombinant adeno-associated virus.") the expression construct optionally comprised in a cell tissue, or animal. (para [0005] "The current disclosure provides artificial expression constructs that selectively drive gene expression in targeted central nervous system cell populations.")

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

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