



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

(13) **A1**

(86) **Date de dépôt PCT/PCT Filing Date:** 2022/05/20
(87) **Date publication PCT/PCT Publication Date:** 2022/11/24
(85) **Entrée phase nationale/National Entry:** 2023/11/15
(86) **N° demande PCT/PCT Application No.:** US 2022/030371
(87) **N° publication PCT/PCT Publication No.:** 2022/246276
(30) **Priorité/Priority:** 2021/05/21 (US63/191,832)

(51) **Cl.Int./Int.Cl. C12N 15/79** (2006.01),
C12N 15/86 (2006.01)
(71) **Demandeur/Applicant:**
ALLEN INSTITUTE, US
(72) **Inventeurs/Inventors:**
DAIGLE, TANYA, US;
LEIN, EDWARD SEBASTIAN, US;
LEVI, BOAZ P., US;
MICH, JOHN K., US;
MITCHELL, AMANDA, US;
TASIC, BOSILJKA, US;
TING, JONATHAN, US;
...
(74) **Agent:** GOWLING WLG (CANADA) LLP

(54) **Titre : CONSTRUCTIONS D'EXPRESSION ARTIFICIELLES POUR LA MODULATION DE L'EXPRESSION GENIQUE DANS DES NEURONES A L'INTERIEUR DU THALAMUS**
(54) **Title: ARTIFICIAL EXPRESSION CONSTRUCTS FOR MODULATING GENE EXPRESSION IN NEURONS WITHIN THE THALAMUS**

FIG. 3A



(57) **Abrégé/Abstract:**

Artificial expression constructs for modulating gene expression in targeted central nervous system cell types are described. The artificial expression constructs can be used to express synthetic genes or modify gene expression in GABAergic or glutamatergic neurons within the thalamus. In some cases, the artificial constructs can also be used to express synthetic genes or modify gene expression in neurons within the thalamus as well as a secondary cell type.

(72) **Inventeurs(suite)/Inventors(continued):** WEED, NATALIE, US; ZENG, HONGKUI, US

Date Submitted: 2023/11/15

CA App. No.: 3219142

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 GABAergic or glutamatergic neurons within the thalamus. In some cases, the artificial constructs can also be used to express synthetic genes or modify gene expression in neurons within the thalamus as well as a secondary cell type.

ARTIFICIAL EXPRESSION CONSTRUCTS FOR
MODULATING GENE EXPRESSION IN NEURONS WITHIN THE THALAMUS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 63/191,832 filed May 21, 2021, which is incorporated herein by reference in its entirety as if fully set forth herein.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

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

FIELD OF THE DISCLOSURE

[0003] 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 the thalamus

REFERENCE TO SEQUENCE LISTING

[0004] 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 A166-0028PCT_ST25.txt. The text file is 178 KB, was created on May 20, 2022, and is being submitted electronically via EFS-Web.

BACKGROUND OF THE DISCLOSURE

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

SUMMARY OF THE DISCLOSURE

[0006] The current disclosure provides artificial expression constructs that drive gene expression in targeted central nervous system cell populations. Targeted central nervous system cell populations include: neurons in the thalamus, including: GABAergic neurons (Gata/Dlx5-6) within thalamus, GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus, cells

of the thalamic reticular nucleus, glutamatergic neurons within the thalamus, glutamatergic neurons (Prkcd-Grin2c (Core, LGN)) within the thalamus, glutamatergic neurons (Rxfp1-Epb4 (Matrix)) within the thalamus, and glutamatergic neurons within the parafascicular (Pf) nuclei of the thalamus. In particular embodiments, in addition to driving gene expression in the thalamus, artificial expression constructs described herein drive gene expression in a secondary cell type. Secondary cell types include: striatal medium spiny neuron-pan, Purkinje cells in the cerebellum, Deep cerebellar nuclei (DCN) cells in the cerebellum, molecular layer interneurons (MLI) cells in the cerebellum, Pvalb neuron cell types, chandelier cells, glutamatergic L5 ET cells in the neocortex, and Vip neurons in the neocortex.

[0007] Particular embodiments of the artificial expression constructs utilize the following enhancers to drive gene expression within targeted central nervous system cell populations as follows (enhancers / targeted cell population):

eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, and eHGT_976h/ glutamatergic neurons within the thalamus;

MGT_E117 and MGT_E118 / glutamatergic neurons (Prkcd-Grin2c (Core, LGN)) within the thalamus;

MGT_E119 and MGT_E120 / GABAergic neurons (Gata/Dlx5-6) within thalamus;

MGT_E121 / glutamatergic neurons (Rxfp1-Epb4 (Matrix)) within the thalamus;

3xCore2-eHGT_367h / glutamatergic neurons within the parafascicular (Pf) nuclei of the thalamus and striatal medium spiny neuron-pan;

eHGT_359h / glutamatergic neurons within the thalamus, Purkinje cells in the cerebellum, and Pvalb neurons in the neocortex;

eHGT_479m / glutamatergic neurons within the thalamus, Purkinje cells in the cerebellum, and chandelier cells in the neocortex;

eHGT_453m / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus, Deep cerebellar nuclei (DCN) cells in the cerebellum, and glutamatergic L5 ET cells in the neocortex;

eHGT_140h / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus and Pvalb neuron cell types;

eHGT_356h / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus, DCN cells in the cerebellum, Vip neurons in the neocortex, and cells of the thalamic reticular

nucleus;

eHGT_128h / glutamatergic neurons within the thalamus and Pvalb neuron cell types;

eHGT_369h / glutamatergic neurons within the thalamus, molecular layer interneurons (MLI) cells in the cerebellum, and Pvalb neurons in the neocortex; and

eHGT_710m / glutamatergic neurons within the thalamus and MLI cells in the cerebellum, chandelier cells, and molecular layer GABAergic interneurons in the cerebellum.

[0008] Particular embodiments utilize multiple concatenated copies of an enhancer or concatenated copies of an enhancer core. Examples include a core or concatenated core of eHGT_367h, eHGT_121h, eHGT_577h, and/or eHGT_606h. These artificial enhancer elements can provide higher levels and more rapid onset of transgene expression compared to a single full length original (native) enhancer.

[0009] In particular embodiments, the enhancer core includes the sequence as set forth in any one of SEQ ID NO: 18, SEQ ID NO: 26, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 41. In particular embodiments, these cores are concatenated and have 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core sequence. In particular embodiments, a three-copy concatemer of the selected enhancer cores include the sequence as set forth in any one of SEQ ID NO: 17, SEQ ID NO: 25, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, and SEQ ID NO: 40.

[0010] Particular embodiments of the enhancer cores utilize Core2-eHGT_367h, coreB_eHGT121h, core2_eHGT_577h, core3_eHGT_577h, core2_eHGT_606h, core3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, and core_eHGT_121h. Particular embodiments of the concatenated enhancer cores utilize 3xCore2-eHGT_367h, eHGT_369h (3xcoreB_eHGT121h), 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, and 3xcore_eHGT_121h. Within the disclosure, eHGT_369h can be used interchangeably with (3xCoreB)eHGT_121h.

[0011] Particular embodiments provide artificial expression constructs including the features of vectors described herein including vectors: CN2415, CN2416, CN2417, CN2418, CN2436, CN3000, CN3001, CN3003, CN3004, CN3005, CN3007, CN3009, AiP1335, AiP1336, AiP1337, AiP1338, AiP1339, CN2555, CN2045, CN2258, CN2251, CN1633, CN2043, CN1621, CN2216, CN2717, CN3639, CN3050, CN3051, CN3056, CN3057, CN4001, CN4003, CN2786, CN2840, CN3460, and CN2650.

BRIEF DESCRIPTION OF THE FIGURES

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

[0013] FIGs. 1A – 1C. Thalamus enhancer candidate selection overview. (1A) Example marker genes for thalamus neuron classes using the Allen Brain Atlas *in situ* hybridization (ISH) data. Both *Rgs16* and *Plekhhg1* show thalamus-specific labeling in the brain. Also shown for comparison are *Gad1* and *Slc17a7*, marker genes for inhibitory and excitatory neuron classes, respectively. (1B) DropVis browser (dropviz.org) single-cell RNA sequencing data from the mouse showing marker gene expression level dot plot summary for excitatory (exc) versus inhibitory (inh) neuron clusters in the thalamus. Darker circles indicate higher gene expression levels. Note that *Rgs16* and *Plekhhg1* are specifically expressed in excitatory neurons but not inhibitory neurons. (1C) Example of candidate peak selection using the Brain Open Chromatin Atlas dataset (bendlj01.u.hpc.mssm.edu/multireg/) as viewed in the UCSC Genome Browser. The open chromatin peak shown (grey highlight) is differentially accessible in mediodorsal thalamus versus numerous regions of neocortex (CTX), amygdala (AMY), hippocampus (HIP), nucleus accumbens (NAC), and putamen (PUT). The peak is located proximal to *Rgs16* locus and was therefore predicted to be an enhancer for thalamus excitatory neuron subclasses.

[0014] FIGs. 2A, 2B. (2A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $1.0E^{12}$ viral genome copies of AAV vector #CN2415 (eHGT_576h). Scale bar: 1 mm. (2B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0015] FIGs. 3A, 3B. (3A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $1.0E^{12}$ viral genome copies of AAV vector #CN2416 (eHGT_577h). Scale bar: 1 mm. (3B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0016] FIGs. 4A, 4B. (4A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $1.0E^{12}$ viral genome copies of AAV vector #CN2417 (eHGT_578h). Scale bar: 1 mm. (4B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0017] FIGs. 5A, 5B. (5A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the brain at 2 months after retro-orbital delivery of approximately $1.0E^{12}$ viral genome copies of AAV vector #CN2216 (3x(coreB)eHGT121h). Mice were transduced via intravenous delivery of AAVs packaged with PHP.eB capsid. Scale bar: 1 mm. (5B) Higher

magnification view of the thalamus region. Scale bar: 500 microns.

[0018] FIGs. 6A, 6B. (6A) Fluorescent reporter expression in mouse brain tissue following retro-orbital injection of CN2555 serotype PHP.eB. Scale bar: 1 mm. (6B) Higher magnification view of SYFP2 expression in the boxed dorsal striatum region from FIG. 6A. Scale bar: 500 microns.

[0019] FIGs. 7A-7G. A C57Bl/6J (wild-type) mouse was injected with $1.0E+12$ viral genomes of CN2555 serotype PHP.eB virus via the retro-orbital sinus. Representative two-photon tomography (TissueCyte) images of the native SYFP2 fluorescence in coronal brain sections are shown four weeks post-injection. FIGs. 7A, 7C, 7E show the full coronal section at different planes in the brain (moving anterior to posterior) and (7B, 7D, 7F) show higher magnification view of the boxed areas in the corresponding FIGs. 7A, 7C, 7E, respectively. (7G) TissueCyte imaging datasets were processed using an established informatics pipeline (Oh et al., Nature 2014, 508: 207-214) and registered to the CCFv3.0 (Wang et al., Cell 2020, 181(4):936-953.e20). Segmented pixel counts or voxels for each brain region were used for analysis and density dot plots of these data in all cortical (left = left hemisphere and right = right hemisphere) and subcortical structures are shown. Dark black, large circles = brain regions with most SYFP2 signal and small, light circles = brain regions with little to no SYFP2 signal. See FIG. 16 for full list of abbreviated brain structures.

[0020] FIGs. 8A-8I. Vector: CN2045 and Enhancer: eHGT_359h. (8A-8E) Animal: mouse 200910-08. (8A) Native SYFP2 fluorescence montage of a sagittal section of a whole mouse brain, (8B) cerebellum, and (8C) higher magnified cerebellum showing selective expression of SYFP2 in cells with Purkinje cell morphology. (8D) Native SYFP2 fluorescence of cerebellum in sagittal section showing labeling of Purkinje cells, and (8E) their overlap (arrows) with Pvalb mRNA expression. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN2045 virus packaged with the PHP.eB capsid. (8F, 8G) Animal: rat 585761. (8F) Native SYFP2 fluorescence montage of a sagittal section of a whole rat brain and the (8G) expanded view of the cerebellum showing selective expression of SYFP2 in cells with Purkinje cell morphology. Virus was administered in one day postnatal rat pups by intracerebroventricular (ICV) injection of CN2045 virus packaged with the PHP.eB capsid. (8H, 8I) Animal: monkey Q21.26.022. (8H) Native SYFP2 fluorescence montage of a sagittal section of a macaque cerebellum and the (8I) expanded view of the cerebellum showing selective expression of SYFP2 in cells with Purkinje cell morphology. Virus was administered by intraparenchymal injection of CN2045 virus packaged with the PHP.eB capsid.

[0021] FIGs. 9A-9D. (9A) Mouse genome coordinates for enhancer eHGT_453m (shaded region), with corresponding single cell ATAC-seq peaks by cell class and subclass. Arrow denotes

open chromatin peaks for the L5 ET neuron subclass. (9B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of 7.00E¹¹GC of AAV vector #CN2251. Scale bar: 100 microns. (9C) Mapping of single cell transcriptomic profiles of SYFP2+ cells sorted from the visual cortex region of the mouse brain after retro-orbital injection of CN2251 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_453m enhancer driven reporter expression occurs selectively in all 4 types of L5 ET neurons of the visual cortex region. The text along the bottom from left to right reads: 169 L2/3 IT VISp Rrad, 168 L2/3 IT VISp Adamts2, 167 L2/3 IT VISp Agmay, 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 Fezf1, 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, L5 CT Nxph2 Sla, 130 L5 CT VISp Krt80 Sla, L5 CT VISp Nxph2.Wls, 127 L5 CT VISp Ctxn3 Brinp3, 126 L5 CT VISp Ctxn3 Sla, 122 L5 CT VISp Gpr139, 120 L6b Col8a1 Rprm, 119 L6b VISp Mup5, 118 L6b VISp Col8a1 Rxfp1, 115 L6b P2ry12, L6b VISp Crh, 110 Lamp5 Krt73, Lamp5 Fam19a1 Pax6, 108 Lamp5 Fam19a1 Tmem182, 106 Lamp5 Ntn1 Npy2r, 105 Lamp5 Plch2 Dock5, 101 Lamp5 Lsp1, 100 Lamp5 Lhx6, Sncg Slc17a8, 96 Sncg Vip Nptx2, 95 Sncg Gpr50, 93 Sncg Vip Itih5, 90 Serpinf1 Clrn1, 89 Serpinf1 Aqp5 Vip, 85 Vip Igfbp6 Car10, 84 Vip Igfbp6 Pltp, Vip Lmo1 Fam159b, 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 Lect1 Oxtr, 70 Vip Rspo1 Itga4, 67 Vip Chat Htr1f, 66 Vip Pygm C1ql1, 61 Vip Crisp1d2 Htr2c, 60 Vip Crisp1d2 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 Chrna1 Glra3, 42 Sst Myh8 Fibin, 40 Sst Chrna2 Ptgdr, 39 Sst Tac2 Myn4, 37 Sst Hpse Sema3c, 36 Sst Hpse Cbln4, 34 Sst Crh2 Efemp1, 33 Sst Crh2 4930553C11Rik, 31 Sst Esm1, 29 Sst Tac2 Tacstd2, 28 Sst Rxfp1 Eya1, 27 Sst Rxfp1 Prdm8, 23 Sst Nts, 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 Tpbpg, 4 Pvalb Vlpr2, Meis2 Adamts19, 170 Astro Aqp4, 171 OPC Pdgfra Grm5, Oligo Serpinb1a, 174 Oligo Synpr, VLMC Osr1 Cd74, VLMC Osr1 Mc5r, VLMC Spp1 Col15a1, Peri Kcnj8, SMC Acta2, Endo Ctla2a, and 181 Microglia Siglech. (9D) Epifluorescence microscopy image of native SYFP2 fluorescence in a fixed brain slice from macaque primary motor cortex at 64 days after stereotaxic injection of

enhancer AAV vector CN2251 serotype PHP.eB. The genome coordinates are for build Hg38 and the views shown are from the UCSC Genome Browser. Scale bar: 500 μ m.

[0022] FIGs. 10A, 10B. Vector: CN2251, Enhancer: eHGT_453m, and Animal: mouse. (10A) Native SYFP2 fluorescence montage of a coronal section of mouse cerebellum and (10B), expanded view of the cerebellum showing selective expression of SYFP2 in cells in the deep cerebellar nuclei. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN2251 virus packaged with the PHP.eB capsid.

[0023] FIGs. 11A-11C. (11A) Fluorescence expression of CN1633 (eHGT_140h), in black, shown in whole mouse brain in sagittal section. (11B) High resolution images showing overlap of CN1633 SYFP2 fluorescence with GABAergic markers Gad1 and Pvalb mRNA expression. The arrows identify SYFP2-labeled cells. (11C) Single cell transcriptomic characterization of SYFP2 fluorescent cells isolated from mouse V1. After single cell gene expression analysis, cells were mapped to an existing taxonomy of mouse V1 cell types. Note that nearly all cells are types of Pvalb neurons.

[0024] FIGs. 12A-12C. (12A) Fluorescence expression of CN1621 (eHGT_128h), in black, shown in whole mouse brain in sagittal section. (12B) High resolution images showing overlap of CN1621 SYFP2 fluorescence with GABAergic marker Pvalb mRNA expression. The arrows identify SYFP2-labeled cells. (12C) Single cell transcriptomic characterization of SYFP2 fluorescent cells isolated from mouse V1. After single cell gene expression analysis, cells were mapped to an existing taxonomy of mouse V1 cell types. Note that nearly all cells are types of Pvalb neurons.

[0025] FIGs. 13A, 13B. CN2717 (eHGT_710m) in mouse neocortex. (13A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the neocortex 40 days after retro-orbital delivery of 6.0×10^{11} viral genome copies of AAV vector #CN2717 (eHGT_710m). Scale bar: 200 microns. (13B) Higher magnification view showing a sparse cell bodies and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0026] FIGs. 14A, 14B. CN2717 (eHGT_710m) in macaque frontal cortex. (14A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the superior frontal cortex region 43 days after stereotaxic injection of 3.46×10^{11} viral genome copies of AAV vector #CN2717 (eHGT_710m) in adult macaque in vivo. Scale bar: 200 microns. (14B) Higher magnification view showing a cell body and signature chandelier cell axon cartridges. Scale bar: 50 microns.

[0027] FIGs. 15A, 15B. Vector: CN2717, Enhancer: eHGT_710m, and Animal: mouse C57BL6J-560070. (15A) Native SYFP2 fluorescence montage of a sagittal section of mouse cerebellum and (15B) expanded view of the cerebellum showing selective expression of SYFP2 in cells with

small interneuron morphology in the molecular layer. Virus was administered in adult mice by intravascular (IV) injection (retro-orbital) of CN2717 virus packaged with the PHP.eB capsid.

[0028] FIGs. 16A, 16B: (16A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the brain at 2 months after retro-orbital delivery of approximately $1.0E^{12}$ viral genome copies of AAV vector #CN2437 (eHGT_607h). Mice were transduced via intravenous delivery of AAVs packaged with PHP.eB capsid. Scale bar: 1 mm. (16B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0029] FIGs. 17A, 17B: (17A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the brain at 2 months after retro-orbital delivery of approximately $1.0E^{12}$ viral genome copies of AAV vector #CN3000 (eHGT_827h). Mice were transduced via intravenous delivery of AAVs packaged with PHP.eB capsid. Scale bar: 1 mm. (17B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0030] FIGs. 18A, 18B: (18A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the brain at 2 months after retro-orbital delivery of approximately $1.0E^{12}$ viral genome copies of AAV vector #CN3003 (eHGT_830h). Mice were transduced via intravenous delivery of AAVs packaged with PHP.eB capsid. Scale bar: 1 mm. (18B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0031] FIGs. 19A, 19B: (19A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the brain at 2 months after retro-orbital delivery of approximately $1.0E^{12}$ viral genome copies of AAV vector #CN3005 (eHGT_832h). Mice were transduced via intravenous delivery of AAVs packaged with PHP.eB capsid. Scale bar: 1 mm. (19B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0032] FIGs. 20A, 20B: (20A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the brain at 2 months after retro-orbital delivery of approximately $1.0E^{12}$ viral genome copies of AAV vector #CN3007 (eHGT_834h). Mice were transduced via intravenous delivery of AAVs packaged with PHP.eB capsid. Scale bar: 1 mm. (20B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0033] FIGs. 21A, 21B: (21A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the brain at 2 months after retro-orbital delivery of approximately $1.0E^{12}$ viral genome copies of AAV vector #CN3009 (eHGT_836h). Mice were transduced via intravenous delivery of AAVs packaged with PHP.eB capsid. Scale bar: 1 mm. (21B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0034] FIGs. 22A, 22B: (22A) Epifluorescence micrograph image (inverted) showing native SYFP2 expression in the brain at 2 months after retro-orbital delivery of approximately $1.0E^{12}$ viral

genome copies of AAV vector #CN2786 (3xcore_eHGT_121h). Mice were transduced via intravenous delivery of AAVs packaged with PHP.eB capsid. Scale bar: 1 mm. (22B) Higher magnification view of the thalamus region. Scale bar: 500 microns.

[0035] FIG. 23. Table of abbreviations of brain structures.

[0036] FIG. 24. Sequences supporting the disclosure as follows: eHGT_576h (SEQ ID NO: 1), eHGT_577h (SEQ ID NO: 2), eHGT_578h (SEQ ID NO: 3), eHGT_579h (SEQ ID NO: 4), eHGT_606h (SEQ ID NO: 5), eHGT_827h (SEQ ID NO: 137), eHGT_828h (SEQ ID NO: 138), eHGT_830h (SEQ ID NO: 6), eHGT_831h (SEQ ID NO: 7), eHGT_832h (SEQ ID NO: 8), eHGT_834h (SEQ ID NO: 9), eHGT_836h (SEQ ID NO: 10), MGT_E117 (SEQ ID NO: 11), MGT_E118 (SEQ ID NO: 12), MGT_E119 (SEQ ID NO: 13), MGT_E120 (SEQ ID NO: 14), MGT_E121 (SEQ ID NO: 15), eHGT_717h (SEQ ID NO: 16), 3xCore2-eHGT_367h (SEQ ID NO: 17), Core of eHGT_367h (SEQ ID NO: 18), eHGT_359h (SEQ ID NO: 19), eHGT_479m (SEQ ID NO: 20), eHGT_453m (SEQ ID NO: 21), eHGT_140h (SEQ ID NO: 22), eHGT_356h (SEQ ID NO: 23), eHGT_128h (SEQ ID NO: 24), eHGT_369h (also referred to as 3xcoreB_eHGT121h) (SEQ ID NO: 25), coreB_eHGT121h (SEQ ID NO: 26), 300 bp core of eHGT_369h (2xcoreB_eHGT121h) (SEQ ID NO: 27), eHGT_710m (SEQ ID NO: 28), eHGT_895h (SEQ ID NO: 29), 3xcore2_eHGT_577h (SEQ ID NO: 30), core2_eHGT_577h (SEQ ID NO: 31), 3xcore3_eHGT_577h (SEQ ID NO: 32), core3_eHGT_577h (SEQ ID NO: 33), 3xcore2_eHGT_606h (SEQ ID NO: 34), core2_eHGT_606h (SEQ ID NO: 35), 3xcore3_eHGT_606h (SEQ ID NO: 36), core3_eHGT_606h (SEQ ID NO: 37), core4_eHGT_577h (SEQ ID NO: 38), core6_eHGT_606h (SEQ ID NO: 39), 3xcore_eHGT_121h (SEQ ID NO: 40), core_eHGT_121h (SEQ ID NO: 41), eHGT_590m (SEQ ID NO: 42), eHGT_976h (SEQ ID NO: 43), Beta-Globin Minimal Promoter (pBGmin/minBGlobin/minBGprom) (SEQ ID NO: 45), minCMV Promoter (SEQ ID NO: 46), Mutated minCMV Promoter (Sacl RE site removed) (SEQ ID NO: 47), minRho Promoter (SEQ ID NO: 48), minRho* Promoter (SEQ ID NO: 49), Hsp68 minimal Promoter (proHsp68) (SEQ ID NO: 50), SYFP2 (SEQ ID NO: 51), EGFP (SEQ ID NO: 52), Optimized Flp recombinase (FlpO) (SEQ ID NO: 53), Improved Cre recombinase (iCre) (SEQ ID NO: 54), SP10 insulator (SP10ins) (SEQ ID NO: 55), 3xSP10ins (SEQ ID NO: 56), WPRE3 (SEQ ID NO: 57), WPRE (SEQ ID NO: 58), BGHpA (SEQ ID NO: 59), HGHPA (SEQ ID NO: 60), 3XFLAG (SEQ ID NO: 61), hsA2 (SEQ ID NO: 62), 10a.a. (SEQ ID NO: 63), H2B (SEQ ID NO: 64), P2A (SEQ ID NOs: 65 or 66), T2A (SEQ ID NO: 67), E2A (SEQ ID NO: 68), F2A (SEQ ID NO: 69), Exemplary Plasmid Backbone 1 – Left ITR (SEQ ID NO: 70), Exemplary Plasmid Backbone 1 – Right ITR (SEQ ID NO: 71), Exemplary Plasmid Backbone 2 – Left ITR (SEQ ID NO: 72), Exemplary Plasmid Backbone 2 – Right ITR (SEQ ID NO: 73), PHP.eB

capsid (SEQ ID NO: 74), AAV9 VP1 capsid protein (SEQ ID NO: 75), tet-Transactivator version 2 (tTA2) (SEQ ID NO: 76), GTPase HRas [Homo sapiens] (SEQ ID NO: 77), Substance P is position 58-68 of Protachykinin-1 [Homo sapiens] (SEQ ID NO: 78), Oxytocin is position 20-28 of Oxytocin-neurophysin 1 [Homo sapiens] (SEQ ID NO: 79), GCaMP6m (SEQ ID NO: 80), GCaMP6s (SEQ ID NO: 81), GCaMP6f (SEQ ID NO: 82), CN2415 (SEQ ID NO: 83), CN2416 (SEQ ID NO: 84), CN2417 (SEQ ID NO: 85), CN2418 (SEQ ID NO: 86), CN2436 (SEQ ID NO: 87), CN3000 (SEQ ID NO: 139), CN3001 (SEQ ID NO: 140), CN3003 (SEQ ID NO: 88), CN3004 (SEQ ID NO: 89), CN3005 (SEQ ID NO: 90), CN3007 (SEQ ID NO: 91), CN3009 (SEQ ID NO: 92), AiP1335 (SEQ ID NO: 93), AiP1336 (SEQ ID NO: 94), AiP1337 (SEQ ID NO: 95), AiP1338 (SEQ ID NO: 96), AiP1339 (SEQ ID NO: 97), CN2555 (SEQ ID NO: 99), CN2045 (SEQ ID NO: 100), CN2258 (SEQ ID NO: 101), CN2251 (SEQ ID NO: 102), CN1633 (SEQ ID NO: 103), CN2043 (SEQ ID NO: 104), CN1621 (SEQ ID NO: 105), CN2216 (SEQ ID NO: 106), CN2717 (SEQ ID NO: 107), CN3639 (SEQ ID NO: 108), CN3050 (SEQ ID NO: 109), CN3051 (SEQ ID NO: 110), CN3056 (SEQ ID NO: 111), CN3057 (SEQ ID NO: 112), CN4001 (SEQ ID NO: 113), CN4003 (SEQ ID NO: 114), CN2786 (SEQ ID NO: 115), CN2840 (SEQ ID NO: 116), CN3460 (SEQ ID NO: 117), and CN2650 (SEQ ID NO: 118).

DETAILED DESCRIPTION

[0037] To fully understand the biology of the brain, different cell types need to be distinguished and defined and, to further study them, artificial expression constructs that can label and perturb them need to be identified (Tasic, *Curr. Opin. Neurobiol.* 50, 242–249 (2018); Zeng & Sanes, *Nat. Rev. Neurosci.* 18, 530–546 (2017)). In mouse, recombinase driver lines have been used to great effect to label cell populations that share marker gene expression (Daigle et al., *Cell* 174, 465–480.e22 (2018); Taniguchi, et al., *Neuron* 71, 995–1013 (2011); Gong et al., *J. Neurosci.* 27, 9817–9823 (2007)). However, the creation, maintenance, and use of such lines that label cell types with high specificity can be costly, frequently requiring triple transgenic crosses, which yield a low frequency of experimental animals. Furthermore, those tools require germline transgenic animals and thus are not applicable to humans.

[0038] The current disclosure provides artificial expression constructs that drive gene expression in targeted central nervous system cell populations. Targeted central nervous system cell populations include: GABAergic neurons (Gata/Dlx5-6) within thalamus, GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus, cells of the thalamic reticular nucleus, glutamatergic neurons within the thalamus, glutamatergic neurons (Prkcd-Grin2c (Core, LGN)) within the thalamus, glutamatergic neurons (Rxfp1-Epb4 (Matrix)) within the thalamus, and glutamatergic neurons within the parafascicular (Pf) nuclei of the thalamus. In particular

embodiments, artificial expression constructs described herein drive gene expression in a secondary targeted cell type. Exemplary secondary cell types include: striatal medium spiny neuron-pan, Purkinje cells in the cerebellum, Deep cerebellar nuclei (DCN) cells in the cerebellum, molecular layer interneurons (MLI) cells in the cerebellum, Pvalb neuron cell types, chandelier cells, glutamatergic L5 ET cells in the neocortex, and Vip neurons in the neocortex.

[0039] Particular embodiments of the artificial expression constructs utilize the following enhancers to drive gene expression within targeted central nervous system cell populations as follows (enhancer / targeted cell population): eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, and eHGT_976h/ glutamatergic neurons within the thalamus;

MGT_E117 and MGT_E118 / glutamatergic neurons (Prkcd-Grin2c (Core, LGN)) within the thalamus;

MGT_E119 and MGT_E120 / GABAergic neurons (Gata/Dlx5-6) within thalamus;

MGT_E121 / glutamatergic neurons (Rxfp1-Epb4 (Matrix)) within the thalamus;

3xCore2-eHGT_367h / glutamatergic neurons within the parafascicular (Pf) nuclei of the thalamus and striatal medium spiny neuron-pan;

eHGT_359h / glutamatergic neurons within the thalamus, Purkinje cells in the cerebellum, and Pvalb neurons in the neocortex;

eHGT_479m / glutamatergic neurons within the thalamus, Purkinje cells in the cerebellum, and chandelier cells in the neocortex;

eHGT_453m / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus, Deep cerebellar nuclei (DCN) cells in the cerebellum, and glutamatergic L5 ET cells in the neocortex;

eHGT_140h / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus and Pvalb neuron cell types;

eHGT_356h / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus, DCN cells in the cerebellum, Vip neurons in the neocortex, and cells of the thalamic reticular nucleus;

eHGT_128h / glutamatergic neurons within the thalamus and Pvalb neuron cell types;

eHGT_369h / glutamatergic neurons within the thalamus, molecular layer interneurons (MLI) cells in the cerebellum, and Pvalb neurons in the neocortex; and

eHGT_710m / glutamatergic neurons within the thalamus and MLI cells in the cerebellum, chandelier cells, and molecular layer GABAergic interneurons in the cerebellum.

[0040] Particular embodiments utilize multiple concatenated copies of an enhancer or concatenated copies of an enhancer core. Examples include a core or concatenated core of eHGT_367h, eHGT_121h, eHGT_577h, and/or eHGT_606h. These artificial enhancer elements can provide higher levels and more rapid onset of transgene expression compared to a single full length original (native) enhancer.

[0041] In particular embodiments, the enhancer core includes the sequence as set forth in any one of SEQ ID NO: 18, SEQ ID NO: 26, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 41. In particular embodiments, these cores are concatenated and have 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core sequence. In particular embodiments, a three-copy concatemer of the selected enhancer cores include the sequence as set forth in any one of SEQ ID NO: 17, SEQ ID NO: 25, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, and SEQ ID NO: 40.

[0042] Particular embodiments of the enhancer cores utilize Core2-eHGT_367h, coreB_eHGT121h, core2_eHGT_577h, core3_eHGT_577h, core2_eHGT_606h, core3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, and core_eHGT_121h. Particular embodiments of the concatenated enhancer cores utilize 3xCore2-eHGT_367h, eHGT_369h (3xcoreB_eHGT121h), 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core6_eHGT_606h, and 3xcore_eHGT_121h. Within the disclosure, eHGT_369h can be used interchangeably with 3xCoreB-eHGT_121h.

[0043] Particular embodiments provide artificial expression constructs including the features of vectors described herein including vectors: CN2415, CN2416, CN2417, CN2418, CN2436, CN3000, CN3001, CN3003, CN3004, CN3005, CN3007, CN3009, AiP1335, AiP1336, AiP1337, AiP1338, AiP1339, CN2555, CN2045, CN2258, CN2251, CN1633, CN2043, CN1621, CN2216, CN2717, CN3639, CN3050, CN3051, CN3056, CN3057, CN4001, CN4003, CN2786, CN2840, CN3460, and CN2650.

[0044] 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.

[0045] (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.

[0046] 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_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, and eHGT_710m.

[0047] In particular embodiments, a targeted central nervous system cell type enhancer is an enhancer that is uniquely or predominantly utilized by the targeted central nervous system cell type. A targeted central nervous system cell type enhancer enhances expression of a gene in the targeted central nervous system. In certain embodiments, a targeted central nervous system cell type enhancer is also a targeted central nervous system type enhancer that enhances expression of a gene in the targeted central nervous system and does not substantially direct expression of genes in other non-targeted cell types, thus having cell type specific transcriptional activity.

[0048] 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.

[0049] When a heterologous coding sequence is selectively expressed in selected cells, it leads to expression of the administered heterologous coding sequence in the intended cell type and is not substantially expressed in other cell types, as explained in additional detail below. In particular embodiments, not substantially expressed in other cell types is less than 50% expression in a reference cell type as compared to a targeted cell type; less than 40% expression in a reference cell type as compared to a targeted cell type; less than 30% expression in a reference cell type as compared to a targeted cell type; less than 20% expression in a reference cell type as

compared to a targeted cell type; or less than 10% expression in a reference cell type as compared to a targeted cell type. In particular embodiments, a reference cell type refers to non-targeted cells. The non-targeted cells can be within the same anatomical structure as the targeted cells and/or can project to a common anatomical area. In particular embodiments, a reference cell type is within an anatomical structure that is adjacent to an anatomical structure that includes the targeted cell type. In particular embodiments, a reference cell type is a non-targeted cell with a different gene expression profile than the targeted cells.

[0050] In particular embodiments, the product of the coding sequence may be expressed at low levels in non-selected cell types, for example at less than 1% or 1%, 2%, 3%, 5%, 10%, 15% or 20% of the levels at which the product is expressed in selected cells. In particular embodiments, the targeted central nervous system cell type is the only cell type that expresses the right combination of transcription factors that bind an enhancer disclosed herein to drive gene expression. Thus, in particular embodiments, expression occurs exclusively within the targeted cell type.

[0051] 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:

[0052] Thalamus GABAergic neuron classes and subclasses:

- All: Express GABA synthesis genes *Gad1/GAD1* and/or *Gad2/GAD2*.
- Thalamic reticular nucleus (TRN) neurons: Express GABA synthesis genes *Gad1/GAD1* and *Pvalb/PVALB*.

[0053] Thalamus glutamatergic neuron classes and subclasses:

- All glutamatergic neurons: Express glutamate transporters *Slc17a6/SLC17A6* and/or *Slc17a7/SLC17A7*. They lack expression of *Gad1/Gad2* and have expression of one or more of the marker genes *Synpo2/SYNPO2*, *Rgs16/RGS16*, *Plekhg1/PLEKHG1*, and *Prkcd/PRKCD*.
- Glutamatergic neurons within the parafascicular (Pf) nuclei: The Pf nuclei is a posterior component of the intralaminar nuclei of the thalamus. It plays a role in the feedback systems of basal ganglia–thalamo-cortical circuits critically involved in cognitive processes.

[0054] Neocortical GABAergic neuron Subclasses:

- All: Express GABA synthesis genes *Gad1/GAD1* and *Gad2/GAD2*.
- *Lamp5*, *Sncg*, *Serpinf1*, and *Vip* GABAergic neurons: Developmentally derived from

neuronal progenitors from the caudal ganglionic eminence (CGE) or preoptic area (POA).

- *Sst* and *Pvalb* GABAergic neurons: Developmentally derived from neuronal progenitors in the medial ganglionic eminence (MGE).
- *Lamp5* GABAergic neurons: Found in many neocortical layers, especially upper (L1-L2/3), and have mainly neurogliaform and single bouquet morphology.
- *Lamp5_Lhx6* GABAergic neurons: A subset of *Lamp5* GABAergic neurons that co-express *Lamp5* and *Lhx6*.
- *Sncg* GABAergic neurons: Found in many neocortical layers, and have molecular overlaps with *Lamp5* and *Vip* cells, but inconsistent expression of *Lamp5* or vasoactive intestinal peptide (*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 *Vip*.
- *Sst* GABAergic neurons: Found in many neocortical layers, but especially frequent in lower layers (L5-L6). They highly express the neurotransmitter somatostatin (*Sst*), and frequently block dendritic inputs to postsynaptic neurons. Included in this subclass are sleep-active *Sst Chodl* neurons (which also express *Nos1* and *Tacr1*) that are highly distinct from other *Sst* neurons but express some shared marker genes including *Sst*. In human, *SST* gene expression is often detected in layer 1 *LAMP5+* GABAergic neuron subtypes.
- *Pvalb* GABAergic neurons: Found in many neocortical layers, but especially frequent in lower layers (L5-L6). They highly express the calcium-binding protein parvalbumin (*Pvalb*), express neuropeptide *Tac1*, and frequently dampen the output of postsynaptic neurons. Most fast-spiking GABAergic neurons express *Pvalb* strongly. Included in this subclass are chandelier cells, which have distinct, chandelier-like morphology and express the markers *Cpne5* and *Vipr2* in mouse, and *NOG* and *UNC5B* in human.
- *Meis2*: A distinct subclass defined by a single type, only neocortical GABAergic neuron type that expresses *Meis2* gene, and does not express some other genes that are expressed by all other neocortical GABAergic neuron types (for example, *Thy1* and *Scn2b*). This type is found in L6b and subcortical white matter.

[0055] Neocortical glutamatergic neuron subclasses:

- All: Express glutamate transmitters *Slc17a6* and/or *Slc17a7*. They all express *Snap25* and

lack expression of *Gad1/Gad2*.

- L2/3 IT glutamatergic neurons: Primarily reside in Layer 2/3 and have mainly intratelencephalic (cortico-cortical) projections.
- L4 IT glutamatergic neurons: Primarily reside in Layer 4 and mainly have either local or intratelencephalic (cortico-cortical) projections.
- L5 IT glutamatergic neurons: Primarily reside in Layer 5 and have mainly intratelencephalic (cortico-cortical) projections. Also called L5a.
- L5 PT glutamatergic neurons: Primarily reside in Layer 5 and have mainly cortico-subcortical (pyramidal tract or corticofugal) projections. Also called L5b or L5 CF (corticofugal) or L5 ET (extratelencephalic). This subclass includes cells that are located in the primary motor cortex and neighboring areas and are corticospinal projection neurons, which are associated with motor neuron/movement disorders, such as ALS. This subclass includes thick-tufted pyramidal neurons, including distinctive subtypes found only in specialized regions, e.g., Betz cells, Meynert cells, and von Economo cells.
- L5 NP glutamatergic neurons: Primarily reside in Layer 5 and have mainly nearby projections.
- L6 CT glutamatergic neurons: Primarily reside in Layer 6 and have mainly cortico-thalamic projections.
- L6 IT glutamatergic neurons: Primarily reside in Layer 6 and have mainly intratelencephalic (cortico-cortical) projections.
- L6 IT Car3 glutamatergic neurons: Most densely present in claustrum and endopyriform nucleus, and sparsely throughout L6 in many cortical areas including the primary visual cortex. These cells have mainly intratelencephalic (cortico-cortical) projections. Additional marker genes for claustrum enriched neurons include *Gnb4* and *Ntn2*.
- 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*.
- 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.
- Deep cerebellar nucleus (DCN) neurons: neurons located in the deep cerebellar nuclei structures. These include glutamatergic and GABAergic cells that express the gene *Pvalb*.

- Molecular layer interneurons (MLI): neurons within the cerebellum that participate in spatially structured networks via chemical and electrical synapses.
- Chandelier cells are a specialized GABAergic interneuron that selectively innervates pyramidal neurons.
- Striatal medium spiny neurons are the major striatal neuron. They receive synaptic input from both glutamatergic and dopaminergic afferents.

[0056] Non-neuronal Subclasses:

- Astrocytes: Neuroectoderm-derived glial cells which express the marker Aqp4 and often GFAP, but do not express neuronal marker SNAP25. They can have a distinct star-shaped morphology and are involved in metabolic support of other cells in the brain. Multiple astrocyte morphologies are observed in mouse and human
- Oligodendrocytes: Neuroectoderm-derived glial cells, which express the marker Sox10. This category includes oligodendrocyte precursor cells (OPCs). Oligodendrocytes are the subclass that is primarily responsible for myelination of neurons.
- VLMCs: Vascular leptomeningeal cells (VLMCs) are part of the meninges that surround the outer layer of the cortex and express the marker genes Lum and Col1a1.
- Pericytes: Blood vessel-associated cells that express the marker genes Kcnj8 and Abcc9. Pericytes wrap around endothelial cells and are important for regulation of capillary blood flow and are involved in blood-brain barrier permeability.
- SMCs: Specialized smooth-muscle cells which are blood vessel-associated cells that express the marker gene Acta2. SMCs cover arterioles in the brain and are involved in blood-brain barrier permeability.
- Endothelial cells: Cells that line blood vessels of the brain. Endothelial cells express the markers Tek and PDGF-B.
- Microglia: hematopoietic-derived immune cells, which are brain-resident macrophages, and perivascular macrophages (PVMs) that may be transitionally associated with brain tissue or included as a byproduct of brain dissection methods. Microglia are known to express Cx3cr1, Tmem119, and PTPRC (CD45).

[0057] 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.

[0058] Exemplary reporter genes/proteins include those expressed by Addgene ID#s 83894 (pAAV-hDlx-Flex-dTomato-Fishell_7), 83895 (pAAV-hDlx-Flex-GFP-Fishell_6), 83896 (pAAV-

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

[0059] GFP is composed of 238 amino acids (26.9 kDa), originally isolated from the jellyfish *Aequorea victoria*/*Aequorea aequorea*/*Aequorea forskalea* that fluoresces green when exposed to blue light. The GFP from *A. victoria* has a major excitation peak at a wavelength of 395 nm and a minor one at 475 nm. Its emission peak is at 509 nm which is in the lower green portion of the visible spectrum. The GFP from the sea pansy (*Renilla reniformis*) has a single major excitation peak at 498 nm. Due to the potential for widespread usage and the evolving needs of researchers, many different mutants of GFP have been engineered. The first major improvement was a single point mutation (S65T) reported in 1995 in *Nature* by Roger Tsien. This mutation dramatically improved the spectral characteristics of GFP, resulting in increased fluorescence, photostability and a shift of the major excitation peak to 488 nm with the peak emission kept at 509 nm. The addition of the 37°C folding efficiency (F64L) point mutant to this scaffold yielded enhanced GFP (EGFP). EGFP has an extinction coefficient (denoted ϵ), also known as its optical cross section of $9.13 \times 10^4 \text{ m}^2/\text{mole}$, 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.

[0060] 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.

[0061] 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).

[0062] Ion transporters are transmembrane proteins that mediate transport of ions across cell membranes. These transporters are pervasive throughout most cell types and important for regulating cellular excitability and homeostasis. Ion transporters participate in numerous cellular processes such as action potentials, synaptic transmission, hormone secretion, and muscle contraction. Many important biological processes in living cells involve the translocation of cations, such as calcium (Ca^{2+}), potassium (K^{+}), and sodium (Na^{+}) ions, through such ion channels. In particular embodiments, ion transporters include voltage gated sodium channels (e.g., SCN1A), potassium channels (e.g., KCNQ2), and calcium channels (e.g. CACNA1C)).

[0063] Exemplary enzymes, transcription factors, receptors, membrane proteins, cellular trafficking proteins, signaling molecules, and neurotransmitters include enzymes such as lactase, lipase, helicase, alpha-glucosidase, and aromatic L-amino acid decarboxylase (AADC), 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), glial cell line-derived neurotrophic factor (GDNF), platelet-derived growth factor (PDGF), transforming growth factor β ($\text{TGF}\beta$), epidermal growth factor (EGF), GTPase and HRas; and neurotransmitters such as cocaine and amphetamine regulated transcript, substance P, oxytocin, and somatostatin.

[0064] In particular embodiments, functional molecules include reporters of cell function and states such as calcium reporters. Intracellular calcium concentration is an important predictor of numerous cellular activities, which include neuronal activation, muscle cell contraction and second messenger signaling. A sensitive and convenient technique to monitor the intracellular calcium levels is through the genetically encoded calcium indicator (GECI). Among the GECIs, green fluorescent protein (GFP) based calcium sensors named GCaMPs are efficient and widely used tools. The GCaMPs are formed by fusion of M13 and calmodulin protein to N- and C-termini of circularly permuted GFP. Some GCaMPs yield distinct fluorescence emission spectra (Zhao et al., Science, 2011, 333(6051): 1888-1891). Exemplary GECIs with green fluorescence include GCaMP3, GCaMP5G, GCaMP6s, GCaMP6m, GCaMP6f, jGCaMP7s, jGCaMP7c, jGCaMP7b, jGCaMP7f, jGCaMP8s, jGCaMP8m, and jGCaMP8f. Furthermore, GECIs with red fluorescence include jRGECO1a and jRGECO1b. AAV products containing GECIs are commercially available.

For example, Vigene Biosciences provides AAV products including AAV8-CAG-GCaMP3 (Cat. No:BS4-CX3AAV8), AAV8-Syn-FLEX-GCaMP6s-WPRE (Cat. No:BS1-NXSAAV8), AAV8-Syn-FLEX-GCaMP6s-WPRE (Cat. No:BS1-NXSAAV8), AAV9-CAG-FLEX-GCaMP6m-WPRE (Cat. No:BS2-CXMAAV9), AAV9-Syn-FLEX-jGCaMP7s-WPRE (Cat. No:BS12-NXSAAV9), AAV9-CAG-FLEX-jGCaMP7f-WPRE (Cat. No:BS12-CXFAAV9), AAV9-Syn-FLEX-jGCaMP7b-WPRE (Cat. No:BS12-NXBAAV9), AAV9-Syn-FLEX-jGCaMP7c-WPRE (Cat. No:BS12-NXCAAV9), AAV9-Syn-FLEX-NES-jRGECO1a-WPRE (Cat. No:BS8-NXAAAV9), and AAV8-Syn-FLEX-NES-jRCaMP1b-WPRE (Cat. No:BS7-NXBAAV8).

[0065] 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).

[0066] In particular embodiments, functional molecules include modulators of neuronal activity like channelrhodopsins (e.g., channelrhodopsin-1, channelrhodopsin-2, and variants thereof). Channelrhodopsins are a subfamily of retinylidene proteins (rhodopsins) that function as light-gated ion channels. In addition to channelrhodopsin 1 (ChR1) and channelrhodopsin 2 (ChR2), several variants of channelrhodopsins have been developed. For example, Lin et al. (*Biophys J*, 2009, 96(5): 1803-14) describe making chimeras of the transmembrane domains of ChR1 and ChR2, combined with site-directed mutagenesis. Zhang et al. (*Nat Neurosci*, 2008, 11(6): 631-3) describe VChR1, which is a red-shifted channelrhodopsin variant. VChR1 has lower light sensitivity and poor membrane trafficking and expression. Other known channelrhodopsin variants include the ChR2 variant described in Nagel, et al., *Proc Natl Acad Sci USA*, 2003, 100(24): 13940-5), ChR2/H134R (Nagel, G., et al., *Curr Biol*, 2005, 15(24): 2279-84), and ChD/ChEF/ChIEF (Lin, J. Y., et al., *Biophys J*, 2009, 96(5): 1803-14), which are activated by blue light (470 nm) but show no sensitivity to orange/red light. Additional variants are described in Lin, *Experimental Physiology*, 2010, 96.1: 19-25; Knopfel et al., *The Journal of Neuroscience*, 2010, 30(45): 14998-15004; and Mardinly et al., *Nat Neurosci*. 2018, 21(6):881-893).

[0067] 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.

[0068] In particular embodiments, functional molecules include a localizing cassette. In particular embodiments, localizing cassettes are used to localize a molecule (e.g., a vector, a protein, a sensor) to a specific subcellular compartment such as the soma, axon, or dendrite(s) of a neuron. In particular embodiments, localizing cassettes include a soma tag (e.g., soma (EE-RR)) to localize at the soma; an axon tag (e.g., derived from GAP43) or synaptophysin (sy) to localize at the axon; hydrophobic tails to localize at the plasma membrane; and hydrophobicity or alkyl chain to localize at the endoplasmic reticulum. In particular embodiments, localizing cassettes are fused to a sensor molecule such as a GECl. In particular embodiments, fusion proteins of a GECl and a localizing cassette includes soma-jGCaMP8s, axon-jRGECO1a, syGCaMP5G, and soma-jGCaMP7s.

[0069] In particular embodiments, functional molecules include a tag cassette. A tag cassette includes His tag (HHHHHH; SEQ ID NO: 125), Flag tag (DYKDDDDK; SEQ ID NO: 126), Xpress tag (DLYDDDDK; SEQ ID NO: 127), Avi tag (GLNDIFEAQKIEWHE; SEQ ID NO: 128), Calmodulin tag (KRRWKKNFIAVSAANRFKKISSSGAL; SEQ ID NO: 129), Polyglutamate tag, HA tag (YPYDVPDYA; SEQ ID NO: 130), Myc tag (EQKLISEEDL; SEQ ID NO: 131), Strep tag (which refers the original STREP® tag (WRHPQFGG; SEQ ID NO: 132), STREP® tag II (WSHPQFEK; SEQ ID NO: 133 (IBA Institut für Bioanalytik, Germany); see, e.g., US 7,981,632), Softag 1 (SLAELLNAGLGGS; SEQ ID NO: 134), Softag 3 (TQDPSRVG; SEQ ID NO: 135), and V5 tag (GKPIPNPLLGLDST; SEQ ID NO: 136). In particular embodiments, a tag cassette includes a fusion of tag cassettes such as 3XFLAG. In particular embodiments, 3XFLAG includes the sequence set forth in SEQ ID NO: 61.

[0070] 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), glial cell line-derived neurotrophic factor (GDNF) (e.g., NP_001177397.1), platelet-derived growth factor receptor (e.g., GenBank: AAA60049.1), epidermal growth factor receptor (e.g., GenBank: CAA25240.1), vascular endothelial growth factor receptor (e.g., GenBank: AAC16449.2), interleukin 8 receptor alpha (e.g., GenBank: AAB59436.1), caveolin (e.g., GenBank: CAA79476.1), dynamin (e.g., GenBank: AAA88025.1), clathrin heavy chain 1 isoform 1 (e.g., NP_004850.1), clathrin heavy chain 2 isoform 1 (e.g.,

NP_009029.3), clathrin light chain A isoform a (e.g., NP_001824.1), clathrin light chain B isoform a (e.g., NP_001825.1), ras-related protein Rab-4A isoform 1 (e.g., NP_004569.2), ras-related protein Rab-11A (e.g., UniProtKB/Swiss-Prot: P62491.3), platelet-derived growth factor (e.g., GenBank: AAA60552.1), transforming growth factor-beta3 (e.g., GenBank: AAA61161.1), nerve growth factor (e.g., GenBank: CAA37703.1), EGF (e.g., GenBank: CAA34902.2), cocaine and amphetamine regulated transcript (Chain A) (e.g., PDB: 1HY9_A), protachykinin-1 (e.g., UniProtKB - P20366), oxytocin-neurophysin 1 (e.g., UniProtKB - P01178), somatostatin (e.g., GenBank: AAH32625.1), genetically-encoded green calcium indicator NTnC (chain A) [synthetic construct] (e.g., PDB: 5MWC_A), calcium indicator TN-XXL [synthetic construct], (e.g., GenBank: ACF93133.1), BRET-based auto-luminescent calcium indicator [synthetic construct] (e.g., GenBank ADF42668.1), calcium indicator protein OeNL(Ca²⁺)-18u [synthetic construct], ((e.g., GenBank BBB18812.1), myosin light chain kinase, Green fluorescent protein, Calmodulin chimera (Chain A) [synthetic construct] ((e.g., PDB: 3EKJ_A), channelopsin 1 (e.g., *UniProtKB* - F8UVI5), channelopsin 1 (e.g., GenBank: AER58217.1), channelrhodopsin-2 ((e.g., UniProtKB - B4Y105), channel rhodopsin 2 [synthetic construct] ((e.g., GenBank: ABO64386.1), CRISPR-associated protein (Cas) (e.g., GenBank: AKG27598.1), Cas9 [synthetic construct] (e.g., GenBank: AST09977.1), CRISPR-associated endonuclease Cpf1 (e.g., UniProtKB/Swiss-Prot: U2UMQ6.1), ribonuclease 4 or ribonuclease L (e.g., UniProtKB/Swiss-Prot: Q05823.2), deoxyribonuclease II beta (e.g., GenBank: AAF76893.1), sodium channel protein type 1 subunit alpha (e.g., UniProtKB - P35498), potassium voltage-gated channel subfamily KQT member 2 (e.g., UniProtKB - O43526), and voltage-dependent L-type calcium channel subunit alpha-1C (e.g., UniProtKB - Q13936).

[0071] Additional effector elements include Cre, iCre, dgCre, FlpO, and tTA2. iCre refers to a codon-improved Cre. dgCre refers to an enhanced GFP/Cre recombinase fusion gene with an N terminal fusion of the first 159 amino acids of the *Escherichia coli* K-12 strain chromosomal dihydrofolate reductase gene (DHFR or *folA*) harboring a G67S mutation and modified to also include the R12Y/Y100I destabilizing domain mutation. FlpO refers to a codon-optimized form of FLPe that greatly increases protein expression and FRT recombination efficiency in mouse cells. Like the Cre/LoxP system, the FLP/FRT system has been widely used for gene expression (and generating conditional knockout mice, mediated by the FLP/FRT system). tTA2 refers to tetracycline transactivator.

[0072] 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.

[0073] Exemplary self-cleaving peptides include the 2A peptides which lead to the production of two proteins from one mRNA. The 2A sequences are short (e.g., 20 amino acids), allowing more use in size-limited constructs. Particular examples include P2A, T2A, E2A, and F2A. In particular embodiments, the artificial expression constructs include an internal ribosome entry site (IRES) sequence. IRES allow ribosomes to initiate translation at a second internal site on a mRNA molecule, leading to production of two proteins from one mRNA.

[0074] Artificial expression constructs can encode nuclear localization proteins, such as Histone H1, Histone H2A, Histone H2B, Histone H3, Histone H4, histone-like protein HPhA, or H2B*.

[0075] 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.

[0076] 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.

[0077] Promoters can include general promoters, tissue-specific promoters, cell-specific promoters, and/or promoters specific for the cytoplasm. Promoters may include strong promoters, weak promoters, constitutive expression promoters, and/or inducible promoters. Inducible promoters direct expression in response to certain conditions, signals or cellular events. For example, the promoter may be an inducible promoter that requires a particular ligand, small molecule, transcription factor or hormone protein in order to effect transcription from the promoter. Particular examples of promoters include minBglobin (also referred to as minBGprom), CMV, minCMV, minCMV* (minCMV* is minCMV with a SacI restriction site removed), minRho, minRho* (minRho* is minRho with a SacI restriction site removed), SV40 immediately early promoter, the Hsp68 minimal promoter (proHSP68), and the Rous Sarcoma Virus (RSV) long-terminal repeat (LTR) promoter. Minimal promoters have no activity to drive gene expression on their own but

can be activated to drive gene expression when linked to a proximal enhancer element.

[0078] 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.

[0079] 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.

[0080] 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.

[0081] 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.

[0082] 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.

[0083] 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.

[0084] 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.

[0085] 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.

[0086] 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.

[0087] 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.

[0088] 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.

[0089] 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.

[0090] "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.

[0091] 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.

[0092] 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.

[0093] 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.

[0094] 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.

[0095] 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.

[0096] In particular embodiments, a viral vector further includes one or more insulator elements. Insulator 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).

[0097] 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.

[0098] Particular embodiments of vectors disclosed herein include:

Expression Construct Name	Features
CN2415	rAAV-eHGT_576h-minBglobin-SYFP2-WPRE3-BGHpA
CN2416	rAAV-eHGT_577h-minBglobin-SYFP2-WPRE3-BGHpA
CN2417	rAAV-eHGT_578h-minBglobin-SYFP2-WPRE3-BGHpA
CN2418	rAAV-3xSP10ins-eHGT_579h-minRho*-SYFP2-WPRE3-BGHpA
CN2436	rAAV-eHGT_606h-minBglobin-SYFP2-WPRE3-BGHpA

CN3000	pAAV-eHGT_827h-minBglobin-SYFP2-WPRE3-BGHpA
CN3001	pAAV-eHGT_828h-minBglobin-SYFP2-WPRE3-BGHpA
CN3003	pAAV-eHGT_830h-minBglobin-SYFP2-WPRE3-BGHpA
CN3004	pAAV-eHGT_831h-minBglobin-SYFP2-WPRE3-BGHpA
CN3005	pAAV-eHGT_832h-minBglobin-SYFP2-WPRE3-BGHpA
CN3007	pAAV-eHGT_834h-minBglobin-SYFP2-WPRE3-BGHpA
CN3009	pAAV-eHGT_836h-minBglobin-SYFP2-WPRE3-BGHpA
AiP1335	rAAV-MGT_E117-minBglobin-SYFP2-WPRE3-BGHpA
AiP1336	rAAV-MGT_E118-minBglobin-SYFP2-WPRE3-BGHpA
AiP1337	rAAV-MGT_E119-minBglobin-SYFP2-WPRE3-BGHpA
AiP1338	rAAV-MGT_E120-minBglobin-SYFP2-WPRE3-BGHpA
AiP1339	rAAV-MGT_E121-minBglobin-SYFP2-WPRE3-BGHpA
CN2555	rAAV-3xSP10ins-3xcore2-eHGT_367h-minRho*-SYFP2-WPRE3-BGHpA
CN2045	rAAV-3xSP10ins-eHGT_359h-minRho*-SYFP2-WPRE3-BGHpA
CN2258	rAAV-eHGT_479m-minBglobin-SYFP2-WPRE3-BGHpA
CN2251	rAAV-eHGT_453m-minBglobin-SYFP2-WPRE3-BGHpA
CN1633	rAAV-hsA2-eHGT_140h-minRho-SYFP2-WPRE3-BGHpA
CN2043	rAAV-3xSP10ins-eHGT_356h-minRho*-SYFP2-WPRE3-BGHpA
CN1621	rAAV-hsA2-eHGT_128h-minRho-SYFP2-WPRE3-BGHpA
CN2216	pAAV-3xcoreB_eHGT121h-minCMV*-SYFP2-WPRE3-BGHpA
CN2717	rAAV-eHGT_710m-minBglobin-SYFP2-WPRE3-BGHpA
CN3639	pAAV-eHGT_895h-minBglobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN3050	rAAV-3xcore2_eHGT_577h-minBglobin-SYFP2-WPRE3-BGHpA
CN3051	rAAV-3xcore3_eHGT_577h-minBglobin-SYFP2-WPRE3-BGHpA
CN3056	rAAV-3xcore2_eHGT_606h-minBglobin-SYFP2-WPRE3-BGHpA
CN3057	rAAV-3xcore3_eHGT_606h-minBglobin-SYFP2-WPRE3-BGHpA
CN4001	rAAV-3xcore4_eHGT_577h-minBglobin-SYFP2-WPRE3-BGHpA
CN4003	rAAV-core6_eHGT_606h-minBglobin-SYFP2-WPRE3-BGHpA
CN2786	pAAV-3xcore_eHGT_121h-minBglobin-SYFP2-WPRE3-BGHpA
CN2840	rAAV-eHGT_590m-minBglobin-SYFP2-WPRE3-BGHpA
CN3460	pAAV-eHGT_976h-minBglobin-SYFP2-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA
CN2650	rAAV-eHGT_717h-minBglobin-SYFP2-WPRE3-BGHpA

[0099] 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. 24). 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.

[0100] 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: 119) are changed to S-DGTLAVPFK-A (SEQ ID NO: 120). In particular embodiments, PHP.eb refers to SEQ ID NO: 74.

[0101] 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).

[0102] 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.

[0103] 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.

[0104] 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.

[0105] AAV-BR1 is an AAV2 variant displaying the NRGTEWD (SEQ ID NO: 121) 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).

[0106] 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: 122), transduces neurons in the enteric nervous system, and strongly transduces peripheral sensory afferents entering the spinal cord and brain stem.

[0107] 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: 123). It transfers genes throughout the CNS with higher efficiency than AAV9 and transduces the majority of astrocytes and neurons across multiple CNS regions.

[0108] AAV-PPS, an AAV2 variant crated by insertion of the DSPAHPS (SEQ ID NO: 124) epitope into the capsid of AAV2, shows a dramatically improved brain tropism relative to AAV2.

[0109] For additional information regarding capsids that cross the blood brain barrier, see Chan et al., Nat. Neurosci. 2017 Aug; 20(8): 1172-1179.

[0110] (ii) Compositions for Administration. Artificial expression constructs and vectors of the present disclosure (referred to herein as physiologically active components) can be formulated with a carrier that is suitable for administration to a cell, tissue slice, animal (e.g., mouse, non-human primate), or human. Physiologically active components within compositions described herein can be prepared in neutral forms, as freebases, or as pharmacologically acceptable salts.

[0111] 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.

[0112] 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.

[0113] 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).

[0114] 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.

[0115] Compositions may include liposomes, lipids, lipid complexes, microspheres, microparticles, nanospheres, and/or nanoparticles.

[0116] 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).

[0117] 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.

[0118] 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.

[0119] 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.

[0120] 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.

[0121] 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.

[0122] 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.

[0123] 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).

[0124] Supplementary active ingredients can also be incorporated into the compositions.

[0125] 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.

[0126] 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.

[0127] (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.

[0128] 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_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, or eHGT_710m, and/or CN2415, CN2416, CN2417, CN2418, CN2436, CN3000, CN3001, CN3003, CN3004, CN3005, CN3007, CN3009, AiP1335, AiP1336, AiP1337, AiP1338, AiP1339, CN2555, CN2045, CN2258, CN2251, CN1633, CN2043, CN1621, CN2216, CN2717, CN3639, CN3050, CN3051, CN3056, CN3057, CN4001, CN4003, CN2786, CN2840, CN3460, or CN2650, and the cell line is a human, primate, or murine cell. Cell lines which can be utilized for transgenesis in the present disclosure also include primary cell lines

derived from living tissue such as rat or mouse brains and organotypic cell cultures, including brain slices from animals such as rats, mice, non-human primates, or human neurosurgical tissue. The PC12 cell line (available from the American Type Culture Collection, ATCC, Manassas, VA) has been shown to express a number of neuronal marker proteins in response to Neuronal Growth Factor (NGF). The PC12 cell line is considered to be a neuronal cell line and is applicable for use with this disclosure. JAR cells (available from ATCC) are a platelet derived cell-line that express some neuronal genes, such as the serotonin transporter gene, and may be used with embodiments described herein.

[0129] 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.

[0130] 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.

[0131] 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).

[0132] 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

[0133] 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).

[0134] 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_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, or eHGT_710m.

[0135] 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).

[0136] (iv) Transgenic Animals. Another aspect of the disclosure includes transgenic animals, the genome of which contains an artificial expression construct including eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h,

3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, and/or eHGT_710m operatively linked to a heterologous coding sequence. In particular embodiments, the genome of a transgenic animal includes CN2415, CN2416, CN2417, CN2418, CN2436, CN3000, CN3001, CN3003, CN3004, CN3005, CN3007, CN3009, AiP1335, AiP1336, AiP1337, AiP1338, AiP1339, CN2555, CN2045, CN2258, CN2251, CN1633, CN2043, CN1621, CN2216, CN2717, CN3639, CN3050, CN3051, CN3056, CN3057, CN4001, CN4003, CN2786, CN2840, CN3460, and/or CN2650. In particular embodiments, when a non-integrating vector is utilized, a transgenic animal includes an artificial expression construct including eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, and/or eHGT_710m, and/or CN2415, CN2416, CN2417, CN2418, CN2436, CN3000, CN3001, CN3003, CN3004, CN3005, CN3007, CN3009, AiP1335, AiP1336, AiP1337, AiP1338, AiP1339, CN2555, CN2045, CN2258, CN2251, CN1633, CN2043, CN1621, CN2216, CN2717, CN3639, CN3050, CN3051, CN3056, CN3057, CN4001, CN4003, CN2786, CN2840, CN3460, and/or CN2650 within one or more of its cells.

[0137] 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.

[0138] 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.

[0139] 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.

[0140] 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.

[0141] (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.

[0142] 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.

[0143] Particular embodiments include methods of administering to a subject an artificial expression construct that includes eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h,

3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, or eHGT_710m, and/or CN2415, CN2416, CN2417, CN2418, CN2436, CN3000, CN3001, CN3003, CN3004, CN3005, CN3007, CN3009, AiP1335, AiP1336, AiP1337, AiP1338, AiP1339, CN2555, CN2045, CN2258, CN2251, CN1633, CN2043, CN1621, CN2216, CN2717, CN3639, CN3050, CN3051, CN3056, CN3057, CN4001, CN4003, CN2786, CN2840, CN3460, and/or CN2650 as described herein to drive selective expression of a gene in a selected cell type as described herein to drive expression of a gene in a targeted cell type. The subject can be an isolated cell, a network of cells, a tissue slice, an experimental animal, a veterinary animal, or a human.

[0144] 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.

[0145] An "effective amount" is the amount of a composition necessary to result in a desired physiological change in the subject. Effective amounts are often administered for research purposes. Effective amounts disclosed herein can cause a statistically-significant effect in an animal model, human study, *in vivo*, or *in vitro* assay.

[0146] 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.

[0147] 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.

[0148] (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.

[0149] 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.

[0150] (vii) Exemplary Embodiments. 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.

1. An artificial enhancer including a core of an eHGT_367h, eHGT_121h, eHGT_577h, and/or eHGT_606h enhancer.
2. The artificial enhancer of embodiment 1, wherein the core has the sequence as set forth in SEQ ID NO: 18, SEQ ID NO: 26, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, or SEQ ID NO: 41 or a sequence having at least 90% sequence identity to the sequence as set forth in SEQ ID NO: 18, SEQ ID NO: 26, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, or SEQ ID NO: 41.
3. The artificial enhancer of embodiments 1 or 2, wherein the core has 2, 3, 4, 5, 6, 7, 8, 9, or 10

copies of the sequence as set forth in SEQ ID NO: 18, SEQ ID NO: 26, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, or SEQ ID NO: 41 or a sequence having at least 90% sequence identity to the sequence as set forth in SEQ ID NO: 18, SEQ ID NO: 26, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, or SEQ ID NO: 41.

4. The artificial enhancer of embodiment 3, having 3 copies of the sequence as set forth in SEQ ID NO: 18.

5. The artificial enhancer of embodiment 3, having 3 copies of the sequence as set forth in SEQ ID NO: 26.

6. The artificial enhancer of embodiment 3, having 3 copies of the sequence as set forth in SEQ ID NO: 31.

7. The artificial enhancer of embodiment 3, having 3 copies of the sequence as set forth in SEQ ID NO: 33.

8. The artificial enhancer of embodiment 3, having 3 copies of the sequence as set forth in SEQ ID NO: 35.

9. The artificial enhancer of embodiment 3, having 3 copies of the sequence as set forth in SEQ ID NO: 37.

10. The artificial enhancer of embodiment 2, having 1 copy of the sequence as set forth in SEQ ID NO: 38.

11. The artificial enhancer of embodiment 2, having 1 copy of the sequence as set forth in SEQ ID NO: 39.

12. The artificial enhancer of embodiment 3, having 3 copies of the sequence as set forth in SEQ ID NO: 41.

13. The artificial enhancer of embodiment 4, wherein the artificial enhancer has the sequence as set forth in SEQ ID NO: 17 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 17.

14. The artificial enhancer of embodiment 5, wherein the artificial enhancer core has the sequence as set forth in SEQ ID NO: 25 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 25.

15. The artificial enhancer of embodiment 6, wherein the artificial enhancer has the sequence as set forth in SEQ ID NO: 30 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 30.

16. The artificial enhancer of embodiment 7, wherein the artificial enhancer has the sequence as set forth in SEQ ID NO: 32 or a sequence having at least 95% sequence identity to the sequence

as set forth in SEQ ID NO: 32.

17. The artificial enhancer of embodiment 8, wherein the artificial enhancer has the sequence as set forth in SEQ ID NO: 34 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 34.

18. The artificial enhancer of embodiment 9, wherein the artificial enhancer core has the sequence as set forth in SEQ ID NO: 36 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 36.

19. The artificial enhancer of embodiment 12, wherein the artificial enhancer has the sequence as set forth in SEQ ID NO: 40 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 40.

20. An artificial expression construct including (i) an enhancer selected from eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, or eHGT_710m; (ii) a promoter; and (iii) a heterologous coding sequence.

21. The artificial expression construct of embodiment 20, wherein the heterologous encoding sequence encodes an effector element or an expressible element.

22. The artificial expression construct of embodiment 20 or 21, wherein the effector element includes a reporter protein or a functional molecule.

23. The artificial expression construct of embodiment 22, wherein the reporter protein includes a fluorescent protein.

24. The artificial expression construct of embodiment 22 or 23, 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).

25. The artificial expression construct embodiment 21, wherein the expressible element includes a non-functional molecule.

26. The artificial expression construct of embodiment 25, 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.

27. The artificial expression construct of any of embodiments 20-26, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.

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

29. The artificial expression construct of any of embodiments 20-28, wherein the artificial expression construct includes or encodes a skipping element.

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

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

32. The artificial expression construct of any of embodiments 20-31, wherein the artificial expression construct includes or encodes a set of features selected from: eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, eHGT_710m, AAV, scAAV, rAAV, pAAV, minBglobin, CMV, minCMV, minCMV*, minRho, minRho*, fluorescent protein (e.g., EGFP, SYFP, GFP), hsA2, Cre, iCre, dgCre, FlpO, tTA2, SP10ins (e.g., 3xSP10ins), tag cassette, 10aa, nuclear localization protein, self-cleaving peptides, WPRE, WPRE3, hGHpA, and/or BGHpA.

33. The artificial expression construct of any of embodiments 20-32, wherein the artificial expression construct includes or encodes a set of features selected from:

eHGT_576h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_577h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_578h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_579h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_606h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_827h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_828h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_830h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_831h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_832h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_834h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_836h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
MGT_E117-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
MGT_E118-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
MGT_E119-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
MGT_E120-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
MGT_E121-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
3xSP10ins-3xcore2_eHGT_367h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
3xSP10ins-eHGT_359h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_479m-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_453m-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
hsA2-eHGT_140h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
3xSP10ins-eHGT_356h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
hsA2-eHGT_128h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
3xcoreB_eHGT121h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_710m-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_895h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA;
3xcore2_eHGT_577h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
3xcore3_eHGT_577h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
3xcore2_eHGT_606h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
3xcore3_eHGT_606h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
core4_eHGT_577h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
core6_eHGT_606h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
3xcore_eHGT_121h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;
eHGT_590m-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_976h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-WPRE3-BGHpA;

eHGT_717h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_576h-[minimal promoter]-[heterologous coding sequence]-[post-regulatory elements];

eHGT_577h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

eHGT_578h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xSP10ins-eHGT_579h- [minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

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

eHGT_827h-[minimal promoter]-[heterologous coding sequence]-[post-regulatory elements];

eHGT_828h-[minimal promoter]-[heterologous coding sequence]-[post-regulatory elements];

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

eHGT_831h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

eHGT_832h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

eHGT_834h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

eHGT_836h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

MGT_E117-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

MGT_E118-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

MGT_E119-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

MGT_E120-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

MGT_E121-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xSP10ins-3xcore2_eHGT_367h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xSP10ins-eHGT_359h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

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

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

hsA2-eHGT_140h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xSP10ins-eHGT_356h- [minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

hsA2-eHGT_128h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xcoreB_eHGT121h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

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

eHGT_895h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xcore2_eHGT_577h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xcore3_eHGT_577h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xcore2_eHGT_606h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xcore3_eHGT_606h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

core4_eHGT_577h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

core6_eHGT_606h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

3xcore_eHGT_121h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

eHGT_590m-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];

eHGT_976h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
and

eHGT_717h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements].

34. A vector including an artificial expression construct of any of embodiments 20-33.

35. The vector of embodiment 34, wherein the vector includes a viral vector.

36. The vector of embodiment 34 or 35, wherein the viral vector includes a recombinant adeno-associated viral (AAV) vector.

37. An adeno-associated viral (AAV) vector including at least one heterologous coding sequence, wherein the heterologous coding sequence is under the transcriptional control of a promoter and an enhancer selected from eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, 3xCore2-eHGT_367h, eHGT_359h, eHGT_479m, eHGT_453m, eHGT_140h, eHGT_356h, eHGT_128h, eHGT_369h, and eHGT_710m.

38. The AAV vector of embodiment 37, wherein the heterologous coding sequence encodes an effector element or an expressible element.
39. The AAV vector of embodiment 38, wherein the effector element includes a reporter protein or a functional molecule.
40. The AAV vector of embodiment 39, wherein the reporter protein includes a fluorescent protein.
41. The AAV vector 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 DREADD.
42. The AAV vector of embodiment 38, wherein the expressible element includes a non-functional molecule.
43. The AAV vector 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. A transgenic cell including an artificial expression construct or a vector of any of the preceding embodiments.
45. The transgenic cell of embodiment 44, wherein the transgenic cell is a neuron within the thalamus.
46. The transgenic cell of embodiment 44, wherein the transgenic cell is a GABAergic or glutamatergic neuron within the thalamus.
47. The transgenic cell of embodiment 46, wherein the GABAergic neuron includes Gata/Dlx5-6 cells.
48. The transgenic cell of embodiment 46, wherein the glutamatergic neuron includes Prkcd-Grin2c cells.
49. The transgenic cell of embodiment 46, wherein the glutamatergic neuron include Rxfp1-Epb4 cells.
50. The transgenic cell of embodiment 46, wherein the GABAergic neuron includes thalamic reticular nucleus (TRN) cells of the thalamus.
51. The transgenic cell of embodiment 46, wherein the glutamatergic neuron include parafascicular (Pf) nuclei of the thalamus.
52. The transgenic cell of embodiment 44, wherein the transgenic cell is secondary cell type.

53. The transgenic cell of embodiment 52, wherein the secondary cell type is a striatal medium spiny neuron.
54. The transgenic cell of embodiment 52, wherein the secondary cell type is a Purkinje cell in the cerebellum.
55. The transgenic cell of embodiment 52, wherein the secondary cell type is a deep cerebellar nuclei (DCN) cell in the cerebellum.
56. The transgenic cell of embodiment 52, wherein the secondary cell type is a molecular layer interneuron (MLI) in the cerebellum.
57. The transgenic cell of embodiment 52, wherein the secondary cell type is a Pvalb neuron.
58. The transgenic cell of embodiment 52, wherein the secondary cell type is a chandelier cell.
59. The transgenic cell of embodiment 52, wherein the secondary cell type is a glutamatergic L5 ET cell in the neocortex.
60. The transgenic cell of embodiment 52, wherein the secondary cell type is a Vip neuron in the neocortex.
61. The transgenic cell of embodiment 44, wherein the transgenic cell is murine, human, or non-human primate.
62. A non-human transgenic animal including an artificial expression construct, a vector, and/or a transgenic cell of any of the preceding embodiments.
63. The non-human transgenic animal of embodiment 62, wherein the non-human transgenic animal is a mouse or a non-human primate.
64. An administrable composition including an artificial expression construct, a vector, and/or a transgenic cell of any of the preceding embodiments.
65. A kit including an artificial expression construct, a vector, a transgenic cell, a non-human transgenic animal, and/or administrable compositions of any of the preceding embodiments.
66. A method for expressing a gene within a population of cells in vivo or in vitro in or derived from the thalamus, the method including providing the administrable composition of embodiment 64 in a sufficient dosage and for a sufficient time to a sample or subject including the population of cells in or derived from the thalamus thereby expressing the gene within the population of cells.
67. The method of embodiment 66, wherein the gene encodes an effector element or an expressible element.
68. The method of embodiment 67, wherein the effector element includes a reporter protein or a functional molecule.
69. The method of embodiment 68, wherein the reporter protein includes a fluorescent protein.
70. The method of embodiment 68, wherein the functional molecule includes a functional ion

transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/Cas molecule, editase, guide RNA molecule, microRNA, homologous recombination donor cassette, or DREADD.

71. The method of embodiment 67, wherein the expressible element includes a non-functional molecule.

72. The method of embodiment 71, 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.

73. The method of any of embodiments 66 - 72, wherein the providing includes pipetting.

74. The method of embodiment 73, wherein the pipetting is to a brain slice.

75. The method of embodiment 74, wherein the brain slice includes a neuron within the thalamus.

76. The method of embodiment 74, wherein the brain slice includes a GABAergic or glutamatergic neuron within the thalamus.

77. The method of embodiment 74, wherein the brain slice includes a GABAergic Gata/Dlx5-6 neuron.

78. The method of embodiment 74, wherein the brain slice includes a glutamatergic Prkcd-Grin2c neuron.

79. The method of embodiment 74, wherein the brain slice includes a glutamatergic Rxrfp1-Epb4 neuron.

80. The method of embodiment 74, wherein the brain slice includes a GABAergic neuron within the thalamic reticular nucleus (TRN) cells of the thalamus.

81. The method of embodiment 74, wherein the brain slice includes a glutamatergic neuron of the parafascicular (Pf) nuclei of the thalamus.

82. The method of embodiment 74, wherein the brain slice includes a secondary cell type.

83. The method of embodiment 82, wherein the secondary cell type is a striatal medium spiny neuron.

84. The method of embodiment 82, wherein the secondary cell type is a Purkinje cell in the cerebellum.

85. The method of embodiment 82, wherein the secondary cell type is a deep cerebellar nuclei (DCN) cell in the cerebellum.

86. The method of embodiment 82, wherein the secondary cell type is a molecular layer

interneuron (MLI) in the cerebellum.

87. The method of embodiment 82, wherein the secondary cell type is a Pvalb neuron.

88. The method of embodiment 82, wherein the secondary cell type is a chandelier cell.

89. The method of embodiment 82, wherein the secondary cell type is a glutamatergic L5 ET cell in the neocortex.

90. The method of embodiment 82, wherein the secondary cell type is a Vip neuron in the neocortex.

91. The method of embodiment 74, wherein the brain slice is murine, human, or non-human primate.

92. The method of any of embodiments 66-91, wherein the providing includes administering to a living subject.

93. The method of embodiment 92, wherein the living subject is a human, non-human primate, or a mouse.

94. The method of embodiments 92 or 93, wherein the administering to a living subject is through injection.

95. The method of embodiment 94, wherein the injection includes intravenous injection, intraparenchymal injection into brain tissue, intracerebroventricular (ICV) injection, intra-cisterna magna (ICM) injection, or intrathecal injection.

96. An artificial expression construct including a sequence as set forth in SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, or SEQ ID NO: 118 or a sequence having at least 90% sequence identity to a sequence as set forth in SEQ ID NO: 83, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, or SEQ ID NO: 118.

[0151] (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.

[0152] 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.

[0153] 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).

[0154] 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.

[0155] As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: Arg (+3.0); Lys (+3.0); aspartate (+3.0 $\pm$ 1); glutamate (+3.0 $\pm$ 1); Ser (+0.3); Asn (+0.2); Gln (+0.2); Gly (0); Thr (-0.4); Pro (-0.5 $\pm$ 1); Ala (-0.5); His (-0.5); Cys (-1.0); Met (-1.3); Val (-1.5); Leu (-1.8); Ile (-1.8); Tyr (-2.3); Phe (-2.5); Trp (-3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

[0156] 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.

[0157] 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.

[0158] Variants of the protein, nucleic acid, and gene sequences disclosed herein also include sequences with at least 70% sequence identity, 80% sequence identity, 85% sequence identity, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the protein, nucleic acid, or gene sequences disclosed herein.

[0159] "% 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.

[0160] Variants also include nucleic acid molecules that hybridizes under stringent hybridization conditions to a sequence disclosed herein and provide the same function as the reference sequence. Exemplary stringent hybridization conditions include an overnight incubation at 42 °C in a solution including 50% formamide, 5XSSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5XDenhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1XSSC at 50 °C. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, moderately high stringency conditions include an overnight incubation at 37°C in a solution including 6XSSPE (20XSSPE=3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50 °C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5XSSC). Variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO,

heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

[0161] The term concatenate is broadly used to describe linking together into a chain or series. It is used to describe the linking together of nucleotide or amino acid sequences into a single nucleotide or amino acid sequence, respectively. The term “concatamerize” should be interpreted to recite: “concatenate.”

As will be understood by one of ordinary skill in the art, each embodiment disclosed herein can comprise, consist essentially of or consist of its particular stated element, step, ingredient or component. Thus, the terms “include” or “including” should be interpreted to recite: “comprise, consist of, or consist essentially of.” The transition term “comprise” or “comprises” means has, but is not limited to, and allows for the inclusion of unspecified elements, steps, ingredients, or components, even in major amounts. The transitional phrase “consisting of” excludes any element, step, ingredient or component not specified. The transition phrase “consisting essentially of” limits the scope of the embodiment to the specified elements, steps, ingredients or components and to those that do not materially affect the embodiment. A material effect would cause a statistically significant reduction in targeted expression in the targeted cell population as determined by scRNA-Seq and the following enhancer / targeted cell population pairings: eHGT_576h, eHGT_577h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_717h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, and eHGT_976h/ glutamatergic neurons within the thalamus; MGT_E117 and MGT_E118 / glutamatergic neurons (Prkcd-Grin2c (Core, LGN)) within the thalamus; MGT_E119 and MGT_E120 / GABAergic neurons (Gata/Dlx5-6) within thalamus; MGT_E121 / glutamatergic neurons (Rxfp1-Epb4 (Matrix)) within the thalamus; 3xCore2-eHGT_367h / glutamatergic neurons within the parafascicular (Pf) nuclei of the thalamus and striatal medium spiny neuron-pan; eHGT_359h / glutamatergic neurons within the thalamus, Purkinje cells in the cerebellum, and Pvalb neurons in the neocortex; eHGT_479m / glutamatergic neurons within the thalamus, Purkinje cells in the cerebellum, and chandelier cells in the neocortex; eHGT_453m / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus, Deep cerebellar nuclei (DCN) cells in the cerebellum, and glutamatergic L5 ET cells in the neocortex; eHGT_140h / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus and Pvalb neuron cell types; eHGT_356h / GABAergic neurons within the thalamic reticular nucleus (TRN) of the thalamus,

DCN cells in the cerebellum, Vip neurons in the neocortex, and cells of the thalamic reticular nucleus; eHGT_128h / glutamatergic neurons within the thalamus and Pvalb neuron cell types; eHGT_369h / glutamatergic neurons within the thalamus, molecular layer interneurons (MLI) cells in the cerebellum, and Pvalb neurons in the neocortex; and eHGT_710m / glutamatergic neurons within the thalamus and MLI cells in the cerebellum, chandelier cells, and molecular layer GABAergic interneurons in the cerebellum.

[0162] In particular embodiments, artificial means not naturally occurring.

[0163] Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term “about.” Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. When further clarity is required, the term “about” has the meaning reasonably ascribed to it by a person skilled in the art when used in conjunction with a stated numerical value or range, i.e. denoting somewhat more or somewhat less than the stated value or range, to within a range of $\pm 20\%$ of the stated value; $\pm 19\%$ of the stated value; $\pm 18\%$ of the stated value; $\pm 17\%$ of the stated value; $\pm 16\%$ of the stated value; $\pm 15\%$ of the stated value; $\pm 14\%$ of the stated value; $\pm 13\%$ of the stated value; $\pm 12\%$ of the stated value; $\pm 11\%$ of the stated value; $\pm 10\%$ of the stated value; $\pm 9\%$ of the stated value; $\pm 8\%$ of the stated value; $\pm 7\%$ of the stated value; $\pm 6\%$ of the stated value; $\pm 5\%$ of the stated value; $\pm 4\%$ of the stated value; $\pm 3\%$ of the stated value; $\pm 2\%$ of the stated value; or $\pm 1\%$ of the stated value.

[0164] 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.

[0165] 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.

[0166] 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.

[0167] 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.

[0168] 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.

[0169] 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.

[0170] 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.

[0171] 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 eHGT_577h enhancer; (ii) a promoter; and (iii) a heterologous encoding sequence.
2. An artificial enhancer comprising a core of an eHGT_577h, eHGT_606h, and/or eHGT_121h enhancer.
3. The artificial enhancer of claim 2, wherein the core has the sequence as set forth in SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 38, SEQ ID NO: 35, SEQ ID NO:37, SEQ ID NO:39, or SEQ ID NO: 41 or a sequence having at least 90% sequence identity to the sequence as set forth in SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 38, SEQ ID NO: 35, SEQ ID NO:37, SEQ ID NO:39, or SEQ ID NO: 41.
4. The artificial enhancer of claim 2, wherein the core has 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the sequence as set forth in SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 38, SEQ ID NO: 35, SEQ ID NO:37, SEQ ID NO:39, or SEQ ID NO: 41 or a sequence having at least 90% sequence identity to the sequence as set forth in SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 38, SEQ ID NO: 35, SEQ ID NO:37, SEQ ID NO:39, or SEQ ID NO: 41.
5. The artificial enhancer of claim 4, having 3 copies of the sequence as set forth in SEQ ID NO: 31.
6. The artificial enhancer of claim 4, having 3 copies of the sequence as set forth in SEQ ID NO: 33.
7. The artificial enhancer of claim 4, having 1 copy of the sequence as set forth in SEQ ID NO: 38.
8. The artificial enhancer of claim 4, having 3 copies of the sequence as set forth in SEQ ID NO: 35.
9. The artificial enhancer of claim 4, having 3 copies of the sequence as set forth in SEQ ID NO: 37.
10. The artificial enhancer of claim 4, having 1 copies of the sequence as set forth in SEQ ID NO: 39.
11. The artificial enhancer of claim 4, having 3 copies of the sequence as set forth in SEQ ID NO: 41.
12. The artificial enhancer of claim 5, wherein the concatenated enhancer core has the sequence as set forth in SEQ ID NO: 30 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 30.
13. The artificial enhancer of claim 6, wherein the concatenated core has the sequence as set

forth in SEQ ID NO: 32 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 32.

14. The artificial enhancer of claim 8, wherein the concatenated enhancer core has the sequence as set forth in SEQ ID NO: 34 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 34.

15. The artificial enhancer of claim 9, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 36 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 36.

16. The artificial enhancer of claim 11, wherein the concatenated enhancer core has the sequence as set forth in SEQ ID NO: 40 or a sequence having at least 95% sequence identity to the sequence as set forth in SEQ ID NO: 40.

17. An artificial expression construct comprising (i) an enhancer selected from eHGT_577h, eHGT_576h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, eHGT_717h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, or MGT_E121; (ii) a promoter; and (iii) a heterologous coding sequence.

18. The artificial expression construct of claim 17, wherein the heterologous encoding sequence encodes an effector element or an expressible element.

19. The artificial expression construct of claim 17, wherein the effector element comprises a reporter protein or a functional molecule.

20. The artificial expression construct of claim 19, wherein the reporter protein comprises a fluorescent protein.

21. The artificial expression construct of claim 19, 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).

22. The artificial expression construct of claim 18, wherein the expressible element comprises a non-functional molecule.

23. The artificial expression construct of claim 22, 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.

24. The artificial expression construct of claim 17, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.

25. The artificial expression construct of claim 24, wherein the capsid comprises PHP.eB, AAV-BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.

26. The artificial expression construct of claim 17, wherein the artificial expression construct comprises or encodes a skipping element.

27. The artificial expression construct of claim 26, wherein the skipping element comprises a 2A peptide and/or an internal ribosome entry site (IRES).

28. The artificial expression construct of claim 27, wherein the 2A peptide comprises T2A, P2A, E2A, or F2A.

29. The artificial expression construct of claim 17, wherein the artificial expression construct comprises or encodes a set of features selected from: eHGT_577h, eHGT_576h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, eHGT_717h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, MGT_E121, AAV, scAAV, rAAV, pAAV, minBglobin, CMV, minCMV, minCMV*, minRho, minRho*, fluorescent protein, hsA2, Cre, iCre, dgCre, FlpO, tTA2, SP10ins, tag cassette, 10aa, nuclear localization protein, self-cleaving peptides, WPRE, WPRE3, hGHpA, and/or BGHpA.

30. The artificial expression construct of claim 17, wherein the artificial expression construct comprises or encodes a set of features selected from:

eHGT_577h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

3xcore2_eHGT_577h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

3xcore3_eHGT_577h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

core4_eHGT_577h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_576h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_578h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

3xSP10ins-eHGT_579h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_606h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_827h-[minimal promoter]-[heterologous coding sequence]-WPRE3-BGHpA;

eHGT_828h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_830h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_831h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_832h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_834h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_836h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
MGT_E117-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
MGT_E118-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
MGT_E119-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
MGT_E120-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
MGT_E121-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_895h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-
WPRES-BGHPA;
3xcore2_eHGT_606h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
3xcore3_eHGT_606h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
core6_eHGT_606h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
3xcore_eHGT_121h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_590m-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_976h-[minimal promoter]-[heterologous coding sequence]-P2A-3XFLAG-10aa-H2B-
WPRES-BGHPA;
eHGT_717h-[minimal promoter]-[heterologous coding sequence]-WPRES-BGHPA;
eHGT_577h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
3xcore2_eHGT_577h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory
elements];
3xcore3_eHGT_577h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory
elements];
core4_eHGT_577h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory
elements];
eHGT_576h-[minimal promoter]-[heterologous coding sequence]-[post-regulatory elements];
eHGT_578h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
3xSP10ins-eHGT_579h- [minimal promoter]-[heterologous coding sequence]- [post-regulatory
elements];

eHGT_606h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_827h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_828h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_830h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_831h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_832h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_834h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_836h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 MGT_E117-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 MGT_E118-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 MGT_E119-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 MGT_E120-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 MGT_E121-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_895h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 3xcore2_eHGT_606h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 3xcore3_eHGT_606h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 core6_eHGT_606h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 3xcore_eHGT_121h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_590m-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 eHGT_976h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements];
 and
 eHGT_717h-[minimal promoter]-[heterologous coding sequence]- [post-regulatory elements].

31. A vector comprising an artificial expression construct of claim 17.

32. The vector of claim 31, wherein the vector comprises a viral vector.

33. The vector of claim 31, wherein the viral vector comprises a recombinant adeno-associated viral (AAV) vector.

34. An adeno-associated viral (AAV) vector comprising at least one heterologous coding sequence, wherein the heterologous coding sequence is under the transcriptional control of a

promoter and an enhancer selected from eHGT_577h, eHGT_576h, eHGT_578h, eHGT_579h, eHGT_606h, eHGT_827h, eHGT_828h, eHGT_830h, eHGT_831h, eHGT_832h, eHGT_834h, eHGT_836h, eHGT_895h, 3xcore2_eHGT_577h, 3xcore3_eHGT_577h, 3xcore2_eHGT_606h, 3xcore3_eHGT_606h, core4_eHGT_577h, core6_eHGT_606h, 3xcore_eHGT_121h, eHGT_590m, eHGT_976h, eHGT_717h, MGT_E117, MGT_E118, MGT_E119, MGT_E120, and MGT_E121.

35. The AAV vector of claim 34, wherein the heterologous coding sequence encodes an effector element or an expressible element.

36. The AAV vector of claim 35, wherein the effector element comprises a reporter protein or a functional molecule.

37. The AAV vector of claim 36, wherein the reporter protein comprises a fluorescent protein.

38. The AAV vector of claim 36, 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 DREADD.

39. The AAV vector of claim 35, wherein the expressible element comprises a non-functional molecule.

40. The AAV vector of claim 39, 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.

41. A transgenic cell comprising an artificial expression construct of claim 17, a vector of claim 31 and/or an AAV vector of claim 34.

42. The transgenic cell of claim 41, wherein the transgenic cell is a neuron within the thalamus.

43. The transgenic cell of claim 41, wherein the transgenic cell is a GABAergic or glutamatergic neuron within the thalamus.

44. The transgenic cell of claim 43, wherein the GABAergic neuron comprises Gata/Dlx5-6 cells.

45. The transgenic cell of claim 43, wherein the glutamatergic neuron comprises Prkcd-Grin2c cells.

46. The transgenic cell of claim 43, wherein the glutamatergic neuron comprise Rxfp1-Epb4 cells.

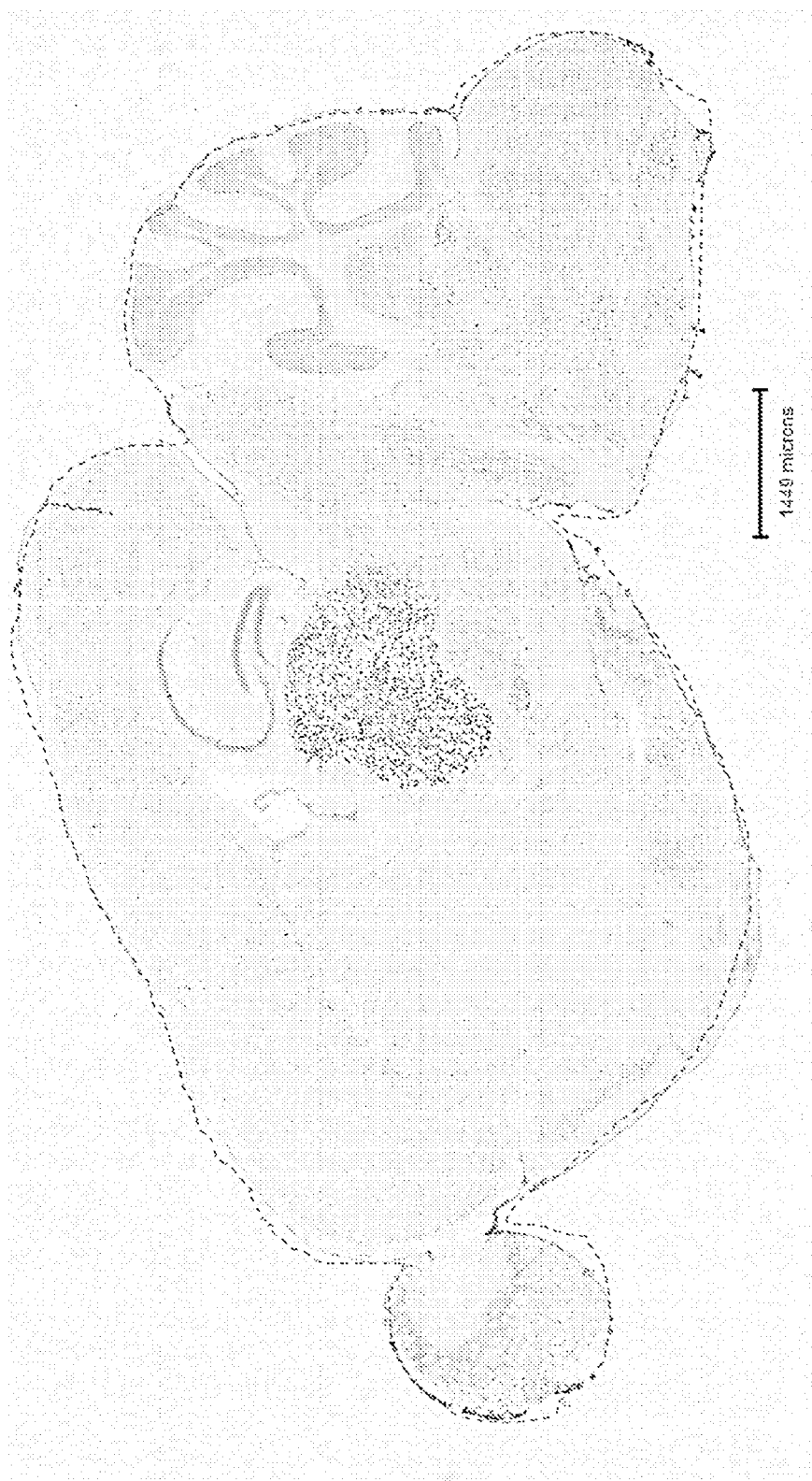
47. The transgenic cell of claim 41, wherein the transgenic cell is murine, human, or non-human primate.

48. A non-human transgenic animal comprising an artificial expression construct of claim 17, a vector of claim 31, and/or a transgenic cell of claim 41.
49. The non-human transgenic animal of claim 48, wherein the non-human transgenic animal is a mouse or a non-human primate.
50. An administrable composition comprising an artificial expression construct of claim 17, a vector of claim 31, and/or a transgenic cell of claim 41.
51. A kit comprising an artificial expression construct of claim 17, a vector of claim 31, a transgenic cell of claim 41, and/or a non-human transgenic animal of claim 48.
52. A method for expressing a gene within a population of cells in vivo or in vitro in or derived from the thalamus, the method comprising providing the administrable composition of claim 50 in a sufficient dosage and for a sufficient time to a sample or subject comprising the population of cells in or derived from the thalamus thereby expressing the gene within the population of cells.
53. The method of claim 52, wherein the gene encodes an effector element or an expressible element.
54. The method of claim 53, wherein the effector element comprises a reporter protein or a functional molecule.
55. The method of claim 54, wherein the reporter protein comprises a fluorescent protein.
56. The method of claim 54, 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 DREADD.
57. The method of claim 53, wherein the expressible element comprises a non-functional molecule.
58. The method of claim 57, 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.
59. The method of claim 52, 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 neuron within the thalamus.
62. The method of claim 60, wherein the brain slice comprises a GABAergic or glutamatergic neuron within the thalamus.

63. The method of claim 60, wherein the brain slice comprises a GABAergic Gata/Dlx5-6 neuron.
64. The method of claim 60, wherein the brain slice comprises a glutamatergic Prkcd-Grin2c neuron.
65. The method of claim 60, wherein the brain slice comprises a glutamatergic Rxfp1-Epb4 neuron.
66. The method of claim 60, wherein the brain slice is murine, human, or non-human primate.
67. The method of claim 52, wherein the providing comprises administering to a living subject.
68. The method of claim 57, wherein the living subject is a human, non-human primate, or a mouse.
69. The method of claim 67, wherein the administering to a living subject is through injection.
70. The method of claim 69, wherein the injection comprises intravenous injection, intraparenchymal injection into brain tissue, intracerebroventricular (ICV) injection, intra-cisterna magna (ICM) injection, or intrathecal injection.
71. An artificial expression construct consisting of or consisting essentially of a sequence as set forth in SEQ ID NO: 84, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, or SEQ ID NO: 118 or a sequence having at least 90% sequence identity to a sequence as set forth in SEQ ID NO: 84, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, or SEQ ID NO: 118.

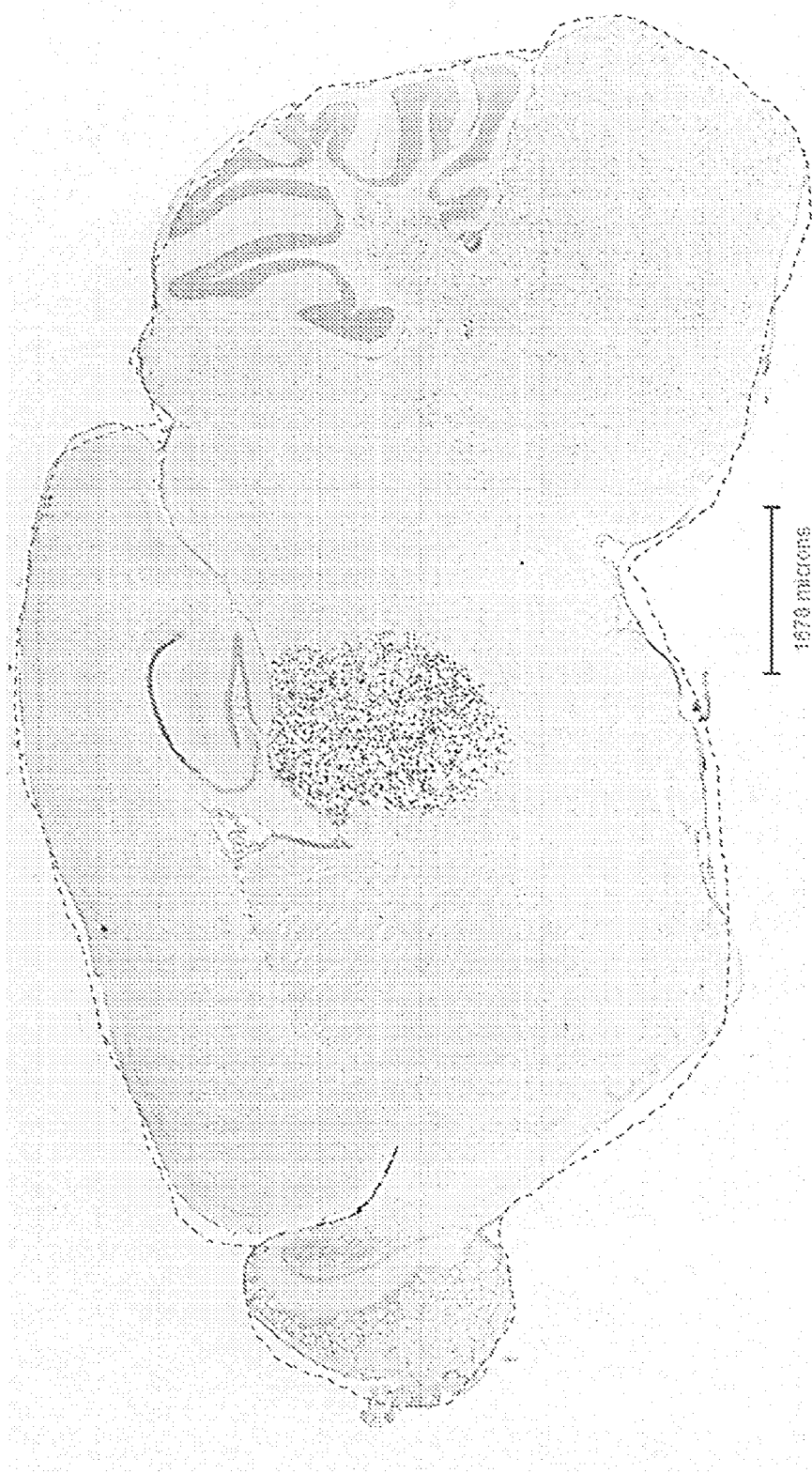
1/112

FIG. 1A
ISH: *Rgs16*



2/112

FIG. 1A cont'd
ISH: *Plekha7*



3/112

FIG. 1A cont'd
ISH: *Gad1*

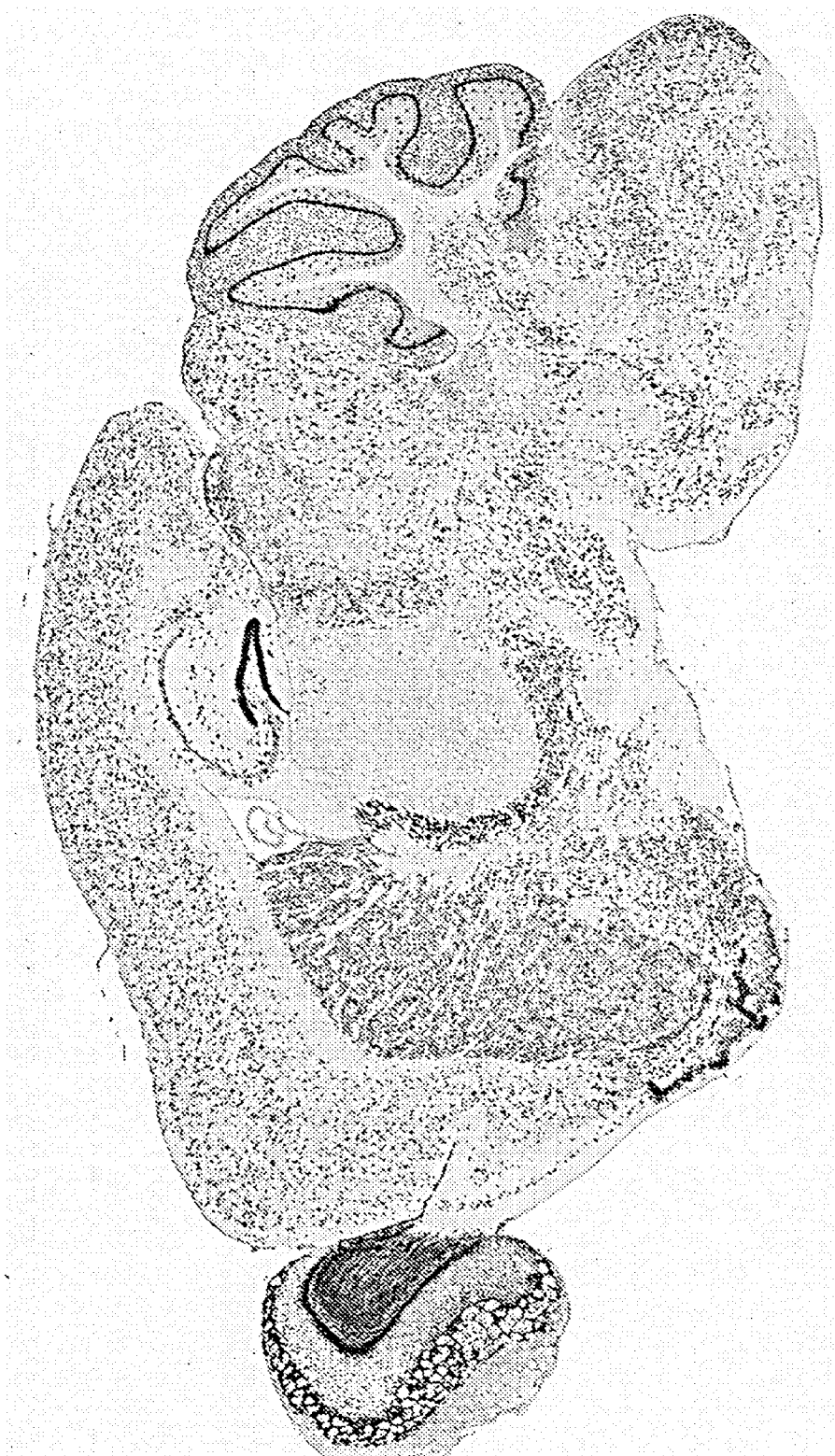


FIG. 1A cont'd
ISH: Slc17a7

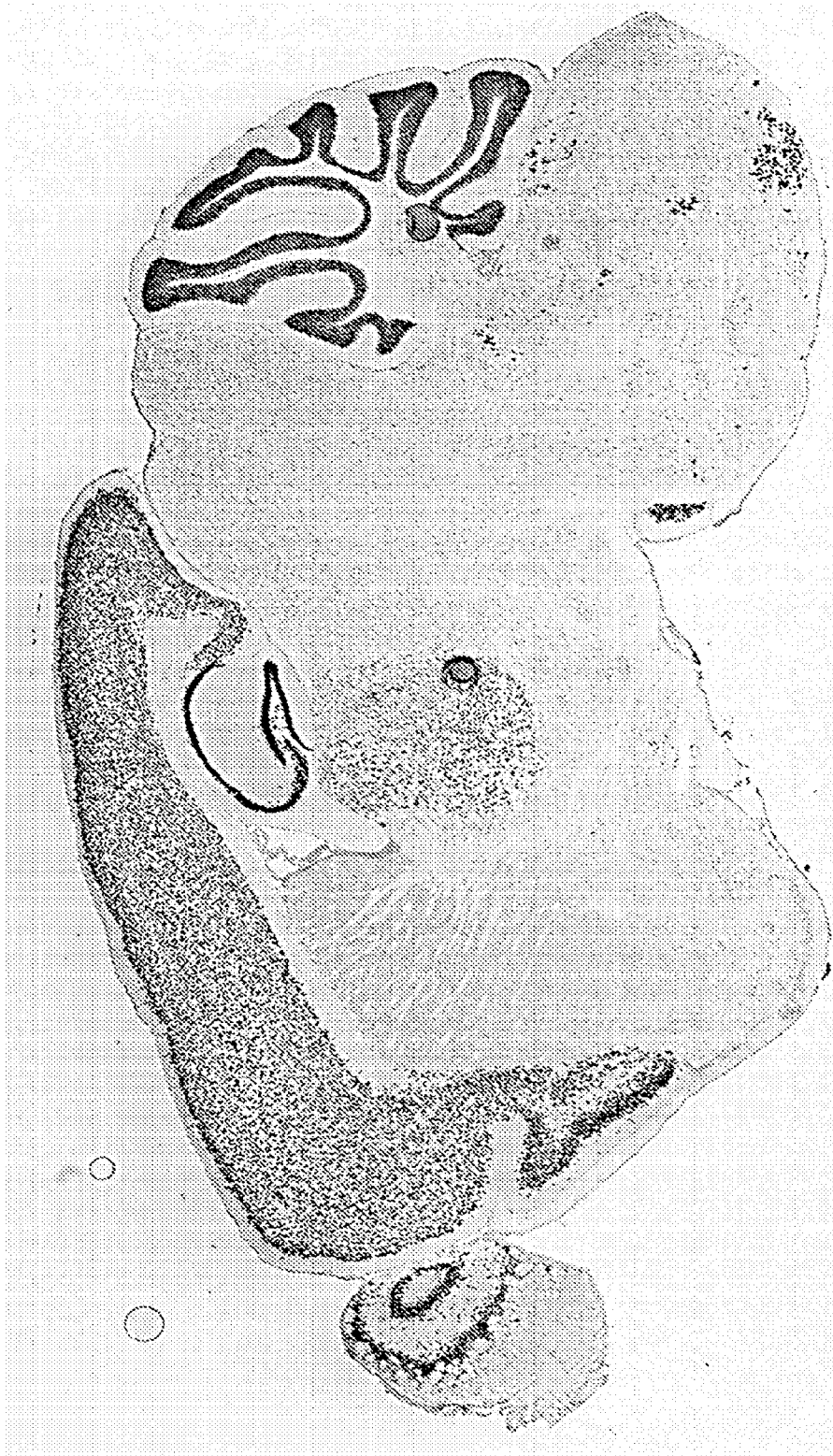


FIG. 1B

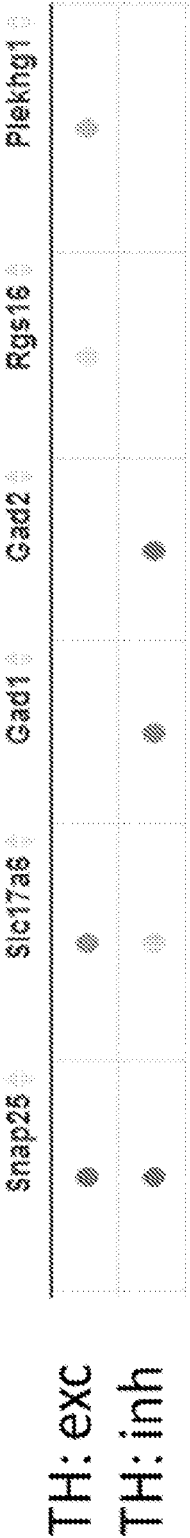


FIG. 1C

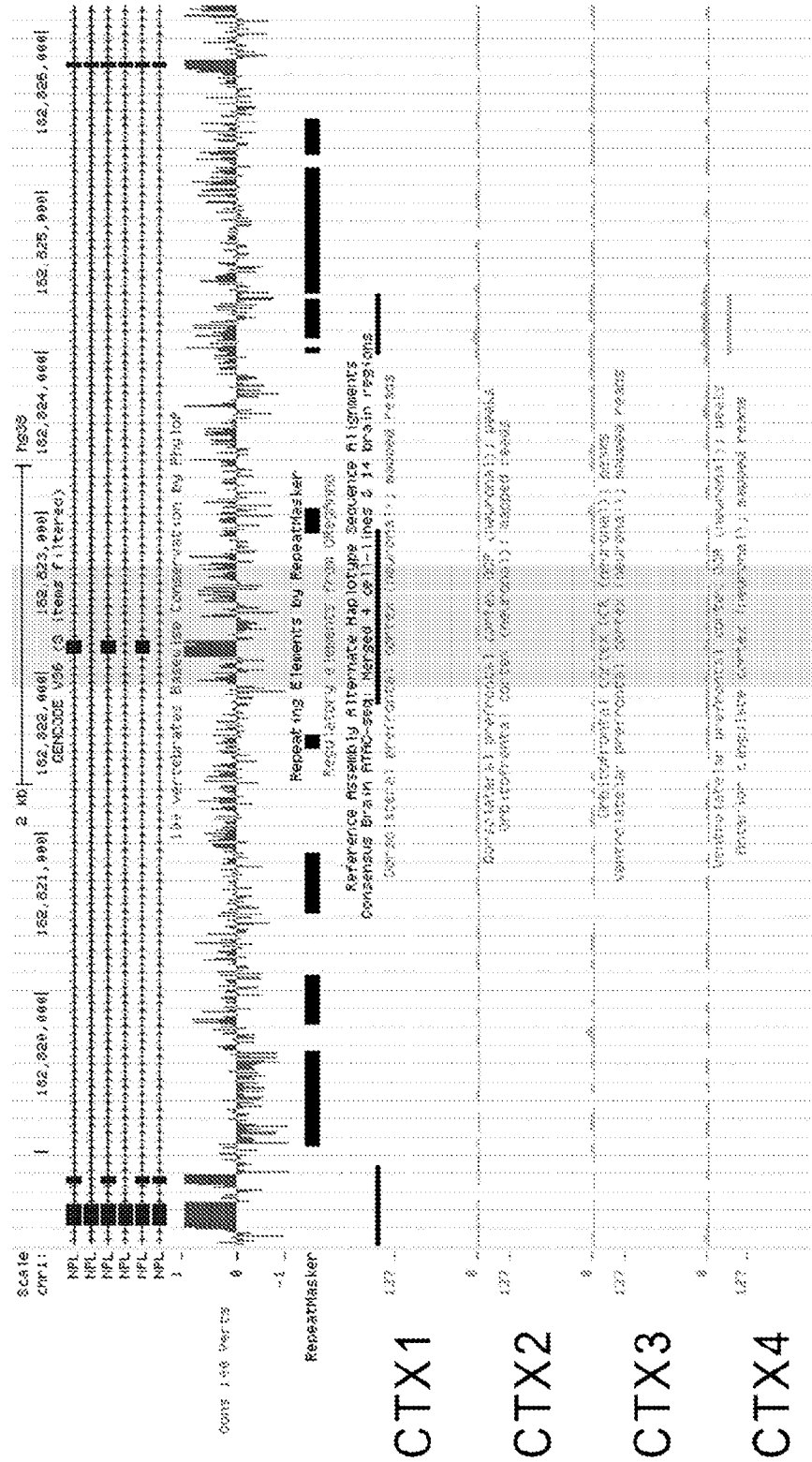
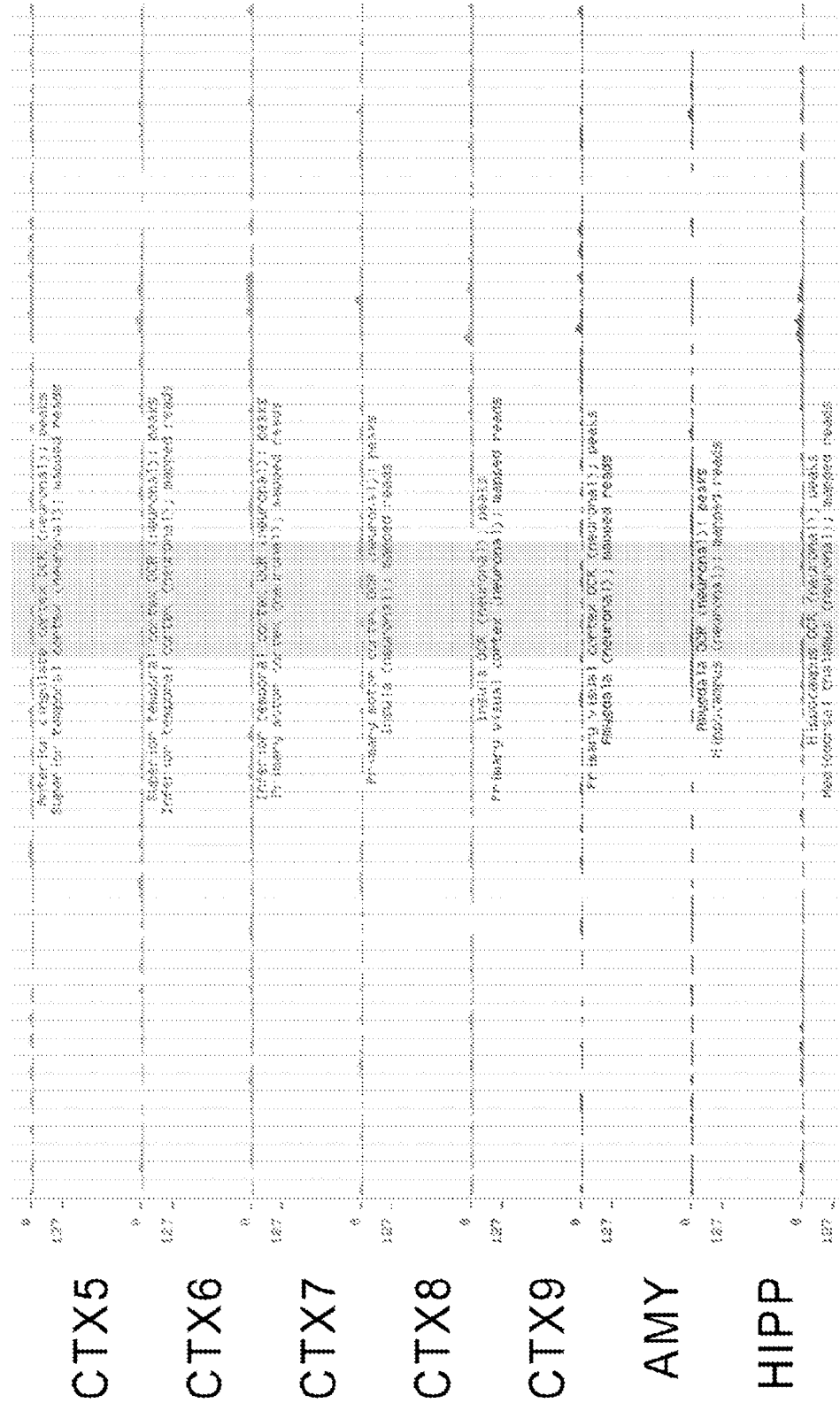
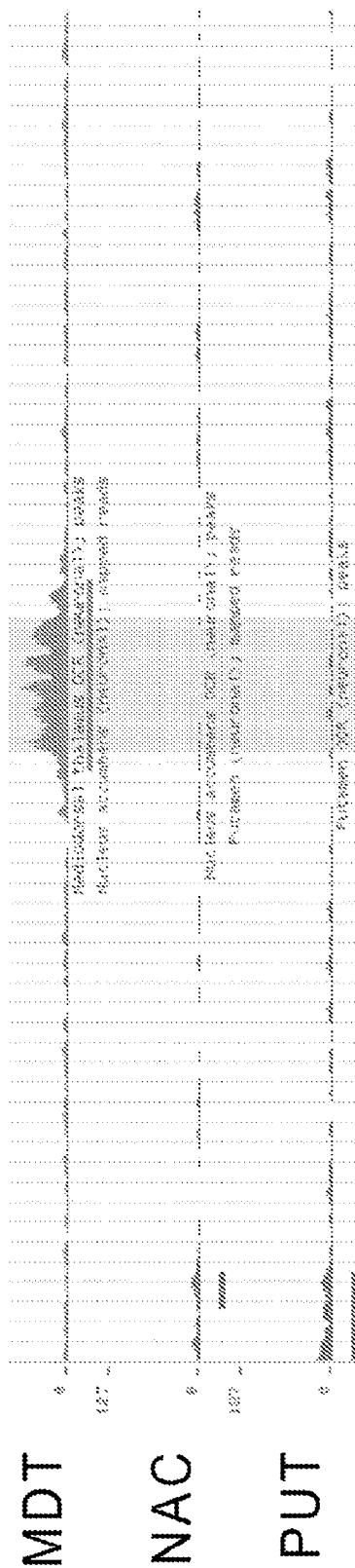


FIG. 1C cont'd



8/112

FIG. 1C cont'd



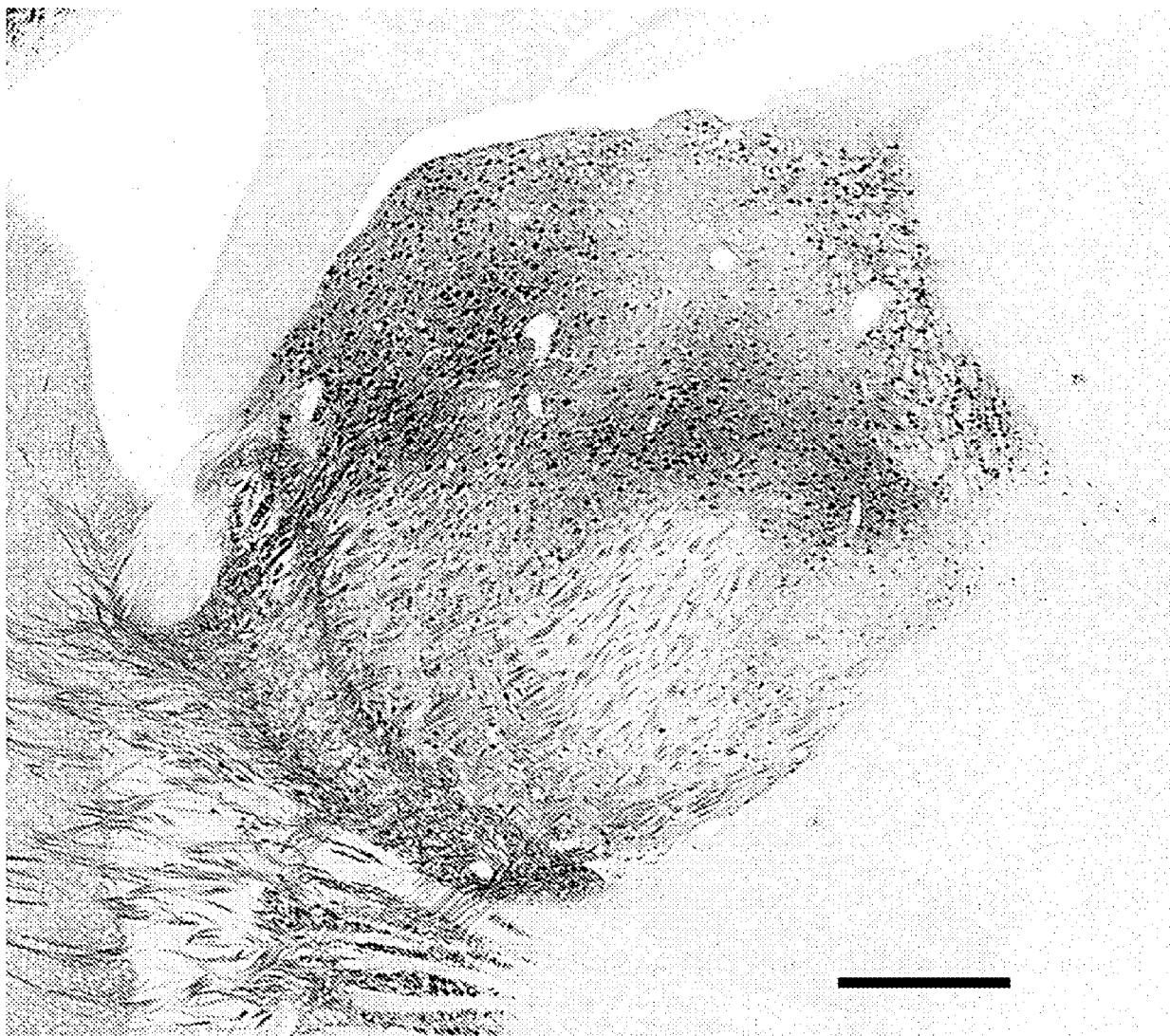
9/112

FIG. 2A



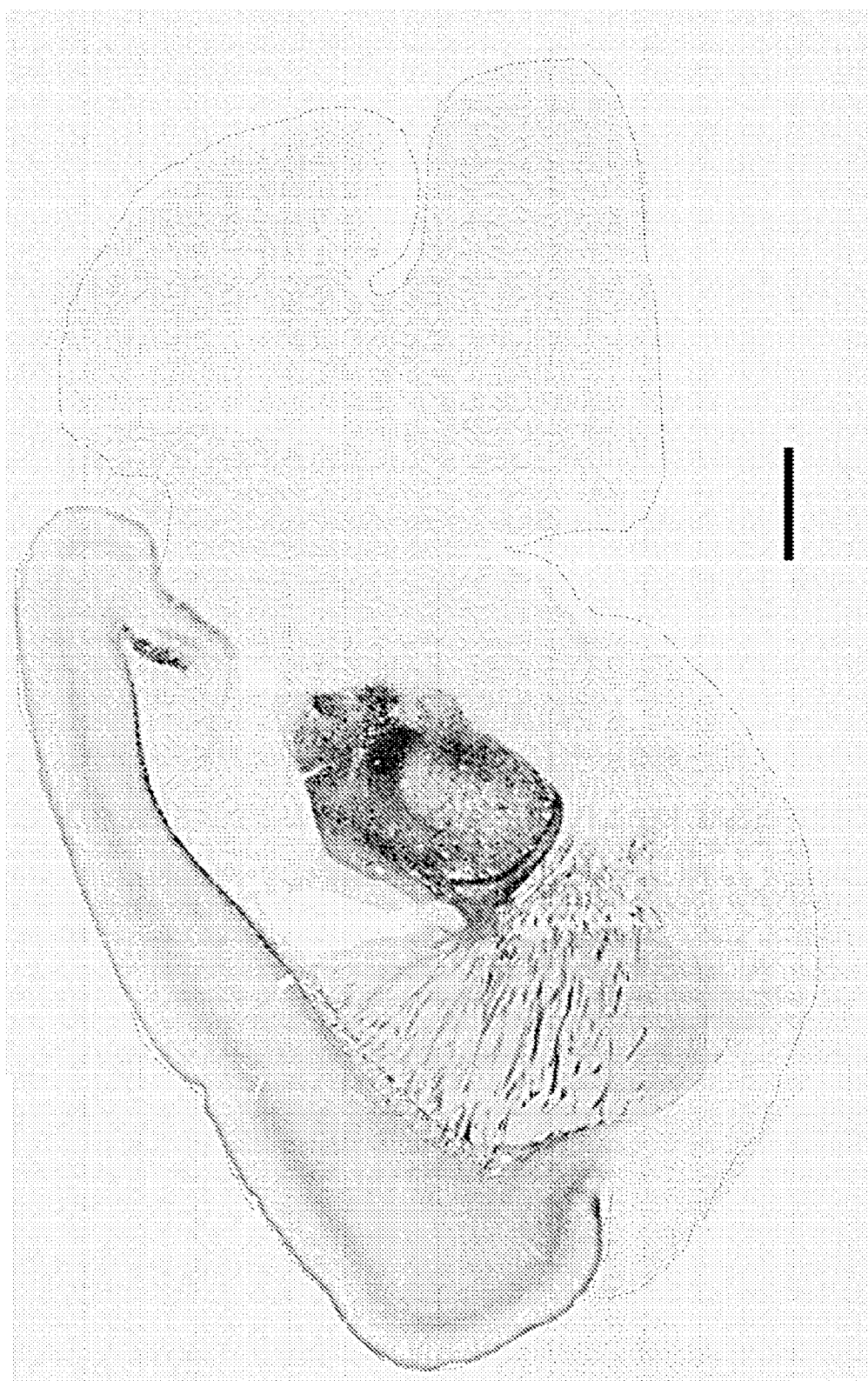
10/112

FIG. 2B



11/112

FIG. 3A



12/112

FIG. 3B

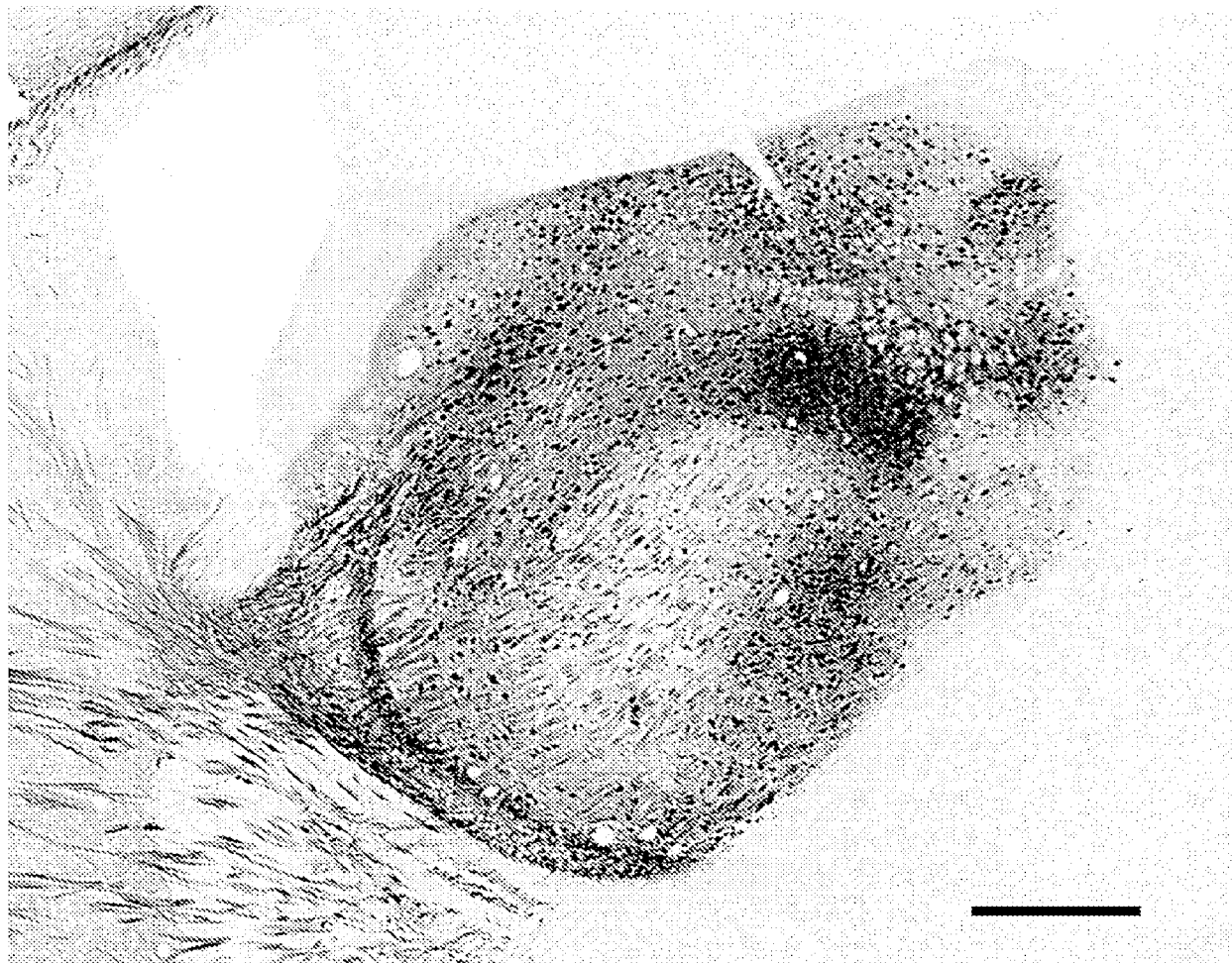


FIG. 4A



14/112

FIG. 4B

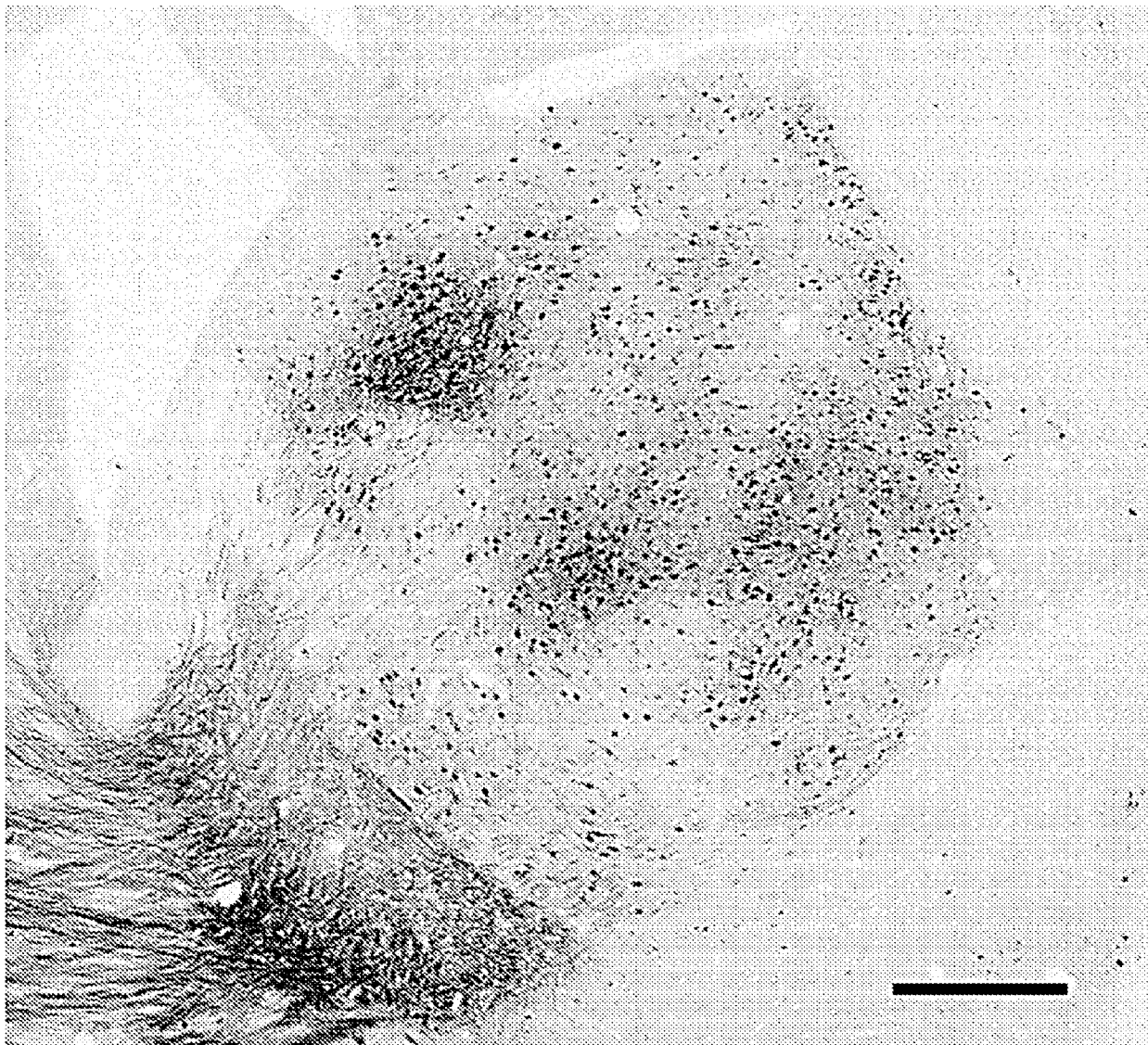
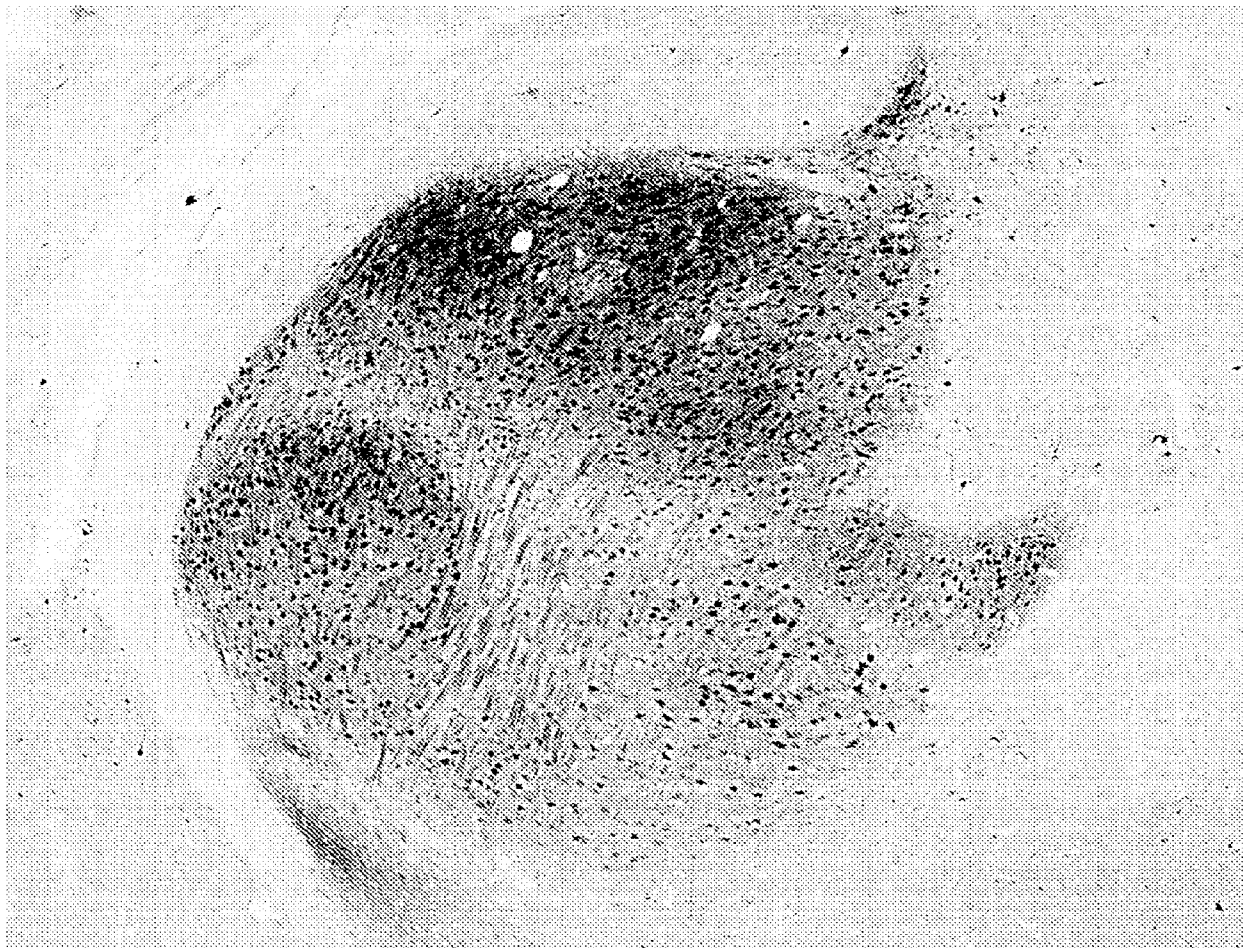


FIG. 5A



16/112

FIG. 5B



17/112

FIG. 6A

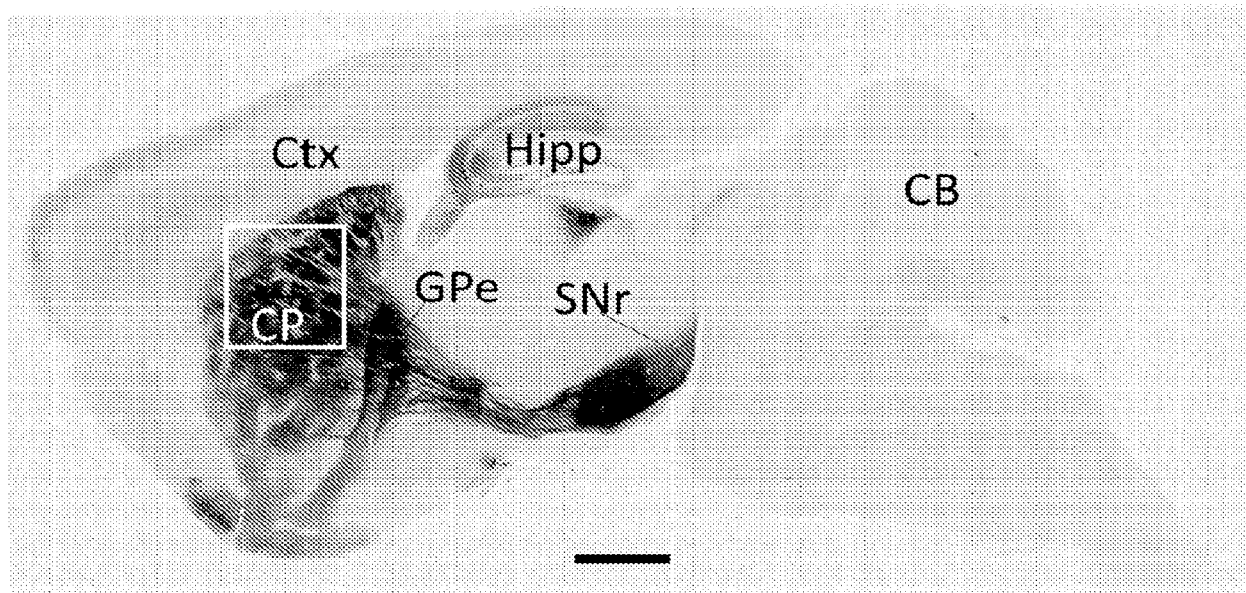
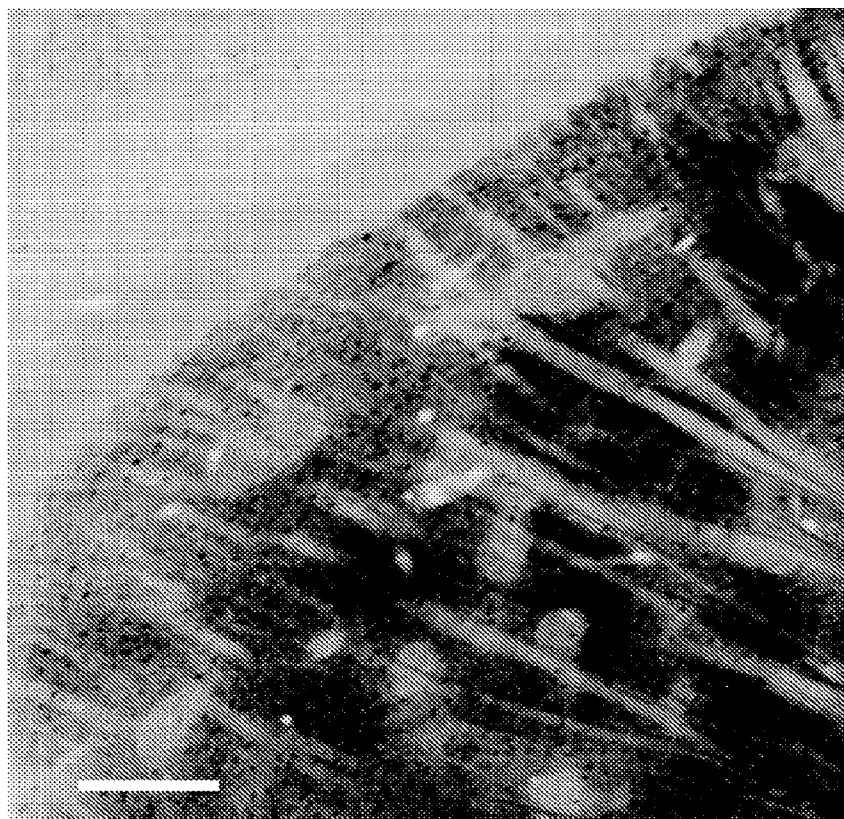


FIG. 6B



18/112

FIG. 7A

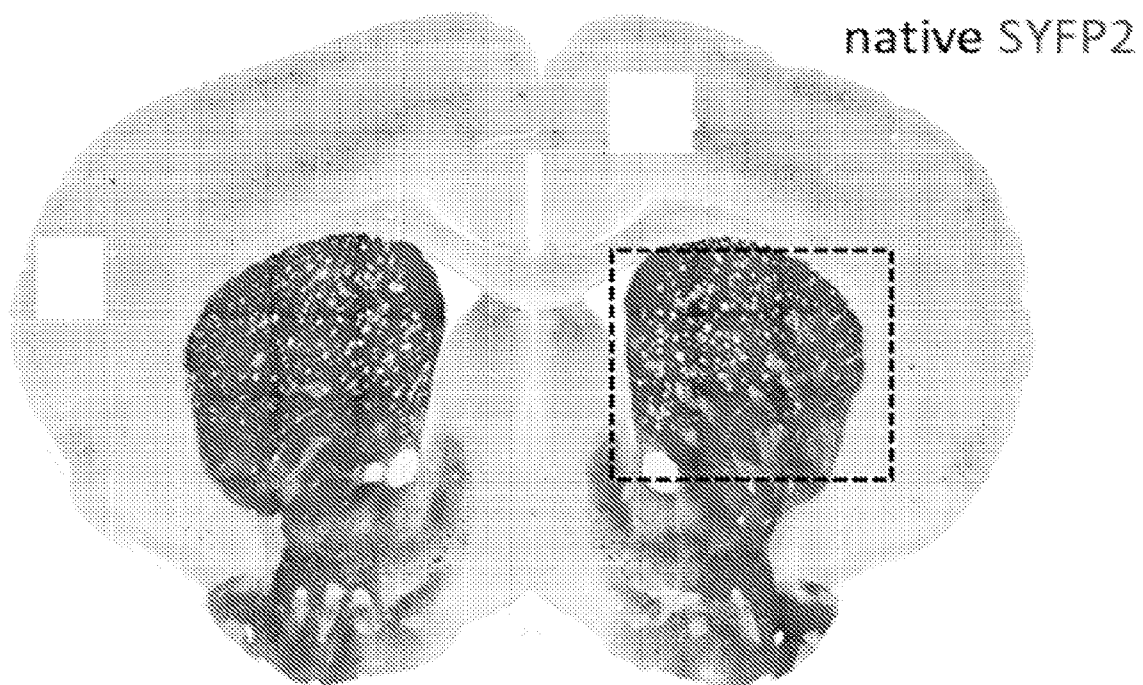
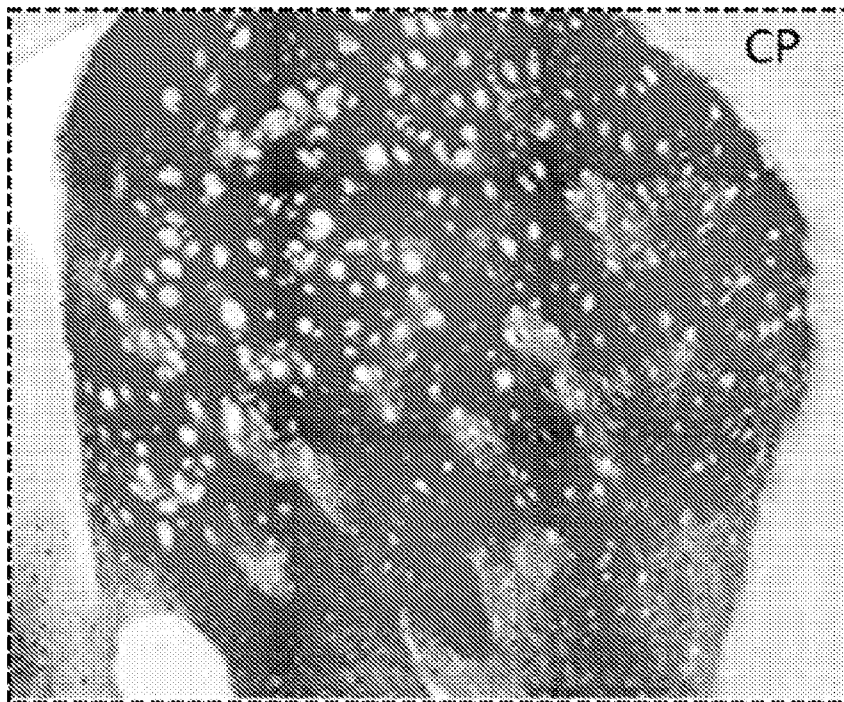


FIG. 7B



19/112

FIG. 7C

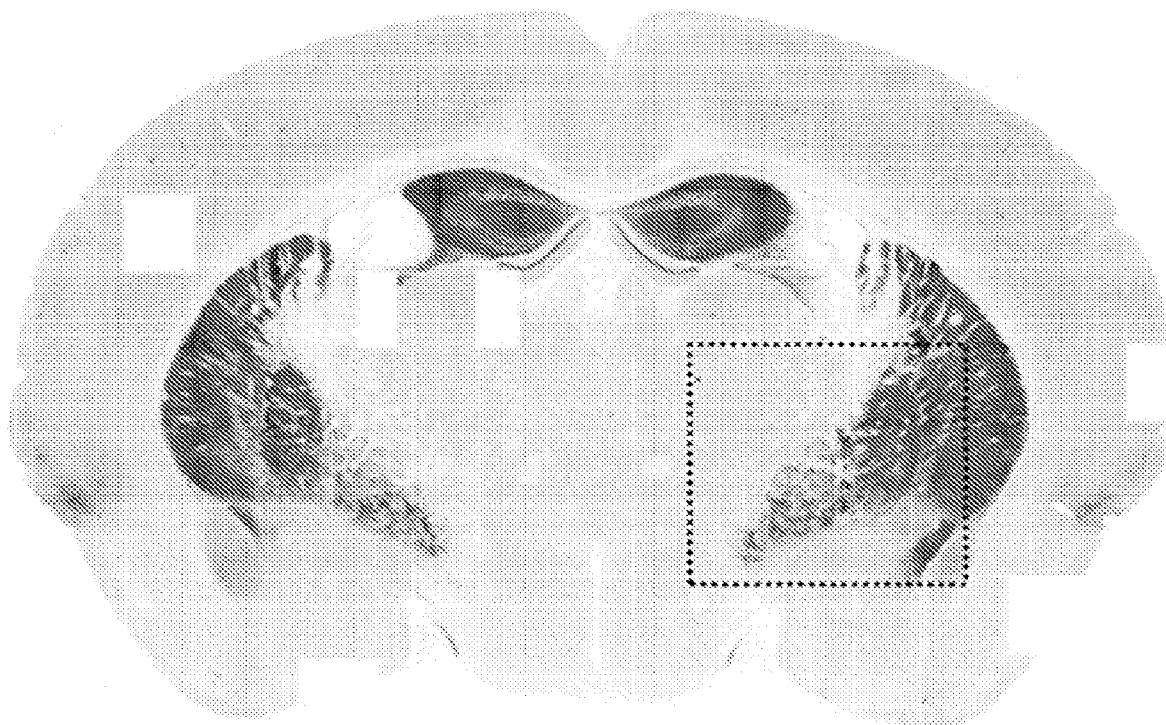
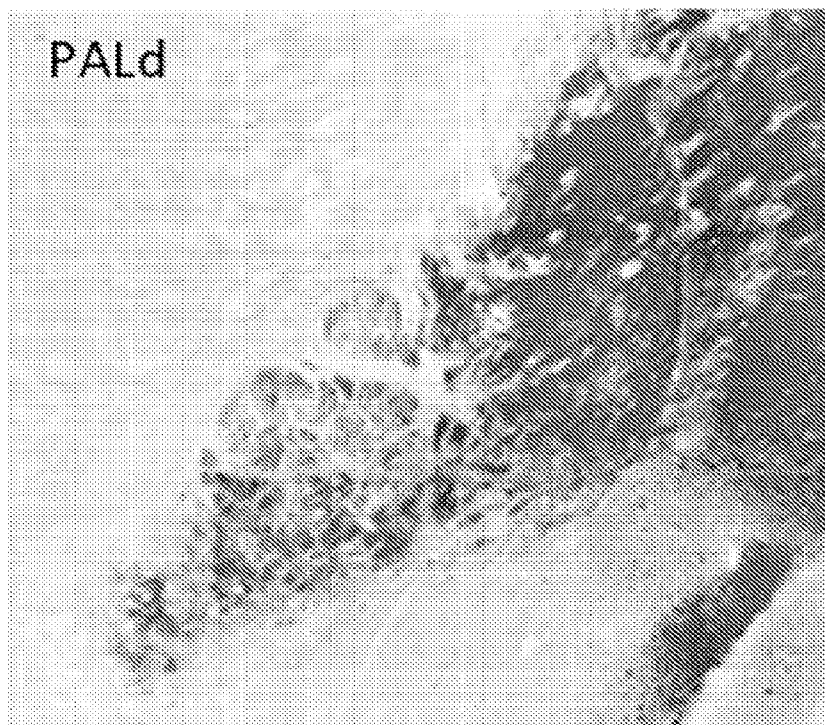


FIG. 7D



20/112

FIG. 7E

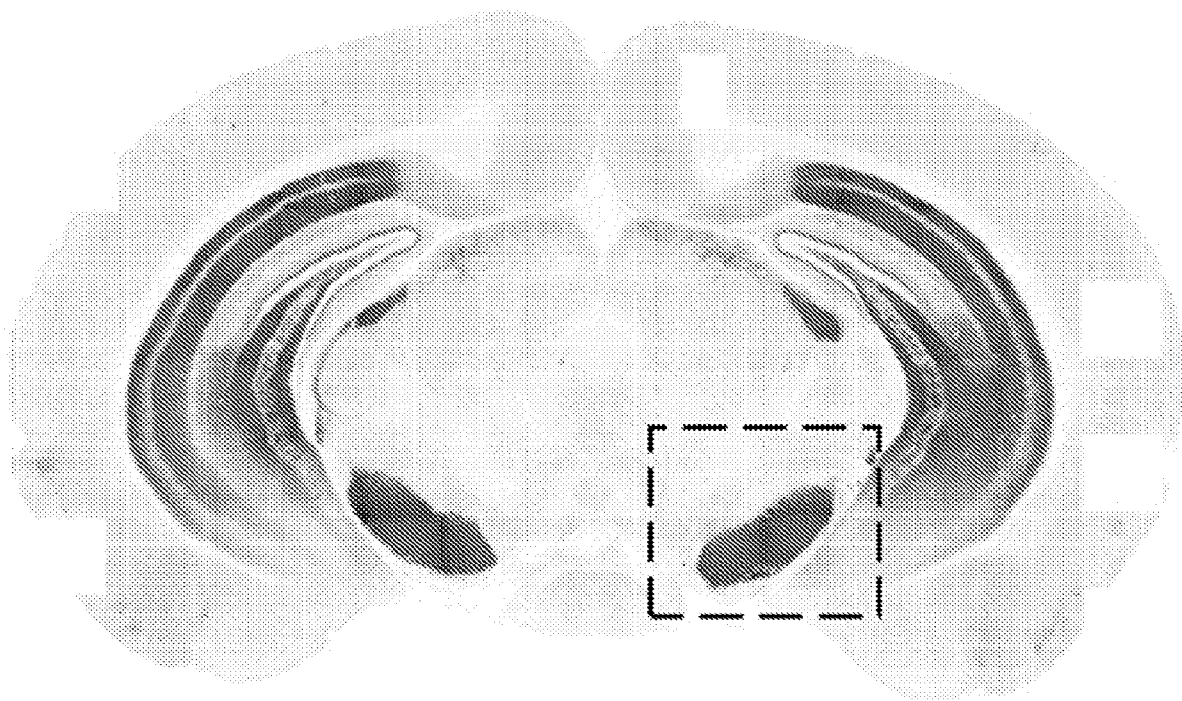
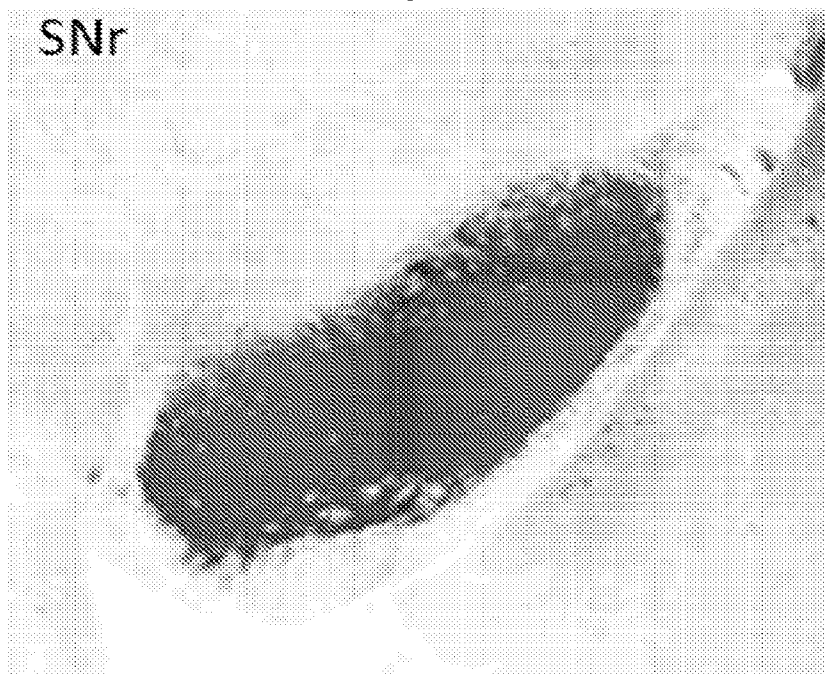


FIG. 7F



21/112

FIG. 7G

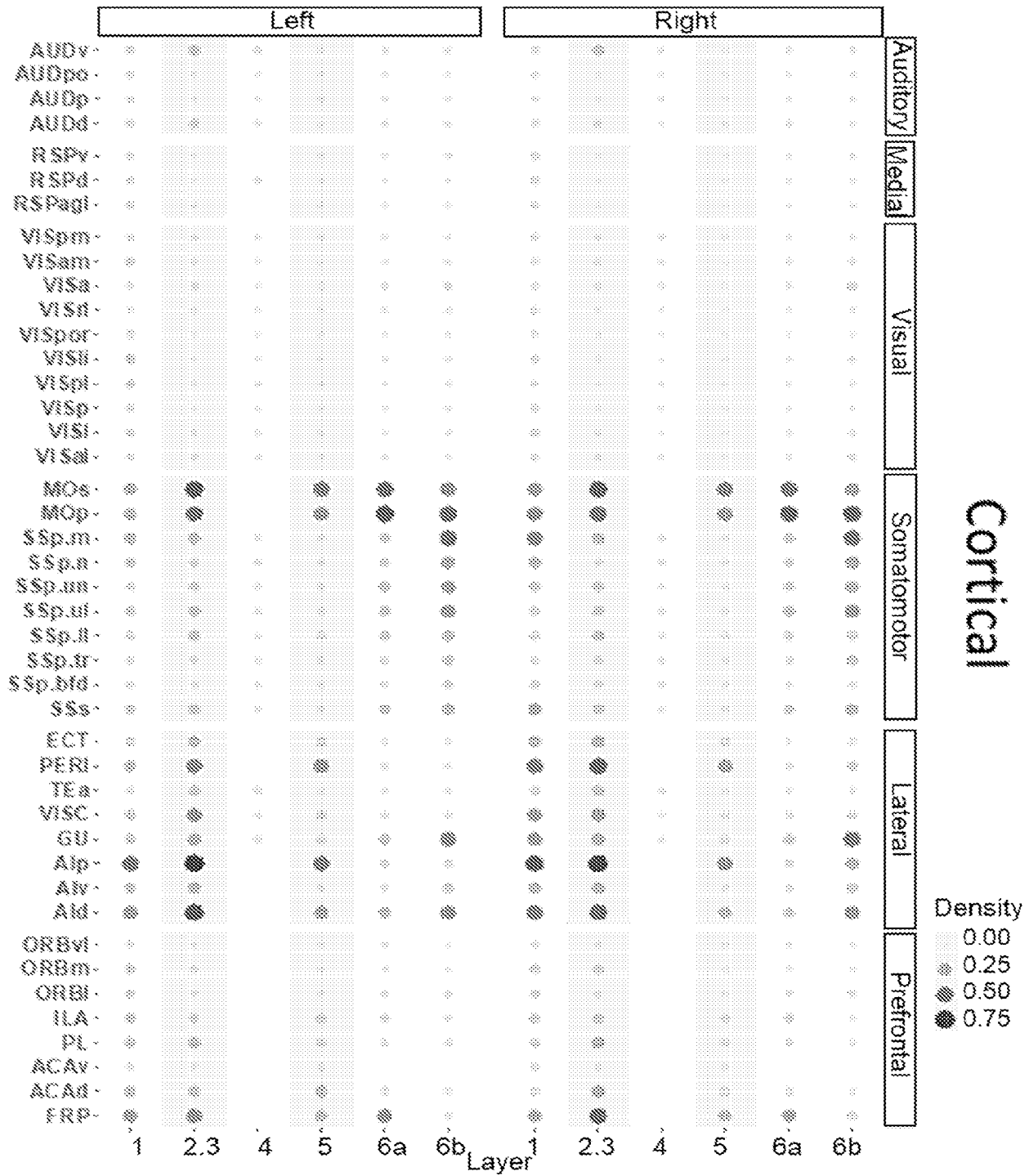
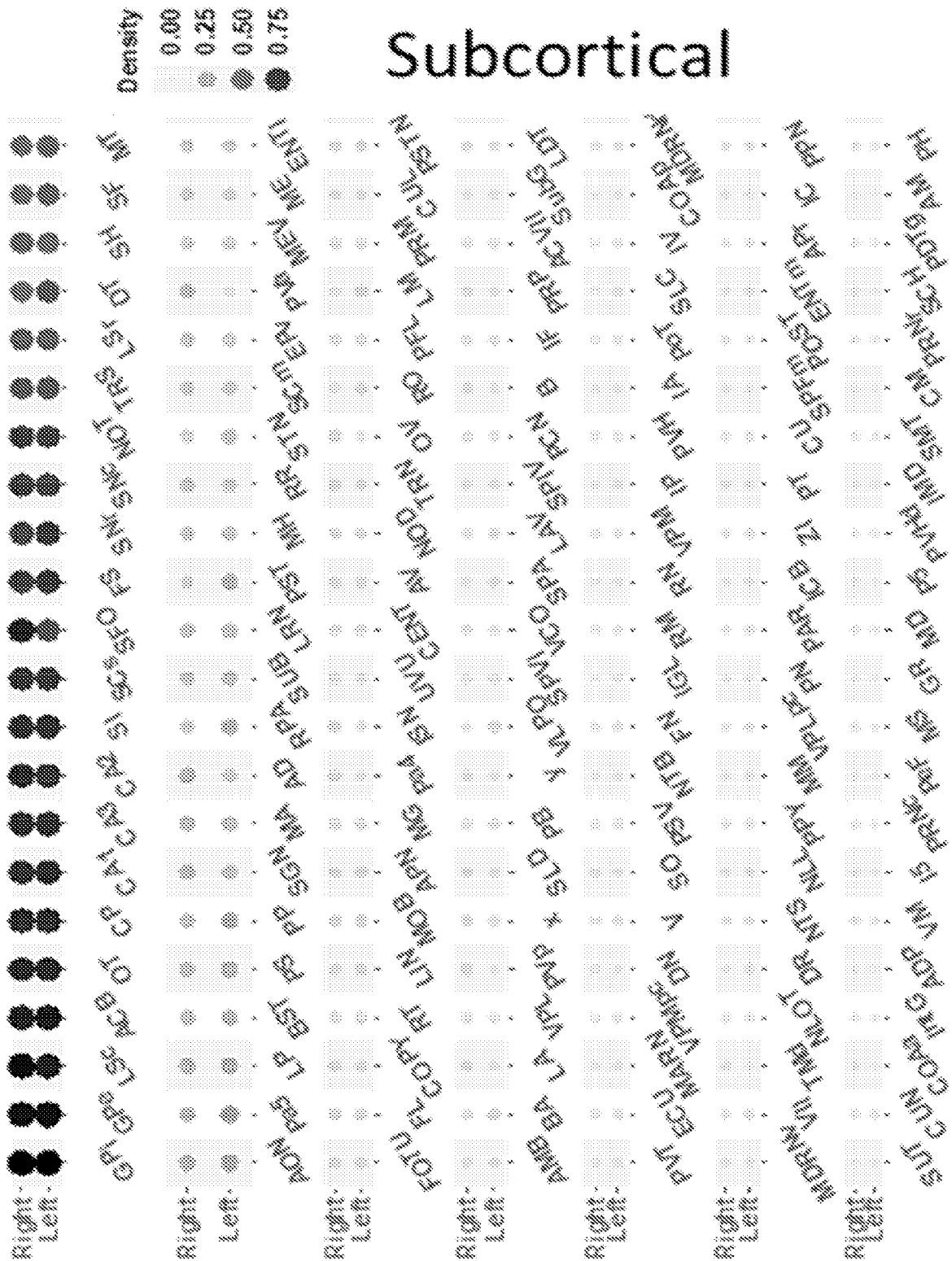
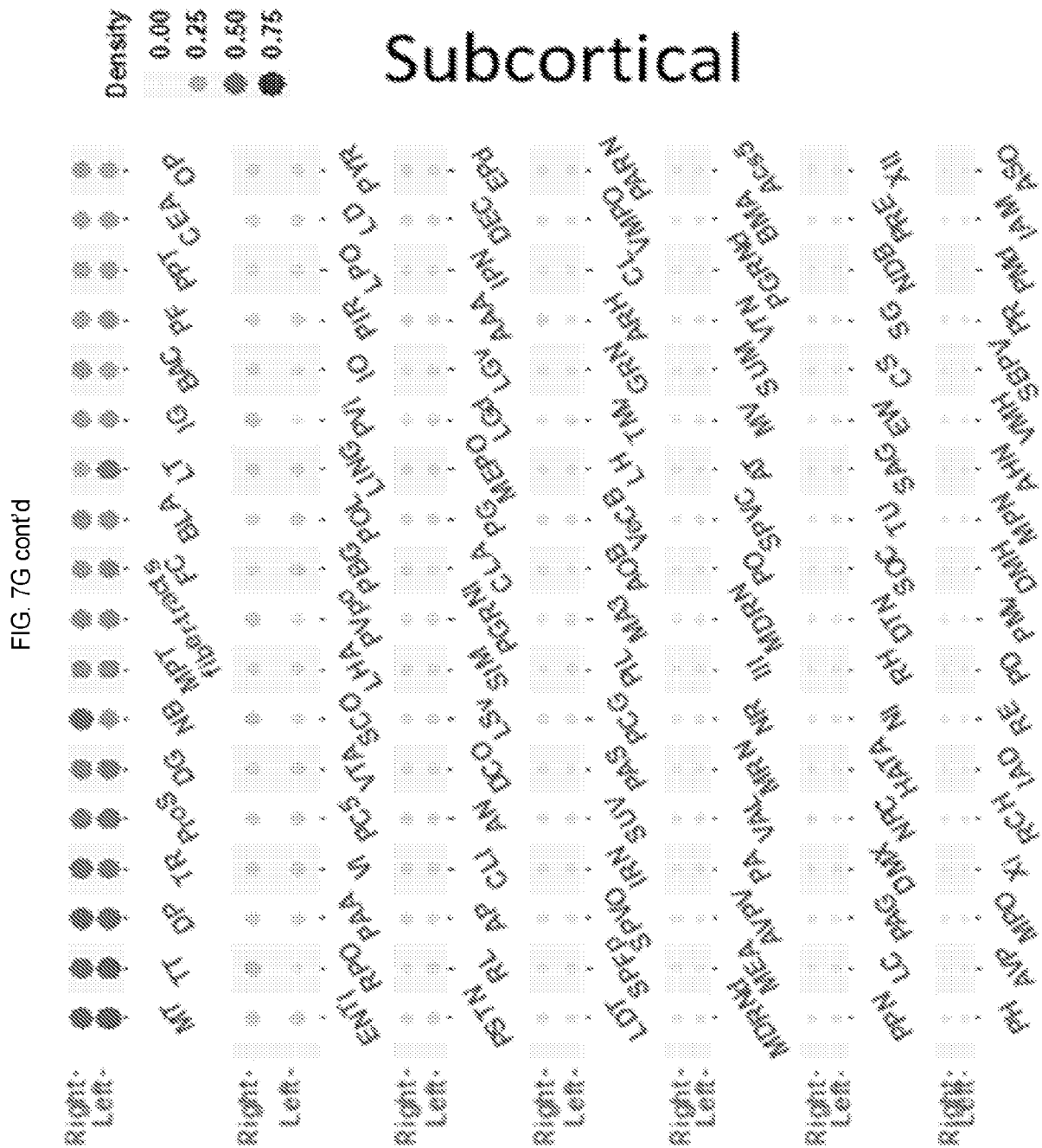


FIG. 7G cont'd





24/112

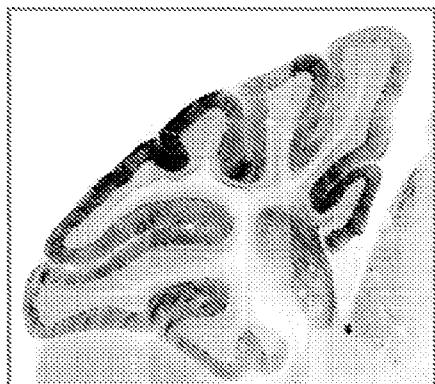
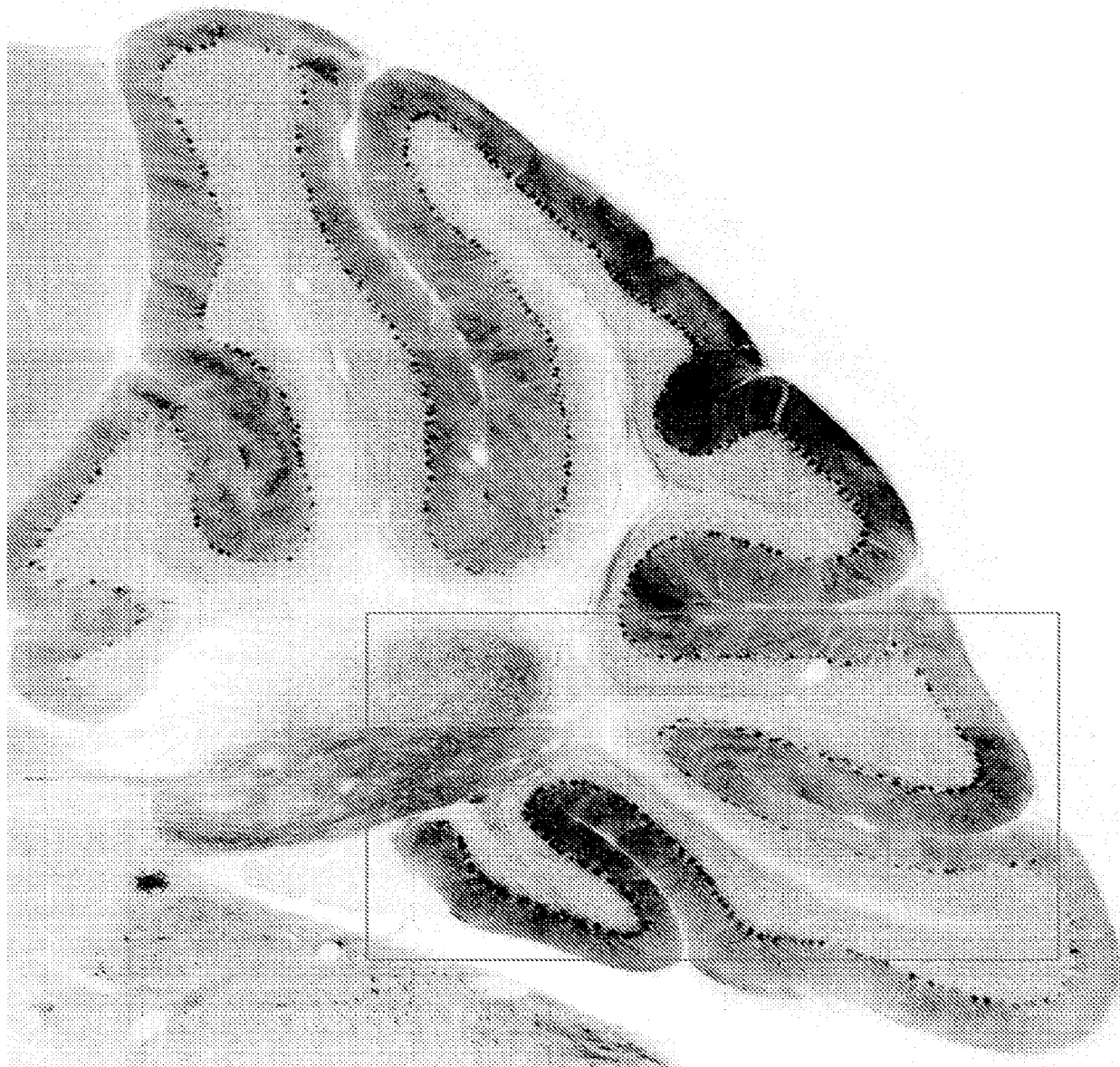


FIG. 8A

25/112

FIG. 8B

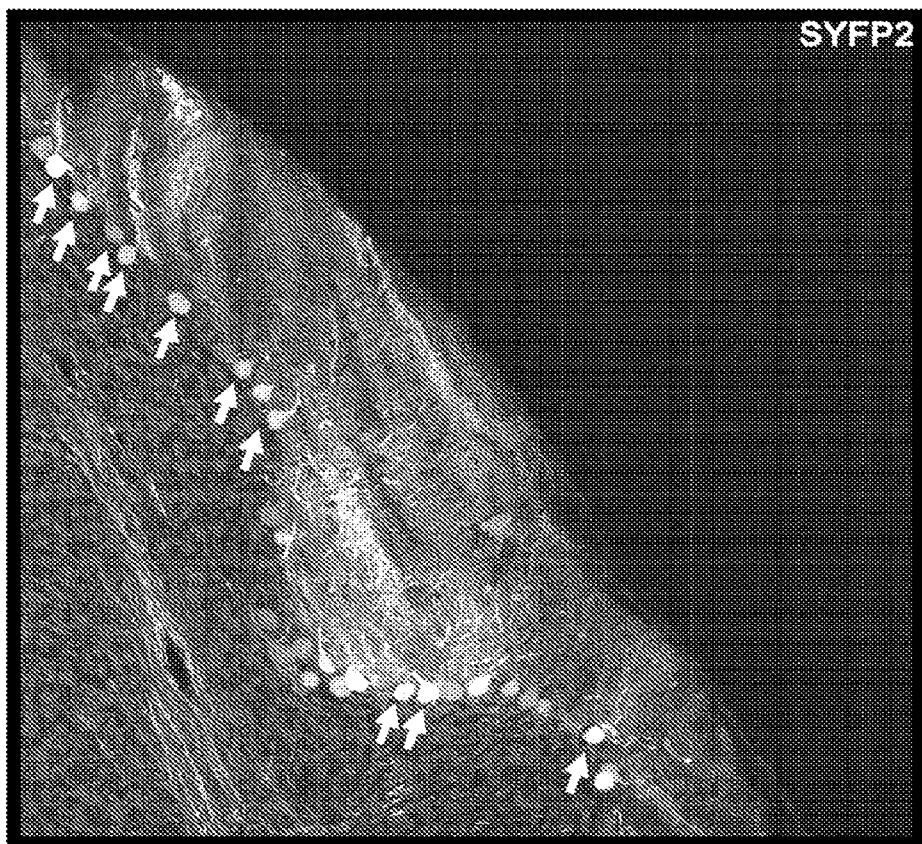


26/112

FIG. 8C



FIG. 8D



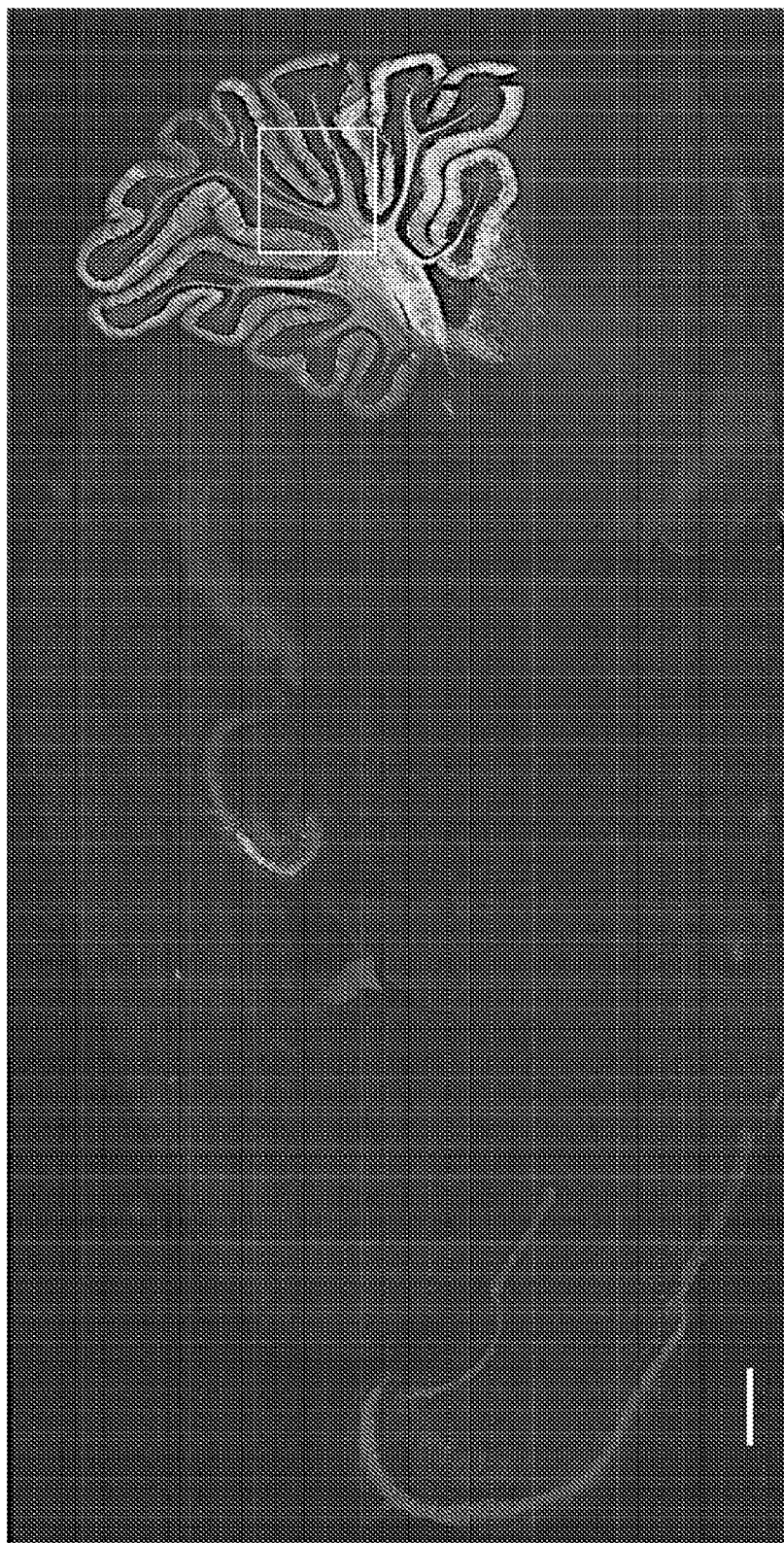
27/112

FIG. 8E



28/112

FIG. 8F



29/112

FIG. 8G

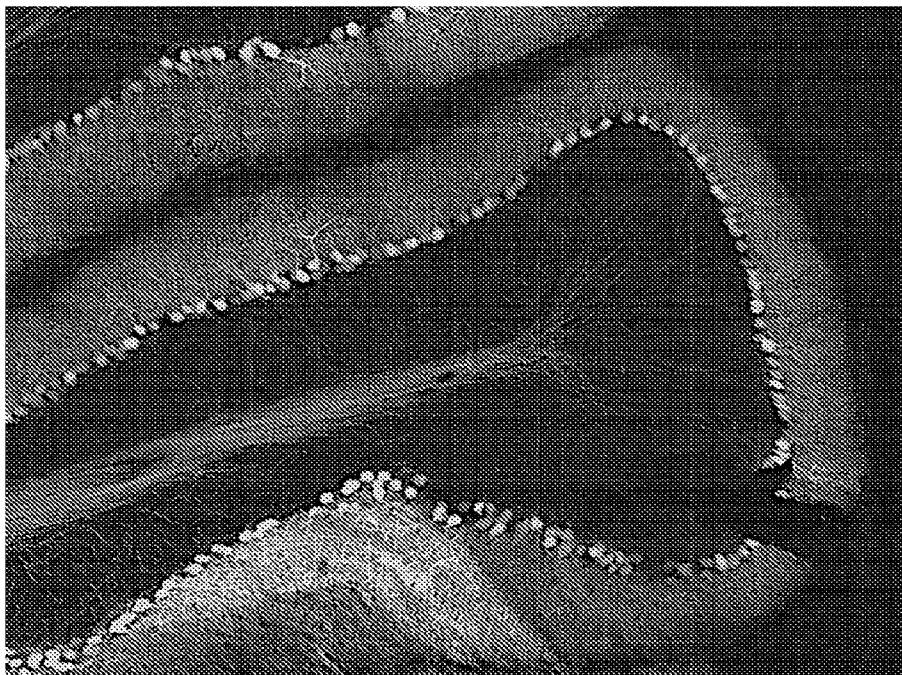
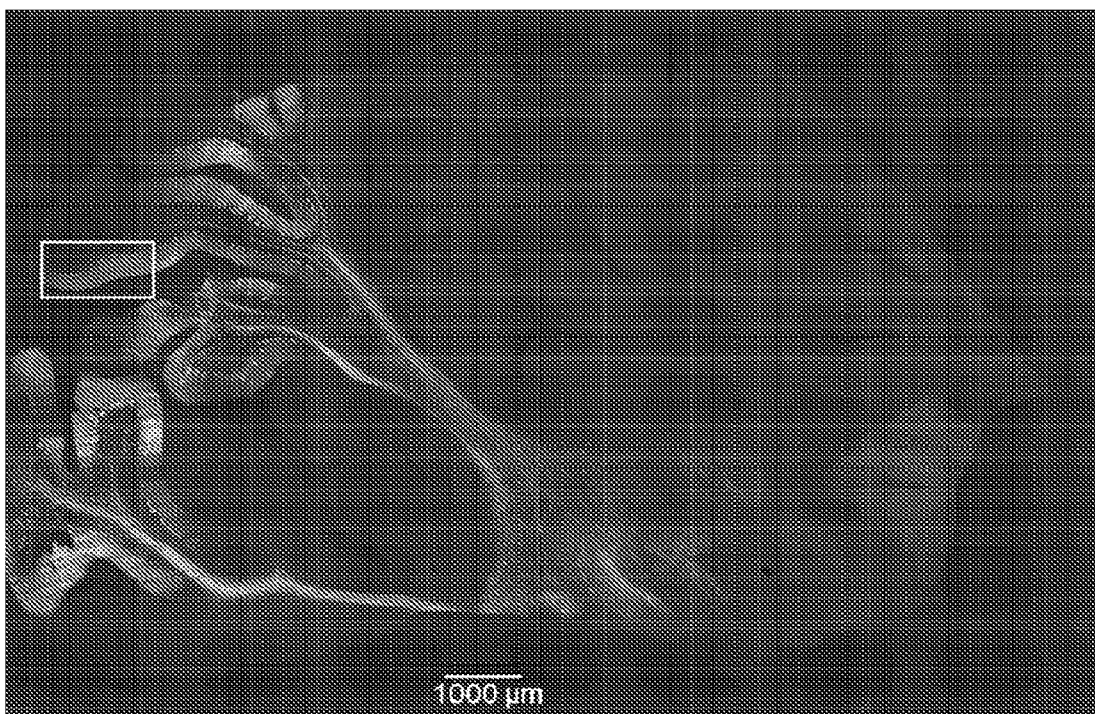
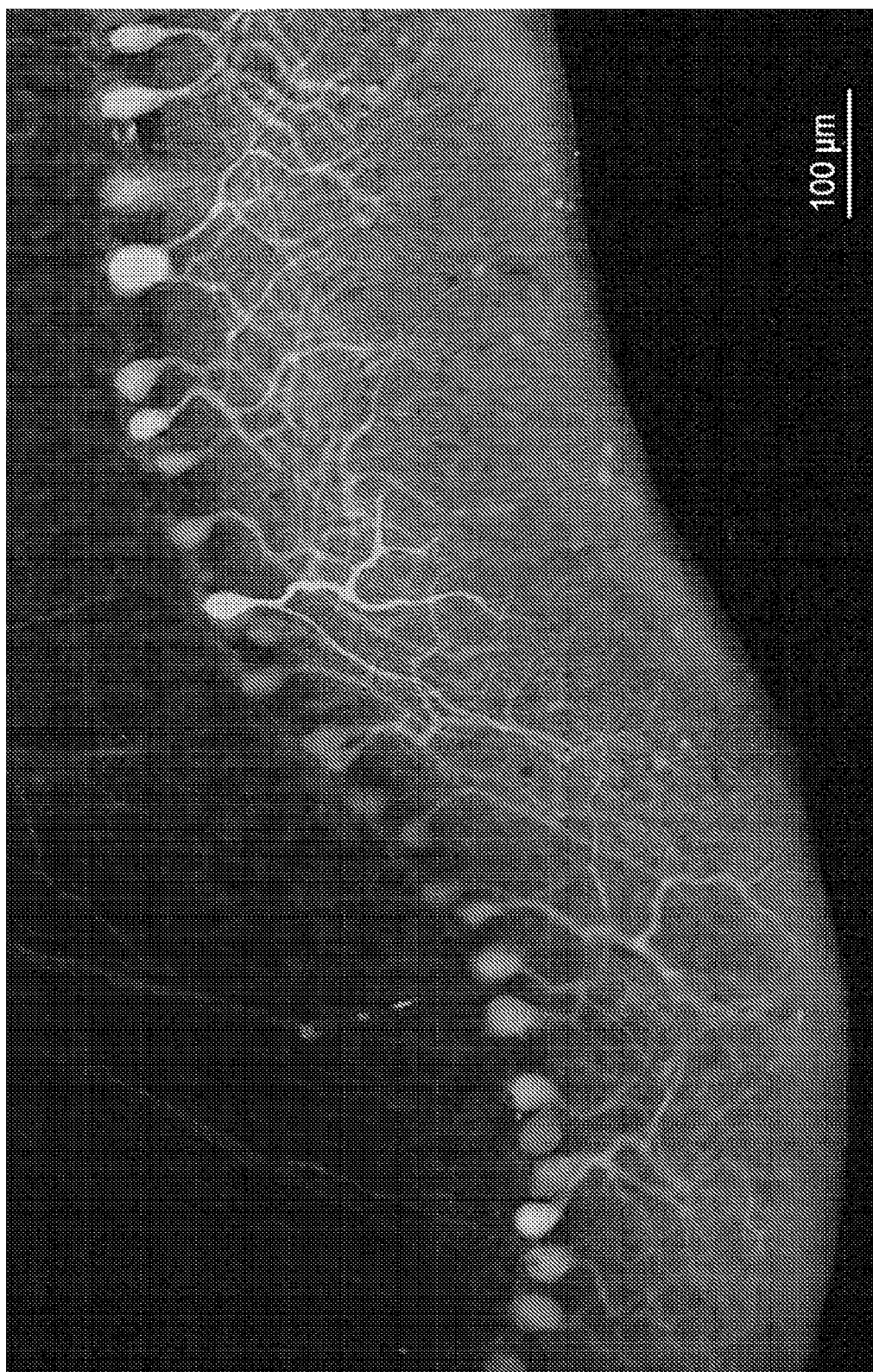


FIG. 8H



30/112

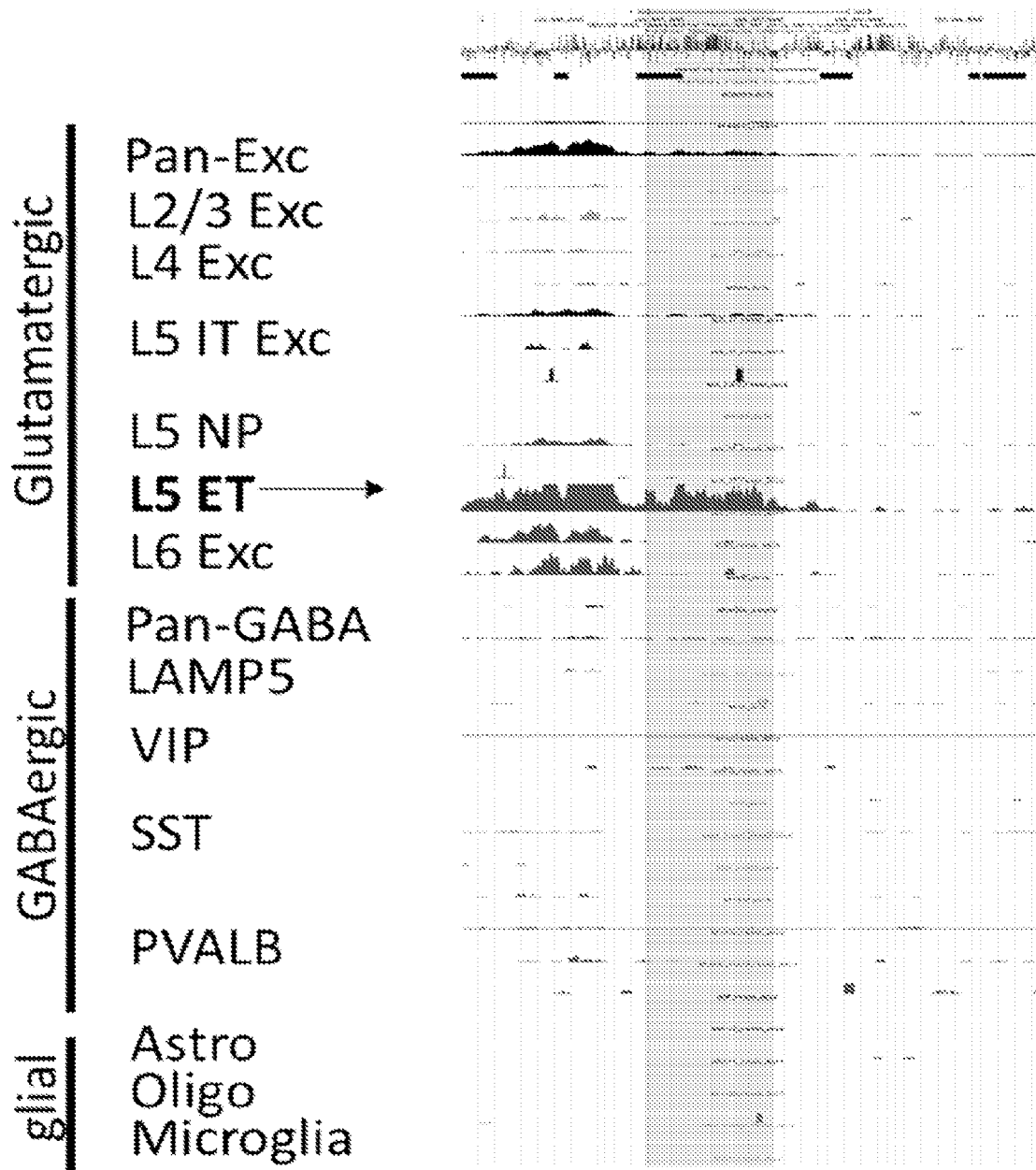
FIG. 8I



31/112

FIG. 9A

eHGT_453m, chr4:123,946,774-123,947,495



32/112

FIG. 9B

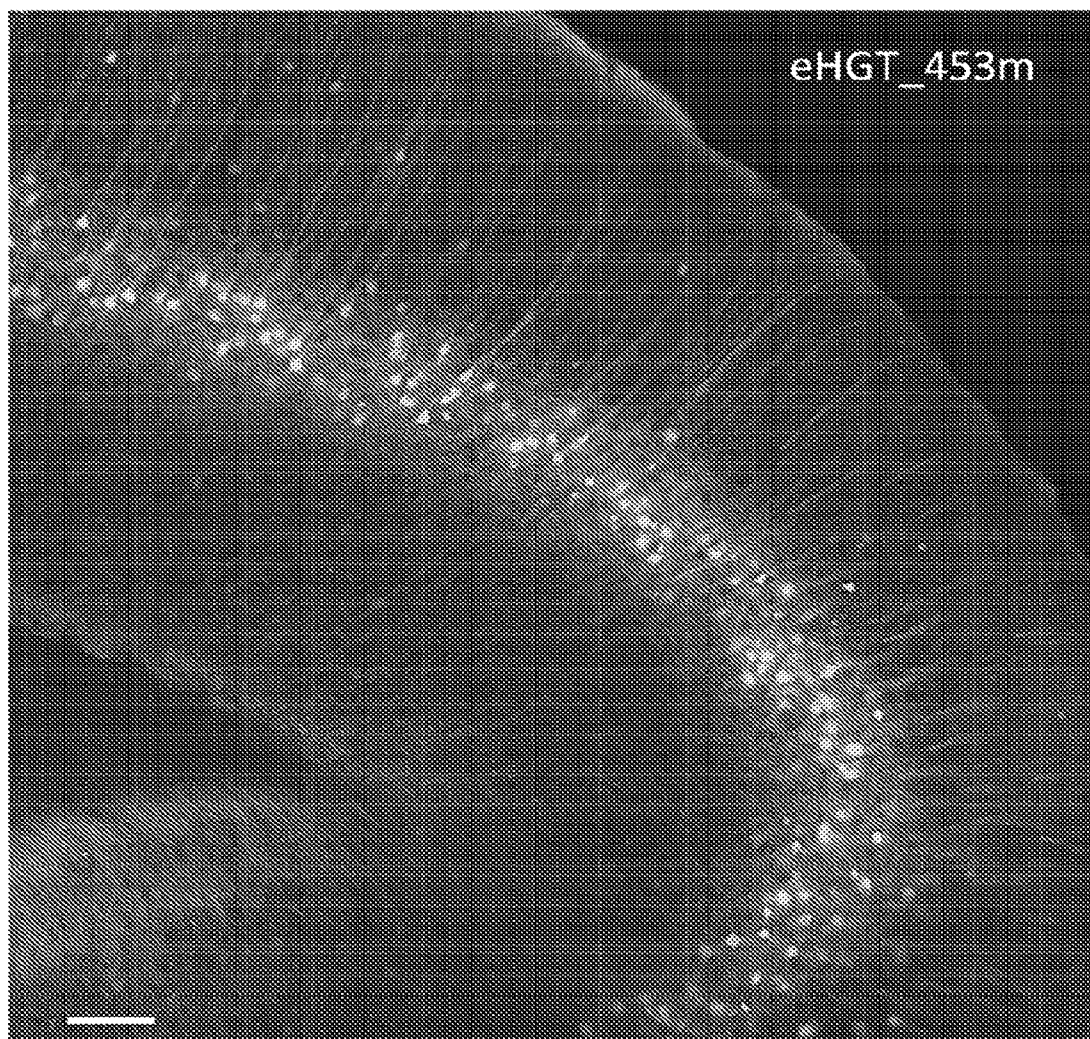
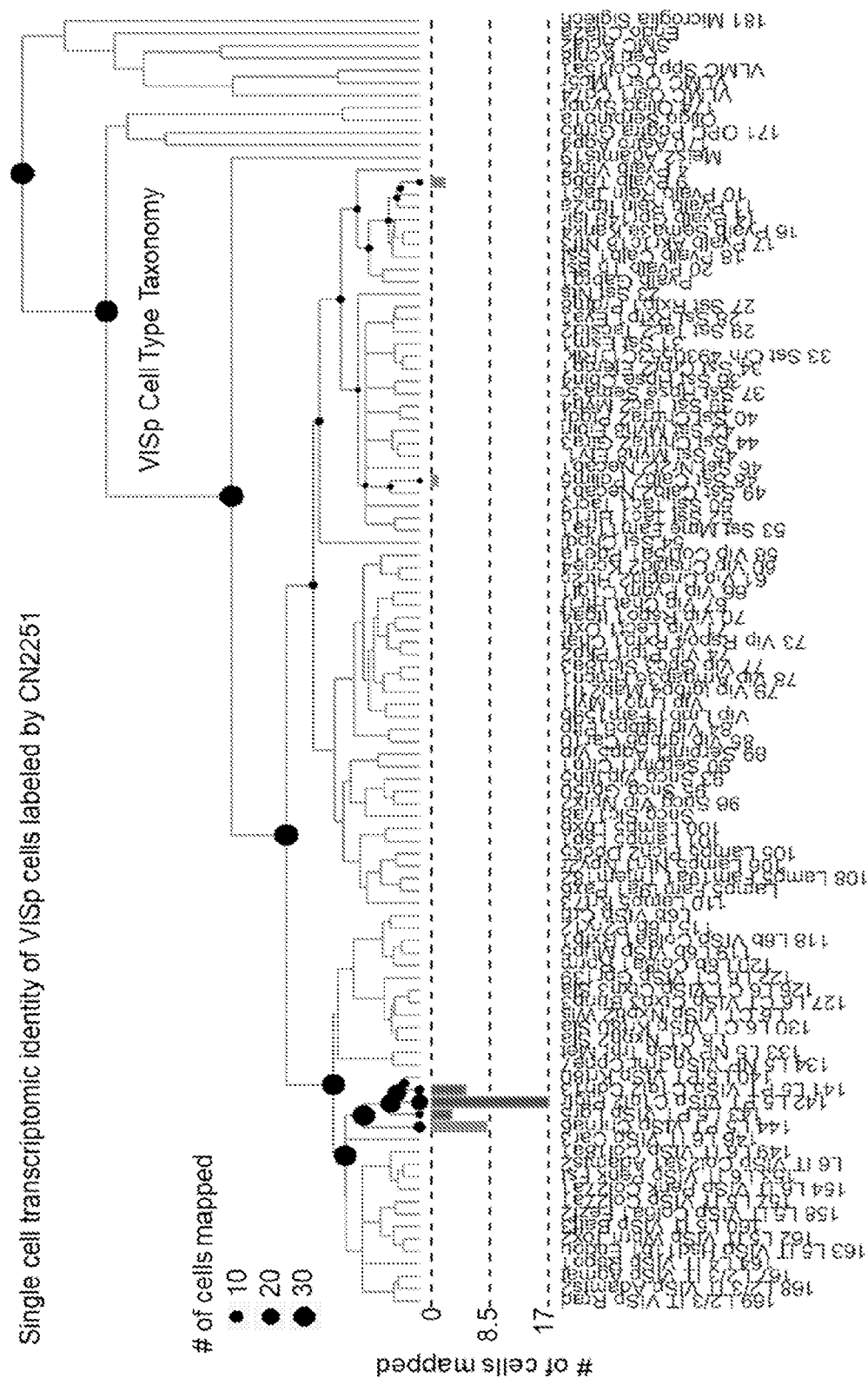


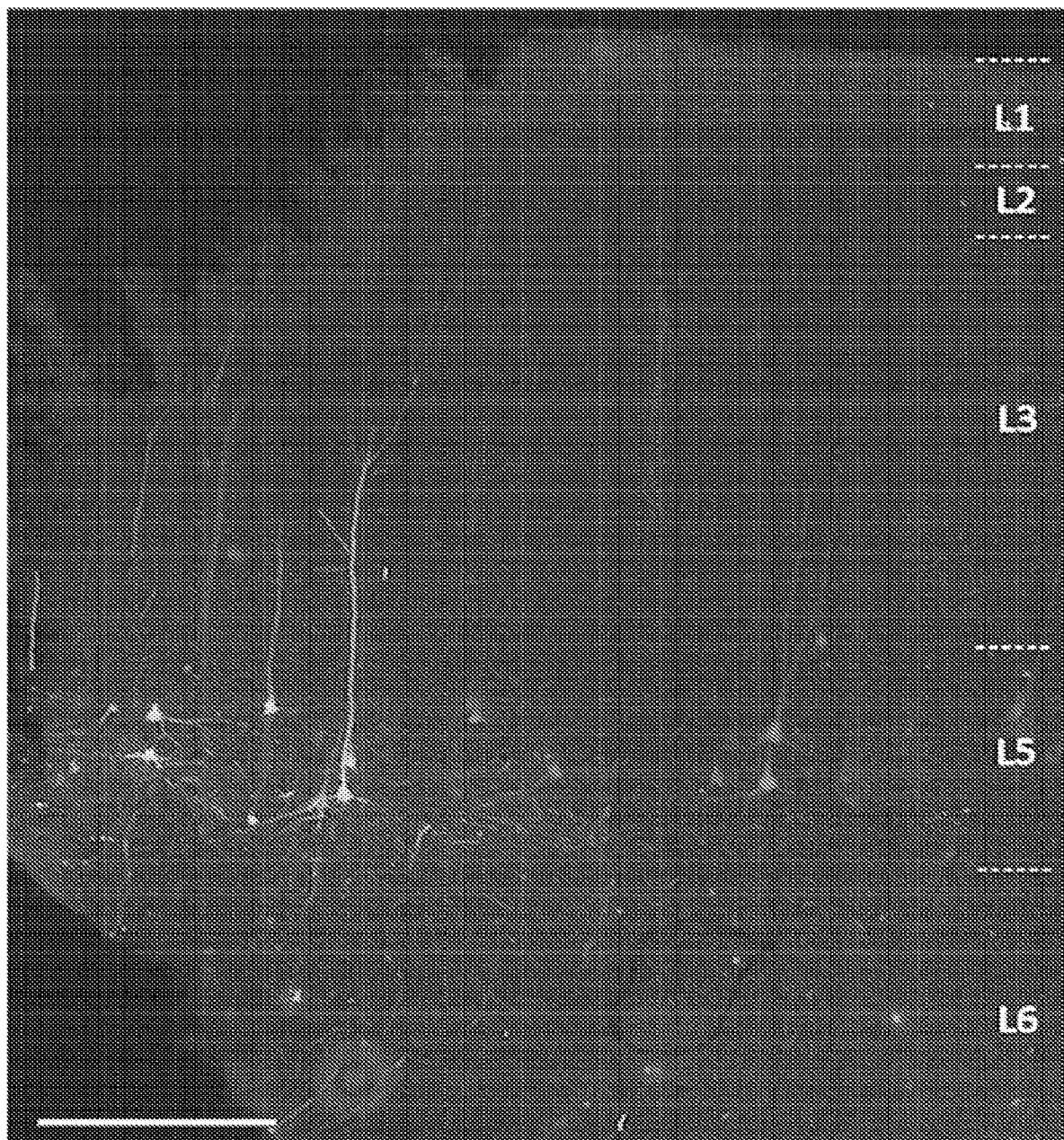
FIG. 9C

Single cell transcriptomic identity of VSp cells labeled by CN2251



34/112

FIG. 9D



35/112

FIG. 10A

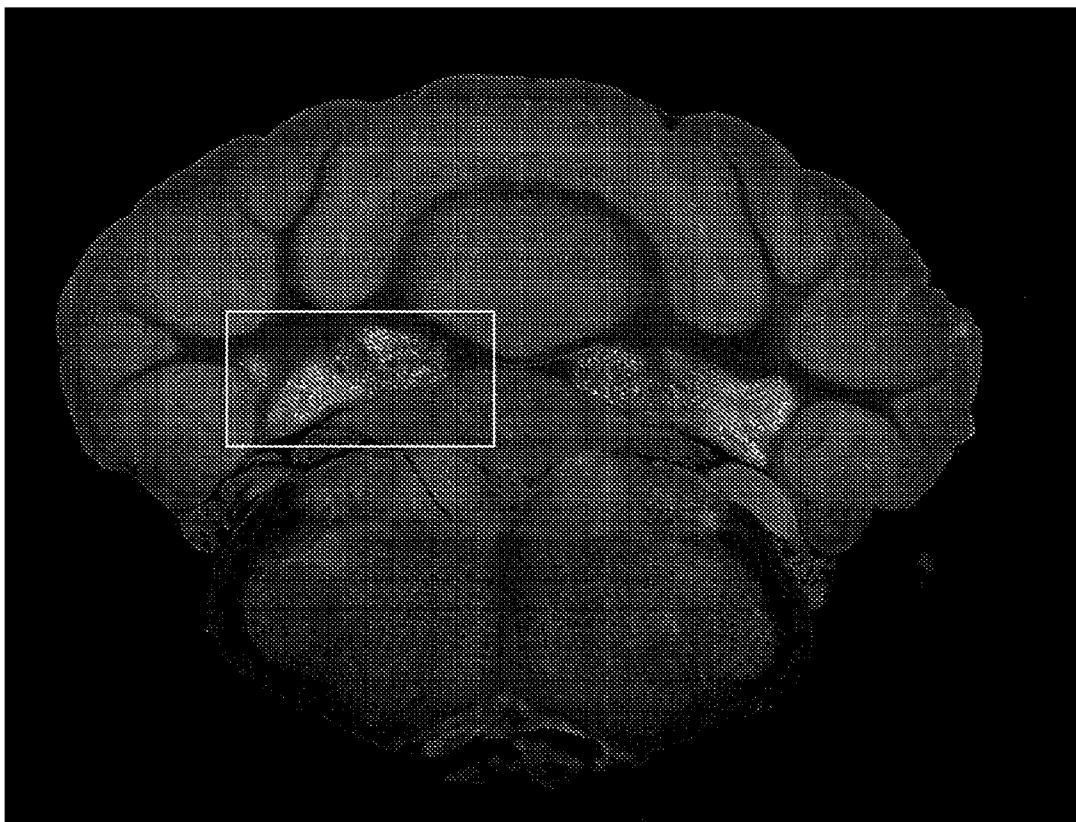
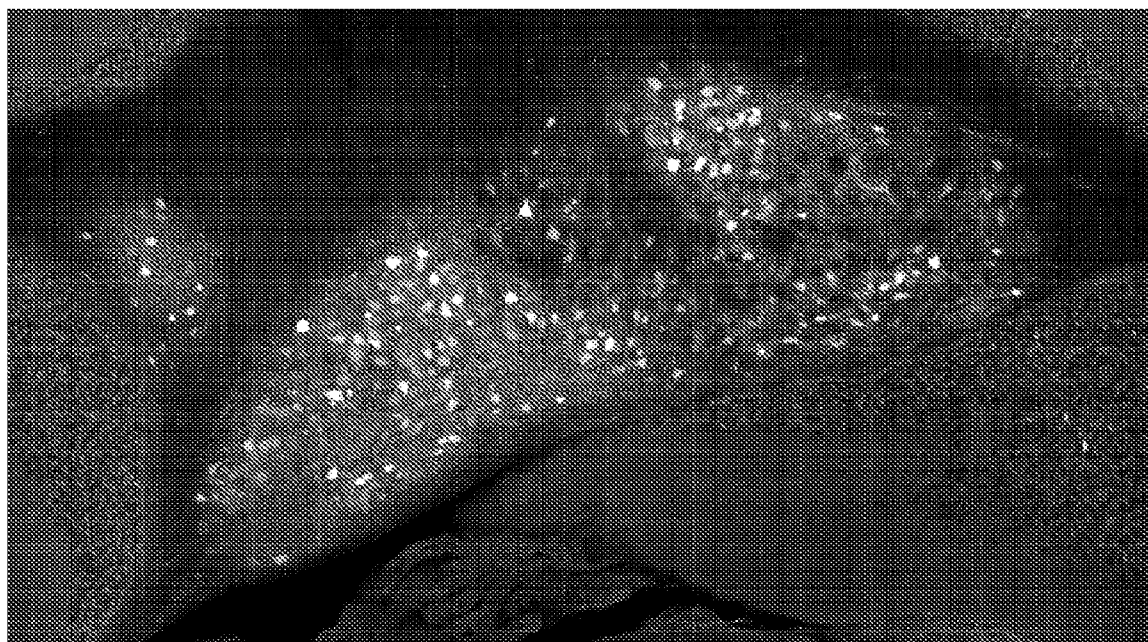


FIG. 10B



36/112

FIG. 11A



37/112

FIG. 11B

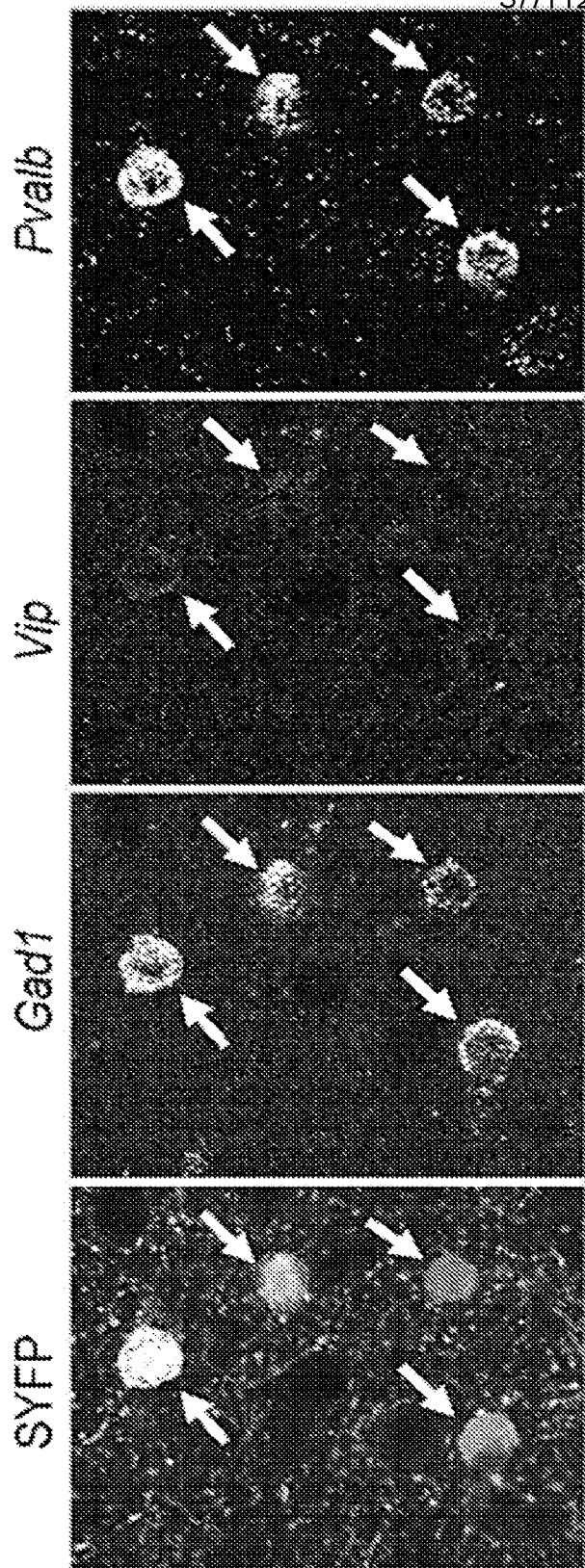
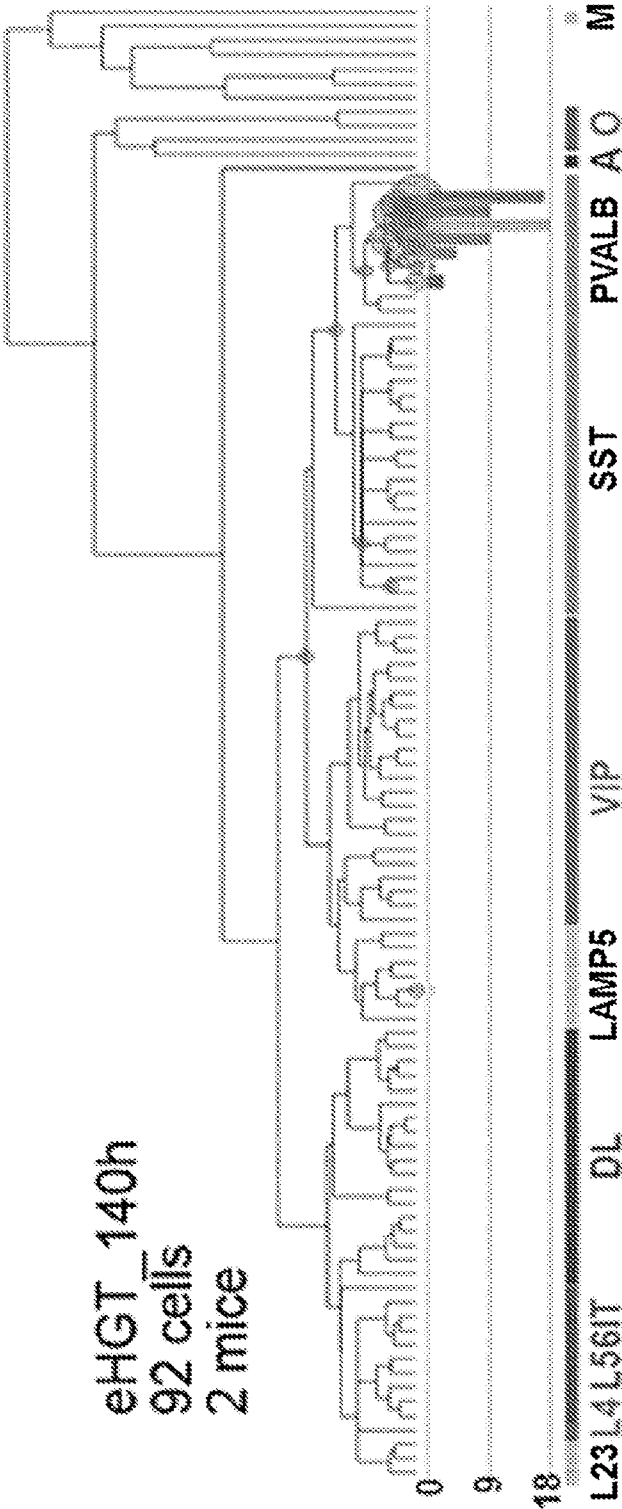


FIG. 11C



39/112

FIG. 12A



40/112

FIG. 12B

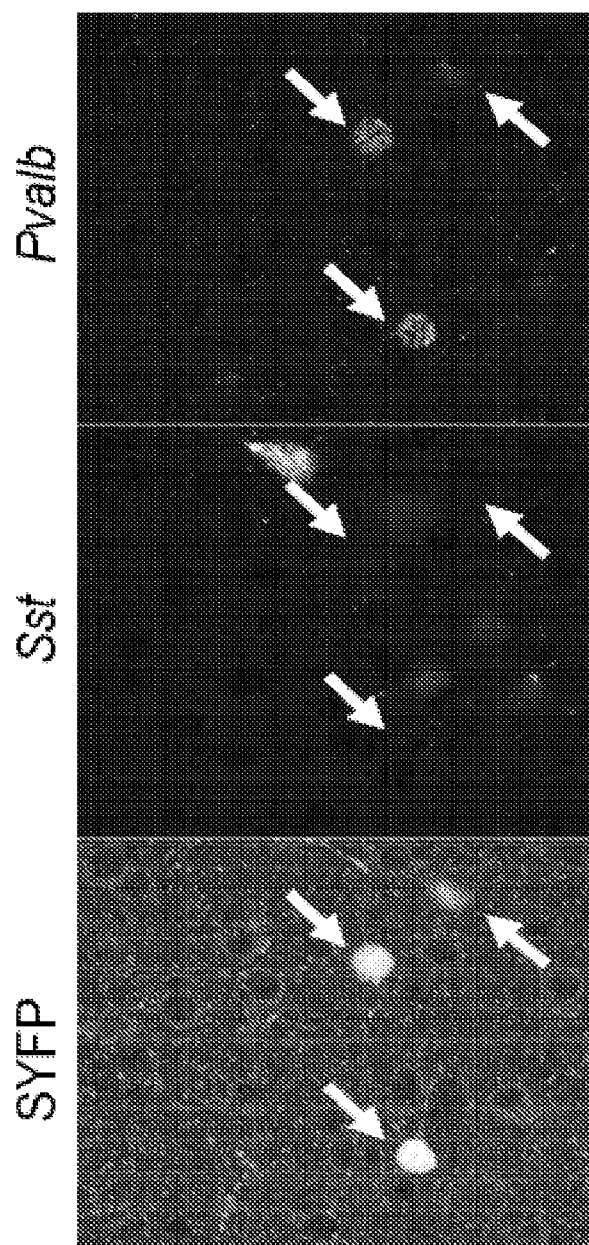
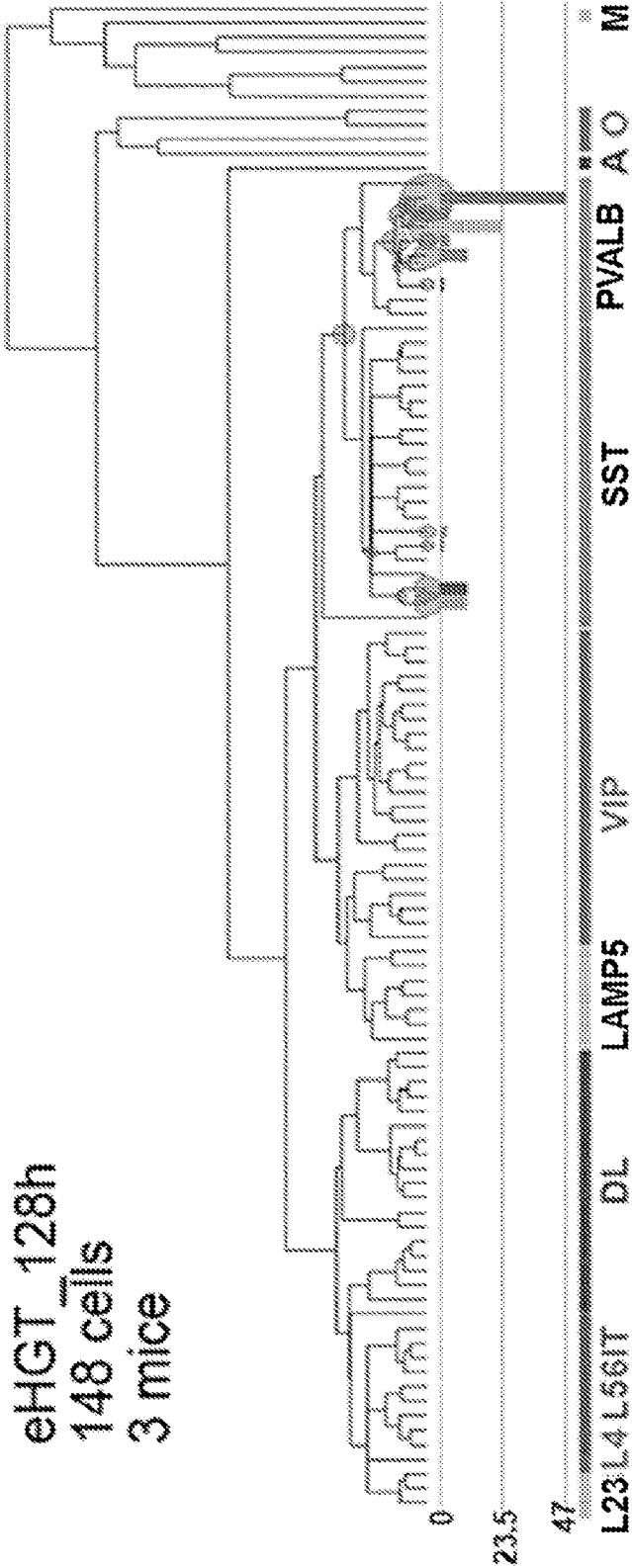
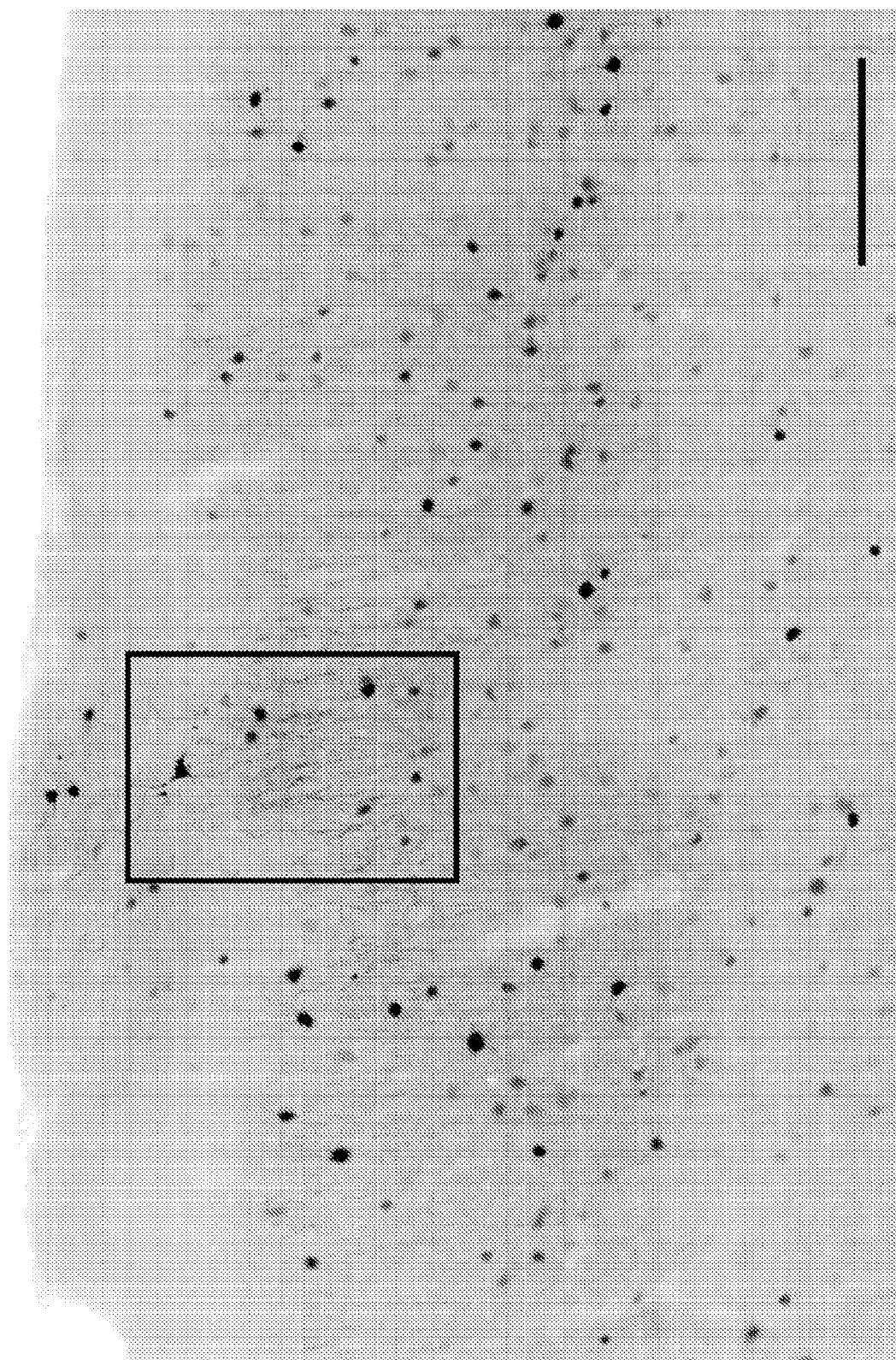


FIG. 12C



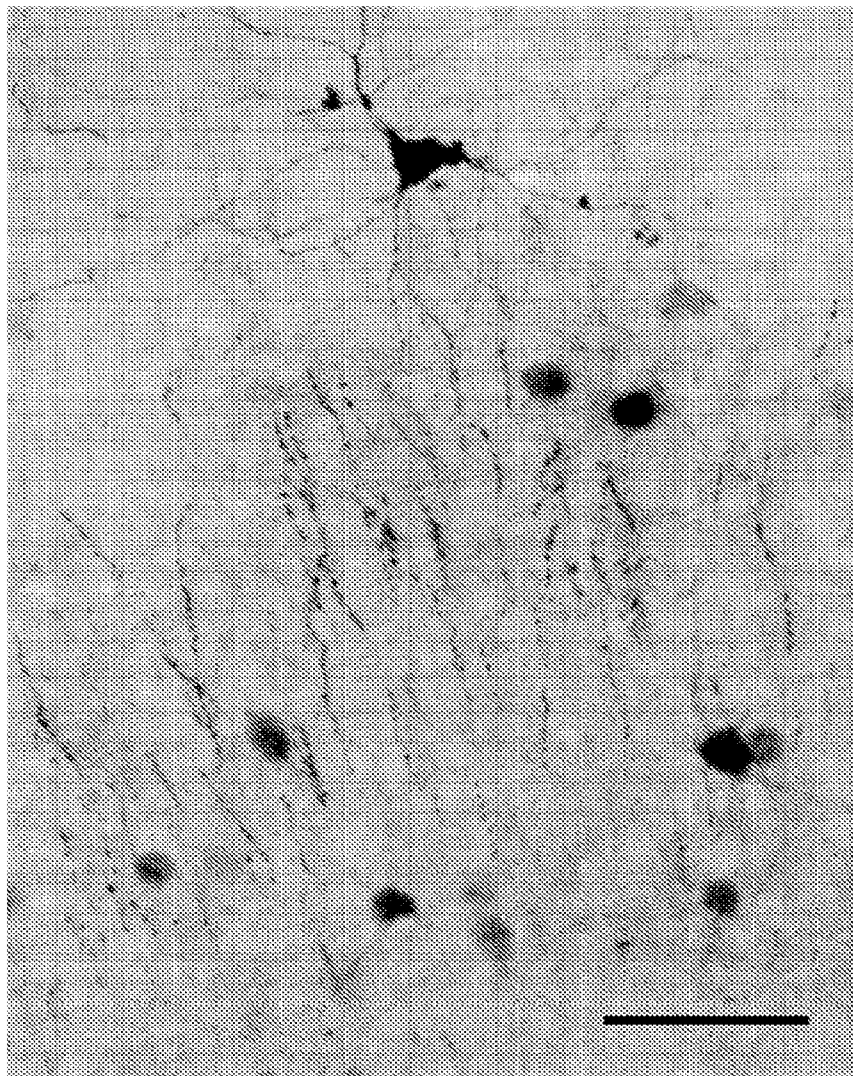
42/112

FIG. 13A



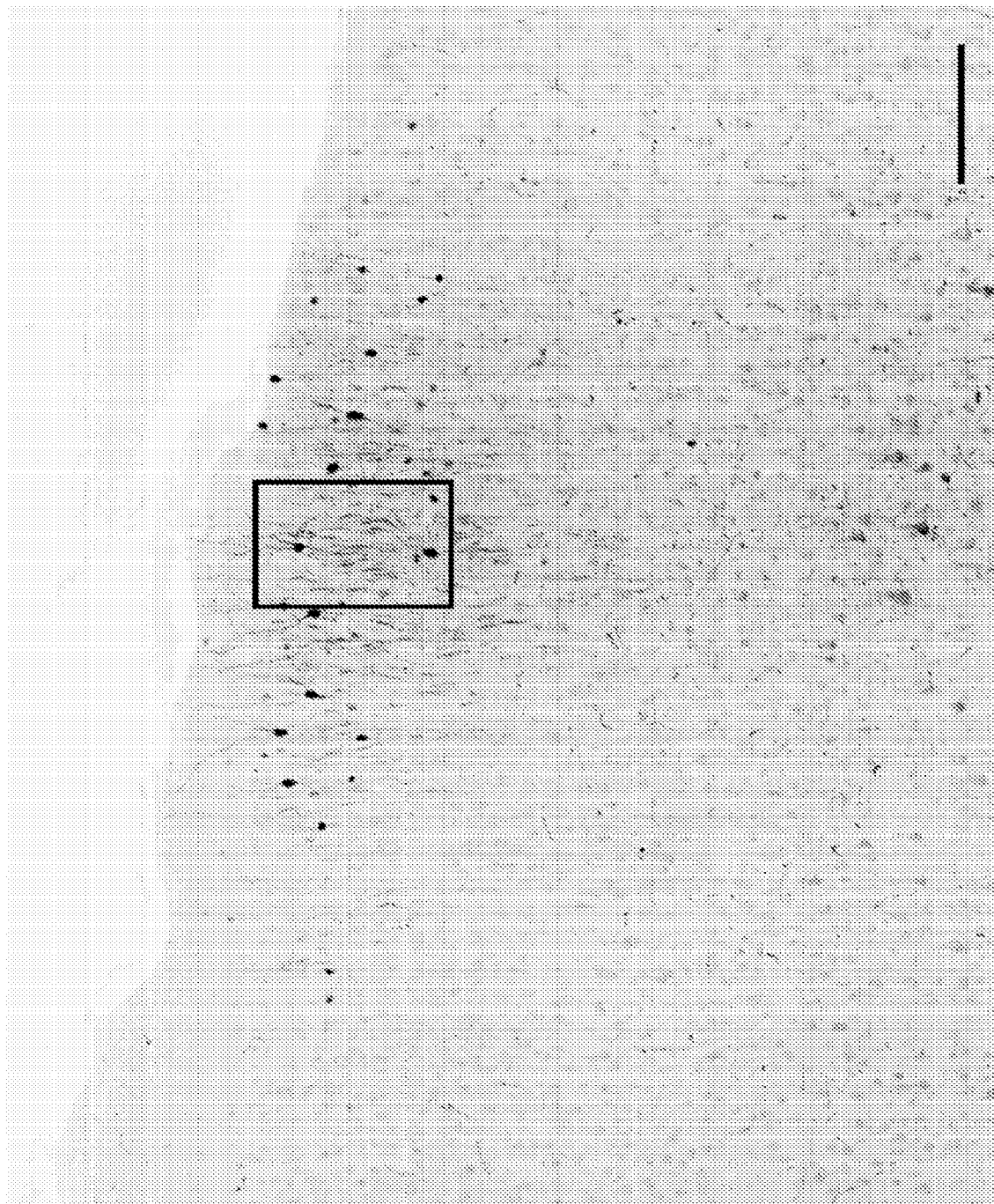
43/112

FIG. 13B



44/112

FIG. 14A



45/112

FIG. 14B



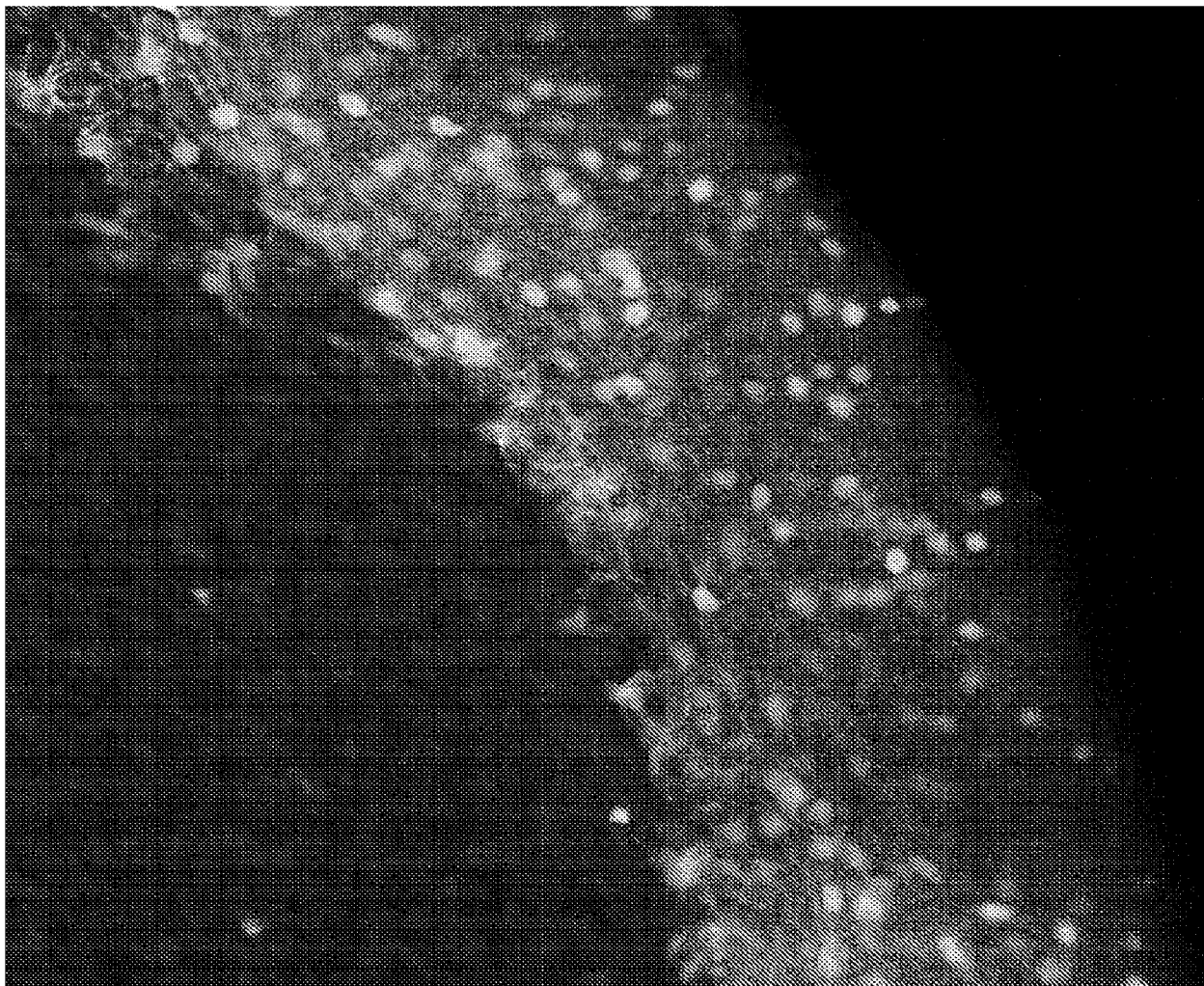
46/112

FIG. 15A



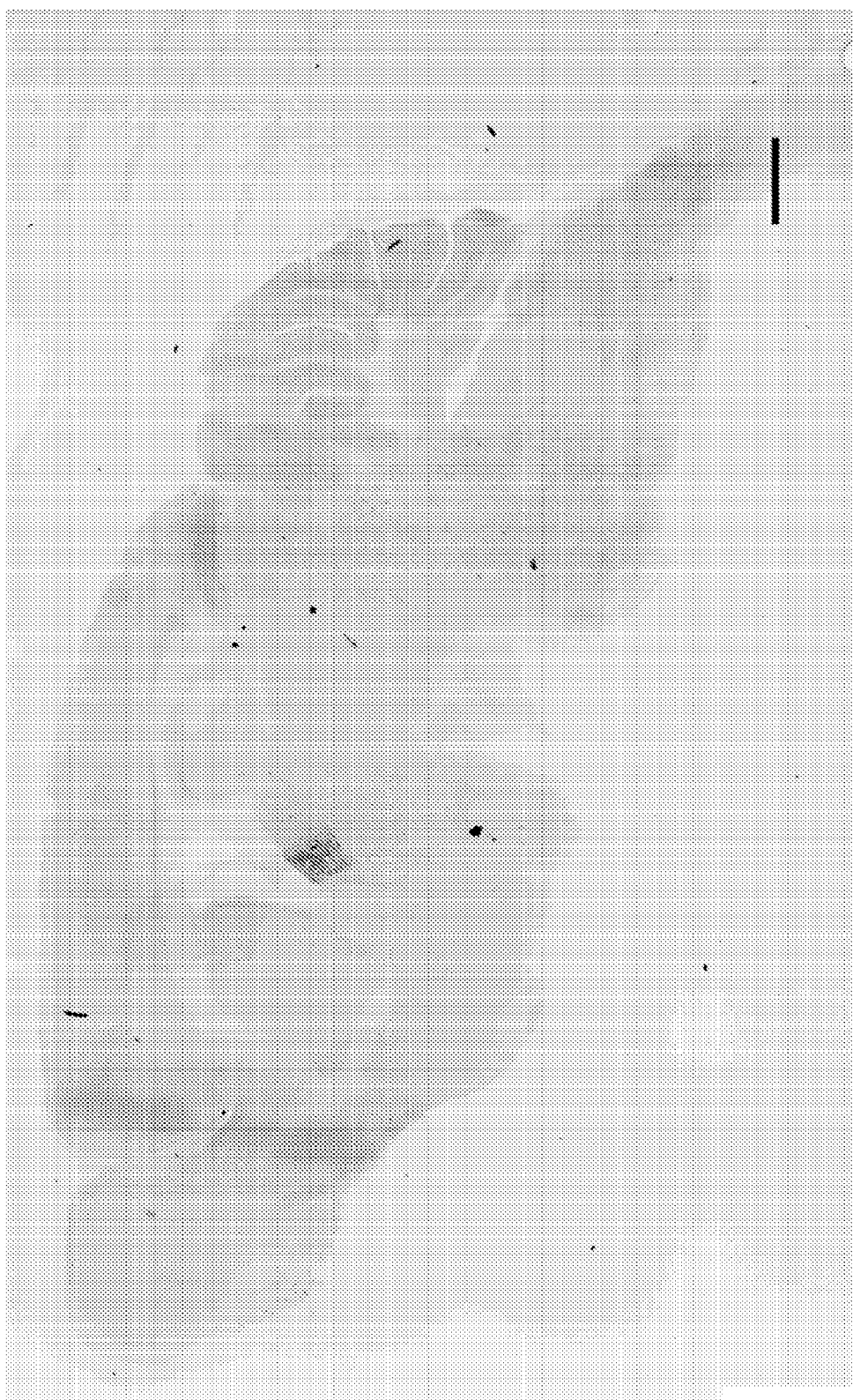
47/112

FIG. 15B



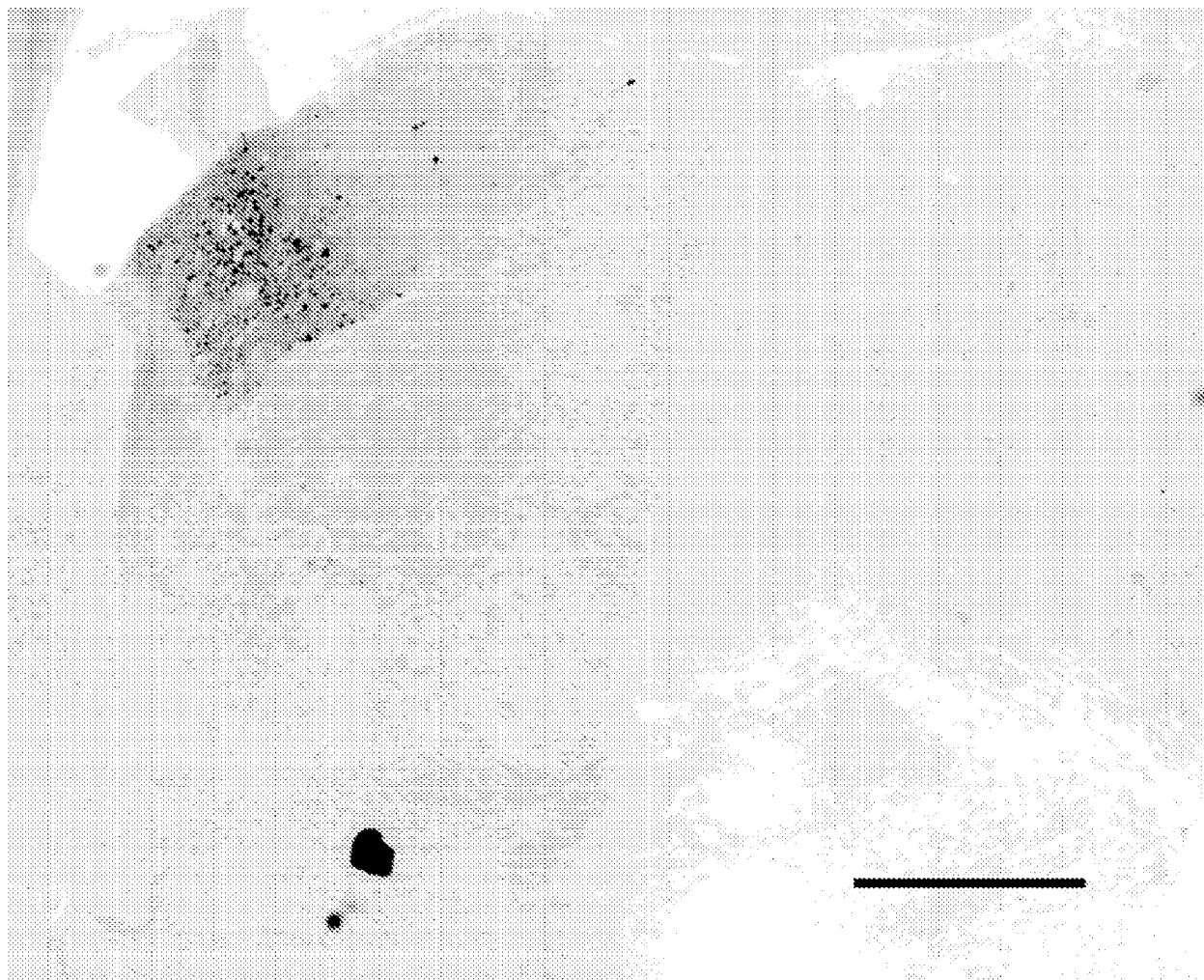
48/112

FIG. 16A



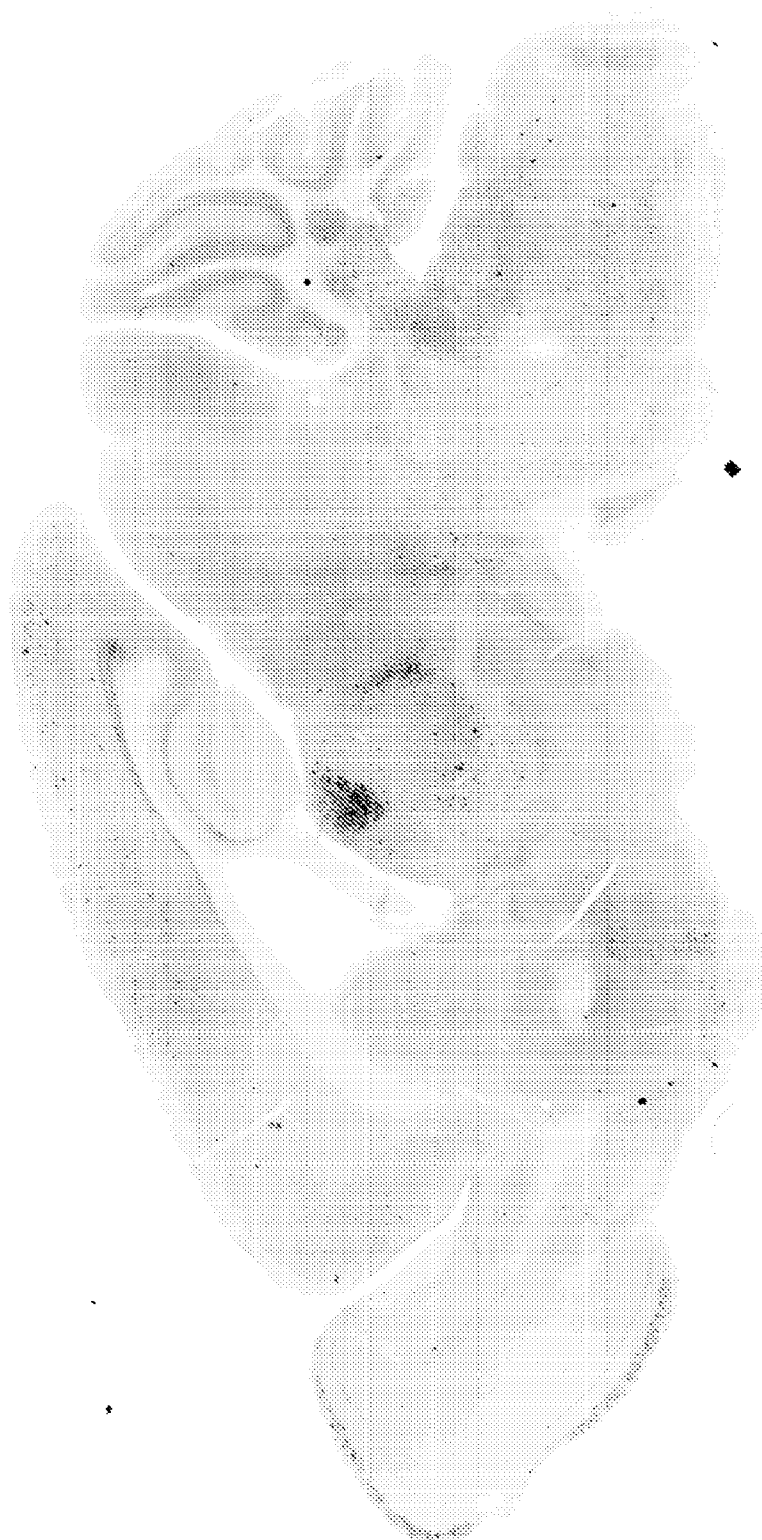
49/112

FIG. 16B



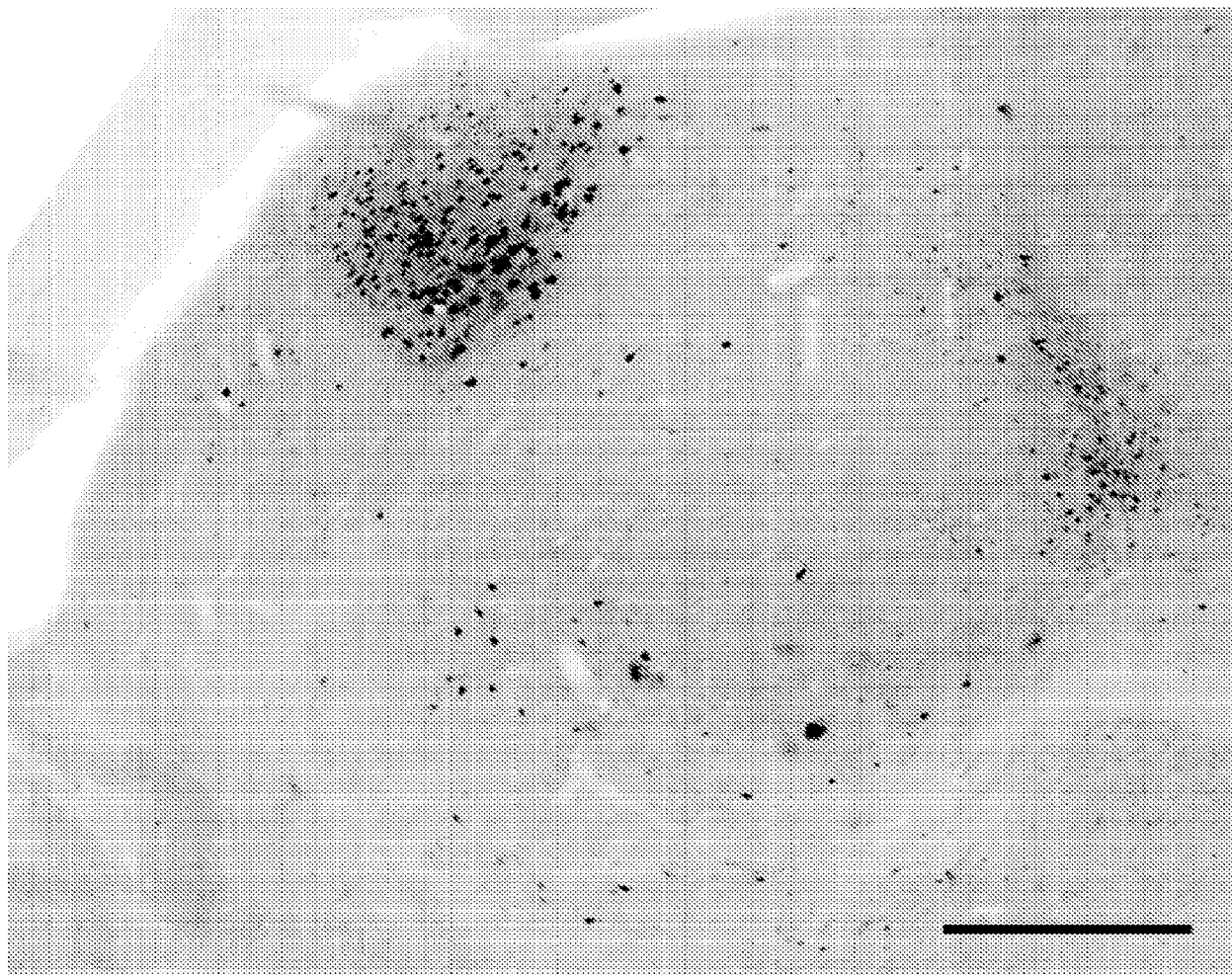
50/112

FIG. 17A



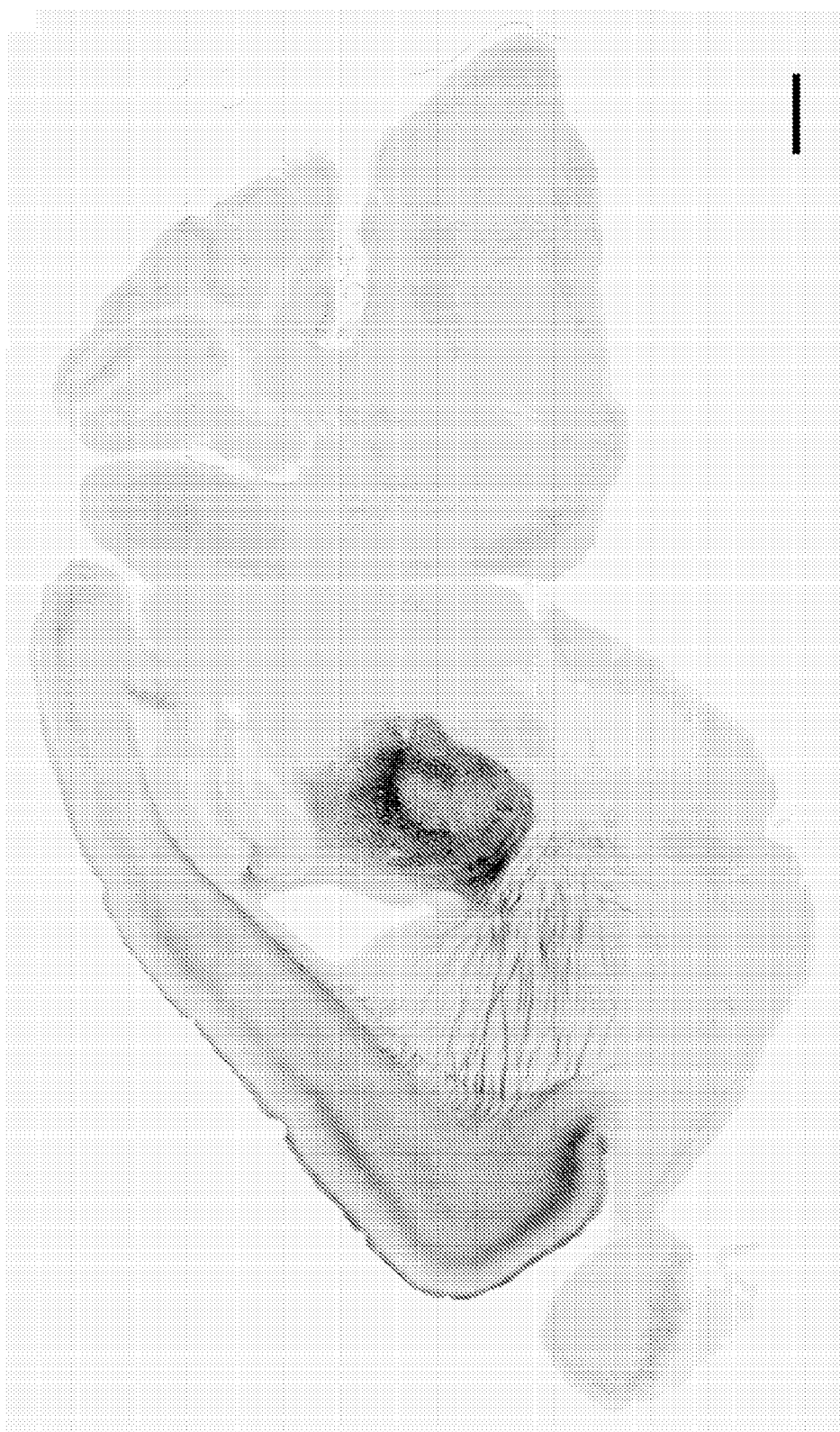
51/112

FIG. 17B



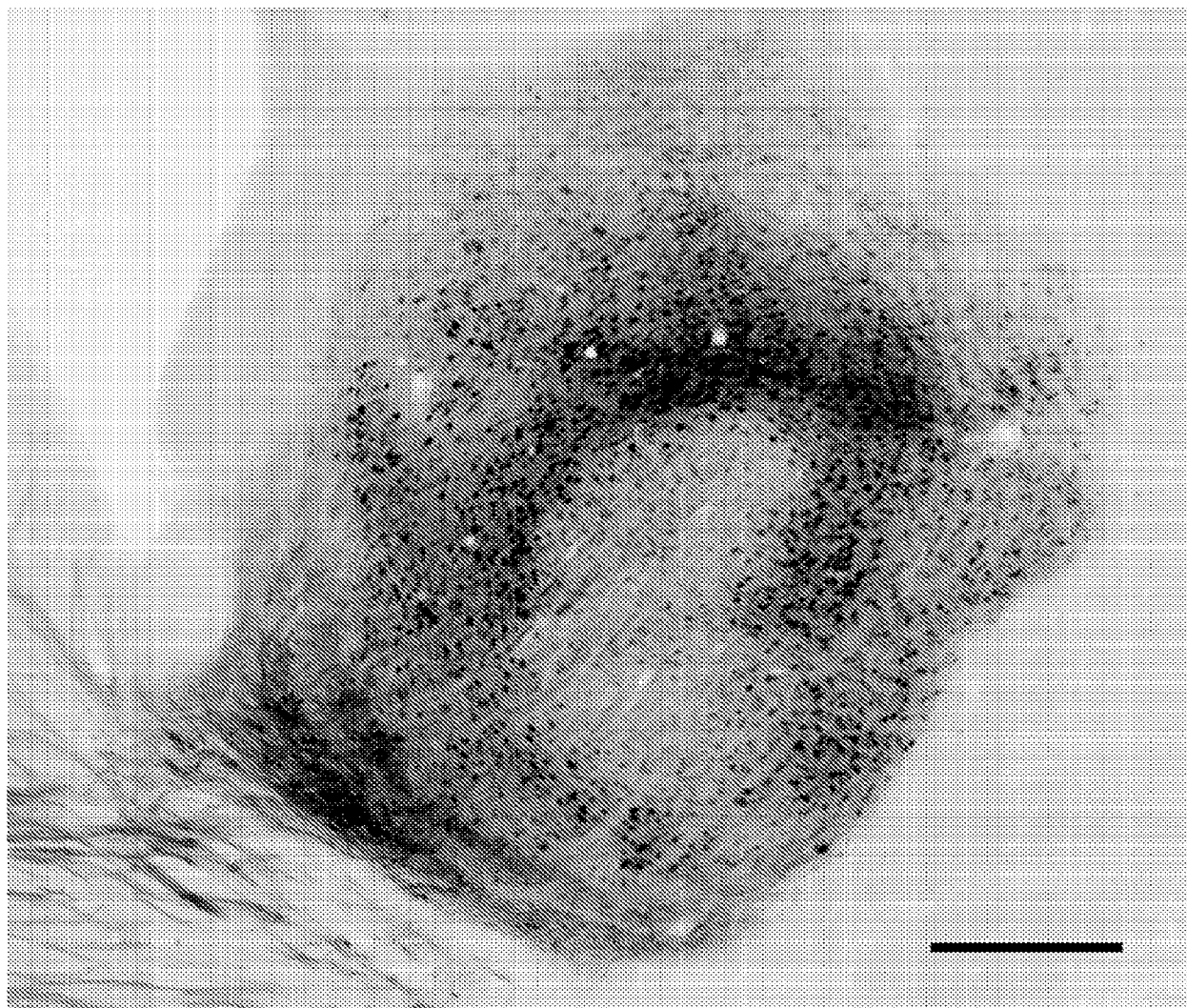
52/112

FIG. 18A



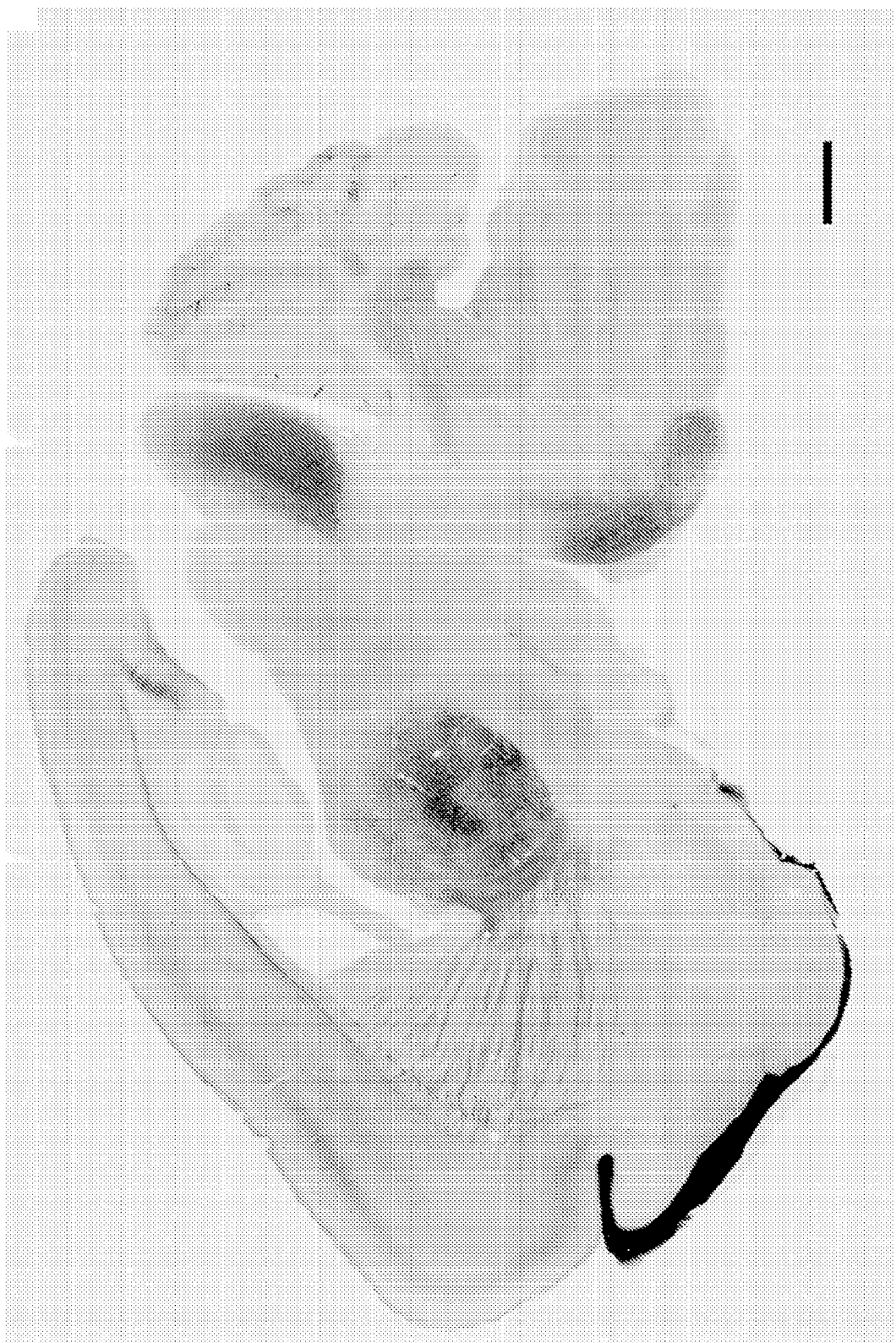
53/112

FIG. 18B



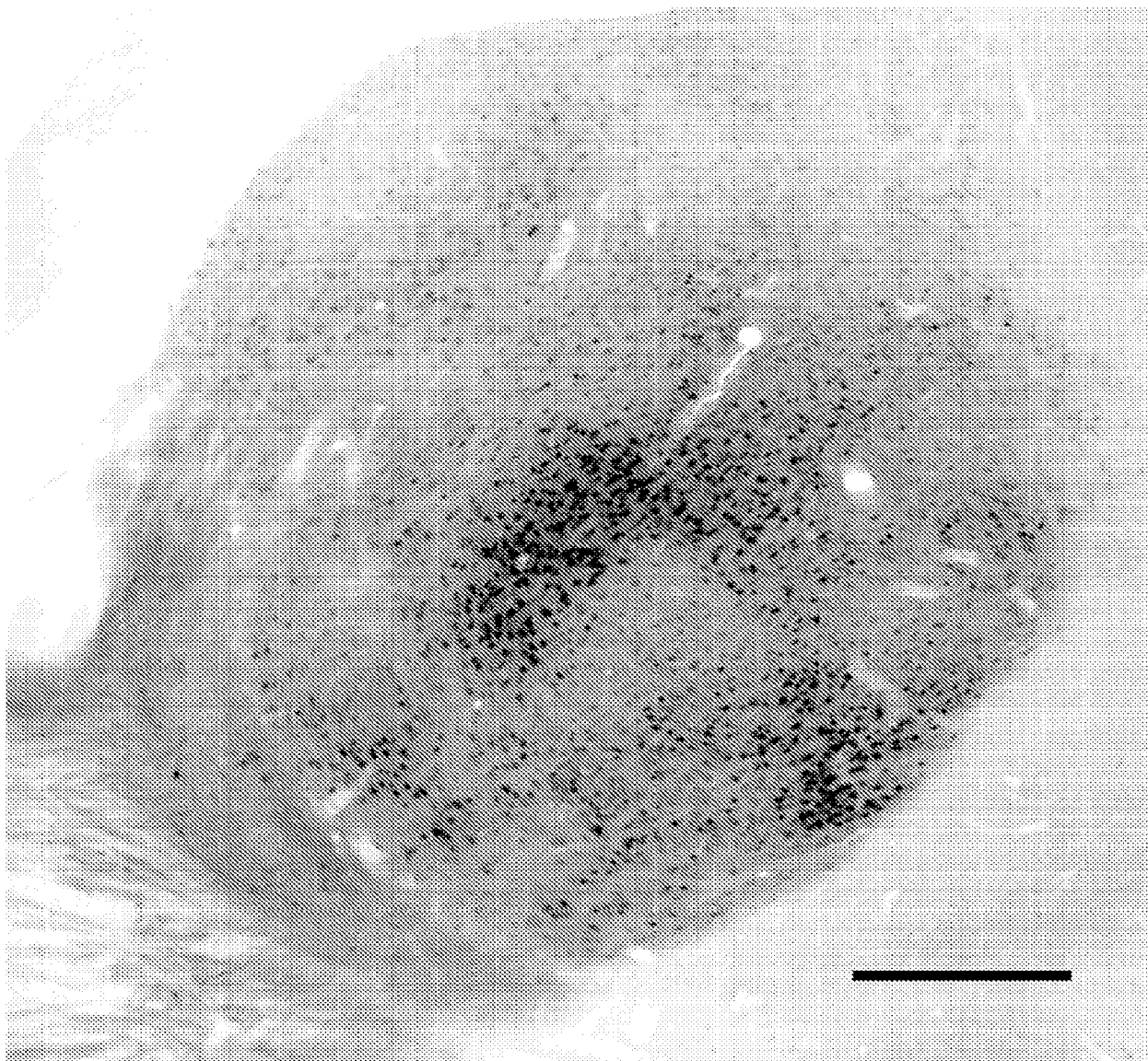
54/112

FIG. 19A



55/112

FIG. 19B



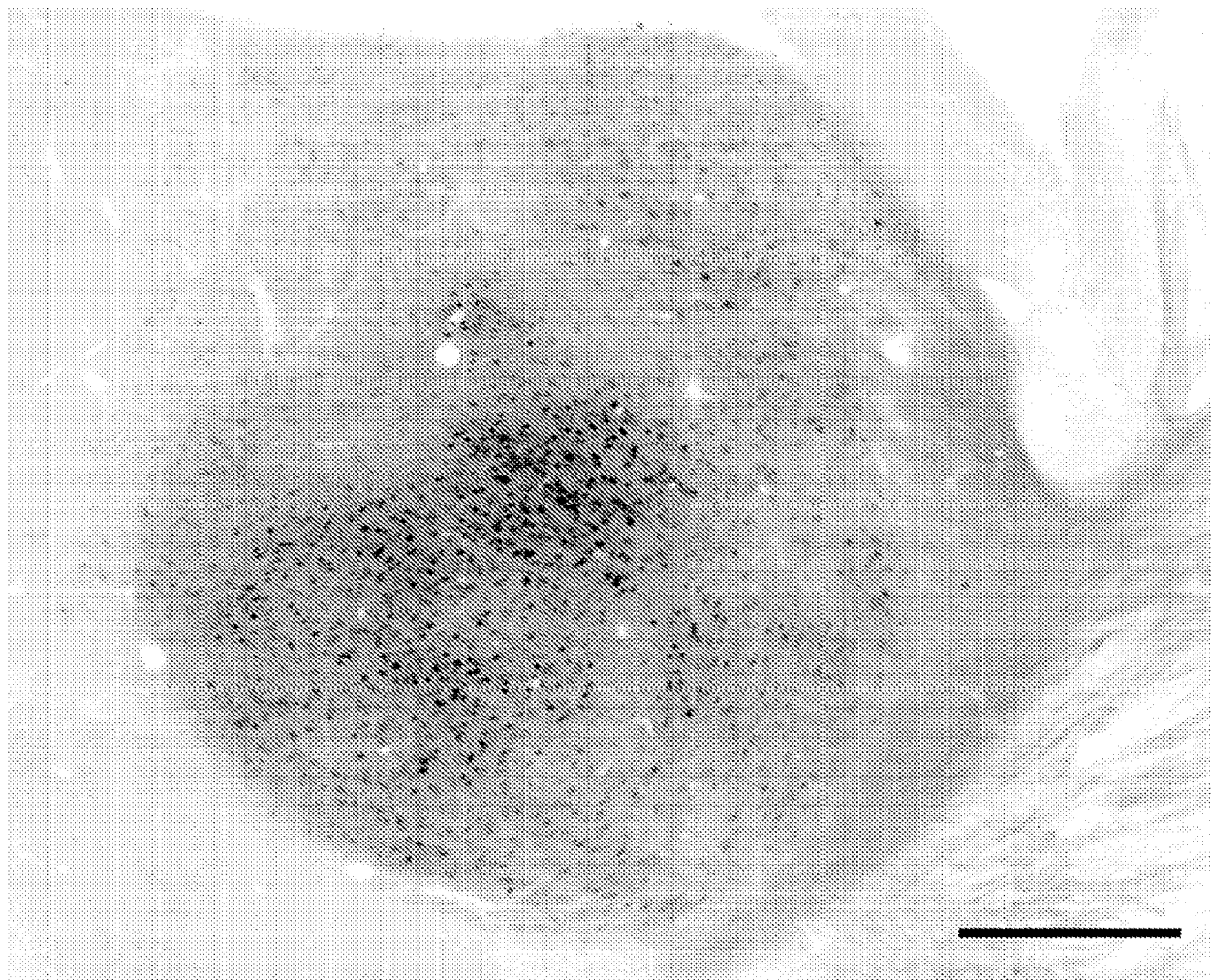
56/112

FIG. 20A



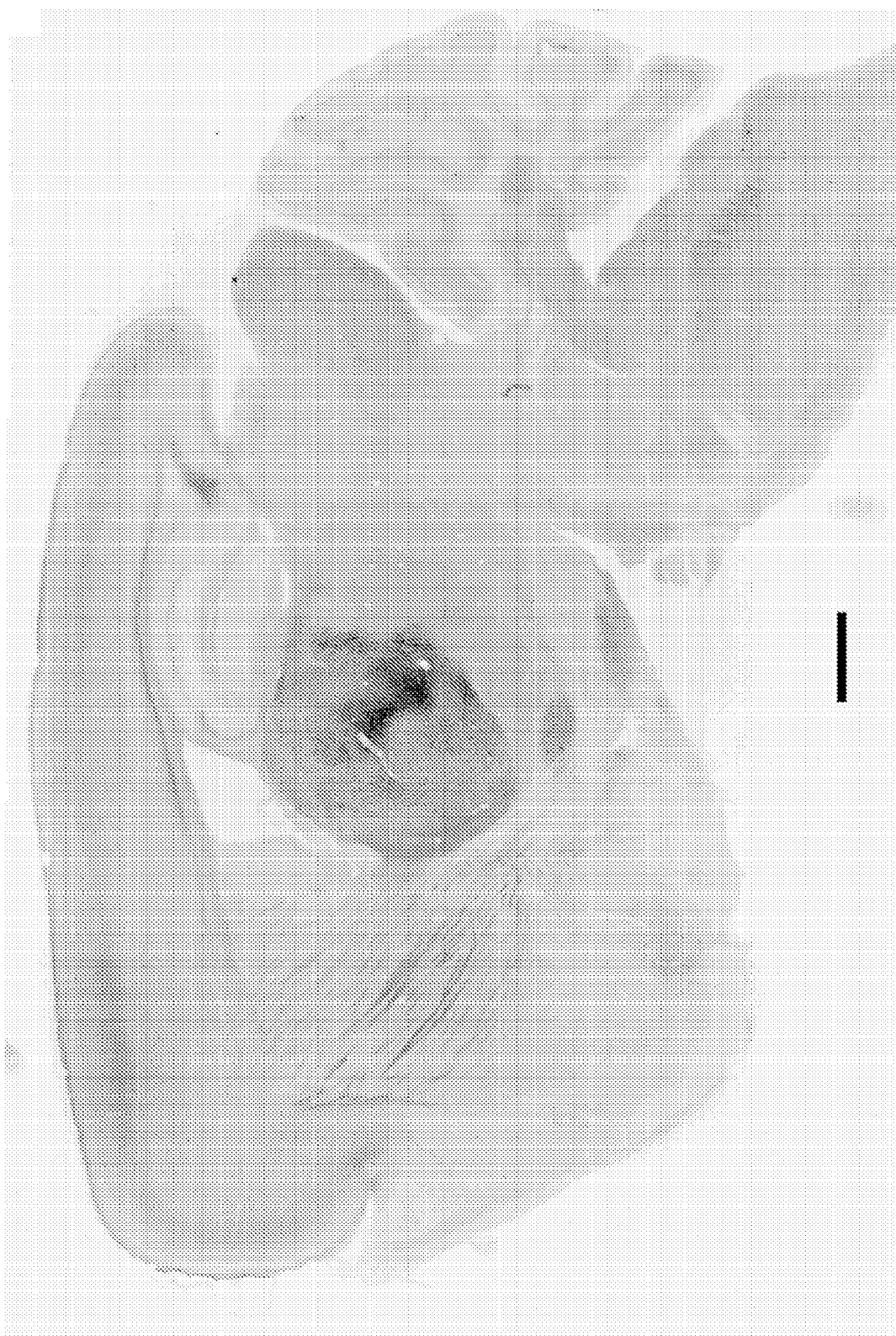
57/112

FIG. 20B



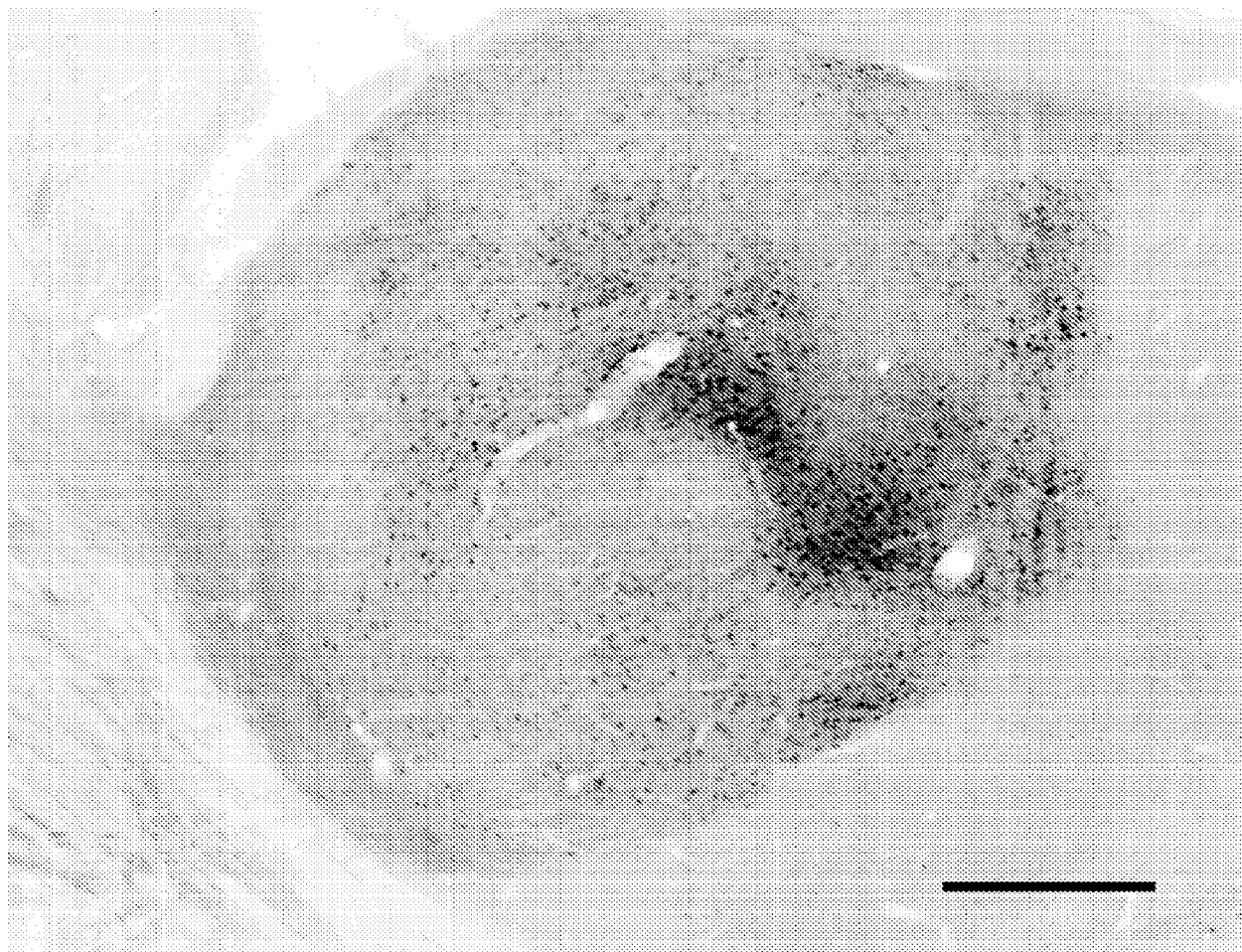
58/112

FIG. 21A



59/112

FIG. 21B



60/112

FIG. 22A

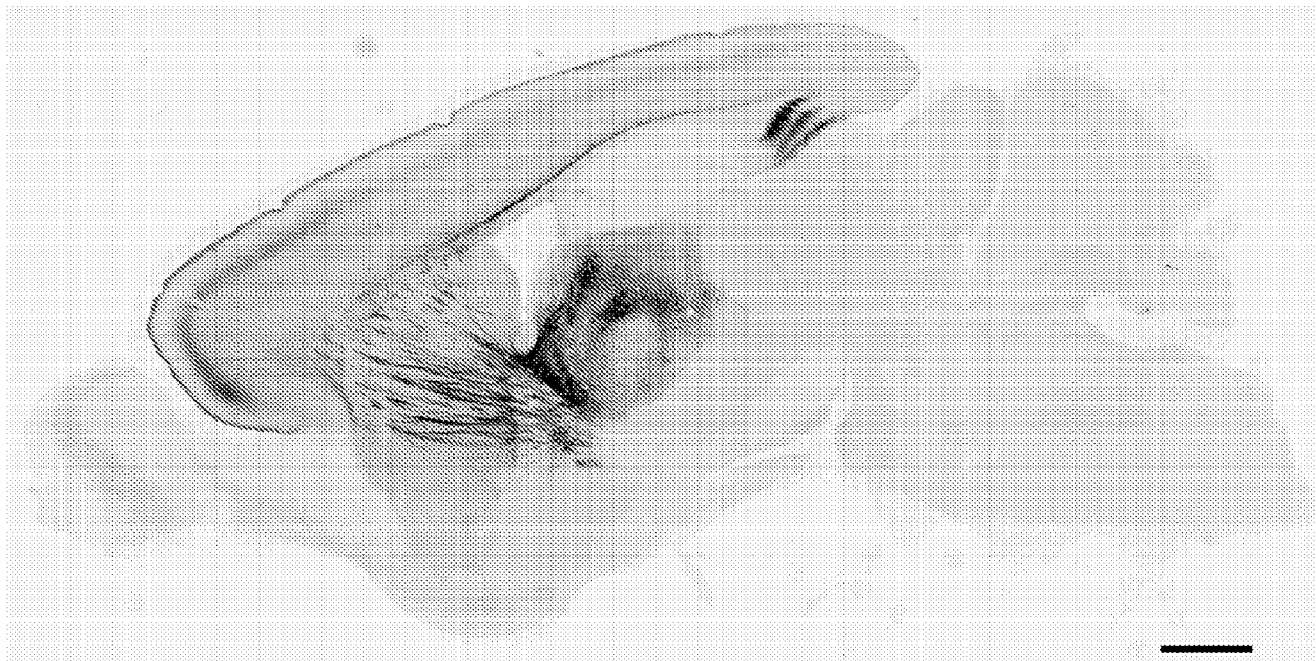
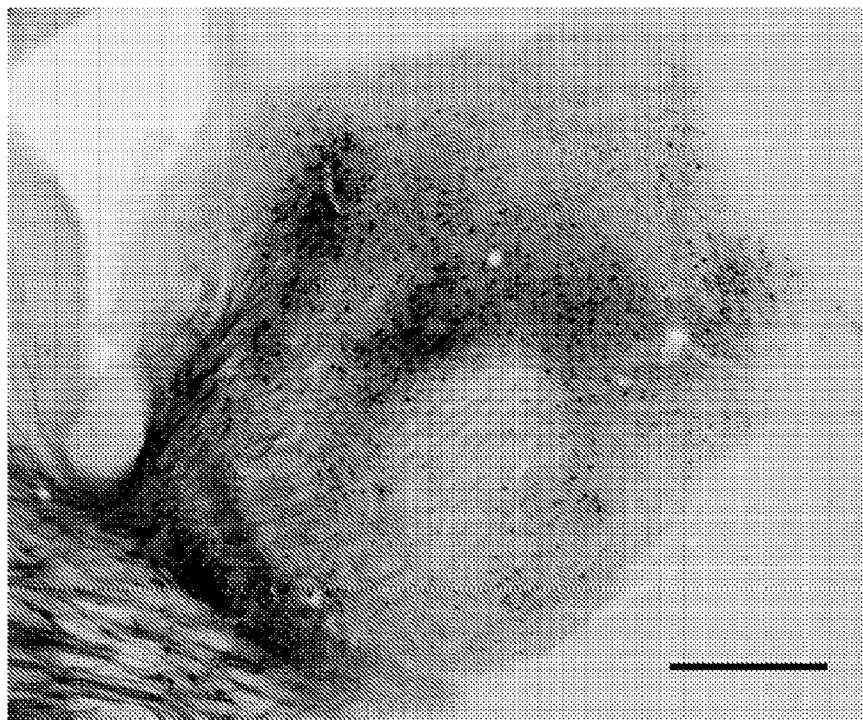


FIG. 22B



61/112

FIG. 23

Abbrev.	Full structure name
AAA	Anterior amygdalar area
ACA _d	Anterior cingulate area, dorsal part
ACA _v	Anterior cingulate area, ventral part
ACB	Nucleus accumbens
Acs5	Accessory trigeminal nucleus
ACVII	Accessory facial motor nucleus
AD	Anterodorsal nucleus
ADP	Anterodorsal preoptic nucleus
AHN	Anterior hypothalamic nucleus
A _d	Agranular insular area, dorsal part
A _p	Agranular insular area, posterior part
A _v	Agranular insular area, ventral part
AM	Anteromedial nucleus
AMB	Nucleus ambiguus
AN	Ansiform lobule
AOB	Accessory olfactory bulb
AON	Anterior olfactory nucleus
AP	Area postrema
APN	Anterior pretectal nucleus
APr	Area prostriata
ARH	Arcuate hypothalamic nucleus
ASO	Accessory supraoptic group
AT	Anterior tegmental nucleus
AUD _d	Dorsal auditory area
AUD _p	Primary auditory area
AUD _{po}	Posterior auditory area
AUD _v	Ventral auditory area
AV	Anteroventral nucleus of thalamus
AVP	Anteroventral preoptic nucleus
AVPV	Anteroventral periventricular nucleus
B	Barrington's nucleus

Abbrev.	Full structure name
BA	Bed nucleus of the accessory olfactory tract
BAC	Bed nucleus of the anterior commissure
BLA	Basolateral amygdalar nucleus
BMA	Basomedial amygdalar nucleus
BST	Bed nuclei of the stria terminalis
CA1	Field CA1
CA2	Field CA2
CA3	Field CA3
CEA	Central amygdalar nucleus
CENT	Central lobule
CL	Central lateral nucleus of the thalamus
CLA	Clastrum
CLI	Central linear nucleus raphe
CM	Central medial nucleus of the thalamus
COA _a	Cortical amygdalar area, anterior part
COA _p	Cortical amygdalar area, posterior part
COPY	Copula pyramidis
CP	Caudoputamen
CS	Superior central nucleus raphe
CU	Cuneate nucleus
CUL	Culmen
CUN	Cuneiform nucleus
DCO	Dorsal cochlear nucleus
DEC	Declive (VI)
DG	Dentate gyrus
DMH	Dorsomedial nucleus of the hypothalamus
DMX	Dorsal motor nucleus of the vagus nerve
DN	Dentate nucleus
DP	Dorsal peduncular area
DR	Dorsal nucleus raphe
DT	Dorsal terminal nucleus of the accessory optic tract

62/112

FIG. 23 cont'd

Abbrev.	Full structure name	Abbrev.	Full structure name
DTN	Dorsal tegmental nucleus	ILA	Infralimbic area
ECT	Ectorhinal area	IMD	Intermediodorsal nucleus of the thalamus
ECU	External cuneate nucleus	IntG	Intermediate geniculate nucleus
ENTI	Entorhinal area, lateral part	IO	Inferior olivary complex
ENTm	Entorhinal area, medial part, dorsal zone	IP	Interposed nucleus
EPd	Endopiriform nucleus, dorsal part	IPN	Interpeduncular nucleus
EPv	Endopiriform nucleus, ventral part	IRN	Intermediate reticular nucleus
EW	Edinger-Westphal nucleus	ISN	Inferior salivatory nucleus
FC	Fasciola cinerea	IV	Trochlear nucleus
fiber tracts	fiber tracts	LA	Lateral amygdalar nucleus
FL	Flocculus	LAV	Lateral vestibular nucleus
FN	Fastigial nucleus	LC	Locus ceruleus
FOTU	Folium-tuber vermis (VII)	LD	Lateral dorsal nucleus of thalamus
FRP	Frontal pole, cerebral cortex	LDT	Laterodorsal tegmental nucleus
FS	Fundus of striatum	LGd	Dorsal part of the lateral geniculate complex
GPe	Globus pallidus, external segment	LGv	Ventral part of the lateral geniculate complex
GPi	Globus pallidus, internal segment	LH	Lateral habenula
GR	Gracile nucleus	LHA	Lateral hypothalamic area
GRN	Gigantocellular reticular nucleus	LIN	Linear nucleus of the medulla
GU	Gustatory areas	LING	Lingula (I)
HATA	Hippocampo-amygdalar transition area	LM	Lateral mammillary nucleus
I5	Intertrigeminal nucleus	LP	Lateral posterior nucleus of the thalamus
IA	Intercalated amygdalar nucleus	LPO	Lateral preoptic area
IAD	Interanterodorsal nucleus of the thalamus	LRN	Lateral reticular nucleus
IAM	Interanteromedial nucleus of the thalamus	LSc	Lateral septal nucleus, caudal (caudodorsal) part
IC	Inferior colliculus	LSr	Lateral septal nucleus, rostral (rostroventral) part
ICB	Infracerebellar nucleus	LSv	Lateral septal nucleus, ventral part
IF	Interfascicular nucleus raphe	LT	Lateral terminal nucleus of the accessory optic tract
IG	Induseum griseum	MA	Magnocellular nucleus
IGL	Intergeniculate leaflet of the lateral geniculate complex	MA3	Medial accessory oculomotor nucleus

63/112

FIG. 23 cont'd

Abbrev.	Full structure name	Abbrev.	Full structure name
III	Oculomotor nucleus	MARN	Magnocellular reticular nucleus
MD	Mediodorsal nucleus of thalamus	ORBI	Orbital area, lateral part
MDRN	Medullary reticular nucleus	ORBm	Orbital area, medial part
MDRN _d	Medullary reticular nucleus, dorsal part	ORBvl	Orbital area, ventrolateral part
MDRN _v	Medullary reticular nucleus, ventral part	OT	Olfactory tubercle
ME	Median eminence	OV	Vascular organ of the lamina terminalis
MEA	Medial amygdalar nucleus	P5	Peritrigeminal zone
MEPO	Median preoptic nucleus	Pa4	Paratrochlear nucleus
MEV	Midbrain trigeminal nucleus	Pa5	Paratrigeminal nucleus
MG	Medial geniculate complex	PAA	Piriform-amygdalar area
MH	Medial habenula	PAG	Periaqueductal gray
MM	Medial mammillary nucleus	PAR	Parasubiculum
MOB	Main olfactory bulb	PARN	Parvicellular reticular nucleus
MO _p	Primary motor area	PAS	Parasolitary nucleus
MO _s	Secondary motor area	PB	Parabrachial nucleus
MPN	Medial preoptic nucleus	PBG	Parabigeminal nucleus
MPO	Medial preoptic area	PCG	Pontine central gray
MPT	Medial pretectal area	PCN	Paracentral nucleus
MRN	Midbrain reticular nucleus	PD	Posterodorsal preoptic nucleus
MS	Medial septal nucleus	PDTg	Posterodorsal tegmental nucleus
MT	Medial terminal nucleus of the accessory optic tract	PeF	Perifornical nucleus
MV	Medial vestibular nucleus	PERI	Perirhinal area
NB	Nucleus of the brachium of the inferior colliculus	PF	Parafascicular nucleus
NDB	Diagonal band nucleus	PFL	Paraflocculus
NI	Nucleus incertus	PG	Pontine gray
NLL	Nucleus of the lateral lemniscus	PGRN _d	Paragigantocellular reticular nucleus, dorsal part
NLOT	Nucleus of the lateral olfactory tract	PGRN _l	Paragigantocellular reticular nucleus, lateral part
NOD	Nodulus (X)	PH	Posterior hypothalamic nucleus
NOT	Nucleus of the optic tract	PIL	Posterior intralaminar thalamic nucleus
NPC	Nucleus of the posterior commissure	PIR	Piriform area
NR	Nucleus of Roller	PL	Prelimbic area
NTB	Nucleus of the trapezoid body	PM _d	Dorsal premammillary nucleus
NTS	Nucleus of the solitary tract	PM _v	Ventral premammillary nucleus

64/112

FIG. 23 cont'd

Abbrev.	Full structure name	Abbrev.	Full structure name
OP	Olivary pretectal nucleus	PN	Paranigral nucleus
PO	Posterior complex of the thalamus	RE	Nucleus of reuniens
POL	Posterior limiting nucleus of the thalamus	RH	Rhomboid nucleus
POST	Postsubiculum	RL	Rostral linear nucleus raphe
PoT	Posterior triangular thalamic nucleus	RM	Nucleus raphe magnus
PP	Peripeduncular nucleus	RN	Red nucleus
PPN	Pedunculo pontine nucleus	RO	Nucleus raphe obscurus
PPT	Posterior pretectal nucleus	RPA	Nucleus raphe pallidus
PPY	Parapyramidal nucleus	RPO	Nucleus raphe pontis
PR	Perireunensis nucleus	RR	Midbrain reticular nucleus, retrorubral area
PRE	Presubiculum	RSPagl	Retrosplenial area, lateral agranular part
PRM	Paramedian lobule	RSPd	Retrosplenial area, dorsal part
PRNc	Pontine reticular nucleus, caudal part	RSPv	Retrosplenial area, ventral part
PRNr	Pontine reticular nucleus	RT	Reticular nucleus of the thalamus
ProS	Prosubiculum	SAG	Nucleus sagulum
PRP	Nucleus prepositus	SBPV	Subparaventricular zone
PS	Parastrial nucleus	SCH	Suprachiasmatic nucleus
PST	Preparasubthalamic nucleus	SCm	Superior colliculus, motor related
PSTN	Parasubthalamic nucleus	SCO	Subcommissural organ
PSV	Principal sensory nucleus of the trigeminal	SCs	Superior colliculus, sensory related
PT	Parataenial nucleus	SF	Septofimbrial nucleus
PVa	Periventricular hypothalamic nucleus, anterior part	SFO	Subfornical organ
PVH	Paraventricular hypothalamic nucleus	SG	Supragenual nucleus
PVHd	Paraventricular hypothalamic nucleus, descending division	SGN	Suprageniculate nucleus
PVi	Periventricular hypothalamic nucleus, intermediate part	SH	Septohippocampal nucleus
PVp	Periventricular hypothalamic nucleus, posterior part	SI	Substantia innominata
PVpo	Periventricular hypothalamic nucleus, preoptic part	SIM	Simple lobule
PVT	Paraventricular nucleus of the thalamus	SLC	Subceruleus nucleus
PYR	Pyramus (VIII)	SLD	Sublaterodorsal nucleus

65/112

FIG. 23 cont'd

Abbrev.	Full structure name	Abbrev.	Full structure name
RCH	Retrochiasmatic area	SMT	Submedial nucleus of the thalamus
SNC	Substantia nigra, compact part	TR	Postpiriform transition area
SNr	Substantia nigra, reticular part	TRN	Tegmental reticular nucleus
SO	Supraoptic nucleus	TRS	Triangular nucleus of septum
SOC	Superior olivary complex	TT	Taenia tecta
SPA	Subparafascicular area	TU	Tuberal nucleus
SPFm	Subparafascicular nucleus, magnocellular part	UVU	Uvula (IX)
SPFp	Subparafascicular nucleus, parvicellular part	V	Motor nucleus of trigeminal
SPIV	Spinal vestibular nucleus	VAL	Ventral anterior-lateral complex of the thalamus
SPVC	Spinal nucleus of the trigeminal, caudal part	VCO	Ventral cochlear nucleus
SPVI	Spinal nucleus of the trigeminal, interpolar part	VeCB	Vestibulocerebellar nucleus
SPVO	Spinal nucleus of the trigeminal, oral part	VI	Abducens nucleus
SSp-bfd	Primary somatosensory area, barrel field	VII	Facial motor nucleus
SSp-II	Primary somatosensory area, lower limb	VISa	Anterior area
SSp-m	Primary somatosensory area, mouth	VISal	Anterolateral visual area
SSp-n	Primary somatosensory area, nose	VISam	Anteromedial visual area
SSp-tr	Primary somatosensory area, trunk	VISC	Visceral area
SSp-ul	Primary somatosensory area, upper limb	VISI	Lateral visual area
SSp-un	Primary somatosensory area, unassigned	VISp	Primary visual area
SSs	Supplemental somatosensory area	VISpl	Posterolateral visual area
STN	Subthalamic nucleus	VISpm	posteromedial visual area
SUB	Subiculum	VISpor	Postrhinal area
SubG	Subgeniculate nucleus	VISrl	Rostrolateral visual area
SUM	Supramammillary nucleus	VLPO	Ventrolateral preoptic nucleus
SUT	Supratrigeminal nucleus	VM	Ventral medial nucleus of the thalamus
SUV	Superior vestibular nucleus	VMH	Ventromedial hypothalamic nucleus
TEa	Temporal association areas	VMPO	Ventromedial preoptic nucleus
TMd	Tuberomammillary nucleus, dorsal part	VPL	Ventral posterolateral nucleus of the thalamus

66/112

FIG. 23 cont'd

Abbrev.	Full structure name
TMv	Tuberomammillary nucleus, ventral part

Abbrev.	Full structure name
VPLpc	Ventral posterolateral nucleus of the thalamus, parvicellular part

Abbrev.	Full structure name
VPM	Ventral posteromedial nucleus of the thalamus
VPMpc	Ventral posteromedial nucleus of the thalamus, parvicellular part
VTA	Ventral tegmental area
VTN	Ventral tegmental nucleus
x	Nucleus x
Xi	Xiphoid thalamic nucleus
XII	Hypoglossal nucleus
y	Nucleus y
ZI	Zona incerta

67/112

FIG. 24

eHGT_576h (438 bp):

ACTGAGCCTTACAGAGATCCAGGGACTTGTCTTGACCAGTGAGCAAGTGGGTGGAGTGG
GCTATGGACTCTGGACTGGAGGAGCCCCGAGCCCGTGCTCTCAGCCGCTGCTCAGCCCT
GCAGCAGGAAGAACATCTGAAGCAAACCTACAAAGAACTTTTCTCCCTGAAAAAGCGATTAG
CCCATAATAGAAGACCTAAGTCCAGCAAAAAGAGCCGCCGTTTCTGCTGCACTTTGTTTCC
TACTGGGTGACGGATGGCGCTGTCATCTCTTCAGGGAGGGAGGCTTACAGACAGGCCAG
CTCGTCAGCCTGCTGGCTCCTTCAAAGTCAGGTCTTCACAGTAGGTTCTTTATTTCTGTCT
CAACAAGTTGAAATCCCCTCAAATTAGACACACACTTTTCTTATGTAATTAAGTACAGTGA
GAAGACAGGCATCA (SEQ ID NO: 1)

eHGT_577h (775 bp):

TGCCTGTCTTCTTCTCTGCAGTACTGCTCAGTATTGCTGGAAACTCAGAACTGAAGACA
GTCTTGTCCCCTAATAGATTAGGATTCAAAAACAGACTTTAGGAGGACTAGGTGTATTTTT
CCCTTTTAAACAGCAGAGGGATTTGAAGATCTGGACCTTTCCTGTTGAGTTATTGAACCTGC
AGTATTTGTTATTGTCTTGCCAGTACCTATAACTACCTGGGAAAAAAGACAAACCAGATGTT
GGAGGCTTTTGAACAAAAGGACTTCTCTTTAGCCCTGAACTATCAGGTAAACGTTTTCTTCT
CTTCCCATAAATCAGAGCCTTTTTTCTTCCCACCTTGAGGATTCTTTTTCTTCTCTACCAT
GCCTGCCCTCATTAAAATTGCCAGCCACTTCTTTCTTTACCTTTCTTCTGTTTTGTACTTTT
TCTTGTTTAAACCAAAGCTGTCCCCATTGTCTGTGTTCCCGACATGTCTAATCAGGACAGG
ACCCATACAGGGATGGCTCCTCTCTTCTCCTAGAAAGGATCTTATTTTTACTCTATCTTCA
AAGTACTTTTGTGACTATAATAAAATGGTCATAGATTTCTGGGAAAATGACCACAGGCTGTC
CCCTTGAGGACATCACATCAGAGTTATCACAGAATATGAGGAAGCTATGAAGATGGTAACC
TAAAATGGGGAAGGCTTCATTCAAAGTCAGGGCAGGTGAGTCTTGTCCACCCTGGAAATGT
GCTGAGAACTAGGAACATCCACCAGGAAGGAA (SEQ ID NO: 2)

eHGT_578h (779 bp):

AGTCCTCACAACAACCCTGTGGGCAACCATGTCAGAATTCATTTATTTTTAAACAGTTAAAG
AATCTGACTCACAGAGAAGGTAAGTGACTCTCCTAAGATTACTGGGCAGTAAATGGCAGCA
TCTGGACTAGAACACACCATTTTCTGATTTTTTGTCTGTTCTTGTCTATGCCAACCTTTTAT
GTAACTTTTTCATGCACAAAAGGAATAGTTCTTCCCGAGAGTTCTGTTCAATTTGCCAACTAGT
GGATAAAAAAAGCTTTAAGGTTTTTACCTAAAGCCCTAACTATTGTCATAATTTTAAATGAA
TAGGCCTATTTTTCTCTCTTTGAGAAAAAAAATCAGAAGTAAATGAGTCAGAATTTCTCT
GAACAAGCCCTCCACCTGAGGTAAGTGAAGGACAGGTTGGCTTGTCTCTCTGTGGGTAC
AGAATAGGAAAGGACTGGAAAGCTTTTTTGAGAATACTTTCAAATTTCTGGCCAAAGCTTCC
TGCTGTGCCGAAGTTTTTTATTTCCAGACTTTTACAAGGATGAAAAAGGTGATGTTGGAG
CAGAACTCATGAACACCTGAGACTCTGAGGTCACCTGGTGGGGCTGTGATTGCAGGAGAT
CCAGGAAGACTGTGGGCAGACATGACTCTAGAGGAGGCTGAAAAAAGGCCAGGTGTGGA
GAGCCCCCTTGGGCCCTTGGGAGAAGGCCAGCAAGACAGGAAGGCCTAGACAAAAGGGC
ATGAGTCTTCCCTTCCCTATCAGGGCTAGAGCTTTAGAACAGGA (SEQ ID NO: 3)

eHGT_579h (791 bp):

ACTATTCGCAGCAGTGGGAATGTAAGTTGGCCCCATTCTCCTGGAAAGTCATTTGGCACTC
TGTGGGCACTGAAGACCAGCAGTGTCTCTGCTGGGGACAGCCTTGGTGGCCTCAAGAGTC
ACACAGATGAGGATGCTCGTTGTCTATTTTGTCTCTGTGCTTGTCTGTGTTCCAAATGA
TCCATAATTTTTTAAAGTGTCTTTTAAAGAATATGTAGGTGTTCTGACTCTTGATCTGATGGC
TGGTTACCCAGGCGTCCTTTTATAATAATTTGTTATACTGTTCTTTTGTCTATGTGCTTTTC
TTTATGGATATGACATTTTACAGTTTTTAAAAATTAGAAAGAATGAGTAGGCGAGACGCTGG
GGTCAGGCAGTCTCAGGCAGAGGAAGCAGCCTGAGCAAAGCCCGGGGAGCTGGGCAAAG
GCAGGGGTGTGCATGACCCTGGGAATCCTGGCCCAACAACCAGGCACCCTGGCCCAAGG

68/112

FIG. 24 cont'd

ACATGCGGTCCCCTGAGGAACAGCATCTCTGCTCCCACCTCTGGCTAGCCATTGGTGCAG
CACCTGCACTGCCGAGGCCAGGCCTAATTTAGTCCCACCTTCAGCCTGTGTGAGGCTGGAT
CAAAAGTGAGCTCTCAGTGAGGGGGAACCAAGTAGAGGTGAGATGATGGGTTTAATAAAGT
AGTTTAACCAACCGCACTGGGTGTGCTAAGAGGCGGTTGCTGTGGCCAGCAGGTGGAGG
TTATAATGAGCCTGGCAGTTGTTGTTGGAGGCGGCAGGCAGTGGCTGTGCTGATGGGAGA
ATT (SEQ ID NO: 4)

eHGT_606h (824 bp):

GGGAAGAATAGTGAGTTCTGGGGTGGAGGGTTTTGATTTTTTAAAGTGGAAAGTCAAGAAA
TGCACCACGGAGAAGGTGGCATTGAGTGAAAACCTCAAAGTGGTTTGAGAGAAAGTGGTG
CAGATATGCGGACAGCGCTTCAGGCAGAGAAAATAACCTGGGATGTCTGAGAAAAGCAAA
GAAGCTGGGGCTACTGCAGTAAGAGCAAGAGAGTTAAGAATAGGTGTAGCCAGAGAGTTA
TCAGAATATAAGATTGCATAGGGCATTTTTTAGGCCAATATAAGGACTGGCATTTTTTAGGTCA
TGATGAGAACTGGCATTCTACTCTGCCTAAAAATAGGAAACCACTGGAGCATTTTGAGCAGAG
GAGTTGAACTGTCTGATTTAAAATTTTTAAAAGGTTCTGGCTACCTTGTGAGAAGAGACTGAA
GTAAGATATGCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAATAATCTAGGCCAAGAGA
GATGGATCAGGGTGATAACAGTAACTTCATGAGAGCATTTCAGTTGCAAAATATATTCCAGA
TTTGCTGACCAGATTTGATAAGGAATATAGAAAAAGAGGAAAGAACAAGAGATAACAGAGT
GTAGGCAGAAAAGAGAAAAGGTTCAAAGACTGAGCCCTGCAGTATTTCAACATTAAGAGCT
TGAATTGCAGAATAGGAACCATCTCAAGGCTTTAAATACAAGTGCTGAACACTCTCAAATTT
ATATTTGTATCCCAGACTTCTCTAAAATTCAAGTTCTAAAATCCAGTTGGTTACTTTATGTCT
CCACTTGGAACAAGTCTTGAAGCAGT (SEQ ID NO: 5)

eHGT_827h (746 bp):

TATCTGTGTCCAGTCCGATTGATGGGCTCATGGATTGGAAAGCTGTGACAGACGCCCTGG
GCTGGGAGGTATAAAACGAAACTGTTTAACTCCAGTGAAAAACAGTTGCAGGGATCCCAG
AAGTGGGAAAAGGTCGCTCTGAGCACCTACCGTGGGGTGGGTGTGATTCTGAGACTAAGG
CTCTGGGGGGCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGG
CGGGCGGATCACGAGGTCAGGAGATCGAGACCTGCTGGCTAACACGGTGAAACCCCTGT
CTCTACTAAAAATACAAAAAATTAGCTAGGCGTGGTGCGAGGTCCTGTAGTCCTAGCTAC
TCGGGAGGTTGAGGCAGGAGAGAATGGTGTGAACCGAGGAGGTGGAGCTTGCAGTGAGCTG
AGATGGCGCCACTGCACTCCAGCCTGGGCAACGGAGCGAGACTGTCTCAAAAAAAAAAAAA
AAAAAAGGCTCTGGGACCCCTGGCCACGAGCCCCAGGCCACGGCCTCCTGCCTGGGAG
GTGTGGCTCCCCATGCAGCCAGCTTCTTCCCTGCCTCAGATCGGTGACCTGCACTCTGTC
CAGGAGGAGGGGAGACTTGCCAGGGACAAACCCGAAAACACGGCAAGGTCAGGGGAG
GAGGGCACTGTTTCCTGTCTGAAGTCCACATTTGTGAAAAACATTGATTGGGCAGAGAGAG
GTGTTGTATTCTGGAATGGAGGTGAGTGG (SEQ ID NO: 137)

eHGT_828h (708 bp):

TCAGTCACTGCCATGGGTTCCATGTTATTCTGACATTGATGGTTACCACACTTATAGCTCCT
GTCCTGATCTCTCCCTTGATCTTGATAACTGCCAGCTTGACATCACCTCTCAGATGTCTGA
CAGGTGTCTCAGATTGAATATGGAAAACCAAGAACCCCTGATTCCCTCCCTTATCCTCACCA
TACCCCTGTCTTCTCCACACCACCAAGTAGTATAGGCATTAATTGCCCTTGAAGGGGATTTA
AAAAGGATATGAAAACCGTTTTTCAATCTTTGTGTGTCACTGTCCTAAAATTTCTATCTCATT
GCTTTTAAAAGAACC AAAATATGGGGGCCAAACATGGTGCCTCACACCTGTAATCCCAGCA
CTTTGGGAGGCCAAGGTGGGCAGATCACTAGGTCAGGAGTTCGACACCAGCATGGCCAAC
ATGGTGAAACCCCTGTCTCTACTAAAGATACAAAAAATTATCTGGGCGTGGTGGCAGGCACC
TGTAATCCCAGCTACTGGGGAGCCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGTGGA
GATTACAGTGAGCTGAGATGGCTCCACTGCACTCCAGCCTGGGCAACAGTGCTAGACTCC

69/112

FIG. 24 cont'd

ATCAAAAAAAAAAAGAAAGAAGGAAACAAAATATGTTCTGGCTCAGATAAAGCCCATGCCTAA
TTTTAAGAAATCCTAAACATAGATCCTCAAGTAG (SEQ ID NO: 138)

eHGT_830h (719 bp):

GCGTCTGTAATCCTAGCTACTTGCGAGGGAGGCTGAGACAGGAGAATCGCTTGAACCCAG
GAGGCGGAGGTTGCAGTGAGCTGAGATTGCGCCACTGCACTCTAGTCTGGGTGACAGAG
CAAGACTCCGTCTCAAAATAATAATAATAATAATAATAATAATAATAAGTAATAATTAT
GCTGAGACAACAGGCACAAAACAGAACTGTCCCAGGCAAACAGGGATATATGGTCACCCA
AGTTCTAACAGCCATGAAAACAAGCTCCACGTCCCTTTGAGACCTACATTGTTCTAGACAC
CAACCTAAAACACTTTTATGTGCGTAGTTCTGTAATCCCCACAATGACTGGGTAGTCATTAT
CTCATTTTACGTATGAGCAAGCAAAGCTGAAGCTCAGAGAGCGTAATAGCTACCATTCTTTTT
GGAGACAAGGTCTCACTCTCTTGCCCAGGCTGGAGTGTGGTGGCACAATCATGACTCACT
GCAGCCTCCACCTCTCAGGCTCAAGTGATCCACCCACCTCAGCCTCTCTGGTAGCTAGGA
CTACAGGTGCACACCACCACGCTGAAATAGGCCTCATAGTCCCATAGACAGTTGTTTTTTT
GTGGGCGTTTTTTTGTTAATTTTTTGTTTTTTGTTTTTTTAGATGGAGTCTCACTCTGTCGCCCA
GGCTGGAGTGCAGTAGTGTGATCTCAGCTTACCACAACCTCCGCCTC (SEQ ID NO: 6)

eHGT_831h (869 bp):

GTGGCTGAAGCCCTACTGCCAGGGTAACATTCTCCTTGCGGAGCCCAGCTGGGCGCCGC
CCCTTCTCCAGCAGCCTCCCCAGTTTTCTGGCGCACTTGACACGCGCTTACAGCAGAG
GCTGCCGTTTTCAAGCTGCGGGGACTGGGCATCCCCAGGTGTCCAGGGCTCCCGCTGCCC
AGGCTCTAAATCGAGCGCGCCCAAGGGGGTTCCCCCTCCCGCGCCCTGCCCGACCCGCT
GCCGCAGCTGCCGGCGGCTTGCGGACAGCGCGCGGCTGGGAAAGGCGGGCATGGGGCT
GTCCCGGGGGGCGAGGCGGGCTGGCCCGGGAGGGGGCAGGAACAGACACCGGCAGCGC
TGTGGGAGAGGAGAACCGAGAGGGCCTCCAGCCGAGCCCGGCTGCCAGCAGCCGGGAG
ACGAGCGAGCCAACCTTCTGGGGGACAGCTGGGCGCGCTGCGCCCCGCTGGCCCATCCC
TGGTGGGGAGGGGCGGGACGGGGGGCTGCCTGTCAATCAGCTCCCCAGGAAGCGCTCA
GAGCCAGCGCATCATCTCCGCGGCAAATTCGGCTCTAGAAATAACTTTTTTCTTTTTTCCC
CTCTTCTGATTCTCCTAGCAAGGGGCTCACCCCTTACTCACCCAGGCAGCTCATGGTGG
CCTCTTTTCGGTGGTGGTCGCCTATGCTCCGCTGCAGAGCCGAGAGCTGCGCCGGGGCTG
GGTGGGCGGGGGGCGGGGAGGGTGGCCAGCCAGTCCCTGACATTGGCCAGG
CCAGAGGGAGGGGCGAGGCTGCGGGGTGAATGGCAGCAAGGGGAGGGGGCGGGCTGA
GTCCGGGAAGAGGGAGGGGAGGGATGGAGGAGCAGTCGCAGGCCAG (SEQ ID NO: 7)

eHGT_832h (669 bp):

TCACCGTGTGAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCACCAGCCTCGGCCT
CCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGCGCCCGGCCATTTTTTAAGAGTGAGA
GTCTCTCTATCTTTTTATTTTATTTTATTTTATTTTATTTTAAAGTTCTGGGATACGTGTGCAGG
ATGTGGCAGGTTTGTACCTAGTTAAACATGCTCTATCTTTAAAAAATAAAAAATAAAGA
CCGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAACGTGGGCAGAT
CACCTGAGGTTAGGAATTCGAGACCAGCCTGACCAAGATGAAGAAACCCTTCTCTACTAAA
AATACAAAATTAGTCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGAGGCTGA
GGCAGAGGAATCGCTTGAACCCGGGATGTGGAGGTTGTGGTGAAGCCGAGATTGCACCATT
GCATTCCAGCCTGGGCAACAAGAGCGAACTCCATCTCAAAAAATAAAATAAATTTTTAA
AACTGGGCCCCGATGAGTCACTCCTGTAATCCCAGCACTTGGGAGGTCAAGGTGGGAGAAT
CACTTGACCCAGGAGTTCGAGACCAGCCTGGGCAACGTCTGACACCCTGTCTCTACA
(SEQ ID NO: 8)

70/112

FIG. 24 cont'd

eHGT_834h (546 bp):

CTATGAAGTGTGGAGTCTGAGAGGGCTGAGGATGGCATCAGTGCAGGGGAAGCGCACG
GTGCAGAAGCCAGAGCTCAAGTGTGGATGAGAGTGTCCATACTCTGGGGTGAAAAACAAGG
AAACAGCAATGACAGGAGATAGGTTGCATAGAGAAGGATGTATTTAATACTTTAAGGGTAA
CTGGAGAGAGATTTCCCACTGTCCAAGAAGGTGTTAAATGCAAATTGGAGGTGAACTCATG
GTCTGCAAAAATAAATATACATATATTCTCAAAGTTCTGTCTACTTAGAGGCCAGAAGAGCA
CAATAATAATGAGCACACATACACCCTCAATCTTGCTTCTAAATACCATTCTCTATTGAAAA
CAAGCAGTGCTCCTTAGAGAAATTACTGACTTCAGTACTGGGGCAGGAAAAAATGAGCCT
GGAACATCCCTTTCTGCCATAAACTAAAAAATTACTCAGAGATTGATGTGGTCTGCCATGAT
CTGAATGTGTCCCCCAAAATTTATATGTTGAACTTAATCCTCACTGTGGTGGTAT (SEQ
ID NO: 9)

eHGT_836h (643 bp):

CAGGTCAGTCAAATCTATCTCATGGAGTGATATAAATCCTACCTGTTTTTCCTGGTTTGAAT
TTTGCATACCATTTCGTTAATTCAAGTAATAGTTATTGGGCACATATGTTATGTGCCTGTCATT
GTTCTACATATAGGGAATATAGCATTAAACAGAACAGACAAAAATTCCTGCCTTTATGGAGT
TTACATCCTAATGAGGGAAACACACAGTAAATGAGAAATGTAAGTAAATAAGTATAATTTT
GATGGTTAAAGTGGTAAGAAGGAAAAGGAAGCATAGAAGGATAAAGGAGAATAAGAAATTA
TGGGGTCTTTTTTAAATAGAGTAGAAAGAGAAGACTACTGGGAATGTGCTACTTGAGTAAA
ATATTTGCAGAGGTTCAAGGCATGAGCTATGAAAACACTTGGCGGAAGAAAGCTCTAGAGAG
AGAGAAAAGTAGATGCAAATATCTAACAATGTGTAATCAGAGCTATCACATTATATAGACAT
CAAGTAACTTCTATGCTTAAATAAATTCTGAACTTTTGCAATTCTTCCAGAATAACATACAA
ATACTTTCTGTGTAATATCCTAAAGTTCATTTCAGGATTCGGTTCTGAGACCCCATTTCTCTT
AGACAGCATATTTCTGCCAGC (SEQ ID NO: 10)

MGT_E117 (627 bp):

CCGGTCCCGCCCCCACTCAGGACCCGCCCCACGGGCGCCCCGCCCCCTTCGAAGCTCCGCC
CACGTGAGCTGGGGGTCCAGCGCGTCTCAGCCTCTTTGAGTTGCAAGTCTTTGGGTCCCC
AGATCTGTTAGCCGTCTCTGGCTCTTGTGGCGCCTCGCCTGCCAGTGGCACGCAGGCCAT
CCCCTCTCAGGGGACTAATTTATAGAGGAGGACTCCGAGGCCAAGTTGGGGGGTGGGGTT
GAGCACTGCACACACAATGGAGCCGAGGAGGACCGACGCGGGCTTCCCCAACTACTTGC
GCCACACCCGGGACTCCCCACTGGCCAGCTGGCCCCGCTTGCCTGGCCTCATTCTCTT
CCTGCTGGCTTGAGCAGAGCCACGCCGCGCAGGCCTGCCGGTTCCAGTCTGCCCCGAGAG
AAGGAGCTGGGAGGTCTCCAGGGCGAGAGGTTCTAGCCAGTTTGTCTCAGCACTGCTCCT
CAGACAATTGTGGGAATTGCCTCTGCCCCCATCTCACTGTCTTTCTGAGCCACGGTTCA
GAGCTCCCCTTGCTCAGTCAAGCTCATAAGAGGGATTTGATTGTCAGCTGTAATCATTCA
CTCATTAAGGGTCCCCTGGGCCCAAAGT (SEQ ID NO: 11)

MGT_E118 (574 bp):

TTTTACAAGTGGACTATTTTAAGTTGTACATTTTTACTATACACTGTGGAGTGACTTAGCTT
AATCGTGGGCTTGGTGCCCTTTTGTGGTTATTCGTCTCTCATGACTGCCCTCTTGCCCTTCT
TCTGATTTAAGGCATGCAGTGTATGTGATTATCTTAGTTTCCTTACTGTTGCTCTGTGATAAA
ATACGCCGACAAAAGCAAGTTAAAGAGAGCTTATTTTAGCTGCCAGGTTTCAGCCCATGGC
TTTGAGGAAGTCATAGCCACACAGGAGCTTTGTGAGCCTTTTACACTGCACCGAATTAGTT
CATACCGCTACCCAGCTCTCTTTGACGCTTTACAAGACCCAGGAATACTTTGCCGAGGG
AATGGTGTCAATTCCTCAGTTAATCTAGTCTAATAACACCCCCCCCCGCTGTGGTCATGCTC
AGAGGCCTATTTATCAGACAGTTCTAGATTTTGTCCAGTGACAATTGGCATTGGCCATCAC
TGTGGCCTTTGTTTCGGTAATTTGTCAAACAAGTTCAAATGGTGTAGTTACTCAAGTCTTG
TTGCTTGCTCTATCTTTC (SEQ ID NO: 12)

71/112

FIG. 24 cont'd

MGT_E119 (621 bp):

CTGTCATAGCTCCCCGGGCGGTGGCTGTTCTTCTACTGCGGCGGCCGCTGCTGCCTGAATT
GGCTGTGGAGCTGCTCTGGTCTAAGCAGACATGGCTGTGGGAGCGAGGGAGGAGGAGGA
AGAGGAGGAGGCGGCGGCGGCGGGGAGGAGGAGGAGGAGGAGGCTGGGAGTCGT
GGAGGCTCTGTCTCCCGCAGGCTGACTGCTGGCTGAGGCGCAGCACAGCCTTGTTAAAT
CCTTAATTGCGCGCTTACGCACAGCGGGGTGGATCGGGTTCCATTGGCCAGGGCTTCGG
GAGCAGAGCAACACCCTAACTCGTCCAAGTAGTTTTAGCACAACAAAGCATTGCTTAAAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
GAGGCGTACAAGGAGAGAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAATATGAGTT
GTTCCCTGTTGGGGGAGGGAGGGTCTCAGTCGATCTCTTTTGTAGTAATAGCTTCGACAAA
CCAAGCAGTCCCAGGGCCTAGCTCCGT (SEQ ID NO: 13)

MGT_E120 (650 bp):

TTTCGCGCGCTCCTTCCCACCGACCTCCCAGGGGTGCACCTCTCCAGCCTCGGTGCGCGCT
TCCGAAACCTTTGGTGCCCCCTTTTCTGTGTCGCGACCCCCACCTCACGCCCCCTGGTC
TGGACAGCATCTCCCCCTCGCCGCCCTCCGCCACGCACCGCCTGACTCCGAGGGGTGC
GAGCGCATTGGGCTGCGCCCCGCGTGGGGGCGCGCGCCAGCCTCGCGTAGCTGTTCTG
ACGCTGCCGTGCGCGCGCCCTCCGCAGCCCAGCCGGCACCCGCACCAGCTCTGCAGTG
CACTCGTCGCCTCTCGGGCCGGTCCACCAAGAGCCAGACTGTGCTGACCGGGGCGCAGC
CTCGAACGTCAGGCGCGAGGGTTCATGAGCCAGAGCGCCCTGGGGCGCCGCGCGGAGAC
CCAGCGGAGATAGCAGTCTCGCTGCCTTGACGCGCGCCCGCCGCGTCCCCAGACCCTT
CTGTCCTTTTCTCCCGCCCCGCGCGCCGCGCATGGCCACCCTGCTCCGCAGCAAGCTGACCA
ATGTGGCCACCTCCGTGTCCAACAAGTCCCAGGCCAAGGTGAGCGGCATGTTTGCCAGGA
TGGGGTTT (SEQ ID NO: 14)

MGT_E121 (711 bp):

ACGCGACCCCCGAGCGGTGCGCGTCCCCCAACTTCTTAGAGGGACAAGTGCGGTTTCAGC
CACCCAGATTGAGCAATAACAGGTCTGTGATGCCCTTAGATGTCCGGGGCTGCACGCGC
GCTACAGTAGTGGCTCAGCGTGTGCCTACCCTACGCCGGCAGGCGCGGGTAACCCGTT
GAACCCCATTCGTGATGGGGATCGGGGATTGCAATTATTCCCATGAACGAGGAATTCCCA
GTAAGTGCGGGCCATAAGCTTGCGTTGATTAAGTCCCTGCCCTTTGTACACACCGCCCGTC
GCTACTACCGATTGGATGGTTTAGTGAGGCCCACGGCCCTGGTGGAGCGCTGAGAAGAC
GGTCGAACCTTGACTATCTAGAGGAAGTAAAAGTCGTAACAAGGTTTCCGTAGGTGAACCTG
CGGAAGGATCATTAACGGGAGACTGTGGAGGAGCGGCGGCGTGGCTCGCTCTCCCGGTC
TTGTGTGTGTCCTCGCCGGGAGGCGCGTGCCTCCCGGGTCCCGTCGCCCGCGTGTGGAG
CGAGGTGTCTGGAGTGAGGTGAGAGAAGGGGTGGGTGGGGTGGTCTGGGTCCGTCTG
GGACCGCCTCCGATTTCCCTCCCTCCCTCTCCCTCGTCCGGCTCTGACCTCGCCAC
CCTACCGCGGCGGCGGCTGCTCGCGGGCGTCTGCCTCTTCCCGTCCGGCTCTT (SEQ
ID NO: 15)

eHGT_717h (644 bp):

ACTTCTTAGGGCGAACGTTCTTTATCACAAGGAGATAAGCTGTAGGAAGTTTGGATATAGC
TTCTGGGATCCTTGCCCGTATCCCAACTTCCTTTTTTTGCTCCTACTTAGTAAGAATTCTAA
AGTCTGCTGGAGGATTGTTATTCAATGTAGGCACGATTGAAAGTAGAAACCATTTTTCTCCC
ATATTCTGTTTCCAATTTATACACCTACGTACGTTGTTCCAGCCTGCTTCTAGTCCATC
CCTTTTGTGAGTTCTGCTTTTCTAAAACCCTGATCTGCAGGGGTTAGAAAACCAGGAAC
TGCACAAAAGGGATGGCATCCCATTAGATCATGTGATCCTCCAGGCTCCAGTAATAGCCCT
TCCTCTGTCCTCATTAACCTACATCTCTCTGTGCCCTTCTTCAAAGTACTGCCCCCTATA
TGGGCCACTGGCAAATTGATTTAAATTAAGTGTTAATAGCATGTAGCAGCCCTGCTTCCCT

72/112

FIG. 24 cont'd

TCTTAAATAATGTTACATATTACATCTGCCACCAACTTGGCTTCTCCTGTTTGCATCTATAT
GTCTTTTTTTTTTTTTTTTTTAAATGTTAGGAAGCGTATTACCTTTGGTTTTGGATAAAATCTCTA
CCAGTGTCTCTTTTCATTCAATGT (SEQ ID NO: 16)

3xCore2-eHGT_367h (480 bp):

TCCAGCTTAAAGGAAAATTCCAGCCCCAGAGATCTTCCGCAGTGAGTGACAGCTCCTTAAT
GTGAGGAGCAGCGTGGGCGTGGTGGGTGGGATGTCAGAGCCATCAAAGCCCTCATTTCT
GATCCCTCTGTGATGTCGGGGTTCCCTGGCAGGCCCACTGTCCAGCTTAAAGGAAAATTC
CAGCCCCAGAGATCTTCCGCAGTGAGTGACAGCTCCTTAATGTGAGGAGCAGCGTGGGCG
TGGTGGGTGGGATGTCAGAGCCATCAAAGCCCTCATTTCTGATCCCTCTGTGATGTCGGG
GTTCCCTGGCAGGCCCACTGTCCAGCTTAAAGGAAAATTCAGCCCCAGAGATCTTCCGC
AGTGAGTGACAGCTCCTTAATGTGAGGAGCAGCGTGGGCGTGGTGGGTGGGATGTCAGA
GCCATCAAAGCCCTCATTTCTGATCCCTCTGTGATGTCGGGGTTCCCTGGCAGGCCCACT
G (SEQ ID NO: 17)

Core of eHGT_367h:

TCCAGCTTAAAGGAAAATTCCAGCCCCAGAGATCTTCCGCAGTGAGTGACAGCTCCTTAAT
GTGAGGAGCAGCGTGGGCGTGGTGGGTGGGATGTCAGAGCCATCAAAGCCCTCATTTCT
GATCCCTCTGTGATGTCGGGGTTCCCTGGCAGGCCCACTG (SEQ ID NO: 18)

eHGT_359h (485 bp):

TCACAGCTCCTTGGTCTGAACTGGTACCCAGGATCAAGATGTCCCTCCCATGATTCTAATC
TCCAGCCCCCTTACACCATAGTCACTTCCAACTCTGGATTTCTTAACATTGCAGTTTGCTT
CCAAAAAAAATTAATAAAAAAACCTCATCTGGGCATTAGAAGTCATTCTGAAGAGGCTGTC
CTATATCTCAGTGCAATTTCTACTTATCCACTCCCGACTATACCCTGCTTTCATCAGATTG
CAACTGACCTAATTTGTAAAGACTTCAGTGAAAAGAGCCTCATTAAAAGCCTTCATCTCTTT
TGCAGCCTCTCTTTCCCTCCCTCAAATGATCCAGGTCAGAATGGCCCAATAAAAATAGCTGG
GAAAATTACAAAGTCACTCAAGACTCTGCCAGGAGTAGAGATTTTAAGAATTAGTGACAGA
CATATACTGTCTAGGAGTAGGGTGGGAGTGGGTTGCATGGCTTAGGGACATAGA (SEQ ID
NO: 19)

eHGT_479m (808 bp):

NCCTGCGGCCGACGCGTTCCCTAACCTAGAAAAACCCACCATCAACTATTGAGGTAATGT
TCCNTAGATGCAGGGTCATCAGTGGGCATTGTTACTTACTGGAAGAATGCTCTCTGGCGGA
TGGGGGAGGGAAGCAGGAAAACCTCTGCGTTAAGCAGGGCATAATCTCTTCAGGGACCTGC
TTCAGTTTCAATCCGACTCAAGTCGTGCTTTGCAGCTGGCCCCAGTTTTGGCTGCTCTGAG
AGTCAGTTGTTGGTTATGCACTGCCTGGGGTGAGGGCCAACCAGGGAGGGGGTTGGAGA
AGGATGCGAATAGTCAGAGCAGCTGGGAGCAAATGTATCCAGTTAAGGGGATCAAACAG
AACGTCAGCAACCCCCGTAATGCCCAAATCTTGCACTAATCTGTGTTATCTAGGCCATCT
CCCCCAACCCACCCCGCTTCTCATTTGGAACAAGGTCTTTCTATGTATGTAGCTAAAGTT
GACCTTGAACCTCATGATCTTCCCTGCCAGCCTCTGAGTGCTAGGGTAAAATAGGCGTGCA
CCATCCTGCCGGGCTTCACCCACTCCCTTTTCAATTTCTGAGACACTTCCAGTTATTCTTCCTC
ATTCTTACTCCATTCACTCATACTGACCACTTCAAATTTCTGGAACCTTCCAGCATGCACTG
GCTGCAGATGAATTGTGCTTTTCTGCTCCCTCATCTGACTCTCCTTCTCGTCTGACTACTTA
GCTCTCATCTTCTCTACGGAGTCACCAGTCTTTTTTTTTTTTTTTTCAAGACAGAGTTTCTC
TGAATAGCCCTGGC (SEQ ID NO: 20)

eHGT_453m (721 bp):

TCACAAATCAGTAAGAATTACACTCCAAAAATAAAACAGGGAAGGGTATTAGTATGTGATAC
GCATGGATAAAGTAGAAGAGGCCAAAAATACAAAGAACCGTAAACTCACTAGTGATGGAAG

73/112

FIG. 24 cont'd

AAACGCTAATTTAAATAATGAGCTTTTCTGACCCTTTATCACATGGGCACATTTTCAAAGGA
 CAAGACTCTGCTGAGGAGAACACAGAACAGGATCCTTCAGACAAGCTTCAGCAACCAATTT
 TAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTAAAAATAACATAGATCTGAAAAGC
 TTTATAACTTTTTGGTTTTTTTTTAAGGACCATATTAACAGCATCTTCCAGTTTTTCTAGAGTAA
 ATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTGTAAAAAGACTATAAA
 AATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGCCTTCAGC
 TCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTCACTGCCCCTGCCTTTGT
 GTGCCCTTGCCTGTCCCTCTGTGACGGTCAACCCTTTCGGTTCTGGCTTGAGCCTGGTCA
 TCAGTGCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTCGTTGTCCTAGTGA
 TGTGACCCTACCTTAGACTTCCCCATCTCTTGAGGTGGCTA (SEQ ID NO: 21)

eHGT_140h (304 bp):

TGTGTGTGAAGATTAAGCCCTTTCTTGCTCTATGTATTTGAGCCTAGAAAAAGACACACTCC
 CAAGAAATTAGAATAAAGGGAAAAATGACTTTGAGATGTACTTTCTTAAATAACTGAAGTA
 ATGACCGCATGAGCCAGCTGATGGGTTTTTAAATGGATACGTTTCTATCAGCCTGCTGCTT
 GATGCCAGAGCCAAATATATATGGAGTGTATTATATCATATCCTGTAGGCAGGAGACTGTT
 GGGGCATTGGGGGACTTAGAGAGGTGAAGTGGCAGGCTTGGCACAGGAATTAACAGCA
 (SEQ ID NO: 22)

eHGT_356h (564 bp):

AGGTGGAGTGAAGACTGAGAGGGGTGGTAGGGGGAGATGAGGCCGGGAGGAGGAGAGT
 ACACCCTCAGGGATGGAGAGCCAAACAAAGATTCTAAACGAGGAGTTCTGCAGTGAGACTG
 GCATTTTGGAACAATCGTCCTGGTTTTGGTTGAGAGCACAACCTTAGAGAACAGAGGCTAGA
 GGTGGCAGGCTGGAGTGACAGCCTACTGCAGCTGTCCAGCCAATTAAGCCCTGTAGGAAG
 AATGGCTGATGCCGCAGTCTATGTGATGCTGAACTGCGGTAGCCAGGCTGACCTGGCCCC
 AATCCCGGGCTGTGACTGACCTTCAGATGCAGGCAGCAGAGAGAAGAGGACAAAAGTGCA
 CCGCTGGGGAAAGGCAAGCAGCTGGCACCCAGGGCACCAAAGACCCGCACAGTGAGCAT
 CAAGGCCAGCAAGGGCACAAAGCCAGCTGTACCTGTCTCTGGGCCAAGCCAAGGGGG
 GCTGGCCTCCTGGCCACCTCCTTGGGGGAGCAGTACCCCCATTGTCATGACTTGTTTACA
 GATTCCCAGACCTCCCAGTGCTATCAA (SEQ ID NO: 23)

eHGT_128h (428 bp):

CTAGTATCATGAGGAGGATCTCAAAGCAGGGAAGGGAAGCCAGCTGAGGGACCCCTGAG
 AGGAGATCAGGGGATAAATACTTTGCAACAGAATCCCCATGGGAAACATGACAAAATTAGT
 AACAGGAATGGAGCCTCCCTCAGAGGATACATGAATCGTTGAGTTGAAAATAACCACAAGG
 TCATCAAATCCAGCTGCCACCAACTGCCTGAATATTTTCAACAAATTTTTGAACAAGCAGTC
 CTCCAGCCTCCTCCAGCTGCATCCAGCAGCTCAGTGCTCCACCTGACAGCCTCTTCTGATA
 GCGTGAAGGCTCAGAAAGCCCCTGTGGGGCCCAGCACATCCTGGGGCAGGCCTCTAGTG
 GGCAGAGGGCAGCATACTGGGGTCTGGATCTAACACACTCCTCCATCCAGGTGGTATATT
 TATAG (SEQ ID NO: 24)

eHGT_369h (also referred to as 3xcoreB_eHGT121h) (450 bp):

ATGAGAGGCTAGAGTTGGAGAGCACAGTTTCTGCTTCCAAATTGGCTTGAAGTGGGGACA
 TCTGCTCTGCATGAAGCCTGGGTCAGGCTGCCCTCTCATCTGTGACCTCCTGTGGTTCGG
 GAGGGACTTGCTCTGGATTATATAAAGGCTATGAGAGGCTAGAGTTGGAGAGCACAGTTC
 CTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTTCAGGCTG
 CCCCTCTCATCTGTGACCTCCTGTGGTTCGGGAGGGAGGCTTGTCTGATTATATAAAGGCTA
 TGAGAGGCTAGAGTTGGAGAGCACAGTTTCTGCTTCCAAATTGGCTTGAAGTGGGGACAT
 CTGCTCTGCATGAAGCCTGGGTTCAGGCTGCCCTCTCATCTGTGACCTCCTGTGGTTCGGG
 AGGGACTTGTCTGGATTATATAAAGGCT (SEQ ID NO: 25)

74/112

FIG. 24 cont'd

coreB_eHGT121h (150bp):

ATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACA
TCTGCTCTGCATGAAGCCTGGGTCAGGCTGCCCCTCTCATCTGTGACCTCCTGTGGTCGG
GAGGGACTTGCTCTGGATTATATAAAGGCT (SEQ ID NO: 26)

300 bp core of eHGT_369h (2xcoreB_eHGT121h):

ATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACA
TCTGCTCTGCATGAAGCCTGGGTCAGGCTGCCCCTCTCATCTGTGACCTCCTGTGGTCGG
GAGGGACTTGCTCTGGATTATATAAAGGCTATGAGAGGCTAGAGTTGGAGAGCACAGTTC
CTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTCAGGCTG
CCCCTCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGGATTATATAAAGGCT
(SEQ ID NO: 27)

eHGT_710m (382 bp):

AGTGACTTGGTGCTATGAGCCATATTTTGCTGTTGCTGTTGTTACTGGTAGTTTTTGTAATT
CTGGGGCTAAACTTGGGGTCTGGTATGCTGTCAATTTACCAAGTGAGCTATACCCTGGATAT
TATGATTTAGATGAATGTGAAATATCACCCAGACATACATACTAAACACTTGGCCCTTG
GCCCATGATGCTAAATGGAGGAGATAGAAGCTTTTGGGGCACAGCCTAGTGGGAAGGAAAT
GAGGTCAAATGACATGTAATCTGAAAGGAATATGGGTATTCTGGGCTTGCCTTATTCTCTCT
CTCCCTCTCTCTCCCTCTCTCTCCCTCTCTCTTTCTCCTTTTCTCTTTCTCTCCTCG
CCTTGTTTTCCA (SEQ ID NO: 28)

eHGT_895h (830 bp):

TTCTTTGTGCAATGGGGAGCACAGAGCTGGATATGCAGCAGTAAATCATGACACAATTATT
GAAGCCCAGCCAGTGCCCCCTGGCACATCAGCACAAAAGGCTAAAATCATTGATCTTACTC
AAGCATTAATGTTGGGACAAGGGAAAAAACTTAACATCTATACAGATTCTAAATATGCATTC
CTTGTGGTTCATGCTCATACCGCCATCAGGAAAGAAAGGGGATTACTTACTAGCAAGCACT
CCCCTATAGAGCATGGACCTGAAATTCTGCAGCTATTGAAGGCAATACACCTGCCAAAGGC
CATAGCTATCATTCAATTGTAGGGGGCATCAAAGGGACTTAACCCCTATAGCACAAAGGGAAT
GGAAAGGCTGACGGAAGCCAAAGCAAAGCCCTCAGCGTGCAATCCCAACAGATCCTAGC
ACTGCTTCCTTTCTACAATTCCCCAGTAGAACCCCAAATACACACCACAGGAAGAAGAGCTA
ATAAAAAAAGCAAGGGAGACAAAAACAAGGATCCTGGTGGCATATGGAGCCAAAAATATAT
CTCCCTCAGGCAGCCCAGTGGAGAGTTATAAAAGCCCTGCATGACTCTTTCCATGTGGGG
AGGGATGCGGCTCTGGCCATGGTAAACAAGTTCTTTACTGGGTTTAACTCAGCTTCAATGG
CTAAGCAGGTCTGTCAAGCCTGCTCACTGTGTGCATTTAACAATCCAGGCATCAAAATGCC
CCCTCTAATAAAACCAGTCCAAAGGAGGGGAAGTTATCCAGGGGAAGACTGGCAGTTAGA
CGTCATCCATATGCCAGCTTGCAGAGGATACAAGCTTTT (SEQ ID NO: 29)

3xcore2_eHGT_577h (876 bp):

ATGTTGGAGGCTTTTGAACAAAAGGACTTCTCTTTAGCCCTGAAGTATCAGGTAAACGTTTT
CTTCTCTTCCCATAAATCACAGCCTTTTTTCTTCCCCACTTGAGGATTCTTTTTCTTCTCTCT
ACCATGCCTGCCCTCATTAAATTTGCCAGCCACTTCTTTCTTTACCTTTCTTCTGTTTTGTA
CTTTTTCTTGTTTAAACCAAAGCTGTCCCCATTGTCTGTGTTCCCGACATGTCTAATCAGG
ACAGGACCCATACAGGGATGGCTCCTCTCTTCCCTCCTAGAAAAATGTTGGAGGCTTTTGAAC
AAAAGGACTTCTCTTTAGCCCTGAAGTATCAGGTAAACGTTTTCTTCTCTTCCCATAAATCA
CAGCCTTTTTTCTTCCCCACTTGAGGATTCTTTTTCTTCTCTCTACCATGCCTGCCCTCATT
AAATTGCCAGCCACTTCTTTCTTTACCTTTCTTCTGTTTTGTACTTTTTCTTGTTTAAACCA

75/112

FIG. 24 cont'd

AAGCTGTCCCCATTGTCTGTGTTCCCGACATGTCTAATCAGGACAGGACCCATACAGGGAT
GGCTCCTCTCTTCCTCCTAGAAAATGTTGGAGGCTTTTGAACAAAAGGACTTCTCTTTAGCC
CTGAACTATCAGGTAAACGTTTTCTTCTCTTCCCATAAATCACAGCCTTTTTTCTTCCCCACT
TGAGGATTCTTTTTCTTCTCTCTACCATGCCTGCCCTCATTAAAATTGCCAGCCACTTCTTT
CTTTACCTTTCTTCTGTTTTGTACTTTTTCTTGTTTAAAACCAAAGCTGTCCCCATTGTCTGT
GTTCCCGACATGTCTAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCCTCCTAG
AAA (SEQ ID NO: 30)

core2_eHGT_577h (292 bp):

ATGTTGGAGGCTTTTGAACAAAAGGACTTCTCTTTAGCCCTGAACTATCAGGTAAACGTTTT
CTTCTCTTCCCATAAATCACAGCCTTTTTTCTTCCCCACTTGAGGATTCTTTTTCTTCTCTCT
ACCATGCCTGCCCTCATTAAAATTGCCAGCCACTTCTTTCTTTACCTTTCTTCTGTTTTGTA
CTTTTTCTTGTTTAAAACCAAAGCTGTCCCCATTGTCTGTGTTCCCGACATGTCTAATCAGG
ACAGGACCCATACAGGGATGGCTCCTCTCTTCCTCCTAGAAA (SEQ ID NO: 31)

3xcore3_eHGT_577h (873 bp):

TAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCCTCCTAGAAAGGATCTTATTT
TACTCTATCTTCAAAGTACTTTTGTGACTATAATAAAATGGTCATAGATTTCTGGGAAAATG
ACCACAGGCTGTCCCCCTTGAGGACATCACATCAGAGTTATCACAGAATATGAGGAAGCTAT
GAAGATGGTAACCTAAAATGGGGAAGGCTTCATTCAAAGTCAGGGCAGGTCAGTCTTGTC
ACCCTGGAAATGTGCTGAGAACTAGGAACATCCACCAGGAAGGAATAATCAGGACAGGAC
CCATACAGGGATGGCTCCTCTCTTCCTCCTAGAAAGGATCTTATTTTTACTCTATCTTCAA
GTACTTTTGTGACTATAATAAAATGGTCATAGATTTCTGGGAAAATGACCACAGGCTGTCCC
CTTGAGGACATCACATCAGAGTTATCACAGAATATGAGGAAGCTATGAAGATGGTAACCTA
AAATGGGGAAGGCTTCATTCAAAGTCAGGGCAGGTCAGTCTTGTCACCCTGGAAATGTG
CTGAGAACTAGGAACATCCACCAGGAAGGAATAATCAGGACAGGACCCATACAGGGATGG
CTCCTCTCTTCCTCCTAGAAAGGATCTTATTTTTACTCTATCTTCAAAGTACTTTTTGTGACTA
TAATAAAATGGTCATAGATTTCTGGGAAAATGACCACAGGCTGTCCCCTTGAGGACATCAC
ATCAGAGTTATCACAGAATATGAGGAAGCTATGAAGATGGTAACCTAAAATGGGGAAGGCT
TCATTCAAAGTCAGGGCAGGTCAGTCTTGTCACCCTGGAAATGTGCTGAGAACTAGGAAC
ATCCACCAGGAAGGAA (SEQ ID NO: 32)

core3_eHGT_577h (291 bp):

TAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCCTCCTAGAAAGGATCTTATTT
TACTCTATCTTCAAAGTACTTTTGTGACTATAATAAAATGGTCATAGATTTCTGGGAAAATG
ACCACAGGCTGTCCCCCTTGAGGACATCACATCAGAGTTATCACAGAATATGAGGAAGCTAT
GAAGATGGTAACCTAAAATGGGGAAGGCTTCATTCAAAGTCAGGGCAGGTCAGTCTTGTC
ACCCTGGAAATGTGCTGAGAACTAGGAACATCCACCAGGAAGGAA (SEQ ID NO: 33)

3xcore2_eHGT_606h (582 bp):

TGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGAGCAAGAGAGTTAAGAATAGGTGT
AGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCATTTTTAGGCCAATATAAGGACTG
GCATTTTTAGGTCATGATGAGAACTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAG
CATTTTGAGCAGTGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGAGCAAGAGAGTT
AAGAATAGGTGTAGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCATTTTTAGGCCA
ATATAAGGACTGGCATTTTTAGGTCATGATGAGAACTGGCATTACTCTGCCTAAAATAGGA

76/112

FIG. 24 cont'd

AACCACTGGAGCATTGAGCAGTGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGA
GCAAGAGAGTTAAGAATAGGTGTAGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCA
TTTTTAGGCCAATATAAGGACTGGCATTGAGTATGATGAGAACTGGCATTACTCTGC
CTAAAATAGGAAACCACTGGAGCATTGAGCAG (SEQ ID NO: 34)

core2_eHGT_606h (194 bp):

TGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGAGCAAGAGAGTTAAGAATAGGTGT
AGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCATTGAGTATGAGTATGAGTATGAGTAT
GCATTTTATAGGTATGATGAGAACTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAG
CATTTTATGAGCAG (SEQ ID NO: 35)

3xcore3_eHGT_606h (582 bp):

CTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAGCATTGAGCAGAGGAGTTGAAC
TGTCTGATTTAAAATTTTAAAAGGTTCTGGCTACCTTGTGAGAAGAGACTGAAGTAAGATAT
GCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAATAATCTAGGCAAGAGAGATGGATC
AGGGTGATAACTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAGCATTGAGCAGAG
GGAGTTGAAGTGTCTGATTTAAAATTTTAAAAGGTTCTGGCTACCTTGTGAGAAGAGACTGA
AGTAAGATATGCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAATAATCTAGGCAAGAG
AGATGGATCAGGGTGATAACTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAGCATT
TTGAGCAGAGGAGTTGAAGTGTCTGATTTAAAATTTTAAAAGGTTCTGGCTACCTTGTGAGA
AGAGACTGAAGTAAGATATGCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAATAATCT
AGGCAAGAGAGATGGATCAGGGTGATAA (SEQ ID NO: 36)

core3_eHGT_606h (194 bp):

CTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAGCATTGAGCAGAGGAGTTGAAC
TGTCTGATTTAAAATTTTAAAAGGTTCTGGCTACCTTGTGAGAAGAGACTGAAGTAAGATAT
GCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAATAATCTAGGCAAGAGAGATGGATC
AGGGTGATAA (SEQ ID NO: 37)

core4_eHGT_577h (400 bp):

CAGGTAAACGTTTTCTTCTCTTCCCATAAATCACAGCCTTTTTCTTCCCCACTTGAGGATTC
TTTTCTTCTCTCTACCATGCCCTGCCCTCATTAAAATTGCCAGCCACTTCTTTCTTTACCTT
TCTTCTGTTTTGTACTTTTTCTTGTTTAAAACCAAAGCTGTCCCATTTGTCTGTGTTCCCGAC
ATGTCTAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCTCCTAGAAAGGATCT
TATTTTACTCTATCTTCAAAGTACTTTTGTGACTATAATAAAATGGTCATAGATTTCTGGGA
AAATGACCACAGGCTGTCCCCTTGAGGACATCACATCAGAGTTATCACAGAATATGAGGAA
GCTATGAAGATGGTAACCTAAAATGG (SEQ ID NO: 38)

core6_eHGT_606h (264 bp):

TGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGAGCAAGAGAGTTAAGAATAGGTGT
AGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCATTGAGTATGAGTATGAGTATGAGTAT
GCATTTTATAGGTATGATGAGAACTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAG
CATTTTATGAGCAGAGGAGTTGAAGTGTCTGATTTAAAATTTTAAAAGGTTCTGGCTACCTTGT
GAGAAGAGACTGAAGTAAGA (SEQ ID NO: 39)

77/112

FIG. 24 cont'd

3xcore_eHGT_121h (900 bp):

CTATCTGGTTTATAATTAGAAGTGAAATGATTTTAAATGGCTTCCACATGGCTTAGTTTTCCA
AGCAGTTCCAGTTACATAGA AACTTCATATAAACAGATAATGAGAGGCTAGAGTTGGAGAGC
ACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTC
AGGCTGCCCCCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGGATTATATAA
AGGCTTCACAAGAGTACTGGAAGCGGGGACGGAATCAACTTGAGCAAAAAACCTCTATCT
GGTTTATAATTAGAAGTGAAATGATTTTAAATGGCTTCCACATGGCTTAGTTTTCCAAGCAG
TTCCAGTTACATAGA AACTTCATATAAACAGATAATGAGAGGCTAGAGTTGGAGAGCACAGTT
CCTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTCAGGCT
GCCCCCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGGATTATATAAAGGCT
TCACAAGAGTACTGGAAGCGGGGACGGAATCAACTTGAGCAAAAAACCTCTATCTGGTTTA
TAATTAGAAGTGAAATGATTTTAAATGGCTTCCACATGGCTTAGTTTTCCAAGCAGTTCCAG
TTACATAGA AACTTCATATAAACAGATAATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGC
TTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTCAGGCTGCCCC
TCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGGATTATATAAAGGCTTCACA
AGAGTACTGGAAGCGGGGACGGAATCAACTTGAGCAAAAAACCT (SEQ ID NO: 40)

core_eHGT_121h (300 bp):

CTATCTGGTTTATAATTAGAAGTGAAATGATTTTAAATGGCTTCCACATGGCTTAGTTTTCCA
AGCAGTTCCAGTTACATAGA AACTTCATATAAACAGATAATGAGAGGCTAGAGTTGGAGAGC
ACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTC
AGGCTGCCCCCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGGATTATATAA
AGGCTTCACAAGAGTACTGGAAGCGGGGACGGAATCAACTTGAGCAAAAAACCT (SEQ ID
NO: 41)

eHGT_590m (259 bp):

GTGTCCTTAGCCATTTTCAACCTCAGAGAAACACTCATTGGAAACATCTGAAATTGCCAGAT
GCCCTGAGAATTTTATTTTTAAAAACCTTTTCCCAAGAAAGTTATTTATAATACTGTGTCAAAA
TGGGACACTTCAACTGTGAGTGCTAATTCCTGATGTATGTTATCATTTTCTCCATTTGTCATA
TTTTTTCTAAAGGGCATTAAATTTAGCTTTTGACTCTAAATTTAAACCCCTCTATTCTGTCT
CTCATTTG (SEQ ID NO: 42)

eHGT_976h (442 bp):

GGTTCACCTGGATTTGAATGGCGTTTGAGAAGGGGAGAAGGTGGCGCAAGTGAACCAAAGC
AACTTAACCTCCATCTGCCTTCGATTCCAAGGCTGAGAAAACATTTCCAGAAGGCCTGGTGC
ATAAGAACCTGATGTGGTGACAGGATGTTCTGTCTTCTCTTGTACACAAGCAAGCGCGC
TGAGTCTCGGCAGGCTCATAAATCCTGAGGGCCTTTTCCAGCAGCCCCGTTTCGGGTGCGC
TACCCCCATTCCAGGAAAGGGACGCTTTGAAACCTCATCACAAGAGAACATTGTGTCTTT
GAACATCCGGCGCCACTGGGGTTTGACTTCTCCAGCGCAGGTAGGATGGGCCAGCCAG
TGAGGTCAGATGTGATTTATGCTTCTAAAGCTCTTGCTTTTCCAGAAATTTTGAAGAAAATAC
AAGCTGCCCAATACTGCA (SEQ ID NO: 43)

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

GGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTGCTTCTG (SEQ ID
NO: 45)

78/112

FIG. 24 cont'd

minCMV Promoter:

GAGGTAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGG (SEQ ID NO: 46)

Mutated minCMV Promoter (Sacl RE site removed):

GAGGTAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGCCTGG (SEQ ID NO: 47)

minRho Promoter:

GATTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTGAGGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCC (SEQ ID NO: 48)

minRho* Promoter:

GATTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTGAGGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCC (SEQ ID NO: 49)

Hsp68 minimal Promoter (proHsp68):

CAGGAACATCCAACTGAGCAGCCGGGGTCCCCCCCACCCCCACCCCCGCCCCACGCGGCAACTTTGAGCCTGTGCTGGGACAGAGCCTCTAGTTCTCTAAATTAGTCCATGAGGTCAGAGGCAGCACTGCCATTGTAACGCGATTGGAGAGGATCACGTCACCGGACACGCCCCCAGGCATCTCCCTGGGTCTCCTAACTTGCGGGGAGAAAGTTTTAGCCCTTAAGTTTTAGCCTTTAACCCCATATTCAGAACTGTGCGAGTTGGCGAAACCCCAAAATCACAACTGTACACAACACCGAGCTAGAGGTGATCTTTCTTGTCATTCCACACAGGCCTTAGTAATGCGTCGCCATAGCAACAGTGTCACTAGTAGCACCAGCACTTCCCCACACCCTCCCCCTCAGGAATCCGTACTCTCCAGTGAACCCCAAGAACCTCTGGAGAGTTCTGGACAAGGGCGGAACCCCAAACTCCGATTACTCAAGGGAGGCGGGGAAGCTCCACCAGACGCGAACTGCTGGAAGATTCTGGCCCCAAGGCCTCTCCGGCTCGCTGATTGGCCCAGCGGAGAGTGGGCGGGGCGCGTGAGACTCTTAAAGCGCGAGGCGGCGAGCAGGTCAACGAGCTGACAGCTACTCAGAACCAATCTGGTTCTCAGAGACAAAGCAGAGCAAGAGAGAGAGAGCGAGCGGCGCGTTCCCGATCCTCGGCCAGGACCAGCCTTCCCCAGAGCATCCCTGCCGCGGAGCGCAACCTTCCCAGGAGCATCCCTGCCGCGGAGCGCAACTTTCCCCGGAGCATCCACGCCGCGGAGCGCAGCCTTCCAGAAGCAGAGCGCGGCGCC (SEQ ID NO: 50)

SYFP2:

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCACTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATTCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGCAACAAGCTGGAGTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA (SEQ ID NO: 51)

79/112

FIG. 24 cont'd

EGFP:

ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGGCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGTTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCA
CCCTCGTGACCACCCTGACCTACGGCGTGCAAGTCTTCAGCCGCTACCCCGACCACATGA
AGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCT
TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAA
GAACGGCATCAAGGTGAAGTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCT
CGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACA
ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACA
TGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAA (SEQ ID NO: 52)

Optimized Flp recombinase (FlpO):

ATGGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTTCGACATCCTGTGCAAGACCCCC
CCCAAGGTGCTGGTGCGGCAGTTCTGTGGAGAGATTTCGAGAGGGCCAGCGGCGAGAAGAT
CGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGATGATCACCCACAACGGCACCG
CCATCAAGAGGGCCACCTTCATGAGCTACAACACCATCATCAGCAACAGCCTGAGCTTCGA
CATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGACCCAGAAGGGCCACCATCCTGGAGGC
CAGCCTGAAGAAGCTGATCCCCGCCTGGGAGTTACCATCATCCCTTACAACGGCCAGAA
GCACCAGAGCGACATCACCGACATCGTGTCCAGCCTGCAGCTGCAGTTCGAGAGCAGCG
AGGAGGCCGACAAGGGCAACAGCCACAGCAAGAAGATGCTGAAGGCCCTGCTGTCCGAG
GGCGAGAGCATCTGGGAGATCACCGAGAAGATCCTGAACAGCTTCGAGTACACCAGCAGG
TTCACCAAGACCAAGACCCTGTACCAGTTCCTGTTCTGGCCACATTCATCAACTGCGGCA
GGTTCAGCGACATCAAGAACGTGGACCCCAAGAGCTTCAAGCTGGTGCAGAACAAAGTACC
TGGGCGTGATCATTCAAGTGCCTGGTGACCGAGACCAAGACAAGCGTGTCCAGGCACATCT
ACTTTTTTCAGCGCCAGAGGCAGGATCGACCCCTGGTGTACCTGGACGAGTTCCTGAGGA
ACAGCGAGCCCGTGTCTGAAGAGAGTGAACAGGACCGGCAACAGCAGCAACAAGCAG
GAGTACCAGCTGCTGAAGGACAACCTGGTGCGCAGCTACAACAAGGCCCTGAAGAAGAAC
GCCCCCTACCCCATCTTCGCTATCAAGAACGGCCCTAAGAGCCACATCGGCAGGCACCTG
ATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAGCTGACAAACGTGGTGGGCAACTGG
AGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCACCTACACCCACCAGATCACCGCCAT
CCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTACGCCTACGACCCCATCAGCAAGGA
GATGATCGCCCTGAAGGACGAGACCAACCCCATCGAGGAGTGGCAGCACATCGAGCAGC
TGAAGGGCAGCGCCGAGGGCAGCATCAGATACCCCGCCTGGAACGGCATCATCAGCCAG
GAGGTGCTGGACTACCTGAGCAGCTACATCAACAGGCGGATCTGA (SEQ ID NO: 53)

Improved Cre recombinase (iCre):

ATGGTGCCCAAGAAGAAGAGGAAAGTCTCCAACCTGCTGACTGTGCACCAAAACCTGCCT
GCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAGGAAGAACCTGATGGACATGTTTCAGG
GACAGGCAGGCCTTCTCTGAACACACCTGGAAGATGCTCCTGTCTGTGTGCAGATCCTGG
GCTGCCTGGTGAAGCTGAACAACAGGAAATGGTTCCCTGCTGAACCTGAGGATGTGAGG
GACTACCTCCTGTACCTGCAAGCCAGAGGCCTGGCTGTGAAGACCATCCAACAGCACCTG
GGCCAGCTCAACATGCTGCACAGGAGATCTGGCCTGCCTCGCCCTTCTGACTCCAATGCT
GTGTCCCTGGTGTGAGGAGAATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAA
GCAGGCCCTGGCCTTTGAACGCACTGACTTTGACCAAGTCAGATCCCTGATGGAGAACTC
TGACAGATGCCAGGACATCAGGAACCTGGCCTTCTGGGCATTGCCTACAACACCCTGCT
GCGCATTGCCGAAATTGCCAGAATCAGAGTGAAGGACATCTCCCGCACCGATGGTGGGAG

80/112

FIG. 24 cont'd

AATGCTGATCCACATTGGCAGGACCAAGACCCTGGTGTCCACAGCTGGTGTGGAGAAGGC
CCTGTCCCTGGGGGTTACCAAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGCTGA
TGACCCCAACAACACTACCTGTTCTGCCGGGTGAGAAAGAATGGTGTGGCTGCCCCCTTCTGC
CACCTCCCAACTGTCCACCCGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGAT
CTATGGTGCCAAGGATGACTCTGGGCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAG
AGTGGGTGCTGCCAGGGACATGGCCAGGGCTGGTGTGTCCATCCCTGAAATCATGCAGG
CTGGTGGCTGGACCAATGTGAACATTGTGATGAACTACATCAGAAACCTGGACTCTGAGAC
TGGGGCCATGGTGAGGCTGCTCGAGGATGGGGACTAA (SEQ ID NO: 54)

SP10 insulator (SP10ins):

GAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAG (SEQ ID NO: 55)

3xSP10ins:

GAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGGAAGCTACCCC
TAACACACTATTCTACACACAGAAAATGCTCTTCACTAGGAAGCTACCCCTAACACACTATT
CTACACACAGAAAATGCTCTTCACTAG (SEQ ID NO: 56)

WPRE3:

ATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTC
CTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG
GCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCGC
CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGG
(SEQ ID NO: 57)

WPRE:

GCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTA
TGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAG
TTGTGGCCCCGTTGTGAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCC
ACTGTTTGGGGCATTGCCACCACTGTCAGCTCCTTTCCGGGACTTTGCTTTCCCCCTCC
CTATTGCCACGGCGGAACATCATCGCCGCTGCTGCTTCCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCATCGTCCTTTCTTGGCTGC
TCGCCTATGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCC
TCAATCCAGCGGACCTTCCCTTCCCGCGGCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTC
TTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCATCGATACC
G (SEQ ID NO: 58)

BGHpA:

CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGAC
CCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGT
CTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGGAGGA
TTGGGAAGACAATAGCAGGCATG (SEQ ID NO: 59)

HGHpA:

ACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTC
CAGTGCCCAACAGCCTTGTCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCC
TTCTATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACA
ACCTGTAGGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGG
CTCACTGCAATCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGT
TGGGATTCCAGGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTGGTAGAGACGGGG

81/112

FIG. 24 cont'd

TTTCACCATATTGGCCAGGCTGGTCTCCAACCTCCTAATCTCAGGTGATCTACCCACCTTGG
CCTCCCAAATTGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTT (SEQ ID
NO: 60)

3XFLAG:

GACTACAAAGACCATGACGGAGATTATAAAGATCATGACATCGATTACAAGGATGACGATG
ACAAG (SEQ ID NO: 61)

hsA2:

TTAGAACAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTAGTTGCTGTTCTTCCAC
GTCAGAAGAGGCACAGACAAATTACCACCAGGTGGCGCTCAGAGTCTGCGGAGGCATCAC
AACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGACCTTTTACTACCTGACT
AGCT (SEQ ID NO: 62)

10a.a.:

TCCGGACTCAGATCTGGAGGCTCCGGAGGC (SEQ ID NO: 63)

H2B:

CCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCGGTGACTAA
GGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCT
ATGTGTACAAGTTTCTGAAGCAGGTCCACCCTGACACCGGCATTTTCGTCCAAGGCCATGG
GCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGAGAGGCTTCCCGCCT
GGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCCGTGCGCC
TGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGCCATCACC
AAGTACACCAGCGCTAAGTAA (SEQ ID NO: 64)

P2A:

GGCAGCGGCGCCACCAACTTCAGCCTGCTGAAGCAGGCCGGCGACGTGGAGGAGAACCC
CGGCCCCGGAGCTAGCGGA (SEQ ID NO: 65)

GGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGGGGATGTAGAGGAGAATCCA
GGGCCTGGTGCTAGTGGA (SEQ ID NO: 66)

T2A:

(GSG)EGRGSLLTCGDVEENPGP (SEQ ID NO: 67)

E2A:

(GSG)QCTNYALLKLAGDVESNPGPP (SEQ ID NO: 68)

F2A:

(GSG)VKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 69)

Exemplary Plasmid Backbone 1 – Left ITR:

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTCGGGCGACCTT
TGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCA
CTAGGGGTTTCT (SEQ ID NO: 70)

82/112

FIG. 24 cont'd

Exemplary Plasmid Backbone 1 – Right ITR:

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCGGGCG
TCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTG
GCCAACTCCATCACTAGGGGTTTCCT (SEQ ID NO: 71)

Exemplary Plasmid Backbone 2 – Left ITR:

CATGTCCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCG
GGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGG
AGTGGCCAACTCCATCACTAGGGGTTTCCT (SEQ ID NO: 72)

Exemplary Plasmid Backbone 2 – Right ITR:

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGG
CCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGA
GCGAGCGCGCAGCTGCCTGCAGGGGCGCCTG (SEQ ID NO: 73)

PHP.eB capsid:

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLPGNGLDK
GEPVNAADAAALEHDKAYDQQLKAGDNPYLYKNHADADEFQERLKEDTSFGGNLGRAVFQAKK
RLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDP
QPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRTVTTSTRT
WALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWG
FRPKRLNFKLFNIQVKEVTDNNGVKTIANNLTSTVQVFTDSYQLPYVLGSAHEGCLPPFPADV
FMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLD
RLMNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
NNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK
VMITNEEEIKTTNPVATESYGQVATNHQSDGTLAVPFKAQAQTGWVQNNQILPGMVWQDRDV
YLQGPWAKIPHTDGNFHPSPLMGGFGMKHPPQILIKNTVPADPPTAFNKDKLNSFITQYST
GQVSVEIEWELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL
(SEQ ID NO: 74)

AAV9 VP1 capsid protein:

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLPGNGLDK
GEPVNAADAAALEHDKAYDQQLKAGDNPYLYKNHADADEFQERLKEDTSFGGNLGRAVFQAKK
RLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDP
QPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRTVTTSTRT
WALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWG
FRPKRLNFKLFNIQVKEVTDNNGVKTIANNLTSTVQVFTDSYQLPYVLGSAHEGCLPPFPADV
FMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLD
RLMNPLIDQYLYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
NNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK
VMITNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGWVQNNQILPGMVWQDRDVYLQGPW
AKIPHTDGNFHPSPLMGGFGMKHPPQILIKNTVPADPPTAFNKDKLNSFITQYSTGQVSVEIE
WELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL (SEQ ID NO:
75)

tet-Transactivator version 2 (tTA2):

ATGTCTAGACTGGACAAGAGCAAAGTCATAAACTCTGCTCTGGAATTACTCAATGAAGTCG
GTATCGAAGGCCTGACGACAAGGAACTCGCTCAAAAGCTGGGAGTTGAGCAGCCTACCC
TGTA CTGGCACGTGAAGAACAAGCGGGGCCCTGCTCGATGCCCTGGCAATCGAGATGCTGG
ACAGGCATCATACCCACTTCTGCCCCCTGGAAGGCGAGTCATGGCAAGACTTTCTGCGGA

83/112

FIG. 24 cont'd

ACAACGCCAAGTCATTCCGCTGTGCTCTCCTCTCACATCGCGACGGGGCTAAAGTGCATCT
 CGGCACCCGCCCAACAGAGAAACAGTACGAAACCCCTGGAAAATCAGCTCGCGTTCCCTGTG
 TCAGCAAGGCTTCTCCCTGGAGAACGCACTGTACGCTCTGTCCGCCGTGGGCCACTTTAC
 ACTGGGCTGCGTATTGGAGGATCAGGAGCATCAAGTAGCAAAAGAGGAAAGAGAGACACC
 TACCACCGATTCTATGCCCCCACTTCTGAGACAAGCAATTGAGCTGTTTCGACCATCAGGGA
 GCCGAACCTGCCTTCTTTTCGGCCTGGAATAATCATATGTGGCCTGGAGAAACAGCTAA
 AGTGCGAAAGCGGCGGGCCGGCCGACGCCCTTGACGATTTTGAAGTAGACATGCTCCAG
 CCGATGCCCTTGACGACTTTGACCTTGATATGCTGCCTGCTGACGCTCTTGACGATTTTGA
 CCTTGACATGCTCCCCGGGTAA (SEQ ID NO: 76)

GTPase HRas [Homo sapiens]:

MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQVVIDGETCLLDILDITAGQEEYS
 AMRDQYMRGTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDLAARTVESR
 QAQDLARSYGIPYIETSAKTRQGVDAFYTLVREIRQHKLRKLNPPDESGPGCMSCKCVLS
 (SEQ ID NO: 77)

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

RPKPQQFFGLM (SEQ ID NO: 78)

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

CYIQNCPLG (SEQ ID NO: 79)

GCaMP6m:

ATGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
 GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTTCGACTCATCACGTC
 GTAAGTGGAAATAAGACAGGTACGCAAGTACAGAGCTATAGGTTCGGCTGAGCTCACTCGAGA
 ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCCGAACCTCAAGATCCGCC
 ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
 GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTTCG
 AAAGACCCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCTGACCGCCGCCGG
 GATCACTCTCGGCATGGACGAGCTGTACAAAGGCCGTACCGGAGGGAGCATGGTGAGCA
 AGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTA
 AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
 GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGAC
 CACCCTGACCTACGGCGTGCAAGTGTTCAGCCGCTACCCCGACCATGAAGCAGCACGA
 CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
 GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCG
 CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
 AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCTTTCTC
 CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
 GGTCTCTGGGGCAGAACCCCAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
 CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCTGACAATGATGGCAAGAAAAGGGA
 GCTACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCCGGTGTGTTTGATAAGGATGGCA
 ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGAGAGAGAAGTTAA
 CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
 ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 80)

GCaMP6s:

ATGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
 GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTTCGACTCATCACGTC

84/112

FIG. 24 cont'd

GTAAGTGGAAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCTGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCCGAACCTCCACATCCGCC
ACAACATCGAGGACGGCGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTTCG
AAAGACCCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCTTTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCTGACAATGATGGCAAGAAAAATGA
AATACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGAGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 81)

GCaMP6f:

ATGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCTGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCCGAACCTCAAGATCCGCC
ACAACATCGAGGACGGCGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTTCG
AAAGACCCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGAATTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCTGACAATGATGGCAAGAAAAATGA
AATACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGAGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 82)

CN2415 (1796 bp between ITRs):

GCGGCCGCACGCGTACTGAGCCTTACAGAGATCCAGGGACTTGTCCTTGACCAGTGAGCA
AGTGGGTGGAGTGGGCTATGGACTCTGGACTGGAGGAGCCCCGAGCCCGTGCTCTCAGC

85/112

FIG. 24 cont'd

CGCTGCTCAGCCCTGCAGCAGGAAGAACATCTGAAGCAAACCTACAAAGAACTTTTTCTCCCT
 GAAAAAGCGATTAGCCCATAATAGAAGACCTAAGTCCAGCAAAAAGAGCCGCGTTTCTGC
 TGCATTTTGTTCCTACTGGGTGACGGATGGCGCTGTCATCTCTTCAGGGAGGGAGGCTT
 ACAGACAGGCCAGCTCGTCAGCCTGCTGGCTCCTTCAAAGTCAGGTCTTCACAGTAGGTT
 CCTTTATTTCTGTCTCAACAAGTTGAAATCCCCTCAAATTAGACACACACTTTTTCTTATGTAA
 TTAAAGTACAGTGAGAAGACAGGCATCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGA
 GCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTC
 GCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGA
 GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
 GCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCC
 CTGGCCCAACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCG
 ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
 GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
 GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
 ATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCACCGCCGAC
 AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGC
 GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
 GCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCG
 CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGA
 GCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCT
 GGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG
 TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTC
 CTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCC
 CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTC
 GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCTTCTTGAAC
 CTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
 TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
 TGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
 83)

CN2416 (2133 bp between ITRs):

GCGGCCGCACGCGTTGCCTGTCTTCTCTCTGCAGTACTGCTCAGTATTGCTGGAAAAC
 CAGAACTGAAGACAGTCTTGTCCTTAATAGATTAGGATTCAAAAACAGACTTTAGGAGGA
 CTAGGTGTATTTTTTCCCTTTTAAACAGCAGAGGGATTTGAAGATCTGGACCTTTCCTGTTG
 AGTTATTGAACCTGCAGTATTTGTTATTGTCTTGCCAGTACCTATAACTACCTGGGAAAAAA
 GACAAACCAGATGTTGGAGGCTTTTGAACAAAAGGACTTCTCTTTAGCCCTGAACTATCAG
 GTAAACGTTTTCTTCTTCCCATAAATCACAGCCTTTTTTCTTCCCACTTGAGGATTCTTT
 TTCTTCTCTCTACCATGCCTGCCCTCATTAAAAATTGCCAGCCACTTCTTTCTTTACCTTTCT
 TCTGTTTTGTACTTTTTCTTGTTTAAACCAAAGCTGTCCCCATTGTCTGTGTTCCCGACATG
 TCTAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCTCCTAGAAAGGATCTTAT
 TTTTACTCTATCTTCAAAGTACTTTTGTGACTATAATAAAATGGTCATAGATTTCTGGGAAAA
 TGACCACAGGCTGTCCCCTTGAGGACATCACATCAGAGTTATCACAGAATATGAGGAAGCT
 ATGAAGATGGTAACCTAAAATGGGGAAGGCTTCATTCAAAGTCAGGGCAGGTGAGTCTTGT
 CCACCCTGGAAATGTGCTGAGAACTAGGAACATCCACCAGGAAGGAAGAGCTCGGGCTGG
 GCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTC
 GAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGG
 TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCC
 GGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCAC
 CGGCAAGCTGCCCGTGCCCTGGCCCAACCCTCGTGACCACCCTGGGCTACGGCGTGCAGT
 GCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG

86/112

FIG. 24 cont'd

AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAAC
GTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCAC
AACATCGAGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGG
CGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGCGGA
TCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCGCGCGCGCGGAATT
CGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT
GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTC
CCGTATGGCTTTTCAATTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAAC
TCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATT
CCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCT
CCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAAATGA
GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCA
GGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACC
GAGCGGCCGC (SEQ ID NO: 84)

CN2417 (2137 bp between ITRs):

GCGGCCGCACGCGTAGTCCTCACAACAACCCTGTGGGCAACCATGTCAGAATTCATTTATT
TTTAAACAGTTAAAGAATCTGACTCACAGAGAAGGTAAGTGAAGTCTCCTAAGATTACTGGGC
AGTAAATGGCAGCATCTGGACTAGAACACACCATTTCTGATTTTTTGTCTGTTCTTGTCT
ATGCCAACCTTTTATGTAACTTTTATGCACAAAAGGAATAGTTCTTCCCGAGAGTTCTGTT
CATTTGCCAACTAGTGGATAAAAAAACTTTAAGGTTTTTGACCTAAAGCCCTAACTATTGT
CATAATTTTAAATGAATAGGCCTATTTTCTCCTTTGAGAAAAAAAATCAGAAGTAAAAT
GAGTCAGAATTTCTGAACAAGCCCTCCACCTGAGGTAAGTGAAGGACAGGTTGGCTTGTC
TCTCTGTGGGTTACAGAATAGGAAAGGACTGGAAAGCTTTTTTGAGAAATACTTTCAAATTC
TGGCCAAAGCTTCTGCTGTGCCGAAGTTTTTATTTCCAGACTTTTCACAAGGATGAAAA
AGGTGATGTTGGAGCAGAAGTCAAGAACCTGAGACTCTGAGGTACCTGGTGGGGCTG
TGATTGCAGGATCCAGGAAGACTGTGGGCGAGCATGACTCTAGAGGAGGCTGAAAAA
GGCCAGGTGTGGAGAGCCCTTGGGGCCCTGGGAGAAAGGCCAGCAAGACAGGAAGGC
CTAGACAAAAGGGCATGAGTCTTCCCTTCCCTATCAGGGCTAGAGCTTTAGAACAGGAGAG
CTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATC
CAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGT
TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTC
AGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGAT
CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACG
GCGTGCAAGTCTTCCCGCTACCCCGACCATGAAGCAGCAGGACTTCTTCAAGTCCG
CCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG
GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAAC
AGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAG
ATCCGCCACAACATCGAGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC
CCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAA
GCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCGCGCGG
CGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATT
CTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCT
ATTGCTTCCCGTATGGCTTTTCAATTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCAC
GGCGGAATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCA

87/112

FIG. 24 cont'd

CTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGG
GTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACG
TGCGGACCGAGCGGCCGC (SEQ ID NO: 85)

CN2418 (2385 bp between ITRs):

GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTACTATTC
GCAGCAGTGGGAATGTAAGTTGGCCCCATTCTCCTGGAAAGTCATTTGGCACTCTGTGGG
CACTGAAGACCAGCAGTGTCTCTGCTGGGGACAGCCTTGGTGGCCTCAAGAGTCACACAG
ATGAGGATGCTCGTTGTCTATTTTAGTCTCTGTGCTTGTCTCTGTGTTCCAAATGATCCATA
ATTTTTTAAAGTGTCTTTTTAAAGAATATGTAGGTGTTCTGACTCTTGATCTGATGGCTGGTTA
CCCAGGCGTCCTTTTTATAATAATTTGTTATACTGTTCTTTTGTCTATGTGCTTTTTCTTTATG
GATATGACATTTTACAGTTTTTAAAAATTAGAAAGAATGAGTAGGCGAGACGCTGGGGTCA
GGCAGTCTCAGGCAGAGGAAGCAGCCTGAGCAAAGCCCCGGGGAGCTGGGCAAAGGCAG
GGGTGTGCATGACCCTGGGAATCCTGGCCCAACAACCAGGCACCCTGGCCCAAGGACAT
GCGGTCCCCTGAGGAACAGCATCTCTGCTCCACCTCTGGCTAGCCATTGGTGCAGCACC
TGCACTGCCGAGGCCAGGCCTAATTTAGTCCCACTTCAGCCTGTGTGAGGCTGGATCAAA
AGTGAGCTCTCAGTGAGGGGGGAACCAGTAGAGGTGAGATGATGGGTTTAATAAAGTAGTTT
AACCACCGCCACTGGGTGTGCTAAGAGGCGGTTGCTGTGGCCAGCAGGTGGAGGTTATAA
TGAGCCTGGCAGTTGTTGTTGGAGGCGGCAGGCAGTGGCTGTGCTGATGGGAGAATTCTT
AAGGAGCTCGATTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCTTGGGG
GGGGAGTTGGAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCA
GAAGCATCCCCGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCACCATGGTGAGC
AAGGGCGAGGAGCTGTTACCCGGGGTGGTGGCCATCCTGGTCGAGCTGGACGGCGACGT
AAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGC
TGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGCCCCACCCTCGTGA
CCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGTACCCCGACCACATGAAGCAGCAC
GACCTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAG
GACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAA
CCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGC
TGGAGTACAACACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCA
TCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGAC
CACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTA
CCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT
GCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGT
CGACGGCGCGCCGCGGCCGGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGT
GAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTT
AATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTTCTCCTCCTTGTATAAATC
CTGGTTAGTTCTTGCCACGGCGGAACATCGCCGCCTGCCTTGCCCGCTGCTGGACAGG
GGCTCGGCTGTTGGGCACTGACAATTCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTA
GTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCA
CTCCCACTGTCCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAT
TCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAG
CAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 86)

88/112

FIG. 24 cont'd

CN2436 (2182 bp between ITRs):

GCGGCCGCACGCGTGGGAAGAATAGTGAGTTCTGGGGTGGAGGGTTTTGATTTTTTAAAG
TGGAAAGTCAAGAAATGCACCACGGAGAAGGTGGCATTGAGTGAAAACCTCAAAGTGGTTT
GAGAGAAAGTGGTGCAGATATGCGGACAGCGCTTCAGGCAGAGAAAATAACCTGGGATGT
CTGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGAGCAAGAGAGTTAAGAATAGGTG
TAGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCATTTTTAGGCCAATATAAGGACTG
GCATTTTTAGGTCATGATGAGAACTGGCATTACTCTGCCTAAAATAGGAAACCACTGGAG
CATTTTTGAGCAGAGGAGTTGAACTGTCTGATTTAAAATTTTAAAAGGTTCTGGCTACCTTGT
GAGAAGAGACTGAAGTAAGATATGCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAAT
AATCTAGGCAAGAGAGATGGATCAGGGTGATAACAGTAACTTCATGAGAGCATTTCAGTT
GCAAATATATTCCAGATTTGCTGACCAGATTTGATAAGGAATATAGAAAAAGAGGAAAGAAC
AAGAGATAACAGAGTGTAGGCAGAAAAGAGAAAAGGTTCAAAGACTGAGCCCTGCAGTATT
TCAACATTAAGAGCTTGAATTGCAGAATAGGAACCATCTCAAGGCTTTAAATACAAGTGCTG
AACACTCTCAAATTTATATTTGTATCCCAGACTTCTCTAAAATTCAAGTTCTAAAATCCAGTT
GGTTACTTTATGTCTCCACTTGGAAACAAGTCTTGAAGCAGTGAGCTCGGGCTGGGCATAA
AAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCT
AGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTG
CCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAG
GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAA
GCTGCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAAGTGCTTCG
CCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCT
ACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG
GAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTAT
ATCACCGCCGACAAGCAGAAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATC
GAGGACGGCGGCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGG
CCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCT
CGGATCGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGCGCGAATTGATATC
ATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATTTGCTC
CTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG
GCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCATCG
CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGG
CTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCG
TGCTTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAAT
TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAG
CAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGG
CCGC (SEQ ID NO: 87)

CN3000 (2104 bp between ITRs):

GCGGCCGCACGCGTTATCTGTGTCCAGTCCGATTGATGGGCTCATGGATTGGAAAGCTGT
GACAGACGCCCTGGGCTGGGAGGTATAAAACGAACTGTTTAACTCCAGTGAAAAACAGT
TGCAGGGATCCCAGAAGTGGGAAAAGGTGCTCTGAGCACCTACCGTGGGGTGGGTGTG
ATTCTGAGACTAAGGCTCTGGGGGCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACT
TTGGGAGGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCGAGACCATCCTGGCTAACA
CGGTGAAACCCTGTCTCTACTAAAAATACAAAAAATTAGCTAGGCGTGGTGGCAGGTGCCT
GTAGTCCTAGCTACTCGGGAGGTTGAGGCAGGAGAATGGTGTGAACCAGGGAGGTGGAG
CTTGCAGTGAGCTGAGATGGCGCCACTGCACTCCAGCCTGGGCAACGGAGCGAGACTGT
CTCAAAAAAAAAAAAAAAAAAGGCTCTGGGACCCCTGGCCACGAGCCCCAGGCCACGGC
CTCCTGCCTGGGAGGTGTGGCTCCCCATGCAGCCAGCTTCTTCCCTGCCTCAGATCGGTG

89/112

FIG. 24 cont'd

ACCTGCACTCTGTCCAGGAGGAGGGGAGACTTGCCCAGGGACAAACCCGAAAACACGGC
AAGGTCAGGGGAGGAGGGGCACTGTTTCCTGTGCAAGTCCACATTTGTGAAAAAACATTGAT
TGGGCAGAGAGAGGTGTTGTATTCTGGAATGGAGGTGAGTGGGAGCTCGGGCTGGGCAT
AAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTTCAAG
CTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACACGGGGTGGT
GCCCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCG
AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGC
AAGCTGCCCCTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGCACTGCTT
CGCCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGG
CTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTA
TATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACAT
CGAGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACG
GCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACT
CTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGGCCGGAATTGATA
TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGC
TCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTA
TGGCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATC
GCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCGGT
GGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCC
CGTGCCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGGAA
ATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGAC
AGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGC
GGCCGC (SEQ ID NO: 139)

CN3001 (2066 bp between ITRs):

GCGGCCGCACGCGTTCACTGCACTGCCATGGGTTCCATGTTATTTCGTACATTGATGGTTACC
ACACTATAGCTCCTGTCTGATCTCTCCCTTGATCTTGGATAACTGCCAGCTTGACATCAC
CTCTCAGATGTCTGACAGGTGTCTCAGATTGAATATGAAAAACCAGAACCCTGATTCCCT
CCCTTATCCTCACCATAACCCTGTCTTCTCCACACCACCAAGTAGTATAGGCATTAATTGCC
CTTGAAGGGGATTTAAAAAGGATATGAAAACCGTTTTTCAATCTTTGTGTGCTACTGTCTTA
AAATTTCTATCTCATTGCTTTTAAAAAGAACCAAAATATGGGGGCCAAACATGGTGCCTCACA
CCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCACTAGGTCAGGAGTTCGAC
ACCAGCATGGCCAACATGGTGAAACCTGTCTCTACTAAAGATACAAAAAATTATCTGGGC
GTGGTGGCAGGCACCTGTAATCCCAGCTACTGGGGAGCCTGAGGCAGGAGAATTGCTTGA
ACCCGGGAGGTGGAGATTACAGTGAGCTGAGATGGCTCCACTGCACTCCAGCCTGGGCA
ACAGTGCTAGACTCCATCAAAAAAAAAAAGAAAGAAAGAAACAAAATATGTTCTGGCTCAGAT
AAAGCCCATGCCTAATTTTAAAGAAATCCTAAAACATAGATCCTCAAGTAGGAGCTCGGGCT
GGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTT
TCGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACACGG
GGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGT
CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACC
ACCGGCAAGCTGCCCCTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGCA
GTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCC
CGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACA
ACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCC

90/112

FIG. 24 cont'd

ACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCTGCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAG
CAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCG
GGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCCGCCGCGA
ATTGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAC
TATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCT
TCCCGTATGGCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGA
ACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACA
ATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCC
CCTCCCCCGTGCTTTCCTTGACCTGGAAGGTGCCACTCCCCTGTCTTTTCTAATAAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGGTGGGGTGGG
GCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGG
ACCGAGCGGCCGC (SEQ ID NO: 140)

CN3003 (2077 bp between ITRs):

GCGGCCGCACGCGTGCCTGTGTAATCCTAGCTACTTGCGAGGGAGGCTGAGACAGGAGA
ATCGCTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCTGAGATTGCGCCACTGCACTCTAG
TCTGGGTGACAGAGCAAGACTCCGTCTCAAAATAATAATAATAATAATAATAATAAT
AATAGTAATAATTATGCTGAGACAACAGGCACAAAACAGAACTGTCCCAGGCAAACAGGGA
TATATGGTCACCCAAGTTCTAACAGCCATGAAAACAAGCTCCACGTCCCTTTGAGACCTAC
ATTGTTCTAGACACCAACCTAAAACACTTTATGTGCGTAGTTCTGTAATCCCCACAATGACT
GGGTTAGTCATTATCTCATTTTACGTATGAGCAAGCAAACCTGAAGCTCAGAGAGCGTAATA
GCTACCATTTCTTTTGGAGACAAGGTCTCACTCTCTTGCCAGGCTGGAGTGTGGTGGCAC
AATCATGACTCACTGCAGCCTCCACCTCTCAGGCTCAAGTGATCCACCCACCTCAGCCTCT
CTGGTAGCTAGGACTACAGGTGCACACCACCGCTGAAATAGGCCTCATAGTCCCATAG
ACAGTTGTTTTTTTGTGGGCGTTTTTTGGTTAATTTTTGTTTTTGTGTTTTTAGATGGAGTCTC
ACTCTGTCGCCCAGGCTGGAGTGCAGTAGTGTGATCTCAGCTTACCACAACCTCCGCCTC
GAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGG
GATCCAGATCTTTTGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGGAGGA
GCTGTTACCCGGGTGGTGCCCATCTGGTCGAGCTGGACGGCGACGTAACGGGCCACA
AGTTTCAGCGTGTCCGGCGGAGGCGGCGATGCCACCTACGGCAAGCTGACCTGAAG
CTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGAOCACCCTGGG
CTACGGCGTGCACTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAA
GTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAA
CTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
ACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAAGCGGCATCAAGGCCAACT
TCAAGATCCGCCACAACATCGAGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAG
AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCA
GTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT
GACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGC
CGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACT
GGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTAT
CATGCTATTGCTTCCCGTATGGCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTT
GCCACGGCGGAACCTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTT
GGGCACTGACAATCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCAT
CTGTTGTTTGCCCTCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCTT
TTCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGG
GTGGGGTGGGGCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATC
TCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 88)

91/112

FIG. 24 cont'd

CN3004 (2227 bp between ITRs):

GCGGCCGCACGCGTGTGGCTGAAGCCCTACTGCCAGGGTAACATTCTCCTTGCGGAGCC
CAGCTGGGCGCCGCCCTTCCCTCCAGCAGCCTCCCCAGTTTTCTTGCGCACTTGACACG
CGCTTACAGCAGAGGCTGCCGTTTCAAGCTGCGGGGACTGGGCATCCCCAGGTGTCCAG
GGCTCCCGCTGCCCAGGCTCTAAATCGAGCGCGCCCAAGGGGGTTCCCCCTCCCGCGCC
CTGCCCGACCCGCTGCCGCAGCTGCCGCGGCTTGGGGACCAGCGGCGGCTGGGAAAG
GCGGGCATGGGGCTGTCCCGGGGGGCGAGGCGGGCTGGCCCGGGAGGGGGCAGGAACA
GACACCGGCAGCGCTGTGGGAGAGGAGAACCGAGAGGGCCTCCAGCCGAGCCCGGCTG
CCAGCAGCCGGGAGACGAGCGAGCCAACCTTCTGGGGGACAGCTGGGCCGCTGCGCC
CCGCTGGCCCATCCCTGGTGGGGAGGGGCGGGACGGGGGGCTGCCTGTCAATCAGCTC
CCCAGGAAGCGCTCAGAGCCAGCGCATCATCTCCGCGGCAAATTCGGCTCTAGAAATAAC
TTTTTTCTTTTTTCCCCCTCTTCTGATTCTCCTAGCAAGGGGCTCACCCCTTACTCACCCCA
GGCAGCTCATGGTGGCCTCTTTCGGTGGTGGTGCCTATGCTCCGCTGCAGAGCCGAGA
GCTGCGCCGGGGCTGGGGTGGGCGGGGGGGCGGCGGGGAGGGTGGCCAGCCAGTCC
CTGACATTGGCCAGGCCAGAGGGAGGGGCGAGGCTGCGGGGTGAATGGCAGCAAGGGG
AGGGGGCGGGCTGAGTCCGGGAAGAGGGAGGGGAGGGATGGAGGAGCAGTCGCAGGC
CCAGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTT
CTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCACCATGGTGAGCAAGGGCGA
GGAGCTGTTACACGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCC
ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG
AAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCAACCTCGTGACCACCT
GGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTT
CAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCC
AACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTA
CCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTT
CGTGACCGCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCG
CGCCGCGGCGCGAATTCGATATCATAATCAAACCTCTGGATTACAAAATTTGTGAAAGATT
GACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGTATCATGCTATTGCTTCCCGTATGGCTTTTATTTTCTCCTCCTTGTATAAATCCTGGTTAG
TTCTTGCCACGGCGGAACCTCATCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAG
CCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTG
TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTG
GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
AGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 89)

CN3005 (2027 bp between ITRs):

GCGGCCGCACGCGTTACCGTGTCAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCC
ACCAGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGCGCCCGGCCATT
TTTTAAGAGTGAGAGTCTCTCTATCTTTTTATTTTATTTTATTTTATTTTAAAGTTCTGGG
ATACGTGTGCAGGATGTGGCAGGTTTGTTACCTAGTTAAACATGCTCTATCTTTAAAAAA
TAAAAAATAAAAGACCGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCC
AACGTGGGCAGATCACCTGAGGTTAGGAATTCGAGACCAGCCTGACCAAGATGAAGAAAC
CCTTCTCTACTAAAAATACAAAATTAGTCAGGCATGGTGGCACATGCCTGTAATCCCAGCTA
CTCAGGAGGCTGAGGCAGAGGAATCGCTTGAACCCGGGATGTGGAGGTTGTGGTGAGCC

92/112

FIG. 24 cont'd

GAGATTGCACCATTGCATTCCAGCCTGGGCAACAAGAGCGAAACTCCATCTCAAAAAAATA
AAATAAATTTTTAAAAACTGGGCCCGATGAGTCACTCCTGTAATCCCAGCACTTGGGAGGT
CAAGGTGGGAGAATCACTTGACCCAGGAGTTCCGAGACCAGCCTGGGCAACGTCGTGACA
CCCTGTCTCTACAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTA
CATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCACCATGGTGAG
CAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACG
TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT
GACCACCCTGGGCTACGGCGTGCAAGTGTTCGCCCGCTACCCCGACCACATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGA
ACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGC
ATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
ACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAG
TCGACGGCGCGCCGCGGCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTG
TGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTT
TAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAAT
CCTGGTTAGTTCTTGCCACGGCGGAACATCGCCGCTGCCTTGCCCGCTGCTGGACAG
GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCT
AGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCTTCTTACCCTGGAAGGTGCC
ACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCA
TTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA
GCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 90)

CN3007 (1904 bp between ITRs):

GCGGCCGCACGCGTCTATGAAGTGTGGAGTCTGAGAGGGCTGAGGATGGCATCAGTGC
AGGGGAAGCGCACGGTGCAGAAGCCAGAGCTCAAGTGTGGATGAGAGTGTCCATACTCT
GGGGTGAACAAGGAAACAGCAATGACAGGAGATAGGTTGCATAGAGAAGGATGTATTT
AATACTTTAAGGGTAACCTGGAGAGAGATTTCCCACTGTCCAAGAAGGTGTTAAATGCAAATT
GGAGGTGAACTCATGGTCTGCAAAAATAAATATACATATATTCTCAAAGTTCTGTCTACTTA
GAGGCCAGAAGAGCACAATAAATGAGCACACATACACCCTCAATCTTGCTTCTAAATA
CCATTCTCTATTGAAACAAGCAGTGCTCCTTAGAGAAATTACTGACTTCAGTACTGGGGCA
GGAAAAAATGAGCCTGGAACATCCCTTTCTGCCATAAACTAAAAAATTACTCAGAGATTGA
TGTGGTCTGCCATGATCTGAATGTGTCCCCCAAATTTATATGTTGAACTTAATCCTCAC
TGTGGTGGTATGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACA
TTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCACCATGGTGAGCAA
GGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAA
ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCTGGGCTACGGCGTGCAAGTGTTCGCCCGCTACCCCGACCACATGAAGCAGCAGC
ACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGG
ACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCT
GGAGTACAACACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCAT
CAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACC
ACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTAC
CTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTCTG

93/112

FIG. 24 cont'd

CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTC
GACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTG
AAAGATTGACTGGTATTCTTAACCTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTA
ATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCATTTTCTCCTCCTTGTATAAATCC
TGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGG
GCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGCCCCCTCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACT
CCCCTGTCTTTCTTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCT
TATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCA
GGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 91)

CN3009 (2001 bp between ITRs):

GCGGCCGCACGCGTCAGGTCAAGTCAAATCTATCTCATGGAGTGATATAAATCCTACCTGTT
TTTCTGTTTGAATTTTGCATACCATTCTGTTAATTCAAGTAATAGTTATTGGGCACATATGT
TATGTGCCTGTCATTGTTCTACATATAGGGAATATAGCATTAAACAGAACAGACAAAAATTC
CTGCCCTTATGGAGTTTACATCCTAATGAGGGAAACACACAGTAAATGAGAAATGTAAGTAA
AATAAGTATAATTTTGTATGGTTAAAGTGGTAAGAAGGAAAAGGAAGCATAGAAGGATAAAG
GAGAATAAGAAATTATGGGGTCTTTTTTAAATAGAGTAGAAAGAGAAGACTACTGGGAATG
TGCTACTTGAGTAAATATTTGCAGAGGTTTCAGGCATGAGCTATGAAAACACTTGGCGGAA
GAAAGCTCTAGAGAGAGAGAAAAGTAGATGCAAATATCTAACAATGTGTAATCAGAGCTAT
CACATTATATAGACATCAAGTAACTTCTATGCTTAAATAAATTCTGAAACTTTTGCAATTCTT
CCAGAATAACATACAAATACTTTCTGTGTAATATCCTAAAGTTCATTTCAGGATTCGGTTCTGA
GACCCCATTTCTCTTAGACAGCATATTTCTGCCAGCGAGCTCGGGCTGGGCATAAAAGT
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCG
CTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTTCACCGGGGTGGTGCCCAT
CCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCG
AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTG
CCCGTGCCCTGGCCACCTCGTGACCACCTGGGCTACGGCGTGCAAGTGTCTTCGCCCG
CTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT
CCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA
GTTTCGAGGGCGACACCTGTTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCA
CCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGG
ACGGCGGCGTGCAAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCC
GTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCTGCTGGAGTTCTGTGACCGCCGCGGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCGCGAATTCGATATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTT
TACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTT
TCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCC
TGCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCG
AGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCTTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 92)

AiP1335 (1994 bp between ITRs):

GCGGCCGCACGCGTATAGGTACCGAGCTCCCGGTCCCGCCCCCACTCAGGACCCGCCCC
ACGGGCGCCCCGCCCTTCGAAGCTCCGCCACGTGAGCTGGGGGTCCAGCGCGTCTCAG

94/112

FIG. 24 cont'd

CCTCTTTGAGTTGCAAGTCCTTGGGTCCCCAGATCTGTTAGCCGTCTCTGGCTCTTGTGGC
 GCCTCGCCTGCCAGTGGCACGCAGGCCATCCCATCTCAGGGGACTAATTTATAGAGGAGG
 ACTCCGAGGCCAAGTTGGGGGGTGGGGTTGAGCACTGCACACACAATGGAGCCCAGGAG
 GCACCGACGGGGCTTCCCCAACTACTTGCGCCACACCCGGGACTCCCCACTGGCCAGC
 TGGCCCGCCTTGCCTGGCCTCATTCTCTTCTGCTGGCTTGAGCAGAGCCACGCCGCCA
 GGCCTGCCGGTTCCAGTCCTGCCCGAGAGAAGGAGCTGGGAGGTCTCCAGGGCGAGAG
 GTTCTAGCCAGTTTGTCTCAGCACTGCTCCTCAGACAATTGTGGGAATTGCCTCTGCCCCC
 ATCTCACTGTCTTTCCTGAGCCCACGGTTCAGAGCTCCCCTTGCTCAGTCAAGCTCATAAG
 AGGGATTTGATTGTCAGCTGTAACCTATTCACTCATTAAAGGGTCCCCTGGGCCCAAAGTGA
 GCTCGGGCTGGGCATAAAAGTCAGGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGAT
 CCAGATCTTTTGAAGCTAGCGCTACCGGTGCCACCATGGTGAGCAAGGGCGAGGAGCT
 GTTACCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGT
 TCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTG
 ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACCCTGGGCTA
 CGGCGTGCACTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTC
 CGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
 CAAGACCCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGA
 AGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACA
 ACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCA
 AGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
 ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTC
 CAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTGTGAC
 CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGC
 GGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGT
 ATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCAT
 GCTATTGCTTCCCGTATGGCTTTTCTCTCTCTGATAAATCCTGGTTAGTTCTTGCC
 ACGGCGGAATCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGG
 CACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGT
 TGTTTGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTCC
 TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTG
 GGGTGGGGCAGGACGAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGC
 GGACCGAGCGGCCGC (SEQ ID NO: 93)

AiP1336 (1988 bp between ITRs):

GCGGCCGCACGCGTATAGGTACCGAGCTCTTTTACAAGTGGACTATTTTAAGTTGTCACAT
 TTTTACTATACACTGTGGAGTGAAGTCTTAATCGTGGGCTTGGTGCCCTTTTGTGGTTAT
 TCGTCTCTCATGACTGCCCTCTTGCCCTTCTTCTGATTTAAGGCATGCAGTGTATGTGATTA
 TCTTAGTTTTCTTACTGTTGCTCTGTGATAAAATACGCCGACAAAAGCAAGTTAAAGAGAGC
 TTATTTTAGCTGCCAGGTTTTAGCCCATGGCTTTGAGGAAGTCATAGCCACACAGGAGCTT
 TGTCAGCCTTTTACACTGCACCGAATTAGTTCATACCGCTACCCCAAGCTCTCTTTCGACGC
 CTTTACAAGACCCAGAATACTTTGCCGAGGGAATGGTGTCAATCCTCAGTTAATCTAGTCTA
 ATAACACCCCCCCCCGCTGTGGTCATGCTCAGAGGCCTATTTATCAGACAGTTCTAGATTT
 TGTCCCAAGTGACAATTGGCATTGGCCATCACTGTGGCCTTTGTTTCGGTAATTTTGTCAAAC
 AAGTTCAAATGGTGTAGTTACTCAAGTCTTGTGCTTGTCTATCTTTCGAGCTCGGGCTGG
 GCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTC
 GAAGCTAGCGCTACCGGTGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGG
 TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTTCAGCGTGTCC
 GGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCAC
 CGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGT
 GCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG

95/112

FIG. 24 cont'd

AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
CCGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAAC
GTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCAC
AACATCGAGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGG
CGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGCGGA
TCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGGCCGCGAATT
CGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT
GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTC
CCGTATGGCTTTTCAATTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAAC
TCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATT
CCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCT
CCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAAATGA
GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCA
GGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCG
GCCGC (SEQ ID NO: 94)

AiP1337 (1988 bp between ITRs):

GCGGCCGCACGCGTATAGGTACCGAGCTCCTGTCATAGCTCCCGGGCGGTGGCTGTTCTT
CTACTGCGGGCGGCCGCTGCTGCCTGAATTGGCTGTGGAGCTGCTCTGGTCTAAGCAGACA
TGGCTGTGGGAGCGAGGGAGGAGGAGGAAGAGGAGGAGGGCGGCGGCGGCGGCGGGGA
GGAGGAGGAGGAGGAGGCTGGGAGTCGTGGAGGCTCTGTCTCCCGCAGGCTGACTGCT
GGCTGAGGCGCAGCACAGCCTTGTTAAATCCTTAATTGCGCGCTTACGCACAGCGGGGT
GGATCGGGTTCCATTGGCCAGGGCTTCGGGAGCAGAGCAACACCCTAACTCGTCCAACCTA
GTTTTAGCACAAACAAAGCATTGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
GATCTCTTTTTAGTGAATAGCTTCGACAACCAAGCAGTCCCAGGGCCTAGCTCCGTGAG
CTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATC
CAGATCTTTTGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGT
TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTC
AGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGAT
CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACG
GCGTGCACTGCTTCGCCCCGCTACCCCGACCATGAAGCAGCAGGACTTCTTCAAGTCCG
CCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG
GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAAC
AGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAG
ATCCGCCACAACATCGAGGACGGCGGGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC
CCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAA
GCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGC
CGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATT
CTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCT
ATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCAC
GGCGGAACTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCA
CTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTA

96/112

FIG. 24 cont'd

ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGG
GTGGGGCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGA
CCGAGCGGCCGC (SEQ ID NO: 95)

AiP1338 (1967 bp between ITRs):

GCGGCCGCACGCGTATAGGTACCGAGCTCTTTCGCGCGCTCCTTCCCACCGACCTCCCAG
GGGTGCACCTCTCCAGCCTCGGTGCGCGCTTCCGAAACCTTTGGTGCCCCCTTTTCTGGT
CCCGACCCCCCACCTCACGCCCCCTGGTCTGGACAGCATCTCCCCCTCGCCGCCCTCCG
CCCACGCACCGCCTGACTCCGAGGGGTGCGAGCGCATTGGGCTGCGCCCGCGTGGGGG
CGCCGCGCCAGCCTCGCGTAGCTGTTCTGACGCTGCCGTGCGCGCCGCCCTCCGCAGCC
CAGCCGGCACCCGCACCAGCTCTGCAGTGCACCTCGTCGCCTCTCGGGCCGGTCCCACCA
AGAGCCAGACTGTCGTGACCGGGGGCCAGCCTCGAACGTGAGGCGCGAGGGTGCATGAGCC
AGAGCGCCCTGGGGCGCCGCGCGGAGACCCAGCGGAGATAGCAGTCCTCGCTGCCTTGA
CGCGCGCCCGCCGCGTCCCCAGACCCTTCTGTCTTTTCTCCCGCCCCGCCGCCCAT
GGCCACCCTGCTCCGCAGCAAGCTGACCAATGTGGCCACCTCCGTGTCCAACAAGTCCCA
GGCCAAGGTGAGCGGCATGTTTGCCAGGATGGGGTTTGAGCTCGGGCTGGGCATAAAAG
TCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGC
GCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCA
TCCTGGTTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGG
GAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCT
GCCCCGTGCCCTGGCCCCACCCTCGTGACCACCCTGGGCTACGGCGTGACGTGCTTCGCCC
GCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG
TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAG
GACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATC
ACCGCCGACAAGCAGAAGAACGGCATCAAGGGCAACTTCAAGATCCGCCACAACATCGAG
GACGGCGGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCC
CGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCCA
CGAGAAGCGCGATCACATGGTCTCTGAGTTCGTGACCGCCGCGCGCAATTTCGATATCATA
ATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAGTATTGTTGCTCCTT
TTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCT
TTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGC
CTGCCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTC
GAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGC
ATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA
GGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ
ID NO: 96)

AiP1339 (2078 bp between ITRs):

GCGGCCGCACGCGTATAGGTACCGAGCTCACGCGACCCCCGAGCGGTGCGCGTCCCCCA
ACTTCTTAGAGGGACAAGTGGCGTTCAGCCACCCGAGATTGAGCAATAACAGGTCTGTGAT
GCCCTTAGATGTCCGGGGCTGCACGCGCGCTACACTGACTGGCTCAGCGTGTGCCTACC
CTACGCCGGCAGGCGCGGGTAACCCGTTGAACCCCATTCGTGATGGGGATCGGGGATTG
CAATTATTCCCATGAACGAGGAATTCCCAGTAAGTGCGGGGCCATAAGCTTGCCTTGATTA
AGTCCCTGCCCTTTGTACACACCGCCCGTGCCTACTACCGATTGGATGGTTTAGTGAGGC
CCACGGCCCTGGTGGAGCGCTGAGAAGACGGTCGAACCTGACTATCTAGAGGAAGTAAAA
GTCGTAACAAGTTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGAGACTGTGGAG
GAGCGGCGGCGTGGCTCGCTCTCCCCGTCTTGTGTGTGTCTCGCCGGGAGGCGCGTGC

97/112

FIG. 24 cont'd

GTCCCGGGTCCCGTCGCCCCGCTGTGGAGCGAGGTGTCTGGAGTGAGGTGAGAGAAGG
GGTGGGTGGGGTCTGGGTCCGTCTGGGACCGCCTCCGATTTCCCTCCCTCCCTCC
CTCTCCCTCGTCCGGCTCTGACCTCGCCACCCTACCGCGCGCGGGCTGCTCGCGGGCG
TCGTGCCTCTTTCCCGTCCGGCTCTTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGC
CATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCAAGCTAGCGCTACCGGTGCG
CACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCACCCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCACCGCCGACAAGC
AGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGTCGACGGCGCGCGCGCGCGCAATTCGATATCATAATCAACCTCTGGA
TTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGG
ATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTC
CTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGC
TGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGAC
TGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCGCGTGCCTTCCTTGACCTG
GAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGGAGGATTGG
GAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 97)

CN2555 (2068 bp between ITRs):

GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTTCCAGCT
TAAAGGAAAATTCCAGCCCCAGAGATCTTCCGCAGTGAGTGACAGCTCCTTAATGTGAGGA
GCAGCGTGGGCGTGGTGGGTGGGATGTCAGAGCCATCAAAGCCCTCATTTCTGATCCCTC
TGTGATGTCGGGGTTCCCTGGCACGCCCACTGTCCAGCTTAAAGGAAAATTCCAGCCCCA
GAGATCTTCCGCAGTGAGTGACAGCTCCTTAATGTGAGGAGCAGCGTGGGCGTGGTGGGT
GGGATGTCAGAGCCATCAAAGCCCTCATTTCTGATCCCTCTGTGATGTCGGGGTTCCCTG
GCACGCCCACTGTCCAGCTTAAAGGAAAATTCCAGCCCCAGAGATCTTCCGCAGTGAGTG
ACAGCTCCTTAATGTGAGGAGCAGCGTGGGCGTGGTGGGTGGGATGTCAGAGCCATCAA
GCCCTCATTTCTGATCCCTCTGTGATGTCGGGGTTCCCTGGCACGCCCACTGGAGCTCGA
TTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTGGA
GCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCCCC
GGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCACCATGGTGAGCAAGGGCGAGG
AGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCAC
AAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAA
GCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCTGG
GCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCA
AGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA
ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG
CTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
TACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAAC

98/112

FIG. 24 cont'd

TTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCA
GAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACC
AGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCCG
TGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCG
CCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGA
CTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGT
ATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTC
TTGCCACGGCGGAACATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTG
TTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCC
TTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGG
GGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGAT
CTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 99)

CN2045 (2073 bp between ITRs):

GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTTTACAG
CTCCTTGGTCTGAAGTGGTACCCAGGATCAAGATGTCCCTCCCATGATTCTAATCTCCAGC
CCCCTTACACCATAGTCACTTCCAACTCTGGATTTCTTAACATTGCAGTTTGTCTCCAAAA
AAAAATTAATAAAAAAACCTCATCTGGGCATTAGAAGTCATTCTGAAGAGGCTGTCTATAT
CTCAGTGCAATTTCTACTTATCCACTCCCGACTATACCCTGCTTTTCATCAGATTGCAACTG
ACCTAATTTGTAAAGACTTCAGTGAAAAGAGCCTCATTAAAAGCCTTCATCTCTTTTGCAGC
CTCTCTTTCTCCCTCAAATGATCCAGGTCAGAATGGCCCAATAAAAAATAGCTGGGAAAATT
ACAAAGTCACTCAAGACTCTGCCAGGAGTAGAGATTTTAAGAATTAGTGACAGACATATACT
GTCTAGGAGTAGGGTGGGAGTGGGTTGCATGGCTTAGGGACATAGAGAGCTCGATTACAGC
CGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTGGAGCCACG
AGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCC
AGATCTTTTGAAGCTAGCGCTACCCGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGT
CACCGGGGTGGTGCCCATCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAAGTTCA
CGGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATC
TGCAACACCGGCAAGCTGCCCCTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGG
CGTGCAAGTGTTCGCCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGC
CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAA
GACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG
GCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACA
GCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGA
TCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACC
CCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAA
GCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCTGTACCCG
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGC
CGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATT
CTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCT
ATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCAC
GGCGGAATCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCA
CTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTCTTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGG
GTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACG
TGCGGACCGAGCGGCCGC (SEQ ID NO: 100)

99/112

FIG. 24 cont'd

CN2258 (2166 bp between ITRs):

GCGGCCGCACGCGTNCCTGCGGCCGCACGCGTTCCCTAACCTAGAAAAACCCACCATCAA
CTATTGAGGTAATGTTCCNTAGATGCAGGGTGCATCAGTGGGCATTGTTACTTACTGGAAGA
ATGCTCTCTGGCGGATGGGGGAGGGAAGCAGGAAACTCTGCGTTAAGCAGGGCATAATC
TCTTCAGGGACCTGCTTCAGTTTCAATCCGACTCAAGTCGTGCTTTGCAGCTGGCCCCAGT
TTTGGCTGCTCTGAGAGTCAGTTGTTGGTTATGCACTGCCTGGGGTGAGGGCCAACCAGG
GAGGGGGTTGGAGAAGGATGCGAATAGTCAGAGCAGCTGGGAGCAAATGTATCCAGTTAA
GGGGATCAAAACAGAACGTCAGCAACCCCCGTAATGCCCAAATCTTGCATAACTCTGTGT
TATCTAGGCCATCTCCCCCAACCCCAACCCCGCTTCTCATTTGGAACAAGGTCTTTCTATGTA
TGTAAGCTAAAGTTGACCTTGAACCTCATGATCTTCCCTGCCCAGCCTCTGAGTGCTAGGGTA
AAATAGGCGTGCACCATCCTGCCGGGCTTCACCCACTCCCTTTTCAATTTCTGAGACACTTCC
AGTTATTCTTCTCATTCTTACTCCATTCACTCATACTGACCACTTCAAAATTCTGGAACCTT
CCAGCATGCACTGGCTGCAGATGAATTGTGCTTTTCTGCTCCCTCATCTGACTCTCCTTCT
CGTCTGACTACTTAGCTCTCATCTTCTCTACGGAGTCACCAGTCTTTTTTTTTTTTTTTTCA
AGACAGAGTTTCTCTGAATAGCCCTGGCGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGA
GCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTC
GCCACCATGGTGAGCAAGGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCC
CTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAAGTCTTCCGCCGCTACCCCG
ACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCACCGCCGAC
AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCCACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGCGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAAGTCGACGGCGCGCCGCGCGCGAATTCGATATCATAATCAACCTCT
GGATTACAAAATTTGTGAAGATTGACTGGTATTCTTAACCTATGTTGCTCCTTTTACGCTATG
TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTC
CTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCCTGCCTTGCC
CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTC
GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCTTCTTGACC
CTGGAAGGTGCCACTCCCACTGTCCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGGAGGAT
TGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
101)

CN2251 (2079 bp between ITRs):

GCGGCCGCACGCGTTCACAAATCAGTAAGAATTACACTCCAAAAATAAAACAGGGGAAGGGT
ATTAGTATGTGATACGCATGGATAAAGTAGAAGAGGCCAAAAATACAAAGAACCGTAAACT
CACTAGTGATGGAAGAAACGCTAATTAATAAATAGAGCTTTCCTGACCCTTTATCACATGGG
CACATTTTCAAAGGACAAGACTCTGCTGAGGAGAACACAGAACAGGATCCTTCAGACAAGC
TTCAGCAACCAATTTTAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTAAAAATAACA
TAGATCTGAAAAGCTTTATAACTTTTGGTTTTTTTTTAAGGACCATATTAACAGCATCTTCCA
GTTTTTCTAGAGTAAATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTG
TAAAAAGACTATAAAAAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGA
CTAAGGCCTTCAGCTCAGCAGGGGGCAGAGTTTATCATCTGTAAGTGTACAC

100/112

FIG. 24 cont'd

TGCCCTGCCTTTGTGTGCCCTTGCCCTGTCCCTCTGTGACGGTCACCCCTTTCGGTTCTGGCT
TGGAGCCTGGTCATCAGTGCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTC
GTTGTCCCTAGTGATGTGACCCCTACCTTAGACTTCCCCATCTCTTGAGGTGGCTAGAGCTCG
GGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGA
TCTTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTTAC
CGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
TGTCGGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGC
ACCACCGGCAAGCTGCCCCTGCCCTGGCCACCCCTCGTGACCACCCCTGGGCTACGGCGT
GCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCAT
GCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC
CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCC
ACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCC
GCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCCTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTG
AGCAAAGACCCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGGCCGC
GAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTA
ACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATT
GCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGC
GGAATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTG
ACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATA
AATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGGTGGGGTG
GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 102)

CN1633 (2011 bp between ITRs):

GCGGCCGCAACGCGTTTTAGAACAAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTA
GTTGCTGTTCTTCCACGTGAGAAGAGGACAGACAAATTACCACCAGGTGGCGCTCAGAG
TCTGCGGAGGCATCACAAACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGA
CCTTTTACTACCTGACTAGCTGTTTGTGATTTTAGGTGTTTGTTTATCTATACCTCTGAGAA
AGCCATTTTTTTTTAACATGAAGAAGTATAAAATATTGTGTGTGAAGATTAAGCCCTTCTTGC
TCTATGTATTTGAGCCTAGAAAAAGACACACTCCCAAGAAATTAGAATAAAGGGAAAAATGA
CTTTGAGATGTACTTTCTTAAAATAACTGAAGTAATGACCGCATGAGCCAGCTGATGGGTTT
TTAAATGGATACGTTTCTATCAGCCTGCTGCTTGATGCCAGAGCCAAATATATATGGAGTGT
ATTATATCATATCCTGTAGGCAGGAGACTGTTGGGGCATTGGGGGACTTAGAGAGGTGAA
GTGGCAGGCTTGGCACAGGAATTAACAGCATGGCTACCGTCAACCAGAATTCGATATCATA
ATCAACCATAGGTACCGAGCTCGGGATTACAGCCGGGAGCTTAGGGAGGGGAGGTCACTTC
ATAAGGGCCTGGGGGGGGAGTTGGAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGC
TGAGCTCCGGCCTCAGAAGCATCCCCGGGTTGGATCCTTCGAAGCTAGCGCTACCGGTGCG
CCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAG
CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC
CACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCTG
GCCACCCCTCGTGACCACCCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCACCGCCGACAAGC
AGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTG

101/112

FIG. 24 cont'd

CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGTCGACATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGG
TATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCA
TGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTG
CACGGCGGAACTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGG
GCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTG
TTGTTTGCCCCCTCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTC
CTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTG
CGGACCGAGCGGCCGC (SEQ ID NO: 103)

CN2043 (2152 bp between ITRs):

GCGGCCGCACGCGCCGGTACCGAAGCTACCCCTAACACACTATTCTACACACAGAAAATG
CTCTTCACTAGGAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGG
AAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGACGCGTAGGTGG
AGTGAAGACTGAGAGGGGGTGGTAGGGGGGAGATGAGGCCGGGAGGAGGAGAGTACACCC
TCAGGGATGGAGAGCCAACAAAGATTCTAAACGAGGAGTTCTGCAGTGAGACTGGCATT
GGAACAATCGTCCTGGTTTTGGTTGAGAGCACAACCTTAGAGAACAGAGGCTAGAGGTGGC
AGGCTGGAGTGACAGCCTACTGCAGCTGTCCAGCCAATTAAGCCCTGTAGGAAGAATGGC
TGATGCCGCACTCTATGTGATGCTGAACTGCGGTAGCCAGGCTGACCTGGCCCCAATCCC
GGGCTGTGACTGACCTTCAGATGCAGGCAGCAGAGAGAAGAGGACAAAAGTGACCCGCT
GGGGAAAGGCAAGCAGCTGGCACCCAGGGCACCAAGACCCGCACAGTGAGCATCAAGG
CCAGCAAGGGCACAAGCCAGCTGTCACCTGTCTCTGGGCCAAGCCAAGGGGGGGCTGG
CCTCCTGGCCACCTCCTTGGGGGAGCAGTACCCCCATTGTCATGACTTGTTCACAGATTCC
CAGACCTCCCAGTGCTATCAAGAGCTCGATTACGCCGGGAGCTTAGGGAGGGGAGGTCA
CTTCATAAGGGCTTGGGGGGGGAGTTGGAGCCACGAGTCGTCCAGCCGGAGCCCCGTGT
GGCTGTGCTCCGGCCTCAGAAGCATCCCCGGATCCAGATCTTTCGAAGCTAGCGCTACCG
GTGCGCACCATGGTGAGCAAGGGCGAGGAGTGTTCACCGGGGTGGTGCCCTCCTGGT
CGAGCTGGACGCGACGTAACCGGCCACAAGTTACCGCTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTG
CCCTGGCCACCCCTCGTGACCACCTGGGCTACGGCGTGCACTGCTTCGCCCGCTACCC
CGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCACCGCCG
ACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCG
GCGTGCACTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
CTGCCCGACAACCACTACCTGAGCTACCAAGTCCAAGCTGAGCAAAGACCCCAACGAGAAG
CGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGCGGCGGAATTCGATATCATAATCAACC
TCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCT
ATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTT
CTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCATCGCCGCTGCCTT
GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA

102/112

FIG. 24 cont'd

GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 104)

CN1621 (2170 bp between ITRs):

GCGGCCGCAACGCGTTTAGAACAAATGGCTGGCCCATAGTAAATGCCGTGTTAGTGTGTTA
GTTGCTGTTCTTCCACGTCAGAAGAGGCACAGACAAATTACCACCAGGTGGCGCTCAGAG
TCTGCGGAGGCATCACAACAGCCCTGAATTTGAATCCTGCTCTGCCACTGCCTAGTTGAGA
CCTTTTACTACCTGACTAGCTGTTTGTGTATTTTAGGTGTTTGTTCATCAAGGTCAGGGGC
CTTACCTCTCCTAGTATCATGAGGAGGATCTCAAAGCAGGGAAGGGAAGCCAGCTGAGGG
ACCCCTGAGAGGAGATCAGGGGATAAATACTTTGCAACAGAATCCCCATGGGAAACATGA
CAAAATTAGTAACAGGAATGGAGCCTCCCTCAGAGGATACATGAATCGTTGAGTTGAAAAT
AACCACAAGGTCATCAAATCCAGCTGCCACCAACTGCCTGAATATTTTCAACAAATTTTGA
ACAAGCAGTCCTCCAGCCTCCTCCAGCTGCATCCAGCAGCTCAGTGCTCCACCTGACAGC
CTCTTCTGATAGCGTGAAGGCTCAGAAAGCCCCTGTGGGGCCAGCACATCCTGGGGCAG
GCCTCTAGTGGGCAGAGGGCAGCATACTGGGGTCTGGATCTAACACACTCCTCCATCCAG
GTGGTATATTTATAGCCCTGTGCTTTACAGAAAAGACTTCTGATGAGATCAAGACTTTAGGC
TGTTTATGTGGCTTCTTTTATGAGCCAACTGAATTCGATATCATAATCAACCATAGGTACCG
AGCTCGGGATTACAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGG
GGAGTTGGAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGA
AGCATCCCCGGGTTGGATCCTTCGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAG
GGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA
CCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCCACCTCGTGACC
ACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA
CGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACC
GCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG
GAGTACAACACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATC
AAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGCGCTGCAGCTCGCCGACCA
CTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCT
GAGCTACCAAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCT
GGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGA
CATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATGTT
GCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCG
TATGGCTTTTCAATTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACCTCA
TCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCC
GTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCC
CCCGTGCCTTCTTGAACCTGGAAGGTGCCACTCCCACTGTCCTTTCTAATAAAATGAGG
AAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG
ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCC
GC (SEQ ID NO: 105)

CN2216 (1818 bp between ITRs):

GCGGCCGCACGCGTATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGC
TTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTCAGGCTGCCCCCTCTCATCTGTGA
CCTCCTGTGGTCGGGAGGGACTTGCTCTGATTATATAAAGGCTATGAGAGGCTAGAGTT
GGAGAGCACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAG
CCTGGGTGAGGCTGCCCCCTCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGG
ATTATATAAAGGCTATGAGAGGCTAGAGTTGGAGAGCACAGTTCCTGCTTCCAAATTGGCT
TGAAGTGGGGACATCTGCTCTGCATGAAGCCTGGGTCAGGCTGCCCCCTCTCATCTGTGAC

103/112

FIG. 24 cont'd

CTCCTGTGGTTCGGGAGGGACTTGCTCTGGATTATATAAAGGCTGAGCTCAGAGGTAGGCG
TGTACGGTGGGAGGCCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCGCCTGGGG
ATCCAGATCTTTTGAAGCTAGCGCTACCGGTCCGCCACCATGGTGAGCAAGGGCGAGGAGC
TGTTTACCGGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGGCCACAAG
TTCAGCGTGTCGGGCGAGGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGCT
GATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCT
ACGGCGTGCACTGCTTCGCCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGT
CCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACT
ACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTAC
AACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTC
AAGATCCGCCACAACATCGAGGACGGCGGCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCACT
CACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCACT
CCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA
CCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCG
CGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGG
TATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCA
TGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGC
CACGGCGGAATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGG
GCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTG
TTGTTTGCCCCCTCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTT
CTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCACTTCTATTCTGGGGGGT
GGGGTGGGGCGAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTG
CGGACCGAGCGGCCGC (SEQ ID NO: 106)

CN2717 (1740 bp between ITRs):

GCGGCCGCACGCGTAGTGACTTGGTGCTATGAGCCATATTTTGTCTGTTGCTGTTGTTACTG
GTAGTTTTTGTAACTTCTGGGGCTAAAACCTTGGGGTCTGGTATGCTGTCACTTACCAGTGAG
CTATACCTGGATATTATGATTTAGATGAATGTGAAATATCACCCAGACATACATACTAA
ACACTTGGCCCTTGGCCCATGATGCTAAATGGAGGAGATAGAAGCTTTTGGGGCAGAGCC
TAGTGGAAGGAAATGAGGTCAAATGACATGTACTCTGAAAGGAATATGGGTATTCTGGGCT
TGCGTTATTCTCTCTCTCCCTCTCTCTCCCTCTCTCTCCCTCTCTCTTTTCTCCTTTT
CTCTTTCTCTCCTCGCCTTGTTTTCCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGC
CATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCG
CACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCACCCCTCGTGACCACCCTGGGCTACGGCGTGCACTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCACCGCCGACAAGC
AGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCTACCACTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGTCGACGGCGCGCCGCGGGCGGAATTCGATATCATAATCAACCTCTGGA
TTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGG
ATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTC
CTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCTGCCTTGCCCGC

104/112

FIG. 24 cont'd

TGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGAC
TGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTCCCCCTCCCCCGTGCCTTCCTTGACCCTG
GAAGGTGCCACTCCCCTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGGAGGATTGG
GAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 107)

CN3639 (2734 bp between ITRs):

GCGGCCGCACGCGTTTTCTTTGTCTGAATGGGGAGCACAGAGCTGGATATGCAGCAGTAAAT
CATGACACAATTATTGAAGCCCAGCCAGTGCCCCCTGGCACATCAGCACAAAAGGCTAAAA
TCATTGATCTTACTCAAGCATTAAATGTTGGGACAAGGGAAAAAACTTAACATCTATACAGAT
TCTAAATATGCATTCTTGTGGTTCATGCTCATACCGCCATCAGGAAAGAAAGGGGATTACT
TACTAGCAAGCACTCCCCTATAGAGCATGGACCTGAAATTCTGCAGCTATTGAAGGCAATA
CACCTGCCAAAGGCCATAGCTATCATTCAATTGTAGGGGGCATCAAAGGGACTTAACCCCTA
TAGCACAAGGGGAATGGAAAGGCTGACGGAAGCCAAAGCAAAAGCCCTCAGCGTGCAATCC
CAACAGATCCTAGCACTGCTTCCTTTCTACAATTCCCAGTAGAACCCAAATACACACCACA
GGAAGAACAGCTAATAAAAAAAGCAAGGGAGACAAAAACAAGGATCCTGGTGGCATATGG
AGCCAAAAATATATCTCCCTCAGGCAGCCAGTGGAGAGTTATAAAAGCCCTGCATGACTC
TTTCCATGTGGGGAGGGATGCGGCTCTGGCCATGGTAAACAAGTTCTTTACTGGGTTTAAC
TCAGCTTCAATGGCTAAGCAGGTCTGTCAAGCCTGCTCACTGTGTGCATTTAACAATCCAG
GCATCAAAATGCCCCCTCTAATAAAACCAGTCCAAAGGAGGGGAACTTATCCAGGGGAAG
ACTGGCAGTTAGACGTCATCCATATGCCAGCTTGACAGAGGATACAAGCTTTTGAGCTCGGG
CTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATC
TTTCGAAGCTAGCGCTACCGGTGCGCCACCATGGTCTCCAAGGGCGAGGAGCTGTTACCG
GGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGATGTCAACGGCCACAAGTTCAGCGTG
TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCAC
CACCGGCAAGCTGCCCCGTGCCCTGGCCACCCTCGTGACCACCCTGGGCTACGGCGTGC
AGTGCTTCGCCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGC
CCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAAGACGACGGCAACTACAAGACCC
GCGCCGAGGTGAAGTTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC
GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCAC
AACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGC
CACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCAT
CGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGA
GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCC
GGGATCACTCTCGGCATGGACGAGCTGTACAAGGGCTCTGGTGCTACCAACTTCTCACTG
TTGAAACAGGCAGGGGATGTAGAGGAGAATCCAGGGCCTGGTGCTAGTGGAGACTACAAA
GACCATGACGGAGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTCCG
GACTCAGATCTGGAGGCTCCGGAGGCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAA
AAGGGCTCCAAGAAGGCGGTGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCG
CAGCCGCAAGGAGAGCTATTCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGA
CACCGGCATTTTCGTCCAAGGCCATGGGCATCATGAATTCGTTTGTGAACGACATTTTCGAG
CGCATCGCAGGAGAGGCTTCCCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCC
CGGGAGATCCAGACGGCCGTGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGT
GTCCGAGGGTACTAAGGCCATCACCAAGTACACCAGCGCTAAGTAATGAGGCGCGCCGC
GGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGT
ATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCAT

105/112

FIG. 24 cont'd

GCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCC
ACGGCGGAACATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGG
CACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGT
TGTTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTCC
TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTG
GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCA
CGTGCGGACCGAGCGGCCGC (SEQ ID NO: 108)

CN3050 (2228 bp between ITRs):

GCGGCCGCACGCGTATGTTGGAGGCTTTTGAACAAAAGGACTTCTCTTTAGCCCTGAACTA
TCAGGTAAACGTTTTCTTCTCTTCCATAAATCACAGCCTTTTTTCTTCCCCACTTGAGGATT
CTTTTTCTTCTCTCTACCATGCCTGCCCTCATTAAAATTGCCAGCCACTTCTTTCTTTACCT
TTCTTCTGTTTTGTACTTTTTCTTGTAAAACCAAAGCTGTCCCCATTGTCTGTGTTCCCGA
CATGTCTAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCTCCTAGAAAATGTT
GGAGGCTTTTGAACAAAAGGACTTCTCTTTAGCCCTGAACTATCAGGTAAACGTTTTCTTCT
CTTCCCATAAATCACAGCCTTTTTTCTTCCCCACTTGAGGATTCTTTTTCTTCTCTACCAT
GCCTGCCCTCATTAAAATTGCCAGCCACTTCTTTCTTTACCTTTCTTCTGTTTTGTACTTTT
TCTTGTAAAACCAAAGCTGTCCCCATTGTCTGTGTTCCCGACATGTCTAATCAGGACAGG
ACCCATACAGGGATGGCTCCTCTCTTCTCCTAGAAAATGTTGGAGGCTTTTGAACAAAAG
GACTTCTCTTTAGCCCTGAACTATCAGGTAAACGTTTTCTTCTCTTCCCATAAATCACAGCC
TTTTTTCTTCCCCACTTGAGGATTCTTTTTCTTCTCTCTACCATGCCTGCCCTCATTAAAATT
GCCAGCCACTTCTTTCTTTACCTTTCTTCTGTTTTGTACTTTTTCTTGTAAAACCAAAGC
TGTCCTCCATTGTCTGTGTTCCCGACATGTCTAATCAGGACAGGACCCATACAGGGATGGCT
CCTCTCTTCTCCTAGAAAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATT
GCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGG
TGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCT
CGTGACCACCCTGGGCTACGGCGTGCAAGTGTTCGCCCGCTACCCCGACCATGAAGC
AGCAGGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT
TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCTG
GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAA
CGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGGCGTGAGCTCG
CCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAAC
CACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG
GTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAA
GTAAGTCGACGGCGCGCCGCGGCGGAATTTCGATATCATAATCAACCTCTGGATTACAA
AATTTGTGAAAGATTGACTGGTATTCTTAAGTATGTTGCTCCTTTTACGCTATGTGGATACG
CTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCAATTTCTCCTCCTTGT
ATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCT
GGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTG
CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCTTCTTGACCCTGGAAG
GTGCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 109)

106/112

FIG. 24 cont'd

CN3051 (2225 bp between ITRs):

GCGGCCGCACGCGTTAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCCTCCTA
GAAAGGATCTTATTTTTACTCTATCTTCAAAGTACTTTTGTGACTATAATAAAATGGTCATAG
ATTTCTGGGAAAATGACCACAGGCTGTCCCTTGAGGACATCACATCAGAGTTATCACAGA
ATATGAGGAAGCTATGAAGATGGTAACCTAAAATGGGGAAGGCTTCATTCAAAGTCAGGGC
AGGTCAGTCTTGTCCACCCTGGAAATGTGCTGAGAACTAGGAACATCCACCAGGAAGGAA
TAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCCTCCTAGAAAAGGATCTTATTT
TACTCTATCTTCAAAGTACTTTTGTGACTATAATAAAATGGTCATAGATTTCTGGGAAAATG
ACCACAGGCTGTCCCTTGAGGACATCACATCAGAGTTATCACAGAATATGAGGAAGCTAT
GAAGATGGTAACCTAAAATGGGGAAGGCTTCATTCAAAGTCAGGGCAGGTCAGTCTTGTCC
ACCCTGGAAATGTGCTGAGAACTAGGAACATCCACCAGGAAGGAATAATCAGGACAGGAC
CCATACAGGGATGGCTCCTCTCTTCCTCCTAGAAAAGGATCTTATTTTTACTCTATCTTCAA
GTACTTTTGTGACTATAATAAAATGGTCATAGATTTCTGGGAAAATGACCACAGGCTGTCC
CTTGAGGACATCACATCAGAGTTATCACAGAATATGAGGAAGCTATGAAGATGGTAACCTA
AAATGGGGAAGGCTTCATTCAAAGTCAGGGCAGGTCAGTCTTGTCCACCCTGGAAATGTG
CTGAGAACTAGGAACATCCACCAGGAAGGAAGAGCTCGGGCTGGGCATAAAAGTCAGGG
CAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACC
GGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTG
CCCTGGCCACCCCTCGTGACCACCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCC
CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACACAGCCACAACGTCTATATCACCGCCG
ACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCG
GCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
CTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAG
CGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGA
CGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGGAATTCGATATCATAATCAACC
TCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAGTATGTTGCTCCTTTTACGCT
ATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTT
CTCCTCCTTGATAAATCCTGGTTAGTTCTTGCCACGGCGGAATCATCGCCGCCTGCCTT
GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCCTGTCTTTCTTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCTTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
110)

CN3056 (1934 bp between ITRs):

GCGGCCGCACGCGTTGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGAGCAAGAGA
GTTAAGAATAGGTGTAGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCATTTTTAGG
CCAATATAAGGACTGGCATTTTTAGGTCATGATGAGAACTGGCATTACTCTGCCTAAAATA
GGAAACCACTGGAGCATTTTGAGCAGTGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTA

107/112

FIG. 24 cont'd

AGAGCAAGAGAGTTAAGAATAGGTGTAGCCAGAGAGTTATCAGAATATAAGATTGCATAGG
GCATTTTTAGGCCAATATAAGGACTGGCATTTTTAGGTCATGATGAGAACTGGCATTCTACTC
TGCCTAAAATAGGAAACCACTGGAGCATTTTGAGCAGTGAGAAAAGCAAAGAAGCTGGGG
CTACTGCAGTAAGAGCAAGAGAGTTAAGAATAGGTGTAGCCAGAGAGTTATCAGAATATAA
GATTGCATAGGGCATTTTTAGGCCAATATAAGGACTGGCATTTTTAGGTCATGATGAGAACT
GGCATTCTACTCTGCCTAAAATAGGAAACCACTGGAGCATTTTGAGCAGGAGCTCGGGCTG
GGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTT
CGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGG
GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC
CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCA
CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCGAG
TGCTTCGCCCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC
GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAA
CGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCA
CAACATCGAGGACGGCGGCGTGCGAGCTCGCCGACCCTACCAGCAGAACACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCCCGACAACCCTACCTGAGCTACCAGTCCAAGCTGAGC
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAA
TTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAC
ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAA
CTCATCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAA
TTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCC
CTCCCCCGTGCTTCTTGAACCTGGAAGGTGCCACTCCCCTGCTTCTCCTAATAAAAT
GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGC
GGCCGC (SEQ ID NO: 111)

CN3057 (1934 bp between ITRs):

GCGGCCGCACGCGTCTGGCATTCTACTCTGCCTAAAATAGGAAACCACTGGAGCATTCTTGA
GCAGAGGAGTTGAACTGTCTGATTTAAAATTTTAAAAGGTTCTGGCTACCTTGTGAGAAGA
GACTGAAGTAAGATATGCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCAATAATCTAGG
CAAGAGAGATGGATCAGGGTGATAACTGGCATTCTACTCTGCCTAAAATAGGAAACCACTGG
AGCATTCTTGAAGCAGAGGAGTTGAACTGTCTGATTTAAAATTTTAAAAGGTTCTGGCTACCTT
GTGAGAAGAGACTGAAGTAAGATATGCATGGAAGTAGGTCATCAGTTATTAGGTTATTGCA
ATAATCTAGGCAAGAGAGATGGATCAGGGTGATAACTGGCATTCTACTCTGCCTAAAATAGG
AAACCACTGGAGCATTCTTGAAGCAGAGGAGTTGAACTGTCTGATTTAAAATTTTAAAAGGTTCT
TGGCTACCTTGTGAGAAGAGACTGAAGTAAGATATGCATGGAAGTAGGTCATCAGTTATTA
GGTTATTGCAATAATCTAGGCAAGAGAGATGGATCAGGGTGATAAGAGCTCGGGCTGGGC
ATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTTCAA
GCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGGTGG
TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCGGC
GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCAACCGG
CAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCGAGTGCT

108/112

FIG. 24 cont'd

TCGCCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCT
ATATCACCGCCGACAAGCAGAAGAACGGGCATCAAGGCCAACTTCAAGATCCGCCACAACA
TCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC
GGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGAC
CCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCAC
TCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGGCCGCAATTCTGAT
ATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTG
CTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGT
ATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCAT
CGCCGCTTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCG
TGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCC
CCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGA
AATTGCATCGCATTGTCTGAGTAGGTGTATTCTATTCTGGGGGGTGGGGTGGGGCAGGA
CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCC
GC (SEQ ID NO: 112)

CN4001 (1758 bp between ITRs):

GCGGCCGCACGCGTCAGGTAAACGTTTTCTTCTCTTCCCATAAATCACAGCCTTTTTTCTTC
CCCACTTGAGGATTCTTTTTCTTCTCTTACCATGCCTGCCCTCATTAATAATTGCCAGCCA
CTTCTTTCTTTACCTTTCTTCTGTTTTGTACTTTTTCTTGTTTAAAACCAAAGCTGTCCCCATT
GTCTGTGTTCCCGACATGTCTAATCAGGACAGGACCCATACAGGGATGGCTCCTCTCTTCC
TCCTAGAAAGGATCTTATTTTTACTCTATCTTCAAAGTACTTTTGTGACTATAATAAAATGGT
CATAGATTTCTGGGAAAATGACCACAGGCTGTCCCCTTGAGGACATCACATCAGAGTTATC
ACAGAATATGAGGAAGCTATGAAGATGGTAACCTAAAATGGGAGCTCGGGCTGGGCATAA
AAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCT
AGCGCTACCGGTCGCCACCATGGTGAAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGC
CCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAG
GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAA
GCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAAGTCTTCG
CCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCT
ACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG
GAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTAT
ATCACCGCCGACAAGCAGAAGAACGGGCATCAAGGCCAACTTCAAGATCCGCCACAACATC
GAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGG
CCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAAGCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCT
CGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCGCGAATTCGATATC
ATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTC
CTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG
GCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGC
CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGG
CTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCG

109/112

FIG. 24 cont'd

TGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAAT
TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAG
CAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGG
CCGC (SEQ ID NO: 113)

CN4003 (1622 bp between ITRs):

GCGGCCGCACGCGTTGAGAAAAGCAAAGAAGCTGGGGCTACTGCAGTAAGAGCAAGAGA
GTTAAGAATAGGTGTAGCCAGAGAGTTATCAGAATATAAGATTGCATAGGGCATTTTTAGG
CCAATATAAAGACTGGCATTTTTAGGTCATGATGAGAACTGGCATTACTCTGCCTAAAATA
GGAAACCACTGGAGCATTGAGCAGAGGAGTTGAACTGTCTGATTAAATTTTAAAGGT
TCTGGCTACCTTGTGAGAAGAGACTGAAGTAAGAGAGCTCGGGCTGGGCATAAAAGTCAG
GGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTA
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCC
CGTGCCCTGGCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCCGCT
ACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCC
AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGT
TCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
GGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCACC
GCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGAC
GGCGGCGTGAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGT
GCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCAT
GGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGCGCGAATTCGATATCATAATCA
ACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTAC
GCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCA
TTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAATCATCGCCGCCTG
CCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCGTGGCTCGAG
AGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTT
CCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATC
GCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGGCCGC
(SEQ ID NO: 114)

CN2786 (2258 bp between ITRs):

GCGGCCGCACGCGTCTATCTGGTTTATAATTAGAAGTGAAATGATTTTAAATGGCTTCCACA
TGGCTTAGTTTTCCAAGCAGTTCCAGTTACATAGAAGTTTCAATATAAACAGATAATGAGAGGC
TAGAGTTGGAGAGCACAGTTCTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGC
ATGAAGCCTGGGTCAGGCTGCCCCCTCTCATCTGTGACCTCCTGTGGTTCGGGAGGGACTTG
CTCTGGATTATATAAAGGCTTCACAAGAGTACTGGAAGCGGGGACGGAATCAACTTGAGCA
AAAAACCTCTATCTGGTTTATAATTAGAAGTGAAATGATTTTAAATGGCTTCCACATGGCTTA
GTTTTCCAAGCAGTTCCAGTTACATAGAAGTTTCAATATAAACAGATAATGAGAGGCTAGAGTT
GGAGAGCACAGTTCCTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAG
CCTGGGTCAGGCTGCCCCCTCTCATCTGTGACCTCCTGTGGTTCGGGAGGGACTTGCTCTGG
ATTATATAAAGGCTTCACAAGAGTACTGGAAGCGGGGACGGAATCAACTTGAGCAAAAAAC

110/112

FIG. 24 cont'd

CTCTATCTGGTTTATAATTAGAAAGTGAATGATTTTAAATGGCTTCCACATGGCTTAGTTTTTC
CAAGCAGTTCCAGTTACATAGAACTTCATATAAACAGATAATGAGAGGCTAGAGTTGGAGA
GCACAGTTCCCTGCTTCCAAATTGGCTTGAAGTGGGGACATCTGCTCTGCATGAAGCCTGG
GTCAGGCTGCCCCCTCATCTGTGACCTCCTGTGGTCGGGAGGGACTTGCTCTGGATTAT
ATAAAGGCTTCACAAGAGTACTGGAAGCGGGGACGGAATCAACTTGAGCAAAAAACCTGA
GCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGAT
CCAGATCTTTTGAAGCTAGCGCTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCT
GTTACACGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGT
TCAGCGTGTCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTG
ATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCACCTCGTGACCACCTGGGCTA
CGGCGTGCAAGTGTTCGCCCGCTACCCCGACCATGAAGCAGCAGACTTCTTCAAGTC
CGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGA
AGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACA
ACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCA
AGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
ACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTC
CAAGCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC
CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGC
GGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGT
ATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCAT
GCTATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCTTGATAAATCCTGGTTAGTTCTTGCC
ACGGCGGAACCTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGG
CACTGACAATTCGGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGT
TGTTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCC
TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCA
CGTGCGGACCGAGCGGCCGC (SEQ ID NO: 115)

CN2840 (1617 bp between ITRs):

GCGGCCGCACGCGTGTGTCTTAGCCATTTTCAACCTCAGAGAAACACTCATTGGAAACAT
CTGAAATTGCCAGATGCCCTGAGAATTTTATTTTTAAAAACCTTTTCCCAAGAAGTTATTTAT
AATACTGTGTCAAAATGGGACACTTCAACTGTGAGTGCTAATTCCTGATGTATGTTATCATT
TTCTCCATTTGTCAATTTTTTCTAAAGGGCATTAAATTTAGCTTTTGA CTCTAAATTTAAACC
CCTCTATTTCCTGTCTCTCATTTTGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATC
TATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTGCGCCACC
ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCC
ACCCTCGTGACCACCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCAT
CTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAG
AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG

111/112

FIG. 24 cont'd

ACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
ACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATT
ACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGAT
ACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTTCTCCTCCT
TGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCATCGCCGCCTGCCTTGCCCGCTG
CTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCGTGGCTCGAGAGATCTTCGACTG
TGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCTTGACCCTGGA
AGGTGCCACTCCCCTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTA
GGTGTCAATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 116)

CN3460 (2346 bp between ITRs):

GCGGCCGCACGCGTGGTTCACTGGATTTGAATGGCGTTTGAGAAGGGGAGAAGGTGGCG
CAAGTGAACCAAAGCAACTTAACCTCATCTGCCTTCGATTCCAAGGCTGAGAAAACATTTT
CAGAAGGCCTGGTGCATAAGAACCTGATGTGGTGACAGGATGTTCTGTCTTCTCTTGTTA
CACAAGCAAGCGCGCTGAGTCTCGGCAGGCTCATAAATCCTGAGGGCCTTTTCCAGCAGC
CCCGTTTCGGGTGCGCTACCCCCATTCCAGGAAAGGGACGCTTTGAAACCCTCATCACAAG
AGAACATTGTGTCTTTGAACATCCGGCGCCACTGGGGTTTGACTTCTCCAGCGCAGGTA
GGATGGGCCAGCCAGTGAGGTCAGATGTGATTTATGCTTCTAAAGCTCTTGCTTTTTCAGAA
TTTTGGAAGAAAATACAAGCTGCCCAATACTGCAGAGCTCGGGCTGGGCATAAAAGTCAG
GGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTTGAAGCTAGCGCTA
CCGGTCGCCACCATGGTCTCCAAGGGCGAGGAGCTGTTCAACGGGGTGGTGCCCATCCT
GGTCGAGCTGGACGGCGATGTCAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGG
GCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCC
GTGCCCTGGCCACCTCGTGACCACCTGGGCTACGGCGTGCAAGTGTTCGCCCGCTA
CCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA
GGAGCGCACCATCTTCTTCAAAGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTT
CGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTATATCACCG
CCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACG
GCGGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTG
CTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGGGCTCTGGTGCTACCAACTTCTCACTGTTGAAACAGGCAGGGGAT
GTAGAGGAGAATCCAGGGCCTGGTGCTAGTGAGACTACAAAGACCATGACGGAGATTAT
AAAGATCATGACATCGATTACAAGGATGACGATGACAAGTCCGGACTCAGATCTGGAGGCT
CCGGAGGCCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGCTCCAAGAAGGCG
GTGACTAAGGCGCAGAAGAAAGGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTA
TTCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACACCGGCATTTTCTCCAAG
GCCATGGGCATCATGAATTCGTTTGTGAACGACATTTTTCGAGCGCATCGCAGGAGAGGCTT
CCCGCCTGGCGCATTACAACAAGCGCTCGACCATCACCTCCCGGGAGATCCAGACGGCC
GTGCGCCTGCTGCTGCCTGGGGAGTTGGCCAAGCACGCCGTGTCCGAGGGTACTAAGGC
CATACCAAGTACACCAGCGCTAAGTAATGAGGCGCGCCGCGGCCGCGAATTCGATATCA
TAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATATGTTGCTCC
TTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGG

112/112

FIG. 24 cont'd

CTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCATCGCC
GCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCTGTGGC
TCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGT
GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATT
GCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGC
AAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGC
CGC (SEQ ID NO: 117)

CN2650 (2002 bp between ITRs):

GCGGCCGCACGCGTACTTCTTAGGGCGAACGTTCTTTATCACAAGGAGATAAGCTGTAGG
AAGTTTGGATATAGCTTCCTGGGATCCTTGCCCGTATCCCAACTTCCTTTTTTTGCTCCTAC
TTAGTAAGAATTCTAAAGTCTGCTGGAGGATTGTTATTCAATGTAGGCACGATTGAAAGTAG
AAACCATTTTTCTCCCATATTCTGTTTCCAATTTTATACACCTACGTACGTTGTTCCAGCCT
GCTTCCTAGTCCATCCCTTTTGTGCAGTTCCTGCCTTCTAAAACCTGATCTGCAGGGGTT
AGAAAACCAGGAACTGCACAAAAGGGATGGCATCCCATTAGATCATGTGATCCTCCAGGCT
CCAGTAATAGCCCTTCCTCTGTCCTCATTAACCTTACATCTCTCCTGTGCCCTTCCTTCAAAG
TACTGCCCCCTATATGGGCCACTGGCAAATTGATTTAAATTAAGTGTTAATAGCATGTAGC
AGCCCTGCTTCCCTTCTTAAATAATGTTACATATTACATCTGCCACCAACTTGGCTTCTCC
TGTTTGCATCTATATGTCTTTTTTTTTTTTTTTTAAATGTTAGGAAGCGTATTACCTTTGGTTT
TGGATAAAATCTCTACCAGTGTCTTTTTCATTCAATGTGAGCTCGGGCTGGGCATAAAAGT
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCG
CTACCGGTGCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCAT
CCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCG
AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTG
CCCGTGCCCTGGCCACCCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCG
CTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT
CCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA
GTTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCACAACGTCTATATCA
CCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGG
ACGGCGGGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCATCGGGCAGCGGCCCC
GTGCTGCTGCCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCCGCGAATTCGATATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTT
TACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTT
TCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACATCATCGCCGCC
TGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCTGTGGCTCG
AGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCCG
(SEQ ID NO: 118)

FIG. 3A

