Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3175993 A1 2021/10/28

(21) 3 175 993

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2021/04/21
(87) **Date publication PCT/PCT Publication Date:** 2021/10/28
(85) **Entrée phase nationale/National Entry:** 2022/10/18
(86) **N° demande PCT/PCT Application No.:** US 2021/028489
(87) **N° publication PCT/PCT Publication No.:** 2021/216778
(30) **Priorité/Priority:** 2020/04/21 (US63/013,342)

(51) **Cl.Int./Int.Cl.** ***C12N 15/67*** (2006.01),
C12N 15/864 (2006.01)

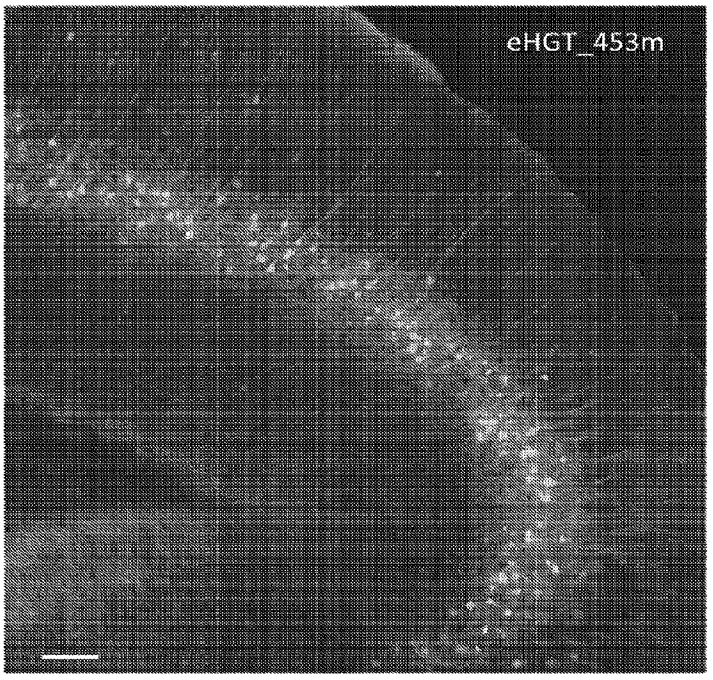
(71) **Demandeur/Applicant:**
ALLEN INSTITUTE, US

(72) **Inventeurs/Inventors:**
DAIGLE, TANYA, US;
GRAYBUCK, LUCAS T., US;
KALMBACH, BRIAN EDWARD, US;
LEIN, EDWARD SEBASTIAN, US;
LEVI, BOAZ P., US;
MICH, JOHN K., US;
SEDENO CORTES, ADRIANA ESTELA, US;
...

(74) **Agent:** GOWLING WLG (CANADA) LLP

(54) **Titre : CONSTRUCTIONS D'EXPRESSION ARTIFICIELLES POUR MODULER SELECTIVEMENT L'EXPRESSION GENIQUE DANS DES NEURONES GLUTAMATERGIQUES DE LA COUCHE NEOCORTICALE 5**
(54) **Title: ARTIFICIAL EXPRESSION CONSTRUCTS FOR SELECTIVELY MODULATING GENE EXPRESSION IN NEOCORTICAL LAYER 5 GLUTAMATERGIC NEURONS**

FIG. 4B



(57) **Abrégé/Abstract:**
Artificial expression constructs for selectively modulating gene expression in selected central nervous system cell types are described. Particularly, the artificial expression constructs can be used to selectively express synthetic genes and / or modify gene expression in neocortical glutamatergic layer 5 neurons, such as glutamatergic layer 5 extratelencephalic-projecting (L5 ET) neurons.

(21) **3 175 993**

(13) **A1**

(72) **Inventeurs(suite)/Inventors(continued):** TASIC, BOSILJKA, US; TING, JONATHAN, US; ZENG, HONGKUI, US

Date Submitted: 2022/10/18

CA App. No.: 3175993

Abstract:

Artificial expression constructs for selectively modulating gene expression in selected central nervous system cell types are described. Particularly, the artificial expression constructs can be used to selectively express synthetic genes and / or modify gene expression in neocortical glutamatergic layer 5 neurons, such as glutamatergic layer 5 extratelencephalic-projecting (L5 ET) neurons.

ARTIFICIAL EXPRESSION CONSTRUCTS FOR SELECTIVELY MODULATING GENE EXPRESSION IN NEOCORTICAL LAYER 5 GLUTAMATERGIC NEURONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to US Provisional Patent Application No. 63/013,342 filed April 21, 2020 which is incorporated herein by reference in its entirety as if fully set forth herein.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under grants MH114126 and MH121274 awarded by the National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO SEQUENCE LISTING

[0003] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is A166-0024PCT_ST25.txt. The text file is 211 KB, was created on April 21, 2021, and is being submitted electronically via EFS-Web.

FIELD OF THE DISCLOSURE

[0004] The current disclosure provides artificial expression constructs for selectively modulating gene expression in selected central nervous system cell types. The artificial expression constructs can be used to selectively express synthetic genes or modify gene expression in neocortical layer 5 glutamatergic neurons, such as layer 5 glutamatergic extratelencephalic-projecting (L5 ET) neurons.

BACKGROUND OF THE DISCLOSURE

[0005] To fully understand the biology of the brain, different cell types need to be distinguished and defined and, to further study them, artificial expression constructs that can selectively 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 selectively drive gene expression in targeted central nervous system cell populations. Targeted central nervous system cell populations include: neocortical glutamatergic layer 5 neurons, such as neocortical glutamatergic layer 5 extratelencephalic-projecting (L5 ET) neurons. In certain applications, aspects of the current disclosure can be used to deliver therapeutic payloads into upper motor neurons, for example, in the context of Amyotrophic Lateral Sclerosis (ALS) or other neurodegenerative diseases involving L5 ET neurons.

[0007] Particular embodiments of the artificial expression constructs utilize the following enhancers to selectively drive protein expression within L5 glutamatergic neurons: eHGT_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, and/or eHGT_583m.

[0008] In particular embodiments, the artificial enhancer elements include a concatenated core of an enhancer. Examples include a concatenated core of eHGT_459m, eHGT_453m, eHGT_462m, eHGT_461m, eHGT_464m, eHGT_452m, eHGT_451m, and/or eHGT_460m. 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 NOs: 45-52. In particular embodiments, these cores are concatenated and have 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core sequence. SEQ ID NOs: 53-60 provide three-copy concatemers of the selected enhancer cores.

[0010] Particular embodiments of the artificial expression constructs utilize 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and/or 3xcore2_eHGT_460m to selectively drive protein expression within L5 glutamatergic neurons.

[0011] Particular embodiments provide artificial expression constructs including the features of vectors described herein including vectors: CN2248, CN2249, CN2250, CN2251, CN2252,

CN2253, CN2254, CN2255, CN2256, CN2315, CN2423, CN2424, CN2690, CN2695, CN2691, CN2692, CN2944, CN2886, CN2755, CN2938, CN2239, CN2410, CN2411, CN2412, CN2963, CN2623, CN2624, CN2601, CN2843, CN2851, CN2853, CN2860, CN2828, CN2870, CN2871, CN2847, CN2686, CN2687, CN2878, CN2946, CN2943, CN3319, CN2918, CN3018, CN3030, CN3027, CN3032, CN3015, CN3012, and/or CN3024.

BRIEF DESCRIPTION OF THE FIGURES

[0012] Some of the drawings submitted herein are 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] FIG. 1: Overview of exemplary enhancer discovery for viral tools. To build cell type-specific labeling tools, cells from adult mouse cortex can be isolated and a single cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) can be performed. Samples can be clustered and compared to single cell RNA sequencing (scRNA-seq) datasets to identify the clusters. Single cells matching the same transcriptomic types can be then pooled and the genome can be searched for type-specific putative enhancers. Alternately, some enhancers were selected using a human motor cortex SNARE-seq dataset (Bakken et al., Nature. 2021 doi: https://doi.org/10.1101/2020.03.31.016972). The selected putative enhancer regions can be cloned upstream of a minimal promoter in an AAV genomic backbone, which can be used to generate self-complementary adeno-associated viral vectors (scAAVs) or recombinant adeno-associated viral vectors (rAAVs). These viral tools can be delivered retro-orbitally to label specific neocortical populations. In cells with a matching cell type, enhancers recruit their cognate transcription factors to drive cell type-specific expression. In other cells, viral genomes are present, but transcripts are not expressed.

[0014] FIGs. 2A, 2B. (2A) Mouse genome coordinates for enhancer eHGT_451m (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. (2B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $6.00E^{11}$GC of AAV vector #CN2249. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser. Scale bar: 100 microns.

[0015] FIGs. 3A, 3B. (3A) Mouse genome coordinates for enhancer eHGT_452m (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. (3B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of

1.45E^{12} GC of AAV vector #CN2250. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser. Scale bar: 100 microns.

[0016] FIGs. 4A-4D. (4A) 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. (4B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of 7.00E^{11}GC of AAV vector #CN2251. Scale bar: 100 microns. (4C) Mapping of single cell transcriptomic profiles of SYPF2+ 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 Crispld2 Htr2c, 60 Vip Crispld2 Kcne4, 58 Vip Col15a1 Pde1a, 54 Sst Chodl, 53 Sst Mme Fam114a1, 52 Sst Tac1 Htr1d, 50 Sst Tac1 Tacr3, 49 Sst Calb2 Necab1, 48 Sst Calb2 Pdlim5, 46 Sst Nr2f2 Necab1, 45 Sst Myh8 Etv1, 44 Sst 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 Tpbg, 4 Pvalb Vipr2,

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. (4D) 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 show are from the UCSC Genome Browser. Scale bar: 500 µm.

[0017] FIGs. 5A-5D. (5A) Mouse genome coordinates for enhancer eHGT_459m (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. (5B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $1.00E^{12}$GC of AAV vector #CN2254. Scale bar: 100 microns. (5C) Mapping of single cell transcriptomic profiles of SYPF2+ cells sorted from the visual cortex region of the mouse brain after retro-orbital injection of CN2254 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_459m enhancer driven reporter expression occurs predominantly in only 1 of 4 types of L5 ET neurons in the visual cortex region. The text along the bottom from left to right reads the same as described for FIG. 4C. (5D) Epifluorescence microscopy image of native SYFP2 fluorescence in a fixed brain slice from macaque primary motor cortex at 113 days after stereotaxic injection of enhancer AAV vector CN2254 serotype PHP.eB. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser. Scale bar: 500 µm.

[0018] FIGs. 6A, 6B. (6A) Mouse genome coordinates for enhancer eHGT_460m (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. (6B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $6.50E^{11}$GC of AAV vector #CN2255. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser. Scale bar: 100 microns.

[0019] FIGs. 7A, 7B. (7A) Mouse genome coordinates for enhancer eHGT_462m (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. (7B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $6.50E^{11}$GC of AAV vector #CN2256. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser. Scale bar: 100 microns.

[0020] FIGs. 8A-8D. (8A) Mouse genome coordinates for enhancer eHGT_464m (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. (8B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $1.00E^{12}$GC of AAV vector #CN2315. Scale bar: 100 microns. (8C) Mapping of single cell transcriptomic profiles of SYPF2+ cells sorted from the visual cortex region of the mouse brain after retro-orbital injection of CN2315 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_464m enhancer driven reporter expression occurs selectively in multiple types of L5 ET neurons in the visual cortex region. The text along the bottom from left to right reads the same as described for FIG. 4C. (8D) 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 CN2315 serotype PHP.eB. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser. Scale bar: 500 µm.

[0021] FIGs. 9A, 9B. (9A) Mouse genome coordinates for enhancer eHGT_461m (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 $5.00E^{11}$GC of AAV vector #CN2423. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser. Scale bar: 100 microns.

[0022] FIGs. 10A-10C. (10A) Mouse genome coordinates for enhancer eHGT_463m (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. (10B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 2 months after retro-orbital delivery of $7.30E^{11}$GC of AAV vector #CN2424. Scale bar: 100 microns. (10C) Mapping of single cell transcriptomic profiles of SYPF2+ cells sorted from the visual cortex region of the mouse brain after retro-orbital injection of CN2424 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_463m enhancer driven reporter expression occurs selectively in 1 of 4 types of L5 ET neurons in the visual cortex region. The text along the bottom from left to right reads the same as described for FIG. 4C. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser.

[0023] FIGs. 11A, 11B. (11A) Mouse genome coordinates for enhancer eHGT_647m (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. (11B) Epifluorescence micrograph image showing native SYFP2 expression in the visual cortex 1.5 months after retro-orbital delivery of 7.00E^{11} GC of AAV vector #CN2847. The genome coordinates are for build Hg38 and the views show are from the UCSC Genome Browser. Scale bar: 100 microns

[0024] FIGs. 12A-12C. Sequences supporting the disclosure. (12A) enhancer sequences disclosed herein. Each artificial expression construct described herein includes at least one of the enhancer sequences provided in FIG. 12A; (12B) exemplary components that can be included within (or encoded by) artificial expression constructs generated with enhancers of (12A); (12C) representative artificial expression constructs generated with the enhancers of (12A) and selected components of (12B).

DETAILED DESCRIPTION

[0025] 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 selectively 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.

[0026] The current disclosure provides artificial expression constructs that selectively drive gene expression in targeted central nervous system cell populations. Targeted central nervous system cell populations include: neocortical glutamatergic layer 5 neurons, such as neocortical glutamatergic extratelencephalic-projecting (L5 ET) neurons. In certain applications, aspects of the current disclosure can be used to deliver therapeutic payloads into upper motor neurons, for example, in the context of Amyotrophic Lateral Sclerosis (ALS) or other neurodegenerative diseases involving L5 ET neurons.

[0027] Particular embodiments of the artificial expression constructs utilize the following enhancers to selectively drive gene expression within L5 ET neurons: eHGT_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h,

eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, and/or eHGT_583m.

[0028] In particular embodiments, the artificial enhancer elements include a concatenated core of an enhancer disclosed herein, such as a concatenated core of eHGT_459m, eHGT_453m, eHGT_462m, eHGT_461m, eHGT_464m, eHGT_452m, eHGT_451m, and/or eHGT_460m. These artificial enhancer elements provide higher levels and more rapid onset of transgene expression compared to a single full length original (native) enhancer.

[0029] In particular embodiments, the enhancer core can include a core sequence as set forth in any one of SEQ ID NOs: 45-52. In particular embodiments, the cores are concatenated and include 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the core. SEQ ID NOs: 53-60 provide three-copy concatemers of selected enhancer cores.

[0030] Particular embodiments of the artificial expression constructs utilize 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and/or 3xcore2_eHGT_460m to selectively drive protein expression within L5 glutamatergic neurons.

[0031] Particular embodiments provide artificial expression constructs including the features of vectors described herein including vectors: CN2248, CN2249, CN2250, CN2251, CN2252, CN2253, CN2254, CN2255, CN2256, CN2315, CN2423, CN2424, CN2690, CN2695, CN2691, CN2692, CN2944, CN2886, CN2755, CN2938, CN2239, CN2410, CN2411, CN2412, CN2963, CN2623, CN2624, CN2601, CN2843, CN2851, CN2853, CN2860, CN2828, CN2870, CN2871, CN2847, CN2686, CN2687, CN2878, CN2946, CN2943, CN3319, CN2918, CN3018, CN3030, CN3027, CN3032, CN3015, CN3012, and/or CN3024.

[0032] Aspects of the disclosure are now described with the following additional options and detail: (i) Artificial Expression Constructs & Vectors for Selective Expression of Genes in Selected 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.

[0033] (i) Artificial Expression Constructs & Vectors for Selective Expression of Genes in Selected Cell Types. Artificial expression constructs disclosed herein include (i) an enhancer sequence that leads to selective 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.

[0034] 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_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m and concatenated cores, such as 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and 3xcore2_eHGT_460m.

[0035] 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 cell type but does not substantially direct expression of genes in other non-targeted cell types, thus having cell type specific transcriptional activity.

[0036] When a coding sequence is selectively expressed in selected cells and is not substantially expressed in other cell types, the product of the coding sequence is preferentially expressed in the selected cell type. In particular embodiments, preferential expression is greater than 50% expression as compared to a reference cell type; greater than 60% expression as compared to a reference cell type; greater than 70% expression as compared to a reference cell type; greater than 80% expression as compared to a reference cell type; or greater than 90% expression as compared to a reference 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.

[0037] 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.

[0038] 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:

[0039] Neocortical GABAergic Subclasses:

- All: Express GABA synthesis genes Gad1/GAD1 and Gad2/GAD2.
- Lamp5, Sncg, Serpinf1, and *Vip*: Developmentally derived from neuronal progenitors from the caudal ganglionic eminence (CGE) or preoptic area (POA).
- *Sst* and Pvalb: Developmentally derived from neuronal progenitors in the medial ganglionic eminence (MGE).
- *Lamp5*: Found in many neocortical layers, especially upper (L1-L2/3), and have mainly neurogliaform and single bouquet morphology.
- Sncg: Found in many neocortical layers, and have molecular overlaps with Lamp5 and *Vip* cells, but inconsistent expression of Lamp5 or *Vip*, with more consistent expression of Sncg.
- Serpinf1: 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*: Found in many neocortical layers, but especially frequent in upper layers (L1-L4), and highly express the neurotransmitter vasoactive intestinal peptide (*Vip*).
- *Sst*: 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+ cells.
- *Pvalb*: 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 cells 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 type that expresses *Meis2* gene, and does not express some other genes that are expressed by all other neocortical GABAergic types (for example, *Thy1* and *Scn2b*). This type is found in L6b and subcortical white matter.

[0040] Neocortical Glutamatergic Subclasses:

- All: Express glutamate transmitters Slc17a6 and/or Slc17a7. They all express Snap25 and lack expression of Gad1/Gad2.
- L2/3 IT: Primarily reside in Layer 2/3 and have mainly intratelencephalic (cortico-cortical) projections.
- L4 IT: Primarily reside in Layer 4 and mainly have either local or intratelencephalic (cortico-cortical) projections.
- L5 IT: Primarily reside in Layer 5 and have mainly intratelencephalic (cortico-cortical) projections. Also called L5a.
- L5 ET: Primarily reside in Layer 5 and have mainly cortico-subcortical projections. Also called L5b or L5 CF (corticofugal) or L5 PT (pyramidal tract projecting). 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: Primarily reside in Layer 5 and have mainly nearby projections.
- L6 CT: Primarily reside in Layer 6 and have mainly cortico-thalamic projections.
- L6 IT: Primarily reside in Layer 6 and have mainly intratelencephalic (cortico-cortical) projections. Included in this subclass are L6 IT Car3 cells, which are highly similar to intracortical-projecting cells in the claustrum.
- L6b: 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: A distinct subclass defined by a single type in L1, Cajal-Retzius cells express distinct molecular markers Lhx5 and Trp73.

[0041] 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.

[0042] Deep cerebellar nuclear neurons: neurons located in the deep cerebellar nuclear structure. These include glutamatergic and GABAergic cells that express the gene *Pvalb*.

[0043] Non-neuronal Subclasses:

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

[0044] 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.

[0045] Exemplary reporter genes/proteins include those expressed by Addgene ID#s 83894 (pAAV-hDlx-Flex-dTomato-Fishell_7), 83895 (pAAV-hDlx-Flex-GFP-Fishell_6), 83896 (pAAV-hDlx-GiDREADD-dTomato-Fishell-5), 83898 (pAAV-mDlx-ChR2-mCherry-Fishell-3), 83899 (pAAV-mDlx-GCaMP6f-Fishell-2), 83900 (pAAV-mDlx-GFP-Fishell-1), and 89897 (pcDNA3-FLAG-mTET2 (N500)). Exemplary reporter genes particularly can include those which encode an expressible fluorescent protein, or expressible biotin; blue fluorescent proteins (*e.g.* eBFP, eBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire); cyan fluorescent proteins (*e.g.* eCFP, Cerulean, CyPet, AmCyanl, Midoriishi-Cyan, mTurquoise); green fluorescent proteins (*e.g.* GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green (mAzamigreen), CopGFP, AceGFP, avGFP, ZsGreenl, Oregon Green™(Thermo Fisher Scientific)); Luciferase; orange fluorescent proteins (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato, dTomato); red fluorescent proteins (mKate, mKate2, mPlum, DsRed monomer, mCherry, mRuby, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRedl, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred, Texas Red™ (Thermo Fisher Scientific)); far red fluorescent proteins (e.g., mPlum and mNeptune); yellow fluorescent proteins (*e.g.*, YFP, eYFP, Citrine, SYFP2, Venus, YPet, PhiYFP, ZsYellowl); and tandem conjugates.

[0046] GFP is composed of 238 amino acids (26.9 kDa), originally isolated from the jellyfish Aequorea victoria/Aequorea aequorea/Aequorea forskalea that fluoresces green when exposed to blue light. The GFP from A. victoria has a major excitation peak at a wavelength of 395 nm and a minor one at 475 nm. Its emission peak is at 509 nm which is in the lower green portion of the visible spectrum. The GFP from the sea pansy (Renilla reniformis) has a single major excitation peak at 498 nm. Due to the potential for widespread usage and the evolving needs of researchers, many different mutants of GFP have been engineered. The first major improvement was a single point mutation (S65T) reported in 1995 in Nature by Roger Tsien. This mutation dramatically improved the spectral characteristics of GFP, resulting in increased fluorescence, photostability and a shift of the major excitation peak to 488 nm with the peak emission kept at 509 nm. The addition of the 37°C folding efficiency (F64L) point mutant to this scaffold yielded enhanced GFP (EGFP). EGFP has an extinction coefficient (denoted ε), also known as its optical cross section of 9.13X10-21 m^2/molecule, also quoted as 55,000 L/(mol•cm). Superfolder GFP, a series of mutations that allow GFP to rapidly fold and mature even when fused to poorly folding peptides, was reported in 2006.

[0047] 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.

[0048] Exemplary functional molecules include functioning ion transporters, cellular trafficking proteins, enzymes, transcription factors, neurotransmitters, calcium reporters, channelrhodopsins, guide RNA, nucleases, or designer receptors exclusively activated by designer drugs (DREADDs).

[0049] Ion transporters are transmembrane proteins that mediate transport of ions across cell membranes. These transporters are pervasive throughout most cell types and important for regulating cellular excitability and homeostasis. Ion transporters participate in numerous cellular processes such as action potentials, synaptic transmission, hormone secretion, and muscle contraction. Many important biological processes in living cells involve the translocation of cations, such as calcium (Ca2+), potassium (K+), and sodium (Na+) ions, through such ion channels. In particular embodiments, ion transporters include voltage gated sodium channels (e.g., SCN1A), potassium channels (e.g., KCNQ2), and calcium channels (e.g. CACNA1C)).

[0050] Exemplary enzymes, transcription factors, receptors, membrane proteins, cellular trafficking proteins, signaling molecules, and neurotransmitters include enzymes such as lactase, lipase, helicase, alpha-glucosidase, amylase; transcription factors such as SP1, AP-1, Heat shock factor protein 1, C/EBP (CCAA-T/enhancer binding protein), and Oct-1; receptors such as transforming growth factor receptor beta 1, platelet-derived growth factor receptor, epidermal growth factor receptor, vascular endothelial growth factor receptor, and interleukin 8 receptor alpha; membrane proteins, cellular trafficking proteins such as clathrin, dynamin, caveolin, Rab-4A, and Rab-11A; signaling molecules such as nerve growth factor (NGF), platelet-derived growth factor (PDGF), transforming growth factor β (TGFβ), epidermal growth factor (EGF), GTPase and HRas; and neurotransmitters such as cocaine and amphetamine regulated transcript, substance P, oxytocin, and somatostatin.

[0051] In particular embodiments, functional molecules include reporters of cell function and states such as calcium reporters. Intracellular calcium concentration is an important predictor of numerous cellular activities, which include neuronal activation, muscle cell contraction and second messenger signaling. A sensitive and convenient technique to monitor the intracellular calcium levels is through the genetically encoded calcium indicator (GECI). Among the GECIs, green fluorescent protein (GFP) based calcium sensors named GCaMPs are efficient and widely used tools. The GCaMPs are formed by fusion of M13 and calmodulin protein to N- and C-termini of circularly permutated GFP. Some GCaMPs yield distinct fluorescence emission spectra (Zhao et al.,Science, 2011, 333(6051): 1888-1891). Exemplary GECIs with green fluorescence include GCaMP3, GCaMP5G, GCaMP6s, GCaMP6m, GCaMP6f, jGCaMP7s, jGCaMP7c, jGCaMP7b, and jGCaMP7f. Furthermore, GECIs with red fluorescence include jRGECO1a and jRGECO1b.

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

[0052] 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).

[0053] In particular embodiments, functional molecules include modulators of neuronal activity like channelrhodopsins (e.g., channelrhodopsin-1, channelrhodopsin-2, and variants thereof). Channelrhodopsins are a subfamily of retinylidene proteins (rhodopsins) that function as light-gated ion channels. In addition to channelrhodopsin 1 (ChR1) and channelrhodopsin 2 (ChR2), several variants of channelrhodopsins have been developed. For example, Lin et al. (*Biophys J,* 2009, 96(5): 1803-14) describe making chimeras of the transmembrane domains of ChR1 and ChR2, combined with site-directed mutagenesis. Zhang et al. (*Nat Neurosci,* 2008, 11(6): 631-3) describe VChR1, which is a red-shifted channelrhodopsin variant. VChR1 has lower light sensitivity and poor membrane trafficking and expression. Other known channelrhodopsin variants include the ChR2 variant described in Nagel, et al., *Proc Natl Acad Sci USA,* 2003, 100(24): 13940-5), ChR2/H134R (Nagel, G., et al., *Curr Biol,* 2005, 15(24): 2279-84), and ChD/ChEF/ChIEF (Lin, J. Y., et al., *Biophys J,* 2009, 96(5): 1803-14), which are activated by blue light (470 nm) but show no sensitivity to orange/red light. Additional variants are described in Lin, *Experimental Physiology,* 2010, 96.1: 19-25 and Knopfel et al., *The Journal of Neuroscience,* 2010, 30(45): 14998-15004).

[0054] 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 or homologous recombination donor cassettes.

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

[0056] Additional effector elements include Cre, iCre, dgCre, FlpO, and tTA2. iCre refers to a

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

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

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

[0059] 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.

[0060] 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.

[0061] Promoters can include general promoters, tissue-specific promoters, cell-specific promoters, and/or promoters specific for the cytoplasm. Promoters may include strong promoters,

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

[0062] 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.

[0063] 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.

[0064] 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.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 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.

[0069] 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.

[0070] AAVs stand out for use within the current disclosure because of their superb safety profile and because their capsids and genomes can be tailored to allow expression in selected cell populations. scAAV refers to a self-complementary AAV. pAAV refers to a plasmid adeno-associated virus. rAAV refers to a recombinant adeno-associated virus.

[0071] 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.

[0072] 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.

[0073] 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.

[0074] "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.

[0075] 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.

[0076] 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.

[0077] 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.

[0078] 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, WPRE3, or HPRE. In particular embodiments, vectors lack or do not include a posttranscriptional regulatory element such as a WPRE, WPRE3, or HPRE.

[0079] 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 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.

[0080] In particular embodiments, a viral vector further includes one or more insulator elements. Insulators elements may contribute to protecting viral vector-expressed sequences, *e.g.*, effector elements or expressible elements, from integration site effects, which may be mediated by cis-acting elements present in genomic DNA and lead to deregulated expression of transferred sequences (*i.e.*, position effect; see, *e.g.*, Burgess-Beusse *et al.*, *PNAS., USA*, 99:16433, 2002; and Zhan *et al.*, *Hum. Genet.*, 109:471, 2001). In particular embodiments, viral transfer vectors include one or more insulator elements at the 3' LTR and upon integration of the provirus into the host genome, the provirus includes the one or more insulators at both the 5' LTR and 3' LTR, by virtue of duplicating the 3' LTR. Suitable insulators for use in particular embodiments include the chicken β-globin insulator (see Chung *et al.*, *Cell* 74:505, 1993; Chung *et al.*, *PNAS USA* 94:575, 1997; and Bell *et al.*, *Cell* 98:387, 1999), SP10 insulator (Abhyankar et al., *JBC* 282:36143, 2007), or other small CTCF recognition sequences that function as enhancer blocking insulators (Liu et al., *Nature Biotechnology*, 33:198, 2015).

[0081] Beyond the foregoing description, a wide range of suitable expression vector types will be known to a person of ordinary skill in the art. These can include commercially available expression vectors designed for general recombinant procedures, for example plasmids that contain one or more reporter genes and regulatory elements required for expression of the reporter gene in cells. Numerous vectors are commercially available, e.g., from Invitrogen, Stratagene, Clontech, etc., and are described in numerous associated guides. In particular embodiments, suitable expression vectors include any plasmid, cosmid or phage construct that is capable of supporting expression

of encoded genes in mammalian cell, such as pUC or Bluescript plasmid series.

[0082] Particular embodiments of vectors disclosed herein include:

Expression Construct Name	Features
CN2248	rAAV-eHGT_450m-minBglobin-SYFP2-WPRE3-BGHpA
CN2249	rAAV-eHGT_451m-minBglobin-SYFP2-WPRE3-BGHpA
CN2250	rAAV-eHGT_452m-minBglobin-SYFP2-WPRE3-BGHpA
CN2251	rAAV-eHGT_453m-minBglobin-SYFP2-WPRE3-BGHpA
CN2252	rAAV-eHGT_457m-minBglobin-SYFP2-WPRE3-BGHpA
CN2253	rAAV-eHGT_458m-minBglobin-SYFP2-WPRE3-BGHpA
CN2254	rAAV-eHGT_459m-minBglobin-SYFP2-WPRE3-BGHpA
CN2255	rAAV-eHGT_460m-minBglobin-SYFP2-WPRE3-BGHpA
CN2256	rAAV-eHGT_462m-minBglobin-SYFP2-WPRE3-BGHpA
CN2315	rAAV-eHGT_464m-minBglobin-SYFP2-WPRE3-BGHpA
CN2423	rAAV-eHGT_461m-minBglobin-SYFP2-WPRE3-BGHpA
CN2424	rAAV-eHGT_463m-minBglobin-SYFP2-WPRE3-BGHpA
CN2690	rAAV-eHGT_692h-minBglobin-SYFP2-WPRE3-BGHpA
CN2695	rAAV-eHGT_693h-minBglobin-SYFP2-WPRE3-BGHpA
CN2691	rAAV-eHGT_694h-minBglobin-SYFP2-WPRE3-BGHpA
CN2692	rAAV-eHGT_695h-minBglobin-SYFP2-WPRE3-BGHpA
CN2944	rAAV-eHGT_667m-minBglobin-SYFP2-WPRE3-BGHpA
CN2886	rAAV-eHGT_668m-minBglobin-SYFP2-WPRE3-BGHpA
CN2755	rAAV-eHGT_696m-minBglobin-SYFP2-WPRE3-BGHpA
CN2938	rAAV-eHGT_455m-minBglobin-SYFP2-WPRE3-BGHpA
CN2239	rAAV-eHGT_457h-minBglobin-SYFP2-WPRE3-BGHpA
CN2410	rAAV-eHGT_603m-minBG-SYFP2-WPRE3-BGHpA
CN2411	rAAV-eHGT_604m-minBG-SYFP2-WPRE3-BGHpA
CN2412	rAAV-eHGT_605m-minBG-SYFP2-WPRE3-BGHpA
CN2963	rAAV-eHGT_661m-minBglobin-SYFP2-WPRE3-BGHpA
CN2623	rAAV-eHGT_766m-minBglobin-SYFP2-WPRE3-BGHpA
CN2624	rAAV-eHGT_767m-minBglobin-SYFP2-WPRE3-BGHpA
CN2601	rAAV-eHGT_768m-minBglobin-SYFP2-WPRE3-BGHpA
CN2843	rAAV-eHGT_663m-minBglobin-SYFP2-WPRE3-BGHpA
CN2851	rAAV-eHGT_660m-minBglobin-SYFP2-WPRE3-BGHpA
CN2853	rAAV-eHGT_665m-minBglobin-SYFP2-WPRE3-BGHpA
CN2860	rAAV-eHGT_700m-minBglobin-SYFP2-WPRE3-BGHpA
CN2828	rAAV-eHGT_456m-minBglobin-SYFP2-WPRE3-BGHpA
CN2870	rAAV-eHGT_698m-minBglobin-SYFP2-WPRE3-BGHpA
CN2871	rAAV-eHGT_699m-minBglobin-SYFP2-WPRE3-BGHpA
CN2847	rAAV-eHGT_647m-minBglobin-SYFP2-WPRE3-BGHpA
CN2686	rAAV-eHGT_648m-minBglobin-SYFP2-WPRE3-BGHpA
CN2687	rAAV-eHGT_649m-minBglobin-SYFP2-WPRE3-BGHpA
CN2878	rAAV-eHGT_670m-minBglobin-SYFP2-WPRE3-BGHpA
CN2946	rAAV-eHGT_697m-minBglobin-SYFP2-WPRE3-BGHpA
CN2943	rAAV-eHGT_662m-minBglobin-SYFP2-WPRE3-BGHpA
CN3319	rAAV-eHGT_583m-minBglobin-SYFP2-WPRE3-BGHpA
CN2918	rAAV-3xcore2_eHGT_459m-minBglobin-SYFP2-WPRE3-BGHpA
CN3018	rAAV-3xcore2_eHGT_453m-minBglobin-SYFP2-WPRE3-BGHpA

CN3030	rAAV-3xcore2_eHGT_462m-minBglobin-SYFP2-WPRE3-BGHpA
CN3027	rAAV-3xcore2_eHGT_461m-minBglobin-SYFP2-WPRE3-BGHpA
CN3032	rAAV-3xcore2_eHGT_464m-minBglobin-SYFP2-WPRE3-BGHpA
CN3015	rAAV-3xcore2_eHGT_452m-minBglobin-SYFP2-WPRE3-BGHpA
CN3012	rAAV-3xcore2_eHGT_451m-minBglobin-SYFP2-WPRE3-BGHpA
CN3024	rAAV-3xcore2_eHGT_460m-minBglobin-SYFP2-WPRE3-BGHpA

[0083] Subcomponent sequences within the larger vector sequences can be readily identified by one of ordinary skill in the art and based on the contents of the current disclosure (see FIG. 12C). Nucleotides between identifiable and enumerated subcomponents reflect restriction enzyme recognition sites used in assembly (cloning) of the constructs, and in some cases, additional nucleotides do not convey any identifiable function. These segments of complete vector sequences can be adjusted based on use of different cloning strategies and/or vectors. In general, short 6-nucleotide palindromic sequences reflect vector construction artifacts that are not important to vector function.

[0084] In particular embodiments vectors (e.g., AAV) with capsids that cross the blood-brain barrier (BBB) are selected. In particular embodiments, vectors are modified to include capsids that cross the BBB. Examples of AAV with viral capsids that cross the blood brain barrier include AAV9 (Gombash et al., Front Mol Neurosci. 2014; 7:81), AAVrh.10 (Yang, et al., Mol Ther. 2014; 22(7): 1299-1309), AAV1R6, AAV1R7 (Albright et al., Mol Ther. 2018; 26(2): 510), rAAVrh.8 (Yang, et al., supra), AAV-BR1 (Marchio et al., EMBO Mol Med. 2016; 8(6): 592), AAV-PHP.S (Chan et al., Nat Neurosci. 2017; 20(8): 1172), AAV-PHP.B (Deverman et al., Nat Biotechnol. 2016; 34(2): 204), AAV-PPS (Chen et al., Nat Med. 2009; 15: 1215), and PHP.eB. In particular embodiments, the PHP.eB capsid differs from AAV9 such that, using AAV9 as a reference, amino acids starting at residue 586: S-AQ-A (SEQ ID NO: 144) are changed to S-DGTLAVPFK-A (SEQ ID NO: 145). In particular embodiments, PHP.eb refers to SEQ ID NO: 85.

[0085] AAV9 is a naturally occurring AAV serotype that, unlike many other naturally occurring serotypes, can cross the BBB following intravenous injection. It transduces large sections of the central nervous system (CNS), thus permitting minimally invasive treatments (Naso et al., BioDrugs. 2017; 31(4): 317), for example, as described in relation to clinical trials for the treatment of spinal muscular atrophy (SMA) syndrome by AveXis (AVXS-101, NCT03505099) and the treatment of CLN3 gene-Related Neuronal Ceroid-Lipofuscinosis (NCT03770572).

[0086] AAVrh.10, was originally isolated from rhesus macaques and shows low seropositivity in humans when compared with other common serotypes used for gene delivery applications (Selot et al., Front Pharmacol. 2017; 8: 441) and has been evaluated in clinical trials LYS-SAF302, LYSOGENE, and NCT03612869.

[0087] AAV1R6 and AAV1R7, two variants isolated from a library of chimeric AAV vectors (AAV1 capsid domains swapped into AAVrh.10), retain the ability to cross the BBB and transduce the CNS while showing significantly reduced hepatic and vascular endothelial transduction.

[0088] rAAVrh.8, also isolated from rhesus macaques, shows a global transduction of glial and neuronal cell types in regions of clinical importance following peripheral administration and also displays reduced peripheral tissue tropism compared to other vectors.

[0089] AAV-BR1 is an AAV2 variant displaying the NRGTEWD (SEQ ID NO: 146) epitope that was isolated during in vivo screening of a random AAV display peptide library. It shows high specificity accompanied by high transgene expression in the brain with minimal off-target affinity (including for the liver) (Körbelin et al., EMBO Mol Med. 2016; 8(6): 609).

[0090] AAV-PHP.S (Addgene, Watertown, MA) is a variant of AAV9 generated with the CREATE method that encodes the 7-mer sequence QAVRTSL (SEQ ID NO: 147), transduces neurons in the enteric nervous system, and strongly transduces peripheral sensory afferents entering the spinal cord and brain stem.

[0091] AAV-PHP.B (Addgene, Watertown, MA) is a variant of AAV9 generated with the CREATE method that encodes the 7-mer sequence TLAVPFK (SEQ ID NO: 148). It transfers genes throughout the CNS with higher efficiency than AAV9 and transduces the majority of astrocytes and neurons across multiple CNS regions.

[0092] AAV-PPS, an AAV2 variant crated by insertion of the DSPAHPS (SEQ ID NO: 149) epitope into the capsid of AAV2, shows a dramatically improved brain tropism relative to AAV2.

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

[0094] (ii) Compositions for Administration. Artificial expression constructs and vectors of the present disclosure (referred to herein as physiologically active components) can be formulated with a carrier that is suitable for administration to a cell, tissue slice, animal (e.g., mouse, non-human primate), or human. Physiologically active components within compositions described herein can be prepared in neutral forms, as freebases, or as pharmacologically acceptable salts.

[0095] Pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

[0096] Carriers of physiologically active components can include solvents, dispersion media,

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

[0097] The phrase "pharmaceutically-acceptable carriers" refer to carriers that do not produce an allergic or similar untoward reaction when administered to a human, and in particular embodiments, when administered intravenously (e.g. at the retro-orbital plexus).

[0098] In particular embodiments, compositions can be formulated for intravenous, intraparenchymal, intraocular, intravitreal, parenteral, subcutaneous, intracerebro-ventricular, intramuscular, intrathecal, intraspinal, intraperitoneal, oral or nasal inhalation, or by direct injection in or application to one or more cells, tissues, or organs.

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

[0100] The formation and use of liposomes is generally known to those of skill in the art. Liposomes have been developed with improved serum stability and circulation half-times (see, for instance, U.S. Pat. No. 5,741,516). Further, various methods of liposome and liposome like preparations as potential drug carriers have been described (see, for instance U.S. Pat. Nos. 5,567,434; 5,552,157; 5,565,213; 5,738,868; and 5,795,587).

[0101] The disclosure also provides for pharmaceutically acceptable nanocapsule formulations of the physiologically active components. Nanocapsules can generally entrap compounds in a stable and reproducible way (Quintanar-Guerrero *et al.*, *Drug Dev Ind Pharm* 24(12):1113-1128, 1998; Quintanar-Guerrero *et al.*, *Pharm Res.* 15(7):1056-1062, 1998; Quintanar-Guerrero *et al.*, J. *Microencapsul.* 15(1):107-119, 1998; Douglas *et al.*, *Crit Rev Ther Drug Carrier Syst* 3(3):233-261, 1987). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles can be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use in the present disclosure. Such particles can be easily made, as described in Couvreur *et al.*, *J Pharm Sci* 69(2):199-202, 1980; Couvreur *et al.*, *Crit Rev Ther Drug Carrier Syst.* 5(1)1-20, 1988; zur Muhlen *et al.*, *Eur J Pharm Biopharm*, 45(2):149-155, 1998; Zambaux *et al.*, *J Control Release* 50(1-3):31-40, 1998; and U.S. Pat. No. 5,145,684.

[0102] Injectable compositions can include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (U.S. Pat. No. 5,466,468). For delivery via injection, the form is sterile and fluid to the extent that it can be delivered by syringe. In particular embodiments, it is stable under the conditions of

manufacture and storage, and optionally contains one or more preservative compounds against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion, and/or by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and/or antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In various embodiments, the preparation will include an isotonic agent(s), for example, sugar(s) or sodium chloride. Prolonged absorption of the injectable compositions can be accomplished by including in the compositions of agents that delay absorption, for example, aluminum monostearate and gelatin. Injectable compositions can be suitably buffered, if necessary, and the liquid diluent first rendered isotonic with sufficient saline or glucose.

[0103] Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. As indicated, under ordinary conditions of storage and use, these preparations can contain a preservative to prevent the growth of microorganisms.

[0104] Sterile compositions can be prepared by incorporating the physiologically active component in an appropriate amount of a solvent with other optional ingredients (e.g., as enumerated above), followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized physiologically active components into a sterile vehicle that contains the basic dispersion medium and the required other ingredients (e.g., from those enumerated above). In the case of sterile powders for the preparation of sterile injectable solutions, preferred methods of preparation can be vacuum-drying and freeze-drying techniques which yield a powder of the physiologically active components plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0105] Oral compositions may be in liquid form, for example, as solutions, syrups or suspensions, or may be presented as a drug product for reconstitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g., lecithin or acacia); non-aqueous vehicles (e.g., almond oil, oily esters, or fractionated vegetable oils); and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically

acceptable excipients such as binding agents (e.g., pregelatinized maize starch, polyvinyl pyrrolidone or hydroxypropyl methylcellulose); fillers (e.g., lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g., magnesium stearate, talc or silica); disintegrants (e.g., potato starch or sodium starch glycolate); or wetting agents (e.g., sodium lauryl sulphate). Tablets may be coated by methods well-known in the art.

[0106] Inhalable compositions can be delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

[0107] Compositions can also include microchip devices (U.S. Pat. No. 5,797,898), ophthalmic formulations (Bourlais *et al.*, *Prog Retin Eye Res*, 17(1):33-58, 1998), transdermal matrices (U.S. Pat. No. 5,770,219 and U.S. Pat. No. 5,783,208) and feedback-controlled delivery (U.S. Pat. No. 5,697,899).

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

[0109] Typically, compositions can include at least 0.1% of the physiologically active components or more, although the percentage of the physiologically active components may, of course, be varied and may conveniently be between 1 or 2% and 70% or 80% or more or 0.5-99% of the weight or volume of the total composition. Naturally, the amount of physiologically active components in each physiologically-useful composition may be prepared in such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of compositions and dosages may be desirable.

[0110] In particular embodiments, for administration to humans, compositions should meet sterility, pyrogenicity, and the general safety and purity standards as required by United States Food and Drug Administration (FDA) or other applicable regulatory agencies in other countries.

[0111] (iii) Cell Lines Including Artificial Expression Constructs. The present disclosure includes cells including an artificial expression construct described herein. A cell that has been transformed with an artificial expression construct can be used for many purposes, including in neuroanatomical studies, assessments of functioning and/or non-functioning proteins, and drug

screens that assess the regulatory properties of enhancers.

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_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and/or 3xcore2_eHGT_460m and/or CN2248, CN2249, CN2250, CN2251, CN2252, CN2253, CN2254, CN2255, CN2256, CN2315, CN2423, CN2424, CN2690, CN2695, CN2691, CN2692, CN2944, CN2886, CN2755, CN2938, CN2239, CN2410, CN2411, CN2412, CN2963, CN2623, CN2624, CN2601, CN2843, CN2851, CN2853, CN2860, CN2828, CN2870, CN2871, CN2847, CN2686, CN2687, CN2878, CN2946, CN2943, CN3319, CN2918, CN3018, CN3030, CN3027, CN3032, CN3015, CN3012, and/or CN3024, and the cell line is a human, primate, or murine cell. 1, 2, 4, 5, 6, 7, 8, 9, or 10 copy concatemers of disclosed enhancer cores can also be used. Cell lines which can be utilized for transgenesis in the present disclosure also include primary cell lines derived from living tissue such as rat or mouse brains and organotypic cell cultures, including brain slices from animals such as rats or mice. The PC12 cell line (available from the American Type Culture Collection, ATCC, Manassas, VA) has been shown to express a number of neuronal marker proteins in response to Neuronal Growth Factor (NGF). The PC12 cell line is considered to be a neuronal cell line and is applicable for use with this disclosure. JAR cells (available from ATCC) are a platelet derived cell-line that express some neuronal genes, such as the serotonin transporter gene, and may be used with embodiments described herein.

[0112] 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.

[0113] 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.

[0114] 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).

[0115] 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

[0116] 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).

[0117] 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_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, and their cores.

[0118] 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).

[0119] (iv) Transgenic Animals. Another aspect of the disclosure includes transgenic animals, the genome of which contains an artificial expression construct including eHGT_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and/or 3xcore2_eHGT_460m operatively linked to a heterologous coding sequence. 1, 2, 4, 5, 6, 7, 8, 9, or 10 copy concatemers of disclosed enhancer cores can also be used. In particular embodiments, the genome of a transgenic animal includes CN2248, CN2249, CN2250, CN2251, CN2252, CN2253, CN2254, CN2255, CN2256, CN2315, CN2423, CN2424, CN2690, CN2695, CN2691, CN2692, CN2944, CN2886, CN2755, CN2938, CN2239, CN2410, CN2411, CN2412, CN2963, CN2623, CN2624, CN2601, CN2843, CN2851, CN2853, CN2860, CN2828, CN2870, CN2871, CN2847, CN2686, CN2687, CN2878, CN2946, CN2943, CN3319, CN2918, CN3018, CN3030, CN3027, CN3032, CN3015, CN3012, and/or CN3024. In particular embodiments, when a non-integrating vector is utilized, a transgenic animal includes an artificial

expression construct eHGT_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and/or 3xcore2_eHGT_460m and/or CN2248, CN2249, CN2250, CN2251, CN2252, CN2253, CN2254, CN2255, CN2256, CN2315, CN2423, CN2424, CN2690, CN2695, CN2691, CN2692, CN2944, CN2886, CN2755, CN2938, CN2239, CN2410, CN2411, CN2412, CN2963, CN2623, CN2624, CN2601, CN2843, CN2851, CN2853, CN2860, CN2828, CN2870, CN2871, CN2847, CN2686, CN2687, CN2878, CN2946, CN2943, CN3319, CN2918, CN3018, CN3030, CN3027, CN3032, CN3015, CN3012, and/or CN3024 within one or more of its cells. 1, 2, 4, 5, 6, 7, 8, 9, or 10 copy concatemers of disclosed enhancer cores can also be used.

[0120] 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.

[0121] 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.

[0122] 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.

[0123] 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 selected 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.

[0124] (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.

[0125] 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.

[0126] Particular embodiments include methods of administering to a subject an artificial expression construct that includes eHGT_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and/or 3xcore2_eHGT_460m and/or CN2248, CN2249, CN2250, CN2251, CN2252, CN2253,

CN2254, CN2255, CN2256, CN2315, CN2423, CN2424, CN2690, CN2695, CN2691, CN2692, CN2944, CN2886, CN2755, CN2938, CN2239, CN2410, CN2411, CN2412, CN2963, CN2623, CN2624, CN2601, CN2843, CN2851, CN2853, CN2860, CN2828, CN2870, CN2871, CN2847, CN2686, CN2687, CN2878, CN2946, CN2943, CN3319, CN2918, CN3018, CN3030, CN3027, CN3032, CN3015, CN3012, and/or CN3024 as described herein to drive selective expression of a gene in a selected cell type. 1, 2, 4, 5, 6, 7, 8, 9, or 10 copy concatemers of disclosed enhancer cores can also be used. The subject can be an isolated cell, a network of cells, a tissue slice, an experimental animal, a veterinary animal, or a human.

[0127] 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.

[0128] An "effective amount" is the amount of a composition necessary to result in a desired physiological change in the subject. Effective amounts are often administered for research purposes. Effective amounts disclosed herein can cause a statistically-significant effect in an animal model or in vitro assay.

[0129] 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.

[0130] 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.

[0131] (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.

[0132] 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.

[0133] 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.

[0134] (vii) Exemplary Embodiments.

1. A concatenated core of an enhancer disclosed herein.
2. A concatenated core of the eHGT_459m, eHGT_453m, eHGT_462m, eHGT_461m, eHGT_464m, eHGT_452m, eHGT_451m, or eHGT_460m enhancer.
3. The concatenated core of embodiment 1 or 2, wherein the concatenated core includes a sequence as set forth in SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, and/or SEQ ID NO: 52.
4. The concatenated core of embodiment 1 or 2, wherein the concatenated core includes a sequence as set forth in SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, or SEQ ID NO: 52.
5. The concatenated core of any of embodiments 1-4, wherein the concatenated core includes 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the enhancer core.

6. The concatenated core of embodiment 5, including 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, and/or SEQ ID NO: 52.
7. The concatenated core of embodiment 5, including 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, or SEQ ID NO: 52.
8. The concatenated core of embodiment 5, including
   2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 45;
   2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 46;
   2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 47;
   2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 48;
   2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 49;
   2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 50;
   2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 51; or
   2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 52.
9. The concatenated core of embodiment 5, including 3 copies of SEQ ID NO: 45; 3 copies of SEQ ID NO: 46; 3 copies of SEQ ID NO: 47; 3 copies of SEQ ID NO: 48; 3 copies of SEQ ID NO: 49; 3 copies of SEQ ID NO: 50; 3 copies of SEQ ID NO: 51; or 3 copies of SEQ ID NO: 52.
10. The concatenated core of embodiment 5, wherein the concatenated core includes SEQ ID NO: 53; SEQ ID NO: 54; SEQ ID NO: 55; SEQ ID NO: 56; SEQ ID NO: 57; SEQ ID NO: 58; SEQ ID NO: 59; or SEQ ID NO: 60.
11. An isolated enhancer selected from eHGT_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, and eHGT_583m.
12. An artificial expression construct including (i) an enhancer or concatenated core thereof of any of embodiments 1-11; (ii) a promoter; and (iii) a heterologous encoding sequence.
13. The artificial expression construct of embodiment 12, wherein the heterologous encoding sequence encodes an effector element or an expressible element.

14. The artificial expression construct of embodiment 13, wherein the effector element includes a reporter protein or a functional molecule.
15. The artificial expression construct of embodiment 14, wherein the reporter protein includes a fluorescent protein.
16. The artificial expression construct of embodiments 13 or 14, 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, homologous recombination donor cassette, a designer receptor exclusively activated by designer drug (DREADD), a growth factor, or pro-survival gene.
17. The artificial expression construct of any of embodiments 13-16, wherein the expressible element includes a non-functional molecule.
18. The artificial expression construct of embodiment 17, 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, homologous recombination donor cassette, a DREADD, a growth factor, or pro-survival gene.
19. The artificial expression construct of any of embodiments 12-18, wherein the artificial expression construct is associated with a capsid that crosses the blood brain barrier.
20. The artificial expression construct of embodiment 19, wherein the capsid includes PHP.eB, AAV - BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.
21. The artificial expression construct of any of embodiments 12-20, wherein the artificial expression construct includes or encodes a skipping element.
22. The artificial expression construct of embodiment 21, wherein the skipping element includes a 2A peptide and/or an internal ribosome entry site (IRES).
23. The artificial expression construct of embodiment 22, wherein the 2A peptide includes T2A, P2A, E2A, or F2A.
24. The artificial expression construct of any of embodiments 12-23, wherein the artificial expression construct includes or encodes a set of features selected from: eHGT_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m,

eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, 3xcore2_eHGT_460m, AAV, scAAV, rAAv, minBglobin, CMV, minCMV, minRho, minRho*, fluorescent protein (e.g., EGFP, SYFP, SYFP2, GFP), Cre, iCre, dgCre, FlpO, tTA2, SP10, WPRE, WPRE3, and/or BGHpA.

25. The artificial expression construct of any of embodiments 12-24, wherein the artificial expression construct includes or encodes a set of features selected from:
eHGT_450m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_451m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_452m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_453m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_457m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_458m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_459m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_460m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_462m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_464m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_461m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_463m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_692h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_693h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_694h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_695h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_667m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_668m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_696m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_455m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_457h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_603m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_604m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_605m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_661m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;

eHGT_766m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_767m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_768m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_663m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_660m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_665m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_700m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_456m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_698m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_699m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_647m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_648m-minBglobin-Heterologous coding sequence WPRE3-BGHpA;
eHGT_649m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_670m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_697m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_662m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_583m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_459m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_453m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_462m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_461m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_464m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_452m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_451m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_460m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_450m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_451m-minimal promoter-Heterologous coding sequence- WPRE3-BGHpA;
eHGT_452m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_453m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_457m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_458m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_459m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_460m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_462m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;

eHGT_464m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_461m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_463m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_692h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_693h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_694h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_695h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_667m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_668m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_696m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_455m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_457h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_603m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_604m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_605m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_661m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_766m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_767m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_768m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_663m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_660m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_665m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_700m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_456m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_698m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_699m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_647m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_648m- minimal promoter -Heterologous coding sequence WPRE3-BGHpA;
eHGT_649m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
eHGT_670m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
eHGT_697m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
eHGT_662m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
eHGT_583m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_459m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;

3xcore2_eHGT_453m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_462m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_461m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_464m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_452m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_451m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
or
3xcore2_eHGT_460m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA.

26. A vector including an artificial expression construct of any of embodiments 12-25.
27. The vector of embodiment 26, wherein the vector includes a viral vector.
28. The vector of embodiments 26 or 27, wherein the viral vector includes a recombinant adeno-associated viral (AAV) vector.
29. An adeno-associated viral (AAV) vector including at least one heterologous encoding sequence, wherein the heterologous encoding sequence is under control of a promoter and an enhancer selected from eHGT_450m, eHGT_451m, eHGT_452m, eHGT_453m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and/or 3xcore2_eHGT_460m.
30. A transgenic cell including an expression construct or vector of any of the preceding embodiments.
31. The transgenic cell of embodiment 30, wherein the transgenic cell is an L5 glutamatergic neuron.
32. The transgenic cell of embodiment 31, wherein the L5 glutamatergic neuron is an L5 ET neuron.
33. A non-human transgenic animal including an artificial expression construct, vector, or transgenic cell of any of the preceding embodiments.
34. The non-human transgenic animal of embodiment 33, wherein the non-human transgenic animal is a mouse or a non-human primate.

35. An administrable composition including an expression construct, vector, or transgenic cell of any of the preceding embodiments.
36. A kit including an artificial expression construct, vector, transgenic cell, transgenic animal, and/or administrable compositions of any of the preceding embodiments.
37. A method for selectively expressing a heterologous gene within a population of cells in vivo or in vitro, the method including providing the administrable composition of embodiment 35 in a sufficient dosage and for a sufficient time to a sample or subject including the population of cells thereby selectively expressing the gene within the population of cells.
38. The method of embodiment 37, wherein the heterologous gene encodes an effector element or an expressible element.
39. The method of embodiment 38, wherein the effector element includes a reporter protein or a functional molecule.
40. The method of embodiment 39, wherein the reporter protein includes a fluorescent protein.
41. The method of embodiments 39 or 40, 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, homologous recombination donor cassette, a DREADD, a growth factor, or a pro-survival gene.
42. The method of any of embodiments 37-41, wherein the expressible element includes a non-functional molecule.
43. The method of embodiment 42 wherein the non-functional molecule includes a non-functional ion transporter, enzyme, transcription factor, receptor, membrane protein, cellular trafficking protein, signaling molecule, neurotransmitter, calcium reporter, channelrhodopsin, CRISPR/CAS molecule, editase, guide RNA molecule, homologous recombination donor cassette, DREADD, growth factor, or pro-survival gene.
44. The method of any of embodiments 37 - 43, wherein the providing includes pipetting.
45. The method of embodiment 44, wherein the pipetting is to a brain slice.
46. The method of embodiment 45, wherein the brain slice includes an L5 ET neuron.
47. The method of embodiment 45 or 46, wherein the brain slice is murine, human, or non-human primate.
48. The method of any of embodiments 37-43, wherein the providing includes administering to a living subject.
49. The method of embodiment 48, wherein the living subject is a human, non-human primate, or a mouse.

50. The method of embodiments 48 or 49, wherein the administering to a living subject is through injection.
51. The method of embodiment 50, wherein the injection includes intravenous injection, intraparenchymal injection into brain tissue, intracerebroventricular (ICV) injection, intra-cisterna magna (ICM) injection, or intrathecal injection.
52. The method of any of embodiments 48-51, wherein the living subject has a neurodegenerative disease.
53. The method of any of embodiments 48-52, wherein the living subject has ALS or multiple sclerosis.
54. An artificial expression construct including CN2248, CN2249, CN2250, CN2251, CN2252, CN2253, CN2254, CN2255, CN2256, CN2315, CN2423, CN2424, CN2690, CN2695, CN2691, CN2692, CN2944, CN2886, CN2755, CN2938, CN2239, CN2410, CN2411, CN2412, CN2963, CN2623, CN2624, CN2601, CN2843, CN2851, CN2853, CN2860, CN2828, CN2870, CN2871, CN2847, CN2686, CN2687, CN2878, CN2946, CN2943, CN3319, CN2918, CN3018, CN3030, CN3027, CN3032, CN3015, CN3012, and/or CN3024.

[0135] (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.

[0136] 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.

[0137] 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).

[0138] 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.

[0139] As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: Arg (+3.0); Lys (+3.0); aspartate (+3.0±1); glutamate (+3.0±1); Ser (+0.3); Asn (+0.2); Gln (+0.2); Gly (0); Thr (−0.4); Pro (−0.5±1); Ala (−0.5); His (−0.5); Cys (−1.0); Met (−1.3); Val (−1.5); Leu (−1.8); Ile (−1.8); Tyr (−2.3); Phe (−2.5); Trp (−3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

[0140] 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.

[0141] 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.

[0142] Variants of the protein, nucleic acid, and gene sequences disclosed herein also include sequences with at least 70% sequence identity, 80% sequence identity, 85% sequence, 90% sequence identity, 95% sequence identity, 96% sequence identity, 97% sequence identity, 98% sequence identity, or 99% sequence identity to the protein, nucleic acid, or gene sequences disclosed herein.

[0143] "% 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.

[0144] Variants also include nucleic acid molecules that hybridizes under stringent hybridization

conditions to a sequence disclosed herein and provide the same function as the reference sequence. Exemplary stringent hybridization conditions include an overnight incubation at 42 °C in a solution including 50% formamide, 5XSSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5XDenhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1XSSC at 50 °C. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, moderately high stringency conditions include an overnight incubation at 37°C in a solution including 6XSSPE (20XSSPE=3M NaCl; 0.2M NaH_2PO_4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50 °C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5XSSC). Variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

[0145] 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 terms “concatamerize” and “concatenate” can be used interchangeably.

[0146] 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 selective expression in L5 glutamatergic neurons, including L5 ET neurons.

[0147] In particular embodiments, artificial means not naturally occurring.

[0148] Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. When further clarity is required, the term "about" has the meaning reasonably ascribed to it by a person skilled in the art when used in conjunction with a stated numerical value or range, i.e. denoting somewhat more or somewhat less than the stated value or range, to within a range of ±20% of the stated value; ±19% of the stated value; ±18% of the stated value; ±17% of the stated value; ±16% of the stated value; ±15% of the stated value; ±14% of the stated value; ±13% of the stated value; ±12% of the stated value; ±11% of the stated value; ±10% of the stated value; ±9% of the stated value; ±8% of the stated value; ±7% of the stated value; ±6% of the stated value; ±5% of the stated value; ±4% of the stated value; ±3% of the stated value; ±2% of the stated value; or ±1% of the stated value.

[0149] 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.

[0150] 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.

[0151] 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.

[0152] 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.

[0153] 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.

[0154] 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.

[0155] 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.

[0156] Definitions and explanations used in the present disclosure are meant and intended to be

controlling in any future construction unless clearly and unambiguously modified in the following examples or when application of the meaning renders any construction meaningless or essentially meaningless. In cases where the construction of the term would render it meaningless or essentially meaningless, the definition should be taken from Webster's Dictionary, 3rd Edition or a dictionary known to those of ordinary skill in the art, such as the Oxford Dictionary of Biochemistry and Molecular Biology (Ed. Anthony Smith, Oxford University Press, Oxford, 2004).

CLAIMS

What is claimed is:

1. An artificial expression construct comprising (i) an enhancer selected from eHGT_453m, eHGT_450m, eHGT_451m, eHGT_452m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, or 3xcore2_eHGT_460m; (ii) a promoter; and (iii) a heterologous encoding sequence.
2. The artificial expression construct of claim 1, wherein the heterologous encoding sequence encodes an effector element or an expressible element.
3. The artificial expression construct of claim 2, wherein the effector element comprises a reporter protein or a functional molecule.
4. The artificial expression construct of claim 3 wherein the reporter protein comprises a fluorescent protein.
5. The artificial expression construct of claim 3, 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, homologous recombination donor cassette, designer receptor exclusively activated by designer drug (DREADD), growth factor, or pro-survival gene.
6. The artificial expression construct of claim 2, wherein the expressible element comprises a non-functional molecule.
7. The artificial expression construct of claim 6, 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, homologous recombination donor cassette, DREADD, growth factor, or pro-survival gene.
8. The artificial expression construct of claim 1, wherein the artificial expression construct is

associated with a capsid that crosses the blood brain barrier.

9. The artificial expression construct of claim 8, wherein the capsid comprises PHP.eB, AAV - BR1, AAV-PHP.S, AAV-PHP.B, or AAV-PPS.

10. The artificial expression construct of claim 1, wherein the artificial expression construct comprises or encodes a skipping element.

11. The artificial expression construct of claim 10, wherein the skipping element comprises a 2A peptide and/or an internal ribosome entry site (IRES).

12. The artificial expression construct of claim 11, wherein the 2A peptide comprises T2A, P2A, E2A, or F2A.

13. The artificial expression construct of claim 1, wherein the artificial expression construct comprises or encodes a set of features selected from: eHGT_453m, eHGT_450m, eHGT_451m, eHGT_452m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m, eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, 3xcore2_eHGT_460m, AAV, scAAV, rAAv, minBglobin, CMV, minCMV, minRho, minRho*, fluorescent protein (e.g., EGFP, SYFP, SYFP2, GFP), Cre, iCre, dgCre, FlpO, tTA2, SP10, WPRE, WPRE3, and/or BGHpA.

14. The artificial expression construct of claim 1, wherein the artificial expression construct comprises or encodes a set of features selected from:

eHGT_453m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_450m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_451m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_452m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_457m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_458m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_459m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_460m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_462m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_464m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;

eHGT_461m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_463m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_692h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_693h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_694h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_695h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_667m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_668m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_696m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_455m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_457h-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_603m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_604m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_605m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_661m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_766m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_767m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_768m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_663m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_660m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_665m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_700m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_456m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_698m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_699m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_647m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_648m-minBglobin-Heterologous coding sequence WPRE3-BGHpA;
eHGT_649m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_670m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_697m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_662m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_583m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_459m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_453m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;

3xcore2_eHGT_462m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_461m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_464m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_452m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_451m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_460m-minBglobin-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_453m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_450m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_451m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_452m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_457m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_458m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_459m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_460m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_462m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_464m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_461m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_463m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_692h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_693h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_694h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_695h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_667m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_668m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_696m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_455m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_457h-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_603m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_604m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_605m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_661m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_766m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_767m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_768m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;

eHGT_663m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_660m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_665m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_700m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_456m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_698m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_699m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_647m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_648m- minimal promoter -Heterologous coding sequence WPRE3-BGHpA;
eHGT_649m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
eHGT_670m- minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
eHGT_697m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
eHGT_662m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
eHGT_583m- minimal promoter -Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_459m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_453m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_462m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_461m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_464m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_452m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA;
3xcore2_eHGT_451m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA; or
3xcore2_eHGT_460m-minimal promoter-Heterologous coding sequence-WPRE3-BGHpA.

15. A vector comprising an artificial expression construct of claim 1.

16. The vector of claim 15, wherein the vector comprises a viral vector.

17. The vector of claim 15, wherein the viral vector comprises a recombinant adeno-associated viral (AAV) vector.

18. An adeno-associated viral (AAV) vector comprising at least one heterologous encoding sequence, wherein the heterologous encoding sequence is under control of a promoter and an enhancer selected from eHGT_453m, eHGT_450m, eHGT_451m, eHGT_452m, eHGT_457m, eHGT_458m, eHGT_459m, eHGT_460m, eHGT_462m, eHGT_464m, eHGT_461m, eHGT_463m, eHGT_692h, eHGT_693h, eHGT_694h, eHGT_695h, eHGT_667m, eHGT_668m, eHGT_696m, eHGT_455m, eHGT_457h, eHGT_603m, eHGT_604m, eHGT_605m, eHGT_661m, eHGT_766m, eHGT_767m, eHGT_768m, eHGT_663m, eHGT_660m, eHGT_665m, eHGT_700m, eHGT_456m, eHGT_698m, eHGT_699m, eHGT_647m,

eHGT_648m, eHGT_649m, eHGT_670m, eHGT_697m, eHGT_662m, eHGT_583m, 3xcore2_eHGT_459m, 3xcore2_eHGT_453m, 3xcore2_eHGT_462m, 3xcore2_eHGT_461m, 3xcore2_eHGT_464m, 3xcore2_eHGT_452m, 3xcore2_eHGT_451m, and/or 3xcore2_eHGT_460m.

19. A transgenic cell comprising an artificial expression construct of claim 1.

20. The transgenic cell of claim 19, wherein the transgenic cell is a L5 ET neuron.

21. A non-human transgenic animal comprising an artificial expression construct of claim 1.

22. The non-human transgenic animal of claim 21, wherein the non-human transgenic animal is a mouse or a non-human primate.

23. An administrable composition comprising an artificial expression construct of claim 1.

24. A kit comprising an artificial expression construct of claim 1.

25. A method for selectively expressing a heterologous gene within a population of cells in vivo or in vitro, the method comprising providing the administrable composition of claim 23 in a sufficient dosage and for a sufficient time to a sample or subject comprising the population of cells thereby selectively expressing the gene within the population of cells.

26. The method of claim 25, wherein the heterologous gene encodes an effector element or an expressible element.

27. The method of claim 26, wherein the effector element comprises a reporter protein or a functional molecule.

28. The method of claim 27, wherein the reporter protein comprises a fluorescent protein.

29. The method of claim 27, 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, homologous recombination donor cassette, DREADD growth factor, or pro-survival gene.

30. The method of claim 26, wherein the expressible element comprises a non-functional molecule.

31. The method of claim 30, 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, homologous recombination donor cassette, DREADD, growth factor, or pro-survival gene.

32. The method of claim 25, wherein the providing comprises pipetting.

33. The method of claim 32, wherein the pipetting is to a brain slice.

34. The method of claim 33, wherein the brain slice comprises an L5 ET neuron.

35. The method of claim 33, wherein the brain slice is murine, human, or non-human primate.

36. The method of claim 25, wherein the providing comprises administering to a living subject.

37. The method of claim 36, wherein the living subject is a human, non-human primate, or a mouse.

38. The method of claim 36, wherein the administering to a living subject is through injection.

39. The method of claim 38, wherein the injection comprises intravenous injection, intraparenchymal injection into brain tissue, intracerebroventricular (ICV) injection, intra-cisterna magna (ICM) injection, or intrathecal injection.

40. The method of claim 36, wherein the living subject has a neurodegenerative disease.

41. The method of claim 36, wherein the living subject has Amyotrophic Lateral Sclerosis (ALS) or multiple sclerosis.

42. An artificial expression construct comprising CN2251, CN2248, CN2249, CN2250, CN2252, CN2253, CN2254, CN2255, CN2256, CN2315, CN2423, CN2424, CN2690, CN2695, CN2691, CN2692, CN2944, CN2886, CN2755, CN2938, CN2239, CN2410, CN2411, CN2412, CN2963, CN2623, CN2624, CN2601, CN2843, CN2851, CN2853, CN2860, CN2828, CN2870, CN2871, CN2847, CN2686, CN2687, CN2878, CN2946, CN2943, CN3319, CN2918, CN3018, CN3030, CN3027, CN3032, CN3015, CN3012, or CN3024.

43. A concatenated core of the eHGT_459m, eHGT_453m, eHGT_462m, eHGT_461m, eHGT_464m, eHGT_452m, eHGT_451m, or eHGT_460m enhancer.

44. The concatenated core of claim 43, wherein the concatenated core has a sequence set forth in SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 47, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 50, SEQ ID NO: 51, or SEQ ID NO: 52.

45. The concatenated core of claim 43, wherein the concatenated core has 2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of the enhancer core.

46. The concatenated core of claim 43, having

2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 45,

2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 46, SEQ ID NO: 47,

2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 48,

2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 49,

2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 50,

2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 51, or

2, 3, 4, 5, 6, 7, 8, 9, or 10 copies of SEQ ID NO: 52.

55. The concatenated core of claim 43, having 3 copies of SEQ ID NO: 45; 3 copies of SEQ ID

NO: 46; 3 copies of SEQ ID NO: 47; 3 copies of SEQ ID NO: 48; 3 copies of SEQ ID NO: 49; 3 copies of SEQ ID NO: 50; 3 copies of SEQ ID NO: 51; or 3 copies of SEQ ID NO: 52.

56. The concatenated core of claim 43, wherein the concatenated core has the sequence as set forth in SEQ ID NO: 53; SEQ ID NO: 54; SEQ ID NO: 55; SEQ ID NO: 56; SEQ ID NO: 57; SEQ ID NO: 58; SEQ ID NO: 59; or SEQ ID NO: 60.

FIG. 1

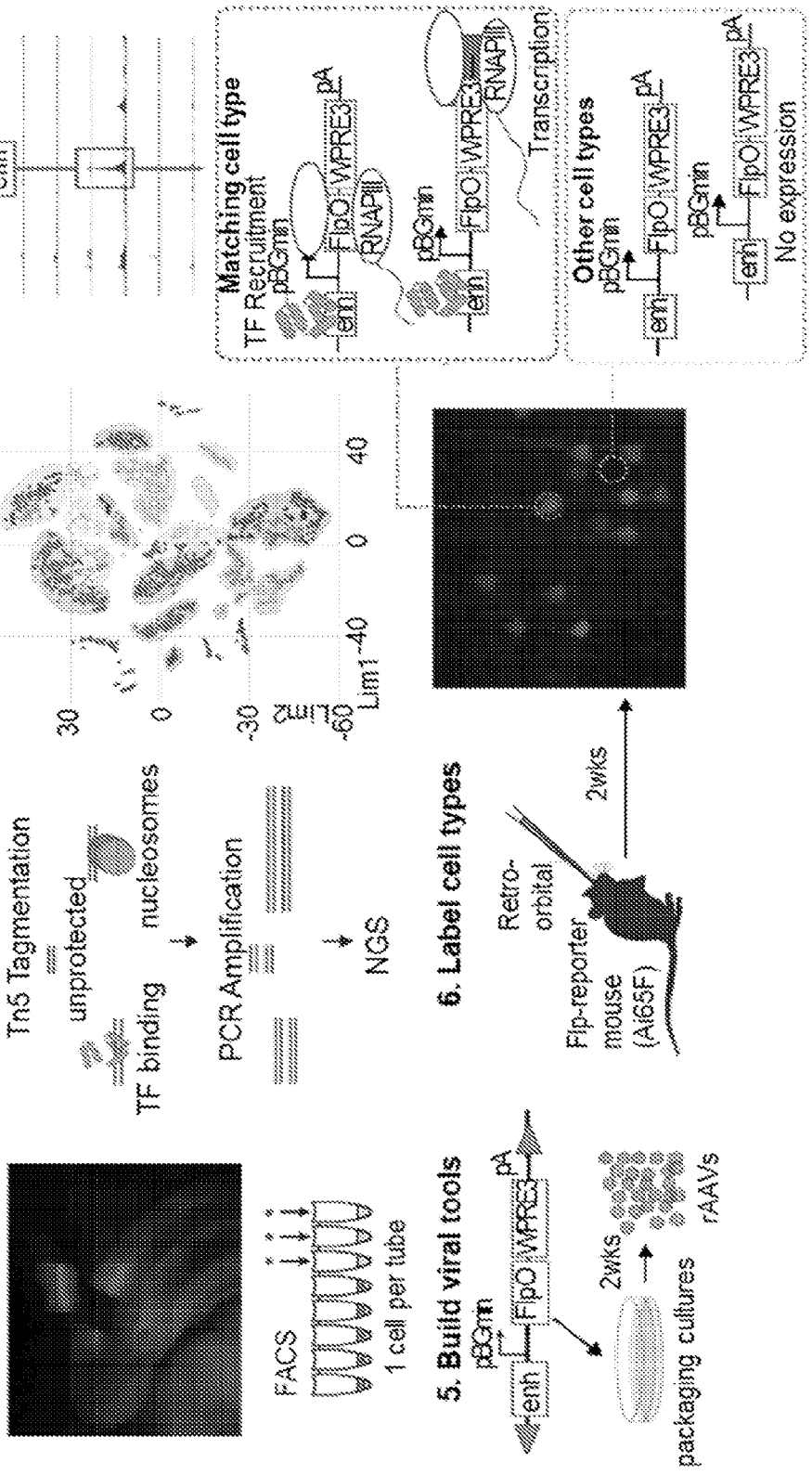

1. Isolate cells
FACS
1 cell per tube
2. Perform scATAC-seq
Tn5 Tagmentation
unprotected
nucleosomes
TF binding
PCR Amplification
NGS
3. Identify cell types
4. Locate enhancers
enh
Matching cell type
TF Recruitment
pBGmin
FlpO
WPRE3
pA
RNAPII
Transcription
Other cell types
No expression
5. Build viral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

FIG. 2A

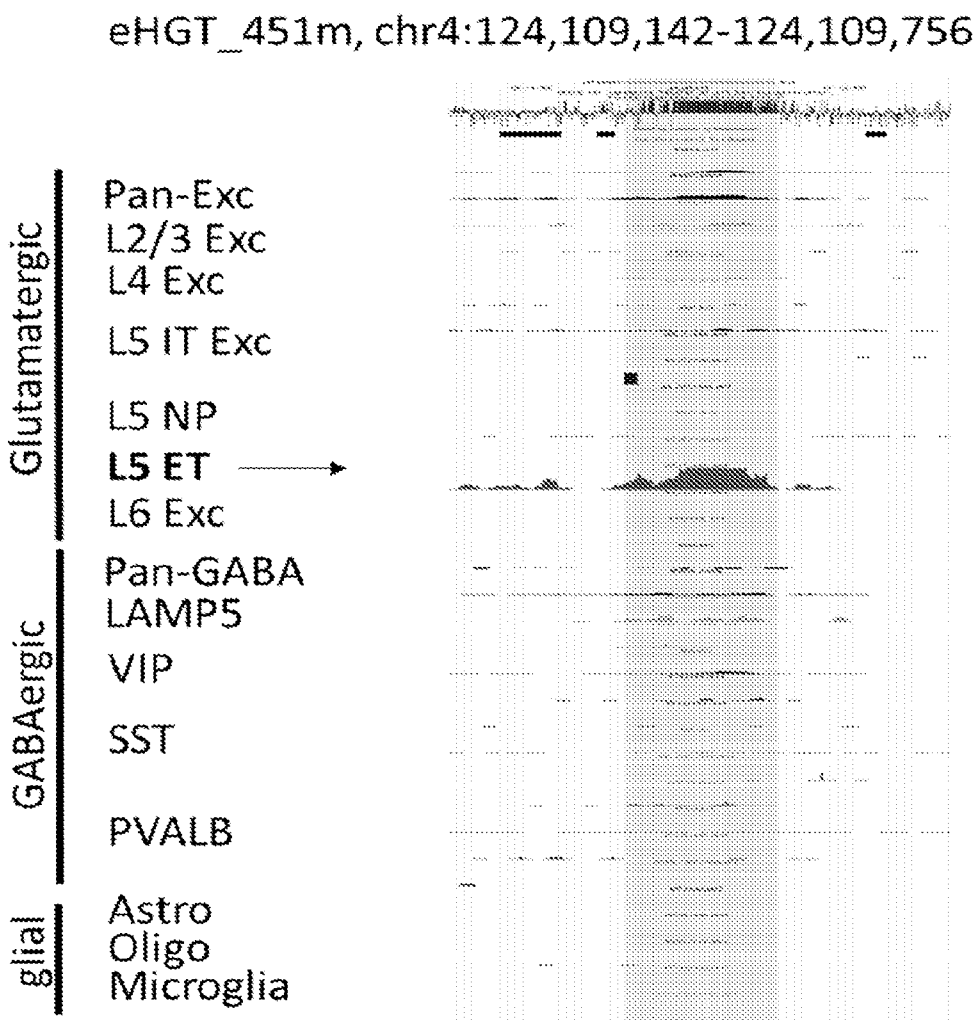

eHGT_451m, chr4:124,109,142-124,109,756
Glutamatergic
GABAergic
glial
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
Pan-GABA
LAMP5
VIP
SST
PVALB
Astro
Oligo
Microglia

FIG. 2B

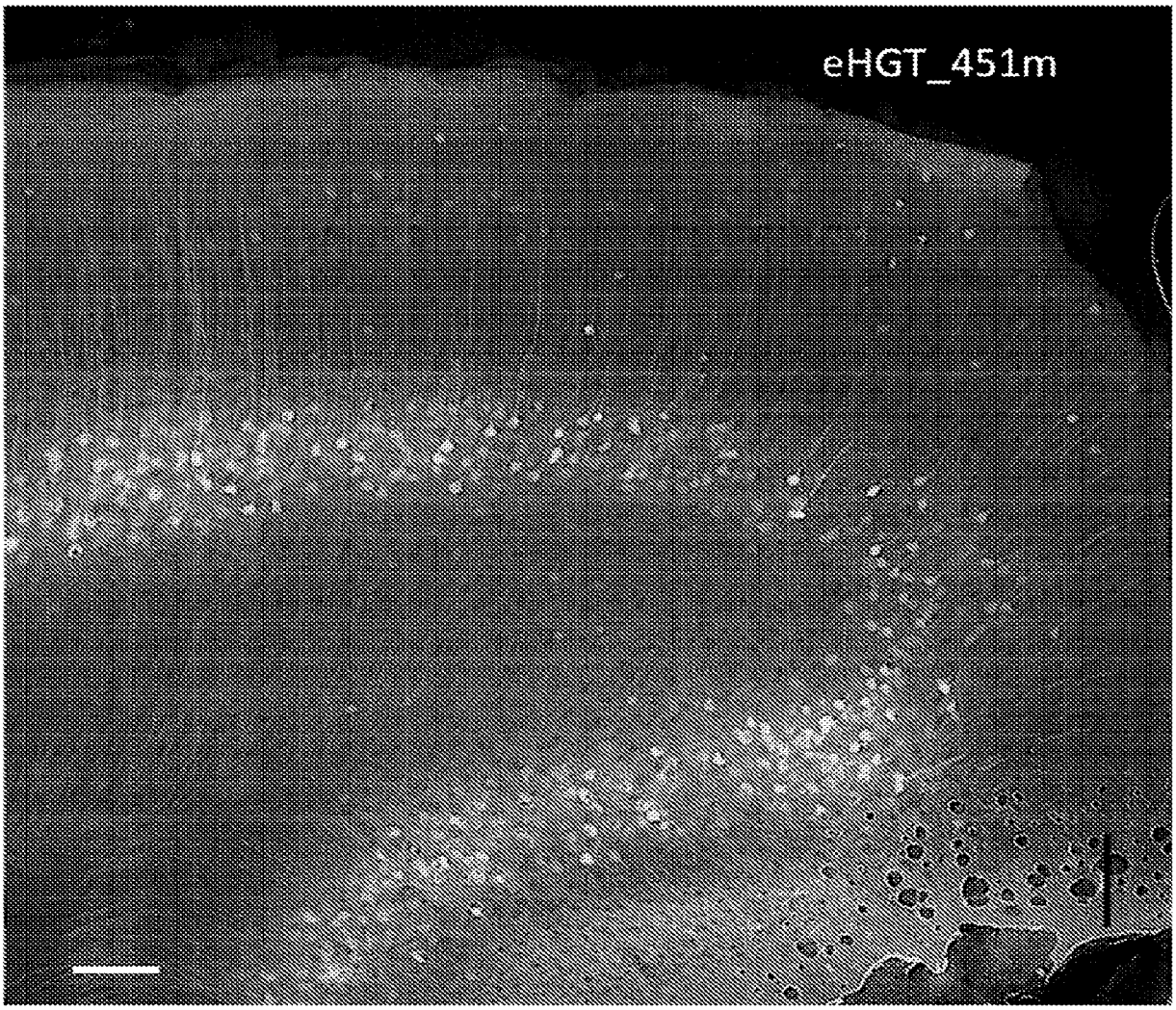


FIG. 3A

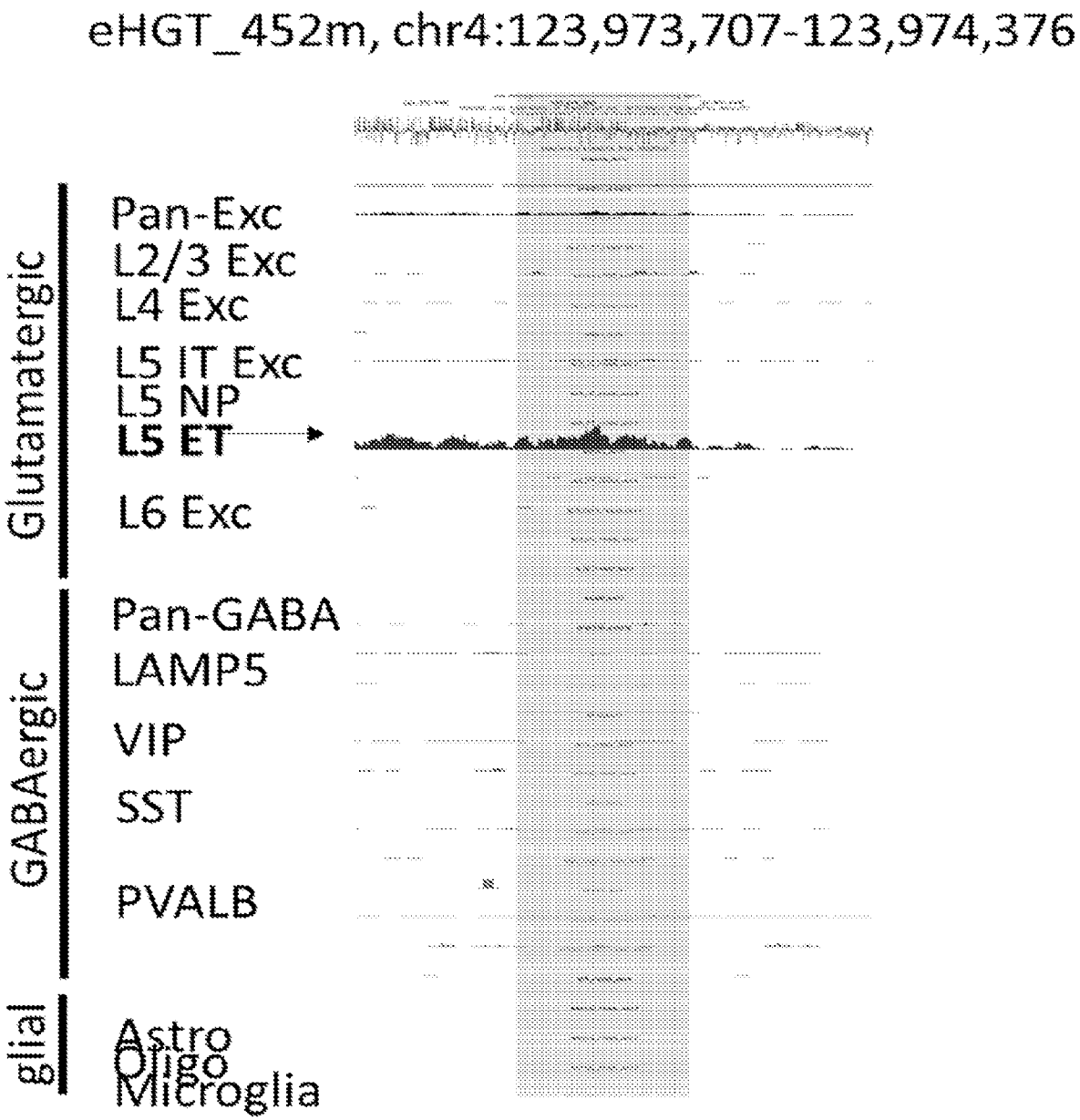

eHGT_452m, chr4:123,973,707-123,974,376
Glutamatergic
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
GABAergic
Pan-GABA
LAMP5
VIP
SST
PVALB
glial
Astro
Oligo
Microglia

FIG. 3B

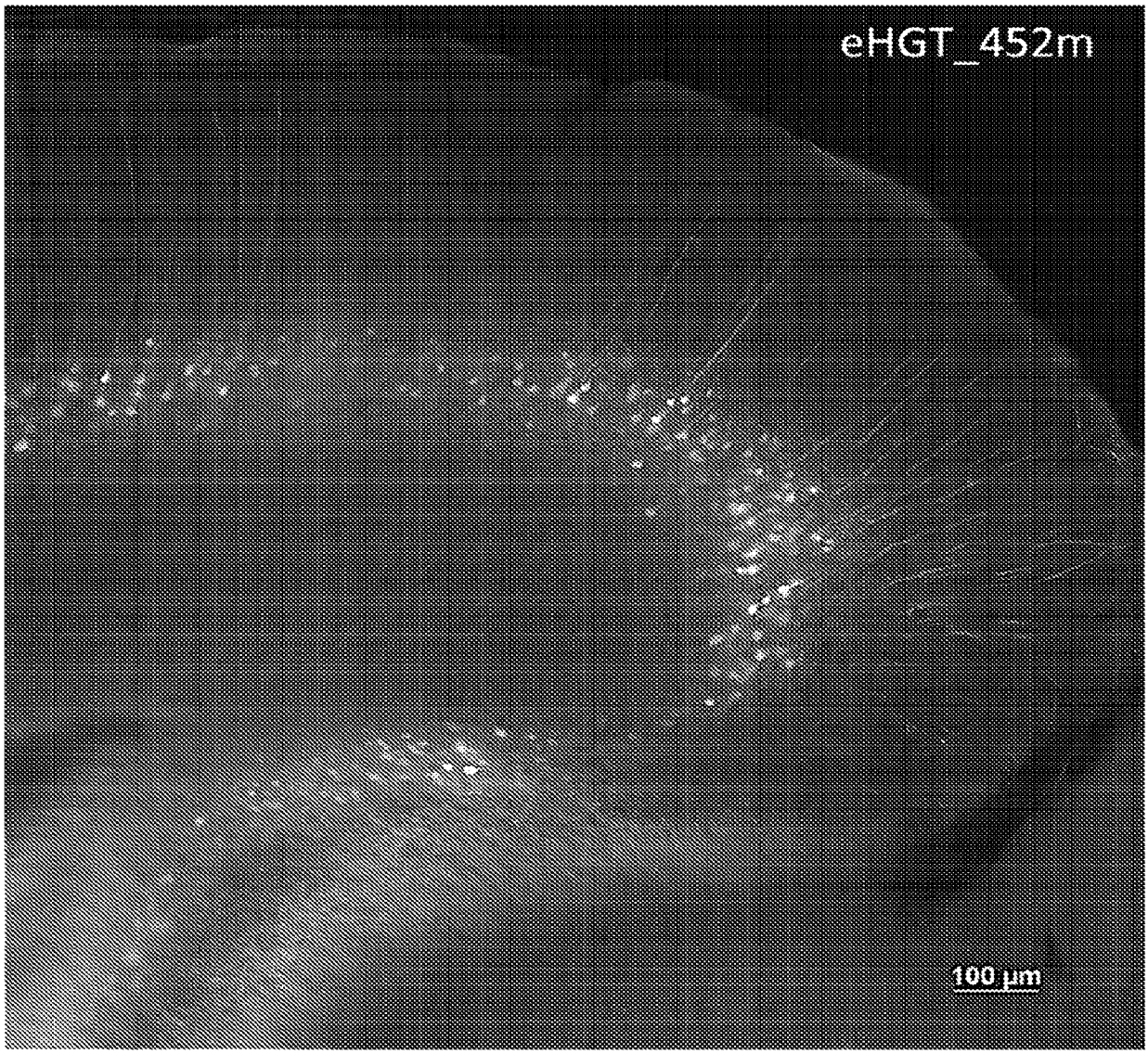


FIG. 4A

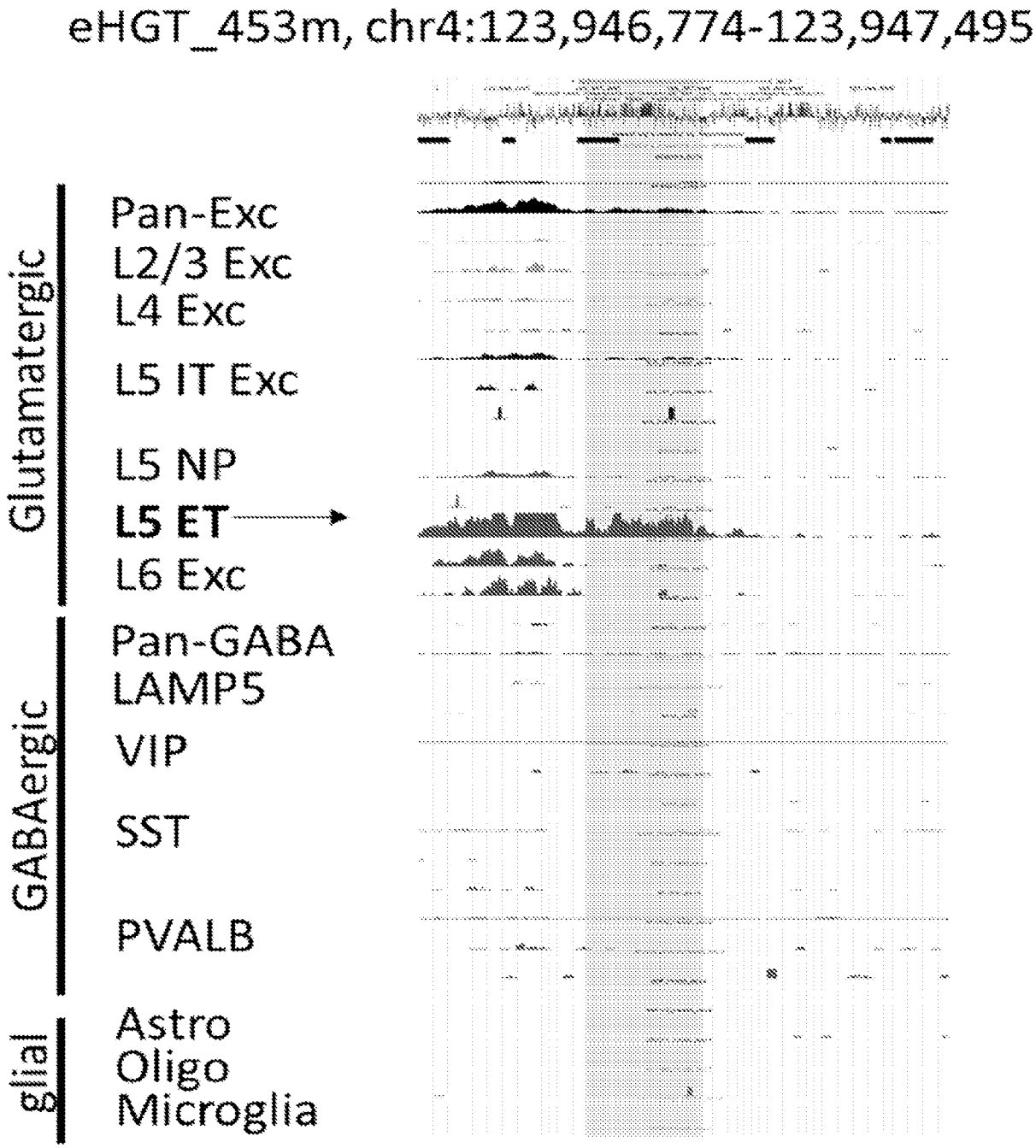
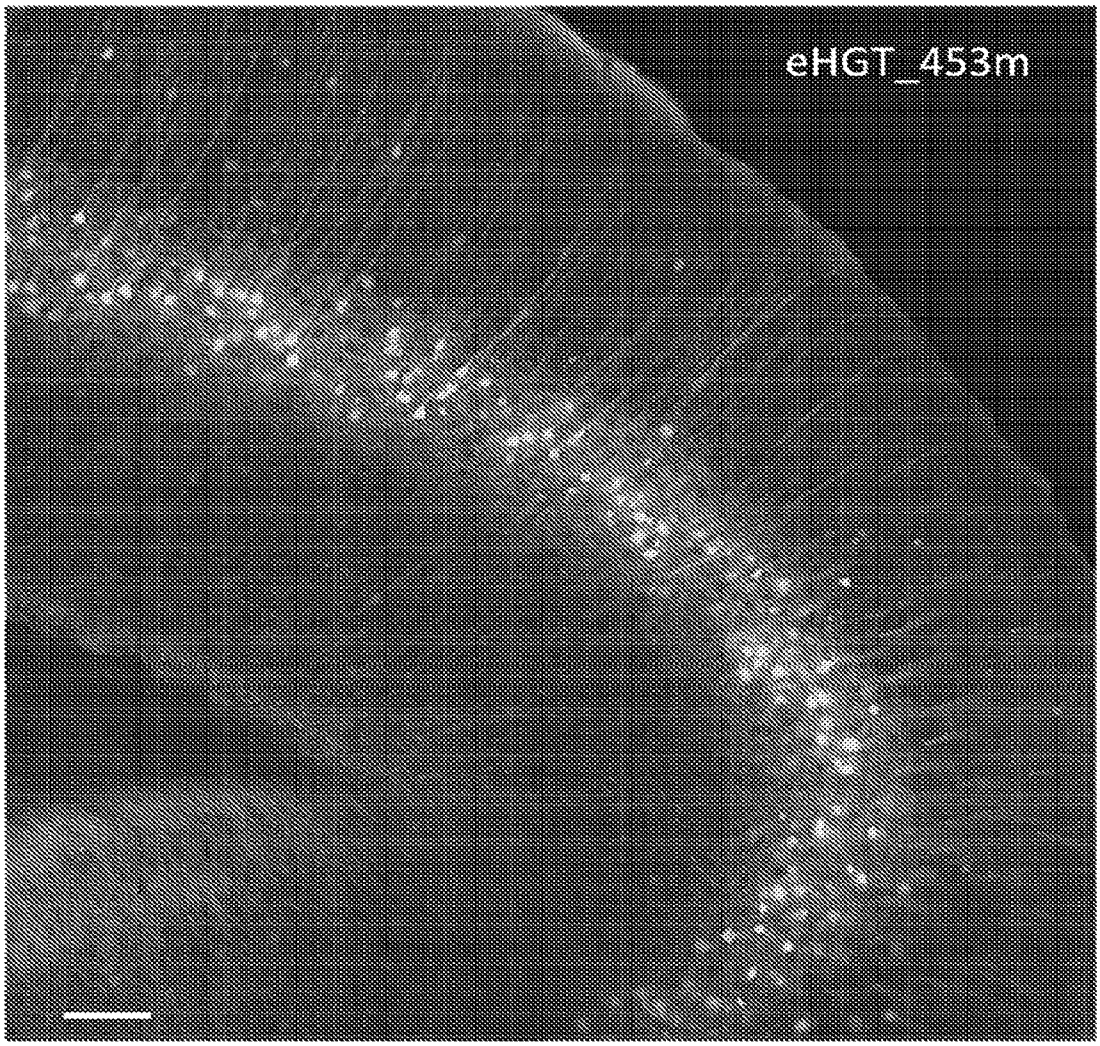

eHGT_453m, chr4:123,946,774-123,947,495
Glutamatergic
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
GABAergic
Pan-GABA
LAMP5
VIP
SST
PVALB
glial
Astro
Oligo
Microglia

FIG. 4B



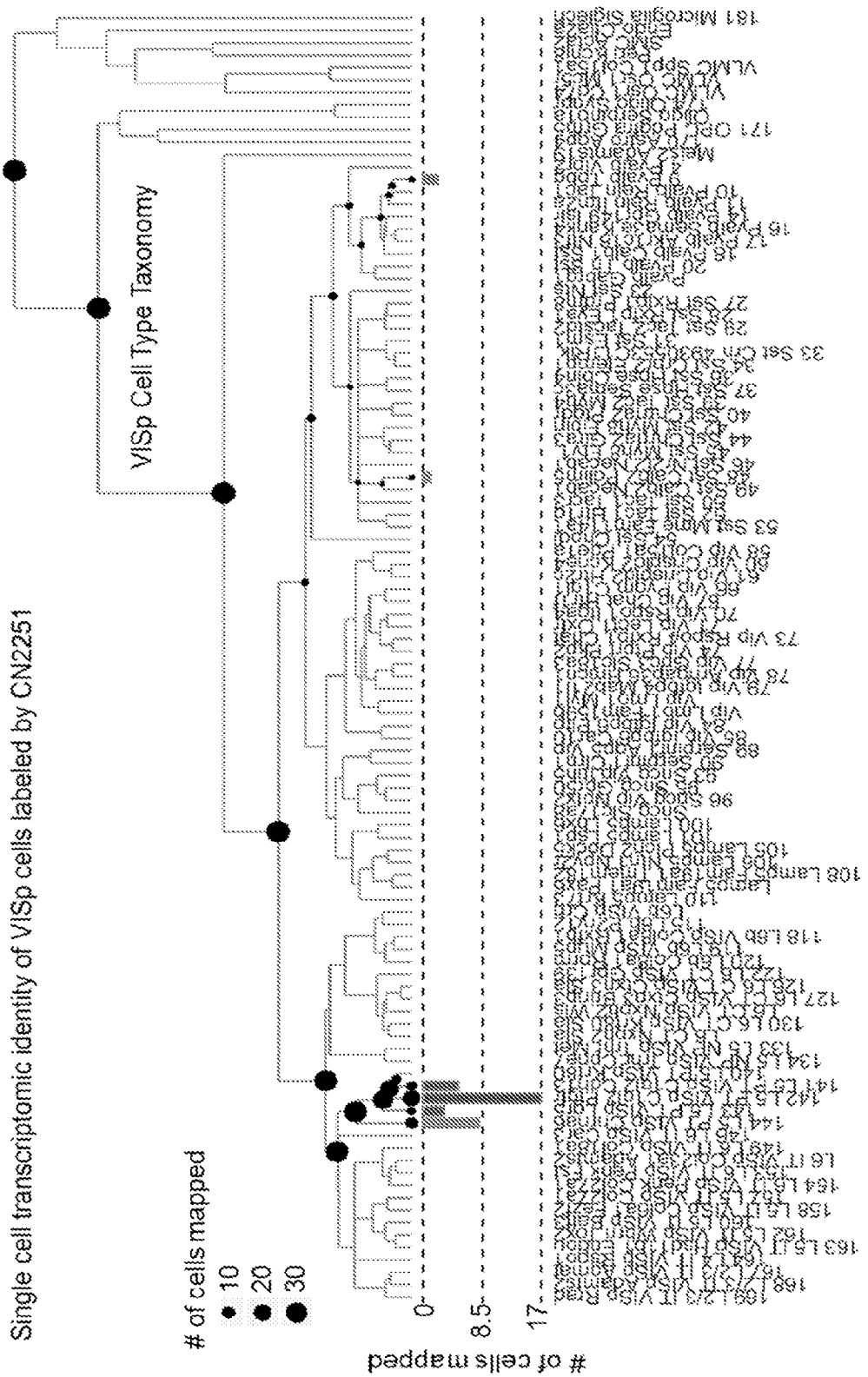

Single cell transcriptomic identity of VISp cells labeled by CN2251
VISp Cell Type Taxonomy
# of cells mapped
10
20
30
0
8.5
17
# of cells mapped

FIG. 4C

FIG. 4D

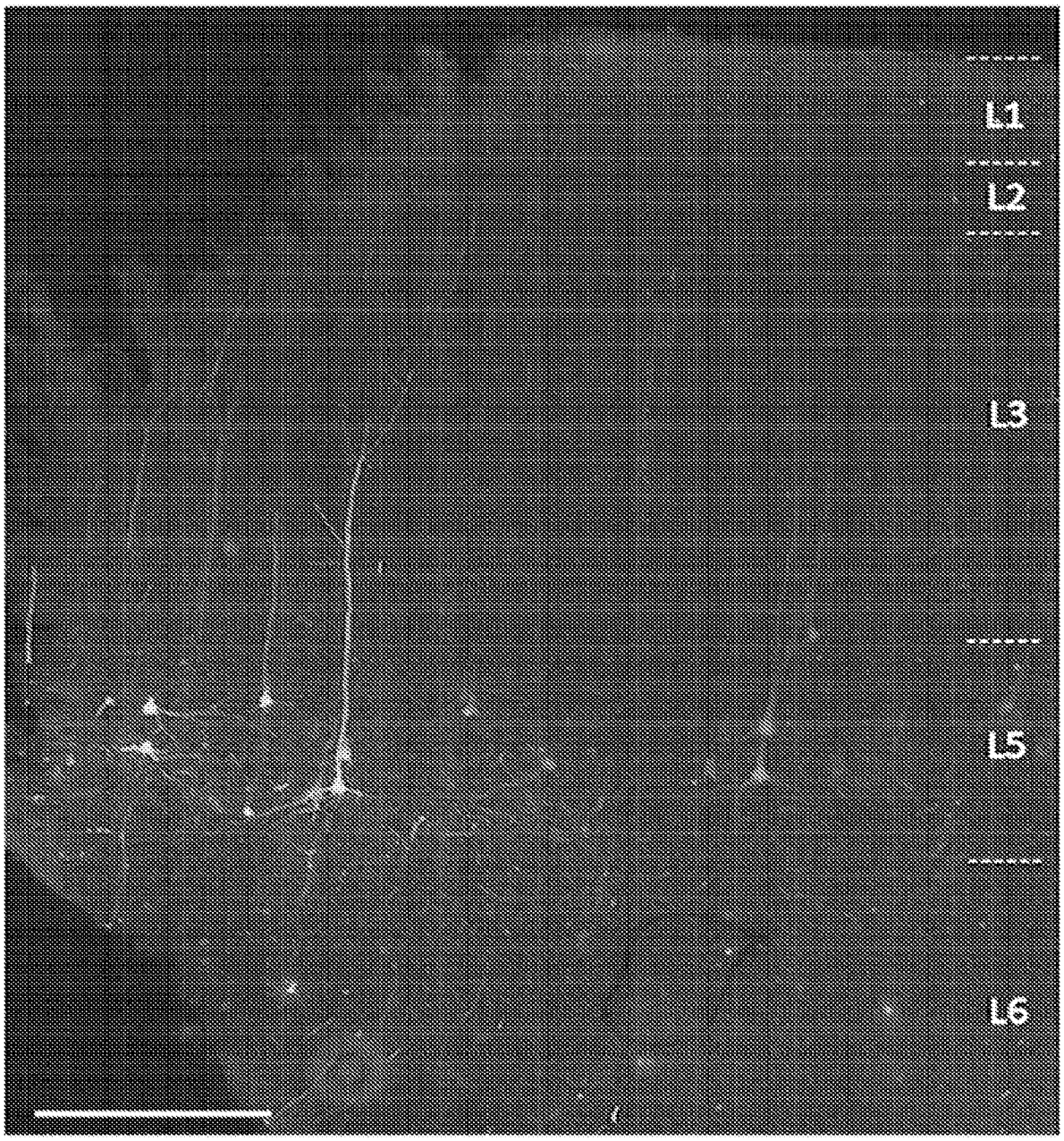


FIG. 5A

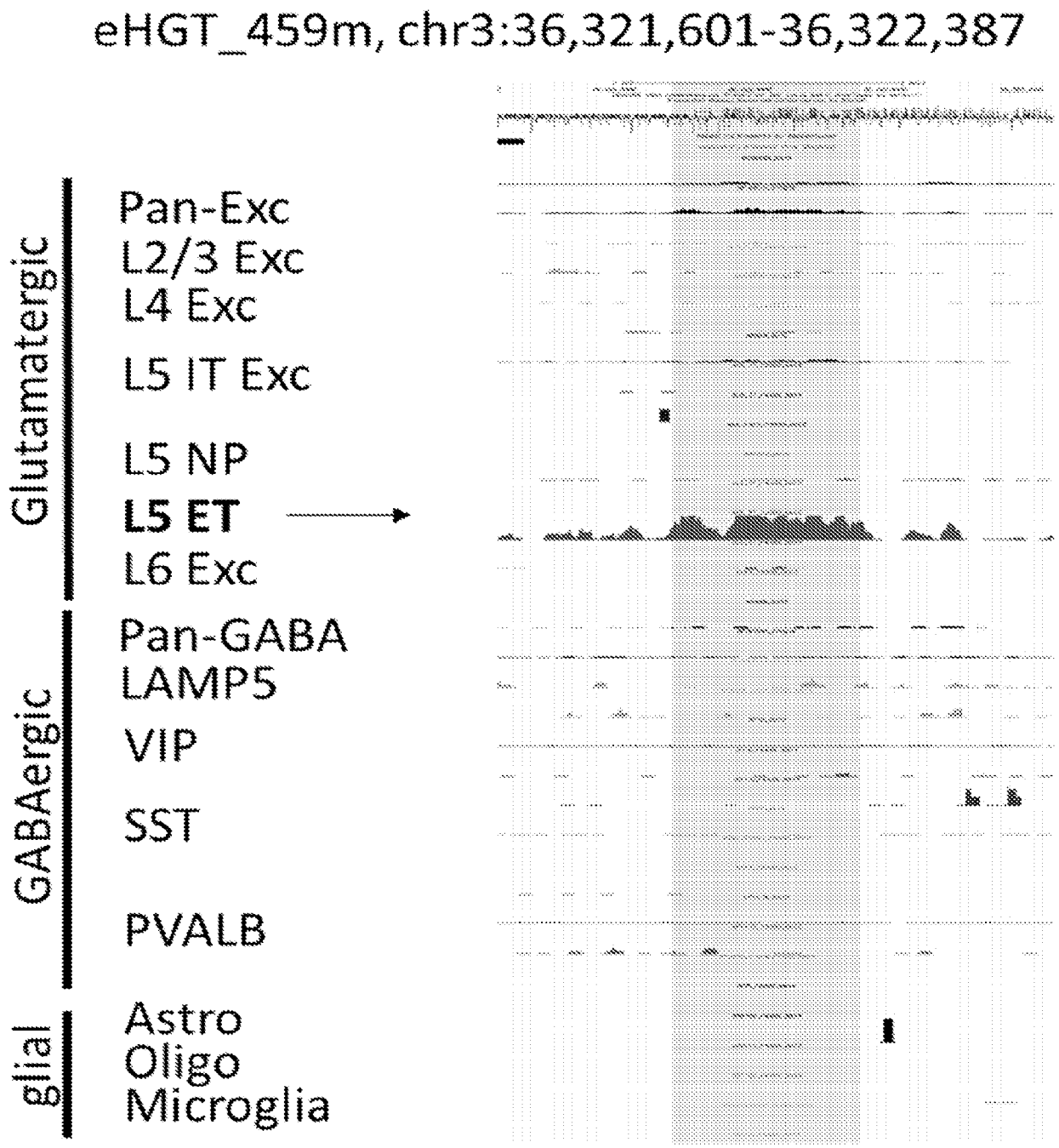
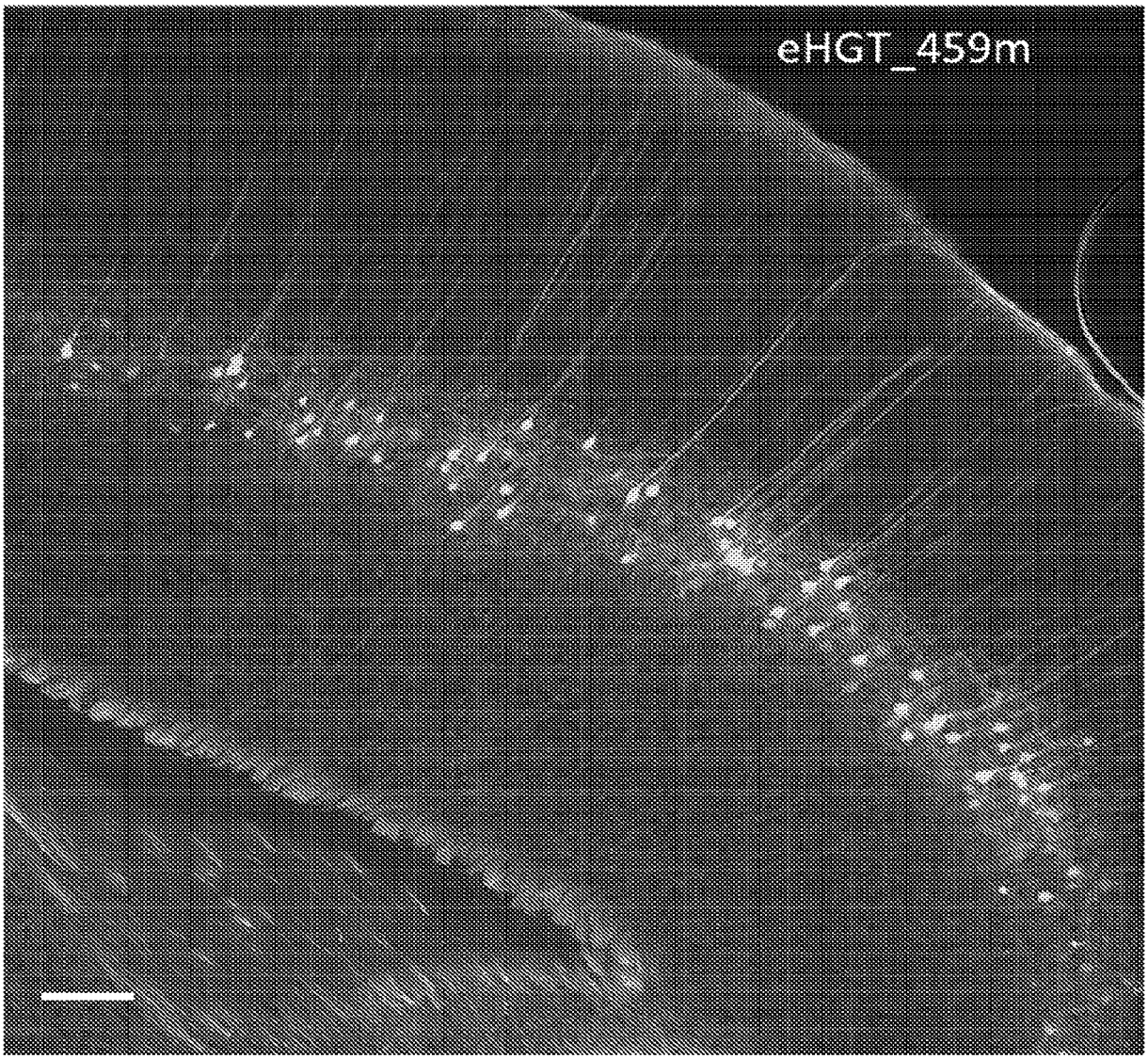

eHGT_459m, chr3:36,321,601-36,322,387
Glutamatergic
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
GABAergic
Pan-GABA
LAMP5
VIP
SST
PVALB
glial
Astro
Oligo
Microglia

FIG. 5B



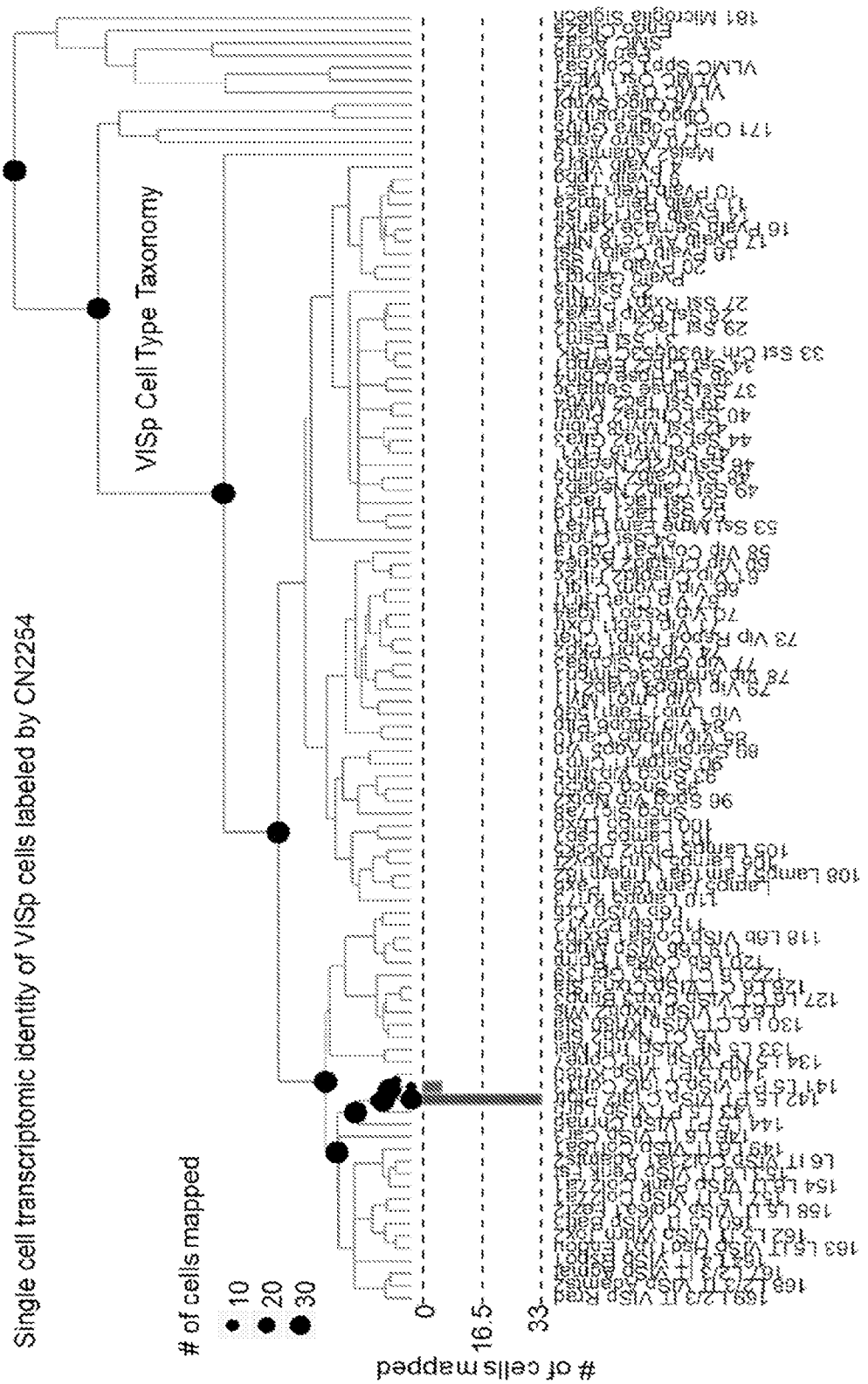

Single cell transcriptomic identity of VISp cells labeled by CN2254
VISp Cell Type Taxonomy
# of cells mapped
10
20
30
# of cells mapped
0
16.5
33

FIG. 5C

FIG. 5D

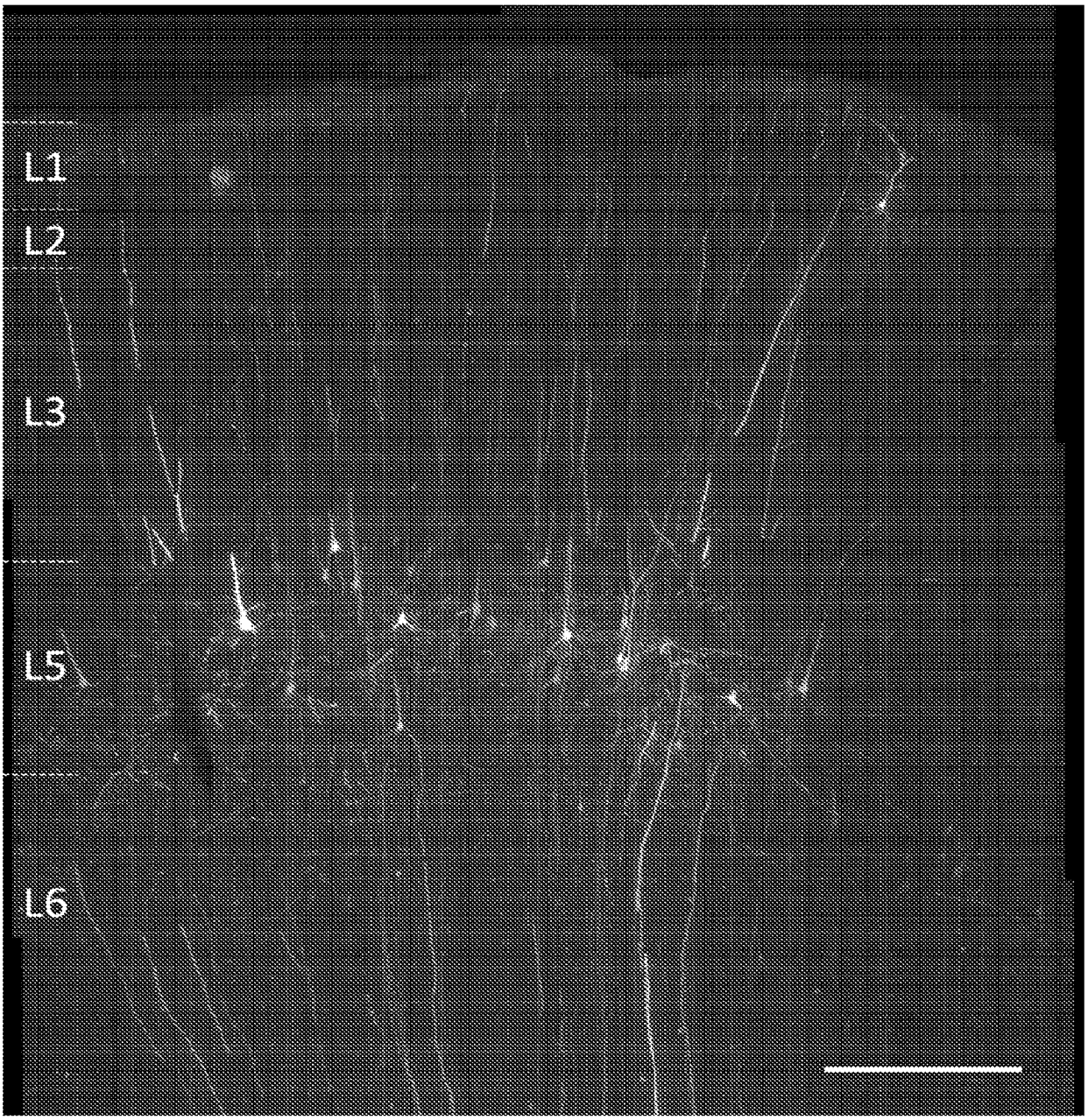


FIG. 6A

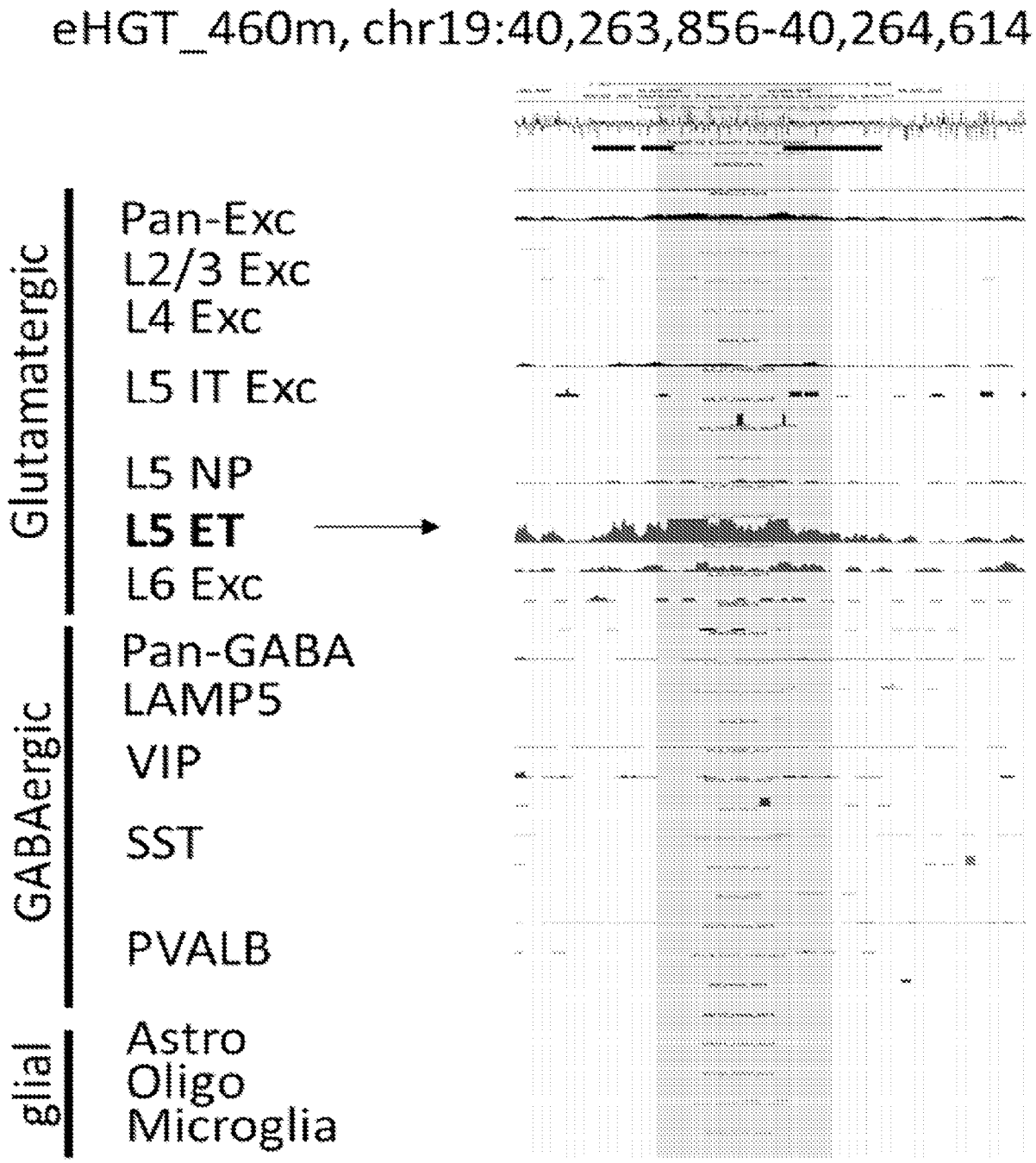

eHGT_460m, chr19:40,263,856-40,264,614
Glutamatergic
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
GABAergic
Pan-GABA
LAMP5
VIP
SST
PVALB
glial
Astro
Oligo
Microglia

FIG. 6B

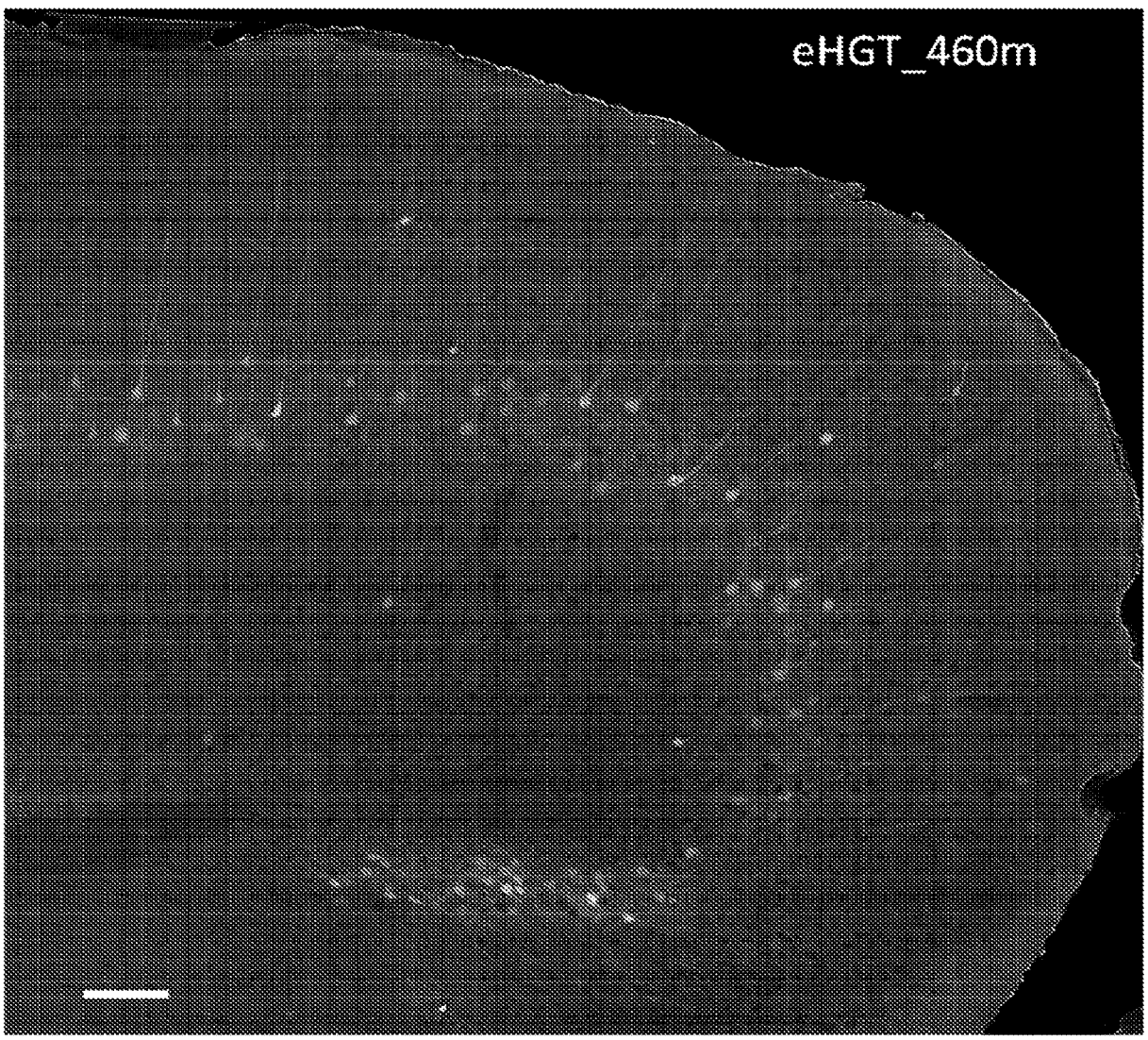


FIG. 7A

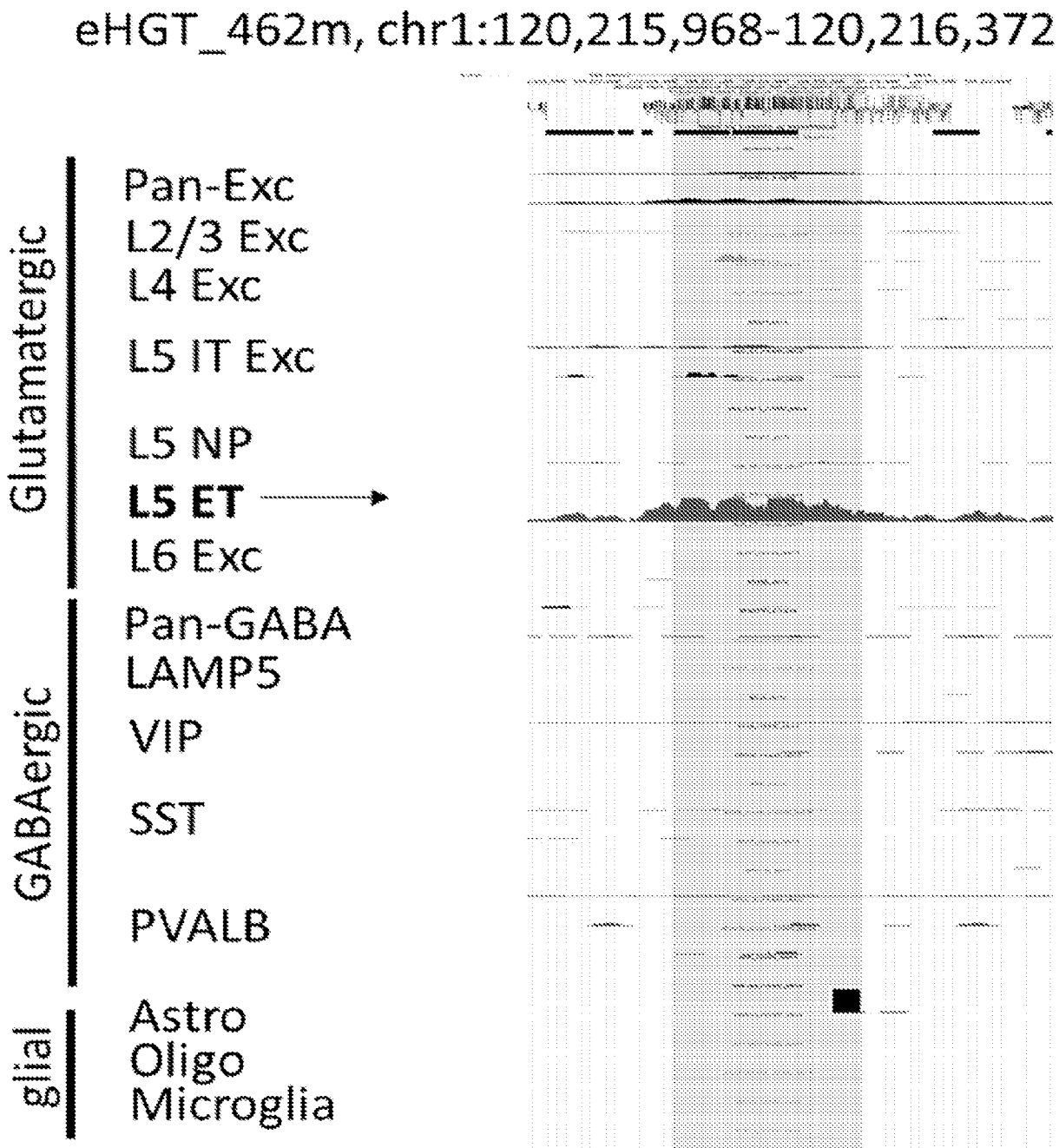

eHGT_462m, chr1:120,215,968-120,216,372
Glutamatergic
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
GABAergic
Pan-GABA
LAMP5
VIP
SST
PVALB
glial
Astro
Oligo
Microglia

FIG. 7B

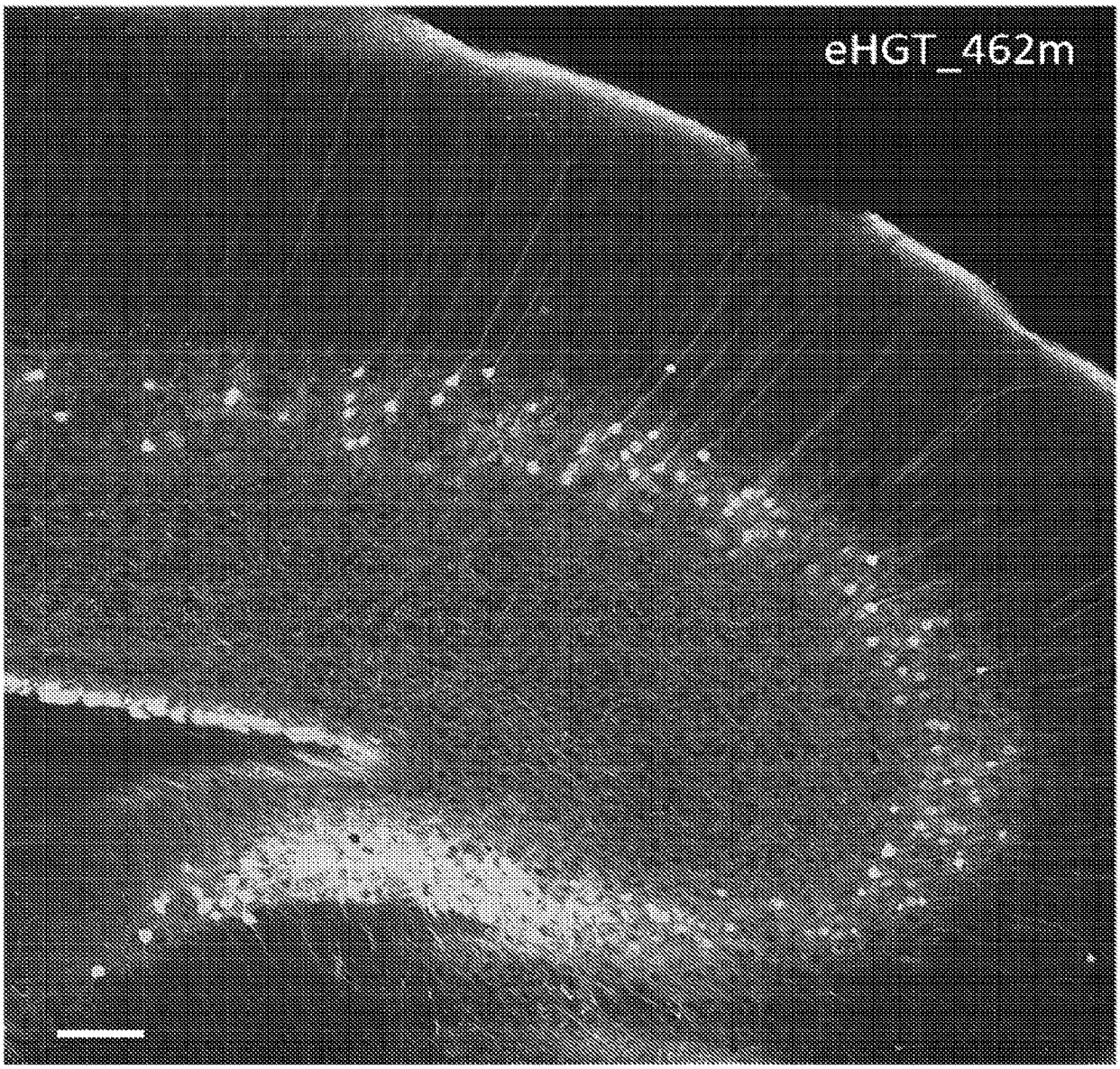


FIG. 8A

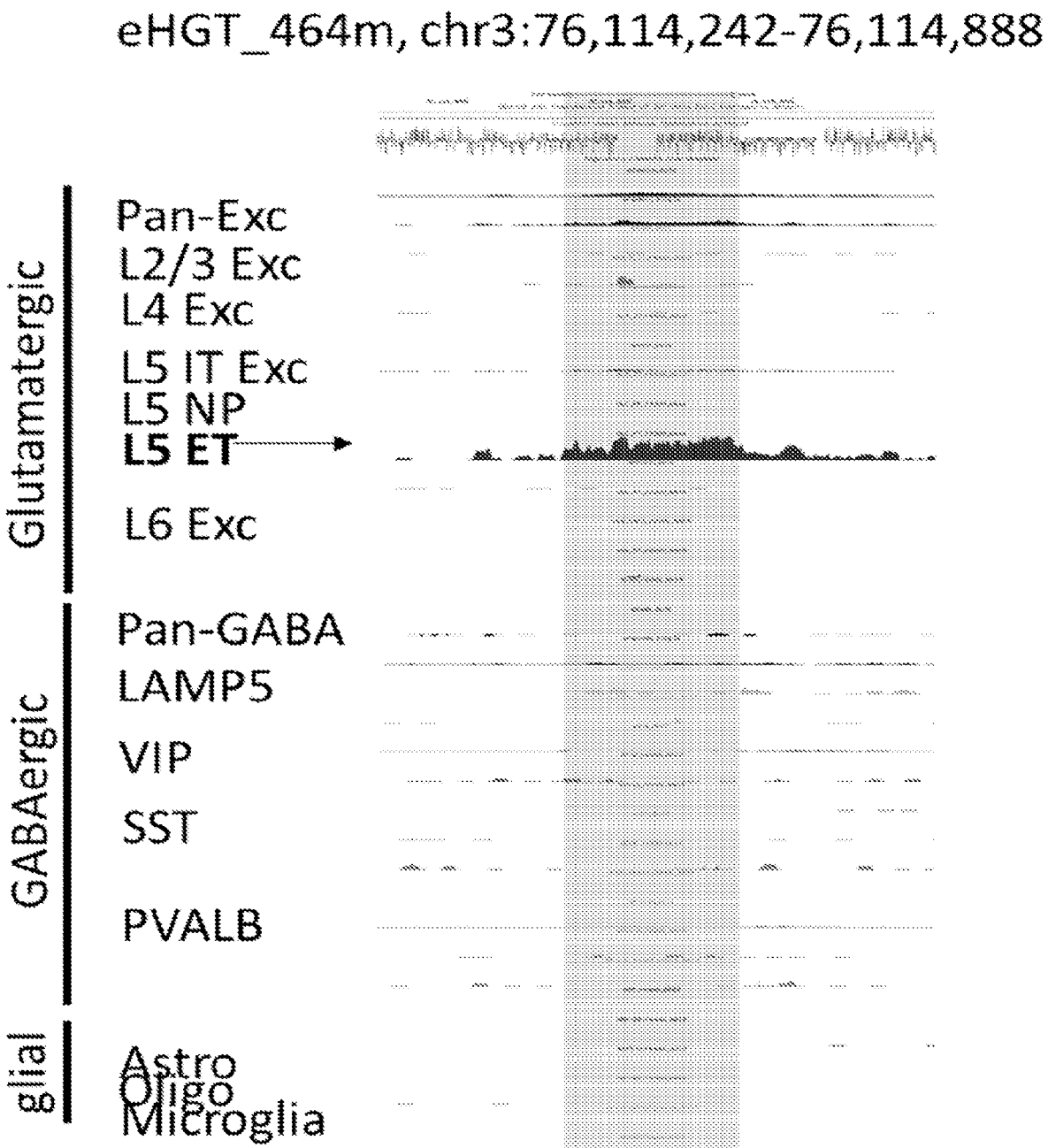

eHGT_464m, chr3:76,114,242-76,114,888
Glutamatergic
GABAergic
glial
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
Pan-GABA
LAMP5
VIP
SST
PVALB
Astro
Oligo
Microglia

FIG. 8B

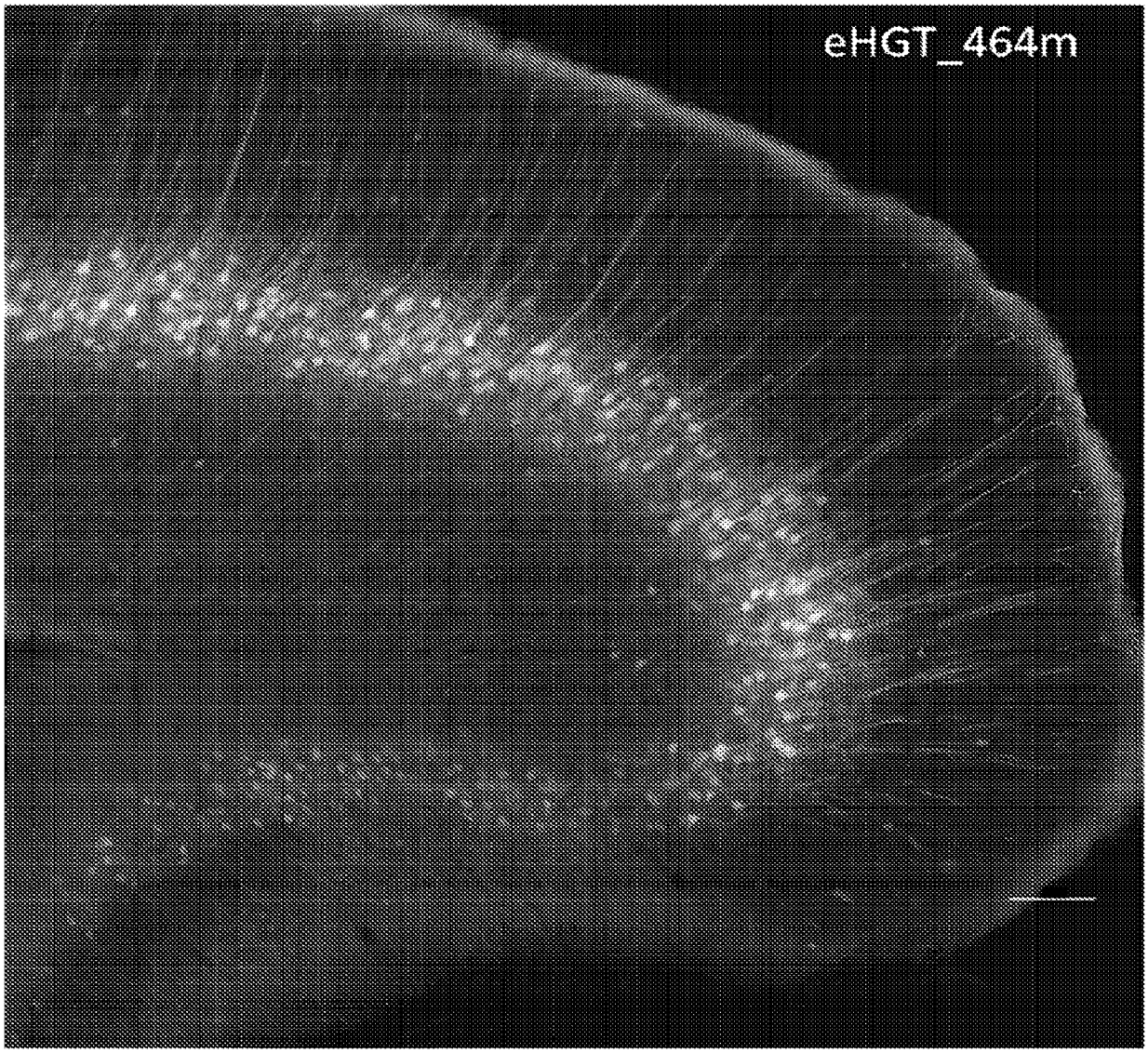


FIG. 8C

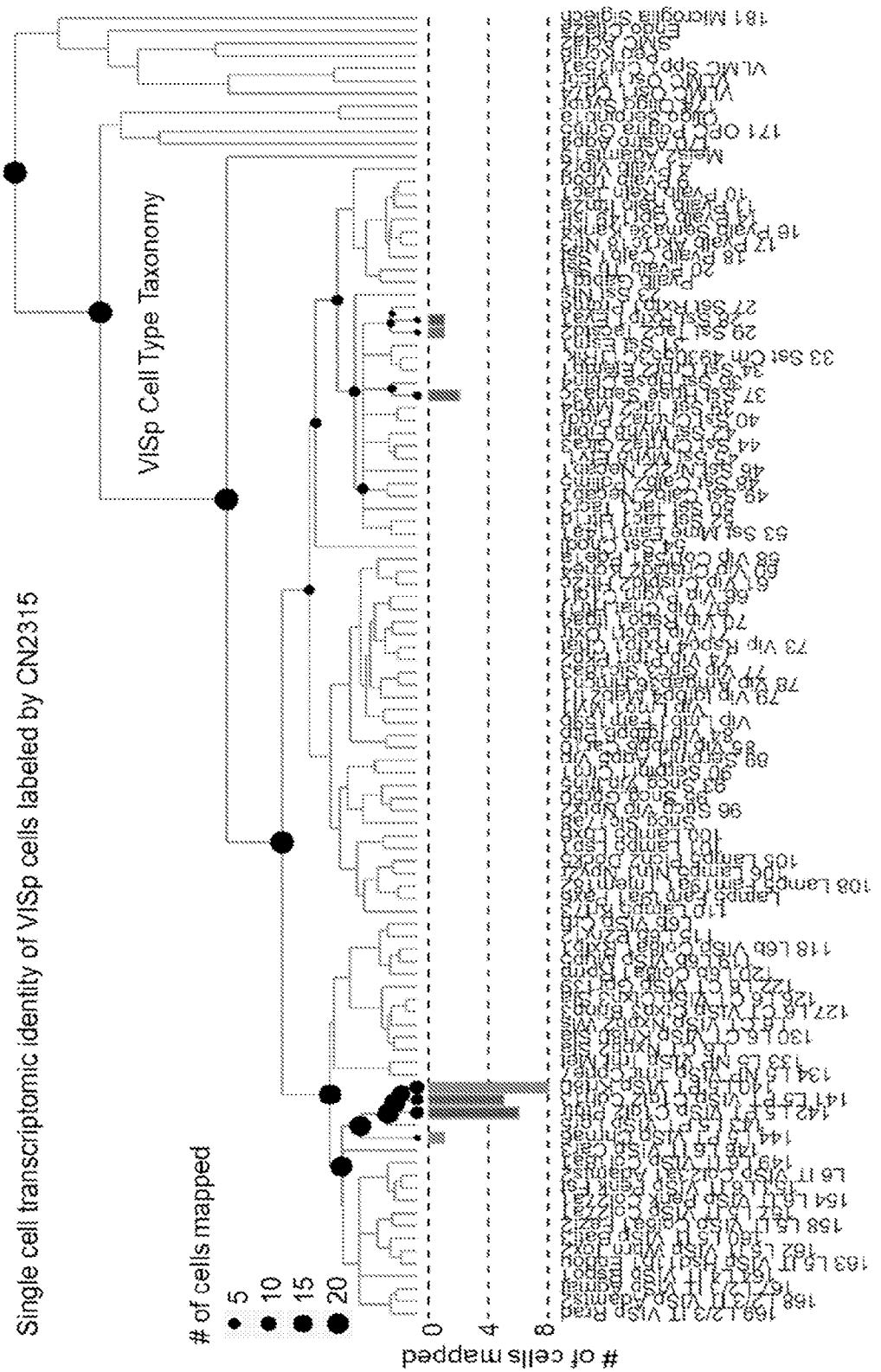

Single cell transcriptomic identity of VISp cells labeled by CN2315
VISp Cell Type Taxonomy
# of cells mapped
5
10
15
20
0
4
8
# of cells mapped

FIG. 8D

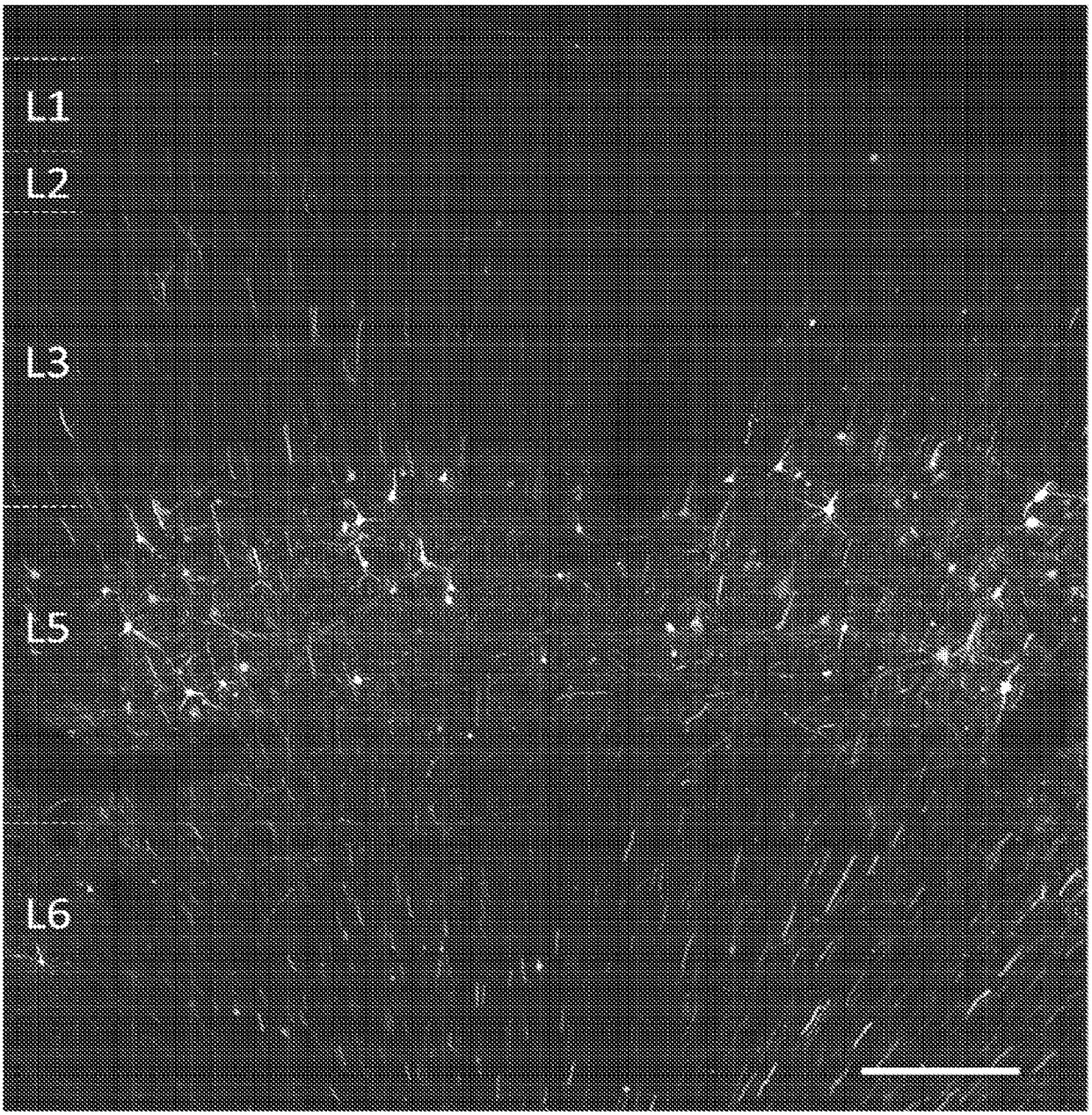


FIG. 9A

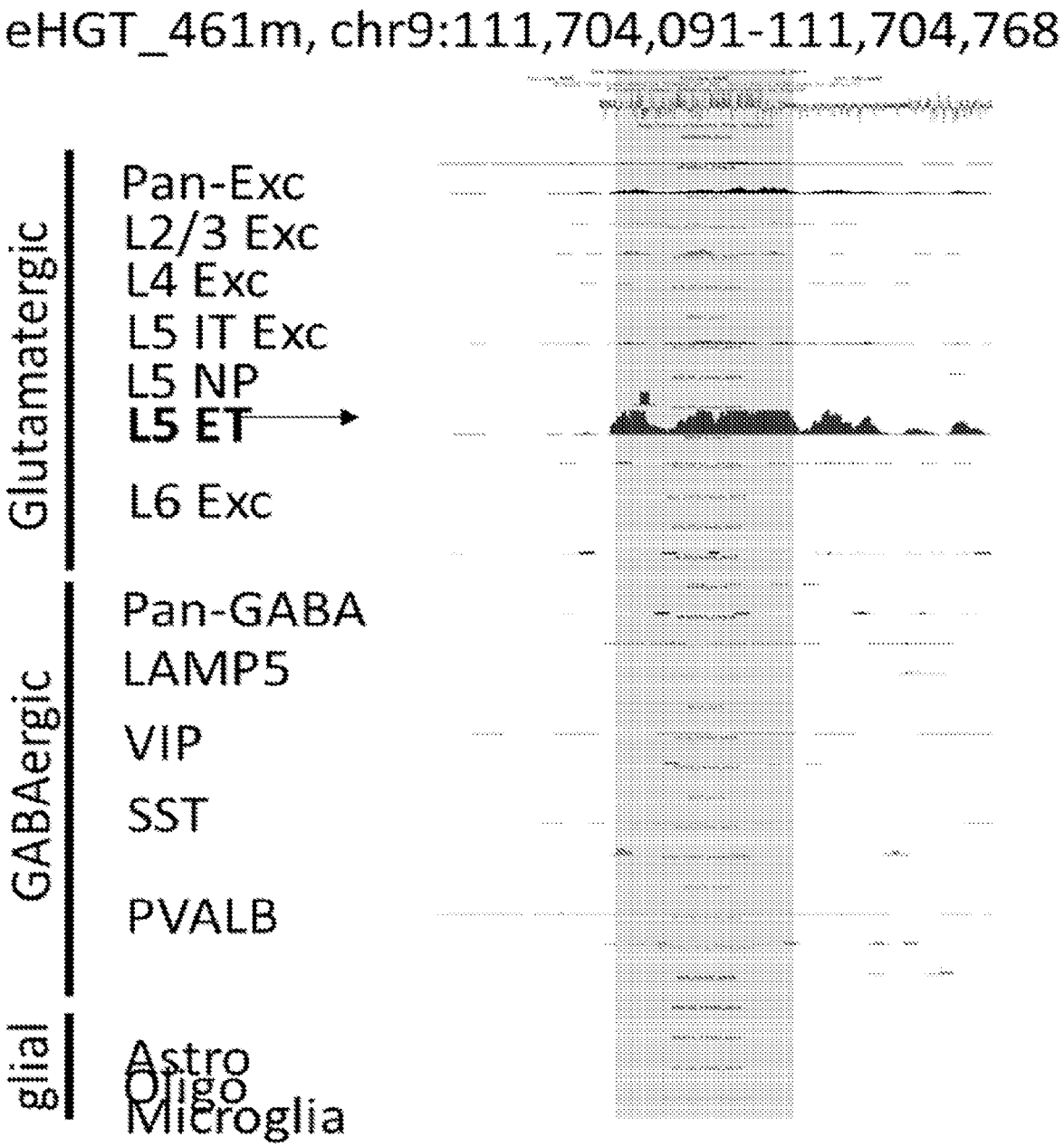

eHGT_461m, chr9:111,704,091-111,704,768
Glutamatergic
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
GABAergic
Pan-GABA
LAMP5
VIP
SST
PVALB
glial
Astro
Oligo
Microglia

FIG. 9B

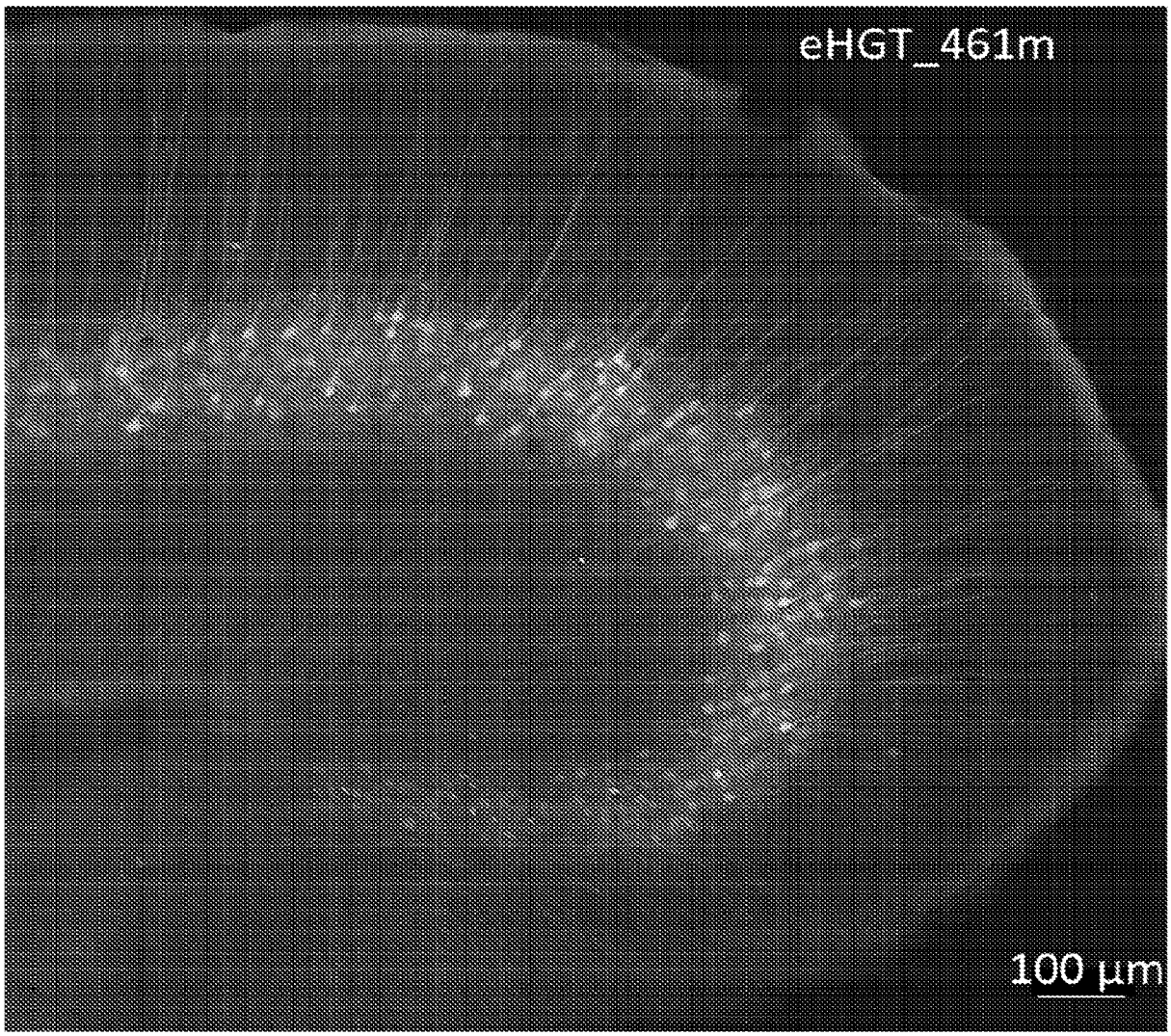


FIG. 10A

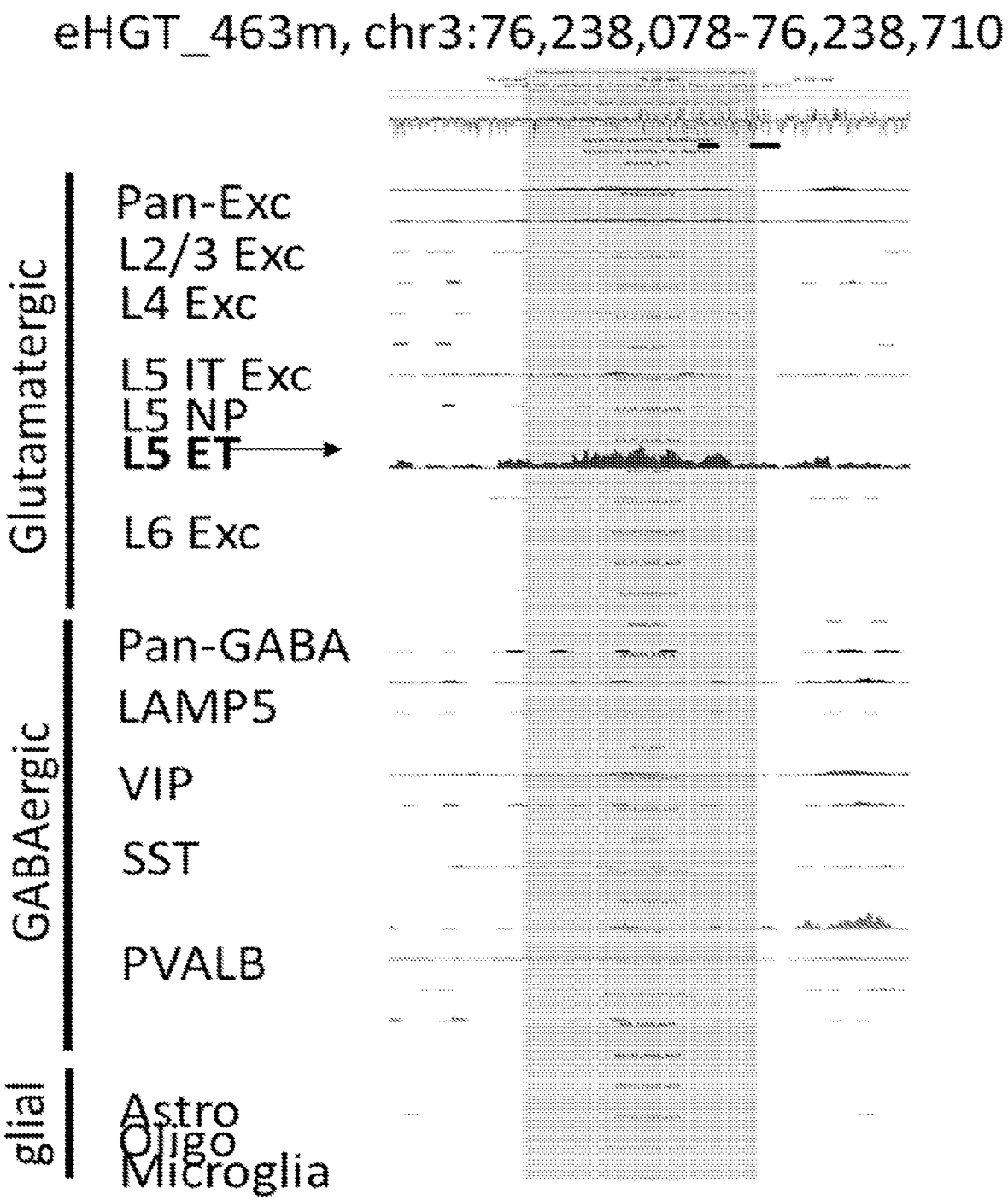
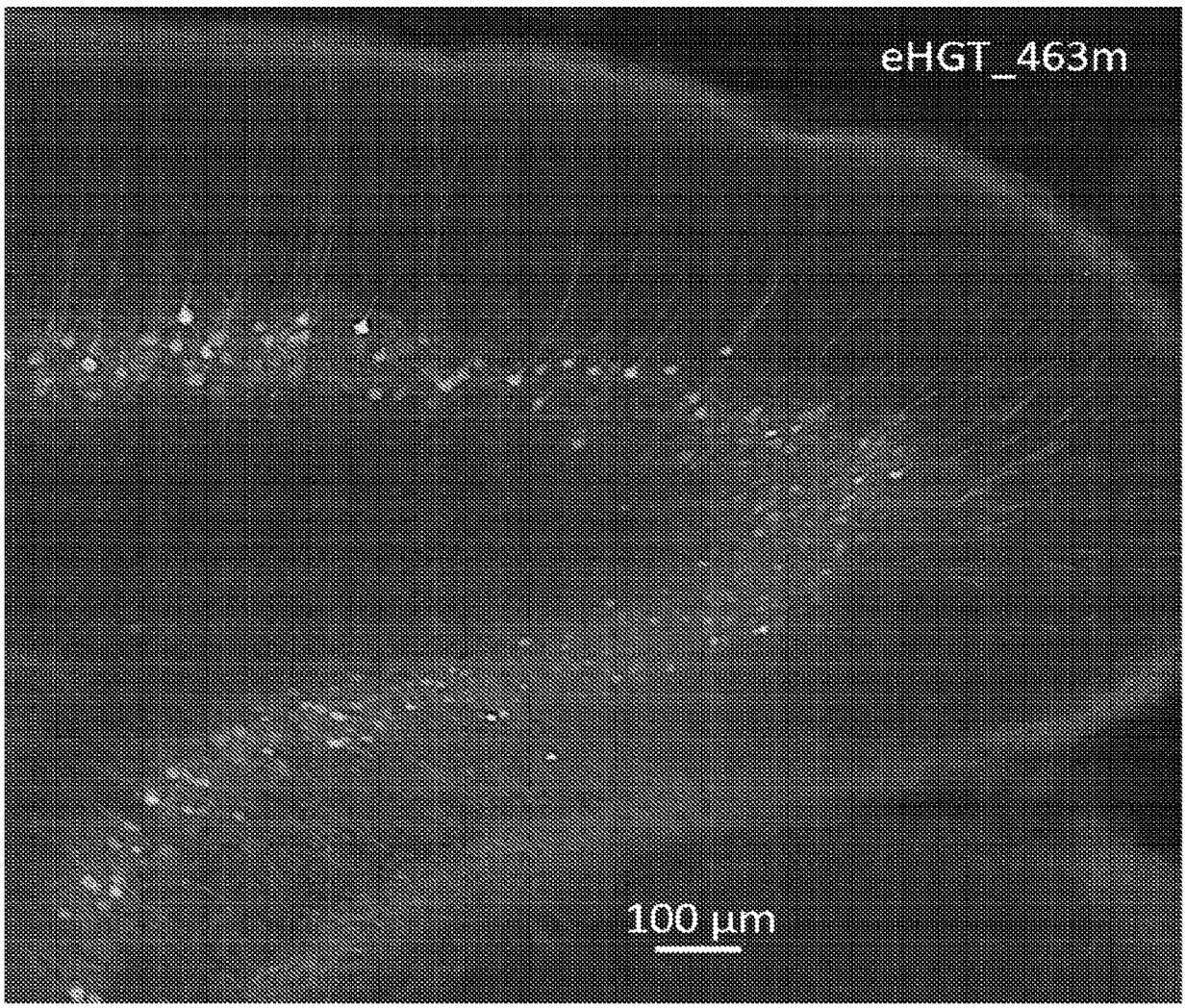

eHGT_463m, chr3:76,238,078-76,238,710
Glutamatergic
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
GABAergic
Pan-GABA
LAMP5
VIP
SST
PVALB
glial
Astro
Oligo
Microglia

FIG. 10B



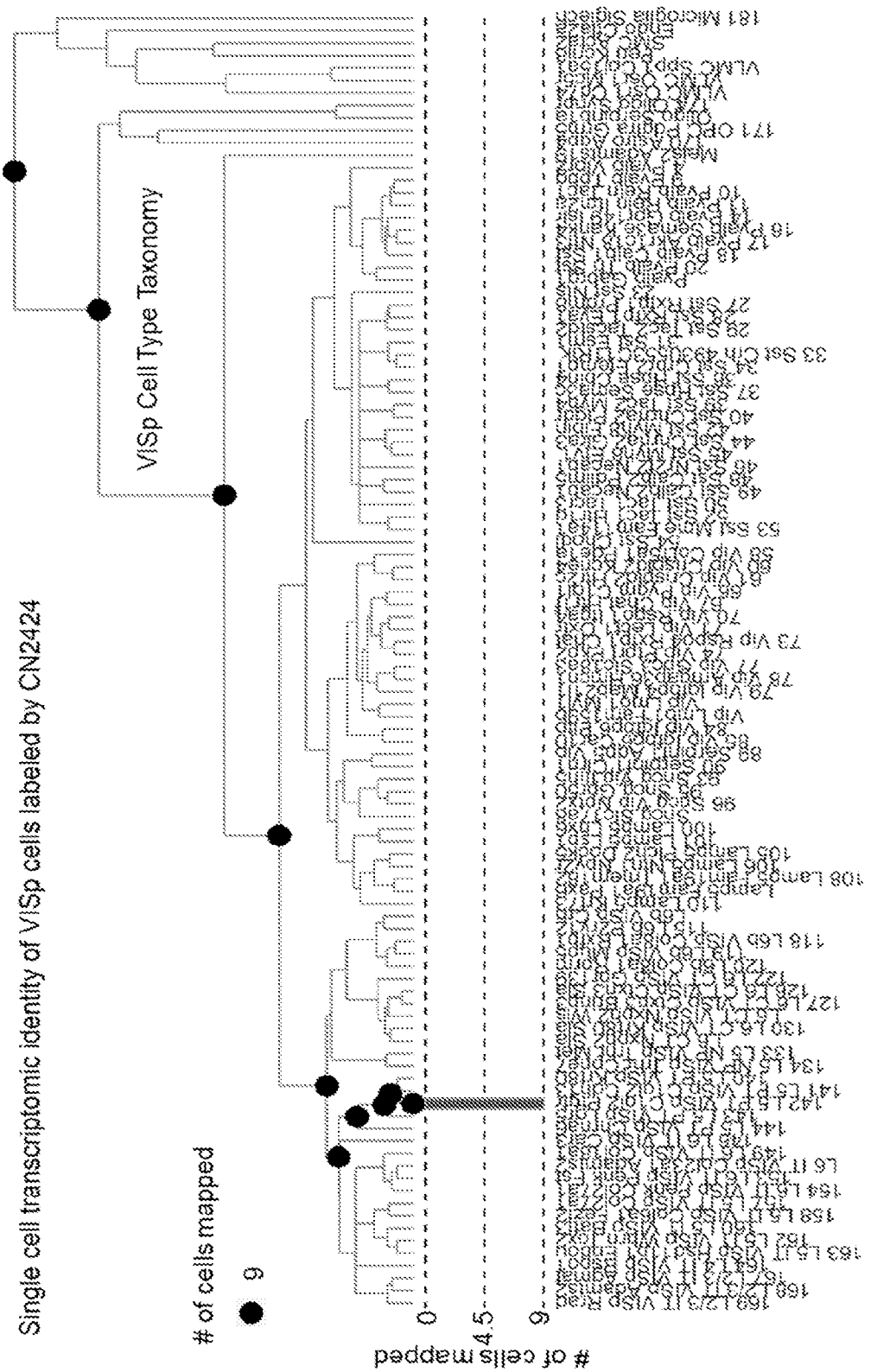

Single cell transcriptomic identity of VISp cells labeled by CN2424
VISp Cell Type Taxonomy
# of cells mapped
9
0
4.5
9
# of cells mapped

FIG. 10C

FIG. 11A

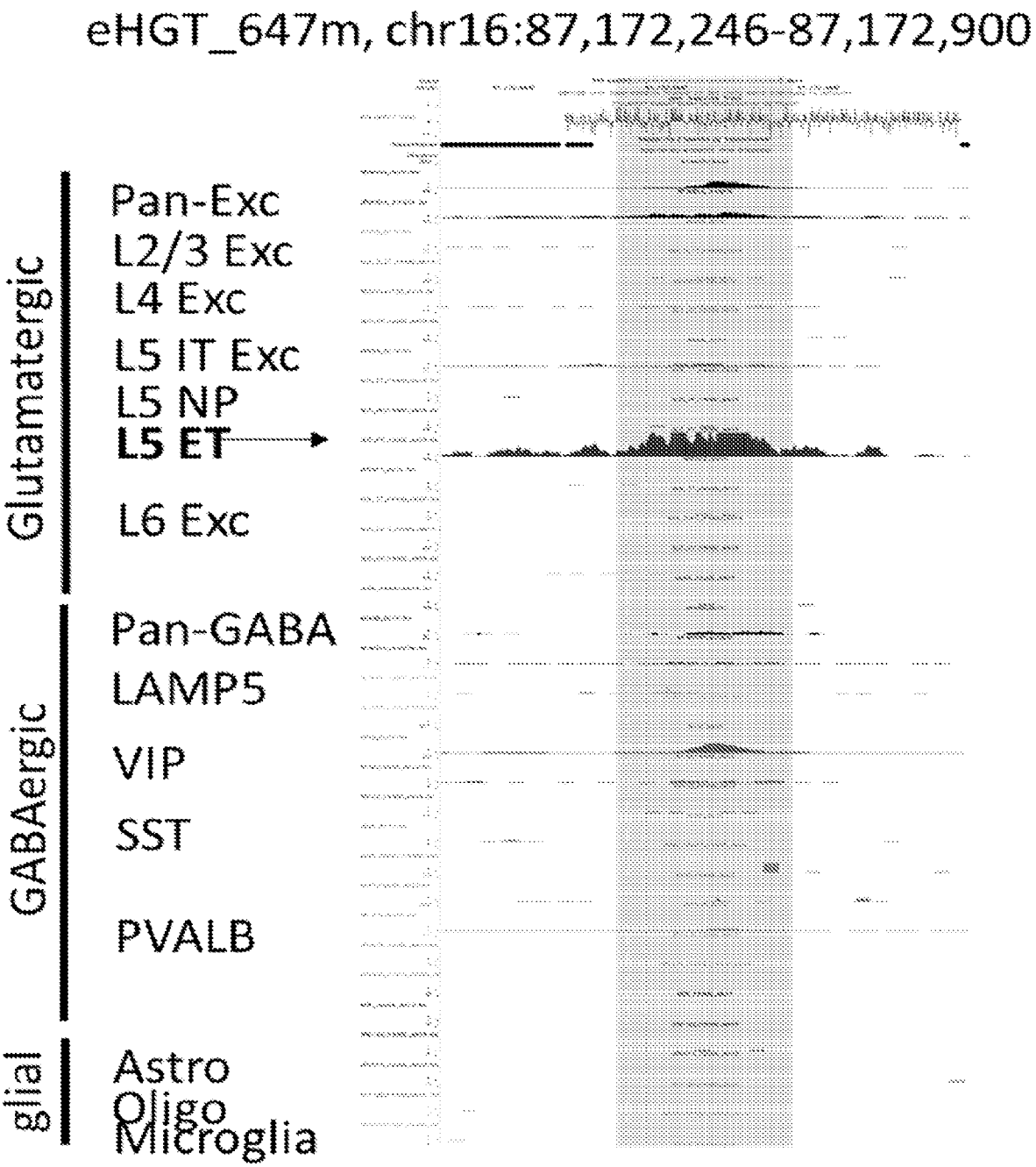

eHGT_647m, chr16:87,172,246-87,172,900
Glutamatergic
Pan-Exc
L2/3 Exc
L4 Exc
L5 IT Exc
L5 NP
L5 ET
L6 Exc
GABAergic
Pan-GABA
LAMP5
VIP
SST
PVALB
glial
Astro
Oligo
Microglia

FIG. 11B

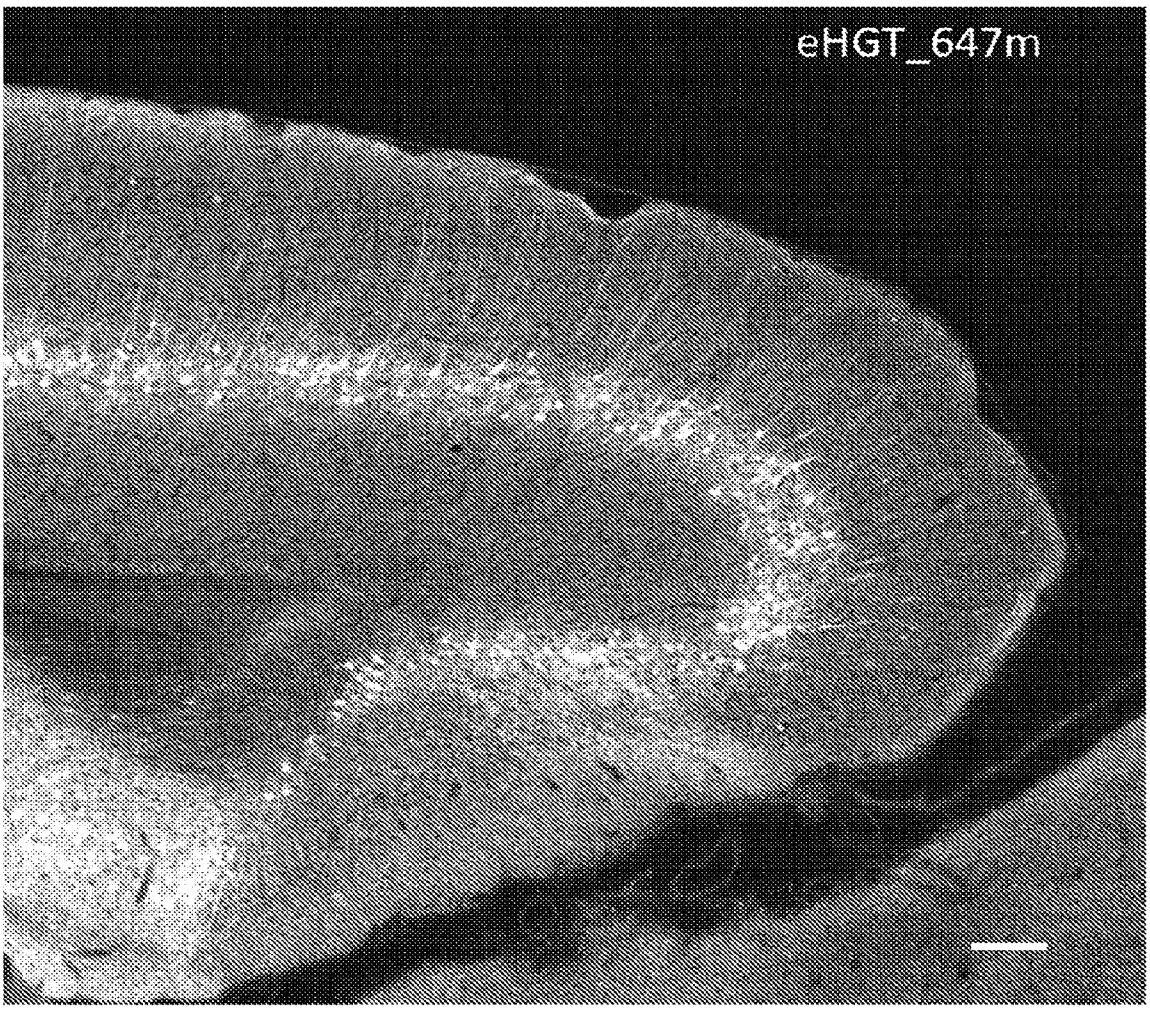


FIG. 12A

eHGT_450m (484 bp):
TTCCGCGTAGACCTGGGGTACTTTTCTAAGACACCCTTTATTGTTTGTTCCTAACCCTAGCC
CTGTATCCATGCAGCGTTCTCCATCACCTATGTAAAACTTAGCAGATCTGAGTCTTCCTCTG
AGGAAACAACATCCTGTGCTGTTAGACTGTATTTTCCGTCTCCAGATTCTCTACTTCGTCCC
TTCCCCCGCCCCAATTTCCCTGTCTCCAGCGAGCAATATTATCTTTCATCTGAGCCTCTCTC
CCTGCCTCCAGTGACTGCCTCCCCTGATGAATTCTCTCCGGGTGACTTTGGCATTTGCATA
TTTCCTCGCTTGGATATTCTGAGATGTTGGAACCGGAGAGCAGGTTGACGCGATCAGGTAC
TTAGCTGAACGAGGCTTTCTCGGGAACTGCACAAATATTTGGAGTCGGGCATGTGCCAATA
GGCAATTAATTGAAATGTGGACACAGCAAGCTTCCGGAGGAGGGAAGCTGAAG (SEQ ID
NO: 1)

eHGT_451m (615 bp):
GCTAAAGACAACATTGGCCTGAAATTCAGGAGACCTAAGTTCTAGATCAATCTGTGGCATG
GTCTTAGACCACAGATGGGCTTTCCTGCCTGCATTTACCCAAAGAGCCCAAAGAGGAGAC
CACATGGGGTACCATGGGACTGACCTTCCCAGTTCAAAGCTTCTTTACACAACGCTGCTGA
ATCACCCGCTAGCTCGCTCTCTTATCATTCCTTTTCATTCCTGGAGGTAAAACCAACCTTAG
AATCGTGGAGCCTGATAGACTTTGAAATTGGCCTTCATTTTAATTCTCAAGTTCTTAGCATG
AGAGCAACATTCAATATGGTTCTGAAATATCTCATGCAAAAAAGTGCCAGAGTCATAGACTG
AAATATTTATTTGAATATAAATGAACTCAGATTTGCCAAAATGCGCTCCCCCTGTTGTCATG
GACACAGATGCCTGCCTTTCTGATTTGCATTTTGCTCTAATGGGTATAATGTGTTTCAGCTC
CTCATTGGTGCTGGGATTGGATTTGAGAAACTCACCAGTCCATGAGTTCCCCCAAAATGTC
AGCCAGTGGGAAACTTTTGTGTCTTTGATTCTGTTTCATCTAGCCAGCTGGAAACAGACTG
C (SEQ ID NO: 2)

eHGT_452m (669 bp):
TGCTCAGAAACTGGCCTCGATGCTGTCCTGCTTTCACCAATTAATAATATATGTGGAGACC
GTTCTCTACTGGCATATGTGTGTGCATCACATAATTATATGTATATATGCAATCCTTTTTCTT
GTTGTCTTTTAGAGTCAATTCTATATATATATTTTTTATTAAGACATTTAAAAGTTCCAAAGTA
AAAATAACAAAATCAAGGATACATTGACTCATTTTTATTTTGTGAGGAAGAAATTTTCCATTT
GTCTGTGTAATTACATTTGCTAATTTCTAAAACAATTCCTGGGATTTATATCTATATATGCCT
CAGGAACATACAGGGAAGATCATACTGTTCTCTGAATTTCAAATTATGTTTTCTGTTTTATTT
CATGAATTTATTATTTTTCTTTGTCTCTGTCTGTCTCTCTCTGTGTGTATGTGTATGCGTGTA
TATGCACGCATGTGCTGTGCACTTGTGTGTATACATTTTTTTGCACGTGAGAAGTGTACCTA
CGAGGCTGTGCCTGGGTGTGGAAGCCCAGGGTTGGTGGGGAAGACCATACTCCAAGGCT
CTCCAACCTTATTCAGTGAGGCAGGCAGTCACATCTGTCTGACATTTCTGATCATTACACGT
GTGCAGTGGGTGTTTTATCCACCGAGCCACCTTCCCAGCCATGGAT (SEQ ID NO: 3)

eHGT_453m (721 bp):
TCACAAATCAGTAAGAATTACACTCCAAAAATAAAACAGGGAAGGGTATTAGTATGTGATAC
GCATGGATAAAGTAGAAGAGGCCAAAAATACAAAGAACCGTAAACTCACTAGTGATGGAAG
AAACGCTAATTAAAATAATGAGCTTTCCTGACCCTTTATCACATGGGCACATTTTCAAAGGA
CAAGACTCTGCTGAGGAGAACACAGAACAGGATCCTTCAGACAAGCTTCAGCAACCAATTT
TAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTAAAAATAACATAGATCTGAAAAGC
TTTATAACTTTTGGTTTTTTTTTAAGGACCATATTAACAGCATCTTCCAGTTTTTCTAGAGTAA
ATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTGTAAAAAGACTATAAA
AATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGCCTTCAGC
TCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTCACACTGCCCTGCCTTTGT
GTGCCCTTGCCTGTCCCTCTGTGACGGTCACCCTTTCGGTTCTGGCTTGGAGCCTGGTCA
TCAGTGCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTCGTTGTCCTAGTGA
TGTGACCCTACCTTAGACTTCCCCATCTCTTGAGGTGGCTA (SEQ ID NO: 4)

FIG. 12A cont'd

eHGT_457m (472 bp):
ACGGATGTCAAGTTAGGCAAAAGTTTCATATAAATGTGGTTAAAGTAATTCTAAAGTTTGTTT
TTACATCACAGAACTGAGCAAATGAACAAATGTAATGATAACATTGGGAGCTGATTATTCTG
AGCAAACAGCAGAGAAATCCAAATCCAGACAGTTGGGGGGAAAGCCACAGCTACTCAGAA
AACACACTGTGATGGAAAACAACCTACAGGTTATTATTAGACTAAAGAGACATAAAATTAGA
TATGCCACTTCCAGAGACTATAGGTTAGACGATAATATATGCTGAGTGAAGTGCCCTGTTTT
GGTAACTGCAGTAAGGCTGTGTATCAGAATGTTCCTAGCTACATATTTAGGTGTAAAGCAG
AGTGAGTCAAAATATATTCCAGGGTTAAAAAAAAAAAAAACAATCTTTACTTCTCTTTTCTTT
AAACTAAATAATTTCAATTATTCTCCTATGTTTAGGGAG (SEQ ID NO: 5)

eHGT_458m (375 bp):
GACTGGAAAAAGGTAGCTATGGCAGCCTCTGCGCCTATATTTTATCCATATTCTGATTTTGT
TTTCCAAGGTCTCCTAGTCAATCATAATCTGAATCAATGGAAATTTCCACGGATAAATAATTC
ATGAGACTTGAGGAAAATGTGTGCTAGTCCTCAATACCTTATAAAATGTATTATTCATTTTGA
ATTAGATTAAAATGTTTAAAAAATGCTATCTATAAGACACTGTGGGAGCCAAAAAATTTATAC
TTCTTTATAAAATCCTTAACTGCCTGGAAATAAACTAGGTGTTAAGAGAATTGAAGTTAAAAC
ACAGAAAAATAATCCCACTCTTTCATAAACTCACAACTCCCAAATAATGAGAACCAAAGGT
(SEQ ID NO: 6)

eHGT_459m (787 bp):
TGACTGTGTCATGACTCTCAGTGTCTCTTTTTTCCACGGTTCTGGGATCATCATTTCTTTCT
CTTCAGACAGTCTAGATGAGCAGCCGAGTGCTTATCTTTAAGGCTGACTCTAGACAGCCAG
TTGTTATTTCATAAATCTGTTAAAACTCACCGTGCAAACACCAGCGTTCCTTAGCTGACCAT
CCTCAAGCACACGTGGAGTGTGACCTCTGCTTCCTTCTCTCATTTCACATATTAGTAATGTG
CGTCTATGAATAGTTACAAGTGAGAAAGAGAAATGGAGATTTTACCATTTAATCATCTCCGA
AGAGAAGGTGAGAAGAAATTGAGAAAGGTAATTTGCATACCATGCATAGGCATTAGAGAAA
GCATTTTTTTTCTTAACTCTCCATTTGTTTACAGATATTTCCATATGGCCTTTTAAAAACTTCA
ACAGAGAAATTCGAACACAACTAACTGGAATTTCAATTTTCATATAAAACCAGCATCATAGA
TATATATGTAGCTACTCTCCCACTTCACATTATTCAGTTAGTATATTTTCTAAGCTAATTTCCA
ATCCATTTTAAGACTATGTACAGCTCATATGAGGCCCCCATATGTTCGTTCACGTTACTTTG
CATCCGTTTACATGATTAAACCCCAAGAACAAATCGATTAGAAGCCAAGAGCAAATGCTTTT
GGCTCCACATTAGAAAACTCAACAGAGAAACCTTTACAAGTGATAGTCATCTGGGTAAGAG
ACATGCAAACATAATTACCATTATTCGTGGACTTAATTCCCT (SEQ ID NO: 7)

eHGT_460m (759 bp):
TTAACTTGAGGAAATCTGCCACAAGATGTTCTCAAGTGCAAATGTCTCTTGATGCATACACG
CATGCATTTCTATTTAAGTATCCAACGAGAATTGTGGGGTTGAGGTAGGTATTTTAGCTTAT
AAAAAAAAAATACCTTATGAGCTTATTTTGGATCATATAAAAGTATGAGCACTTTTAGAACAC
TTTTAATAATGCAAATCATCTCTGCATCATCCAAATTTAATCCATGAACATTTAGTCATCGTT
ATCTTACATGAGTCATATCTGTGTCCTTATTGTCTTATCTTAAATGCATTAATTAATTCGATA
GAGCATCTATAAAAATGAAACGCTATTCAAAAGCTGGCCCATGACAAGATCAGTGTGTTGG
TTTCTATGCCTATAGTGGTGCAATGGAAGGAGTGCGCTGTACCCCACAGACTTTGTGATTG
ATGAGTTGTCTACTTTATTCACAATGAAGCTAAAAACAATCTAACAACCATCAGGGTGCTAC
AGCAGCACAGAAATGACTAGCTCTTTTGTGTCAAGTTATTGCCGGTTGTCTTAGTCACTGTC
CGACTGATGTGCAGAGACACCATGACTAAGGCAACGCTTATCAAAGAAAGAGTCTAGTTGA
GAGCTTGCTTGTAGTTCCAGAGGTTTGTTTAGTCTAGCATGGTGGCACACACTGTGCTAGA
ACAATATTGAGAGCTACAACCTGATCTGAAAAGCAGAGAGAGATAGAGACAGAAAGAGGCA
GAGAGACACTGGCCCTG (SEQ ID NO: 8)

FIG. 12A cont'd

eHGT_462m (405 bp):
GCTGGTGGGTGGGAAGAGTTTAGTAACTGCTCACCAATGGTGCTGTAGACATTATCAGTTG
CAAGAAGTGTTTTAATTTCAGAGAAATTAAAATATGCAAAATATGTGTCTCTGAATCAACAAA
ACATGGTGCCTCAGGTTTCCGTATCTGTGTGGGCCTGTTTTTAGGTTTTCTGTTCTATCTAT
TCCTGGGCTTCCACCACACTGTGTTAATGAGGATAACTTTAAAATACATCTTGATATTTAGT
AGAGCGCATTTCCACGCCTGATTCATCACGAGCACCTTAACTTGTCTTGGTTCTTCTGCGG
AGATGTTGGCACCAGCTTCTTCACCCGTGACTCCAGCTTCTGGGGAACCTGTGCTCTCACC
CCTGCCCCGCCTGTGCCTGCCAGGCACACCTTAAT (SEQ ID NO: 9)

eHGT_464m (647 bp):
AGTTTTGGAAAAACACCATACTGGTTATTTTCAAATAACTCTATTTCACAACAGTATACTAGA
ATTTTATGCTAAAATCTTTCTGTGACTTTGTGTTCACTGATCAACAATTCCTCTTTTATGATA
GCCTATGCTGTCTACCATCTATTCCAAACTTTTATGAGATCAGTAATTTTAGATTATGAATGA
GTTTGATCATATAGTACATACTTTTGTGTCTGCTATTGCATTTACATAGTGACATAGGCTTAC
AGGATATTGTAAATGTCATCCCTAGATACATTTATGTAACCTTTTTTTTAAAATAAAACTCTAA
CAAATGTTAGTTATTGTGCACTATGGTTTTGAATATTTGAATTTGCTTTGCTCAGATCTAAAT
AACTTCTGATAAGTTTGGCATTCAGGAAAGAGTAAATGTAAATGTAGAAATGCTTGGAATAT
AAAACAAGTAATTTTTGTGATAATTTTTCTGAAAATGCTAATGTATATTTTCAGCAGGCCTCC
TTTCTTGTTTAGAACAGTATATTTCTTAGCTTTAATTAGTACTTACTTTATATATCTATGCATC
CATTTATAGTTCAGCATTAGATTGAATATGTATGCATGTATGTATGATTTATTTATATTTTTGT
CTCCCAGTCCACACCC (SEQ ID NO: 10)

eHGT_461m (678 bp):
GGGAGATATAAACAAGAAGAAACAGGACTGAATAAATGTGTGCAGAAGGATCCTGTAGCAG
CGTCGTTCTTCCTGGCCAGTTGAGCGCGCGCAGCACTTAAGATTGGTGCTCATCCATGCC
TATTCATTCTTGTGATGTGTTGTGTTAGATATTATTGGTAATTATTACTTAGAGAGGTGGCAA
AGTAGTGAATAGGTGGAAAGGCACTGAGACCATATTGTCACCGCAGCTTTGGAAAAGACCT
TCTTATTTTAACTTATGTAATTAACCATCCCAAATTTACCCTGATCCTCACTGGAAATGTCTT
CCCTTTCCTGATATTTACCACAGTTTTAAAAAGTCTTAGATGGGTATGCATATTTGAGAAATA
CGTAGAGTTAATTAGTGCGTAGATGTGTTTTTAAGTTGCATAAAACATATTGTGCTTTGGCA
TAGCTCCCTTTGTAACTTATTTTAGGAAACCAAGTTTGATTTTGACAACAGCATTAGAAGTAA
AGTGTGTGAAGATGATTTAGAATCCATGAATATCCACATTGCTGAATATAATCCTTCCTGGC
TACAGAGCTCCTATTTCTGCTCCTGTAAGCAGCCATTCACCCACACTTCTTAAAGTAACCCC
AATAAACTCACTCATTAAGTTAGATTTTGGCAGTGTCCTTCTTTGATCTGTTGTCCGT (SEQ
ID NO: 11)

eHGT_463m (633 bp):
TTTGCTTGTGTGCCAGCAGGGAGGAGCAAGAGGAATTTTGTCACTTCTCTTTCACTTTCTTC
ATTATTTTAAGTTGGATCTGCTGTTCTGTTCACTTTTAAAATATAACCTTCAGTATTTATGAAC
GTGATTCATCTTTTACAATATGTTGTTTAACATGAATGCATCTTCTTCTTGCCTTTCAGTGCA
TCTTTATGCCATTGGCAGCCACAATCAAGTGTTTCCATTAAAGTTCACAGTAACAGTGCAGT
GTTGTCTGTCCCAACTCTGCTAACTTTTACCACCCACTATTTAGTTTTATCTACCATGTGAAT
TTATTTGTAAGATCTTCTTTCTCTTATCCTAGTTAATTATGGGGAGGACTTGCCAGGATATTA
ATTTACATGGCAATTTTGGATAAGTTATGCATGTATTCATCTCCTATATACAATAAGTTTTTCT
TGTGAACGTCAAAAGAAAGGATAATGCATATGACTGTATATGATATAAACTGGAATGTAATA
GTTGCCAATAGGATCTGAATCAAATTTCAGTTTTTAACTATGTTCTATATATCTATGTGCATA
TTTATAGCATGTGTGACATATGTATAATTTAAATATTTCTATAAACTTATACAGAAACAAAATA
TGT (SEQ ID NO: 12)

FIG. 12A cont'd

eHGT_692h (601 bp):

CTGAAGTTCTGGGGGCAGACCGCAATTGCTGCCACAGAGCCCCTCCTGGCAGGTGCAGC
CCACAGGAAGACCAGCACCCTGGCCACAGAGACAAAAGGTCTTATTGACTTGCAGTAAAAA
TGTCCTTATAACCTGTCCAGAACAAAAGGGGAGGATAAGGTTCAACCAGACACCGAATTCA
GGGAGACAGCATCAGTCAGAGCTTTACCACTGCCAGCAATTTTGAAGCTGCCCATATTTCT
GCTCAGGCAAGATGTTTTCTTGCTCTGAAACCTTTTAGCAGATGCAAGGTATTGCTCCGGG
TGTATCCCAGAACACAAGTCTTTGCTAAAGTGGGTAGACAGCCACTTTCCTTCAGAGCCAG
GGACTGCCCTGGGAGTAAAACCATTTGGAGAGCATTTTATATTTGCAAAATGTTGCTGTGA
CATGTTATCATTTATTCCTTTCAACAGCGCAGTGAGATTGGAGAGCATCCTCCGATGCACT
GTACATTTGGGAGGAAGAGAAAGGATAAGAGCCCTGCTGATTCAGTGGTGGTCATTATCAT
TGTCAATGGAGCTGTCACTGTTATCGTCATTTTCGTTTGCGGTGAGGTATGTGA (SEQ ID
NO: 13)

eHGT_693h (518 bp):

AGAAAGCAAGAGCAAAGGGCAGGTCTTCCAGGAGAGACAGACAGACAAAGAAAAAGCAAG
GGAGGTGGAGAGTGTGTGAGAGAGGGAGTGAGAAAGGGCTCAGCCATTTGTCTGGGCAG
GTCTGCGGAACTCAAATCCGGCTCATCTGCATCTCCTCAAATAAACTTCTGGGAAGTGAAT
GCCTGAAGCAGGCAGAGTGCAGAGAAGTCAGCAAATATGTGCGTAGTGAATTCCAAGCCT
TCTGTTCCCAGCTCCTGGGAGGGGCCGCAGTGCCTTGGGGTGTCCTGCCGTGTCTCCATC
GGGTCCCCTGCCTGGCTGCTCTAACCCTGAGGAATGGCCTCCATTGTGGGCCTGTGGAAA
CAGGGAGAAGAATCTCTTATCCAGCAGGTCTGAGGGCCACAGGAACATAAGTAACACTGA
AGAGTGAGGCTGATCCCTGGGGTTTGCGCGAGCCTGCAAATGCATGATTCAACAGCACTC
GCCTGGTGCCTCTGGGAGGCAAGCTGGATGAGGATGTT (SEQ ID NO:14)

eHGT_694h (613 bp):

AGAAATCCCCTTCCCTACCTTTTATCTATCAATCATATCTATCTATATCCCCCAAAACTTAGC
TGATACTAACAATTTAAGGATCCCCAATCCCAAAGCATTTAGATTTTAAGACAAATAAAACA
GTTGGAATCCTGGTCCCATGGAATTTCAGAGTTGGGAGAAAGTGAAGAGATTTCCTGGCTA
ACCCCTGCAGTTAACAGATGGCAAAAAGAAGTACAGAGAATTGTCTGAAGTCCCAGAGCAA
GGGAGTGGAAGCGCGGCCGCAACGAGGCATGCATAAGATTCTGCAGAAACCACGTTCTCC
CGCTTCACAGCGGTCTGCTCTTTTCAGGAGAGCCTGGGAAAACTTCTGACTGTGCAATTTG
CTTAAAAACAACAGAGAAACATCATAGAAGAGGATCAATTCCTGACCATGCATTTGAAATTA
GACACTTACAGGCATTCCCTGCAGAGTGCTCTCCACTGGGGCCCAGTGGAAGCAGCATCA
CCCATCAGCCGCTGGGGGTGGCTGGTGCCACCAGGTCAGGCCTGGAGCGGGGCAGGAG
TCTCCTCAGGCTCGATGCTCTTGACCTGCTGCTGGTGCCAAGAGTTAGAAGCAAAGCACC
CACCA (SEQ ID NO: 15)

eHGT_695h (406 bp):

TTCCTTGGAACACTATTTGGCATATATGCATATTTTGGGGCTCAGTGTCTTTCATCGTTATG
TTTATGGTTATAATGACCACATCTCTCACCCATTAGCTGTGTGGCTTTTGAGGTCAGTAAGC
TATTCTTCCCAGAGTGCTTGACATAGGGCCTGCGGCATGGTAAATACCCACAGTTTTTGAG
TTTCCTGTGGGTTGAGCATTTAGAATCATCTGTTTGTTTAATCAGGACATATTTCTCCAGTAC
CTTCAGGGGTCCCAGTACTATGTTTAGGAACTTAGCCATCACTTGCACAGTAAACAGTATCT
GTGGATTTTGTCATTCAACAAACACTGCCATTGTTCAGTCACTCAACAAACGCTGTTACGGC
ATCCACCCTGGGGAAAGAAGGGTCTCTTGGCAGA (SEQ ID NO: 16)

FIG. 12A cont'd

eHGT_667m (335 bp):
TCCATCTTTCTGACTTTAGGGCCATTTTTAGAATGGCCTTTGTTACCCCAAGATTACATAAAT
ATTCACCCAAACTTCCTCCAGTACTTTTATGGTTTCCTTCACAGGATTTAATTCTTAAATCCA
TCTGGAATGTATTTCTGAAAAAAAAAAAAAAAGTAAAATAGTGCCAGCTATTTTGCTTCAGA
GCCAATGTTTCCTAAGTTGAAAACTTTAGCTATATTCAGTTGTGTGATGTCATTAAAGATAAT
AAGATTTGGGACCAGATTCTGAGTTCAGTTGCACTCTCTGTATAGGGTATAGAAAATCATTC
AACTTTCTGAGAAATGTGCAGA (SEQ ID NO: 17)

eHGT_668m (304 bp):
ACACACACACACACACACACACACAAATAAAAACAAAAACAACTCTTAAAAACCTAAGAAAG
GGATAAATACAACATCAAAGACTGTTGGGGTTATAAAACATACACATTTATAGACATTTCCA
AGATGCTACTAAAATTACATACAAATGAGTATCCAAAATGTCACAGAACAAATTATTCTTGTG
CAACATAAACAACTAGAGAGCATAGCAAAGGAAGCCCAGACCATGAGCTTTCCCGATCTGC
TTCCCAAGCTAACATATGTTGGCATCCAATCCTGAAAACTACAGAAGACATGAGGT (SEQ ID
NO: 18)

eHGT_696m (524 bp):
TTTCCAGCACTGCAGGAGTCTTCAGTATGAAGTAGAACCAGTTGGGTGACGGGGCAGACA
CTGGAGACAGACAGAAAAGGTTCTAATATCCATGTTCTGTTTGATTTTGGAAGGATGCTGAA
TGCTAATCAGCTTTCTTATTTTAAGAAAGACTATCTTAAATTATTATACTTGCCTCATATGATT
TGGAGGATGATTGCATGCTGTTCCCTCAGCAAACAGGAGCTTAGAAAAGTTTTTGAATAATT
TCACAAAAGGCCCCAAAGTATTCTCAGTATGAAGATGGAAATTGGGCTGAGTCATCTTATCT
AGAGGAAGTATCAGAAAACTTCGTTGAAATAGTTTTGAGTAGGCCTTTGGCACAGAAGAAG
CCATGTGGATCAGAAGTGAAAAAAATGACGCTTGCCAATGCTTGCGGTGAGATCTGACATA
GAGTGGCCATCTTTCGGGATGCAGTGTTAGAATGTGTAGTAGAAAATAAGGTGGAAGAGAC
GTGTTCGGTTCCATCTCAAGATTTGGGGGCA (SEQ ID NO: 19)

eHGT_455m (700 bp):
TGCATGGATTCTAAGTTCCTAAGAGGAATTCAGAAAATACCAATTTTTATTTAGCATATAAGA
TAATGTCTTCCATCATAACATTCACAATCACACACAAACATACACACACACACATCATACT
TCATTCATATTCACTGTATTACTATCGTCTGCTAATCTTGCCAACTCTTACACTGAGAACCTT
TCTTGACTTAAATTATGGATATGAGAGACACTGTGATATTTGTTTTTCTCCACCTACCTCTGC
TTTTATATAAGGCATATTCCATGCATGCTAATGATATTGGAGCACAATGTTTATGCATGTATA
ATCATATATTTCTTTCCTGTGCCTGTGAAATCATAGTTATAACTAAATTTCTTTCTCTTATGAT
TCCATAAAAAATTCTCAGGAATATAAAGTATATATAAGCCAAAATTCTATAAAATTAATAAATT
ATGATGATCAGAAAGGTACAACAATGTCTTTCAGAAGTCTATAATGACTTTTGACCCAGCAT
GTTCATGCCCTACATGCTGTTAGTGAAATTAACTTTCCAATTCAAATTAGTACATTGAGTTGA
TTGAACATATTTCATCGTTCACATTGACTCACACTGTGGGATTCTTATAAAAGACAATTTCCC
TTAAATATAACCAGCTTTTAGTTTTTTCATGAAAGCCAGTCCCTTGTTACTAGTGTATTTAGA
TGCATC (SEQ ID NO: 21)

eHGT_457h (640 bp):
CTGTGTCTCAGGCCAGAAGATTTTCTATATTGGCTGGAGCTTCTGTGGGATGCAGGCAGTG
GATGAGAAGGGAAACGACAGCAAATAGAGGAAGCACAAGTCTCTGGCCTCAGTGTTGGGT
GCACACAAAGCAAATGGGAGAAGGAGGAGAGGGGCTCCCTTCCCCTGCCCCAAGGCCCA
GCCAGGCAGGCCCCCAGCAGCCTGGCTGTCTCCCTCCCTTGCGGGCTGGAGCTGCCTCT

FIG. 12A cont'd

CCGCCTCTGCCCGCCTGATCGGGGCAGAGCTCGGCTTCATGTATTATTAATAGCACTGCC
AGCCTTCCTGTTTTCACGCTGATAACCTGCAGGGGAGGGCTGGTGGGAGGGAGGGAGCA
AGGTGGCAAACCCTGGAGCTTCTAGGGTCCCTCCCCCCTCAGTAGAGCAGCTCTGCGGG
GGTTGGCCCCAAGTCTCTAAAGTTGGTTTCTAGAGCCTCCCCCATCCAAGCCCACGGGAA
GGGACCAGGGCCTCCTAAAGGGAAATCACATGGGAGGTTGGAACGTGGGGTCTGGGAAG
GGCCCTGTTACTCAATCCATCTAGGACACCAATGGCTCAAGTCACCTTGGCCCTCCCCAGA
GGCAGAAGTGCCCATGTCTAGGGCACATGGGGCTCAGAGCTGA (SEQ ID NO: 22)

eHGT_603m (493 bp):
ATGTGCATACCTGCCTCTGAGAAGGACCAAAGTCAGATGCACTCTCAATGTAAGGGCTTGG
GGGTGGAACACCTCAAGTTCCACACACACCCCAATCTCCTACAGGAGTTTCCTTTAGAGGA
CCTGTCTCCTCCAAGGGGCTCTGTTAGACCAAGTGGGAAAGTCTCTAGGAACTGACTTCAG
AGGTTTGGGGCTGTGCCAGTCCCGCGGCGCACTCTGCCACGGAGGAGACCTGGAGCTCC
TGGGTTGGCCACCTTGCTCCCTGCCTCCTCCCACCTCCTCCCCAGCAGGTTATCAGCATG
AAAACAGGAAGGCTGGCAGTGCTATTAATAAAACATGAAGCAGAGATCTGCCCCATCAGGC
AGCGGAGGCAGGCGGCAGCTCGGCAGCCCCGAGGGAGGGAGAGAACCCCCGCATCTGG
GCCTGGGCAGGGGTATTCTCTTCTCACAGATGCCACACCCCAGGAGCCACCTCCATTTTG
CGGGACTGCTGT (SEQ ID NO: 23)

eHGT_604m (466 bp):
AGCTAACCTACCAGTCCCACCGTGATATCTGGATGCTGATATAAAGGAGAATGTATCAAAT
AATGTAATGCTTGGAAGTTATTTGAATCTGTTATTCTTAAGAGCACACTGATATTACTCATTG
TGTGAGGCACACAAAAAAGATCAGAACACAAAGGCATATTTGTGCTGTGTGTCCAACACTG
AGGGGCGGGAGGAGGTAATACTTTTTTGGAAAAGACAGGAAAGAATGAAAATAGATGGTTT
GTAAAGATAGTAAGCTTGTGGGTGGCTTTTTATTTCCATAATCATGGTATTATGTTTTTGTAT
AATTTTAAGAAAAACTTATGTTAATCAGAAACTTTGAATAACACCTTAACAAAAATGTGCTGC
ACACCGCAGCCATCCCAGAGAAGCTAACACTCAAGATGTCCTGTTCAGGCTTCGCGCATTT
GAGTCCAGGGAAGATGTCTCTCGTTGGCACCAA (SEQ ID NO: 24)

eHGT_605m (429 bp):
TGCTCAGGGTCTGCAAACTTTTTCTTATGGTTAGTACAAGAAGAATGCATAAACATCACAAA
CCAAAGTACTTCTTTCAACCAATCAAACAAATTATATACAGGGTAACACACAAAATCAAACT
GAGCAAATCTAGCCAATGGGTCCCATATGCCACATTTTTCATTTCCTTATTCCAGCCCTCAG
GAATGTAGGTGTTTGCTCTGAGTAGGCCTACCATAGTATGTAATTCTGCTATCTTATGCACC
GATAATTTTTCTTGGTTCAGGAGTGTTTGATGTATGCTCCATAGTTGTCTCTCATTGTGCTCT
GCTGTCTCATTGATTTCTCTAAGGTTTTAACAAATTTATCCATGTGGTTTTTCTCTAATTTGA
AATCAAATTGCATGTCAGACAAGAAGTAATATTTTTGTTGGTAGTGTACTGTGGT (SEQ ID
NO: 25)

eHGT_661m (633 bp):
TCTAACCATGTGTGTTTCCTTTGAAATAATTCTATATACTCCTTTTAATGAATTTACATTAAAT
CAATAGCTTGTATTGACAACTATATTATCATCAAAGCATAAACAAAACTTTCTTTGATTCCTT
GCTTCAGAAGGCTTGATCATTTGTTTCCCAGGGAAATGGAATATAGGTAAGGACAAAGCTA
TTTAAACACAATTCTCTCTTTTAATGCAAACTATTACATCATGCAGTTTATTTATATTATGAAT
NCAAAATCAGGGATGACATATATAACATATGGACATGAATATTACTCTTTAGATTTCATATTA
AACTTATCAGTTTTTTTTTATCTCAAAGCATGGAATCAAATTCAAGACACAGATGGACACAAC

FIG. 12A cont'd

ATAAGTTTGGAATGGCTTTTAAATATTTTTACAGGAATTTTTCCATTTTGGCAAAGTATTTAG
TATATAGTGAGTTGCTTCTATGATTATATTAGATGTGTTTTTGAAGAATGTTGTTTTGCTAAA
TTATTCTAACGCTCAATTTACATTCTGAACATTTTCAATCTGAGCGCGCAGAGAGGGAGTGG
CCAACTCCATCACTAGGGGTTCCTGCGGCCGCACGCGTTCTAACCATGTGTGTTTCTCTTT
TGGACT (SEQ ID NO: 26)

eHGT_766m (440 bp):
AACATGTATTTGTGGAGTGGCTAAATTGAGATATATCACCTTAAAATCAACTCCCAGCAATT
TGAAGAGCACAGTAAGTAAGACCTTGTTCCTAACCACAGTCGCCATGCAGTGCAGTAGAGC
CCAGATCTTACACTTCCTGTGTAACTGGAGTCTTGTGTTCTTTAGCCAAGCTCTTCCCACAT
GCGAGTCATTTTCAATTCAAAATACTAGTTATTTTCAGTTTAAGAATCAGTCTGATTTCTTGA
GCATAGTGCTTATCTTATATGGTATGGCTTTCAGCTCCCTGGAGATTTCCTGGGTGTGCTG
CTGTGCCTTTGCCTCCAGGATAAATGCCATGCTGAGAAGAGCATGACTCTGTCAGCACAG
GCAGACGGTGGAAACTTTACCTCTGCGACCTCTCTGCAGCCTCAGGGTAGCAGGGGATGG
CATGGTTTCAT (SEQ ID NO: 27)

eHGT_767m (475 bp):
TGTATGCCTAAAGGGTTTGCTTCTATGTTAATTCCCAACAACAATCATGATACTTCCTTACTT
CTGCAGGACTAGGCTGGCTCACTAGTGAGGTTTTCTCCAAGGATGTGCTCCAATATCTGCT
CCACAATACTTAAATAGCTTCACAATACATAAATATTTAAGTACTCCTCTGTAAGGAGCGTTC
TGTTTCCATATGTCAAGGCAGAATCTTCACATAAAGCATGCATTTTGCTCCCTGTTTTCAAT
GGTTTTACTCTGTAGAGTTTCTCAAATTAATTCAATACCCACTTTAAATTAAATGGAAACTTT
AATGCTATGGTTTGCTGTGCCATTTAGATTTTATGTGTTGAGATGCAATTCCAACTCTACTT
GGCACGAAGTATAGACACTGAATCCAACTATGATGTATAAGCAAGGGTCCTTGTGAGGTAG
GTAATGAGAATGGAGCCTGGAACGTGTAACACTGGTGACT (SEQ ID NO: 28)

eHGT_768m (547 bp):
TGCAGATTATAATTCTATGGAAACTGACTGTTTTAAACATAAATTTTAGACCTCTTCTCAAGG
TGTGGCAATCCTCAACAGCTTCTGGTATGCATAGACCACATTTGTATTCACCATTAATGCAT
AAGAAAAACATTTTGGTTGCCAGTTTGGAGCTAAATTGATAAATCTGCTACAAGCATTTGTG
CATGAGTTTGTGTGTAAATATGCTTTTTACTTACTTAGGATGAATATACAAGGATAGCATTGT
TATATTGCTTGACATTTGCATATTCAATTTTATAGAGAATGCACAGCAATTTTGGTATCTGTC
TGTACCCTTTCATATTTTATGCTCAACATTTGAGTCACTTGGTTTCTTGGCATCATTATGAAT
TTTTAATGGTCTTACTTTTAAAAAATATGCTTTAACTCATCTTAGAATTATATCACATTTCCCC
TTCCAGTCTCTCCTAGACATCCCTCCCTCAAACAGCCCTTGTGTGCTCCTGCCTTCAGTTC
ATAGCCTGTTTTACTTTATTATTATTATTGCTCTCTGTGTGTGCAC (SEQ ID NO: 29)

eHGT_663m (621 bp):
GCTTCTGAATACCTCCACGCATACACTCACTTAAATTATAAATAAACCTGAAAAGCAATGTA
GCAAAAATGAAAGAGATTTCTGCCACCAGCTTCTATAATTATGTAATCATAATGTTTTGAAAG
TTAAACTCCCCTTTTCAACAAGAAAATTATAATATATATGCATATATGCATATATATTCACACA
CATAACATATATTTATAAATCAATTTAAACTGCCTTTGTGCAGTGTACTAGGCAAAAGTAATG
TATGTATCTATCTCTGTGAATGCCCTCTAATTCTGGAAATAAATAATGGGACATCTCAGTATT
TTCAAGAAGTCTGTAAGCCACAAATAGGCTTATGCCTATTGTTCTGTCATAATTATTAATGGT
CTAGTACTTGGTCTCAAGCATGTCTGGTATGGATGATAAAACTAAGAGTCCAAAAATAGGTC
CCCTGTGATCACTATCAGGAAACTCAAAAGCACAGAATTCATGGAGTATTTCTAACCAGAAA

FIG. 12A cont'd

CACAAACACTACACATACTCATACACATATGTGGATATCTGTTATATAAGATGTAACTTTATA
TAAAACCCCTAGTCTGGTAATGTTCTTAACCTTATCTGGTGACGTTACATGACACA (SEQ ID NO: 31)

eHGT_660m (312 bp):

TACTGTCCAGGCCAGCTTTGAACAGACTCTGTATTCCAGAAAGACTTTCAGCTTGTGCTTCT
TTTCCCCAAACCTGCCTTGAGTAGCTGACATCATAGGCTTGCACTACCAGGCCTGGCTTAA
AATACATACTCCCAATTTACATATCTGATGAGTAACATTCTCATAGAAGCACAAAAGTACATA
CATGGATTCTAATATAGGTTATTTTCTTGGGCATTACACACTACAGATATACTATTCAATATT
ATTTGTGGCCTACTAGAGACTGAGCCTAAGGCTATAAACATGTTAGTCGTGCTCTACCAGT
GA (SEQ ID NO: 32)

eHGT_665m (768 bp):

TGGAACTTGGGACTTCTGAGCACAGGGGGTGGATGGCTGGAAATGTGAGCAGAACCCATG
GTGTGAGAACTGAGATAAGTTTCTGCTGAGGCTAGCCTAGAGGCACAGGTGGTGAGACAG
CTAGGGCAGAACATATCAGTGAATAGTGTAGGATGATGCCACTTTTAAAATATTTCCCGCAA
TATGAACCAATTTCTCAAAAGGTATAAATTACCATAACTCTCTTAAATACAAAATGAGTAATC
TGAATAGCTGCAAATTGTAAGAAAATTGAACATAATCATAACCCATCGGCAAAAATAAATCT
TTGGGGAAATGCCACTTTATTAGAGAATTCTGCTATACATTTGATGAATATATTACAAAGCTT
AAAAAGCTCTGTGGGTGTTCATTTTGCATATTGAAAACTAAAATTTTATATAAGCTAAAAAAG
GCCAATTGTCTCTGTCTGTCTTTCTGTCTCTATCTCTGTCTCTCTCTGTCTCTCTGTCTCTCT
CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCCTTTCTTTTTCT
CTCTCTGTGTATCTAACTATTCTTCTTTTGCAATTTCTGAAATTTCTTCTCCAGGGCTTTCTTT
CTTCACAAGTGTAAGCACAGCATGTGTGCTCCCACTGCCAAGATGCTGTGCCATGCTATAA
CTTCTTTTTGACATCATATGTAGCTTATACATATTAAAATTTACTTCAGCACTCAACAGTCAC
CACTGTCATGACTGGAGGTT (SEQ ID NO: 33)

eHGT_700m (598 bp):

TTGGCTTTCCTTCCAAGTTGTATTTTGGTGAAGTTTGACTCCTGGAATCTCACTTCTGCACT
TTAGATTAGCATACTTTAAAAAGGAACATTACAGTACAATGGAAAAGTAAAGATCTGGGTTT
TATCTCAGCTTTAGTTAGTATTTATAGGAAGCATGGATTACATAAAGAGATGACATTATTTCA
GTCTGAAATTGATTTGTACTATACGACATTTCATTTTGTTTGTGCACATTGATTAGATAATTC
CTCAAATCAGTTTGAAAATTTCATCTCTGTTCATAGCTGAAGATAACCACCAAATTGCTTTTA
ACTAATTTACATAATAAGTATAGTAGAGGTTTATTCTATATAAAGTATCAAATATAAATATTCA
TCTCAAATTCTAAAGGATTTAGCAATGTAAATTACTTATGAATTACATACACATTTGTATTTAA
ATCTTTTTTCACTTTTATTAAAAATAGATCCTTTTCTCATACAATATATCCTGATTGCAGTTCC
CTCCCCTCTACTCCTCCCAGTTCCTTCCCACCTCCCTTTCCATCTGGATCTCTCTCCTATTA
AAGGTTGGACAGAAGAGATGGCACTGCATGT (SEQ ID NO: 34)

eHGT_456m (795 bp):

TGGTTAGCACTATTTTGAATGGACACTTCTATAAAATTTGTTTTAAAAACAATATGGAAGAGT
CATAAGATGTCTGGGCACAGTGCCCCCAGTCAACTCATCAATGGGATTCTGGAATCCAAAG
GTTATATGGTGTAAAAACACAGAGAGCAAAATGCACAAGGATCTTCATAGAACCATAATTTT
TAAAAAGCAAACACTTCCCCCAAAGGAGCATGTAAGACAACTTGAAGGTATTAAAATGGGA
TTGTGGTTTTATTAGTTAGAGAACATCATGCAAATTATATTATCTAGGCATTATAATGCAAAC
TTTAAAAGGAAATGCTCATTATGTAATGTTAAGTGAAAAAAATCAGGCCATAAAATGAAATGT

FIG. 12A cont'd

GCATTATGCATGCATATGATGAAAAGACAAGAAATAGATAAGAATATGCAGAAATACAAATA
GTTGGCTGTGGTGTGGGCTTCTGGCTTGTTTTCTGTTTTAAAGCCATTACATCATCTTTCCA
ATGACAAACTACATTTTAAGTTATTGAAAACAAAATCTCGTTAGCCTTGTCCTTGCATAAATT
CCCAGAGAACACATTATTCTTGTTCCCTTCTACCTCTGCTACTTGTCTTTGATTTGGACCTG
CTCCACCTCCTCTTACTTTTCAGTCCTGAACTAGATGGAAAAAAGAAATAGACCCTCAGGAA
ACAAGGAAGCAAAACCACCCAACTGTGTGGAAGCCGTGCATTCAAGCTGCAAACGAGTGG
AGTGGCTTCAATACCTTAACTTATCCCCATTGTTGCAAAGGGAGACCCAGA (SEQ ID NO:
35)

eHGT_698m (475 bp):

TGTATGCCTAAAGGGTTTGCTTCTATGTTAATTCCCAACAACAATCATGATACTTCCTTACTT
CTGCAGGACTAGGCTGGCTCACTAGTGAGGTTTTCTCCAAGGATGTGCTCCAATATCTGCT
CCACAATACTTAAATAGCTTCACAATACATAAATATTTAAGTACTCCTCTGTAAGGAGCGTTC
TGTTTCCATATGTCAAGGCAGAATCTTCACATAAAGCATGCATTTTGCTCCCTGTTTTCAAT
GGTTTTACTCTGTAGAGTTTCTCAAATTAATTCAATACCCACTTTAAATTAAATGGAAACTTT
AATGCTATGGTTTGCTGTGCCATTTAGATTTTATGTGTTGAGATGCAATTCCAACTCTACTT
GGCACGAAGTATAGACACTGAATCCAACTATGATGTATAAGCAAGGGTCCTTGTGAGGTAG
GTAATGAGAATGGAGCCTGGAACGTGTAACACTGGTGACT (SEQ ID NO: 36)

eHGT_699m (546 bp):

TGCAGATTATAATTCTATGGAACTGACTGTTTTAAACATAAATTTTAGACCTCTTCTCAAGGT
GTGGCAATCCTCAACAGCTTCTGGTATGCATAGACCACATTTGTATTCACCATTAATGCATA
AGAAAAACATTTTGGTTGCCAGTTTGGAGCTAAATTGATAAATCTGCTACAAGCATTTGTGC
ATGAGTTTGTGTGTAAATATGCTTTTTACTTACTTAGGATGAATATACAAGGATAGCATTGTT
ATATTGCTTGACATTTGCATATTCAATTTTATAGAGAATGCACAGCAATTTTGGTATCTGTCT
GTACCCTTTCATATTTTATGCTCAACATTTGAGTCACTTGGTTTCTTGGCATCATTATGAATT
TTTAATGGTCTTACTTTTAAAAAATATGCTTTAACTCATCTTAGAATTATATCACATTTCCCCT
TCCAGTCTCTCCTAGACATCCCTCCCTCAAACAGCCCTTGTGTGCTCCTGCCTTCAGTTCA
TAGCCTGTTTTACTTTATTATTATTATTGCTCTCTGTGTGTGCAC (SEQ ID NO: 37)

eHGT_647m (655 bp):

TCCACTTATACCCCATCAGCATCTCTATCATTGTCTAAGACATTTTTAATTTGAAATAAATGG
GGAAACTCATTATCATGGCCTCAAGGAAATACAGAAACCAAGGATTGTCCTTTTAGTGCAC
AGGTGTCTTCTAAATTCACATATTCTGGATTTAACAAGGAAAATAAAATCTGACATAGGCTG
AATTTGAAAATGGTAAGGATATTTTTAACTAAACAAAAAAAATAGGTAATAGAACATTTCATA
TTGAACTCCCAGGATAGCATTTATCATAATCCACCAGTTGAATTATTCATGCGCTGATAAAT
AACCTTAGCCTCCGGCACAGTAATCCTCTCACTGAATAATCACCAGGGCTCCCTGCCCGTG
GTGTTTTGCTTTACAACATAAATTTTAAAATTATTCTGTGTTGGCATACATATGTGAAAGCTC
ATTAACAGATACTGGCATTCTGATTGGTTTGGCACTAAATTTAAACTGTTTTGGAAAGAACT
GAAACACTGAAGGATATGCCATTCTGTTTCTTTATACCTTATCGATGTTTCCTTAACAATCG
GTTGAGATTTATGCATCTTTTGATAGATTTTATCTGAGATAGCTGATATGCATTGTGATGAGT
TCCTTTTTAACGCTGTAATGTTTCATTGGCCCT (SEQ ID NO: 38)

eHGT_648m (347 bp):

AGGCACCACTTTAAGGGCCTTTATCATTATCCACAACAGCAGCAGCAGCATGAGTAGCATT
AGGAATGCGTTTCCTCCTGTGCTTTGAGGCTTACGTAGGTCTTTCTTCTGCATCGTTTCATG

FIG. 12A cont'd

AAATTCTCACAAGAACACTATGAGGTAGGTGTCAGATGTCCATTTTACAGGACAGAAAAGA
GATTAAAAAAAAAGAGAGAGAGAGAGGAAAAAGGGCTTTATTCAAAGCCAAGTCAGTTCATA
TGGGCAGAGCCAGGACTTGAACGCAGGCCCCACAGTCTTAACTCCAATATTTTCAGCTGC
CTTGGGAACTTCTTGGAAGAGCCCATTCCATTTTCCTGCCAT (SEQ ID NO: 39)

eHGT_649m (412 bp):

GCGGCAGGAATTTAAAGAGGCAGAAGTTCATTTCCTGTGATTTATGGGACTTTGCTGACTT
TCACGCTCCGGTTCTAGGGACAGGGAGACCCTCCCTGCGGGAGCCGCTGCTAGAGGGAA
AGGCAGCGTTGGATCGGATGCCGGCCCCAGCCTGGGAGACGTGCTTCCTCTGCAGCCTG
TGATCAGGAGCGAAGTGTAGAGACCAGAATCGTTCAGAGGCAGGGCTGGTCCCTGTTGAT
TTATGACTCAGGCAAACAGTCCATTCATCCGTCAGCGGCTCCACTTCTATCCCCAGCTTAC
TGGAGGCAGAAGAATGTCCCCTGGAAGTGGTCTGAGCTGCTTCCTGTGGTGGCCCCATTC
CTCCGGGTGGGGGATCCTATTTTTAGAGGGCCACCATTGCCATGAAGGTCCA (SEQ ID NO:
40)

eHGT_670m (740 bp):

AGGACAACTTGACTTATGCAAATGAACAATATAAATTTTAGAACTATGTATAAATAATTAAAT
AAGTACATGGACTTACAGGAAGGAATGAATACCTTCAGAGTATGCTGTAGTATATTTCCTGA
ATGTATAGATTAAGCTATATAATTGTCAATGAGCTGTTGTATTACGTATAATTCAGGGTTTAA
ACAAAATAAAAATGTGTGCATATGTTTTATGCAAATGCATGTATTTTTGAATCAATGATAATC
CCACACTTTTTATAATGATTCATACCCGCTCAAAGTGATGAAAATAATTGACCTGTAAATGTT
TATTTTTAAGTAGAAAAAAATCTCTGTGAGGCTCAGAGAACATGTTGATACAGGGGGAGAG
GGAATGTACAGAGTAGCATGCTCTTGTAAGTTGCTCCAAGGCAACGCATGACTACTGCGGT
CTTGACCTAACAGCTGCACAGACCTTCAGAAGATTGTATCCATCTACATTTCATTATGGAGC
CAAGAATGGCTCATGAAACTTCTGCCCTCCCTGGGAGCATGTAGGCAGTTAATTTGAATGG
TAATTTTATGTACAAGGAGATAGAATATTTTCAGTGATGTCACCACTGGTTTGTCACAATGC
TTTCTATAAATGCCCAAATCCTGTAAGTATTCTTAATGAAAACCATTGACTTACGCACTAATA
GGAAATTAAAGTGGAAACAGATGTTCTATTTTTATTTATGCTGCTGTGATCAACAC (SEQ ID
NO: 41)

eHGT_697m (440 bp):

AACATGTATTTGTGGAGTGGCTAAATTGAGATATATCACCTTAAAATCAACTCCCAGCAATT
TGAAGAGCACAGTAAGTAAGACCTTGTTCCTAACCACAGTCGCCATGCAGTGCAGTAGAGC
CCAGATCTTACACTTCCTGTGTAACTGGAGTCTTGTGTTCTTTAGCCAAGCTCTTCCCACAT
GCGAGTCATTTTCAATTCAAAATACTAGTTATTTTCAGTTTAAGAATCAGTCTGATTTCTTGA
GCATAGTGCTTATCTTATATGGTATGGCTTTCAGCTCCCTGGAGATTTCCTGGGTGTGCTG
CTGTGCCTTTGCCTCCAGGATAAATGCCATGCTGAGAAGAGCATGACTCTGTCAGCACAG
GCAGACGGTGGAAACTTTACCTTTGCGACCTCTCTGCAGCCTCAGGGTAGCAGGGGATGG
CATGGTTTCAT (SEQ ID NO: 42)

eHGT_662m (390 bp):

TCTCAACACTAAGAAAACTGCAGTATGAAGATATGGAATTTGAGGCCAGTTTGAGATAAAAC
AGAAAGATATGAGCAAAAGACAAGGATGTGAGAAACATAAATAAAAACTATGCATATTCATG
AAGCACCGCATAATAGTTTCATATGGATACACATCCCATGATGTTTAAAGCAGGCTTATCAT
TTTTATCTTTTCAAACATTTAACACTTATGTTAATATATGTAAACTCAGTCATTTTTAGCCTTC
TGAGAGAACCTGCATCATTTTTAATCACTCTTCCGTTCAGTAGGATTCTTGAATTATTATCAT

FIG. 12A cont'd

AGTTAACCATAGCTGAGTAATCAAACCTCAATTGTAAATTTGAAAATTTACCCTTCTATTCTT
TTCTTGCAATAAAA (SEQ ID NO: 43)

eHGT_583m (711 bp):

CAACAAACCAGGGAATCTACTGTATAGTGAAGCAGGATTGTGAAAAGGCCATTCATATAAG
AGAACTCCAGATGGCTGACACAGATATGGTAGTGTTAAATCTTTCTGATAACCAAGAGGCT
CAAGTTAAACTGAGATTTCCCTTTATACACTGTGAATCTGTAAACATTACAACCTTCCAGAAT
ACCAAGTACTGTCTAGACTGAGACATAAGAACCTTGTTACAAGGTGTGTGGCTTTCTAGAAT
GAAGTCAGTTTACAATATTTGAAAAAAATTAGTATGCATACCCTAGACCTAGCAATGGCATC
TACTGTTACTTTATAGGTATGGGAACTGGTGACAGATAAGATTCAATGTGAAAAGAATGGAT
AACTAAAATATATTGGAAGCACAGTATGAAATAATATGCAGTGGGTAGGTGGAATGAGTGA
TGACGTATAACTCCATGCTAACTATAGTGTGGAGAGAATGCAAAGTACGCCAGAGAACAGG
ATGTATTGTTAAGTATGTAAATTAATTCCATGCACATAATAAGCTTACTTATTTTTTCAATGAC
AAATACATTTATATATTCAATGACATGCATGGAAATTTCTAGAATGTCTAGAGCGGAGGAGG
GGAAATGGATCTGAGTTCAGGAATGAATTAGCAAAAGAATAAAAAATGCACATGTGAAGGT
ATTAACATAAGGCCTTAGAAGCCTGTGTGTC (SEQ ID NO: 44)

Core of eHGT_459m enhancer (265 bp):

CTATGAATAGTTACAAGTGAGAAAGAGAAATGGAGATTTTACCATTTAATCATCTCCGAAGA
GAAGGTGAGAAGAAATTGAGAAAGGTAATTTGCATACCATGCATAGGCATTAGAGAAAGCA
TTTTTTTTCTTAACTCTCCATTTGTTTACAGATATTTCCATATGGCCTTTTAAAAACTTCAACA
GAGAAATTCGAACACAACTAACTGGAATTTCAATTTTCATATAAAACCAGCATCATAGATATA
TATGTAGCTACTCTC (SEQ ID NO: 45)

Core of eHGT_453m enhancer (251 bp):

ATTTTAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTAAAAATAACATAGATCTGAAA
AGCTTTATAACTTTTGGTTTTTTTTTTAAGGACCATATTAACAGCATCTTCCAGTTTTTCTAGA
GTAAATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTGTAAAAAGACTA
TAAAAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGA
(SEQ ID NO: 46)

Core of eHGT_462m enhancer (195 bp):

CTGTTCTATCTATTCCTGGGCTTCCACCACACTGTGTTAATGAGGATAACTTTAAAATACAT
CTTGATATTTAGTAGAGCGCATTTCCACGCCTGATTCATCACGAGCACCTTAACTTGTCTTG
GTTCTTCTGCGGAGATGTTGGCACCAGCTTCTTCACCCGTGACTCCAGCTTCTGGGGAAC
CTGTGCTCTCA (SEQ ID NO: 47)

Core of eHGT_461m enhancer (240 bp):

CCGCAGCTTTGGAAAAGACCTTCTTATTTTAACTTATGTAATTAACCATCCCAAATTTACCCT
GATCCTCACTGGAAATGTCTTCCCTTTCCTGATATTTACCACAGTTTTAAAAAGTCTTAGATG
GGTATGCATATTTGAGAAATACGTAGAGTTAATTAGTGCGTAGATGTGTTTTTAAGTTGCAT
AAAACATATTGTGCTTTGGCATAGCTCCCTTTGTAACTTATTTTAGGAAACC (SEQ ID NO:
48)

FIG. 12A cont'd

Core of eHGT_464m enhancer (227 bp):
TTTGTGTCTGCTATTGCATTTACATAGTGACATAGGCTTACAGGATATTGTAAATGTCATCC
CTAGATACATTTATGTAACCTTTTTTTAAAATAAAACTCTAACAAATGTTAGTTATTGTGCACT
ATGGTTTTGAATATTTGAATTTGCTTTGCTCAGATCTAAATAACTTCTGATAAGTTTGGCATT
CAGGAAAGAGTAAATGTAAATGTAGAAATGCTTGGAAT (SEQ ID NO: 49)

Core of eHGT_452m enhancer (235 bp):
CATTGACTCATTTTTATTTTGTGAGGAAGAAATTTTCCATTTGTCTGTGTAATTACATTTGCT
AATTTCTAAAACAATTCCTGGGATTTATATCTATATATGCCTCAGGAACATACAGGGAAGAT
CATACTGTTCTCTGAATTTCAAATTATGTTTTCTGTTTTATTTCATGAATTTATTATTTTTCTTT
GTCTCTGTCTGTCTCTCTCTGTGTGTATGTGTATGCGTGTATATG (SEQ ID NO: 50)

Core of eHGT_451m enhancer (213 bp):
TTTTCATTCCTGGAGGTAAAACCAACCTTAGAATCGTGGAGCCTGATAGACTTTGAAATTGG
CCTTCATTTTAATTCTCAAGTTCTTAGCATGAGAGCAACATTCAATATGGTTCTGAAATATCT
CATGCAAAAAAGTGCCAGAGTCATAGACTGAAATATTTATTTGAATATAAATGAACTCAGAT
TTGCCAAAATGCGCTCCCCCTGTTGT (SEQ ID NO: 51)

Core of eHGT_460m enhancer (278 bp):
CATGAACATTTAGTCATCGTTATCTTACATGAGTCATATCTGTGTCCTTATTGTCTTATCTTA
AATGCATTAATTAATTCGATAGAGCATCTATAAAAATGAAACGCTATTCAAAAGCTGGCCCA
TGACAAGATCAGTGTGTTGGTTTCTATGCCTATAGTGGTGCAATGGAAGGAGTGCGCTGTA
CCCCACAGACTTTGTGATTGATGAGTTGTCTACTTTATTCACAATGAAGCTAAAAACAATCT
AACAACCATCAGGGTGCTACAGCAGCACAG (SEQ ID NO: 52)

3xcore2_eHGT_459m (795 bp):
CTATGAATAGTTACAAGTGAGAAAGAGAAATGGAGATTTTACCATTTAATCATCTCCGAAGA
GAAGGTGAGAAGAAATTGAGAAAGGTAATTTGCATACCATGCATAGGCATTAGAGAAAGCA
TTTTTTTTCTTAACTCTCCATTTGTTTACAGATATTTCCATATGGCCTTTTAAAAACTTCAACA
GAGAAATTCGAACACAACTAACTGGAATTTCAATTTTCATATAAAACCAGCATCATAGATATA
TATGTAGCTACTCTCCTATGAATAGTTACAAGTGAGAAAGAGAAATGGAGATTTTACCATTT
AATCATCTCCGAAGAGAAGGTGAGAAGAAATTGAGAAAGGTAATTTGCATACCATGCATAG
GCATTAGAGAAAGCATTTTTTTTTCTTAACTCTCCATTTGTTTACAGATATTTCCATATGGCCT
TTTAAAAACTTCAACAGAGAAATTCGAACACAACTAACTGGAATTTCAATTTTCATATAAAAC
CAGCATCATAGATATATATGTAGCTACTCTCCTATGAATAGTTACAAGTGAGAAAGAGAAAT
GGAGATTTTACCATTTAATCATCTCCGAAGAGAAGGTGAGAAGAAATTGAGAAAGGTAATTT
GCATACCATGCATAGGCATTAGAGAAAGCATTTTTTTTTCTTAACTCTCCATTTGTTTACAGAT
ATTTCCATATGGCCTTTTAAAAACTTCAACAGAGAAATTCGAACACAACTAACTGGAATTTCA
ATTTTCATATAAAACCAGCATCATAGATATATATGTAGCTACTCTC (SEQ ID NO: 53)

3xcore2_eHGT_453m (749 bp):
ATTTTAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTTAAAAATAACATAGATCTGAAA
AGCTTTATAACTTTTGGTTTTTTTTTTTAAGGACCATATTAACAGCATCTTCCAGTTTTTCTAGA
GTAAATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTGTAAAAAGACTA
TAAAAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGAATTT
TAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTTAAAAATAACATAGATCTGAAAAGC

FIG. 12A cont'd

TTTATAACTTTTGGTTTTTTTTTTAAGGACCATATTAACAGCATCTTCCAGTTTTTCTAGAGTA
AATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTGTAAAAAGACTATAA
AAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTAAGGAATTTTA
GGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTAAAAATAACATAGATCTGAAAAGCTT
TATAACTTTTGGTTTTTTTTTTAAGGACCATATTAACAGCATCTTCCAGTTTTTCTAGAGTAAA
TGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTGTAAAAAGACTATAAAA
ATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGACTA (SEQ ID NO: 54)

3xcore2_eHGT_462m (585 bp):

CTGTTCTATCTATTCCTGGGCTTCCACCACACTGTGTTAATGAGGATAACTTTAAAATACAT
CTTGATATTTAGTAGAGCGCATTTCCACGCCTGATTCATCACGAGCACCTTAACTTGTCTTG
GTTCTTCTGCGGAGATGTTGGCACCAGCTTCTTCACCCGTGACTCCAGCTTCTGGGGAAC
CTGTGCTCTCACTGTTCTATCTATTCCTGGGCTTCCACCACACTGTGTTAATGAGGATAACT
TTAAAATACATCTTGATATTTAGTAGAGCGCATTTCCACGCCTGATTCATCACGAGCACCTT
AACTTGTCTTGGTTCTTCTGCGGAGATGTTGGCACCAGCTTCTTCACCCGTGACTCCAGCT
TCTGGGGAACCTGTGCTCTCACTGTTCTATCTATTCCTGGGCTTCCACCACACTGTGTTAAT
GAGGATAACTTTAAAATACATCTTGATATTTAGTAGAGCGCATTTCCACGCCTGATTCATCA
CGAGCACCTTAACTTGTCTTGGTTCTTCTGCGGAGATGTTGGCACCAGCTTCTTCACCCGT
GACTCCAGCTTCTGGGGAACCTGTGCTCTCA (SEQ ID NO: 55)

3xcore2_eHGT_461m (720 bp):

CCGCAGCTTTGGAAAAGACCTTCTTATTTTAACTTATGTAATTAACCATCCCAAATTTACCCT
GATCCTCACTGGAAATGTCTTCCCTTTCCTGATATTTACCACAGTTTTAAAAAGTCTTAGATG
GGTATGCATATTTGAGAAATACGTAGAGTTAATTAGTGCGTAGATGTGTTTTTAAGTTGCAT
AAAACATATTGTGCTTTGGCATAGCTCCCTTTGTAACTTATTTTAGGAAACCCCGCAGCTTT
GGAAAAGACCTTCTTATTTTAACTTATGTAATTAACCATCCCAAATTTACCCTGATCCTCACT
GGAAATGTCTTCCCTTTCCTGATATTTACCACAGTTTTAAAAAGTCTTAGATGGGTATGCAT
ATTTGAGAAATACGTAGAGTTAATTAGTGCGTAGATGTGTTTTTAAGTTGCATAAAACATATT
GTGCTTTGGCATAGCTCCCTTTGTAACTTATTTTAGGAAACCCCGCAGCTTTGGAAAAGAC
CTTCTTATTTTAACTTATGTAATTAACCATCCCAAATTTACCCTGATCCTCACTGGAAATGTC
TTCCCTTTCCTGATATTTACCACAGTTTTAAAAAGTCTTAGATGGGTATGCATATTTGAGAAA
TACGTAGAGTTAATTAGTGCGTAGATGTGTTTTTAAGTTGCATAAAACATATTGTGCTTTGG
CATAGCTCCCTTTGTAACTTATTTTAGGAAACC (SEQ ID NO: 56)

3xcore2_eHGT_464m (681 bp):

TTTGTGTCTGCTATTGCATTTACATAGTGACATAGGCTTACAGGATATTGTAAATGTCATCC
CTAGATACATTTATGTAACCTTTTTTTAAAATAAAACTCTAACAAATGTTAGTTATTGTGCACT
ATGGTTTTGAATATTTGAATTTGCTTTGCTCAGATCTAAATAACTTCTGATAAGTTTGGCATT
CAGGAAAGAGTAAATGTAAATGTAGAAATGCTTGGAATTTTGTGTCTGCTATTGCATTTACA
TAGTGACATAGGCTTACAGGATATTGTAAATGTCATCCCTAGATACATTTATGTAACCTTTTT
TTAAAATAAAACTCTAACAAATGTTAGTTATTGTGCACTATGGTTTTGAATATTTGAATTTGCT
TTGCTCAGATCTAAATAACTTCTGATAAGTTTGGCATTCAGGAAAGAGTAAATGTAAATGTA
GAAATGCTTGGAATTTTGTGTCTGCTATTGCATTTACATAGTGACATAGGCTTACAGGATAT
TGTAAATGTCATCCCTAGATACATTTATGTAACCTTTTTTTAAAATAAAACTCTAACAAATGTT
AGTTATTGTGCACTATGGTTTTGAATATTTGAATTTGCTTTGCTCAGATCTAAATAACTTCTG

FIG. 12A cont'd

ATAAGTTTGGCATTCAGGAAAGAGTAAATGTAAATGTAGAAATGCTTGGAAT (SEQ ID NO: 57)

3xcore2_eHGT_452m (705 bp):

CATTGACTCATTTTTATTTTGTGAGGAAGAAATTTTCCATTTGTCTGTGTAATTACATTTGCT
AATTTCTAAAACAATTCCTGGGATTTATATCTATATATGCCTCAGGAACATACAGGGAAGAT
CATACTGTTCTCTGAATTTCAAATTATGTTTTCTGTTTTATTTCATGAATTTATTATTTTTCTTT
GTCTCTGTCTGTCTCTCTCTGTGTGTATGTGTATGCGTGTATATGCATTGACTCATTTTTATT
TTGTGAGGAAGAAATTTTCCATTTGTCTGTGTAATTACATTTGCTAATTTCTAAAACAATTCC
TGGGATTTATATCTATATATGCCTCAGGAACATACAGGGAAGATCATACTGTTCTCTGAATT
TCAAATTATGTTTTCTGTTTTATTTCATGAATTTATTATTTTTCTTTGTCTCTGTCTGTCTCTCT
CTGTGTGTATGTGTATGCGTGTATATGCATTGACTCATTTTTATTTTGTGAGGAAGAAATTTT
CCATTTGTCTGTGTAATTACATTTGCTAATTTCTAAAACAATTCCTGGGATTTATATCTATATA
TGCCTCAGGAACATACAGGGAAGATCATACTGTTCTCTGAATTTCAAATTATGTTTTCTGTTT
TATTTCATGAATTTATTATTTTTCTTTGTCTCTGTCTGTCTCTCTCTGTGTGTATGTGTATGCG
TGTATATG (SEQ ID NO: 58)

3xcore2_eHGT_451m (639 bp):

TTTTCATTCCTGGAGGTAAAACCAACCTTAGAATCGTGGAGCCTGATAGACTTTGAAATTGG
CCTTCATTTTAATTCTCAAGTTCTTAGCATGAGAGCAACATTCAATATGGTTCTGAAATATCT
CATGCAAAAAAGTGCCAGAGTCATAGACTGAAATATTTATTTGAATATAAATGAACTCAGAT
TTGCCAAAATGCGCTCCCCCTGTTGTTTTTCATTCCTGGAGGTAAAACCAACCTTAGAATCG
TGGAGCCTGATAGACTTTGAAATTGGCCTTCATTTTAATTCTCAAGTTCTTAGCATGAGAGC
AACATTCAATATGGTTCTGAAATATCTCATGCAAAAAAGTGCCAGAGTCATAGACTGAAATA
TTTATTTGAATATAAATGAACTCAGATTTGCCAAAATGCGCTCCCCCTGTTGTTTTTCATTCC
TGGAGGTAAAACCAACCTTAGAATCGTGGAGCCTGATAGACTTTGAAATTGGCCTTCATTTT
AATTCTCAAGTTCTTAGCATGAGAGCAACATTCAATATGGTTCTGAAATATCTCATGCAAAA
AAGTGCCAGAGTCATAGACTGAAATATTTATTTGAATATAAATGAACTCAGATTTGCCAAAAT
GCGCTCCCCCTGTTGT (SEQ ID NO: 59)

3xcore2_eHGT_460m (834 bp):

CATGAACATTTAGTCATCGTTATCTTACATGAGTCATATCTGTGTCCTTATTGTCTTATCTTA
AATGCATTAATTAATTCGATAGAGCATCTATAAAAATGAAACGCTATTCAAAAGCTGGCCCA
TGACAAGATCAGTGTGTTGGTTTCTATGCCTATAGTGGTGCAATGGAAGGAGTGCGCTGTA
CCCCACAGACTTTGTGATTGATGAGTTGTCTACTTTATTCACAATGAAGCTAAAAACAATCT
AACAACCATCAGGGTGCTACAGCAGCACAGCATGAACATTTAGTCATCGTTATCTTACATG
AGTCATATCTGTGTCCTTATTGTCTTATCTTAAATGCATTAATTAATTCGATAGAGCATCTAT
AAAAATGAAACGCTATTCAAAAGCTGGCCCATGACAAGATCAGTGTGTTGGTTTCTATGCCT
ATAGTGGTGCAATGGAAGGAGTGCGCTGTACCCCACAGACTTTGTGATTGATGAGTTGTCT
ACTTTATTCACAATGAAGCTAAAAACAATCTAACAACCATCAGGGTGCTACAGCAGCACAG
CATGAACATTTAGTCATCGTTATCTTACATGAGTCATATCTGTGTCCTTATTGTCTTATCTTA
AATGCATTAATTAATTCGATAGAGCATCTATAAAAATGAAACGCTATTCAAAAGCTGGCCCA
TGACAAGATCAGTGTGTTGGTTTCTATGCCTATAGTGGTGCAATGGAAGGAGTGCGCTGTA
CCCCACAGACTTTGTGATTGATGAGTTGTCTACTTTATTCACAATGAAGCTAAAAACAATCT
AACAACCATCAGGGTGCTACAGCAGCACAG (SEQ ID NO: 60)

FIG. 12B

Beta-Globin Minimal Promoter (pBGmin/minBGlobin/minBGprom):
GGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTG (SEQ ID NO: 61)

minCMV Promoter:
GAGGTAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGA TCGCCTGG (SEQ ID NO: 62)

Mutated minCMV Promoter (SacI RE site removed):
GAGGTAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGA TCGCCTGG (SEQ ID NO: 63)

minRho Promoter:
GATTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCCTGGGGGGGGAGTTG GAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGAGCTCCGGCCTCAGAAGCATCC CC (SEQ ID NO: 64)

minRho* Promoter:
GATTCAGCCGGGAGCTTAGGGAGGGGAGGTCACTTCATAAGGGCTTGGGGGGGGAGTTG GAGCCACGAGTCGTCCAGCCGGAGCCCCGTGTGGCTGTGCTCCGGCCTCAGAAGCATCC CC (SEQ ID NO: 65)

Hsp68 minimal Promoter (proHsp68):
CAGGAACATCCAAACTGAGCAGCCGGGGTCCCCCCCACCCCCCACCCCGCCCCACGCGG CAACTTTGAGCCTGTGCTGGGACAGAGCCTCTAGTTCCTAAATTAGTCCATGAGGTCAGAG GCAGCACTGCCATTGTAACGCGATTGGAGAGGATCACGTCACCGGACACGCCCCCAGGC ATCTCCCTGGGTCTCCTAAACTTGGCGGGGAGAAGTTTTAGCCCTTAAGTTTTAGCCTTTAA CCCCCATATTCAGAACTGTGCGAGTTGGCGAAACCCCACAAATCACAACAAACTGTACACA ACACCGAGCTAGAGGTGATCTTTCTTGTCCATTCCACACAGGCCTTAGTAATGCGTCGCCA TAGCAACAGTGTCACTAGTAGCACCAGCACTTCCCCACACCCTCCCCCTCAGGAATCCGTA CTCTCCAGTGAACCCCAGAAACCTCTGGAGAGTTCTGGACAAGGGCGGAACCCACAACTC CGATTACTCAAGGGAGGCGGGGAAGCTCCACCAGACGCGAAACTGCTGGAAGATTCCTG GCCCCAAGGCCTCCTCCGGCTCGCTGATTGGCCCAGCGGAGAGTGGGCGGGGCCGGTG AAGACTCCTTAAAGGCGCAGGGCGGCGAGCAGGTCACCAGACGCTGACAGCTACTCAGA ACCAAATCTGGTTCCATCCAGAGACAAGCGAAGACAAGAGAAGCAGAGCGAGCGGCGCGT TCCCGATCCTCGGCCAGGACCAGCCTTCCCCAGAGCATCCCTGCCGCGGAGCGCAACCT TCCCAGGAGCATCCCTGCCGCGGAGCGCAACTTTCCCCGGAGCATCCACGCCGCGGAGC GCAGCCTTCCAGAAGCAGAGCGCGGCGCC (SEQ ID NO: 66)

FIG. 12B cont'd

SYFP2:

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCAT
CTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAG
AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
ACAAGTAA (SEQ ID NO: 67)

EGFP:

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCA
CCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA
AGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCT
TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAA
GAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCT
CGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACA
ACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACA
TGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAA (SEQ ID NO: 68)

Optimized Flp recombinase (FlpO):

ATGGCTCCTAAGAAGAAGAGGAAGGTGATGAGCCAGTTCGACATCCTGTGCAAGACCCCC
CCCAAGGTGCTGGTGCGGCAGTTCGTGGAGAGATTCGAGAGGCCCAGCGGCGAGAAGAT
CGCCAGCTGTGCCGCCGAGCTGACCTACCTGTGCTGGATGATCACCCACAACGGCACCG
CCATCAAGAGGGCCACCTTCATGAGCTACAACACCATCATCAGCAACAGCCTGAGCTTCGA
CATCGTGAACAAGAGCCTGCAGTTCAAGTACAAGACCCAGAAGGCCACCATCCTGGAGGC
CAGCCTGAAGAAGCTGATCCCCGCCTGGGAGTTCACCATCATCCCTTACAACGGCCAGAA
GCACCAGAGCGACATCACCGACATCGTGTCCAGCCTGCAGCTGCAGTTCGAGAGCAGCG
AGGAGGCCGACAAGGGCAACAGCCACAGCAAGAAGATGCTGAAGGCCCTGCTGTCCGAG
GGCGAGAGCATCTGGGAGATCACCGAGAAGATCCTGAACAGCTTCGAGTACACCAGCAGG
TTCACCAAGACCAAGACCCTGTACCAGTTCCTGTTCCTGGCCACATTCATCAACTGCGGCA
GGTTCAGCGACATCAAGAACGTGGACCCCAAGAGCTTCAAGCTGGTGCAGAACAAGTACC
TGGGCGTGATCATTCAGTGCCTGGTGACCGAGACCAAGACAAGCGTGTCCAGGCACATCT
ACTTTTTCAGCGCCAGAGGCAGGATCGACCCCCTGGTGTACCTGGACGAGTTCCTGAGGA
ACAGCGAGCCCGTGCTGAAGAGAGTGAACAGGACCGGCAACAGCAGCAGCAACAAGCAG
GAGTACCAGCTGCTGAAGGACAACCTGGTGCGCAGCTACAACAAGGCCCTGAAGAAGAAC
GCCCCCTACCCCATCTTCGCTATCAAGAACGGCCCTAAGAGCCACATCGGCAGGCACCTG
ATGACCAGCTTTCTGAGCATGAAGGGCCTGACCGAGCTGACAAACGTGGTGGGCAACTGG
AGCGACAAGAGGGCCTCCGCCGTGGCCAGGACCACCTACACCCACCAGATCACCGCCAT

FIG. 12B cont'd

CCCCGACCACTACTTCGCCCTGGTGTCCAGGTACTACGCCTACGACCCCATCAGCAAGGA
GATGATCGCCCTGAAGGACGAGACCAACCCCATCGAGGAGTGGCAGCACATCGAGCAGC
TGAAGGGCAGCGCCGAGGGCAGCATCAGATACCCCGCCTGGAACGGCATCATCAGCCAG
GAGGTGCTGGACTACCTGAGCAGCTACATCAACAGGCGGATCTGA (SEQ ID NO: 69)

Improved Cre recombinase (iCre):

ATGGTGCCCAAGAAGAAGAGGAAAGTCTCCAACCTGCTGACTGTGCACCAAAACCTGCCT
GCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAGGAAGAACCTGATGGACATGTTCAGG
GACAGGCAGGCCTTCTCTGAACACACCTGGAAGATGCTCCTGTCTGTGTGCAGATCCTGG
GCTGCCTGGTGCAAGCTGAACAACAGGAAATGGTTCCCTGCTGAACCTGAGGATGTGAGG
GACTACCTCCTGTACCTGCAAGCCAGAGGCCTGGCTGTGAAGACCATCCAACAGCACCTG
GGCCAGCTCAACATGCTGCACAGGAGATCTGGCCTGCCTCGCCCTTCTGACTCCAATGCT
GTGTCCCTGGTGATGAGGAGAATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAA
GCAGGCCCTGGCCTTTGAACGCACTGACTTTGACCAAGTCAGATCCCTGATGGAGAACTC
TGACAGATGCCAGGACATCAGGAACCTGGCCTTCCTGGGCATTGCCTACAACACCCTGCT
GCGCATTGCCGAAATTGCCAGAATCAGAGTGAAGGACATCTCCCGCACCGATGGTGGGAG
AATGCTGATCCACATTGGCAGGACCAAGACCCTGGTGTCCACAGCTGGTGTGGAGAAGGC
CCTGTCCCTGGGGGTTACCAAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGCTGA
TGACCCCAACAACTACCTGTTCTGCCGGGTCAGAAAGAATGGTGTGGCTGCCCCTTCTGC
CACCTCCCAACTGTCCACCCGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGAT
CTATGGTGCCAAGGATGACTCTGGGCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAG
AGTGGGTGCTGCCAGGGACATGGCCAGGGCTGGTGTGTCCATCCCTGAAATCATGCAGG
CTGGTGGCTGGACCAATGTGAACATTGTGATGAACTACATCAGAAACCTGGACTCTGAGAC
TGGGGCCATGGTGAGGCTGCTCGAGGATGGGGACTAA (SEQ ID NO: 70)

SP10 insulator (SP10ins):

GAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAG (SEQ ID NO: 71)

3xSP10ins:

GAAGCTACCCCTAACACACTATTCTACACACAGAAAATGCTCTTCACTAGGAAGCTACCCC
TAACACACTATTCTACACACAGAAAATGCTCTTCACTAGGAAGCTACCCCTAACACACTATT
CTACACACAGAAAATGCTCTTCACTAG (SEQ ID NO: 72)

WPRE3:

ATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTC
CTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATG
GCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGC
CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGG
(SEQ ID NO: 73)

WPRE:

GCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTA
TGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAG
TTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCC
ACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCC
CTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGC
TCGCCTATGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCC

FIG. 12B cont'd

TCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTC
TTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCATCGATACC
G (SEQ ID NO: 74)

BGHpA:
CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGAC
CCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGT
CTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGA
TTGGGAAGACAATAGCAGGCATG (SEQ ID NO: 75)

HGHpA:
ACGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTC
CAGTGCCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCC
TTCTATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCAAGTTGGGAAGACA
ACCTGTAGGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCAGTGGCACAATCTTGG
CTCACTGCAATCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGT
TGGGATTCCAGGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGG
TTTCACCATATTGGCCAGGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGG
CCTCCCAAATTGCTGGGATTACAGGCGTGAACCACTGCTCCCTTCCCTGTCCTT (SEQ ID
NO: 76)

P2A:
GGCAGCGGCGCCACCAACTTCAGCCTGCTGAAGCAGGCCGGCGACGTGGAGGAGAACCC
CGGCCCCGGAGCTAGCGGA (SEQ ID NO: 77)

T2A:
(GSG)EGRGSLLTCGDVEENPGP (SEQ ID NO: 78)

E2A:
(GSG)QCTNYALLKLAGDVESNPGPP (SEQ ID NO: 79)

F2A:
(GSG)VKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 80)

Exemplary Plasmid Backbone 1 – Left ITR:
CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTCGGGCGACCTT
TGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCA
CTAGGGGTTCCT (SEQ ID NO: 81)

Exemplary Plasmid Backbone 1 – Right ITR:
CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCG
TCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTG
GCCAACTCCATCACTAGGGGTTCCT (SEQ ID NO: 82)

Exemplary Plasmid Backbone 2 – Left ITR:
CATGTCCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCG
GGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGG
AGTGGCCAACTCCATCACTAGGGGTTCCT (SEQ ID NO: 83)

FIG. 12B cont'd

Exemplary Plasmid Backbone 2 – Right ITR:
AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGG
CCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGA
GCGAGCGCGCAGCTGCCTGCAGGGGCGCCTG (SEQ ID NO: 84)

PHP.eB capsid:
MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDK
GEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQERLKEDTSFGGNLGRAVFQAKK
RLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDP
QPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVITTSTRT
WALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWG
FRPKRLNFKLFNIQVKEVTDNNGVKTIANNLTSTVQVFTDSDYQLPYVLGSAHEGCLPPFPADV
FMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLD
RLMNPLIDQYLYYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
NNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK
VMITNEEEIKTTNPVATESYGQVATNHQSDGTLAVPFKAQAQTGWVQNQGILPGMVWQDRDV
YLQGPIWAKIPHTDGNFHPSPLMGGFGMKHPPPQILIKNTPVPADPPTAFNKDKLNSFITQYST
GQVSVEIEWELQKENSKRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL
(SEQ ID NO: 85)

AAV9 VP1 capsid protein:
MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDK
GEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQERLKEDTSFGGNLGRAVFQAKK
RLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTESVPDP
QPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVITTSTRT
WALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWG
FRPKRLNFKLFNIQVKEVTDNNGVKTIANNLTSTVQVFTDSDYQLPYVLGSAHEGCLPPFPADV
FMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLD
RLMNPLIDQYLYYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
NNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK
VMITNEEEIKTTNPVATESYGQVATNHQSAQAQAQTGWVQNQGILPGMVWQDRDVYLQGPIW
AKIPHTDGNFHPSPLMGGFGMKHPPPQILIKNTPVPADPPTAFNKDKLNSFITQYSTGQVSVEIE
WELQKENSKRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLTRNL (SEQ ID NO:
86)

tet-Transactivator version 2 (tTA2):
ATGTCTAGACTGGACAAGAGCAAAGTCATAAACTCTGCTCTGGAATTACTCAATGAAGTCG
GTATCGAAGGCCTGACGACAAGGAAACTCGCTCAAAAGCTGGGAGTTGAGCAGCCTACCC
TGTACTGGCACGTGAAGAACAAGCGGGCCCTGCTCGATGCCCTGGCAATCGAGATGCTGG
ACAGGCATCATACCCACTTCTGCCCCCTGGAAGGCGAGTCATGGCAAGACTTTCTGCGGA
ACAACGCCAAGTCATTCCGCTGTGCTCTCCTCTCACATCGCGACGGGGCTAAAGTGCATCT
CGGCACCCGCCCAACAGAGAAACAGTACGAAACCCTGGAAAATCAGCTCGCGTTCCTGTG
TCAGCAAGGCTTCTCCCTGGAGAACGCACTGTACGCTCTGTCCGCCGTGGGCCACTTTAC
ACTGGGCTGCGTATTGGAGGATCAGGAGCATCAAGTAGCAAAAGAGGAAAGAGAGACACC
TACCACCGATTCTATGCCCCCACTTCTGAGACAAGCAATTGAGCTGTTCGACCATCAGGGA
GCCGAACCTGCCTTCCTTTTCGGCCTGGAACTAATCATATGTGGCCTGGAGAAACAGCTAA
AGTGCGAAAGCGGCGGGCCGGCCGACGCCCTTGACGATTTTGACTTAGACATGCTCCCAG
CCGATGCCCTTGACGACTTTGACCTTGATATGCTGCCTGCTGACGCTCTTGACGATTTTGA
CCTTGACATGCTCCCCGGGTAA (SEQ ID NO: 87)

FIG. 12B cont'd

GTPase HRas [Homo sapiens]:
MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYS
AMRDQYMRTGEGFLCVFAINNTKSFEDIHQYREQIKRVKDSDDVPMVLVGNKCDLAARTVESR
QAQDLARSYGIPYIETSAKTRQGVEDAFYTLVREIRQHKLRKLNPPDESGPGCMSCKCVLS
(SEQ ID NO: 88)

GCaMP6m
ATGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCTTTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCCTGACAATGATGGCAAGAAAAGGGA
GCTACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 89)

GCaMP6s
ATGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCCACATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGCTTTCTC

FIG. 12B cont'd

CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCCTGACAATGATGGCAAGAAAAATGA
AATACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 90)

GCaMP6f

ATGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAATGG
GTCGGGATCTGTACGACGATGACGATAAGGATCTCGCCACCATGGTCGACTCATCACGTC
GTAAGTGGAATAAGACAGGTCACGCAGTCAGAGCTATAGGTCGGCTGAGCTCACTCGAGA
ACGTCTATATCAAGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCTACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCGTGCAGTCCAAACTTTCG
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGGGCGGTACCGGAGGGAGCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGTGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACATCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTGCCGGACCAACTGACTGAAGAGCAGATCGCAGAATTTAAAGAGGAATTCTC
CCTATTTGACAAGGACGGGGATGGGACAATAACAACCAAGGAGCTGGGGACGGTGATGC
GGTCTCTGGGGCAGAACCCCACAGAAGCAGAGCTGCAGGACATGATCAATGAAGTAGATG
CCGACGGTGACGGCACAATCGACTTCCCTGAGTTCCTGACAATGATGGCAAGAAAAATGA
AATACAGGGACACGGAAGAAGAAATTAGAGAAGCGTTCGGTGTGTTTGATAAGGATGGCA
ATGGCTACATCAGTGCAGCAGAGCTTCGCCACGTGATGACAAACCTTGGAGAGAAGTTAA
CAGATGAAGAGGTTGATGAAATGATCAGGGAAGCAGACATCGATGGGGATGGTCAGGTAA
ACTACGAAGAGTTTGTACAAATGATGACAGCGAAGTGA (SEQ ID NO: 91)

FIG. 12C

CN2248 (1842 bp between ITRs):

GCGGCCGCACGCGTTTCCGCGTAGACCTGGGGTACTTTTCTAAGACACCCTTTATTGTTTG
TTCCTAACCCTAGCCCTGTATCCATGCAGCGTTCTCCATCACCTATGTAAAACTTAGCAGAT
CTGAGTCTTCCTCTGAGGAAACAACATCCTGTGCTGTTAGACTGTATTTTCCGTCTCCAGAT
TCTCTACTTCGTCCCTTCCCCCGCCCCAATTTCCCTGTCTCCAGCGAGCAATATTATCTTTC
ATCTGAGCCTCTCTCCCTGCCTCCAGTGACTGCCTCCCCTGATGAATTCTCTCCGGGTGAC
TTTGGCATTTGCATATTTCCTCGCTTGGATATTCTGAGATGTTGGAACCGGAGAGCAGGTT
GACGCGATCAGGTACTTAGCTGAACGAGGCTTTCTCGGGAACTGCACAAATATTTGGAGTC
GGGCATGTGCCAATAGGCAATTAATTGAAATGTGGACACAGCAAGCTTCCGGAGGAGGGA
AGCTGAAGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTT
GCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGG
GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC
CCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCA
CCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGAC
TTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAA
GGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG
AGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG
GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGAC
GGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAA
GATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATG
CCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGG
TTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCT
CGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCC
CACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTA
TTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG
GCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 92)

CN2249 (1973 bp between ITRs):

GCGGCCGCACGCGTGCTAAAGACAACATTGGCCTGAAATTCAGGAGACCTAAGTTCTAGA
TCAATCTGTGGCATGGTCTTAGACCACAGATGGGCTTTCCTGCCTGCATTTACCCAAAGAG
CCCAAAGAGGAGACCACATGGGGTACCATGGGACTGACCTTCCCAGTTCAAAGCTTCTTTA
CACAACGCTGCTGAATCACCCGCTAGCTCGCTCTCTTATCATTCCTTTTCATTCCTGGAGGT
AAAACCAACCTTAGAATCGTGGAGCCTGATAGACTTTGAAATTGGCCTTCATTTTAATTCTC
AAGTTCTTAGCATGAGAGCAACATTCAATATGGTTCTGAAATATCTCATGCAAAAAAGTGCC
AGAGTCATAGACTGAAATATTTATTTGAATATAAATGAACTCAGATTTGCCAAAATGCGCTC
CCCCTGTTGTCATGGACACAGATGCCTGCCTTTCTGATTTGCATTTTGCTCTAATGGGTATA
ATGTGTTTCAGCTCCTCATTGGTGCTGGGATTGGATTTGAGAAACTCACCAGTCCATGAGT
TCCCCCAAAATGTCAGCCAGTGGGAAACTTTTGTGTCTTTGATTCTGTTTCATCTAGCCAGC
TGGAAACAGACTGCGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT
ACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAG
CAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACG
TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT
GACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGC

FIG. 12C cont'd

ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGA
ACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGC
ATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
ACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAG
TCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTG
TGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTT
TAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAAT
CCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAG
GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCT
AGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCC
ACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCA
TTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA
GCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 93)

CN2250 (2027 bp between ITRs):

GCGGCCGCACGCGTTGCTCAGAAACTGGCCTCGATGCTGTCCTGCTTTCACCAATTAATAA
TATATGTGGAGACCGTTCTCTACTGGCATATGTGTGTGCATCACATAATTATATGTATATAT
GCAATCCTTTTTCTTGTTGTCTTTTAGAGTCAATTCTATATATATATTTTTTATTAAGACATTT
AAAAGTTCCAAAGTAAAAATAACAAAATCAAGGATACATTGACTCATTTTTATTTTGTGAGGA
AGAAATTTTCCATTTGTCTGTGTAATTACATTTGCTAATTTCTAAAACAATTCCTGGGATTTAT
ATCTATATATGCCTCAGGAACATACAGGGAAGATCATACTGTTCTCTGAATTTCAAATTATG
TTTTCTGTTTTATTTCATGAATTTATTATTTTTCTTTGTCTCTGTCTGTCTCTCTCTGTGTGTA
TGTGTATGCGTGTATATGCACGCATGTGCTGTGCACTTGTGTGTATACATTTTTTTGCACGT
GAGAAGTGTACCTACGAGGCTGTGCCTGGGTGTGGAAGCCCAGGGTTGGTGGGGAAGAC
CATACTCCAAGGCTCTCCAACCTTATTCAGTGAGGCAGGCAGTCACATCTGTCTGACATTT
CTGATCATTACACGTGTGCAGTGGGTGTTTTATCCACCGAGCCACCTTCCCAGCCATGGAT
GAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGG
GATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGA
GCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACA
AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
CTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGG
CTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAA
GTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAA
CTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACT
ACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACT
TCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAG
AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCA
GTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT
GACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGC
CGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACT
GGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTAT
CATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTT
GCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTT
GGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCAT

FIG. 12C cont'd

CTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCT
TTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGG
GTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATC
TCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 94)

CN2251 (2079 bp between ITRs):

GCGGCCGCACGCGTTCACAAATCAGTAAGAATTACACTCCAAAAATAAAACAGGGAAGGGT
ATTAGTATGTGATACGCATGGATAAAGTAGAAGAGGCCAAAAATACAAAGAACCGTAAACT
CACTAGTGATGGAAGAAACGCTAATTAAAATAATGAGCTTTCCTGACCCTTTATCACATGGG
CACATTTTCAAAGGACAAGACTCTGCTGAGGAGAACACAGAACAGGATCCTTCAGACAAGC
TTCAGCAACCAATTTTAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTAAAAATAACA
TAGATCTGAAAAGCTTTATAACTTTTGGTTTTTTTTTAAGGACCATATTAACAGCATCTTCCA
GTTTTTCTAGAGTAAATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTG
TAAAAAGACTATAAAAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGA
CTAAGGCCTTCAGCTCAGCAGGGGGCAGAGTTTATCATCTGTACTCTCTGAAGTGTCACAC
TGCCCTGCCTTTGTGTGCCCTTGCCTGTCCCTCTGTGACGGTCACCCTTTCGGTTCTGGCT
TGGAGCCTGGTCATCAGTGCCTGGCGGCCTGTGCCTGGGCTGTGCTGAGCTCATTTGGTC
GTTGTCCTAGTGATGTGACCCTACCTTAGACTTCCCCATCTCTTGAGGTGGCTAGAGCTCG
GGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGA
TCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCAC
CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGC
ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGT
GCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT
GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC
CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCC
ACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCC
GCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTG
AGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGC
GAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTA
ACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATT
GCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGC
GGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTG
ACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATA
AAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG
GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 95)

CN2252 (1830 bp between ITRs):

GCGGCCGCACGCGTACGGATGTCAAGTTAGGCAAAAGTTTCATATAAATGTGGTTAAAGTA
ATTCTAAAGTTTGTTTTTACATCACAGAACTGAGCAAATGAACAAATGTAATGATAACATTGG
GAGCTGATTATTCTGAGCAAACAGCAGAGAAATCCAAATCCAGACAGTTGGGGGGAAAGC
CACAGCTACTCAGAAAACACACTGTGATGGAAAACAACCTACAGGTTATTATTAGACTAAAG
AGACATAAAATTAGATATGCCACTTCCAGAGACTATAGGTTAGACGATAATATATGCTGAGT
GAAGTGCCCTGTTTTGGTAACTGCAGTAAGGCTGTGTATCAGAATGTTCCTAGCTACATATT

FIG. 12C cont'd

TAGGTGTAAAGCAGAGTGAGTCAAAATATATTCCAGGGTTAAAAAAAAAAAAAACAATCTTT
ACTTCTCTTTTCTTTAAACTAAATAATTTCAATTATTCTCCTATGTTTAGGGAGGAGCTCGGG
CTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATC
TTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACC
GGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGT
GTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCA
CCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTG
CAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATG
CCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC
CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCAT
CGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCA
CAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCG
CCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA
TCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGA
GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCC
GGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCG
AATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAA
CTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG
CTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCG
GAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGA
CAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTG
CCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAA
AATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG
GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 96)

CN2253 (1733 bp between ITRs):

GCGGCCGCACGCGTGACTGGAAAAAGGTAGCTATGGCAGCCTCTGCGCCTATATTTTATC
CATATTCTGATTTTGTTTTCCAAGGTCTCCTAGTCAATCATAATCTGAATCAATGGAAATTTC
CACGGATAAATAATTCATGAGACTTGAGGAAAATGTGTGCTAGTCCTCAATACCTTATAAAA
TGTATTATTCATTTTGAATTAGATTAAAATGTTTAAAAAATGCTATCTATAAGACACTGTGGG
AGCCAAAAAATTTATACTTCTTTATAAAATCCTTAACTGCCTGGAAATAAACTAGGTGTTAAG
AGAATTGAAGTTAAAACACAGAAAAATAATCCCACTCTTTCATAAACTCACAACTCCCAAATA
ATGAGAACCAAAGGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCT
TACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGA
GCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAA
GCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCG
TGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAG
CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC
AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGT
GAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACA
AGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACG
GCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCC
GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCA
CTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGT
CCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTA
AGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATT
TGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGC

FIG. 12C cont'd

TTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAA
ATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGAC
AGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTT
CTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGC
CACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAAT
AGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 97)

CN2254 (2145 bp between ITRs):

GCGGCCGCACGCGTTGACTGTGTCATGACTCTCAGTGTCTCTTTTTTCCACGGTTCTGGGA
TCATCATTTCTTTCTCTTCAGACAGTCTAGATGAGCAGCCGAGTGCTTATCTTTAAGGCTGA
CTCTAGACAGCCAGTTGTTATTTCATAAATCTGTTAAAACTCACCGTGCAAACACCAGCGTT
CCTTAGCTGACCATCCTCAAGCACACGTGGAGTGTGACCTCTGCTTCCTTCTCTCATTTCA
CATATTAGTAATGTGCGTCTATGAATAGTTACAAGTGAGAAAGAGAAATGGAGATTTTACCA
TTTAATCATCTCCGAAGAGAAGGTGAGAAGAAATTGAGAAAGGTAATTTGCATACCATGCAT
AGGCATTAGAGAAAGCATTTTTTTTCTTAACTCTCCATTTGTTTACAGATATTTCCATATGGC
CTTTTAAAAACTTCAACAGAGAAATTCGAACACAACTAACTGGAATTTCAATTTTCATATAAA
ACCAGCATCATAGATATATATGTAGCTACTCTCCCACTTCACATTATTCAGTTAGTATATTTT
CTAAGCTAATTTCCAATCCATTTTAAGACTATGTACAGCTCATATGAGGCCCCCATATGTTC
GTTCACGTTACTTTGCATCCGTTTACATGATTAAACCCCAAGAACAAATCGATTAGAAGCCA
AGAGCAAATGCTTTTGGCTCCACATTAGAAAACTCAACAGAGAAACCTTTACAAGTGATAGT
CATCTGGGTAAGAGACATGCAAACATAATTACCATTATTCGTGGACTTAATTCCCTGAGCTC
GGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAG
ATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCA
CCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGC
GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTG
CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCG
TGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCA
TGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA
CCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC
ATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGC
CACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATC
CGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCC
CATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCT
GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCG
CCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCG
CGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTT
AACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATT
GCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGC
GGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTG
ACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATA
AAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG
GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 98)

CN2255 (2117 bp between ITRs):

GCGGCCGCACGCGTTTAACTTGAGGAAATCTGCCACAAGATGTTCTCAAGTGCAAATGTCT
CTTGATGCATACACGCATGCATTTCTATTTAAGTATCCAACGAGAATTGTGGGGTTGAGGTA

FIG. 12C cont'd

GGTATTTTAGCTTATAAAAAAAAAATACCTTATGAGCTTATTTTGGATCATATAAAAGTATGA
GCACTTTTAGAACACTTTTAATAATGCAAATCATCTCTGCATCATCCAAATTTAATCCATGAA
CATTTAGTCATCGTTATCTTACATGAGTCATATCTGTGTCCTTATTGTCTTATCTTAAATGCA
TTAATTAATTCGATAGAGCATCTATAAAAATGAAACGCTATTCAAAAGCTGGCCCATGACAA
GATCAGTGTGTTGGTTTCTATGCCTATAGTGGTGCAATGGAAGGAGTGCGCTGTACCCCAC
AGACTTTGTGATTGATGAGTTGTCTACTTTATTCACAATGAAGCTAAAAACAATCTAACAACC
ATCAGGGTGCTACAGCAGCACAGAAATGACTAGCTCTTTTGTGTCAAGTTATTGCCGGTTG
TCTTAGTCACTGTCCGACTGATGTGCAGAGACACCATGACTAAGGCAACGCTTATCAAAGA
AAGAGTCTAGTTGAGAGCTTGCTTGTAGTTCCAGAGGTTTGTTTAGTCTAGCATGGTGGCA
CACACTGTGCTAGAACAATATTGAGAGCTACAACCTGATCTGAAAAGCAGAGAGAGATAGA
GACAGAAAGAGGCAGAGAGACACTGGCCCTGGAGCTCGGGCTGGGCATAAAAGTCAGGG
CAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACC
GGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTG
CCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCC
CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCG
ACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCG
GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
CTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAG
CGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGA
CGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACC
TCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCT
ATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTT
CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTT
GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ
ID NO: 99)

CN2256 (1763 bp between ITRs):

GCGGCCGCACGCGTGCTGGTGGGTGGGAAGAGTTTAGTAACTGCTCACCAATGGTGCTGT
AGACATTATCAGTTGCAAGAAGTGTTTTAATTTCAGAGAAATTAAAATATGCAAAATATGTGT
CTCTGAATCAACAAAACATGGTGCCTCAGGTTTCCGTATCTGTGTGGGCCTGTTTTTAGGTT
TTCTGTTCTATCTATTCCTGGGCTTCCACCACACTGTGTTAATGAGGATAACTTTAAAATACA
TCTTGATATTTAGTAGAGCGCATTTCCACGCCTGATTCATCACGAGCACCTTAACTTGTCTT
GGTTCTTCTGCGGAGATGTTGGCACCAGCTTCTTCACCCGTGACTCCAGCTTCTGGGGAA
CCTGTGCTCTCACCCCTGCCCCGCCTGTGCCTGCCAGGCACACCTTAATGAGCTCGGGCT
GGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTT
TCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGG
GGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGT
CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACC
ACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCA
GTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC

FIG. 12C cont'd

CGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACA
ACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAG
CAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCG
GGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGA
ATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAC
TATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCT
TCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGA
ACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACA
ATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCC
CCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGG
GCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGG
ACCGAGCGGCCGC (SEQ ID NO: 100)

CN2315 (2005 bp between ITRs):

GCGGCCGCACGCGTAGTTTTGGAAAAACACCATACTGGTTATTTTCAAATAACTCTATTTCA
CAACAGTATACTAGAATTTTATGCTAAAATCTTTCTGTGACTTTGTGTTCACTGATCAACAAT
TCCTCTTTTATGATAGCCTATGCTGTCTACCATCTATTCCAAACTTTTATGAGATCAGTAATT
TTAGATTATGAATGAGTTTGATCATATAGTACATACTTTTGTGTCTGCTATTGCATTTACATA
GTGACATAGGCTTACAGGATATTGTAAATGTCATCCCTAGATACATTTATGTAACCTTTTTTT
AAAATAAAACTCTAACAAATGTTAGTTATTGTGCACTATGGTTTTGAATATTTGAATTTGCTTT
GCTCAGATCTAAATAACTTCTGATAAGTTTGGCATTCAGGAAAGAGTAAATGTAAATGTAGA
AATGCTTGGAATATAAAACAAGTAATTTTTGTGATAATTTTTCTGAAAATGCTAATGTATATTT
TCAGCAGGCCTCCTTTCTTGTTTAGAACAGTATATTTCTTAGCTTTAATTAGTACTTACTTTA
TATATCTATGCATCCATTTATAGTTCAGCATTAGATTGAATATGTATGCATGTATGTATGATT
TATTTATATTTTTGTCTCCCAGTCCACACCCGAGCTCGGGCTGGGCATAAAAGTCAGGGCA
GAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGG
TCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCC
GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGA
CAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGG
CGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC
GAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCT
CTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTA
TGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTT
CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTT
GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTT

FIG. 12C cont'd

GACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ
ID NO: 101)

CN2423 (2036 bp between ITRs):

GCGGCCGCACGCGTGGGAGATATAAACAAGAAGAAACAGGACTGAATAAATGTGTGCAGA
AGGATCCTGTAGCAGCGTCGTTCTTCCTGGCCAGTTGAGCGCGCGCAGCACTTAAGATTG
GTGCTCATCCATGCCTATTCATTCTTGTGATGTGTTGTGTTAGATATTATTGGTAATTATTAC
TTAGAGAGGTGGCAAAGTAGTGAATAGGTGGAAAGGCACTGAGACCATATTGTCACCGCA
GCTTTGGAAAAGACCTTCTTATTTTAACTTATGTAATTAACCATCCCAAATTTACCCTGATCC
TCACTGGAAATGTCTTCCCTTTCCTGATATTTACCACAGTTTTAAAAAGTCTTAGATGGGTAT
GCATATTTGAGAAATACGTAGAGTTAATTAGTGCGTAGATGTGTTTTTAAGTTGCATAAAAC
ATATTGTGCTTTGGCATAGCTCCCTTTGTAACTTATTTTAGGAAACCAAGTTTGATTTTGACA
ACAGCATTAGAAGTAAAGTGTGTGAAGATGATTTAGAATCCATGAATATCCACATTGCTGAA
TATAATCCTTCCTGGCTACAGAGCTCCTATTTCTGCTCCTGTAAGCAGCCATTCACCCACAC
TTCTTAAAGTAACCCCAATAAACTCACTCATTAAGTTAGATTTTGGCAGTGTCCTTCTTTGAT
CTGTTGTCCGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACA
TTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAA
GGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA
ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACG
ACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGG
ACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCT
GGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCAT
CAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACC
ACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTAC
CTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTC
GACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTG
AAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTA
ATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCC
TGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGG
GCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACT
CCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTC
TATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 102)

CN2424 (1991 bp between ITRs):

GCGGCCGCACGCGTTTTGCTTGTGTGCCAGCAGGGAGGAGCAAGAGGAATTTTGTCACTT
CTCTTTCACTTTCTTCATTATTTTAAGTTGGATCTGCTGTTCTGTTCACTTTTAAAATATAACC
TTCAGTATTTATGAACGTGATTCATCTTTTACAATATGTTGTTTAACATGAATGCATCTTCTTC
TTGCCTTTCAGTGCATCTTTATGCCATTGGCAGCCACAATCAAGTGTTTCCATTAAAGTTCA
CAGTAACAGTGCAGTGTTGTCTGTCCCAACTCTGCTAACTTTTACCACCCACTATTTAGTTT
TATCTACCATGTGAATTTATTTGTAAGATCTTCTTTCTCTTATCCTAGTTAATTATGGGGAGG
ACTTGCCAGGATATTAATTTACATGGCAATTTTGGATAAGTTATGCATGTATTCATCTCCTAT

FIG. 12C cont'd

ATACAATAAGTTTTTCTTGTGAACGTCAAAAGAAAGGATAATGCATATGACTGTATATGATAT
AAACTGGAATGTAATAGTTGCCAATAGGATCTGAATCAAATTTCAGTTTTTAACTATGTTCTA
TATATCTATGTGCATATTTATAGCATGTGTGACATATGTATAATTTAAATATTTCTATAAACTT
ATACAGAAACAAAATATGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATT
GCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGG
TGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCT
CGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGC
AGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT
TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAA
CGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCG
CCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAAC
CACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG
GTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAA
GTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAA
AATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACG
CTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGT
ATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCT
GGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTG
CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAG
GTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 103)

CN2690 (1959 bp between ITRs):

GCGGCCGCACGCGTCTGAAGTTCTGGGGGCAGACCGCAATTGCTGCCACAGAGCCCCTC
CTGGCAGGTGCAGCCCACAGGAAGACCAGCACCCTGGCCACAGAGACAAAAGGTCTTATT
GACTTGCAGTAAAAATGTCCTTATAACCTGTCCAGAACAAAAGGGGAGGATAAGGTTCAAC
CAGACACCGAATTCAGGGAGACAGCATCAGTCAGAGCTTTACCACTGCCAGCAATTTTGAA
GCTGCCCATATTTCTGCTCAGGCAAGATGTTTTCTTGCTCTGAAACCTTTTAGCAGATGCAA
GGTATTGCTCCGGGTGTATCCCAGAACACAAGTCTTTGCTAAAGTGGGTAGACAGCCACTT
TCCTTCAGAGCCAGGGACTGCCCTGGGAGTAAAACCATTTGGAGAGCATTTTATATTTGCA
AAATGTTGCTGTGACATGTTATCATTTATTCCTTTCAACAGCGCAGTGAGATTGGAGAGCAT
CCTCCGATGCACTGTACATTTGGGAGGAAGAGAAAGGATAAGAGCCCTGCTGATTCAGTG
GTGGTCATTATCATTGTCAATGGAGCTGTCACTGTTATCGTCATTTTCGTTTGCGGTGAGGT
ATGTGAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCT
TCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCG
AGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT
GAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCC
TGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCT
TCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACG
GCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTA
CAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGC

FIG. 12C cont'd

CAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCA
GCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCT
ACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGT
TCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGC
GCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGAT
TGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCT
TTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTA
GTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCG
GCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCA
GCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACT
GTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCT
GGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT
GAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 104)

CN2695 (1876 bp between ITRs):

GCGGCCGCACGCGTAGAAAGCAAGAGCAAAGGGCAGGTCTTCCAGGAGAGACAGACAGA
CAAAGAAAAAGCAAGGGAGGTGGAGAGTGTGTGAGAGAGGGAGTGAGAAAGGGCTCAGC
CATTTGTCTGGGCAGGTCTGCGGAACTCAAATCCGGCTCATCTGCATCTCCTCAAATAAAC
TTCTGGGAAGTGAATGCCTGAAGCAGGCAGAGTGCAGAGAAGTCAGCAAATATGTGCGTA
GTGAATTCCAAGCCTTCTGTTCCCAGCTCCTGGGAGGGGCCGCAGTGCCTTGGGGTGTCC
TGCCGTGTCTCCATCGGGTCCCCTGCCTGGCTGCTCTAACCCTGAGGAATGGCCTCCATT
GTGGGCCTGTGGAAACAGGGAGAAGAATCTCTTATCCAGCAGGTCTGAGGGCCACAGGAA
CATAAGTAACACTGAAGAGTGAGGCTGATCCCTGGGGTTTGCGCGAGCCTGCAAATGCAT
GATTCAACAGCACTCGCCTGGTGCCTCTGGGAGGCAAGCTGGATGAGGATGTTGAGCTCG
GGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGA
TCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCAC
CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGC
ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGT
GCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT
GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC
CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCC
ACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCC
GCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTG
AGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGC
GAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTA
ACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATT
GCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGC
GGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTG
ACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATA
AAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG

FIG. 12C cont'd

GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 105)

CN2691 (1971 bp between ITRs):

GCGGCCGCACGCGTAGAAATCCCCTTCCCTACCTTTTATCTATCAATCATATCTATCTATAT
CCCCCAAAACTTAGCTGATACTAACAATTTAAGGATCCCCAATCCCAAAGCATTTAGATTTT
AAGACAAATAAAACAGTTGGAATCCTGGTCCCATGGAATTTCAGAGTTGGGAGAAAGTGAA
GAGATTTCCTGGCTAACCCCTGCAGTTAACAGATGGCAAAAAGAAGTACAGAGAATTGTCT
GAAGTCCCAGAGCAAGGGAGTGGAAGCGCGGCCGCAACGAGGCATGCATAAGATTCTGC
AGAAACCACGTTCTCCCGCTTCACAGCGGTCTGCTCTTTTCAGGAGAGCCTGGGAAAACTT
CTGACTGTGCAATTTGCTTAAAAACAACAGAGAAACATCATAGAAGAGGATCAATTCCTGAC
CATGCATTTGAAATTAGACACTTACAGGCATTCCCTGCAGAGTGCTCTCCACTGGGGCCCA
GTGGAAGCAGCATCACCCATCAGCCGCTGGGGGTGGCTGGTGCCACCAGGTCAGGCCTG
GAGCGGGGCAGGAGTCTCCTCAGGCTCGATGCTCTTGACCTGCTGCTGGTGCCAAGAGTT
AGAAGCAAAGCACCCACCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTAT
TGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATG
GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGG
CGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACG
GCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC
CTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAA
GCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT
CTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCC
TGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGG
CACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAG
AACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTC
GCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAA
CCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACAT
GGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACA
AAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATAC
GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTG
TATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCT
GGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTG
CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAG
GTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 106)

CN2692 (1764 bp between ITRs):

GCGGCCGCACGCGTTTCCTTGGAACACTATTTGGCATATATGCATATTTTGGGGCTCAGTG
TCTTTCATCGTTATGTTTATGGTTATAATGACCACATCTCTCACCCATTAGCTGTGTGGCTTT
TGAGGTCAGTAAGCTATTCTTCCCAGAGTGCTTGACATAGGGCCTGCGGCATGGTAAATAC
CCACAGTTTTTGAGTTTCCTGTGGGTTGAGCATTTAGAATCATCTGTTTGTTTAATCAGGAC
ATATTTCTCCAGTACCTTCAGGGGTCCCAGTACTATGTTTAGGAACTTAGCCATCACTTGCA
CAGTAAACAGTATCTGTGGATTTTGTCATTCAACAAACACTGCCATTGTTCAGTCACTCAAC
AAACGCTGTTACGGCATCCACCCTGGGGAAAGAAGGGTCTCTTGGCAGAGAGCTCGGGCT

FIG. 12C cont'd

GGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTT
TCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGG
GGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGT
CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACC
ACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCA
GTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC
CGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCG
CGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACA
ACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCC
ACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATC
GGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAG
CAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCG
GGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGA
ATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAAC
TATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCT
TCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGA
ACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACA
ATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCC
CCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGG
GCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGG
ACCGAGCGGCCGC (SEQ ID NO: 107)

CN2944 (1693 bp between ITRs):

GCGGCCGCACGCGTTCCATCTTTCTGACTTTAGGGCCATTTTTAGAATGGCCTTTGTTACC
CCAAGATTACATAAATATTCACCCAAACTTCCTCCAGTACTTTTATGGTTTCCTTCACAGGAT
TTAATTCTTAAATCCATCTGGAATGTATTTCTGAAAAAAAAAAAAAAAGTAAAATAGTGCCAG
CTATTTTGCTTCAGAGCCAATGTTTCCTAAGTTGAAAACTTTAGCTATATTCAGTTGTGTGAT
GTCATTAAAGATAATAAGATTTGGGACCAGATTCTGAGTTCAGTTGCACTCTCTGTATAGGG
TATAGAAAATCATTCAACTTTCTGAGAAATGTGCAGAGAGCTCGGGCTGGGCATAAAAGTC
AGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGC
TACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATC
CTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGA
GGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGC
TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA
GTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCA
CCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGG
ACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTT

FIG. 12C cont'd

TACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTT
TCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCC
TGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCG
AGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 108)

CN2886 (1662 bp between ITRs):

GCGGCCGCACGCGTACACACACACACACACACACACACACAAATAAAAACAAAAACAACTCTT
AAAAACCTAAGAAAGGGATAAATACAACATCAAAGACTGTTGGGGTTATAAAACATACACAT
TTATAGACATTTCCAAGATGCTACTAAAATTACATACAAATGAGTATCCAAAATGTCACAGAA
CAAATTATTCTTGTGCAACATAAACAACTAGAGAGCATAGCAAAGGAAGCCCAGACCATGA
GCTTTCCCGATCTGCTTCCCAAGCTAACATATGTTGGCATCCAATCCTGAAAACTACAGAA
GACATGAGGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACAT
TTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAA
GGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA
ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACG
ACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGG
ACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCT
GGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCAT
CAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACC
ACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTAC
CTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTC
GACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTG
AAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTA
ATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCC
TGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGG
GCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACT
CCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTC
TATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 109)

CN2755 (1882 bp between ITRs):

GCGGCCGCACGCGTTTTCCAGCACTGCAGGAGTCTTCAGTATGAAGTAGAACCAGTTGGG
TGACGGGGCAGACACTGGAGACAGACAGAAAAGGTTCTAATATCCATGTTCTGTTTGATTT
TGGAAGGATGCTGAATGCTAATCAGCTTTCTTATTTTAAGAAAGACTATCTTAAATTATTATA
CTTGCCTCATATGATTTGGAGGATGATTGCATGCTGTTCCCTCAGCAAACAGGAGCTTAGA
AAAGTTTTTGAATAATTTCACAAAAGGCCCCAAAGTATTCTCAGTATGAAGATGGAAATTGG
GCTGAGTCATCTTATCTAGAGGAAGTATCAGAAAACTTCGTTGAAATAGTTTTGAGTAGGCC

FIG. 12C cont'd

TTTGGCACAGAAGAAGCCATGTGGATCAGAAGTGAAAAAAATGACGCTTGCCAATGCTTGC
GGTGAGATCTGACATAGAGTGGCCATCTTTCGGGATGCAGTGTTAGAATGTGTAGTAGAAA
ATAAGGTGGAAGAGACGTGTTCGGTTCCATCTCAAGATTTGGGGGCAGAGCTCGGGCTGG
GCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTC
GAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGG
TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCC
GGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCAC
CGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGT
GCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG
AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAAC
GTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCAC
AACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGG
CGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGA
TCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATT
CGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT
GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTC
CCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAAC
TCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATT
CCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCT
CCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA
GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCA
GGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACC
GAGCGGCCGC (SEQ ID NO: 110)

CN2938 (2058 bp between ITRs):

GCGGCCGCACGCGTTGCATGGATTCTAAGTTCCTAAGAGGAATTCAGAAAATACCAATTTT
TATTTAGCATATAAGATAATGTCTTCCATCATAACATTCACAATCACACACAAACATACACAC
ACACACACATCATACTTCATTCATATTCACTGTATTACTATCGTCTGCTAATCTTGCCAACTC
TTACACTGAGAACCTTTCTTGACTTAAATTATGGATATGAGAGACACTGTGATATTTGTTTTT
CTCCACCTACCTCTGCTTTTATATAAGGCATATTCCATGCATGCTAATGATATTGGAGCACA
ATGTTTATGCATGTATAATCATATATTTCTTTCCTGTGCCTGTGAAATCATAGTTATAACTAA
ATTTCTTTCTCTTATGATTCCATAAAAAATTCTCAGGAATATAAAGTATATATAAGCCAAAATT
CTATAAAATTAATAAATTATGATGATCAGAAAGGTACAACAATGTCTTTCAGAAGTCTATAAT
GACTTTTGACCCAGCATGTTCATGCCCTACATGCTGTTAGTGAAATTAACTTTCCAATTCAA
ATTAGTACATTGAGTTGATTGAACATATTTCATCGTTCACATTGACTCACACTGTGGGATTCT
TATAAAAGACAATTTCCCTTAAATATAACCAGCTTTTAGTTTTTTCATGAAAGCCAGTCCCTT
GTTACTAGTGTATTTAGATGCATCGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCA
TCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCA
CCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG
GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCAC
ATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC

FIG. 12C cont'd

ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAG
AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
ACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATT
ACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGAT
ACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCT
TGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTG
CTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTG
TGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTA
GGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 112)

CN2239 (1998 bp between ITRs):

GCGGCCGCACGCGTCTGTGTCTCAGGCCAGAAGATTTTCTATATTGGCTGGAGCTTCTGT
GGGATGCAGGCAGTGGATGAGAAGGGAAACGACAGCAAATAGAGGAAGCACAAGTCTCT
GGCCTCAGTGTTGGGTGCACACAAAGCAAATGGGAGAAGGAGGAGAGGGGCTCCCTTCC
CCTGCCCCAAGGCCCAGCCAGGCAGGCCCCCAGCAGCCTGGCTGTCTCCCTCCCTTGCG
GGCTGGAGCTGCCTCTCCGCCTCTGCCCGCCTGATCGGGGCAGAGCTCGGCTTCATGTAT
TATTAATAGCACTGCCAGCCTTCCTGTTTTCACGCTGATAACCTGCAGGGGAGGGCTGGTG
GGAGGGAGGGAGCAAGGTGGCAAACCCTGGAGCTTCTAGGGTCCCTCCCCCCTCAGTAG
AGCAGCTCTGCGGGGGTTGGCCCCAAGTCTCTAAAGTTGGTTTCTAGAGCCTCCCCCATC
CAAGCCCACGGGAAGGGACCAGGGCCTCCTAAAGGGAAATCACATGGGAGGTTGGAACG
TGGGGTCTGGGAAGGGCCCTGTTACTCAATCCATCTAGGACACCAATGGCTCAAGTCACC
TTGGCCCTCCCCAGAGGCAGAAGTGCCCATGTCTAGGGCACATGGGGCTCAGAGCTGAG
AGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGA
TCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCT
GTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGT
TCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTG
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTA
CGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTC
CGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGA
AGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACA
ACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCA
AGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTC
CAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC
CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGC
GGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGT
ATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCAT
GCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCC

FIG. 12C cont'd

ACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGG
CACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGT
TGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCC
TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCA
CGTGCGGACCGAGCGGCCGC (SEQ ID NO: 113)

CN2410 (1851 bp between ITRs):

GCGGCCGCACGCGTATGTGCATACCTGCCTCTGAGAAGGACCAAAGTCAGATGCACTCTC
AATGTAAGGGCTTGGGGGTGGAACACCTCAAGTTCCACACACACCCCAATCTCCTACAGG
AGTTTCCTTTAGAGGACCTGTCTCCTCCAAGGGGCTCTGTTAGACCAAGTGGGAAAGTCTC
TAGGAACTGACTTCAGAGGTTTGGGGCTGTGCCAGTCCCGCGGCGCACTCTGCCACGGA
GGAGACCTGGAGCTCCTGGGTTGGCCACCTTGCTCCCTGCCTCCTCCCACCTCCTCCCCA
GCAGGTTATCAGCATGAAAACAGGAAGGCTGGCAGTGCTATTAATAAAACATGAAGCAGAG
ATCTGCCCCATCAGGCAGCGGAGGCAGGCGGCAGCTCGGCAGCCCCGAGGGAGGGAGA
GAACCCCCGCATCTGGGCCTGGGCAGGGGTATTCTCTTCTCACAGATGCCACACCCCAGG
AGCCACCTCCATTTTGCGGGACTGCTGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGA
GCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTC
GCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCC
CTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCG
ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGAC
AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCT
GGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG
TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTC
CTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCC
CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTC
GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC
CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
TGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
114)

CN2411 (1824 bp between ITRs):

GCGGCCGCACGCGTAGCTAACCTACCAGTCCCACCGTGATATCTGGATGCTGATATAAAG
GAGAATGTATCAAATAATGTAATGCTTGGAAGTTATTTGAATCTGTTATTCTTAAGAGCACAC
TGATATTACTCATTGTGTGAGGCACACAAAAAAGATCAGAACACAAAGGCATATTTGTGCTG
TGTGTCCAACACTGAGGGGCGGGAGGAGGTAATACTTTTTTGGAAAAGACAGGAAAGAAT

FIG. 12C cont'd

GAAAATAGATGGTTTGTAAAGATAGTAAGCTTGTGGGTGGCTTTTTATTTCCATAATCATGG
TATTATGTTTTTTGTATAATTTTAAGAAAAACTTATGTTAATCAGAAACTTTGAATAACACCTTA
ACAAAAATGTGCTGCACACCGCAGCCATCCCAGAGAAGCTAACACTCAAGATGTCCTGTTC
AGGCTTCGCGCATTTGAGTCCAGGGAAGATGTCTCTCGTTGGCACCAAGAGCTCGGGCTG
GGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTT
CGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGG
GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC
CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCA
CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAG
TGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC
GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGA
CTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCA
CAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGC
AAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAA
TTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACT
ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTT
CCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAA
CTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAA
TTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCC
CTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAAT
GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGG
CAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGA
CCGAGCGGCCGC (SEQ ID NO: 115)

CN2412 (1787 bp between ITRs):

GCGGCCGCACGCGTTGCTCAGGGTCTGCAAACTTTTTCTTATGGTTAGTACAAGAAGAATG
CATAAACATCACAAACCAAAGTACTTCTTTCAACCAATCAAACAAATTATATACAGGGTAACA
CACAAAATCAAACTGAGCAAATCTAGCCAATGGGTCCCATATGCCACATTTTTCATTTCCTT
ATTCCAGCCCTCAGGAATGTAGGTGTTTGCTCTGAGTAGGCCTACCATAGTATGTAATTCT
GCTATCTTATGCACCGATAATTTTTCTTGGTTCAGGAGTGTTTGATGTATGCTCCATAGTTG
TCTCTCATTGTGCTCTGCTGTCTCATTGATTTCTCTAAGGTTTTAACAAATTTATCCATGTGG
TTTTTCTCTAATTTGAAATCAAATTGCATGTCAGACAAGAAGTAATATTTTTGTTGGTAGTGT
ACTGTGGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTT
GCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGG
GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC
CCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCA
CCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGAC
TTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC
GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG
CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAA

FIG. 12C cont'd

GGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG
AGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG
GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGAC
GGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAA
GATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATG
CCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGG
TTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCT
CGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTG
CCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCC
CACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTA
TTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG
GCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 116)

CN2963 (1991 bp between ITRs):

GCGGCCGCACGCGTTCTAACCATGTGTGTTTCCTTTGAAATAATTCTATATACTCCTTTTAA
TGAATTTACATTAAATCAATAGCTTGTATTGACAACTATATTATCATCAAAGCATAAACAAAA
CTTTCTTTGATTCCTTGCTTCAGAAGGCTTGATCATTTGTTTCCCAGGGAAATGGAATATAG
GTAAGGACAAAGCTATTTAAACACAATTCTCTCTTTTAATGCAAACTATTACATCATGCAGTT
TATTTATATTATGAATNCAAAATCAGGGATGACATATATAACATATGGACATGAATATTACTC
TTTAGATTTCATATTAAACTTATCAGTTTTTTTTTATCTCAAAGCATGGAATCAAATTCAAGAC
ACAGATGGACACAACATAAGTTTGGAATGGCTTTTAAATATTTTTACAGGAATTTTTCCATTT
TGGCAAAGTATTTAGTATATAGTGAGTTGCTTCTATGATTATATTAGATGTGTTTTTGAAGAA
TGTTGTTTTGCTAAATTATTCTAACGCTCAATTTACATTCTGAACATTTTCAATCTGAGCGCG
CAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTGCGGCCGCACGCGTTCTAACCA
TGTGTGTTTCTCTTTTGGACTGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTA
TTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCAT
GGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACG
GCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC
GGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCAC
CCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAA
GCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT
CTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCC
TGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGG
CACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAG
AACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTC
GCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAA
CCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACAT
GGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACA
AAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATAC
GCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTG
TATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCT
GGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTG
CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAG
GTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG

FIG. 12C cont'd

GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 117)

CN2623 (1798 bp between ITRs):

GCGGCCGCACGCGTAACATGTATTTGTGGAGTGGCTAAATTGAGATATATCACCTTAAAAT
CAACTCCCAGCAATTTGAAGAGCACAGTAAGTAAGACCTTGTTCCTAACCACAGTCGCCAT
GCAGTGCAGTAGAGCCCAGATCTTACACTTCCTGTGTAACTGGAGTCTTGTGTTCTTTAGC
CAAGCTCTTCCCACATGCGAGTCATTTTCAATTCAAAATACTAGTTATTTTCAGTTTAAGAAT
CAGTCTGATTTCTTGAGCATAGTGCTTATCTTATATGGTATGGCTTTCAGCTCCCTGGAGAT
TTCCTGGGTGTGCTGCTGTGCCTTTGCCTCCAGGATAAATGCCATGCTGAGAAGAGCATGA
CTCTGTCAGCACAGGCAGACGGTGGAAACTTTACCTCTGCGACCTCTCTGCAGCCTCAGG
GTAGCAGGGGATGGCATGGTTTCATGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGC
CATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGC
CACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGC
AGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGA
TTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGG
ATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTC
CTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGC
TGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGAC
TGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTG
GAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
GAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 118)

CN2624 (1833 bp between ITRs):

GCGGCCGCACGCGTTGTATGCCTAAAGGGTTTGCTTCTATGTTAATTCCCAACAACAATCA
TGATACTTCCTTACTTCTGCAGGACTAGGCTGGCTCACTAGTGAGGTTTTCTCCAAGGATG
TGCTCCAATATCTGCTCCACAATACTTAAATAGCTTCACAATACATAAATATTTAAGTACTCC
TCTGTAAGGAGCGTTCTGTTTCCATATGTCAAGGCAGAATCTTCACATAAAGCATGCATTTT
GCTCCCTGTTTTCAATGGTTTTACTCTGTAGAGTTTCTCAAATTAATTCAATACCCACTTTAA
ATTAAATGGAAACTTTAATGCTATGGTTTGCTGTGCCATTTAGATTTTATGTGTTGAGATGCA
ATTCCAACTCTACTTGGCACGAAGTATAGACACTGAATCCAACTATGATGTATAAGCAAGG
GTCCTTGTGAGGTAGGTAATGAGAATGGAGCCTGGAACGTGTAACACTGGTGACTGAGCT
CGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCA
GATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTC

FIG. 12C cont'd

ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG
CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCT
GCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGC
GTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCC
ATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAG
ACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGG
CATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAG
CCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGAT
CCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAG
CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCC
GCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCC
GCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTC
TTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTA
TTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACG
GCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCAC
TGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT
TTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAA
TAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGG
TGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGT
GCGGACCGAGCGGCCGC (SEQ ID NO: 119)

CN2601 (1905 bp between ITRs):

GCGGCCGCACGCGTTGCAGATTATAATTCTATGGAAACTGACTGTTTTAAACATAAATTTTA
GACCTCTTCTCAAGGTGTGGCAATCCTCAACAGCTTCTGGTATGCATAGACCACATTTGTAT
TCACCATTAATGCATAAGAAAAACATTTTGGTTGCCAGTTTGGAGCTAAATTGATAAATCTG
CTACAAGCATTTGTGCATGAGTTTGTGTGTAAATATGCTTTTTACTTACTTAGGATGAATATA
CAAGGATAGCATTGTTATATTGCTTGACATTTGCATATTCAATTTTATAGAGAATGCACAGCA
ATTTTGGTATCTGTCTGTACCCTTTCATATTTTATGCTCAACATTTGAGTCACTTGGTTTCTT
GGCATCATTATGAATTTTTAATGGTCTTACTTTTAAAAAATATGCTTTAACTCATCTTAGAATT
ATATCACATTTCCCCTTCCAGTCTCTCCTAGACATCCCTCCCTCAAACAGCCCTTGTGTGCT
CCTGCCTTCAGTTCATAGCCTGTTTTACTTTATTATTATTATTGCTCTCTGTGTGTGCACGAG
CTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATC
CAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGT
TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTC
AGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGAT
CTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACG
GCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCG
CCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG
GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC
AGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAG
ATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC
CCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAA
GCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGC

FIG. 12C cont'd

CGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATT
CTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCT
ATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCAC
GGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCA
CTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGG
GTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACG
TGCGGACCGAGCGGCCGC (SEQ ID NO: 120)

CN2843 (1979 bp between ITRs):

GCGGCCGCACGCGTGCTTCTGAATACCTCCACGCATACACTCACTTAAATTATAAATAAAC
CTGAAAAGCAATGTAGCAAAAATGAAAGAGATTTCTGCCACCAGCTTCTATAATTATGTAAT
CATAATGTTTTGAAAGTTAAACTCCCCTTTTCAACAAGAAAATTATAATATATATGCATATAT
GCATATATATTCACACACATAACATATATTTATAAATCAATTTAAACTGCCTTTGTGCAGTGT
ACTAGGCAAAAGTAATGTATGTATCTATCTCTGTGAATGCCCTCTAATTCTGGAAATAAATA
ATGGGACATCTCAGTATTTTCAAGAAGTCTGTAAGCCACAAATAGGCTTATGCCTATTGTTC
TGTCATAATTATTAATGGTCTAGTACTTGGTCTCAAGCATGTCTGGTATGGATGATAAAACT
AAGAGTCCAAAAATAGGTCCCCTGTGATCACTATCAGGAAACTCAAAAGCACAGAATTCAT
GGAGTATTTCTAACCAGAAACACAAACACTACACATACTCATACACATATGTGGATATCTGT
TATATAAGATGTAACTTTATATAAAACCCCTAGTCTGGTAATGTTCTTAACCTTATCTGGTGA
CGTTACATGACACAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTT
ACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAG
CAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACG
TAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT
GACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGA
ACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAG
CTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGC
ATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
ACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAG
TCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTG
TGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTT
TAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAAT
CCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAG
GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCT
AGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCC
ACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCA
TTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA
GCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 122)

FIG. 12C cont'd

CN2851 (1670 bp between ITRs):

GCGGCCGCACGCGTTACTGTCCAGGCCAGCTTTGAACAGACTCTGTATTCCAGAAAGACT
TTCAGCTTGTGCTTCTTTTCCCCAAACCTGCCTTGAGTAGCTGACATCATAGGCTTGCACTA
CCAGGCCTGGCTTAAAATACATACTCCCAATTTACATATCTGATGAGTAACATTCTCATAGA
AGCACAAAAGTACATACATGGATTCTAATATAGGTTATTTTCTTGGGCATTACACACTACAG
ATATACTATTCAATATTATTTGTGGCCTACTAGAGACTGAGCCTAAGGCTATAAACATGTTA
GTCGTGCTCTACCAGTGAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATT
GCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGG
TGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCT
CGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGC
AGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT
TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAA
CGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCG
CCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAAC
CACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG
GTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAA
GTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAA
AATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACG
CTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGT
ATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCT
GGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTG
CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAG
GTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 123)

CN2853 (2126 bp between ITRs):

GCGGCCGCACGCGTTGGAACTTGGGACTTCTGAGCACAGGGGGTGGATGGCTGGAAATG
TGAGCAGAACCCATGGTGTGAGAACTGAGATAAGTTTCTGCTGAGGCTAGCCTAGAGGCA
CAGGTGGTGAGACAGCTAGGGCAGAACATATCAGTGAATAGTGTAGGATGATGCCACTTTT
AAAATATTTCCCGCAATATGAACCAATTTCTCAAAAGGTATAAATTACCATAACTCTCTTAAA
TACAAAATGAGTAATCTGAATAGCTGCAAATTGTAAGAAAATTGAACATAATCATAACCCAT
CGGCAAAAATAAATCTTTGGGGAAATGCCACTTTATTAGAGAATTCTGCTATACATTTGATG
AATATATTACAAAGCTTAAAAAGCTCTGTGGGTGTTCATTTTGCATATTGAAAACTAAAATTT
TATATAAGCTAAAAAAGGCCAATTGTCTCTGTCTGTCTTTCTGTCTCTATCTCTGTCTCTCTC
TGTCTCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
CTCTCCCTTTCTTTTTCTCTCTCTGTGTATCTAACTATTCTTCTTTTGCAATTTCTGAAATTTC
TTCTCCAGGGCTTTCTTTCTTCACAAGTGTAAGCACAGCATGTGTGCTCCCACTGCCAAGA
TGCTGTGCCATGCTATAACTTCTTTTTGACATCATATGTAGCTTATACATATTAAAATTTACTT
CAGCACTCAACAGTCACCACTGTCATGACTGGAGGTTGAGCTCGGGCTGGGCATAAAAGT
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCG
CTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT

FIG. 12C cont'd

CCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCG
AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGT
CCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAA
GTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCA
CCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGG
ACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGG
CATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTT
TACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTT
TCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCC
TGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCG
AGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCC
TTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCAT
CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAG
GGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC
(SEQ ID NO: 124)

CN2860 (1956 bp between ITRs):

GCGGCCGCACGCGTTTGGCTTTCCTTCCAAGTTGTATTTTGGTGAAGTTTGACTCCTGGAA
TCTCACTTCTGCACTTTAGATTAGCATACTTTAAAAAGGAACATTACAGTACAATGGAAAAG
TAAAGATCTGGGTTTTATCTCAGCTTTAGTTAGTATTTATAGGAAGCATGGATTACATAAAGA
GATGACATTATTTCAGTCTGAAATTGATTTGTACTATACGACATTTCATTTTGTTTGTGCACA
TTGATTAGATAATTCCTCAAATCAGTTTGAAAATTTCATCTCTGTTCATAGCTGAAGATAACC
ACCAAATTGCTTTTAACTAATTTACATAATAAGTATAGTAGAGGTTTATTCTATATAAAGTATC
AAATATAAATATTCATCTCAAATTCTAAAGGATTTAGCAATGTAAATTACTTATGAATTACATA
CACATTTGTATTTAAATCTTTTTTCACTTTTATTAAAAATAGATCCTTTTCTCATACAATATATC
CTGATTGCAGTTCCCTCCCCTCTACTCCTCCCAGTTCCTTCCCACCTCCCTTTCCATCTGGA
TCTCTCTCCTATTAAAGGTTGGACAGAAGAGATGGCACTGCATGTGAGCTCGGGCTGGGC
ATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAA
GCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG
TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGC
GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCT
TCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCT
ATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACA
TCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC
GGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGAC
CCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCAC

FIG. 12C cont'd

TCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGAT
ATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTG
CTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGT
ATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCAT
CGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCG
TGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCC
CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGA
AATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGA
CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAG
CGGCCGC (SEQ ID NO: 125)

CN2828 (2153 bp between ITRs):

GCGGCCGCACGCGTTGGTTAGCACTATTTTGAATGGACACTTCTATAAAATTTGTTTTAAAA
ACAATATGGAAGAGTCATAAGATGTCTGGGCACAGTGCCCCCAGTCAACTCATCAATGGGA
TTCTGGAATCCAAAGGTTATATGGTGTAAAAACACAGAGAGCAAAATGCACAAGGATCTTC
ATAGAACCATAATTTTTAAAAAGCAAACACTTCCCCCAAAGGAGCATGTAAGACAACTTGAA
GGTATTAAAATGGGATTGTGGTTTTATTAGTTAGAGAACATCATGCAAATTATATTATCTAGG
CATTATAATGCAAACTTTAAAAGGAAATGCTCATTATGTAATGTTAAGTGAAAAAAATCAGGC
CATAAAATGAAATGTGCATTATGCATGCATATGATGAAAAGACAAGAAATAGATAAGAATAT
GCAGAAATACAAATAGTTGGCTGTGGTGTGGGCTTCTGGCTTGTTTTCTGTTTTAAAGCCAT
TACATCATCTTTCCAATGACAAACTACATTTTAAGTTATTGAAAACAAAATCTCGTTAGCCTT
GTCCTTGCATAAATTCCCAGAGAACACATTATTCTTGTTCCCTTCTACCTCTGCTACTTGTCT
TTGATTTGGACCTGCTCCACCTCCTCTTACTTTTCAGTCCTGAACTAGATGGAAAAAAGAAA
TAGACCCTCAGGAAACAAGGAAGCAAAACCACCCAACTGTGTGGAAGCCGTGCATTCAAG
CTGCAAACGAGTGGAGTGGCTTCAATACCTTAACTTATCCCCATTGTTGCAAAGGGAGACC
CAGAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTC
TGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAG
GAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCA
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGA
AGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTG
GGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTC
AAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA
GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACA
ACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCA
ACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTA
CCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTT
CGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCG
CGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATT
GACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAG
TTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAG
CCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTG
TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTG

FIG. 12C cont'd

GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
AGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 126)

CN2870 (1833 bp between ITRs):

GCGGCCGCACGCGTTGTATGCCTAAAGGGTTTGCTTCTATGTTAATTCCCAACAACAATCA
TGATACTTCCTTACTTCTGCAGGACTAGGCTGGCTCACTAGTGAGGTTTTCTCCAAGGATG
TGCTCCAATATCTGCTCCACAATACTTAAATAGCTTCACAATACATAAATATTTAAGTACTCC
TCTGTAAGGAGCGTTCTGTTTCCATATGTCAAGGCAGAATCTTCACATAAAGCATGCATTTT
GCTCCCTGTTTTCAATGGTTTTACTCTGTAGAGTTTCTCAAATTAATTCAATACCCACTTTAA
ATTAAATGGAAACTTTAATGCTATGGTTTGCTGTGCCATTTAGATTTTATGTGTTGAGATGCA
ATTCCAACTCTACTTGGCACGAAGTATAGACACTGAATCCAACTATGATGTATAAGCAAGG
GTCCTTGTGAGGTAGGTAATGAGAATGGAGCCTGGAACGTGTAACACTGGTGACTGAGCT
CGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCA
GATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTC
ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG
CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCT
GCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGC
GTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCC
ATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAG
ACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGG
CATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAG
CCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGAT
CCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAG
CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCC
GCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCC
GCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTC
TTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTA
TTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACG
GCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCAC
TGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT
TTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAA
TAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGG
TGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGT
GCGGACCGAGCGGCCGC (SEQ ID NO: 127)

CN2871 (1904 bp between ITRs):

GCGGCCGCACGCGTTGCAGATTATAATTCTATGGAACTGACTGTTTTAAACATAAATTTTAG
ACCTCTTCTCAAGGTGTGGCAATCCTCAACAGCTTCTGGTATGCATAGACCACATTTGTATT
CACCATTAATGCATAAGAAAAACATTTTGGTTGCCAGTTTGGAGCTAAATTGATAAATCTGC
TACAAGCATTTGTGCATGAGTTTGTGTGTAAATATGCTTTTTACTTACTTAGGATGAATATAC
AAGGATAGCATTGTTATATTGCTTGACATTTGCATATTCAATTTTATAGAGAATGCACAGCAA
TTTTGGTATCTGTCTGTACCCTTTCATATTTTATGCTCAACATTTGAGTCACTTGGTTTCTTG
GCATCATTATGAATTTTTAATGGTCTTACTTTTAAAAAATATGCTTTAACTCATCTTAGAATTA
TATCACATTTCCCCTTCCAGTCTCTCCTAGACATCCCTCCCTCAAACAGCCCTTGTGTGCTC
CTGCCTTCAGTTCATAGCCTGTTTTACTTTATTATTATTATTGCTCTCTGTGTGTGCACGAGC

FIG. 12C cont'd

TCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCC
AGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTT
CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCA
GCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATC
TGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGG
CGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC
CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAA
GACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG
GCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACA
GCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGA
TCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACC
CCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAA
GCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC
CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGC
CGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATT
CTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCT
ATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCAC
GGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCA
CTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTG
TTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTA
ATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGG
GTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACG
TGCGGACCGAGCGGCCGC (SEQ ID NO: 128)

CN2847 (2013 bp between ITRs):

GCGGCCGCACGCGTTCCACTTATACCCCATCAGCATCTCTATCATTGTCTAAGACATTTTTA
ATTTGAAATAAATGGGGAAACTCATTATCATGGCCTCAAGGAAATACAGAAACCAAGGATTG
TCCTTTTAGTGCACAGGTGTCTTCTAAATTCACATATTCTGGATTTAACAAGGAAAATAAAAT
CTGACATAGGCTGAATTTGAAAATGGTAAGGATATTTTTAACTAAACAAAAAAAATAGGTAAT
AGAACATTTCATATTGAACTCCCAGGATAGCATTTATCATAATCCACCAGTTGAATTATTCAT
GCGCTGATAAATAACCTTAGCCTCCGGCACAGTAATCCTCTCACTGAATAATCACCAGGGC
TCCCTGCCCGTGGTGTTTTGCTTTACAACATAAATTTTAAAATTATTCTGTGTTGGCATACAT
ATGTGAAAGCTCATTAACAGATACTGGCATTCTGATTGGTTTGGCACTAAATTTAAACTGTT
TTGGAAAGAACTGAAACACTGAAGGATATGCCATTCTGTTTCTTTATACCTTATCGATGTTT
CCTTAACAATCGGTTGAGATTTATGCATCTTTTGATAGATTTTATCTGAGATAGCTGATATGC
ATTGTGATGAGTTCCTTTTTAACGCTGTAATGTTTCATTGGCCCTGAGCTCGGGCTGGGCA
TAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAA
GCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG
TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGC
GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCT
TCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCT
ATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACA

FIG. 12C cont'd

TCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC
GGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGAC
CCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCAC
TCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGAT
ATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTG
CTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGT
ATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCAT
CGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCG
TGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCC
CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGA
AATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGA
CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAG
CGGCCGC (SEQ ID NO: 129)

CN2686 (1705 bp between ITRs):

GCGGCCGCACGCGTAGGCACCACTTTAAGGGCCTTTATCATTATCCACAACAGCAGCAGC
AGCATGAGTAGCATTAGGAATGCGTTTCCTCCTGTGCTTTGAGGCTTACGTAGGTCTTTCTT
CTGCATCGTTTCATGAAATTCTCACAAGAACACTATGAGGTAGGTGTCAGATGTCCATTTTA
CAGGACAGAAAAGAGATTAAAAAAAAGAGAGAGAGAGAGGAAAAAGGGCTTTATTCAAAGC
CAAGTCAGTTCATATGGGCAGAGCCAGGACTTGAACGCAGGCCCCACAGTCTTAACTCCA
ATATTTTCAGCTGCCTTGGGAACTTCTTGGAAGAGCCCATTCCATTTTCCTGCCATGAGCTC
GGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAG
ATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCA
CCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGC
GTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTG
CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCG
TGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCA
TGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGA
CCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGC
ATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGC
CACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATC
CGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCC
CATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCT
GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCG
CCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCG
CGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTT
AACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATT
GCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGC
GGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTG
ACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATA
AAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG
GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 130)

FIG. 12C cont'd

CN2687 (1770 bp between ITRs):

GCGGCCGCACGCGTGCGGCAGGAATTTAAAGAGGCAGAAGTTCATTTCCTGTGATTTATG
GGACTTTGCTGACTTTCACGCTCCGGTTCTAGGGACAGGGAGACCCTCCCTGCGGGAGCC
GCTGCTAGAGGGAAAGGCAGCGTTGGATCGGATGCCGGCCCCAGCCTGGGAGACGTGCT
TCCTCTGCAGCCTGTGATCAGGAGCGAAGTGTAGAGACCAGAATCGTTCAGAGGCAGGGC
TGGTCCCTGTTGATTTATGACTCAGGCAAACAGTCCATTCATCCGTCAGCGGCTCCACTTC
TATCCCCAGCTTACTGGAGGCAGAAGAATGTCCCCTGGAAGTGGTCTGAGCTGCTTCCTG
TGGTGGCCCCATTCCTCCGGGTGGGGGATCCTATTTTTAGAGGGCCACCATTGCCATGAA
GGTCCAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGC
TTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGC
GAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGG
CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACC
CTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTC
TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGG
CCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACC
AGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGC
TACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAG
TTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGC
GCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGAT
TGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCT
TTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTA
GTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCG
GCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCA
GCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACT
GTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCT
GGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT
GAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 131)

CN2878 (2098 bp between ITRs):

GCGGCCGCACGCGTAGGACAACTTGACTTATGCAAATGAACAATATAAATTTTAGAACTAT
GTATAAATAATTAAATAAGTACATGGACTTACAGGAAGGAATGAATACCTTCAGAGTATGCT
GTAGTATATTTCCTGAATGTATAGATTAAGCTATATAATTGTCAATGAGCTGTTGTATTACGT
ATAATTCAGGGTTTAAACAAAATAAAAATGTGTGCATATGTTTTATGCAAATGCATGTATTTT
TGAATCAATGATAATCCCACACTTTTTATAATGATTCATACCCGCTCAAAGTGATGAAAATAA
TTGACCTGTAAATGTTTATTTTTAAGTAGAAAAAAATCTCTGTGAGGCTCAGAGAACATGTT
GATACAGGGGGAGAGGGAATGTACAGAGTAGCATGCTCTTGTAAGTTGCTCCAAGGCAAC
GCATGACTACTGCGGTCTTGACCTAACAGCTGCACAGACCTTCAGAAGATTGTATCCATCT
ACATTTCATTATGGAGCCAAGAATGGCTCATGAAACTTCTGCCCTCCCTGGGAGCATGTAG
GCAGTTAATTTGAATGGTAATTTTATGTACAAGGAGATAGAATATTTTCAGTGATGTCACCA
CTGGTTTGTCACAATGCTTTCTATAAATGCCCAAATCCTGTAAGTATTCTTAATGAAAACCAT
TGACTTACGCACTAATAGGAAATTAAAGTGGAAACAGATGTTCTATTTTTATTTATGCTGCTG
TGATCAACACGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATT

FIG. 12C cont'd

TGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAG
GGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGA
CCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACC
ACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGA
CTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA
CGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACC
GCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG
GAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATC
AAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCA
CTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCT
GAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCT
GGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGA
CGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAA
AGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAAT
GCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTG
GTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGG
CTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTT
GCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTC
CCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCT
ATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 132)

CN2946 (1798 bp between ITRs):

GCGGCCGCACGCGTAACATGTATTTGTGGAGTGGCTAAATTGAGATATATCACCTTAAAAT
CAACTCCCAGCAATTTGAAGAGCACAGTAAGTAAGACCTTGTTCCTAACCACAGTCGCCAT
GCAGTGCAGTAGAGCCCAGATCTTACACTTCCTGTGTAACTGGAGTCTTGTGTTCTTTAGC
CAAGCTCTTCCCACATGCGAGTCATTTTCAATTCAAAATACTAGTTATTTTCAGTTTAAGAAT
CAGTCTGATTTCTTGAGCATAGTGCTTATCTTATATGGTATGGCTTTCAGCTCCCTGGAGAT
TTCCTGGGTGTGCTGCTGTGCCTTTGCCTCCAGGATAAATGCCATGCTGAGAAGAGCATGA
CTCTGTCAGCACAGGCAGACGGTGGAAACTTTACCTTTGCGACCTCTCTGCAGCCTCAGG
GTAGCAGGGGATGGCATGGTTTCATGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGC
CATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGC
CACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGC
AGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTG
CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGA
TCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGA

FIG. 12C cont'd

TTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGG
ATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTC
CTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGC
TGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGAC
TGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTG
GAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
GAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 133)

CN2943 (1748 bp between ITRs):

GCGGCCGCACGCGTTCTCAACACTAAGAAAACTGCAGTATGAAGATATGGAATTTGAGGC
CAGTTTGAGATAAAACAGAAAGATATGAGCAAAAGACAAGGATGTGAGAAACATAAATAAAA
ACTATGCATATTCATGAAGCACCGCATAATAGTTTCATATGGATACACATCCCATGATGTTT
AAAGCAGGCTTATCATTTTTATCTTTTCAAACATTTAACACTTATGTTAATATATGTAAACTCA
GTCATTTTTAGCCTTCTGAGAGAACCTGCATCATTTTTAATCACTCTTCCGTTCAGTAGGATT
CTTGAATTATTATCATAGTTAACCATAGCTGAGTAATCAAACCTCAATTGTAAATTTGAAAAT
TTACCCTTCTATTCTTTTCTTGCAATAAAAGAGCTCGGGCTGGGCATAAAAGTCAGGGCAG
AGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGT
CGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCC
CTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCG
ACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG
GCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAAC
ATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGAC
AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCG
CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCT
GGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG
TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTC
CTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCC
CGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTC
GACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC
CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC
TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGAT
TGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
134)

CN3319 (2069 bp between ITRs):

GCGGCCGCACGCGTCAACAAACCAGGGAATCTACTGTATAGTGAAGCAGGATTGTGAAAA
GGCCATTCATATAAGAGAACTCCAGATGGCTGACACAGATATGGTAGTGTTAAATCTTTCTG
ATAACCAAGAGGCTCAAGTTAAACTGAGATTTCCCTTTATACACTGTGAATCTGTAAACATT
ACAACCTTCCAGAATACCAAGTACTGTCTAGACTGAGACATAAGAACCTTGTTACAAGGTGT

FIG. 12C cont'd

GTGGCTTTCTAGAATGAAGTCAGTTTACAATATTTGAAAAAAATTAGTATGCATACCCTAGA
CCTAGCAATGGCATCTACTGTTACTTTATAGGTATGGGAACTGGTGACAGATAAGATTCAAT
GTGAAAAGAATGGATAACTAAAATATATTGGAAGCACAGTATGAAATAATATGCAGTGGGTA
GGTGGAATGAGTGATGACGTATAACTCCATGCTAACTATAGTGTGGAGAGAATGCAAAGTA
CGCCAGAGAACAGGATGTATTGTTAAGTATGTAAATTAATTCCATGCACATAATAAGCTTAC
TTATTTTTTCAATGACAAATACATTTATATATTCAATGACATGCATGGAAATTTCTAGAATGTC
TAGAGCGGAGGAGGGGAAATGGATCTGAGTTCAGGAATGAATTAGCAAAAGAATAAAAAAT
GCACATGTGAAGGTATTAACATAAGGCCTTAGAAGCCTGTGTGTCGAGCTCGGGCTGGGC
ATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAA
GCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG
TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGC
GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCT
TCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCT
ATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACA
TCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC
GGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGAC
CCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCAC
TCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGAT
ATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTG
CTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGT
ATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCAT
CGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCG
TGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCC
CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGA
AATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGA
CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGAGATCTCACGTGCGGACCGAG
CGGCCGC (SEQ ID NO: 135)

CN2918 (2147 bp between ITRs):

GCGGCCGCACGCGTCTATGAATAGTTACAAGTGAGAAAGAGAAATGGAGATTTTACCATTT
AATCATCTCCGAAGAGAAGGTGAGAAGAAATTGAGAAAGGTAATTTGCATACCATGCATAG
GCATTAGAGAAAGCATTTTTTTTTCTTAACTCTCCATTTGTTTACAGATATTTCCATATGGCCT
TTTAAAAACTTCAACAGAGAAATTCGAACACAACTAACTGGAATTTCAATTTTCATATAAAAC
CAGCATCATAGATATATATGTAGCTACTCTCCTATGAATAGTTACAAGTGAGAAAGAGAAAT
GGAGATTTTACCATTTAATCATCTCCGAAGAGAAGGTGAGAAGAAATTGAGAAAGGTAATTT
GCATACCATGCATAGGCATTAGAGAAAGCATTTTTTTTCTTAACTCTCCATTTGTTTACAGAT
ATTTCCATATGGCCTTTTAAAAACTTCAACAGAGAAATTCGAACACAACTAACTGGAATTTCA
ATTTTCATATAAAACCAGCATCATAGATATATATGTAGCTACTCTCCTATGAATAGTTACAAG
TGAGAAAGAGAAATGGAGATTTTACCATTTAATCATCTCCGAAGAGAAGGTGAGAAGAAATT
GAGAAAGGTAATTTGCATACCATGCATAGGCATTAGAGAAAGCATTTTTTTTCTTAACTCTC
CATTTGTTTACAGATATTTCCATATGGCCTTTTAAAAACTTCAACAGAGAAATTCGAACACAA
CTAACTGGAATTTCAATTTTCATATAAAACCAGCATCATAGATATATATGTAGCTACTCTCGA

FIG. 12C cont'd

GCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGAT
CCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCT
GTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGT
TCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTG
ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTA
CGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTC
CGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGA
AGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACA
ACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCA
AGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAAC
ACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTC
CAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGAC
CGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGC
GGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGT
ATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCAT
GCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCC
ACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGG
CACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGT
TGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCC
TAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGC
GGACCGAGCGGCCGC (SEQ ID NO: 136)

CN3018 (2105 bp between ITRs):

GCGGCCGCACGCGTAATTTTAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTAAAAA
TAACATAGATCTGAAAAGCTTTATAACTTTTGGTTTTTTTTTTAAGGACCATATTAACAGCAT
CTTCCAGTTTTTCTAGAGTAAATGAATATGTATCCCTTTTATAATTAACAATATACATATATTC
ATTTTGTAAAAAGACTATAAAAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGC
ATCAGACTAAGGAATTTTAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTAAAAATAA
CATAGATCTGAAAAGCTTTATAACTTTTGGTTTTTTTTTTAAGGACCATATTAACAGCATCTT
CCAGTTTTTCTAGAGTAAATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATT
TTGTAAAAAGACTATAAAAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATC
AGACTAAGGAATTTTAGGATGTTCTTAAATAGATCGATTCTTCTACTTTTTTTAAAAATAACAT
AGATCTGAAAAGCTTTATAACTTTTGGTTTTTTTTTTAAGGACCATATTAACAGCATCTTCCA
GTTTTTCTAGAGTAAATGAATATGTATCCCTTTTATAATTAACAATATACATATATTCATTTTG
TAAAAAGACTATAAAAATAGAAGAGAAAATGAAGAGATGTTTATTTCTTTTACACGCATCAGA
CTAAGGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGC
TTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGC
GAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGG
CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACC
CTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTC
TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT

FIG. 12C cont'd

ACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGG
CCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACC
AGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGC
TACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAG
TTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGC
GCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGAT
TGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCT
TTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTA
GTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCG
GCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCA
GCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACT
GTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCT
GGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT
GCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 137)

CN3030 (1937 bp between ITRs):

GCGGCCGCACGCGTCTGTTCTATCTATTCCTGGGCTTCCACCACACTGTGTTAATGAGGAT
AACTTTAAAATACATCTTGATATTTAGTAGAGCGCATTTCCACGCCTGATTCATCACGAGCA
CCTTAACTTGTCTTGGTTCTTCTGCGGAGATGTTGGCACCAGCTTCTTCACCCGTGACTCC
AGCTTCTGGGGAACCTGTGCTCTCACTGTTCTATCTATTCCTGGGCTTCCACCACACTGTG
TTAATGAGGATAACTTTAAAATACATCTTGATATTTAGTAGAGCGCATTTCCACGCCTGATTC
ATCACGAGCACCTTAACTTGTCTTGGTTCTTCTGCGGAGATGTTGGCACCAGCTTCTTCAC
CCGTGACTCCAGCTTCTGGGGAACCTGTGCTCTCACTGTTCTATCTATTCCTGGGCTTCCA
CCACACTGTGTTAATGAGGATAACTTTAAAATACATCTTGATATTTAGTAGAGCGCATTTCC
ACGCCTGATTCATCACGAGCACCTTAACTTGTCTTGGTTCTTCTGCGGAGATGTTGGCACC
AGCTTCTTCACCCGTGACTCCAGCTTCTGGGGAACCTGTGCTCTCAGAGCTCGGGCTGGG
CATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGA
AGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTG
GTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGG
CGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGC
TTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA
GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCC
GAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGT
CTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCG
ACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAG
ACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
ACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCG
ATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGT
TGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCC
GTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTC
ATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTC
CGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTC
CCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAG

FIG. 12C cont'd

GAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG
GACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGC
CGC (SEQ ID NO: 138)

CN3027 (2072 bp between ITRs):

GCGGCCGCACGCGTCCGCAGCTTTGGAAAAGACCTTCTTATTTTAACTTATGTAATTAACC
ATCCCAAATTTACCCTGATCCTCACTGGAAATGTCTTCCCTTTCCTGATATTTACCACAGTTT
TAAAAAGTCTTAGATGGGTATGCATATTTGAGAAATACGTAGAGTTAATTAGTGCGTAGATG
TGTTTTTAAGTTGCATAAAACATATTGTGCTTTGGCATAGCTCCCTTTGTAACTTATTTTAGG
AAACCCCGCAGCTTTGGAAAAGACCTTCTTATTTTAACTTATGTAATTAACCATCCCAAATTT
ACCCTGATCCTCACTGGAAATGTCTTCCCTTTCCTGATATTTACCACAGTTTTAAAAAGTCTT
AGATGGGTATGCATATTTGAGAAATACGTAGAGTTAATTAGTGCGTAGATGTGTTTTTAAGT
TGCATAAAACATATTGTGCTTTGGCATAGCTCCCTTTGTAACTTATTTTAGGAAACCCCGCA
GCTTTGGAAAAGACCTTCTTATTTTAACTTATGTAATTAACCATCCCAAATTTACCCTGATCC
TCACTGGAAATGTCTTCCCTTTCCTGATATTTACCACAGTTTTAAAAAGTCTTAGATGGGTAT
GCATATTTGAGAAATACGTAGAGTTAATTAGTGCGTAGATGTGTTTTTAAGTTGCATAAAAC
ATATTGTGCTTTGGCATAGCTCCCTTTGTAACTTATTTTAGGAAACCGAGCTCGGGCTGGG
CATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGA
AGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTG
GTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGG
CGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGC
TTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA
GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCC
GAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGT
CTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCG
ACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAG
ACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
ACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCG
ATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGT
TGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCC
GTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTC
ATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTC
CGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTC
CCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAG
GAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG
GACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGC
CGC (SEQ ID NO: 139)

CN3032 (2033 bp between ITRs):

GCGGCCGCACGCGTTTTGTGTCTGCTATTGCATTTACATAGTGACATAGGCTTACAGGATA
TTGTAAATGTCATCCCTAGATACATTTATGTAACCTTTTTTTTAAAATAAAACTCTAACAAATGT
TAGTTATTGTGCACTATGGTTTTGAATATTTGAATTTGCTTTGCTCAGATCTAAATAACTTCT
GATAAGTTTGGCATTCAGGAAAGAGTAAATGTAAATGTAGAAATGCTTGGAATTTTGTGTCT

FIG. 12C cont'd

GCTATTGCATTTACATAGTGACATAGGCTTACAGGATATTGTAAATGTCATCCCTAGATACA
TTTATGTAACCTTTTTTTTAAAATAAAACTCTAACAAATGTTAGTTATTGTGCACTATGGTTTTG
AATATTTGAATTTGCTTTGCTCAGATCTAAATAACTTCTGATAAGTTTGGCATTCAGGAAAGA
GTAAATGTAAATGTAGAAATGCTTGGAATTTTGTGTCTGCTATTGCATTTACATAGTGACATA
GGCTTACAGGATATTGTAAATGTCATCCCTAGATACATTTATGTAACCTTTTTTTTAAAATAAA
ACTCTAACAAATGTTAGTTATTGTGCACTATGGTTTTGAATATTTGAATTTGCTTTGCTCAGA
TCTAAATAACTTCTGATAAGTTTGGCATTCAGGAAAGAGTAAATGTAAATGTAGAAATGCTT
GGAATGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTT
CTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGA
GGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCC
ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG
AAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCT
GGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTT
CAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGG
CAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCG
AGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCC
AACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTA
CCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTT
CGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCG
CGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATT
GACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAG
TTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGG
CTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAG
CCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTG
TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTG
GGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
CACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 140)

CN3015 (2057 bp between ITRs):

GCGGCCGCACGCGTCATTGACTCATTTTTATTTTGTGAGGAAGAAATTTTCCATTTGTCTGT
GTAATTACATTTGCTAATTTCTAAAACAATTCCTGGGATTTATATCTATATATGCCTCAGGAA
CATACAGGGAAGATCATACTGTTCTCTGAATTTCAAATTATGTTTTCTGTTTTATTTCATGAA
TTTATTATTTTTCTTTGTCTCTGTCTGTCTCTCTCTGTGTGTATGTGTATGCGTGTATATGCA
TTGACTCATTTTTATTTTGTGAGGAAGAAATTTTCCATTTGTCTGTGTAATTACATTTGCTAAT
TTCTAAAACAATTCCTGGGATTTATATCTATATATGCCTCAGGAACATACAGGGAAGATCAT
ACTGTTCTCTGAATTTCAAATTATGTTTTCTGTTTTATTTCATGAATTTATTATTTTTCTTTGTC
TCTGTCTGTCTCTCTCTGTGTGTATGTGTATGCGTGTATATGCATTGACTCATTTTTATTTTG
TGAGGAAGAAATTTTCCATTTGTCTGTGTAATTACATTTGCTAATTTCTAAAACAATTCCTGG
GATTTATATCTATATATGCCTCAGGAACATACAGGGAAGATCATACTGTTCTCTGAATTTCA
AATTATGTTTTCTGTTTTATTTCATGAATTTATTATTTTTCTTTGTCTCTGTCTGTCTCTCTCTG
TGTGTATGTGTATGCGTGTATATGGAGCTCGGGCTGGGCATAAAAGTCAGGGCAGAGCCA
TCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGGTCGCCA
CCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG

FIG. 12C cont'd

GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC
CTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCAC
ATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAG
AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCA
GCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
ACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCTCTGGATT
ACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGAT
ACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCT
TGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTG
CTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGATCTTCGACTG
TGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTA
GGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO: 141)

CN3012 (1991 bp between ITRs):

GCGGCCGCACGCGTTTTTCATTCCTGGAGGTAAAACCAACCTTAGAATCGTGGAGCCTGAT
AGACTTTGAAATTGGCCTTCATTTTAATTCTCAAGTTCTTAGCATGAGAGCAACATTCAATAT
GGTTCTGAAATATCTCATGCAAAAAAGTGCCAGAGTCATAGACTGAAATATTTATTTGAATA
TAAATGAACTCAGATTTGCCAAAATGCGCTCCCCCTGTTGTTTTTCATTCCTGGAGGTAAAA
CCAACCTTAGAATCGTGGAGCCTGATAGACTTTGAAATTGGCCTTCATTTTAATTCTCAAGT
TCTTAGCATGAGAGCAACATTCAATATGGTTCTGAAATATCTCATGCAAAAAAGTGCCAGAG
TCATAGACTGAAATATTTATTTGAATATAAATGAACTCAGATTTGCCAAAATGCGCTCCCCCT
GTTGTTTTTCATTCCTGGAGGTAAAACCAACCTTAGAATCGTGGAGCCTGATAGACTTTGAA
ATTGGCCTTCATTTTAATTCTCAAGTTCTTAGCATGAGAGCAACATTCAATATGGTTCTGAAA
TATCTCATGCAAAAAAGTGCCAGAGTCATAGACTGAAATATTTATTTGAATATAAATGAACTC
AGATTTGCCAAAATGCGCTCCCCCTGTTGTGAGCTCGGGCTGGGCATAAAAGTCAGGGCA
GAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAAGCTAGCGCTACCGG
TCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGC
CCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCTTCGCCCGCTACCCC
GACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGA
CAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGG
CGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGAC

FIG. 12C cont'd

GAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGATATCATAATCAACCT
CTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTA
TGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTT
CTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCATCGCCGCCTGCCTT
GCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGCTCGAGAGAT
CTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCAT
TGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCCGC (SEQ ID NO:
142)

CN3024 (2186 bp between ITRs):

GCGGCCGCACGCGTCATGAACATTTAGTCATCGTTATCTTACATGAGTCATATCTGTGTCC
TTATTGTCTTATCTTAAATGCATTAATTAATTCGATAGAGCATCTATAAAAATGAAACGCTATT
CAAAAGCTGGCCCATGACAAGATCAGTGTGTTGGTTTCTATGCCTATAGTGGTGCAATGGA
AGGAGTGCGCTGTACCCCACAGACTTTGTGATTGATGAGTTGTCTACTTTATTCACAATGAA
GCTAAAAACAATCTAACAACCATCAGGGTGCTACAGCAGCACAGCATGAACATTTAGTCAT
CGTTATCTTACATGAGTCATATCTGTGTCCTTATTGTCTTATCTTAAATGCATTAATTAATTC
GATAGAGCATCTATAAAAATGAAACGCTATTCAAAAGCTGGCCCATGACAAGATCAGTGTG
TTGGTTTCTATGCCTATAGTGGTGCAATGGAAGGAGTGCGCTGTACCCCACAGACTTTGTG
ATTGATGAGTTGTCTACTTTATTCACAATGAAGCTAAAAACAATCTAACAACCATCAGGGTG
CTACAGCAGCACAGCATGAACATTTAGTCATCGTTATCTTACATGAGTCATATCTGTGTCCT
TATTGTCTTATCTTAAATGCATTAATTAATTCGATAGAGCATCTATAAAAATGAAACGCTATT
CAAAAGCTGGCCCATGACAAGATCAGTGTGTTGGTTTCTATGCCTATAGTGGTGCAATGGA
AGGAGTGCGCTGTACCCCACAGACTTTGTGATTGATGAGTTGTCTACTTTATTCACAATGAA
GCTAAAAACAATCTAACAACCATCAGGGTGCTACAGCAGCACAGGAGCTCGGGCTGGGCA
TAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGGGATCCAGATCTTTCGAA
GCTAGCGCTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG
TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGC
GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCGTGCAGTGCT
TCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCG
AGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCT
ATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACA
TCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC
GGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGAC
CCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCAC
TCTCGGCATGGACGAGCTGTACAAGTAAGTCGACGGCGCGCCGCGGCCGCGAATTCGAT
ATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTG
CTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGT
ATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTAGTTCTTGCCACGGCGGAACTCAT
CGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCG
TGGCTCGAGAGATCTTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCC
CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGA

FIG. 12C cont'd

AATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGA
CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCACGTGCGGACCGAGCGGCC
GC (SEQ ID NO: 143)

FIG. 4B