



(51) International Patent Classification:

Not classified

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(21) International Application Number:

PCT/US2024/029057

(22) International Filing Date:

13 May 2024 (13.05.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/501,903 12 May 2023 (12.05.2023) US

(71) Applicant: TEMPLE UNIVERSITY - OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION

[US/US]; Broad Street and Montgomery Avenue, Philadelphia, Pennsylvania 19122 (US).

(72) Inventor: HU, Wenhui; 224 Europa Court, Cherry Hill,

New Jersey 08003 (US).

(74) Agent: ZACHARIADES, Nicholas A.; Fox Rothschild

LLP, 997 Lenox Drive, Building 3, Lawrenceville, New Jersey 08648-2311 (US).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every

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

(54) Title: CHIMERIC CAPSID RECOMBINANT ADENO-ASSOCIATED VIRUS WITH BBB PENETRATION AND MICROGLIA TARGETING (AAV-BM)

(57) Abstract: Compositions that target microglia-related diseases such as infectious diseases (e.g. HIV), neurodegenerative diseases (e.g. Alzheimer's, Parkinson's, Huntington's diseases), neurodevelopmental disorders (e.g. autism, intellectual disability) and CNS injury (e.g. TBI, SCI), include gene editing complexes. Novel AAV-BM serotypes that can cross the blood-brain barrier and transduce microglia, and package nucleic acid sequences, inclusive of, gene-editing complexes, shRNAs, sgRNAs, miRNAs and lncRNAs are utilized in the treatment of neurological diseases.



CHIMERIC CAPSID RECOMBINANT ADENO-ASSOCIATED VIRUS WITH BBB PENETRATION and MICROGLIA TARGETING (AAV-BM)

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority under 35 U.S.C. § 119(e) to U.S. Provisional Application No. 63/501,903 filed May 12, 2023. The entire contents of this application are incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under R01 DA056876, and R01 MH130193 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

[0003] Human immunodeficiency virus (HIV) infection of the Central Nervous System (CNS) causes encephalopathy (HPE: HIV-associated progressive encephalopathy), which presents as cognitive, memory and/or behavioral changes. HPE can progress to HIV-associated dementia (HAD/NeuroAIDS), which occurs in end-stage patients with very high viral loads and very low (<200 per mL) CD4⁺ T cell counts. In the post-ART (Antiretroviral Therapy) era, clinicians observe HIV-Associated Neurocognitive Disorder (HAND) even in patients with normal CD4⁺ T cell counts. Thus, CNS cells, mainly microglia (MG), harboring integrated HIV sequences appear to be sufficient to bring about the cognitive decline, opportunistic infections, autoimmunity, and muscle wasting observed in the clinic. Ridding patients of this reservoir of chronically infected cells is the most direct route toward eradication of HAND and HIV.

[0004] Despite their overall effectiveness in relieving immunodeficiency, many of the antiviral drugs present in ART regimens have limited access to the CNS because they cannot cross the Blood-Brain Barrier (BBB) efficiently. On the other hand, the long-lived myeloid cells such as perivascular macrophages and resident microglia in CNS persistently harbor HIV provirus and could contribute to the source of residual viremia during long-term ART or rebounding virus upon ART cessation. On the other hand, the long-lived myeloid cells such as perivascular macrophages and resident microglia in CNS persistently harbor HIV provirus and could contribute to the source

of residual viremia during long-term ART or rebounding virus upon ART cessation. It is undoubtedly and urgently need-ed to develop novel strategies to specifically target CNS myeloid cells for HIV eradication and a cure.

SUMMARY

[0005] The lack of Adeno-Associated Virus (AAV) serotypes that are highly effective and reliable to transduce myeloid cells in the CNS remains the key challenge. Accordingly, we now provide recombinant AAV that cross the blood brain barrier, specifically transduce microglia with a high efficiency.

[0006] Embodiments relate to recombinant Adeno-Associated Virus (AAV) chimeric capsids to treat HIV- and non-HIV-associated diseases of nervous system. Accordingly, this disclosure relates to novel AAV serotypes, in particular serotypes of the AAV having the ability to cross the Blood-Brain Barrier (BBB) and infect myeloid cells including microglia and macrophages (AAV-BM).

[0007] In certain aspects, an adeno-associated virus (AAV) comprises a nucleic acid sequence encoding: (i) a capsid protein comprising a BBB-penetrating peptide, and (ii) a microglia-tropic peptide. In certain embodiments, the AAV transfer vector comprises a microglial tropic promoter, a mRNA expression-boosting 5'-UTR (from SARS-CoV-2), a transgene, a woodchuck hepatitis virus post-transcriptional response element (WPRE) short version WPRE3, 4x miRNA-9 target sites (miR9T), and poly(A) sequence. In certain embodiments, the BBB-penetrating peptide is derived from an AAV serotype (AAV-B) that efficiently crosses the BBB relative to an AAV serotype that lacks ability to cross the BBB. In certain embodiments, the AAV serotype that efficiently crosses the BBB comprises AAV8, AAV9, AAV10 or derivatives thereof. In certain embodiments, the AAV9 derivatives comprise those with reported BBB-penetrating efficiency, such as AAV-PhP.B, AAV-PhP-eB, AAV-F, AAV9-Retro, AAV-Cap-MAC, AAV-C2, AAV9P31, AAV-B10, AAV-B22, AAV-DJ8, AAV-CPP21, AAV-CPP16, AAV-cc47, AAV-Pal1, AAV-MDV1A, AAVBI28, AV-BI62, AAV-BI65, and others, or combinations thereof. In certain embodiments, the BBB penetrating peptide is inserted into a capsid VP1 protein of microglia-tropism AAV (AAV-M), for example, AAV1, AAV2, AAV5, AAV6, AAV-DJ and their derivatives (including mutations). In certain embodiments, the AAV VP1 capsid protein is an AAV2 VP1 capsid protein. In certain embodiments, the BBB penetrating peptide is inserted at

amino acid position N587 of AAV2 VP1 protein. In certain embodiments, the transgene encodes a therapeutic agent. In certain embodiments, the transgene encodes a gene-editing complex. In certain embodiments, the gene editing complex comprises at least two isolated nucleic acid sequences comprising: (i) an isolated nucleic acid sequence encoding a Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease, and (ii) at least one guide RNA (gRNA), the gRNA being complementary to a target sequence in a virus or host genome. In certain embodiments, the gRNA target sequences are different. In certain embodiments, the virus comprises a retrovirus. In certain embodiments, the virus is a human immunodeficiency virus (HIV). In certain embodiments, the first and second target sequences comprise one or more nucleic acid sequences in a HIV comprising: long terminal repeat (LTR) nucleic acid sequences, nucleic acid sequences encoding structural proteins, non-structural proteins, or combinations thereof. In certain embodiments, the sequences encoding structural proteins comprise nucleic acid sequences encoding: Gag, Gag-Pol precursor, Pro (protease), Reverse Transcriptase (RT), integrase (In), Env or combinations thereof, and wherein the sequences encoding non-structural proteins comprise nucleic acid sequences encoding: regulatory proteins, accessory proteins or combinations thereof. In certain embodiments, the regulatory proteins comprise: Tat, Rev or combinations thereof, and wherein accessory proteins comprise Nef, Vpr, Vpu, Vif or combinations thereof. In certain embodiments, the endonuclease comprises Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, ARMAN 1, ARMAN 4, saCas9, cjCas9, Cas12a, Cas12f, mutants, variants, high-fidelity variants, orthologs, analogs, fragments, or combinations thereof. In certain embodiments, the amino acid sequence encoding the BBB penetrating peptide comprises SAQLNTTKPI (SEQ ID NO: 1) for AAV2-Rep-Cap-MAC (R2-MAC). In certain embodiments, the BBB penetrating peptide comprises poly-L-arginine, a transferrin or lactoferrin ligand, transactivator of transcription (tat), penetratin, insulin peptide or lipoprotein. In certain embodiments, the therapeutic agent is a treatment for neurological disorders, e.g. AAV-saCas9, AAV-cjCas9, AAV-Cas12f, and others. In certain embodiments, the neurological disorders comprise: AIDS dementia complex, Alzheimer's disease, amyotrophic lateral sclerosis, adrenoleukodystrophy, Alexander disease, Alper's disease, ataxia telangiectasia, Batten disease, bovine spongiform encephalopathy (BSE), Canavan disease, corticobasal degeneration, Creutzfeldt-Jakob disease, dementia with Lewy bodies, fatal familial insomnia, frontotemporal

lobar degeneration, Huntington's disease, Kennedy's disease, Krabbe disease, Lyme disease, Machado-Joseph disease, multiple sclerosis, multiple system atrophy, neuroacanthocytosis, Niemann-Pick disease, Parkinson's disease, Pick's disease, primary lateral sclerosis, progressive supranuclear palsy, Refsum disease, Sandhoff disease, diffuse myelinoclastic sclerosis, spinocerebellar ataxia, subacute combined degeneration of spinal cord, tabes dorsalis, Tay-Sachs disease, toxic encephalopathy, transmissible spongiform encephalopathy, or wobbly hedgehog syndrome. In certain embodiments, the therapeutic agent comprises antibodies, aptamers, peptides, proteins, natural ligands of one or more CNS target(s), modified versions of natural ligands of one or more CNS target(s), aptamers, inhibitory nucleic acids, small inhibitory RNAs (siRNA), short hairpin RNAs (shRNA)), miRNA, dsDNA, lncRNA, antisense oligonucleotides ribozymes, and small molecules, or active fragments of thereof. In certain embodiments, the microglia specific peptide comprises M1, M2 or the combination thereof.

[0008] In another aspect, a method of treating a subject suffering from a virus mediated neurological disorder or a neurological disorder comprising: administering to the subject a therapeutically effective amount of an AAV comprising a nucleic acid sequence encoding: (i) a capsid protein comprising a BBB penetrating peptide, (ii) a microglial specific peptide, (iii) a microglial tropic promoter and (iv) a transgene. In certain embodiments, the AAV further comprises a woodchuck hepatitis virus post-transcriptional response element (WPRE). In certain embodiments, the BBB penetrating peptide is derived from an AAV serotype that efficiently crosses the BBB relative to an AAV serotype that lacks ability to cross the blood brain barrier. In certain embodiments, the AAV serotype that efficiently crosses the BBB comprises AAV2, AAV8, AAV9, AAV10 or derivatives thereof. In certain embodiments, the AAV9 derivatives comprise AAV-PhP.B, AAV-PhP-eB, AAV-F, AAV9-Retro, AAV-Cap-MAC, AAV-C2, AAV9P31, AAV-B10, AAV-B22, AAV-DJ8, or combinations thereof. In certain embodiments, the BBB penetrating peptide is inserted into an AAV2, capsid VP1 protein. In certain embodiments, the BBB penetrating peptide is inserted at amino acid position Q587. In certain embodiments, the transgene encodes a therapeutic agent. In certain embodiments, the transgene encodes a gene-editing complex. In certain embodiments, the gene editing complex comprises one or more isolated nucleic acid sequences encoding a CRISPR-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a target sequence in a virus genome. In certain embodiments, the gene editing complex comprises at least two isolated nucleic acid sequences

comprising: (i) a first isolated nucleic acid sequence encoding a first Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a first target sequence in a virus genome, and (ii) a second isolated nucleic acid sequence encoding a second Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a second target sequence in a virus genome. In certain embodiments, the first and second target sequences are different. In certain embodiments, the virus comprises a retrovirus. In certain embodiments, the virus is a human immunodeficiency virus (HIV). In certain embodiments, the first and second target sequences comprise one or more nucleic acid sequences in a human immunodeficiency virus (HIV) comprising: long terminal repeat (LTR) nucleic acid sequences, nucleic acid sequences encoding structural proteins, non-structural proteins, or combinations thereof. In certain embodiments, the sequences encoding structural proteins comprise nucleic acid sequences encoding: Gag, Gag-Pol precursor, Pro (protease), Reverse Transcriptase (RT), integrase (In), Env or combinations thereof, and wherein the sequences encoding non-structural proteins comprise nucleic acid sequences encoding: regulatory proteins, accessory proteins or combinations thereof. In certain embodiments, regulatory proteins comprise: Tat, Rev or combinations thereof, and wherein accessory proteins comprise Nef, Vpr, Vpu, Vif or combinations thereof. In certain embodiments, endonuclease comprises Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4, mutants, variants, high-fidelity variants, orthologs, analogs, fragments, or combinations thereof. In certain embodiments, the nucleic acid sequence encoding the blood brain barrier (BBB) penetrating peptide comprises at least an 80% sequence identity to SAQLNTTKPI (SEQ ID NO: 1). In certain embodiments, the amino acid sequence encoding the blood brain barrier (BBB) penetrating peptide comprises SAQLNTTKPI (SEQ ID NO: 1) for AAV2-Rep-Cap-MAC (R2-MAC). In certain embodiments, the blood brain barrier (BBB) penetrating peptide comprises poly-L-arginine, a transferrin or lactoferrin ligand, transactivator of transcription (tat), penetratin, insulin peptide or lipoprotein. In certain embodiments, the therapeutic agent is a treatment for neurological disorders. In certain embodiments, the neurological disorders comprise: AIDS dementia complex, Alzheimer's disease, amyotrophic lateral sclerosis, adrenoleukodystrophy, Alexander disease, Alper's disease, ataxia telangiectasia, Batten disease,

bovine spongiform encephalopathy (BSE), Canavan disease, corticobasal degeneration, Creutzfeldt-Jakob disease, dementia with Lewy bodies, fatal familial insomnia, frontotemporal lobar degeneration, Huntington's disease, Kennedy's disease, Krabbe disease, Lyme disease, Machado-Joseph disease, multiple sclerosis, multiple system atrophy, neuroacanthocytosis, Niemann-Pick disease, Parkinson's disease, Pick's disease, primary lateral sclerosis, progressive supranuclear palsy, Refsum disease, Sandhoff disease, diffuse myelinoclastic sclerosis, spinocerebellar ataxia, subacute combined degeneration of spinal cord, tabes dorsalis, Tay-Sachs disease, toxic encephalopathy, transmissible spongiform encephalopathy, or wobbly hedgehog syndrome. In certain embodiments, the therapeutic agent comprises antibodies, aptamers, peptides, proteins, natural ligands of one or more CNS target(s), modified versions of natural ligands of one or more CNS target(s), aptamers, inhibitory nucleic acids, small inhibitory RNAs (siRNA), short hairpin RNAs (shRNA), ribozymes, and small molecules, or active fragments of thereof. In certain embodiments, the microglia specific peptide comprises M1, M2 or the combination thereof. In certain embodiments, the treatment further comprises administering to the subject one or more secondary treatments. In certain embodiments, the secondary treatments comprise anti-virus agents, chemotherapeutic agents, immunotherapy, growth inhibitory agents, cytotoxic agents, radiation therapy, surgery, anti-angiogenesis agents, apoptotic agents, anti-tubulin agents and combinations thereof. In certain embodiments, the anti-virus agent comprises antibodies, aptamers, adjuvants, anti-sense oligonucleotides, chemokines, cytokines, immune stimulating agents, immune modulating agents, B-cell modulators, T-cell modulators, NK cell modulators, antigen presenting cell modulators, enzymes, siRNA's, ribavirin, protease inhibitors, helicase inhibitors, polymerase inhibitors, helicase inhibitors, neuraminidase inhibitors, nucleoside reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors, purine nucleosides, chemokine receptor antagonists, interleukins, or combinations thereof.

[0009] In another aspect, an AAV transfer vector for microglia-specific transgene expression comprises a nucleic acid including: (i) a microglial tropic minimal/essential promoter, (ii) a mRNA expression-boosting UTR, (iii) a transgene, a mRNA stabilizing minimal Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element 3 (WPRE3) and (iv) a microglia-illuminating miR9T and (V) a poly(A) sequence.

[0010] In another aspect, a recombinant adeno-associated virus (rAAV) comprises a nucleic acid sequence encoding: (i) a capsid protein comprising a BBB penetrating peptide, (ii) at least two

microglial specific peptides, (iii) a microglial tropic promoter and (iv) a transgene. In certain embodiments, the method further comprises a nucleic acid encoding: (i) a microglial tropic minimal/essential promoter, (ii) a mRNA expression-boosting UTR, (iii) a mRNA stabilizing minimal Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element 3 (WPRES) and (iv) a poly(A) sequence.

[0011] In another aspect, a bar-coded (bc) adeno-associated virus (AAV) library comprises bar-coded AAV transfer vectors wherein each bar coded AAV transfer vector comprises two CD68 or two CD11b promoters one or a plurality of microglial specific peptides. In certain embodiments, each microglial specific peptide specifically targets the bc AAV to microglia. In certain embodiments, each microglial specific peptide specifically binds to different microglial targets. In certain embodiments, the bc AAV comprises adeno-associated virus serotype B (AAV-B).

[0012] In another aspect, a method of screening for adeno-associated virus target specificity comprising, screening the bar coded AAV library against microglial specific markers to determine specificity. In certain embodiments, the microglial specific markers comprise Tmem119, P2ry12, HexB, Fcrls, Sall1, C1q, Gpr34, Olfr13, Mertk, Prosr, Tyro3, Tgfr1 or combinations thereof.

[0013] In another aspect, the amino acid sequence encoding a blood brain barrier (BBB) penetrating peptide comprises an amino acid sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SAQLNTTKPI (SEQ ID NO: 1). In some embodiments, the BBB peptide comprises an amino acid sequence having at least about 80% sequence identity to SEQ ID NO: 1. In some embodiments, the BBB comprises an amino acid sequence having at least about 90% sequence identity to SEQ ID NO: 1. In some embodiments, the BBB comprises an amino acid sequence having at least about 95% sequence identity to SEQ ID NO: 1. In some embodiments, the BBB comprises the amino acid sequence of SEQ ID NO: 1.

[0014] In another aspect, the nucleotide sequence of an AAV-R2-MAC (AAV2-REP-link-VP1(SAQ-MAC)) comprises a sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 2. In some embodiments, the AAV-R2-MAC comprises a nucleotide sequence having at least about 80% sequence identity to SEQ ID NO: 2. In some embodiments, the AAV-R2-MAC comprises a nucleotide sequence having at least about 90% sequence identity to SEQ ID

NO: 2. In some embodiments, the AAV-R2-MAC comprises a nucleotide sequence having at least about 95% sequence identity to SEQ ID NO: 2. In some embodiments, the AAV-R2-MAC comprises a nucleotide sequence of SEQ ID NO: 2.

[0015] In another aspect, the amino acid sequence encoding an R2-MAC, REP protein comprises an amino acid sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 3. In some embodiments, the R2-MAC, REP protein comprises an amino acid sequence having at least about 80% sequence identity to SEQ ID NO: 3. In some embodiments, the R2-MAC, REP protein comprises an amino acid sequence having at least about 90% sequence identity to SEQ ID NO: 3. In some embodiments, the R2-MAC, REP protein comprises an amino acid sequence having at least about 95% sequence identity to SEQ ID NO: 3. In some embodiments, the R2-MAC, REP protein comprises the amino acid sequence of SEQ ID NO: 3.

[0016] In another aspect, the amino acid sequence encoding an R2-MAC, VP1(SAQ-MAC) protein comprises an amino acid sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 4. In some embodiments, an R2-MAC, VP1(SAQ-MAC) comprises an amino acid sequence having at least about 80% sequence identity to SEQ ID NO: 4. In some embodiments, an R2-MAC, VP1(SAQ-MAC) comprises an amino acid sequence having at least about 90% sequence identity to SEQ ID NO: 4. In some embodiments, an R2-MAC, VP1(SAQ-MAC) comprises an amino acid sequence having at least about 95% sequence identity to SEQ ID NO: 4. In some embodiments, an R2-MAC, VP1(SAQ-MAC) comprises the amino acid sequence of SEQ ID NO: 4.

[0017] In another aspect, an AAV-R2e-MAC nucleotide sequence (enhanced version of R2-MAC with 4 YF mutations and 1 TV mutation; (AAV2-REP-link-VP1(4YF/TV and SAQ-MAC)) comprises a nucleotide sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 5. In some embodiments, an AAV-R2e-MAC comprises a nucleotide sequence having at least about 80% sequence identity to SEQ ID NO: 5. In some embodiments, an AAV-R2e-MAC comprises a nucleotide sequence having at least about 90% sequence identity to SEQ ID NO: 5. In some embodiments, an AAV-R2e-MAC comprises a nucleotide sequence having at least about

95% sequence identity to SEQ ID NO: 5. In some embodiments, an AAV-R2e-MAC comprises a nucleotide sequence of SEQ ID NO: 5.

[0018] In another aspect, the amino acid sequence encoding an R2e-MAC, REP protein comprises an amino acid sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 6. In some embodiments, an R2e-MAC, REP protein comprises an amino acid sequence having at least about 80% sequence identity to SEQ ID NO: 6. In some embodiments, the an R2e-MAC, REP protein comprises an amino acid sequence having at least about 90% sequence identity to SEQ ID NO: 6. In some embodiments, the an R2e-MAC, REP protein comprises an amino acid sequence having at least about 95% sequence identity to SEQ ID NO: 6. In some embodiments, an R2e-MAC, REP protein comprises the amino acid sequence of SEQ ID NO: 6.

[0019] In another aspect, the amino acid sequence encoding an R2e-MAC, VP1(4YF/TV and SAQ-MAC) protein comprises an amino acid sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 7. In some embodiments, an R2e-MAC, VP1 comprises an amino acid sequence having at least about 80% sequence identity to SEQ ID NO: 7. In some embodiments, an R2e-MAC, VP1 comprises an amino acid sequence having at least about 90% sequence identity to SEQ ID NO: 7. In some embodiments, an R2e-MAC, VP1 comprises an amino acid sequence having at least about 95% sequence identity to SEQ ID NO: 7. In some embodiments, an R2e-MAC, VP1 comprises the amino acid sequence of SEQ ID NO: 7.

[0020] In another aspect, the nucleotide sequence encoding TP2018 (ITR-pdsAAV-hHexB-134-UTR-EGFP-4xmiR9T-BGH-ITR) comprises a nucleotide sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 8. In some embodiments, the TP2018 comprises a nucleotide sequence having at least about 80% sequence identity to SEQ ID NO: 8. In some embodiments, the TP2018 comprises a nucleotide sequence having at least about 90% sequence identity to SEQ ID NO: 8. In some embodiments, TP2018 comprises a nucleotide sequence having at least about 95% sequence identity to SEQ ID NO: 8. In some embodiments TP2018 comprises a nucleotide sequence of SEQ ID NO: 8.

[0021] In another aspect, the nucleotide sequence encoding an AAV barcode reporter vector (ITR-pDsAAV-hHexB134-UTR-NLS-iCre-HA-WPRE3-bc25-miR9T-bGH-ITR) comprises a nucleotide sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 9. In some embodiments, an AAV barcode reporter vector comprises a nucleotide sequence having at least about 80% sequence identity to SEQ ID NO: 9. In some embodiments, an AAV barcode reporter vector comprises a nucleotide sequence having at least about 90% sequence identity to SEQ ID NO: 9. In some embodiments, an AAV barcode reporter vector comprises a nucleotide sequence having at least about 95% sequence identity to SEQ ID NO: 9. In some embodiments, an AAV barcode reporter vector comprises a nucleotide sequence of SEQ ID NO: 9.

[0022] In another aspect, a nucleotide sequence encoding an AAV barcode reporter vector (pDsAAV-hCD68-135-UTR-NLS-iCre-HA-WPRE3-bc25-miR9T-bGH-ITR) comprises a nucleotide sequence having at least about 70% (such as at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater) sequence identity to SEQ ID NO: 10. In some embodiments, an AAV barcode reporter vector comprises a nucleotide sequence having at least about 80% sequence identity to SEQ ID NO: 10. In some embodiments, an AAV barcode reporter vector comprises a nucleotide sequence having at least about 90% sequence identity to SEQ ID NO: 10. In some embodiments, an AAV barcode reporter vector comprises a nucleotide sequence having at least about 95% sequence identity to SEQ ID NO: 10. In some embodiments, an AAV barcode reporter vector comprises a nucleotide sequence of SEQ ID NO: 10.

[0023] In another aspect, a pharmaceutical composition comprises any one or more SEQ ID NOs: 1-9, 10 and combinations thereof.

[0024] Definitions

[0025] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the disclosure. Unless specifically defined otherwise, all technical and scientific terms used herein shall be taken to have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, and biochemistry).

[0026] As used herein, the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. Furthermore, to the extent that the

terms “including”, “includes”, “having”, “has”, “with”, or variants thereof are used in either the detailed description and/or the claims, such terms are intended to be inclusive in a manner similar to the term “comprising.”

[0027] As used in this specification and the appended claims, the term “or” is generally employed in its sense including “and/or” unless the content clearly dictates otherwise.

[0028] The term “about” is used herein to mean a value that is $\pm 10\%$ of the recited value.

[0029] The term “AAV” refers to adeno-associated virus and may be used to refer to the naturally occurring wild-type virus itself or derivatives thereof. The term covers all subtypes, serotypes and pseudotypes, and both naturally occurring and recombinant forms, except where required otherwise. The AAV genome is built of single stranded DNA and comprises inverted terminal repeats (ITRs) at both ends of the DNA strand, and two open reading frames: rep and cap, encoding replication and capsid proteins, respectively. A foreign polynucleotide can replace the native rep and cap genes. AAVs can be made with a variety of different serotype capsids which have varying transduction profiles or as used herein, “tropism” for different tissue types. As used herein, the term “serotype” refers to an AAV which is identified by and distinguished from other AAVs based on capsid protein reactivity with defined antisera, e.g., AAV serotypes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, DJ or DJ/8. For example, serotype AAV2 is used to refer to an AAV which contains capsid proteins encoded from the cap gene of AAV2 and a genome containing 5' and 3' ITR sequences from the same AAV2 serotype. Pseudotyped AAV as refers to an AAV that contains capsid proteins from one serotype and a viral genome including 5'-3' ITRs of a second serotype. Pseudotyped rAAV would be expected to have cell surface binding properties of the capsid serotype and genetic properties consistent with the ITR serotype. Pseudotyped rAAV are produced using standard techniques described in the art. In certain embodiments, the AAVs comprise mutations in the capsid proteins. These mutations comprise one or more amino acid changes at one or multiple locations of the capsid proteins.

[0030] As used herein, by “administering” is meant a method of giving a dosage of a composition described herein (e.g., an AAV and/or a pharmaceutical composition thereof) to a subject. The compositions utilized in the methods described herein can be administered by any suitable route, including, for example, by inhalation, nebulization, aerosolization, intranasally, intratracheally,

intrabronchially, orally, parenterally (e.g., intravenously, subcutaneously, or intramuscularly), orally, nasally, rectally, topically, or buccally.

[0031] The term “amino acid” as used herein refers to naturally occurring and synthetic α , β , γ , and δ amino acids, and includes but is not limited to, amino acids found in proteins, *i.e.* glycine, alanine, valine, leucine, isoleucine, methionine, phenylalanine, tryptophan, proline, serine, threonine, cysteine, tyrosine, asparagine, glutamine, aspartate, glutamate, lysine, arginine and histidine. Alternatively, the amino acid can be a derivative of alanyl, valinyl, leucinyl, isoleucinyl, prolinyl, phenylalaninyl, tryptophanyl, methioninyl, glycyl, serinyl, threoninyl, cysteinyl, tyrosinyl, asparaginyl, glutaminyl, aspartoyl, glutaroyl, lysinyl, argininyl, histidinyl, β -alanyl, β -valinyl, β -leucinyl, β -isoleucinyl, β -prolinyl, β -phenylalaninyl, β -tryptophanyl, β -methioninyl, β -glycyl, β -serinyl, β -threoninyl, β -cysteinyl, β -tyrosinyl, β -asparaginyl, β -glutaminyl, β -aspartoyl, β -glutaroyl, β -lysinyl, β -argininyl or β -histidinyl. When the term amino acid is used, it is considered to be a specific and independent disclosure of each of the esters of α , β , γ , and δ glycine, alanine, valine, leucine, isoleucine, methionine, phenylalanine, tryptophan, proline, serine, threonine, cysteine, tyrosine, asparagine, glutamine, aspartate, glutamate, lysine, arginine and histidine in the D and L-configurations.

[0032] The term “blood-brain barrier” or “BBB” refers to the physiological barrier between the peripheral circulation and the brain and spinal cord which is formed by tight junctions within the brain capillary endothelial plasma membranes, creating a tight barrier that restricts the transport of molecules into the brain, even very small molecules such as urea (60 Daltons). The BBB within the brain, the blood-spinal cord barrier within the spinal cord, and the blood-retinal barrier within the retina are contiguous capillary barriers within the CNS and are herein collectively referred to as the blood-brain barrier or BBB. The BBB also encompasses the blood-CSF barrier (choroid plexus) where the barrier is comprised of ependymal cells rather than capillary endothelial cells.

[0033] The term “blood-brain barrier receptor” (abbreviated “R/BBB” herein) is an extracellular membrane-linked receptor protein expressed on brain endothelial cells which is capable of transporting molecules across the BBB or be used to transport exogenous administered molecules. Examples of R/BBB herein include transferrin receptor (TfR), insulin receptor, insulin-like growth factor receptor (IGF-R), low density lipoprotein receptors including without limitation low density lipoprotein receptor-related protein 1 (LRP1) and low density lipoprotein receptor-related protein

8 (LRP8), and heparin-binding epidermal growth factor-like growth factor (HB-EGF). An exemplary R/BBB herein is transferrin receptor (TfR). The “transferrin receptor” (“TfR”) is a transmembrane glycoprotein (with a molecular weight of about 180,000) composed of two disulphide-bonded sub-units (each of apparent molecular weight of about 90,000) involved in iron uptake in vertebrates. In certain embodiments, the amino acid sequence encoding the blood brain barrier (BBB) penetrating peptide comprises SAQLNTTKPI (SEQ ID NO: 1).

[0034] The term “central nervous system” or “CNS” refers to the complex of nerve tissues that control bodily function and includes the brain and spinal cord.

[0035] As used herein, the terms “comprising,” “comprise” or “comprised,” and variations thereof, in reference to defined or described elements of an item, composition, apparatus, method, process, system, etc. are meant to be inclusive or open ended, permitting additional elements, thereby indicating that the defined or described item, composition, apparatus, method, process, system, etc. includes those specified elements--or, as appropriate, equivalents thereof--and that other elements can be included and still fall within the scope/definition of the defined item, composition, apparatus, method, process, system, etc.

[0036] A “comparison window” refers to a segment of any one of the number of contiguous positions (*e.g.*, least about 10 to about 100, about 20 to about 75, about 30 to about 50, 100 to 500, 100 to 200, 150 to 200, 175 to 200, 175 to 225, 175 to 250, 200 to 225, 200 to 250) in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. In embodiments, a comparison window is the entire length of one or both of two aligned sequences. In embodiments, two sequences being compared comprise different lengths, and the comparison window is the entire length of the longer or the shorter of the two sequences. In embodiments relating to two sequences of different lengths, the comparison window includes the entire length of the shorter of the two sequences. In embodiments relating to two sequences of different lengths, the comparison window includes the entire length of the longer of the two sequences.

[0037] “Diagnostic” or “diagnosed” means identifying the presence or nature of a pathologic condition. Diagnostic methods differ in their sensitivity and specificity. The “sensitivity” of a diagnostic assay is the percentage of diseased individuals who test positive (percent of “true positives”). Diseased individuals not detected by the assay are “false negatives.” Subjects who

are not diseased and who test negative in the assay, are termed “true negatives.” The “specificity” of a diagnostic assay is 1 minus the false positive rate, where the “false positive” rate is defined as the proportion of those without the disease who test positive. While a particular diagnostic method may not provide a definitive diagnosis of a condition, it suffices if the method provides a positive indication that aids in diagnosis. The term “predisposition” as used herein means that a subject does not currently present with the dysfunction, but is liable to be affected by the dysfunction in time. Methods of diagnosis according to the disclosure are useful to confirm the existence of a dysfunction, or predisposition thereto. Methods of diagnosis are also useful in methods for assessment of clinical screening, prognosis, choice of therapy, evaluation of therapeutic benefit, i.e. for drug screening and drug development.

[0038] The term “microglial activation” as used herein can refer to processes associated with innate activation or adaptive activation of the microglia. Such activation may include morphological changes of the microglial cells, including shortening of cellular processes and enlargement of their soma, as well as the release of proinflammatory cytokines and chemokines, reactive oxygen and/or nitrogen intermediates, proteinases and complement proteins, and upregulation of cell surface activation antigens.

[0039] A “neurological disorder” as used herein refers to a disease or disorder which affects the CNS and/or which has an etiology in the CNS. Exemplary CNS diseases or disorders include, but are not limited to, neuropathy, amyloidosis, cancer, an ocular disease or disorder, viral or microbial infection, inflammation, ischemia, neurodegenerative disease, seizure, behavioral disorders, and a lysosomal storage disease. For the purposes of this application, the CNS will be understood to include the eye, which is normally sequestered from the rest of the body by the blood-retina barrier. Specific examples of neurological disorders include, but are not limited to, neurodegenerative diseases (including, but not limited to, Lewy body disease, postpoliomyelitis syndrome, Shy-Draeger syndrome, olivopontocerebellar atrophy, Parkinson's disease, multiple system atrophy, striatonigral degeneration, tauopathies (including, but not limited to, Alzheimer disease and supranuclear palsy), prion diseases (including, but not limited to, bovine spongiform encephalopathy, scrapie, Creutzfeldt-Jakob syndrome, kuru, Gerstmann-Straussler-Scheinker disease, chronic wasting disease, and fatal familial insomnia), bulbar palsy, motor neuron disease, and nervous system heterodegenerative disorders (including, but not limited to, Canavan disease, Huntington's disease, neuronal ceroid-lipofuscinosis, Alexander's disease, Tourette's syndrome,

Menkes kinky hair syndrome, Cockayne syndrome, Halervorden-Spatz syndrome, lafora disease, Rett syndrome, hepatolenticular degeneration, Lesch-Nyhan syndrome, and Unverricht-Lundborg syndrome), dementia (including, but not limited to, Pick's disease, and spinocerebellar ataxia), cancer (e.g. of the CNS and/or brain, including brain metastases resulting from cancer elsewhere in the body).

[0040] The terms “numbered with reference to” or “corresponding to,” when used in the context of the numbering of a given amino acid or polynucleotide sequence, refers to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence.

[0041] The terms “patient” or “individual” or “subject” are used interchangeably herein, and refers to a mammalian subject to be treated, with human patients being preferred. In some cases, the methods of the disclosure find use in experimental animals, in veterinary application, and in the development of animal models for disease, including, but not limited to, rodents including mice, rats, and hamsters, and primates.

[0042] “Percentage of sequence identity” is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. In embodiments, the percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. The term “identical” or percent “identity,” in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*e.g.*, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity over a specified region, *e.g.*, of an entire polypeptide sequence or an individual domain thereof), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using a sequence comparison algorithm or by manual alignment and visual inspection. In embodiments, two sequences are 100% identical.

In embodiments, two sequences are 100% identical over the entire length of one of the sequences (*e.g.*, the shorter of the two sequences where the sequences have different lengths). In embodiments, identity may refer to the complement of a test sequence. In embodiments, the identity exists over a region that is at least about 10 to about 100, about 20 to about 75, about 30 to about 50 amino acids or nucleotides in length. In embodiments, the identity exists over a region that is at least about 50 amino acids or nucleotides in length, or more preferably over a region that is 100 to 500, 100 to 200, 150 to 200, 175 to 200, 175 to 225, 175 to 250, 200 to 225, 200 to 250 or more amino acids or nucleotides in length.

[0043] Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (*see, e.g.*, *Current Protocols in Molecular Biology* (Ausubel et al., eds. 1995 supplement)).

[0044] Non-limiting examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., *Nuc. Acids Res.* 25:3389-3402 (1997) and Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), respectively. BLAST and BLAST 2.0 may be used, with the parameters described herein, to determine percent sequence identity for nucleic acids and proteins. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (NCBI), as is known in the art. An exemplary BLAST algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of

matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. In embodiments, the NCBI BLASTN or BLASTP program is used to align sequences. In embodiments, the BLASTN or BLASTP program uses the defaults used by the NCBI. In embodiments, the BLASTN program (for nucleotide sequences) uses as defaults: a word size (W) of 28; an expectation threshold (E) of 10; max matches in a query range set to 0; match/mismatch scores of 1,-2; linear gap costs; the filter for low complexity regions used; and mask for lookup table only used. In embodiments, the BLASTP program (for amino acid sequences) uses as defaults: a word size (W) of 3; an expectation threshold (E) of 10; max matches in a query range set to 0; the BLOSUM62 matrix (*see* Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1992)); gap costs of existence: 11 and extension: 1; and conditional compositional score matrix adjustment. For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. In embodiments, when using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Preferably, default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters

[0045] As used herein, “recombinant adeno-associated virus (AAV)” or “rAAV vector” is meant a recombinantly-produced AAV or AAV particle that comprises a polynucleotide sequence not of AAV origin (e.g., a polynucleotide comprising a transgene, which may be operably linked to one or more enhancer and/or promoters) to be delivered into a cell, either *in vivo*, *ex vivo*, or *in vitro*. Non-naturally occurring (e.g., chimeric) capsids may be used in the rAAVs.

[0046] The terms “specifically binds to”, “specific for”, “targets” and related grammatical variants refer to that binding which occurs between such paired species as enzyme/substrate, receptor/agonist, antibody/antigen, and lectin/carbohydrate which may be mediated by covalent or non-covalent interactions or a combination of covalent and non-covalent interactions. When the

interaction of the two species produces a non-covalently bound complex, the binding which occurs is typically electrostatic, hydrogen-bonding, or the result of lipophilic interactions. Accordingly, “specific binding” occurs between a paired species where there is interaction between the two which produces a bound complex having the characteristics of an antibody/antigen or enzyme/substrate interaction. In particular, the specific binding is characterized by the binding of one member of a pair to a particular species and to no other species within the family of compounds to which the corresponding member of the binding member belongs. Thus, for example, an antibody typically binds to a single epitope and to no other epitope within the family of proteins. In some embodiments, specific binding between an antigen and an antibody will have a binding affinity of at least 10^{-6} M. In other embodiments, the antigen and antibody will bind with affinities of at least 10^{-7} M, 10^{-8} M to 10^{-9} M, 10^{-10} M, 10^{-11} M, or 10^{-12} M. As used herein, the terms “specific binding” or “specifically binding” when used in reference to the interaction of an antibody and a protein or peptide means that the interaction is dependent upon the presence of a particular structure (i.e., the epitope) on the protein.

[0047] A “therapeutic gene,” “prophylactic gene,” “target polynucleotide,” “transgene,” “gene of interest,” “exogenous gene” and the like generally refer to a gene or genes to be transferred using a vector. Typically, such genes are located within the rAAV vector (which vector is flanked by inverted terminal repeat (ITR) regions and thus can be replicated and encapsidated into rAAV particles). Target polynucleotides can be used to generate rAAV vectors for a number of different applications. Such polynucleotides include, but are not limited to: (i) polynucleotides encoding proteins useful in other forms of gene therapy to relieve deficiencies caused by missing, defective or sub-optimal levels of a structural protein or enzyme; (ii) polynucleotides that are transcribed into anti-sense molecules; (iii) polynucleotides that are transcribed into decoys that bind transcription or translation factors; (iv) polynucleotides that encode cellular modulators such as cytokines; (v) polynucleotides that can make recipient cells susceptible to specific drugs, such as the herpes virus thymidine kinase gene; (vi) polynucleotides for cancer therapy, such as E1A tumor suppressor genes or p53 tumor suppressor genes for the treatment of various cancers; and (vii) polynucleotides for gene editing (e.g., CRISPR). To effect expression of the transgene in a recipient host cell, it is in one embodiment operably linked to a promoter, either its own or a heterologous promoter. A large number of suitable promoters are known in the art, the choice of which depends on the desired level of expression of the target polynucleotide, whether one desires

constitutive expression, inducible expression, cell-specific or tissue-specific expression, etc. The rAAV vector may also contain a selectable marker.

[0048] The term “variant,” when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to a wild type gene. This definition may also include, for example, “allelic,” “splice,” “species,” or “polymorphic” variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. Of particular utility in the disclosure are variants of wild type gene products. Variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes that give rise to variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

[0049] Any compositions or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0050] FIGS. 1A, 1B demonstrate the AAV serotype screening for infectivity in human ad monkey primary microglia identified 8 AAV serotypes with >80% transduction efficiency at MOI of 10^5 vg/cell (>90% efficiency at MOI 10^6 vg/cell, data not shown). Cells in 96-well plate (5×10^3 /well) were infected with indicated AAV carrying CMV-EGFP (10^8 vg per well for 3 wells. At 3 days postinfection (dpi), fluorescent microscopic imaging (10x objective) and flow cytometry were performed. Data represent mean $\pm$ SD (n=3).

[0051] FIGS. 2A- 2C is a diagram of AAV-BM generation and summary of functional test in vitro. FIG. 2A: Reported peptides from BBB-crossing serotypes (AAV-B) were inserted into the heparan sulfate (HS) binding sites (R585-R588) of AAV2 CAP VP1 and underlined within the

colored inserted amino acid sequences. FIG. 2B: Microglia-specific peptide (highlighted red) was inserted into the 455-loop of AAV-PhP.eB. FIG. 2C: Transduction efficiency of AAV-BM in various cell types *in vitro*.

[0052] FIG. 3 demonstrates that R2-MAC shows highest packaging titer but all tested AAV-BM lost various degrees of transduction efficiency in HEK293T cells. Cells in 96-well were transfected with helper vector, transfer vector TP1152 (scAAV-CB-EGFP) and indicated serotype vectors. At 3 days, 10 μ l of crude supernatant was used to infect HEK293T cells (1×10^4 /well of 96-well). At 3 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m. Note: Two clones of each AAV-BM were sequenced and tested for the functional packaging. The number (red) indicates transduction unit (TU/ml).

[0053] FIG. 4 demonstrates that R2-MAC shows almost similar efficiency in transducing mouse neural stem cells (NSC) to the original AAV2-RC serotype but the other three AAV-BM lost the transduction compared with the original AAV serotype (AAV2 and PhP-eB). HEK293T cells in 96-well were transfected with helper vector, transfer vector TP1921 (AAV-Cre-CD63-MG1MG2-MCP) and indicated serotype vector. At 3 days, 10 or 100 μ l of crude supernatant was used to infect cultured NSCs (1×10^4 /well of 96-well) from Ai14 LoxP-STOP-LoxP (LSL)-tdTomato (tdT) mice. At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m.

[0054] FIG. 5 demonstrates that R2-MAC transduces mouse primary microglia but the other three AAV-BM does not. HEK293T cells in 96-well were transfected with helper vector, transfer vector TP1921 (AAV-Cre-CD63-MG1MG2-MCP) and indicated serotype vector. At 3 days, 10 μ l of crude supernatant was used to infect primary microglia from AI14 mice cultured 2 weeks in 96-well). At 6 or 14 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m.

[0055] FIG. 6 demonstrates that all the tested AAV-BM can transduce mouse fibroblast from AI14 mice. HEK293T cells in 96-well were transfected with helper vector, transfer vector TP1921 (AAV-Cre-CD63-MG1MG2-MCP) and indicated serotype vector. At 3 days, 100 μ l of crude supernatant was used to infect mouse skin primary fibroblast (1×10^4 /well of 96-well) from AI14 LSL-tdT mice. At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m.

[0056] FIGS. 7A-7F demonstrate that R2-MAC shows stronger transduction efficiency than AAV2 (normalized by MOI) in human primary microglia to the original AAV2-RC serotype but the other three AAV-BM lost the transduction. HEK293T cells in 10-cm dish were transfected

with helper vector, transfer vector TP1152 (scAAV-CB-GFP, FIGS. 7A-7E) or TP1987 (scAAV-CD68-135-UTR-EGFP, FIG. 7F) and indicated serotype vector. At 5 days, Polyethylene glycol (PEG)/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 1 μ l was used to infect human primary microglia (1×10^4 /well of 96-well). At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m. No toxicity present in all groups.

[0057] FIGS. 8A-8F demonstrate that R2-MAC shows stronger transduction efficiency than AAV2 (normalized by moi) in human primary microglia to the original AAV2-RC serotype but the other three AAV-BM lost the transduction. HEK293T cells in 10-cm dish were transfected with helper vector, transfer vector TP1152 (scAAV-CB-GFP, FIGS. 8A-8E) or TP1987 (scAAV-CD68-135-UTR-EGFP, FIG. 8F) and indicated serotype vector. At 5 days, Polyethylene glycol (PEG)/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 10 μ l was used to infect human primary microglia (1×10^4 /well of 96-well). At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m. No toxicity present in all groups.

[0058] FIGS. 9A-9D demonstrate that R2-MAC shows high transduction efficiency in mouse NSC to AAV-DJ8. HEK293T cells in 10-cm dish were transfected with helper vector, transfer vector TP1987 (scAAV-hCD68-135-UTR-EGFP) and indicated serotype vectors AAV2-MAC and AAV-DJ8. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 1 and 10 μ l of purified virus (in 1x DPBS) was used to infect mouse NSC (1×10^4 /well of 96-well). At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m.

[0059] FIG. 10 demonstrates that R2-MAC crosses BBB and transduces microglia. P23 mice were injected via jugular vein with purified AAV-R2-MAC carrying hCD68-135-UTR-eGFP (2×10^{10} vg per animal). At 2 weeks postinfection, animals were perfused with 4% PAF and brain and other major organs were immunostained with anti-eGFP antibody. Scale bar = 100 μ m. TCP884 and 885 indicate the AI14 animal code.

[0060] FIG. 11 demonstrates that AAV-DJ8 crosses BBB and transduces microglia. P23 mice were injected via jugular vein with purified AAV-DJ8 carrying hCD68-135-UTR-eGFP (2×10^{10} vg per animal). At 2 weeks postinfection, animals were perfused with 4% PAF and brain and other major organs were immunostained with anti-eGFP antibody. Scale bar = 100 μ m.

[0061] FIGS. 12A-12D demonstrate that R2-MAC shows moderate transduction efficiency in mouse neural stem cells (NSC) compared to AAV-DJ8. HEK293T cells in 1x10-cm dish were transfected with helper vector, transfer vector TP1960 (AAV-Cre-GFP-CD63-MG1MG2-MCP) and indicated serotype vectors AAV2-MAC and AAV-DJ8. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 1 and 10 μ l of purified virus (in 1x DPBS) was used to infect mouse NSC (1×10^4 /well of 96-well). At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m.

[0062] FIGS. 13A-13F demonstrate that intracranial microinjection of AAV-R2-MAC induce effective delivery of Cre to microglia (arrows) and neurons but DJ8 transduced only neurons. R2-MAC (FIGS. 13A-13D) or AAV-DJ8 (FIGS. 13E, 13F) carrying AAV-CMV-Cre (TP1960) were injected stereotactically into striatum and hippocampus of adult AI14 mice (3-8 months old). At 2 weeks postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μ m. Dose=2 μ l/point: 4.2e8 vg for R2-MAC and 3.6e9 vg for DJ8. TP1960 AAV-CMV-Cre-GFP-CD63-MG1MG2-MCP.

[0063] FIGS. 14A-14D demonstrate that an I.V. injection of AAV-R2-MAC can cross BBB and transduce microglia (white arrow), neurons and microglia-like (red arrow). R2-MAC (FIGS. 14A-14C) or AAV-DJ8 (FIG. 14D) carrying AAV-CMV-Cre (TP1960) were injected via jugular vein of adult AI14 mice (3-4 months old). At 2 weeks postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μ m.

[0064] FIGS. 15A-15D demonstrate that R2-MAC shows strong transduction efficiency in mouse neural stem cells (NSC). HEK293T cells in 3x10-cm dishes were transfected with helper vector, TP1989 (R2-MAC) and indicated transfer vectors. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 1 and 9 μ l of purified virus (in 1x DPBS) was used to infect mouse NSC (1×10^4 /well of 96-well). At 6 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m.

[0065] FIG. 16 demonstrates that the higher titer of AAV-R2-MAC-CMV-Cre induce extensive transduction throughout the brain, mainly in microglia-like cells (white arrows), neurons, astrocytes based on morphology. HEK293T cells in 3x10-cm dishes were transfected with helper vector, TP1989 (R2-MAC) and TP1921 transfer vector (AAV-CMV-Cre-CD63-MG1MG2-MCP). At 5 days, PEG/Chloroform purification and buffer exchange via

ultrafiltration (30KD) were performed. Around 70 μ l (5.6e13 vg/kg) of purified virus was injected intravenously in adult AI14 mice (TCP90, male, 21g, 2 months old). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μ m.

[0066] FIG. 17 demonstrates that AAV-R2-MAC-CMV-Cre also induces transduction of macrophage in various organs. Various organs of the same animals as described in FIG. 16 were collected for cryosection and tdTomato positive cells without immunostaining were imaged under confocal microscope.

[0067] FIG. 18 demonstrates that AAV-R2-MAC-CMV-Cre induce transduction throughout the brain, in microglia-like cells (white arrows), neurons, astrocytes based on morphology. The same amount of purified virus as described in Fig. 16 was injected intravenously in adult AI14 mice (TCP91, male, 22g, 2 months old). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μ m.

[0068] FIG. 19 demonstrates that extensive tdT expression throughout the brain was not colocalized with IBA1 expression. Around 70 μ l of purified AAV-R2-MAC-CMV-Cre virus (5.6e13 vg/kg) was injected intravenously in adult AI14 mice (TCP92, male, 21.4g, 2 months old). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were immunostained with anti-IBA1 antibody and counterstained with Hoechst and imaged under confocal microscope. Scale bar = 100 μ m.

[0069] FIG. 20 demonstrates that higher titer of AAV-R2-MAC-hHexB-134-UTR-eGFP induces effective transduction in microglia-like cells. HEK293T cells in 3x10-cm dishes were transfected with helper vector, TP1989 (R2-MAC) and TP1986 transfer vector (scAAV-hHexB-134-UTR-eGFP). At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30K) were performed. Around 50 μ l of purified virus (9.6e13 vg/kg) was injected intravenously of adult wildtype mice (Tg26Q115, 25.9g, male, 3 months old). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were immunostained with Chicken anti-GFP and Rabbit anti-GFP antibodies. Scale bar = 100 μ m.

[0070] FIG. 21 demonstrates that inclusion of miRNA 9 target (miR9T) in AAV-R2-MAC-hHexB-134-UTR-eGFP virus induces specific transgene expression in microglia but not neural stem cells (NSC). HEK293T cells in 6-well plate were transfected with helper vector, TP1989

(R2-MAC) and indicated transfer vector. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30K) were performed. Crude or purified viruses were used to treat mouse NSCs or human primary microglia, and at indicated days, confocal images were taken. Scale bar = 100 μ m.

[0071] FIG. 22 demonstrates that inclusion of miRNA 9 target (miR9T) in AAV-PhP.eB-hHexB-134-UTR-eGFP virus induces specific transgene expression in microglia but not neural stem cells (NSC). HEK293T cells in 6-well plate were transfected with helper vector, TP1190 (AAV-PhP.eB) and indicated transfer vector. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. Crude or purified viruses were used to treat mouse NSCs or human primary microglia, and at indicated days, confocal images were taken. Scale bar = 100 μ m.

[0072] FIG. 23 demonstrates that AAV2 mutations at 4YF (Tyrosine to Phenylalanine) plus TV (Threonine to Valine), designated R2e for enhanced AAV2, robustly increase transgene expression in mouse NSC and human microglia but miR9T addition illuminates microglia expression. HEK293T cells in a 6-well plate were transfected with helper vector, indicated capsids and transfer vectors. On day 5, PEG/Chloroform purification and ultrafiltration (30KD) were performed. Crude or purified viruses were used to treat mouse NSCs or human primary microglia, and at indicated days, confocal images were taken. Scale bar = 100 μ m.

[0073] FIGS. 24A-24C demonstrate that R2-MAC mutations at 4YF plus TV (R2e-MAC) robustly increase transgene expression with miR9T in human primary microglia. (FIG. 24A) Combined mutation (4YF/TV) significantly improved the transduction efficiency. (FIG. 24B) R2e-MAC mutants showed similar transduction efficiency to R2e mutant. (FIG. 24C) Diagram of mutation residues. HEK293T cells in 6-well plate were transfected with three plasmids. At day 5, PEG/Chloroform purification and ultrafiltration (30KD) were performed. Crude viruses were used to treat human microglia, and at day 7, confocal images were taken. Scale bar = 100 μ m.

[0074] FIGS. 25A-25E show that the 4x repeated miRNA-9 target site (miR9T) was added downstream of WPRE3 to illuminate microglia expression. (FIG. 25A) Mouse primary microglia validated by immunocytochemistry with anti-IBA1 antibody. (FIG. 25B) R2-MAC-4YF/TV efficient delivery of scAAV-HexB-UTR-EGFP-miR9T-BGH (TP2018) to mouse microglia with

or without miR9T. (FIG. 25C) Diagram of the improved Cre (iCre)-mediated LoxP recombination to induce tdTomato (tdT) expression in Ai14 transgenic mice. (FIGS. 25D, 25E) R2-MAC-4YF/TV increases transgene expression in mouse NSC and human microglia without miR9T (TP2057: scAAV-hCD68-iCre-3xHA-WPRE3-BGH) but miR9T addition illuminates microglia expression (TP2058: scAAV-hCD68-iCre-3xHA-WPRE3-miR9T-BGH). Equal MOI (1.8×10^6) of the purified virus was added to indicated mouse NSC (FIG. 25D) or microglia (FIG. 25E). On day 8, confocal images were taken. Scale bar = 100 μ m.

[0075] FIG. 26 is a schematic representation showing the screening of AAV-B-bc-Library in mouse microglia cultured from LSL-tdT reporter mice (Ai14) identified 7 AAV-B serotypes (highlighted red) that can transduce microglia at the efficiency of >50 positive cells per well of 96-well plate. Four barcoded AAV transfer vectors with microglia-specific promoter (duplicate) and miR9T were packaged into selected AAV-B at a large scale. Viruses were purified by ultracentrifugation and titer was quantified via qPCR assay. Microglia (1×10^4 cells/well) were infected with 1 μ l of purified AAV-B virus. At 1 week postinfection, the positive tdT cells per well were counted under confocal fluorescent microscopy.

[0076] FIG. 27 shows representative images for AAV-B-bc-Library transduction in mouse microglia cultured from LSL-tdT reporter mice (Ai14). Four bar-coded AAV transfer vectors with microglia-specific promoter (duplicate) and miR9T were packaged into selected AAV-B at a large scale. Viruses were purified by ultracentrifugation and titer was quantified via qPCR assay. Microglia (1×10^4 cells/well) were infected with 1 μ l of purified AAV-B virus. At 1 week postinfection, the positive tdT cells per well were imaged and counted under confocal fluorescent microscopy. Scale bar = 100 μ m

[0077] FIGS. 28A-28D demonstrate that addition of miR9T to barcoded (bc) AAV-B library induces specific transgene expression in microglia *in vivo*. (FIG. 28A) Brain injection of mixed 8 AAV serotype (0.3 - 2.5×10^{10} GC, 2 μ l/site) carrying 4 bc iCre-miR9T vectors in LSL-tdT mice (M, 3 months old) induces microglia tdT expression in the injection sites but not contralateral sites. (FIGS. 28B-28D) Representative images in other brain regions showing transgenes mainly in microglia but also some neuron-like cells. On day 7 after injection, brain sections were immunostained with anti-GFAP and anti-RFP (FIG. 28A) or anti-Cre (FIGS. 28B-28D) antibodies, and confocal images were taken. 3V, third ventricles. Scale bar = 100 μ m.

[0078] FIGS. 29A-29D show that the fluorescent immunohistochemistry verifies the specificity of tdT positive cells. Brain sections as described in FIGS. 28A-28D were immunostained with rat anti-RFP antibody and Alexa-680 donkey anti-Raf secondary antibody (FIGS. 29A, 29B) or chicken anti-mCherry antibody and Alexa-594 donkey anti-chicken secondary antibody (FIGS. 29C, 29D). Confocal images were taken. Scale bar = 100 μ m.

[0079] FIGS. 30A-30B demonstrate that brain injection of individual barcoded (bc) AAV-B induces specific iCre transduction in microglia *in vivo*. Stereotactic injection (1 μ l/site) of AAV1/2 (1.51e10 GC), AAV1 (1.35e10GC), AAV6-TM 6.9e10GC) and AAV-DJ8 (1.36e10GC) in LSL-tdT mice (M, 2 months old). On day 14 after injection, brain sections were stained with hoechst only (FIG. 30A) or immunostained with anti-IBA1 (microglia), anti-mCherry (tdT) and anti-GFAP (astrocytes) antibodies (FIG. 30B) and confocal images were taken.

[0080] FIG. 31 shows a further illustration of microglia transduction by AAV-DJ8 as described in FIGS. 30A-30B. Ten of confocal Z-stack images were maximized by projection and overlay of individual channel shows clear colocalization (white arrows) of tdT expression with microglia marker IBA1 but no astrocyte marker GFAP.

[0081] FIG. 32 demonstrates that systemic (i.v., 50 μ l) injection of individual barcoded (bc) AAV-B induces specific iCre transduction in microglia *in vivo*. Jugular vein injection of AAV-GTX (TCP101, F, 9 months old, 25.7g, 4.71e12 GC), AAV-1RX (TCP100, F, 9-month old, 34.8g, 3.55e12GC), AAV-PhP-C2 (TCP102, F, 9 month old, 29g, 3.98e12GC), AAVv66 (TCP105, F, 9 months old, 32.5g, 3.45e12GC), AAV-B1 (TCP109, F, 9 months old, 25.4g, 2.145e12GC), AAV-Rh39 (TCP115, F, 8 month old, 20.5g, 3.65e12GC), AAV-cc47 (TCP122, M, 4 months old, 32.2g, 4.44e12GC) and AAV1_P5 (TCP127, F, 4 months old, 23.2g, 3.85e12GC) in LSL-tdT mice. On day 10 after injection, brain sections were immunostained with anti-GFAP (astrocytes), anti-mCherry (tdT) and anti-IBA1 (microglia) antibodies and confocal images were taken. Ten of confocal Z-stack images were maximized by projection and overlay of individual channel shows clear colocalization of tdT expression with microglia marker IBA1 but no astrocyte marker GFAP.

[0082] FIGS. 33A-33D demonstrate that systemic (i.v., 150 μ l) injection of individual barcoded (bc) AAV-B induces specific iCre transduction in microglia *in vivo*. Jugular vein

injection of AAVRh10 (TCP107, F, 7 months old, 23.2g, 6.435×10^{12} GC), AAV-Pal2 (TCP114, F, 6 months old, 23g, 1.251×10^{13} GC), AAV6-TM (TCP104, F, 7-month old, 26g, 1.065×10^{13} GC) and AAV-HSC16 (TCP108, F, 7-month old, 24.2g, 1.485×10^{13} GC) in LSL-tdT mice. On day 14 after injection, brain sections were immunostained with anti-RFP (tdT) and anti-GFAP (astrocytes) antibodies and confocal images were taken.

[0083] FIG. 34 demonstrates that AAV-CPP21-iCre brain injection induces expression of tdT mainly in neurons but rarely in astrocytes and microglia. Mouse brain stereotactic injection of AAV-CPP21-hCD68-iCre-3xHA-WPRE3 viruses (6.4×10^9 GC, 2 μ l/site). On day 7, brain sections were immunostained with anti-GFAP (astrocytes) and anti-IBA1 (microglia) antibodies, and confocal images were taken.

[0084] FIG. 35 demonstrates that large scale of packaging and purification of R2e-MAC with different transfer vector retain high transduction in microglia. Two CDMO companies performed large-scale purification and obtained similar yield although lower than the standard AAV2 system (2-5x). SIM-A9 microglia (4×10^4 cells/well) were treated with 10 μ l of purified R2e-MAC viruses and fluorescent images were taken at day 3 after infection. Scale bar = 100 μ m.

[0085] FIGS. 36A-36B demonstrate that systemic (i.v., 150 μ l) injection of AAVone-R2e-MAC-scAAV-CMV-EGFP (3.66×10^{13} GC/ml) in C57BL6 mice (22g, male) induced extensive transduction mainly in microglia-like cells and rarely in astrocytes and neurons. On day 7 after injection, brain sections were immunostained with anti-eGFP, anti-IBA (microglia) and anti-GFAP (astrocytes) antibodies and representative confocal tiling images at 10x objective are shown. Each inset is one of the tiled individual images.

[0086] FIG. 37 is a series of representative confocal Z-stack images demonstrating extensive transduction in microglia-like cells as well as some neurons and astrocytes after i.v. injection as described in FIGS. 36A, 36B. The inset of hippocampus shows colocalization of microglia-like cells with IBA1+ cells in most cases (white arrows). Red arrow shows astrocyte colocalization. Maximized images from 12 Z-stacks at interval thickness of 1 μ m.

[0087] FIG. 38 is a series of representative confocal Z-stack images showing extensive transduction in microglia-like cells as well as some neurons and astrocytes in pre-cortex and olfactory tubercle after i.v. injection as described in FIGS. 36A, 36B. White arrows show the colocalization of eGFP-labeled microglia-like cells with IBA1+ cells in most cases. Red arrow

shows astrocyte colocalization. Maximized images from 12-14 Z-stacks at interval thickness of 1 μm .

[0088] FIG. 39 is a series of representative confocal Z-stack images showing extensive transduction in microglia-like cells as well as some neurons and astrocytes in amygdala and lateral striatum after i.v. injection as described in FIGS. 36A-36B. White arrows show the colocalization of eGFP-labeled microglia-like cells with IBA1+ cells in most cases. Red arrow shows astrocyte colocalization. Maximized images from 12-14 Z-stacks at interval thickness of 1 μm .

[0089] FIG. 40 shows the AAV-R2-MAC sequence (SEQ ID NO: 2) and the plasmid encoding AAV-R2-MAC.

[0090] FIG. 41 shows the R2-MAC, REP amino acid sequence (SEQ ID NO: 3) and the R2-MAC, VP1(SAQ-MAC) protein (SEQ ID NO: 4).

[0091] FIG. 42 shows the AAV-R2e-MAC nucleotide sequence (enhanced version of R2-MAC with 4 YF mutations and 1 TV mutation; SEQ ID NO: 5) and the plasmid encoding AAV-R2e-MAC.

[0092] FIG. 43 shows the R2e-MAC, REP amino acid sequence (SEQ ID NO: 6) and the R2e-MAC, VP1 (4YF/TV and SAQ-MAC) protein (SEQ ID NO: 7).

[0093] FIG. 44 shows the nucleic acid sequence of TP2018 (ITR-pdsAAV-hHexB-134-UTR-EGFP-4xmiR9T-BGH-ITR; SEQ ID NO: 8) and the TP2018 plasmid.

[0094] FIG. 45 the AAV barcode reporter vector (ITR-pDsAAV-hHexB134-UTR-NLS-iCre-HA-WPRE3-bc25-miR9T-bGH-ITR; SEQ ID NO: 9) nucleotide sequence and the plasmid encoding the AAV barcode reporter.

[0095] FIG. 46 the AAV barcode reporter vector (pDsAAV-hCD68-135-UTR-NLS-iCre-HA-WPRE3-bc25-miR9T-bGH-ITR; SEQ ID NO: 10) and the plasmid encoding the AAV barcode reporter.

DETAILED DESCRIPTION

[0096] Many viral and non-viral delivery technologies for gene and cell therapy have been successfully applied to the genome manipulation *in vitro* or *ex vivo* with relatively easy

transfection and high delivery efficiency. Microglia (MG) play essential roles in the pathogenesis of many neurodegenerative diseases and aging. Highly specific targeting on MG is also required for potential novel therapeutics for these diseases. However, the *in vivo* application of currently available deliveries to the cells in animals and human, particularly the myeloid cells (residual microglia and perivascular macrophages) in the central nervous system (CNS), is very limited. Among the viral systems, AAVs are preferred because of their large potential in ongoing clinical trials with thousands of cases, high safety, low immunogenicity, and broad targeting capsid serotypes. Insufficient and non-specific AAV transduction is the major limiting factor for AAV clinical gene therapy.

[0097] Most importantly, we have shown AAV delivery of CRISPR/Cas genome editor to eradicate HIV provirus in humanized mouse and non-human primate infection models, leading to Phase 1/2 clinical trial of AAV9-saCas9 HIV eradicator.

[0098] Increasing evidence has demonstrated that MG plays essential roles in the pathogenesis of many neurodegenerative diseases and aging. Highly specific targeting on MG is also required for potential novel therapeutics for these diseases.

[0099] Accordingly, this disclosure provides novel AAV vectors for the treatment of HIV-Associated Neurocognitive Disorder (HAND) and other neurological disorders. In certain embodiments, the AAVs are engineered to specifically target the microglia. Microglia are a population of Central Nervous System (CNS) cells that serve an immune function. According to Lopez, *et al.*, (Microglia: The Real Foe in HIV-Associated Neurocognitive Disorders? *Biomedicines*. 2021 Jul 30;9(8):925. doi: 10.3390/biomedicines9080925. PMID: 34440127; PMCID: PMC8389599), microglia are the principal targets of HIV infection in the CNS. Development of therapies which can (i) cross the Blood-Brain Barrier (BBB), and (ii) specifically target infected microglia will go some way toward treatment of HAND, and will directly address the pathophysiology of the disease, unlike the current standard of care. HAND, according to a February 2022 study may have a prevalence of up to 42.6% among HIV patients receiving HAART. A 2020 meta-analysis of 5,588 records revealed an estimated 16,145,400 patients suffer with HAND worldwide. This disclosure also provides targeting the root cause of cognitive defects: residual components of the HIV genome in the microglia.

[00100] The instant disclosure also provides for Adeno-Associated Virus serotype 2 (AAV2) that has been mutated to enforce high level expression of the target genes while at the same time restricting expression to only the microglia. In the clinic, intravenous administration of purified virus will lead to migration to within the CNS space, where individual virus particles will attempt to infect all cell types. Infection will be limited to microglia based upon modifications made to the virus capsid protein made by the investigators. Infection of target cells harboring integrated HIV sequences will then engage the Cas9/CRISPR system which will selectively excise the HIV sequences. Descendants of the microglial clones harboring HIV sequences will have had their genomes edited to lack HIV sequences, and this reduce the chance of neuropathology.

[00101] ADENO-ASSOCIATED VIRUSES (AAVs)

[00102] Adeno-associated viruses (AAVs) are small, non-pathogenic, ssDNA packaging viruses, capable of infecting a wide range of vertebrate hosts, including humans. AAVs belong to the Parvovirinae subfamily of the Parvoviridae, and Dependoparvovirus genus. As the name implies, they require co-infection with adeno- or herpesviruses as helpers for replication (Cotmore, S. F. *et al.* ICTV virus taxonomy profile: parvoviridae. *J. Gen. Virol.* 100, 367–368 (2019); Buller, R. M. & Rose, J. A. Characterization of adenovirus-associated virus-induced polypeptides in KB cells. *J. Virol.* 25, 331–338 (1978); McPherson, R. A., *et al.*, Human cytomegalovirus completely helps adeno-associated virus replication. *Virology* 147, 217–222 (1985); Weindler, F. W. & Heilbronn, R., A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. *J. Virol.* 65, 2476–2483 (1991); Weitzman, M. D. & Linden, R. M. Adeno-associated virus biology. *Methods Mol. Biol.* 807, 1–23 (2011)). AAVs package a 4.7 kb genome encoding non-structural (rep), structural (cap), assembly activating (aap), and membrane associated accessory (maap) proteins (Wistuba, A., *et al.*, Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins. *J. Virol.* 69, 5311–5319 (1995); Sonntag, F., Schmidt, K. & Kleinschmidt, J. A. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc. Natl Acad. Sci. USA* 107, 10220–10225 (2010); Ogden, P. J. *et al.*, AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science* 366, 1139–1143 (2019)). AAVs have become widely used for gene therapy applications, with several advantages over other viral vectors, including a lower toxicity and the availability of over 150 naturally occurring genotypes and serotypes (Zinn, E. & Vandenberghe, L. H. Adeno-associated virus: fit to serve. *Curr. Opin. Virol.* 8, 90–97 (2014);

Gao, G. P. *et al.* Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc. Natl Acad. Sci. USA* 99, 11854–11859 (2002); Gao, G. *et al.* Clades of adeno-associated viruses are widely disseminated in human tissues. *J. Virol.* 78, 6381–6388 (2004)). These serotypes differ in their tropism, and thus can target most tissues and cell types for gene delivery (Wu, Z., Asokan, A. & Samulski, R. J. Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol. Ther.* 14, 316–327 (2006)). Recombinant AAVs (rAAVs), packaging a gene of interest (GOI), have been successfully studied in clinical trials for the treatment of a wide variety of rare genetic disorders. Notably, three AAV-based biologics have been approved: Luxturna by the FDA and EMA (FDA STN#125610; EMA/H/C/004451), Zolgensma by the FDA (FDA STN #125694), and Glybera by the EMA; and several other products are presently being reviewed (clinicaltrials.gov/) Naso, M. F. *et al.* Adeno-associated virus (AAV) as a vector for gene therapy. *BioDrugs* 31, 317–334 (2017)). In these applications, a gene of interest (GOI) replaces the natural AAV genome for delivery to tissues or cells to treat a monogenic disease. In addition, rAAVs are widely used research tools for transgene expression in tissue culture and preclinical animal models (Kimura, T., *et al.* Production of adeno-associated virus vectors for in vitro and in vivo applications. *Sci. Rep.* 9 (2019), doi.org/10.1038/s41598-019-49624-w)).

[00103] AAV capsids consist of a total of 60 molecules of viral proteins (VPs); a mixture of the three overlapping gene products, VP1, VP2, and VP3, encoded by the cap open reading frame (ORF) and organized in T = 1 icosahedral symmetry (Chapman, M. & Agbandje-Mckenna, M., In Parvoviruses. 107–123 (CRC Press, 2005; crcnetbase.com/doi/10.1201/b13393-13)). The VPs are generated through alternative splicing of the mRNA and use of an alternate translational start codon (Srivastava, A *et al.* Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J. Virol.* 45, 555–564 (1983)). The VP3 (59–61 kDa, 524–544aa) sequence is shared among all VPs and is referred to as the VP3 common region. VP2 (64–67 kDa, 580–601aa) is approximately 57aa longer than VP3 and the VP2 N-terminal region is referred to as the VP1/VP2 common region. VP1 (79–82 kDa, 713–738aa) is approximately 137 aa longer than VP2 and this region is called the VP1 unique (VP1u) region. The VP3 common region assembles the icosahedral capsid. The VP1u contains an essential phospholipase A2 (PLA2) enzyme, and VP1u and VP1/VP2 common region contain nuclear localization sequences (NLSs) (Sonntag, F *et al.* Adeno-associated virus type 2 capsids with externalized VP1/VP2 trafficking domains are generated prior

to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus. *J. Virol.* 80, 11040–11054 (2006)). These N-terminal extensions of VP1 and VP2 are reported to play crucial roles in endosomal trafficking and escape, nuclear localization, and genome release (Agbandje-McKenna, M. & Kleinschmidt, J., *In Methods in Molecular Biology*. 47–92 (2012)). Specific VP subunits have been targeted for modification, such as the removal of common immunogenic motifs, the integration of sequences encoding for fluorophores or designed nanobodies against receptors on target cells, redirecting tropism (Eichhoff, A. M. *et al.* Nanobody-enhanced targeting of AAV gene therapy vectors. *Mol. Ther.* 15, 211–220 (2019)). As some of these modifications are directed to VP1, and because the unique portions of VP1 and VP1/VP2 are required for infection, it is important to assess how many VP1 and VP2 molecules are in the AAV icosahedral capsid and how they are, if at all, structurally organized. (Wörner, T.P., Bennett, A., Habka, S. *et al.* Adeno-associated virus capsid assembly is divergent and stochastic. *Nat Commun* 12, 1642 (2021). doi.org/10.1038/s41467-021-21935-5)).

[00104] In certain embodiments, the disclosure provides an adeno-associated virus (AAV) comprising a nucleic acid sequence encoding: (i) a capsid protein comprising a blood brain barrier penetrating peptide, (ii) a microglial specific peptide, (iii) a microglial tropic promoter and (iv) a transgene. In certain embodiments, the AAV further comprises a woodchuck hepatitis virus post-transcriptional response element (WPRE). Adeno-associated virus is a member of the Parvoviridae family comprises a linear, single-stranded DNA genome of less than about 5,000 nucleotides. AAV requires co-infection with a helper virus (i.e., an adenovirus or a herpes virus), or expression of helper genes, for efficient replication. AAV vectors used for administration of therapeutic nucleic acids typically have approximately 96% of the parental genome deleted, such that only the terminal repeats (ITRs), which contain recognition signals for DNA replication and packaging, remain. This eliminates immunologic or toxic side effects due to expression of viral genes. In addition, delivering specific AAV proteins to producing cells enables integration of the AAV vector comprising AAV ITRs into a specific region of the cellular genome, if desired (see, e.g., U.S. Pat. Nos. 6,342,390 and 6,821,511). Host cells comprising an integrated AAV genome show no change in cell growth or morphology (see, for example, U.S. Pat. No. 4,797,368).

[00105] The AAV ITRs flank the unique coding nucleotide sequences for the non-structural replication (Rep) proteins and the structural capsid (Cap) proteins (also known as virion proteins (VPs)). The terminal 145 nucleotides are self-complementary and are organized so that an

energetically stable intramolecular duplex forming a T-shaped hairpin may be formed. These hairpin structures function as an origin for viral DNA replication by serving as primers for the cellular DNA polymerase complex. The Rep genes encode the Rep proteins Rep78, Rep68, Rep62, and Rep40. Rep78 and Rep68 are transcribed from the p5 promoter, and Rep 52 and Rep40 are transcribed from the p19 promoter. The Rep78 and Rep68 proteins are multifunctional DNA binding proteins that perform helicase and nickase functions during productive replication to allow for the resolution of AAV termini (see, e.g., Im *et al.*, *Cell*, 51:447 (1990)). These proteins also regulate transcription from endogenous AAV promoters and promoters within helper viruses (see, e.g., Pereira *et al.*, *J. Virol.*, 71:1079 (1997)). The other Rep proteins modify the function of Rep78 and Rep68. The cap genes encode the capsid proteins VP1, VP2, and VP3. The cap genes are transcribed from the p40 promoter.

[00106] The AAV vector may be generated using any AAV serotype known in the art. Several AAV serotypes and over 100 AAV variants have been isolated from adenovirus stocks or from human or nonhuman primate tissues (reviewed in, e.g., Wu *et al.*, *Molecular Therapy*, 14(3): 316 (2006)). Generally, the AAV serotypes have genomic sequences of significant homology at the nucleic acid sequence and amino acid sequence levels, such that different serotypes have an identical set of genetic functions, produce virions which are essentially physically and functionally equivalent and replicate and assemble by practically identical mechanisms. AAV serotypes 1-5 and 7-9 are defined as “true” serotypes, in that they do not efficiently cross-read with neutralizing sera specific for all other existing and characterized serotypes. In contrast, AAV serotypes 6, 10 (also referred to as Rh10), and 11 are considered “variant” serotypes as they do not adhere to the definition of a “true” serotype. AAV serotype 2 (AAV2) has been used extensively for gene therapy due to its lack of pathogenicity, wide range of infectivity, and ability to establish long-term transgene expression (see, e.g., Carter, *Hum. Gene Ther.*, 15:541 (2005)). Genome sequences of various AAV serotypes and comparisons thereof are disclosed in, for example, GenBank Accession numbers U89790, J01901, AF043303, and AF085716: Chiorini *et al.*, *J. Virol.*, 71:6823 (1997); Srivastava *et al.*, *J. Virol.*, 45:565 (1983); Chiorini *et al.*, *J. Virol.*, 23:1309 (1999); Rutledge *et al.*, *J. Virol.*, 22:309 (1998); and Wu *et al.*, *J. Virol.*, 24:8635 (2000)).

[00107] AAV rep and ITR sequences are particularly conserved across most AAV serotypes. For example, the Rep78 proteins of AAV2, AAV3A, AAV3B, AAV4, and AAV6 are reportedly about 89-93% identical (see Bantel-Schaal *et al.*, *J. Virol.*, 73(2):939 (1999)). It has been reported that

AAV serotypes 2, 3A, 3B, and 6 share about 82% total nucleotide sequence identity at the genome level (Bantel-Schaal *et al.*, *supra*). Moreover, the rep sequences and ITRs of many AAV serotypes are known to efficiently cross-complement (e.g., functionally substitute) corresponding sequences from other serotypes during production of AAV particles in mammalian cells.

[00108] Generally, the cap proteins, which determine the cellular tropicity of the AAV particle, and related cap protein-encoding sequences, are significantly less conserved than Rep genes across different AAV serotypes. In view of the ability Rep and ITR sequences to cross-complement corresponding sequences of other serotypes, the AAV vector can comprise a mixture of serotypes and thereby be a 'chimeric' or 'pseudotyped' AAV vector. A chimeric AAV vector typically comprises AAV capsid proteins derived from two or more (e.g., 2, 3, 4, etc.) different AAV serotypes. In contrast, a pseudotyped AAV vector comprises one or more ITRs of one AAV serotype packaged into a capsid of another AAV serotype. Chimeric and pseudotyped AAV vectors are further described in, for example, U.S. Pat. No. 6,723,551; Flotte, *Mol. Ther.*, 12(1):1 (2006); Gao *et al.*, *J. Virol.*, 78:6381 (2004); Gao *et al.*, *Proc. Natl. Acad. Sci. USA*, 23:11854 (2002); De *et al.*, *Mol. Ther.*, 12:67 (2006); and Gao *et al.*, *Mol. Ther.*, 12:77 (2006). In certain embodiments, the AAV comprises one or more mutations in the capsid protein at one or more locations in the capsid protein.

[00109] In one embodiment, the AAV vector is generated using an AAV that infects humans (e.g., AAV2). Alternatively, the AAV vector is generated using an AAV that infects non-human primates, such as, for example, the great apes (e.g., chimpanzees). Old World monkeys (e.g., macaques) and New World monkeys (e.g., marmosets). In one embodiment, the AAV vector is generated using an AAV that infects a non-human primate pseudotyped with an AAV that infects humans. Examples of such pseudotyped AAV vectors are disclosed in, e.g., Cearley *et al.*, *Molecular Therapy*, 12:528 (2006).

[00110] In addition to the nucleic acid sequence encoding a therapeutic gene, the AAV vector may comprise expression control sequences, such as promoters, enhancers, polyadenylation signals, transcription terminators, internal ribosome entry sites (IRES), and the like, that provide for the expression of the nucleic acid sequence in a host cell. Exemplary expression control sequences are known in the art and described in, for example, Goeddel, *Gene Expression Technology. Methods in Enzymology*. Vol. 185. Academic Press. San Diego. Calif. (1990). In

certain embodiments, the AAV further comprises a woodchuck hepatitis virus post-transcriptional response element (WPRE). The woodchuck hepatitis virus post-transcriptional response element (WPRE) is thought to promote nuclear export of viral genomic mRNA (Donello JE et al. Woodchuck hepatitis virus contains a tripartite posttranscriptional regulatory element. J Virol. 1998 Jun;72(6):5085-92). As a result, the WPRE is often included in viral gene transfer vectors where it can improve viral titer and subsequent transgene expression (Zufferey R *et al.* Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. J Virol. 1999 Apr;73(4):2886-922).

[00111] A large number of promoters, including constitutive, inducible, and repressive promoters, from a variety of different sources are well known in the art. Representative sources of promoters include for example, virus, mammal, insect, plant, yeast, and bacteria, and suitable promoters from these sources are readily available, or can be made synthetically, based on sequences publicly available, for example, from depositories such as the ATCC as well as other commercial or individual sources. Promoters can be unidirectional (i.e., initiate transcription in one direction) or bi-directional (i.e., initiate transcription in either a 3' or 5' direction). Non limiting examples of promoters include, for example, the T7 bacterial expression system, pBAD (araA) bacterial expression system, the cytomegalovirus (CMV) promoter, the SV40 promoter, and the RSV promoter. Inducible promoters include, for example, the Tet system (U.S. Pat. Nos. 5,464,758 and 5,814,618), the Ecdysone inducible system (No et al., Proc. Natl. Acad. Sci., J2:3346 (1996)), the T-REX™ system (Invitrogen, Carlsbad, Calif.), LACSWITCH™ System (Stratagene, San Diego, Calif.), and the Cre-ERT tamoxifen inducible recombinase system (Indra *et al.*, *Nuc. Acid. Res.*, 22:4324 (1999); *Nuc. Acid. Res.*, 28:e99 (2000); U.S. Pat. No. 7,112,715; and Kramer & Fuseenegger. *Methods Mol. Bid.*, 308:123 (2005)).

[00112] Typically, AAV vectors are produced using well characterized plasmids. For example, human embryonic kidney 293T cells are transfected with one of the transgene specific plasmids and another plasmid containing the adenovirus helper and AAV rep and cap genes. After 72 hours, the cells are harvested, and the vector is released from the cells by five freeze/thaw cycles. Subsequent centrifugation and berizonase treatment remove cellular debris and unencapsidated DNA. Iodixanol gradients and ion exchange columns may be used to further purify each AAV vector. Next the purified vector is concentrated by a size exclusion centrifuge spin column to the required concentration. Finally, the buffer is exchanged to create the final vector products

formulated (for example) in 1X phosphate buffered saline. The viral titers may be measured by TaqMan™ real-time PCR and the viral purity may be assessed by SDS-PAGE.

[00113] MODIFIED AAV CAPSID PROTEINS AND VIRUS CAPSIDS AND VIRUS VECTORS

[00114] The present disclosure provides AAV capsid proteins comprising at least one mutation (i.e., a modification, which can be a substitution or an insertion or a deletion) in the amino acid sequence and virus capsids and virus vectors comprising the modified AAV capsid protein. The mutated AAV capsid proteins, and the nucleic acids encoding them, are not found in nature (i.e., are non-natural) and have neither the sequence of the wild-type sequences found in nature nor the function of those sequences. In certain embodiments, the modifications at the one or more amino acid positions described can confer one or more desirable properties to virus vectors comprising the modified AAV capsid protein including without limitation: (i) enhancing the penetration of the blood-brain barrier; (ii) enhancing tropism to specifically target a desired cell, e.g. microglia; (iii) enhancing intracellular trafficking to a target cell nucleus; (iv) enhancing the interactions with target cells; (v) enhancing penetration through the mucus gel layer.

[00115] In certain embodiments, the tropism is specific for microglial cells. Any microglial marker can be targeted. *See*, for example Table 1.

[00116] In certain embodiments, the modified AAV capsid protein of the disclosure comprises one or more mutations (i.e., modifications) in the amino acid sequence of the native, capsid protein, such as, for example, a blood brain barrier penetrating. Those skilled in the art will appreciate that for some AAV capsid proteins the corresponding modification will be an insertion and/or a substitution, depending on whether the corresponding amino acid positions are partially or completely present in the virus or, alternatively, are completely absent.

[00117] As used herein, a “mutation” or “modification” in an amino acid sequence includes substitutions, insertions and/or deletions, each of which can involve one, two, three, four, five, six, seven, eight, nine, ten or more amino acids. In certain embodiments, the modification is a substitution. In other embodiments, the modification is an insertion e.g., a blood brain barrier penetrating peptide. In certain embodiments, the modification is a deletion. In certain embodiments, the modification is a modified amino acid.

[00118] It is to be understood that the substitutions and insertions described in the AAV capsid proteins of this disclosure can include substitutions and/or insertions with conservative amino acid residues. Such conservative substitutions are well known in the art and include, e.g., nonpolar amino acids Gly, Ala, Val, Leu, Ile, Met, Phe, Trp and Pro can be substituted for one another; polar amino acids Ser, Thr, Cys, Tyr, Asn and Gln can be substituted for one another; negatively charged amino acids Asp and Glu can be substituted for one another; and positively charged amino acids Lys, Arg and His can be substituted for one another, in any combination.

[00119] The present disclosure also provides an AAV capsid comprising one or more mutations as well as a virus vector comprising one or more mutations. In certain embodiments, recombinant AAV genomes comprise, an adeno-associated virus (AAV) comprises a nucleic acid sequence encoding: (i) a capsid protein comprising a blood brain barrier penetrating peptide (the capsid can be derived from pseudotyped AAV or AAV serotypes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, DJ or DJ/8), (ii) a microglial specific peptide, (iii) a microglial tropic promoter, (iv) a transgene, (v) a woodchuck hepatitis virus post-transcriptional response element (WPRE) and one or more AAV ITRs flanking a nucleic acid molecule. In some embodiments, the capsid protein and microglial specific peptide are related to the AAV serotype. The AAV DNA in the rAAV genomes may be from any AAV serotype for which a recombinant virus can be derived including, but not limited to, AAV serotypes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, DJ or DJ/8. Production of pseudotyped rAAV is disclosed in, for example, WO 01/83692. Other types of rAAV variants, for example rAAV with capsid mutations, are also contemplated. See, for example, Marsic *et al.*, *Molecular Therapy*, 22(11): 1900-1909 (2014).

[00120] In certain embodiments, an AAV transfer vector further comprises a microglial tropic promoter, (iv) a transgene, (v) a woodchuck hepatitis virus post-transcriptional response element (WPRE) and one or more AAV ITRs flanking a nucleic acid molecule.

[00121] Techniques to produce rAAV particles, in which an AAV genome to be packaged, rep and cap genes, and helper virus functions are provided, are standard in the art. Production of rAAV requires that the following components are present within a single cell (denoted herein as a packaging cell): a rAAV genome, AAV rep and cap genes separate from the rAAV genome, and helper virus functions. The AAV rep and cap genes may be from any AAV serotype for which recombinant virus can be derived and may be from a different AAV serotype than the rAAV

genome ITRs, including, but not limited to, AAV serotypes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, DJ or DJ/8. Production of pseudotyped rAAV is disclosed in, for example, WO 01/83692 which is incorporated by reference herein in its entirety.

[00122] Methods of generating a packaging cell comprise creating a cell line that stably expresses all the necessary components for AAV particle production. For example, a plasmid (or multiple plasmids) comprising a rAAV genome lacking AAV rep and cap genes, AAV rep and cap genes separate from the rAAV genome, and a selectable marker, such as a neomycin resistance gene, are integrated into the genome of a cell. AAV genomes have been introduced into bacterial plasmids by procedures such as GC tailing (Samulski et al., 1982, Proc. Natl. Acad. Sci. USA, 79: 2077-2081), addition of synthetic linkers containing restriction endonuclease cleavage sites (Laughlin et al., 1983, Gene, 23: 65-73) or by direct, blunt-end ligation (Senapathy & Carter, 1984, J. Biol. Chem., 259: 4661-4666). The packaging cell line is then infected with a helper virus such as adenovirus. The advantages of this method are that the cells are selectable and are suitable for large-scale production of rAAV. Other examples of suitable methods employ adenovirus or baculovirus rather than plasmids to introduce rAAV genomes and/or rep and cap genes into packaging cells.

[00123] In one embodiment packaging cells may be stably transformed cancer cells such as HeLa cells, 293 cells and PerC.6 cells (a cognate 293 line). In another embodiment, packaging cells are cells that are not transformed cancer cells, such as low passage 293 cells (human fetal kidney cells transformed with EI of adenovirus), MRC-5 cells (human fetal fibroblasts), WI-38 cells (human fetal fibroblasts), Vero cells (monkey kidney cells) and FRhL-2 cells (rhesus fetal lung cells).

[00124] In certain embodiments, an AAV-R2-MAC(AAV2-REP-link-VP1(SAQ-MAC)) nucleotide sequence comprises SEQ ID NO: 2:

[00125] ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAG
CATCTGCCCCGGCATTCTTGACAGCTTTGTGAAGTGGGTGGCCGAGAAGGAATGGGA
GTTGCCCGCCAGATTCTGACATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGT
GGCCGAGAAGCTGCAGCGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCC
CGGAGGCCCTTTTCTTTGTGCAATTTGAGAAGGGAGAGAGCTACTTCCACATGCACG
TGCTCGTGGAACACCGGGGTGAAATCCATGGTTTTTGGGACGTTTCTGAGTCAGA
TTCGCGAAAACTGATTCAGAGAATTTACCGCGGGATCGAGCCGACTTTGCCAACT
GGTTCGCGGTCACAAAGACCAGAAATGGCGCCGGAGGCGGGAACAAGGTGGTGGGA
TGAGTGCTACATCCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGC
GTGGACTAATATGGAACAGTATTTAAGCGCCTGTTTGAATCTCACGGAGCGTAAACG

GTTGGTGGCGCAGCATCTGACGCACGTGTGCGCAGACGCAGGAGCAGAACAAAGAGA
ATCAGAATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAACCTTCAGCCAGGTACA
TGGAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATC
CAGGAGGACCAGGCCTCATACATCTCCTTCAATGCGGCCTCCAACCTCGCGGTCCCAA
ATCAAGGCTGCCTTGGACAATGCGGGAAAGATTATGAGCCTGACTAAAACCGCCCC
CGACTACCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAA
AATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATG
GGCCACGAAAAAGTTCCGGCAAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTA
CCGGGAAGACCAACATCGCGGAGGCCATAGCCCACACTGTGCCCTTCTACGGGTGC
GTAAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTGCGACAAGATGGTGATC
TGGTGGGAGGAGGGGAAGATGACCGCCAAGGTCGTGGAGTCGGCCAAAGCCATTCT
CGGAGGAAGCAAGGTGCGCGTGGACCAGAAATGCAAGTCCTCGGCCCCAGATAGACC
CGACTCCCGTGATCGTCACCTCCAACACCAACATGTGCGCCGTGATTGACGGGAACT
CAACGACCTTCGAACACCAGCAGCCGTTGCAAGACCGGATGTTCAAATTTGAACTC
ACCCGCCGTCTGGATCATGACTTTGGGAAGGTCACCAAGCAGGAAGTCAAAGACTT
TTTCCGGTGGGCAAAGGATCACGTGGTTGAGGTGGAGCATGAATTCTACGTCAAAA
AGGGTGGAGCCAAGAAAAGACCCGCCCCAGTGACGCAGATATAAGTGAGCCCAA
ACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAGACGCGGAAGCTTCGATCA
ACTACGCAGACAGGTACCAAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGATG
CTGTTTCCCTGCAGACAATGCGAGAGAATGAATCAGAATTCAAATATCTGCTTCACT
CACGGACAGAAAGACTGTTTAGAGTGCTTTCCCGTGTGAGAATCTCAACCCGTTTCT
GTCGTCAAAAAGGCGTATCAGAAACTGTGCTACATTCATCATATCATGGGAAAGGT
GCCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCATCTT
TGAACAATAAatgatttaaatcaggtATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGA
CACTCTCTCTGAAGGAATAAGACAGTGGTGGAAGCTCAAACCTGGCCCCACCACCAC
CAAAGCCCCGCAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCCTGGGTAC
AAGTACCTCGGACCCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCGAGA
CGCCGCGGCCCTCGAGCACGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACA
ACCCGTACCTCAAGTACAACCACGCCGACGCGGAGTTTCAGGAGCGCCTTAAAGAA
GATACGTCTTTTGGGGGGCAACCTCGGACGAGCAGTCTTCCAGGCGAAAAAGAGGGT
TCTTGAACCTCTGGGCCTGGTTGAGGAACCTGTTAAGACGGCTCCGGGAAAAAAGA
GGCCGGTAGAGCACTCTCCTGTGGAGCCAGACTCCTCCTCGGGAAACCGGAAAGGCG
GGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGAGACGCAGACTC
AGTACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCCCCTCTGGTCTGGGAAC
TAATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCG
ACGGAGTGGGTAATTCCTCGGGAAATTGGCATTGCGATTCCACATGGATGGGCGAC
AGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCACCTACAACAACCACCT
CTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCTA
CAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCACCACGT
GACTGGCAAAGACTCATCAACAACAACCTGGGGATTCCGACCCAAGAGACTCAACTT
CAAGCTCTTTAACATTCAAGTCAAAGAGGTCACGCAGAATGACGGTACGACGACGA
TTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCC
CGTACGTCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCAGCAGACGTCT
TCATGGTGCCACAGTATGGATAACCTCACCTGAACAACGGGAGTCAGGCAGTAGGA
CGCTCTTCAATTTACTGCCTGGAGTACTTTCCTTCTCAGATGCTGCGTACCGGAAACA
ACTTTACCTTCAGCTACACTTTTGAGGACGTTCCCTTCCACAGCAGCTACGCTCACAG

[00126] In certain embodiments, an R2-MAC, REP protein comprises SEQ ID NO: 3:

[00128] In certain embodiments, an R2-MAC, VP1(SAQ-~~MAC~~) protein comprises SEQ ID NO: 4:

40

[00129] In certain embodiments, an AAV-R2e-MAC nucleotide sequence (enhanced version of R2-MAC with 4 YF mutations and 1 TV mutation; (AAV2-REP-link-VF1(4YF/TV and SAQ-MAC)) comprises SEQ ID NO: 5:

[00130] ATGCCCGGGGTTTTACGAGATTGTGATTAAAGGTCCCCAGCGACCTTGACGAGCATCTGC
CCGGCATTTCCTGACAGCTTTTGTGAAGTGGGTGGCCGAGAAAGGAATGGGAGTTGCCGCCAGATTTC
TGACATGGATCTGAATCTGATTGAGCAGGACCCCTGAACCTGGCCGAGAAAGCTGACAGCGCGAC
TTTCTGACGGAATGGCGCCGCTGTGAGTAAGGCCCCCGGAGGCCCTTTTCTTTGTGCAATTTGAGA
AGGGAGAGAGCTACTTCCACATGCAAGTGTCTGCTGGAAACCACCGGGGTGAAATCCATGGTTTTT
GGGACGTTTCTGAGTCAGATTTCGCGAAAACTGATTTCAGAGAATTTACCGCGGGATCGAGCCG
ACTTTGCCAAACTGGTTTCGCGGTACAAAAGACAGAAATGGCGCCGGAGGCGGGAAACAAGGTGG
TGGATGAGTGTACATCCCCAATTACTTGTCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTG
GACTAATATGGAACAGTATTTAAGCGCCTGTTTGAATCTCACGAGCGTAAACGGTTGGTGGCG
CAGCATCTGACGCACGTGTCCCGAGCGCAGGAGCCAGAACAAAGAGAAATCAGAATCCCAATTCTG
ATGCGCCCGGTGATCAGATCAAAAACCTTCAGCCAGGTACATGGAGCTGGTTCGGGTGGCTCGTGA
CAAGGGGATTAACCTCGGAGAAAGCAGTGGATCCAGGAGGACCCAGGCCCTCATACATCTCCTTCAAT
GCGGCCTCCAACTCGCGGTCCCAATCAAGGOTGCCTTGGACAATGCGGGAAAGATTATGAGCC
TGACTAAAACCGCCCGGACTACCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCG
GATTTATAAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGCTCTTCTGGA
TGGGCCACGAAAAAGTTCCGCAAGAGGAACACCATCTGGCTGTTTGGGCTGCAACTACCGGGA
AGACCAACATCGCGGAGGCCATAGCCACACTGTGCCCTTCTACGGGTGGCTAAACTGGACCAA
TGAGAACTTTCCCTTCAACGACTGTGTGACAAAGATGGTGATCTGGTGGGAGGAGGGGAAGATG
ACCGCCAAAGGTCTGGAGTCCGCCAAAGCCATTCTCGGAGGAAGCAAGGTGGCGCTGGACCAGA
AATGCAAGTCTCGGCCAGATAGACCGEACTCCCGTGATCGTCACCTCCAAACACCAACATGTG
CGCGGTGATTGAACGGAACTCAGCGACCTTTCGAACACCAGCAGCCGTTGCAAGACCGGATGTTT
AAATTTGAACTCACCCCGCGTCTGGATCATGACTTTGGGAAGGTACCCAAGCAGGAAGTCAAAG
ACTTTTTTCGSGTGGGCAAAAGGATCAGGTGGTTGAGGTGGAGCATGAATTCTACGTCAAAAAGGG
TGGAGCCAAGAAAAAGACCCGCCCCCAGTGACGCAGATATAAGTGAGCCCCAAACGGGTGCGCGAG
TCAGTTGGCGAGCCATCGACGTGAGACGCGGAAGCTTCGATCAACTACGCAGACAGGTACCAA
ACAAATGTTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGACAAATGCGAGAGAAT
GAATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAGACTGTTTAGAGTGTCTTCCCCTG
TCAGAATCTCAACCGTTTCTGTGCTCAAAAAGGCGTATCAGAACTGTGCTACATTCAATCATA
TCATGGGAAGGTGCCAGACGCTTGCACTGCGCTGCGATCTGGTCAATGTGGATTTGGATGACTG
CATCTTTGAACAATAAatgatttaaatacaggtATGGCTGCCGATGGTTATCTTCCAGATTGGCT
CGAGGACACTCTCTCTGAAGGAATAAGACAGTGGTGGAAGCTCAAACCTGGCCACCACCACCA
AAGCCCGCAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCCCTGGGTACAAGTACCTCG
GACCCCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCGCGGCCCTCGAGCA
CGACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAAACCGTACCTCAAGTACAACCACGCC
GACGCGGAGTTTCAGGAGCGCCTTAAAGAAGATACGTCTTTTGGGGGCAACCTCGGACGAGCAG
TCTTCCAGGCGAAAAAGAGGGTTCTTGAACCTCTGGGCCTGGTTGAGGAACCTGTTAAGACGGC
TCCGGGAAAAAAGAGGCCGTTAGAGCACTCTCTGTGGAGCCAGACTCTCTCGGGAAACCGGA
AAGGCGGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGAGACGCAGACTCAG
TACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCCCCTCTGGTCTGGGAACATAATACGAT
GGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCC
TCGGGAAATTTGGCATTGCGATTCCACATGGATGGGCGACAGAGTCATCACCCACAGCACCCGAA
CCTGGGCCCCTGCCACCTACAACAACCACTCTACAAACAAATTTCCAGCCAATCAGGAGCCTC

GAAACGACAATCACTTCTTTGGCTACAGCACCCTTGGGGGTATTTTGGACTTCAACAGATTCCAC
 TGCCACTTTTTCACCACGTGACTGGCAAAGACTCATCAACAACAACACTGGGGGATTCCGACCCCAAGA
 GACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTCAACGCAGAATGACGGTACGACGAC
 GATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTAC
 GTCTCTGGCTCGGGCGCATCAAGGATGCTTCCCGCCGTTCACAGCAGACGTCTTCATGGTGCCAC
 AGTATGGATACCTCACCTGAACAACGGGAGTCAGGCAGTAGGACGCTCTTCATTTTACTGCCT
 GGAGTACTTTCTTCTCAGATGCTGCGTACCGGAAACAACCTTTACCTTCAGCTACACTTTTGAG
 GACGTTCTTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCCTCTCA
 TCGACCAGTACCTGTATTCTTTGAGCAGAACAAACACTCCAAAGTGGAAACCACCGCAGTCAAG
 GCTTCAGTTTTCTCAGGCCGAGCGAGTGACATTCCGGGACCAGTCTAGGAACTGGCTTCTCTGGA
 CCTGTATTACCGCCAGCAGCGAGTATCAAAGgtATCTGCGGATAACAACAACAGTGAATCTCTCGT
 GGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCAT
 GGCAAGCCACAAGGACGATGAAGAAAAGTTTTTCTCAGAGCGGGGTTCTCATCTTTGGGAAG
 CAAGGCTCAGAGAAAACAATGTGGACATTGAAAAGGTCATGATTACAGACGAAGAGGAAATCA
 GGACAACCAATCCCGTGGCTACGGAGCAGTATGGTTCTGTATCTACCAACCTCCAGagtggcca
 aCTCAACACCAACGAAGCCGATAAGACAAAGCAGCTACCGCAGATGTCAACACACAAGGCGTTCTT
 CCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCTGGGCAAAGATTCCAC
 ACACGGACGGACATTTTCACCCTCTCCCTCATGGGTGGATTTCGGACTTAAACACCCTCCTCC
 ACAGATTCTCATCAAGAACACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAG
 TTTGCTTCTTTCATCACACAGTACTCCACGGGACAGGTTCAGCGTGGAGATCGAGTGGGAGCTGC
 AGAAGGAAAACAGCAAACGCTGGAATCCCGAAATTCACTACACTTCCAACTACAACAAGTCTGT
 TAATGTGGACTTTACTGTGGACACTAATGGCGTGTATTTCAGAGCCTCGCCCCATTGGCACCAGA
 TCTCCTGACTCGTAATCTGTAA

[00131] In certain embodiments, an R2e-MAC, REP protein comprises SEQ ID NO: 6:

[00132] MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPPDSMDLNLIEQAPLTVAEK
 LQRDFLTEWRRVSKAPEALFFVQFEKGESYFHMHLVETTGVKSMVLGRFLSQIREKLIQRIYR
 GIEPTLPNWFVAVTKTRNGAGGGNKVVDECYIPNYLLPKTQPELQWAWTNMEQYLSACNLNTERK
 RLVAQHLTHVSQTQEONKENQNPNSDAPVIRSKTSARYMELVGWLVDKGITSEKQWIQEDQASY
 ISFNAASNSRSQIKAALDNAGKIMSLTKTAPDYLVGQQPVEDISSNRIYKILELNGYDPQYAA
 S VFLGWATKKFGKRNTIWLFGPATTTGKTNIAEAIHTVPFYGCNVNWTNENFPFNDCCVDMVIWWE
 EGKMTAKVVESAKAILGGSKVRVDQCKSSAQIDPTPVIVTSNTNMCAVIDGNSTTFEHQQPLQ
 DRMFKFELTRRLDHDGFKVTKQEVKDFFRWAKDHVVEVEHEFYVKKGGAKKRPPAPSDADISEPK
 RVRESVAQPSTSDAEASINYADRYQNKCSRHVGMNLMFLPCRQCERMNQNSNICFTHGQKDCLE
 CFPVSESQPVSVVKKAYQKLCYIHHIMGKVPDACTACDLVNVDLDDCIFEQ

[00133] In certain embodiments, an R2e-MAC, VP1(4YF/ and SAQ- protein comprises
 SEQ ID NO: 7:

[00134] MAADGYLPDWLEDLTLSEGIRQWWKLKPGPPPKPAERHKDDSRGLVLPGYKYLGPFGN
 LDKGEPVNEADAAALEHDKAYDRQLDSGDNPYLKYNHADAEEFQERLKEDTSFGGNLGRAVFQAK
 KRVLEPLGLVEEPVKTA PGKKRPVEHSPVEPDSSSGTGKAGQQPARKRLNFGQTGDADSVDPDQ
 PLGQPPAAPSGLGTNTMATGSGAPMADNNEGADGVGNSSGNWHCDSTWMGDRVITTTSTRTWALP
 TYNNHLYKQISSQSGASNDNHFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWGFPRKRLNFK
 LFNIQVKEVTQNDGTTTIANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVEMVPQYGYL

TLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFTFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYL
YFLSRTNTPSGTTTQSRLQFSQAGASDIRDQSRNWLPGPCYRQQRVSKSADNNNSEFSWTGAT
KYHLNGRDSLNVNPGPAMASHKDDEEKFFPQSGVLI FGKQGSEKTNVDIEKVMITDEEEI RTTNP
VATEQYGSVSTNLOEACDNTTKPQAATADVNTQGVLPGMVWQDRDVYLQGP IWAKI PHTDGH
FHPSPLMGGFGLKHPPQILIKNTVPFANPSTTFSAKFASFITQYSTGQVSVEIEWELQKENS
KRWNPEIQYTSNYSNVNVDFTVD TNGVYSEPRPIGTRFLTRNL

[00135] In certain embodiments, TP2018 (ITR-pdsAAV-hHexB-134-ITR-EGFP-4xmiR9T-BGH-ITR) comprises SEQ ID NO: 8:

[00136] ctgcgcgcctcgcctcgcctcactgagggcgccccgggcaaagcccgggcgctcgggcgacct
ttggctcgcccgccctcagtgagcgagcgagcgcgagagggagtggaattcacgcgtggatc
tgaattcaattcacgcgtggatccGTCTGGGGGGCGGGCGCGCAGTCATCTGACTCGGTGACTC
ACCCGCGGCCGCGCTTCTCTGATCCGGGCGGGGCGGGAAGTCGGGTCCCGAGGCTCCGGCTCG
GCAGACCGGGCGGAAAGCAGCCGAGCGGCCaTTAAAGGTTTATACCTCCGAGGTAACAACCC
AACCAACTTTCGATCTCTTGTAGATCTTCTCTTAACGAACCTTAAATCTGTGTGGCTGTCA
TTCGGCTCATGCTTACTGCACTCAGCGAGTAAATTAATAACTAATTACTGTCTGCTCAGGGA
CAGGAGTAACTCGTCTATCTTCTGCAAGCTGCTTACGGTTTGGTCCGCTGTCAGCCGATCATC
AGCACATCTAGGTTTCTGTCGGGTGTGACCGAAAGGTAAGaccggtcgccaccatgggtgagcaa
ggcgagagagctgttcacccggggtgggtgcccacccctgggtcgagctggagcgcgacgtaaaacggc
cacaagttcagcgtgttcggcgagggcgagggcgatgccacctacgggaagctgacctgaagt
tcacttgaccacccgggaagctgcccgtgcccgtggcccacccctcgtgaccacccctgacctacgg
cgtgagctgttcaagccgctaccccgaccccatgaagcagcaggaattcttcaactcccccattg
cccgaaagctacgtcccgagcgccaccatcttcttcaaggacgacggcgaactacaagacccggc
ccgaggtgaagttcgagggcgacacccctgggtgaacccgcatcgagctgaaggggcatcgacttcaa
ggaggacggcaacatccctggggcacaagctggagtacaactacaacagccacaacgtctatatc
atggccgacaagcagaagaacggcatcaaggtgaactcaagatccggcacaacatcgaggacg
gcagcgtgcagctcgccgaccactaccagcagaacaccccccaccccgagcgacggccccctgctgct
gcccgcacaaccactacctgagcaccacagtcggcgccctgagcaaaagaccccacccgagaagcgcgat
cacatggctcctgctggagttcgtgacccggcgccgggatcactctcgccatggacgagctgtaca
agtaaagcgggccatcaagctTCATACAGCTAGATAACCAAGCGGATTTCATACAGCTAGATAACCA
AAGAGCGGTTTCATACAGCTAGATAACCAAGCTCACTCATAACAGCTAGATAACCAAGGgtcgac
tagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcc
cccgtgccttcccttgaccctggaaggtgccactcccactgtcctttcctaataaaaatgaggaaa
ttgcatcgcatgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaa
gggggaggattgggaagacaatagcaggcatgctggggagagatTgatctGAGGAACCCCTAGT
GATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAGGTC
GCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGG
CC

[00137] In certain embodiments, an AAV barcode reporter vector (ITR-pDsAAV-hHexB134-ITR-NLS-iCre-HA-WPRE3-bc25-miR9T-bGH-ITR) comprises SEQ ID NO: 9:

[00138] ctgcgcgcctcgcctcgcctcactgagggcgccccgggcaaagcccgggcgctcgggcgacctttggctcgcccgccctcagtgag
cgagcgagcgcgagagagggagtggaattcacgcgtggatctgaattcaattcacgcgtggatccGTCTGGGGGGCGGGCG
CGCGCAGTCATCTGACTCGGTGACTCACCCGCGGCCGCGCTTCTCTGATCCGGGCC
GGGCGGGAAGTCGGGTCCCGAGGCTCCGGCTCGGCAGACCGGGCGGAAAGCAGCC

[illegible]

[00139] In certain embodiments, an AAV barcode reporter vector (pDsAAV-hCD68-135-ITR-~~SL-S-ICre-ITR~~-WPRES-bc25-miR9T-bGH-ITR) comprises SEQ ID NO: 10:

[00140] ctgcgcgctcgcctcgtcactgaggccgcccggggcaaaagcccgggcgctcgggcgacctttggctcgcccggcctcagtgag
cgagcgcgagcgcgcagagagggagtggaattcacgcgtggatctgaaltcaattcacgcgtggtaccgcccctgagtcaggctgtgggtg
ggatcatctccagtacaggaagtgcagactttcatttctccttcccaagagagggctgaggggagcaggggtgagcaactgggtgcagacagc
ctagctggactttgggtgaggcggttcagccatctcgtcgacgCCACCGTTAACGCTAGCAATTAAAGGTTTAT
ACC TTCCCAAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCCTTAAAC
GAACTTTTAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGT
ATAATTAATAACTAATTACTGTTCGTTGACAGGACACGAGTAACTCGCTATCTTCTG
CAGGCTGCTTACGGTTTTCGTCCGTGTTGCAGCCGATCATCAACACATCTAGGTTTCGT
CCGGGTGTGACCGAAAGGTAAGaccggctgccaccatgCCcaagaagaagCggaagtggaagtctgcaacctg
ctgactctgcaacaaaacctgcctgcctccctgtggatgccacctctgatgaagtcaggaagaacctgatggacatcttcagggaacagcc
aggcctctctgaaacacacctggaagatgctcctgctctgtgtgcagatctgggctgcctggtgcaagctgaacaacaggaaatgultccct
gctgaacctgaggatgtgagggactacctctgtacctgcaagccagaggcctggctgtgaagaccatccaacagcacctgggcccagctc
aacatgctgcacaggagatctggcctgcctcgcctctgactccaaltgctgtgtctctgtggtgatgaggagaatcagaaggagaatgtgg
atgctggggagagagccaagcaggccctggccttgaacgcactgacttggaccagtcagatccctgatggagaactctgacagatgcc
aggacateaggaacctggccttccctggcattgectacaacacctgctgcgcattggcgaaattgccagaatcagagtgaaggacatctc
ccgcaccgatgggtggagaatgctgatccacaltggcaggaccaagacctggctgtccacagctgggtgtggagaaggccctgtccctgg

```

gggtfaccgaagctggggagagatggatctctgtgtctgggtggctgatgaccccaacaactacctggtctgcccgggicagaagaatggg
gtggctgcccctctgctacacctcccaactgtccacccggggccctggaaggatctttgagggccaccaccggctgatctatgggtccaagg
atgactctgggcaagataacctggcctggtctggccactctgccagagtggtgctgcccaguuacatggccaguuactggtgtgtccatcc
ctgaaatcatgcaggctgggtggctggaccaatgtgaacattgtgatgaactacatcagaaacctggactctgagactggggccatggtgag
gctgctcgaggatgggtgactaccatagatgttccagattacgcttaagaattcgataatcaagcttatcgataatcaacctctggattacaaa
ttgtgaaagattgactggtatcttaactatgttctcttttacgtatgtggatacgtctgtttaatgctttgtatcatgctatgcttccgtatg
gcttcatcttctctctgtataaaatcctgggtAGTTCTtggccacgggggaacatcgcgcgcctgcccgtgcccgtgctggacagggggc
tcggctgttgggactgacaattccgtggNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAagtaaagcgccatcaag
ctTCATACAGCTAGATAACCAAAGCGATTCATACAGCTAGATAACCAAAGACGCGTT
CATACAGCTAGATAACCAAAGTCACTCATACAGCTAGATAACCAAAGgtcgactagagctcg
ctgatcagcctcgactgtgacctctagtgtccagccatctgtgtttgcccctcccccgctgaccttcttgaccttggaaaggtgccactcccactg
tcttctctaataaaaatgaggaaattgcacgcattgtctgagtaggtgtcattctattctgggggggtgggggtggggcaggacagcaagggggg
aggattgggaagacaatagcaggcatgctggggagagatcgatctaggaacccctagtgtgaggttggccactcctctctgctgcgcctc
gctgctcactgaggccgcccggggcaagccggggcgctggggcgaccttgggtgcccggcctcagtgagcgcagcgcgcgcagag
agggaglggcca

```

[00141] MICROGLIAL SPECIFIC MARKERS

[00142] Inflammatory processes and microglia activation accompany most of the pathophysiological diseases in the central nervous system. Glial pathology precedes and even drives the development of multiple neurodegenerative conditions. These resident brain immune cells are divergent from the peripherally infiltrated macrophages. Microglial heterogeneity in the brain is especially visible in their morphology and cell density in particular brain structures but also in the expression of cellular markers. This often determines their role in physiology or pathology of brain functioning. Furthermore, due to activation, microglia show a broad spectrum of phenotypes ranging from the pro-inflammatory, potentially cytotoxic M1 to the anti-inflammatory, scavenging, and regenerative M2.

[00143] Microglia markers include surface, intracellular (cytosolic proteins, gene transcripts), and released molecules. The general microglia markers can be detected irrespective of current cell phenotype. The most widely used markers are ionized calcium-binding adapter molecule 1 (IBA-1), cluster of differentiation receptors (CD68, CD11b, CD14, CD45, CD80, and CD115), fractalkine receptor (CX3CR1), ferritin, F4/80, high-affinity immunoglobulin epsilon receptor subunit gamma (FCER1G) and vimentin. Many of those markers can be detected in other cells also, like CD68 in infiltrating macrophages and vimentin in astroglia. The most specific general microglia markers are transmembrane protein 119 (TMEM119) and purinergic receptor P2Y12R (Jurga AM, *et al.* "Overview of General and Discriminating Markers of Differential Microglia

Phenotypes.” *Front Cell Neurosci.* 2020 Aug 6;14:198. doi: 10.3389/fncel.2020.00198. PMID: 32848611; PMCID: PMC7424058).

[00144] Table 1: Overview of microglial cell marker proteins, their alternative names, function, and localization.

Marker	Full name	Synonyms	Functions
<i>CD11B</i>	cluster of differentiation 11b	integrin subunit alpha M, ITGAM; complement receptor 3 alpha, CR3A	alpha subunit of an integrin complement receptor part 3 (MAC-1); involved in adhesion processes and uptake of complement-coated molecules
<i>CD14</i>	cluster of differentiation 14	myeloid cell-specific leucine-rich glycoprotein, monocyte differentiation antigen CD14	co-receptor for transmembrane TLR4 and endosomal TLR7/9 presenting antigens to them
<i>CD16</i>	cluster of differentiation 16	low-affinity immunoglobulin gamma Fc region receptor III, Fc-gamma RIII	Fc receptor detecting immunoglobulin gamma (IgG) antibodies; engaged in phagocytosis processes
<i>CD40</i>	cluster of differentiation 40	tumor necrosis factor receptor superfamily member 5	transduces signals activating ERK kinase
<i>CD45</i>	cluster of differentiation 45	receptor-type tyrosine-protein phosphatase C, PTPRC	includes enzymatic subunit; positive regulator of T-cell activation
<i>CD80</i>	cluster of differentiation 80	T-lymphocyte activation antigen CD80, Activation B7-1 antigen	together with CD86, CD28, and ICAM1, generates the co-stimulatory signals after MHCII activation
<i>CD68</i>	cluster of differentiation 68	macrosialin	strongly upregulated during inflammation; it able to internalize from cell surface to endosomes immediately after stimulation

Marker	Full name	Synonyms	Functions
<i>CD115</i>	cluster of differentiation 115	macrophage colony-stimulating factor 1 receptor, M-CSFR; colony-stimulating factor 1 receptor, CSF-1R	recognizes pro-inflammatory ligands like IL-34 or CSF-1—cytokines controlling proliferation, differentiation and general functioning of macrophages/microglia
<i>CX3CR1</i>	CX3C chemokine receptor 1	fractalkine receptor, C-X3-C CKR-1, Beta chemokine receptor-like 1, CMK-BRL-1, G-protein coupled receptor 13, V28	microglia migration and adhesion
<i>TMEM119</i>	transmembrane protein 119	osteoblast induction factor (OBIF)	uncertain
<i>F4/80</i>	cell surface glycoprotein F4/80	adhesion G protein-coupled receptor E1, EGF-like module-containing mucin-like hormone receptor-like 1, EMRI	cell surface glycoprotein, expressed in mice, not confirmed in human
<i>FCER1G</i>	high-affinity immunoglobulin epsilon receptor subunit gamma	FcRgamma, Fc-epsilon RI-gamma, FcεRI gamma	associates with pattern recognition, C-type lectin-like receptor [pattern recognition receptors (PRRs), C-type lectin-like receptor (CLEC)]; induces downstream signaling leading to maturation of APCs
<i>FCRL5</i>	Fe receptor-like 5, scavenger receptor		involved in microglial maintenance
<i>Sirpa</i>	signal regulatory protein alpha	CD172a, SHPS-1, BIT	inhibitory receptor; interacts with a broadly expressed CD47; also called the “do not eat me” signal
<i>Siglec</i>	sialic acid-binding immunoglobulin type lectins		family of proteins, pro-inflammatory immune responses and phagocytosis are turned down in microglia by

Marker	Full name	Synonyms	Functions
			inhibitory Siglec signaling; less rodent homologs than human
<i>Glut5</i>	glucose transporter 5	SLC2A5	exclusively microglial glucose transporter
<i>P2Y12</i>	P2Y purinoceptor 12	ADPG-R, P2T(AC), P2Y(AC), P2Y(cyc), P2Y12 platelet ADP receptor, short name: P2Y(ADP), SP1999	detecting nucleotides like ATP released during injuries
<i>Pu.1</i>	transcription factor Pu.1	Spi- 1	plays a crucial role in determining macrophage lineages and microglial genesis and is a major factor in selecting the set of enhancers expressed by microglia.
<i>IBA-1</i>	ionized calcium binding adapter molecule 1	allograft inflammation factor 1 (AIF-1), microglia response factor (MRF-1), daintain	reorganization of microglial cytoskeleton, supporting the phagocytosis process
<i>HexB</i>	β -hexosaminidase subunit β	N-acetyl-beta-glucosaminidase subunit beta	responsible for the degradation of GM2 gangliosides and other molecules containing terminal N-acetyl hexosamines
<i>VIMENTIN</i>	vimentin	fibroblast intermediate filament	key controller for microglia activation
<i>Ferritin</i>	ferritin		responsible for iron storage and its homeostasis, which is downregulated in inflammation
<i>Sal1</i>	Sal-like protein 1		maintenance of microglia homeostasis; its inactivation resulted in the conversion of microglia from resting

Marker	Full name	Synonyms	Functions
			tissue macrophages into inflammatory phagocytes
<i>CD16</i>	cluster of differentiation 16	low-affinity immunoglobulin gamma Fc region receptor III	Fc receptor detecting immunoglobulin gamma (IgG) antibodies; engaged in phagocytosis processes
<i>CD32</i>	cluster of differentiation 32	low-affinity immunoglobulin gamma Fc region receptor II	membrane receptor for the Fc region of IgG; induces inflammatory signals
<i>CD40</i>	cluster of differentiation 40	tumor necrosis factor receptor superfamily member 5	transduces signals activating ERK kinase
<i>CD86</i>	cluster of differentiation 86	T-lymphocyte activation antigen CD86 or B7-2	membrane co-stimulatory receptor responsible for immune cell proliferation and IL-2 production
<i>MHC II</i>	major histocompatibility complex II		mobilizes immune cells to inflammatory response in pathological situation; reacts to TGFβ1 in rodent but not in human
<i>CD163</i>	cluster of differentiation 163	scavenger receptor cysteine-rich type I protein M130, hemoglobin scavenger receptor	clears oxidative Hb, which in consequence leads to subsequent degradation of heme by heme oxygenase-1 (HO-1); produces Fe ²⁺ , CO and the anti-inflammatory metabolites
<i>CD206</i>	cluster of differentiation 206	macrophage mannose receptor 1 (MRC-1), C-type lectin domain family 13 member D	endocytosis processes <i>via</i> detection of pathogenic glycoproteins and polysaccharide chains
<i>TSPO</i>	translocator protein	peripheral benzodiazepine receptor (PBR)	immunomodulation, regulation of apoptosis, cell proliferation

Marker	Full name	Synonyms	Functions
<i>iNOS</i>	inducible nitric oxide synthase	hepatocyte NOS, NOS type II, Peptidyl-cysteine S-nitrosylase NOS2	enzyme producing NO from L-arginine, promotes response against tumors and pathogens <i>via</i> NO production; promotes synthesis of inflammatory factors (IL-6) and is related with expression of transcription factors, e.g., IRF-1 and NF-κB
<i>ARG1</i>	arginase 1	type I arginase, liver-type arginase	enzyme converting an amino acid arginine into ornithine and urea further metabolized to proline and polyamides; needed for wound healing or tissue remodeling, expressed in mice, not confirmed in human
<i>Ym1</i>	chitinase-like protein 3	beta-N-acetylhexosaminidase Ym1, Chitinase-3-like protein 3, Chi3l3, ECF-L, Eosinophil chemotactic cytokine	heparin-binding lectin; prevents degradation of extracellular matrix components; expressed in mice, not confirmed in human
<i>FIZZ1</i>	restin-like alpha	cysteine-rich secreted protein FIZZ1, RELMalpha	mediates interactions between sensory nerves and inflammatory cells in lung; blocks nerve growth factor-induced survival of dorsal root ganglion neurons; expressed in mice, not confirmed in human
<i>MMP9</i>	matrix metalloproteinase 9	gelatinase B	regulates bioavailability of cytokines and chemokines in inflammation; promotes pro-inflammatory IL-1β maturation
<i>MMP12</i>	matrix metalloproteinase 12	macrophage metalloelastase	regulates bioavailability of cytokines and chemokines in inflammation

[00145] Resident CNS microglia and macrophages from the periphery have different origin. One of the distinct molecules is the CD44 marker reported to be expressed only by infiltrating cells and not on resident microglia (Bennett et al., 2016). Another study examining the gene transcription

of adult microglia compared to peripheral cells has suggested that microglia lack CD169 (Butovsky O., *et al.* (2012). Modulation of inflammatory monocytes with a unique microRNA-gene signature ameliorates ALS mice. *J. Neuroimmunol.* 253:63 10.1172/JCI62636). Siglec-H was also indicated as a marker for microglia in mice, absent from CNS-associated macrophages and CNS-infiltrating monocytes except for a minor subset of cells (Konishi H., *et al.* (2017). Siglec-H is a microglia-specific marker that discriminates microglia from CNS-associated macrophages and CNS-infiltrating monocytes. *Glia* 65, 1927–1943. 10.1002/glia.23204 (2017).

[00146] *Quantitative Markers*: In some cases, the quantitative differentiation of markers is conducted. For example, in the adult brain, subtle differences in CD45 protein amount can be detected between microglia expressing CD11b⁺/CD45^{low} and macrophages expressing CD11b⁺/CD45^{high} (Grabert K., *et al.* (2016). Microglial brain region-dependent diversity and selective regional sensitivities to ageing. *Nat. Neurosci.* 19, 504–516. 10.1038/nn.4222).

[00147] A comparison of different phenotypes of human CNS-resident microglia and peripheral immune cells showed characteristic patterns of markers. Three main markers were chosen to distinguish perivascular macrophages (CD11b⁺/CD206^{high}/CD163⁺) from resident microglia (CD11b⁺/CD206^{low/-}/CD163⁻). Despite undetectable levels in physiological conditions, CD206 and CD163 are expressed in the activated M2 anti-inflammatory microglial phenotype (Böttcher C., *et al.* (2019). Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. *Nat. Neurosci.* 22, 78–90. 10.1038/s41593-018-0290-2).

[00148] *Ontogeny and Transcription Factors*: Transforming growth factor β (TGF β)-dependent signaling was proposed to allow microglia to be distinguished from peripheral macrophages and other immune cells in animal models. This includes expression of several factors, such as FCRLS (Fc receptor-like S, scavenger receptor), HexB (β -hexosaminidase subunit β), P2Y12R, or TMEM119. It was shown that the development of microglia and maintenance of their identity relies on the transcription factors SALL1 and Pu.1 (for an in depth review see Yeh H., Ikezu T. (2019). Transcriptional and epigenetic regulation of microglia in health and disease. *Trends Mol. Med.* 25, 96–111. 10.1016/j.molmed.2018.11.004). This was confirmed in human microglial cells, which exclusively expressed P2Y12R and TMEM119 with additional high expression of CD64, CX3CR1, TGF β , TREM2, CD115, CCR5, CD32, CD172a, and CD91 and low to absent

expression of CD44, CCR2, CD45, CD206, CD163, and CD274 (PD-L1; Böttcher *et al.*, 2019). The transcription factors Pu.1 and Myb were also indicated to allow microglia (Pu.1-dependent transcription) to be distinguished from peripheral macrophages (Butovsky O., Weiner H. L. (2018). Microglial signatures and their role in health and disease. *Nat. Rev. Neurosci.* 19, 622–635. 10.1038/s41583-018-0057-5). P2Y₁₂R is a metabotropic purinoceptor detecting nucleotides like ATP, being released during injuries. It is considered along with TMEM119 as one of the most specific microglial markers.

[00149] The additional significant difference between microglia and macrophages that helps in their discrimination during the CNS inflammation is that microglia activation is detectable very quickly (within 24 h) while peripheral macrophage infiltration is detectable within the next few days.

[00150] *Markers Involved in Mechanisms of Microglia Activation:* Microglia may be activated by various factors present in their surroundings and spotted during the “surveilling” process. Those factors may be exogenous, such as pathogen-associated molecular patterns (PAMPs), bacterial LPS, or pathogen genetic material or viruses. Activating signals can also be endogenous, presented by stressed surrounding cells such as danger/damage-associated molecular patterns (DAMPs, for example, nucleotides) and protein aggregates like amyloid β (A β) senile plaques or can be released by other microglial cells and astrocytes. Recognition of harmful vs. healthy signals is the most complex step, as it requires microglia to distinguish very specific information and not fight the healthy tissue, as happens in some CNS diseases. This requires “find-me” signals released by pathogen or apoptotic cells and specific receptors present on microglia [e.g., vitronectin receptor (VNR), MER receptor tyrosine kinase (MerTK), CX3CR1, or complement receptor 3 (CR3)]. Microglia also respond to so-called “eat-me” signals from the target cells. These are phosphatidylserine or calreticulin present in, e.g., disrupted neuronal membrane, secreted fractalkine or opsonins (growth arrest-specific 6, Gas6), milk fat globule epidermal growth factor 8 (MFG-E8), or complement factors. In opposition, the target cells can also send “do not eat me” signals protecting them from being phagocytosed by microglia. Those signals are recognized by specific microglial receptors. For instance, neuronal CD47 and sialylation are cell-surface proteins recognized by microglial signal regulatory protein α (SIRP α ; Barclay A. N., Van den Berg T. K. (2014). The interaction between signal regulatory protein alpha (SIRP α) and CD47: structure, function, and therapeutic target. *Annu. Rev. Immunol.* 32, 25–50. 10.1146/annurev-immunol-

032713-120142) and Siglec receptors, respectively (Linnartz-Gerlach B., Kopatz J., Neumann H. (2014). Siglec functions of microglia. *Glycobiology* 24, 794–799. 10.1093/glycob/cwu044).

[00151] TRANSGENE

[00152] In certain embodiments, the transgene encodes a therapeutic agent. In certain embodiments, the transgene encodes a gene-editing complex. In certain embodiments, the gene editing complex comprises at least two isolated nucleic acid sequences comprising: (i) a first isolated nucleic acid sequence encoding a first Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a first target sequence in a virus genome, and (ii) a second isolated nucleic acid sequence encoding a second Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a second target sequence in a virus genome.

[00153] The CRISPR-Cas system includes a gene editing complex comprising a CRISPR-associated nuclease, e.g., Cas9, and a guide RNA complementary to a target sequence situated on a DNA strand, such as a target sequence in proviral DNA integrated into a mammalian genome, a receptor used by a virus to infect a cell, e.g., HIV. The gene editing complex can cleave the DNA within the target sequence. This cleavage can in turn cause the introduction of various mutations into the proviral DNA, resulting in inactivation of HIV provirus. The mechanism by which such mutations inactivate the provirus can vary. For example, the mutation can affect proviral replication, and viral gene expression. The mutations may be located in regulatory sequences or structural gene sequences and result in defective production of HIV. The mutation can comprise a deletion. The size of the deletion can vary from a single nucleotide base pair to about 10,000 base pairs. In some embodiments, the deletion can include all or substantially all of the integrated retroviral DNA sequence. In some embodiments the deletion can include the entire integrated retroviral DNA sequence. The mutation can comprise an insertion, that is, the addition of one or more nucleotide base pairs to the pro-viral sequence. The size of the inserted sequence also may vary, for example from about one base pair to about 300 nucleotide base pairs. The mutation can comprise a point mutation, that is, the replacement of a single nucleotide with another nucleotide. Useful point mutations are those that have functional consequences, for example, mutations that

result in the conversion of an amino acid codon into a termination codon or that result in the production of a nonfunctional protein.

[00154] In certain embodiments, the CRISPR/Cas system can be a type I, a type II, or a type III system. Non-limiting examples of suitable CRISPR/Cas proteins include Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, saCas9, cjCas9, Cas12a, Cas12f, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, ARMAN 1, ARMAN 4, Cas3, Cas4, Cas5, Cas5e (or CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, saCas9, cjCas9, Cas12a, Cas12f, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (or CasA), Cse2 (or CasB), Cse3 (or CasE), Cse4 (or CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4, and Cu1966.

[00155] The Cas9 can be orthologous. Six smaller Cas9 orthologues have been used and reports have shown that Cas9 from *Staphylococcus aureus* (SaCas9) can edit the genome with efficiencies similar to those of SpCas9, while being more than 1 kilobase shorter. In addition to the wild type and variant Cas9 endonucleases described, embodiments of the disclosure also encompass CRISPR systems including “enhanced specificity” *S. pyogenes* Cas9 variants (eSpCas9), which dramatically reduce off target cleavage. These variants are engineered with alanine substitutions to neutralize positively charged sites in a groove that interacts with the non-target strand of DNA. This aim of this modification is to reduce interaction of Cas9 with the non-target strand, thereby encouraging re-hybridization between target and non-target strands. The effect of this modification is a requirement for more stringent Watson-Crick pairing between the gRNA and the target DNA strand, which limits off-target cleavage (Slaymaker, I.M. *et al.* (2015) DOI:10.1126/science.aad5227).

[00156] In certain embodiments, three variants found to have the best cleavage efficiency and fewest off-target effects: SpCas9 (K855A), SpCas9 (K810A/K1003A/R1060A) (a.k.a. eSpCas9 1.0), and SpCas9(K848A/K1003A/R1060A) (a.k.a. eSPCas9 1.1) are employed in the compositions. The disclosure is by no means limited to these variants, and also encompasses all Cas9 variants (Slaymaker, I.M. *et al. Science*. 2016 Jan 1;351(6268):84-8. doi: 10.1126/science.aad5227. Epub 2015 Dec 1). The present disclosure also includes another type of enhanced specificity Cas9 variant, “high fidelity” spCas9 variants (HF-Cas9). Examples of

high-fidelity variants include SpCas9-HF1 (N497A/R661A/Q695A/Q926A), SpCas9-HF2 (N497A/R661A/Q695A/Q926A/D1135E), SpCas9-HF3 (N497A/R661A /Q695A/ Q926A/L169A), SpCas9-HF4 (N497A/R661A/Q695A/Q926A/Y450A). Also included are all SpCas9 variants bearing all possible single, double, triple and quadruple combinations of N497A, R661A, Q695A, Q926A or any other substitutions (Kleinstiver, B. P. *et al.*, 2016, *Nature*. DOI: 10.1038/nature16526).

[00157] As used herein, the term “Cas” is meant to include all Cas molecules comprising variants, mutants, orthologues, high-fidelity variants and the like.

[00158] In one embodiment, the endonuclease is derived from a type II CRISPR/Cas system. In other embodiments, the endonuclease is derived from a Cas9 protein and includes Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4, mutants, variants, high-fidelity variants, orthologs, analogs, fragments, or combinations thereof. The Cas9 protein can be from *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus* sp., *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas* sp., *Crocospaera watsonii*, *Cyanothece* sp., *Microcystis aeruginosa*, *Synechococcus* sp., *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccsii*, *Candidatus Desulfurudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natronaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter* sp., *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc* sp., *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira* sp., *Lyngbya* sp., *Microcoleus chthonoplastes*, *Oscillatoria* sp., *Petrotoga mobilis*, *Thermosipho africanus*, or *Acaryochloris marina*. Included are Cas9 proteins encoded in genomes of the nanoarchaea ARMAN-1 (*Candidatus Micrarchaeum acidiphilum* ARMAN-1) and ARMAN-4 (*Candidatus Parvarchaeum acidiphilum* ARMAN-4), CasY (*Kerfeldbacteria*, *Vogelbacteria*, *Komeilibacteria*, *Katanobacteria*), CasX (*Planctomycetes*, *Deltaproteobacteria*).

[00159] In general, CRISPR/Cas proteins comprise at least one RNA recognition and/or RNA binding domain. RNA recognition and/or RNA binding domains interact with guide RNAs. CRISPR/Cas proteins can also comprise nuclease domains (i.e., DNase or RNase domains), DNA binding domains, helicase domains, RNase domains, protein-protein interaction domains, dimerization domains, as well as other domains. Active DNA-targeting CRISPR-Cas systems use 2 to 4 nucleotide protospacer-adjacent motifs (PAMs) located next to target sequences for self versus non-self discrimination. ARMAN-1 has a strong 'NGG' PAM preference. Cas9 also employs two separate transcripts, CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA), for RNA-guided DNA cleavage. Putative tracrRNA was identified in the vicinity of both ARMAN-1 and ARMAN-4 CRISPR-Cas9 systems (Burstein, D. *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature*. 2017 Feb 9;542(7640):237-241. doi: 10.1038/nature21059. Epub 2016 Dec 22).

[00160] Embodiments of the disclosure also include a new type of class 2 CRISPR-Cas system found in the genomes of two bacteria recovered from groundwater and sediment samples. This system includes Cas1, Cas2, Cas4 and an approximately ~ 980 amino acid protein that is referred to as CasX. The high conservation (68% protein sequence identity) of this protein in two organisms belonging to different phyla, Deltaproteobacteria and Planctomycetes, suggests a recent cross-phyla transfer. The CRISPR arrays associated with each CasX has highly similar repeats (86% identity) of 37 nucleotides (nt), spacers of 33–34 nt, and a putative tracrRNA between the Cas operon and the CRISPR array. Distant homology detection and protein modeling identified a RuvC domain near the CasX C-terminal end, with organization reminiscent of that found in type V CRISPR-Cas systems. The rest of the CasX protein (630 N-terminal amino acids) showed no detectable similarity to any known protein, suggesting this is a novel class 2 effector. The combination of tracrRNA and separate Cas1, Cas2 and Cas4 proteins is unique among type V systems, and phylogenetic analyses indicate that the Cas1 from the CRISPR-CasX system is distant from those of any other known type V. Further, CasX is considerably smaller than any known type V proteins: 980 aa compared to a typical size of about 1,200 amino acids for Cpf1, C2c1 and C2c3 (Burstein, D. *et al.*, 2017 *supra*).

[00161] Another new class 2 Cas protein is encoded in the genomes of certain candidate phyla radiation (CPR) bacteria. This approximately 1,200 amino acid Cas protein, termed CasY, appears to be part of a minimal CRISPR-Cas system that includes Cas1 and a CRISPR array. Most of the

CRISPR arrays have unusually short spacers of 17–19 nt, but one system, which lacks Cas1 (CasY.5), has longer spacers (27–29 nt). Accordingly, in some embodiments of the disclosure, the CasY molecules comprise CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, mutants, variants, analogs or fragments thereof.

[00162] The CRISPR/Cas-like protein can be a wild type CRISPR/Cas protein, a modified CRISPR/Cas protein, or a fragment of a wild type or modified CRISPR/Cas protein. The CRISPR/Cas-like protein can be modified to increase nucleic acid binding affinity and/or specificity, alter an enzymatic activity, and/or change another property of the protein. For example, nuclease (i.e., DNase, RNase) domains of the CRISPR/Cas-like protein can be modified, deleted, or inactivated. Alternatively, the CRISPR/Cas-like protein can be truncated to remove domains that are not essential for the function of the fusion protein. The CRISPR/Cas-like protein can also be truncated or modified to optimize the activity of the effector domain of the fusion protein.

[00163] In some embodiments, the CRISPR/Cas-like protein can be derived from a wild type Cas protein or fragment thereof. In other embodiments, the CRISPR/Cas-like protein can be derived from modified Cas proteins. For example, the amino acid sequence of the Cas9 protein can be modified to alter one or more properties (e.g., nuclease activity, affinity, stability, etc.) of the protein. Alternatively, domains of the Cas9 protein not involved in RNA-guided cleavage can be eliminated from the protein such that the modified Cas9 protein is smaller than the wild type Cas9 protein.

[00164] In some embodiments, the CRISPR-associated endonuclease can be a sequence from another species, for example, other bacterial species, bacteria genomes and archaea, or other prokaryotic microorganisms. Alternatively, the wild type Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, ARMAN 1, ARMAN 4, sequences can be modified. The nucleic acid sequence can be codon optimized for efficient expression in mammalian cells, *i.e.*, “humanized.” A humanized Cas9 nuclease sequence can be for example, the Cas9 nuclease sequence encoded by any of the expression vectors listed in GENBANK accession numbers KM099231.1, GI:669193757; KM099232.1 GI:669193761; or KM099233.1 GI:669193765. Alternatively, the Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, ARMAN 1, ARMAN 4, sequences can be for example, the sequence contained within a commercially available vector such as PX330 or PX260 from Addgene (Cambridge, MA). In some embodiments, the Cas9

endonuclease can have an amino acid sequence that is a variant or a fragment of any of the Cas9 endonuclease sequences of GENBANK accession numbers KM099231.1 GI:669193757; KM099232.1 GI:669193761; or KM099233.1 GI:669193765, or Cas9 amino acid sequence of PX330 or PX260 (Addgene, Cambridge, MA).

[00165] The wild type Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, ARMAN 1, ARMAN 4, sequences can be a mutated sequence. For example, the Cas9 nuclease can be mutated in the conserved HNH and RuvC domains, which are involved in strand specific cleavage. In another example, an aspartate-to-alanine (D10A) mutation in the RuvC catalytic domain allows the Cas9 nickase mutant (Cas9n) to nick rather than cleave DNA to yield single-stranded breaks, and the subsequent preferential repair through HDR can potentially decrease the frequency of unwanted indel mutations from off-target double-stranded breaks. The sequences of Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4, mutants, variants, high-fidelity variants, orthologs, analogs, fragments, or combinations thereof, can be modified to encode biologically active variants, and these variants can have or can include, for example, an amino acid sequence that differs from a wild type by virtue of containing one or more mutations (*e.g.*, an addition, deletion, or substitution mutation or a combination of such mutations). One or more of the substitution mutations can be a substitution (*e.g.*, a conservative amino acid substitution). For example, a biologically active variant of a Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4, polypeptides can have an amino acid sequence with at least or about 50% sequence identity (*e.g.*, at least or about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence identity) to a wild type Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4 polypeptides. Conservative amino acid substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine, and leucine; aspartic acid and glutamic acid; asparagine, glutamine, serine and threonine; lysine, histidine and arginine; and phenylalanine and tyrosine. The amino acid residues in the Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4, amino acid sequence can be non-naturally occurring amino acid residues. Naturally

occurring amino acid residues include those naturally encoded by the genetic code as well as non-standard amino acids (*e.g.*, amino acids having the D-configuration instead of the L-configuration). The present peptides can also include amino acid residues that are modified versions of standard residues (*e.g.* pyrrolysine can be used in place of lysine and selenocysteine can be used in place of cysteine). Non-naturally occurring amino acid residues are those that have not been found in nature, but that conform to the basic formula of an amino acid and can be incorporated into a peptide. These include D-alloisoleucine(2R,3S)-2-amino-3-methylpentanoic acid and L-cyclopentyl glycine (S)-2-amino-2-cyclopentyl acetic acid. For other examples, one can consult textbooks or the worldwide web (a site currently maintained by the California Institute of Technology displays structures of non-natural amino acids that have been successfully incorporated into functional proteins).

[00166] Two nucleic acids or the polypeptides they encode may be described as having a certain degree of identity to one another. For example, a Cas9 protein and a biologically active variant thereof may be described as exhibiting a certain degree of identity. Alignments may be assembled by locating short Cas9 sequences in the Protein Information Research (PIR) site (pir.georgetown.edu), followed by analysis with the “short nearly identical sequences” Basic Local Alignment Search Tool (BLAST) algorithm on the NCBI website (ncbi.nlm.nih.gov/blast).

[00167] A percent sequence identity to Cas9 can be determined and the identified variants may be utilized as a CRISPR-associated endonuclease and/or assayed for their efficacy as a pharmaceutical composition. A naturally occurring Cas9 can be the query sequence and a fragment of a Cas9 protein can be the subject sequence. Similarly, a fragment of a Cas9 protein can be the query sequence and a biologically active variant thereof can be the subject sequence. To determine sequence identity, a query nucleic acid or amino acid sequence can be aligned to one or more subject nucleic acid or amino acid sequences, respectively, using the computer program ClustalW (version 1.83, default parameters), which allows alignments of nucleic acid or protein sequences to be carried out across their entire length (global alignment). See Chenna *et al.*, *Nucleic Acids Res.* 31:3497-3500, 2003.

[00168] In some embodiments, the isolated nucleic acids sequences can be encoded by the same construct with one or more isolated nucleic acids sequences directed toward a first and second

retroviral target sequence, and one or more isolated nucleic acids sequences directed toward one or more target sequences.

[00169] In certain embodiments, the first and second target sequences comprise one or more nucleic acid sequences in a human immunodeficiency virus (HIV) comprising: long terminal repeat (LTR) nucleic acid sequences, nucleic acid sequences encoding structural proteins, non-structural proteins, or combinations thereof. In certain embodiments, the sequences encoding structural proteins comprise nucleic acid sequences encoding: Gag, Gag-Pol precursor, Pro (protease), Reverse Transcriptase (RT), integrase (In), Env or combinations thereof, and wherein the sequences encoding non-structural proteins comprise nucleic acid sequences encoding: regulatory proteins, accessory proteins or combinations thereof. In certain embodiments, the regulatory proteins comprise: Tat, Rev or combinations thereof, and wherein accessory proteins comprise Nef, Vpr, Vpu, Vif or combinations thereof.

[00170] Accordingly, the disclosure features compositions for use in inactivating a proviral DNA integrated into a host cell, including an isolated nucleic acid sequence encoding a CRISPR-associated endonuclease and one or more isolated nucleic acid sequences encoding one or more gRNAs complementary to a target sequence in HIV or another retrovirus. A second isolated nucleic acid sequence encoding a CRISPR-associated endonuclease and one or more isolated nucleic acid sequences encoding one or more gRNAs complementary to a target sequence encoding a second target virus sequence. The isolated nucleic acid can include one gRNA, two gRNAs, three gRNAs etc. Furthermore, the isolated nucleic acid can include one or more gRNAs complementary to target sequences in the retrovirus and a second isolated nucleic acid can include one or more gRNAs complementary to target virus sequences such as LTR, Gag, pol etc. One of ordinary skill in the art would only be limited by their imagination with respect to the various combinations of gRNAs.

[00171] In some embodiments, at least one gRNA is complementary to a target sequence in the integrated retroviral DNA and at least one gRNA is complementary to a target sequence in a gene encoding for at least one receptor used by a retrovirus for attachment and/or infection of a cell. In another embodiment, two or more gRNAs are complementary to two or more different target sequences in the integrated retroviral DNA and two or more guide RNAs (gRNAs), are

complementary to two or more target sequences in a gene encoding for at least one receptor used by a retrovirus for attachment and/or infection of a cell *in vitro* or *in vivo*.

[00172] In some embodiments, the isolated nucleic acid encodes at least one gRNA complementary to a target sequence in the integrated retroviral DNA and at least a first gRNA that is complementary to a first target sequence in a gene encoding for at least one receptor used by a retrovirus for attachment and/or infection of a cell; and a second gRNA that is complementary to a second target sequence in a gene encoding for at least one receptor used by a retrovirus for attachment and/or infection of a cell.

[00173] In some embodiments, the isolated nucleic acid encodes at least a first gRNA that is complementary to a first target sequence in the integrated retroviral DNA and at least a second gRNA that is complementary to a second target sequence in the integrated retroviral DNA. Accordingly, any number and combinations of gRNAs with different target sequences can be used to target desired target sequences. In some embodiments, gRNA targets comprise one or more target sequences in an LTR region of an HIV proviral DNA and one or more targets in a structural gene of the HIV proviral DNA; or, one or more targets in a second gene; or, one or more targets in a first gene and one or more targets in a second gene; or, one or more targets in a first gene and one or more targets in a second gene and one or more targets in a third gene; or, one or more targets in a second gene and one or more targets in a third gene or fourth gene; or, any combinations thereof.

[00174] In some embodiments, gRNA targets comprise one or more target sequences in a gene encoding at least one receptor used by a retrovirus for attachment and/or infection of a cell and one or more targets in another gene associated with a viral infection; or, one or more targets in a second gene; or, one or more targets in a first gene and one or more targets in a second gene; or, one or more targets in a first gene and one or more targets in a second gene and one or more targets in a third gene; or, one or more targets in a second gene and one or more targets in a third gene or fourth gene; or, any combinations thereof.

[00175] In certain embodiments, a composition for preventing or treating a retroviral infection *in vitro* or *in vivo*, the composition comprises at least two isolated nucleic acid sequences wherein the first isolated nucleic acid sequences encodes a first Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the

gRNA being complementary to a target sequence in the integrated retroviral DNA; the second isolated nucleic acid sequences encodes a second Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a target sequence in a gene encoding for at least one receptor used by a retrovirus for attachment and/or infection of a cell *in vitro* or *in vivo*.

[00176] In certain embodiments, the target sequence comprises one or more nucleic acid sequences in coding and non-coding nucleic acid sequences of the retrovirus genome.

[00177] In certain embodiments, the target sequences comprise one or more nucleic acid sequences in HIV comprising: long terminal repeat (LTR) nucleic acid sequences, nucleic acid sequences encoding structural proteins, non-structural proteins or combinations thereof. In certain embodiments, the sequences encoding structural proteins comprise nucleic acid sequences encoding: Gag, Gag-Pol precursor, Pro (protease), Reverse Transcriptase (RT), integrase (In), Env or combinations thereof. In certain embodiments, the sequences encoding non-structural proteins comprise nucleic acid sequences encoding: regulatory proteins, accessory proteins or combinations thereof. In certain embodiments, the regulatory proteins comprise: Tat, Rev or combinations thereof. In certain embodiments, the accessory proteins comprise Nef, Vpr, Vpu, Vif or combinations thereof.

[00178] In certain embodiments, the gRNA target sequences comprise one or more target sequences in an LTR region of an HIV proviral DNA and one or more target sequences in a structural gene of the HIV proviral DNA; or, one or more targets in a second gene; or, one or more targets in a first gene and one or more targets in a second gene; or, one or more targets in a first gene and one or more targets in a second gene and one or more targets in a third gene; or, one or more targets in a second gene and one or more targets in a third gene or fourth gene; or, any combinations thereof.

[00179] *Guide RNA Sequences:* The compositions and methods of the present disclosure may include a sequence encoding a guide RNA that is complementary to a target sequence in HIV. The genetic variability of HIV is reflected in the multiple groups and subtypes that have been described. A collection of HIV sequences is compiled in the Los Alamos HIV databases and compendiums (hiv.lanl.gov). The methods and compositions of the disclosure can be applied to HIV from any of those various groups, subtypes, and circulating recombinant forms. These include for example,

the HIV major group (often referred to as Group M) and the minor groups, Groups N, O, and P, as well as but not limited to, any of the following subtypes, A, B, C, D, F, G, H, J and K. or group (for example, but not limited to any of the following Groups, N, O and P) of HIV.

[00180] A gRNA includes a mature crRNA that contains about 20 base pairs (bp) of unique target sequence (called spacer) and a trans-activated small RNA (tracrRNA) that serves as a guide for ribonuclease III-aided processing of pre-crRNA. The crRNA:tracrRNA duplex directs Cas9 to target DNA via complementary base pairing between the spacer on the crRNA and the complementary sequence (called protospacer) on the target DNA. Cas9 recognizes a trinucleotide (NGG) protospacer adjacent motif (PAM) to specify the cut site (the 3rd nucleotide from PAM). In the present disclosure, the crRNA and tracrRNA can be expressed separately or engineered into an artificial fusion gRNA via a synthetic stem loop (AGAAAU) to mimic the natural crRNA/tracrRNA duplex. Such gRNA can be synthesized or *in vitro* transcribed for direct RNA transfection or expressed from U6 or H1-promoted RNA expression vector.

[00181] In the compositions of the present disclosure, each gRNA includes a sequence that is complementary to a target sequence in a retrovirus. The exemplary target retrovirus is HIV, but the compositions of the present disclosure are also useful for targeting other retroviruses, such as HIV-2 and simian immunodeficiency virus (SIV)-1. The guide RNA can be a sequence complimentary to a coding or a non-coding sequence (*i.e.*, a target sequence). For example, the guide RNA can be a sequence that is complementary to a HIV long terminal repeat (LTR) region.

[00182] Some of the exemplary gRNAs of the present disclosure are complimentary to target sequences in the long terminal repeat (LTR) regions of HIV. The LTRs are subdivided into U3, R and U5 regions. LTRs contain all of the required signals for gene expression and are involved in the integration of a provirus into the genome of a host cell. For example, the basal or core promoter, a core enhancer and a modulatory region is found within U3 while the transactivation response element is found within R. In HIV, the U5 region includes several sub-regions, for example, TAR or trans-acting responsive element, which is involved in transcriptional activation; Poly A, which is involved in dimerization and genome packaging; PBS or primer binding site; Psi or the packaging signal; DIS or dimer initiation site. Accordingly, in some embodiments, gRNA targets comprise one or more target sequences in an LTR region of an HIV proviral DNA and one or more targets in a structural gene of the HIV proviral DNA; or, one or more targets in a second

gene; or, one or more targets in a first gene and one or more targets in a second gene; or, one or more targets in a first gene and one or more targets in a second gene and one or more targets in a third gene; or, one or more targets in a second gene and one or more targets in a third gene or fourth gene; or, any combinations thereof.

[00183] Some of the exemplary gRNAs of the present disclosure target sequences in the coding and non-coding protein coding genome of HIV. gRNAs complementary to LTR target sequences include LTR 1, LTR 2, LTR 3, LTR A, LTR B, LTR B', LTR C, LTR D, LTR E, LTR F, LTR G, LTR H, LTR I, LTR J, LTR K, LTR L, LTR M, LTR N, LTR O, LTR P, LTR Q, LTR R, LTR S, AND LTR T. gRNAs complementary to Gag target sequences include Gag A, Gag B, Gag C, and Gag D. gRNAs complementary to pol target sequences include Pol A and Pol B. Accordingly, the compositions of the present disclosure include these exemplary gRNAs, but are not limited to them, and can include gRNAs complimentary to any suitable target site in the protein coding genes of HIV, including but not limited to those encoding the envelope protein env, the structural protein tat, and the accessory proteins vif, willef (negative factor) vpu (Virus protein U) and tev.

[00184] Guide RNA sequences according to the present disclosure can be sense or anti-sense sequences. The guide RNA sequence generally includes a proto-spacer adjacent motif (PAM). The sequence of the PAM can vary depending upon the specificity requirements of the CRISPR endonuclease used. In the CRISPR-Cas system derived from *S. pyogenes*, the target DNA typically immediately precedes a 5'-NGG proto-spacer adjacent motif (PAM). Thus, for the *S. pyogenes* Cas9, the PAM sequence can be AGG, TGG, CGG or GGG. Other Cas9 orthologs may have different PAM specificities. For example, Cas9 from *S. thermophilus* requires 5'-NNAGAA for CRISPR 1 and 5'-NGGNG for CRISPR 3) and *Neisseria meningitidis* requires 5'-NNNNGATT). The specific sequence of the guide RNA may vary, but, regardless of the sequence, useful guide RNA sequences will be those that minimize off-target effects while achieving high efficiency and complete ablation of the genomically integrated HIV provirus. The length of the guide RNA sequence can vary from about 20 to about 60 or more nucleotides, for example about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 36, about 37, about 38, about 39, about 40, about 45, about 50, about 55, about 60 or more nucleotides. Useful selection methods identify regions having extremely low homology between the foreign viral genome and host cellular genome including endogenous retroviral DNA, include bioinformatic screening using 12-bp+NGG target-

selection criteria to exclude off-target human transcriptome or (even rarely) untranslated-genomic sites; avoiding transcription factor binding sites within the HIV LTR promoter (potentially conserved in the host genome); and WGS, Sanger sequencing and SURVEYOR assay, to identify and exclude potential off-target effects.

[00185] The guide RNA sequence can be configured as a single sequence or as a combination of one or more different sequences, *e.g.*, a multiplex configuration. Multiplex configurations can include combinations of two, three, four, five, six, seven, eight, nine, ten, or more different guide RNAs.

[00186] Combinations of gRNAs are especially effective when expressed in multiplex fashion, that is, simultaneously in the same cell. In many cases, the combinations produce excision of the HIV provirus extending between the target sites. The excisions are attributable to deletions of sequences between the cleavages induced by the endonuclease at each of the multiple target sites. These combinations pairs of gRNAs, with one member being complementary to a target site in an LTR of the retrovirus, and the other member being complementary to a gRNA complementary to a target site in a structural gene of the retrovirus. Exemplary effective combinations include Gag D combined with one of LTR 1, LTR 2, LTR 3, LTR A, LTR B, LTR C, LTR D, LTR E, LTR F, LTR G, LTR H, LTR I, LTR J, LTR K, LTR L, LTR M, LTR N, LTR O, LTR P, LTR Q, LTR R, LTR S, or LTR T. Exemplary effective combinations also include LTR 3 combined with one of LTR-1, Gag A, Gag B, Gag C, Gag D, Pol A, or Pol B.

[00187] These are only meant as examples and are not to be construed as limiting the disclosure in any way. When the compositions are administered as a nucleic acid or are contained within an expression vector, the CRISPR endonuclease can be encoded by the same nucleic acid or vector as the guide RNA sequences. Alternatively, or in addition, the CRISPR endonuclease can be encoded in a physically separate nucleic acid from the gRNA sequences or in a separate vector.

[00188] The gRNA sequences according to the present disclosure can be complementary to either the sense or anti-sense strands of the target sequences. They can include additional 5' and/or 3' sequences that may or may not be complementary to a target sequence. They can have less than 100% complementarity to a target sequence, for example 75% complementarity. The gRNA sequences can be employed as a combination of one or more different sequences, *e.g.*, a multiplex

configuration. Multiplex configurations can include combinations of two, three, four, five, six, seven, eight, nine, ten, or more different guide RNAs.

[00189] *Modified or Mutated Nucleic Acid Sequences:* In some embodiments, any of the nucleic acid sequences may be modified or derived from a native nucleic acid sequence, for example, by introduction of mutations, deletions, substitutions, modification of nucleobases, backbones and the like. The nucleic acid sequences include the vectors, gene-editing agents, gRNAs, etc. Examples of some modified nucleic acid sequences envisioned for this disclosure include those comprising modified backbones, for example, phosphorothioates, phosphotriesters, methyl phosphonates, short chain alkyl or cycloalkyl intersugar linkages or short chain heteroatomic or heterocyclic intersugar linkages. In some embodiments, modified oligonucleotides comprise those with phosphorothioate backbones and those with heteroatom backbones, $\text{CH}_2\text{--NH--O--CH}_2$, $\text{CH}_2\text{--N(CH}_3\text{)--O--CH}_2$ [known as a methylene(methylimino) or MMI backbone], $\text{CH}_2\text{--O--N(CH}_3\text{)--CH}_2$, $\text{CH}_2\text{--N(CH}_3\text{)--N(CH}_3\text{)--CH}_2$ and $\text{O--N(CH}_3\text{)--CH}_2\text{--CH}_2$ backbones, wherein the native phosphodiester backbone is represented as O--P--O--CH_2 . The amide backbones disclosed by De Mesmaeker *et al. Acc. Chem. Res.* 1995, 28:366-374) are also embodied herein. In some embodiments, the nucleic acid sequences having morpholino backbone structures (Summerton and Weller, U.S. Pat. No. 5,034,506), peptide nucleic acid (PNA) backbone wherein the phosphodiester backbone of the oligonucleotide is replaced with a polyamide backbone, the nucleobases being bound directly or indirectly to the aza nitrogen atoms of the polyamide backbone (Nielsen *et al. Science* 1991, 254, 1497). The nucleic acid sequences may also comprise one or more substituted sugar moieties. The nucleic acid sequences may also have sugar mimetics such as cyclobutyls in place of the pentofuranosyl group.

[00190] The nucleic acid sequences may also include, additionally or alternatively, nucleobase (often referred to in the art simply as “base”) modifications or substitutions. As used herein, “unmodified” or “natural” nucleobases include adenine (A), guanine (G), thymine (T), cytosine (C) and uracil (U). Modified nucleobases include nucleobases found only infrequently or transiently in natural nucleic acids, *e.g.*, hypoxanthine, 6-methyladenine, 5-Me pyrimidines, particularly 5-methylcytosine (also referred to as 5-methyl-2' deoxycytosine and often referred to in the art as 5-Me-C), 5-hydroxymethylcytosine (HMC), glycosyl HMC and gentobiosyl HMC, as well as synthetic nucleobases, *e.g.*, 2-aminoadenine, 2-(methylamino)adenine, 2-(imidazolylalkyl)adenine, 2-(aminoalkylamino)adenine or other heterosubstituted alkyladenines,

2-thiouracil, 2-thiothymine, 5-bromouracil, 5-hydroxymethyluracil, 8-azaguanine, 7-deazaguanine, N⁶ (6-aminohexyl)adenine and 2,6-diaminopurine. Kornberg, A., DNA Replication, W. H. Freeman & Co., San Francisco, 1980, pp75-77; Gebeyehu, G., *et al. Nucl. Acids Res.* 1987, 15:4513). A “universal” base known in the art, *e.g.*, inosine may be included. 5-Me-C substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2°C. (Sanghvi, Y. S., in Crooke, S. T. and Lebleu, B., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, pp. 276-278).

[00191] Another modification of the nucleic acid sequences of the disclosure involves chemically linking to the nucleic acid sequences one or more moieties or conjugates which enhance the activity or cellular uptake of the oligonucleotide. Such moieties include but are not limited to lipid moieties such as a cholesterol moiety, a cholesteryl moiety (Letsinger *et al.*, *Proc. Natl. Acad. Sci. USA* 1989, 86, 6553), cholic acid (Manoharan *et al. Bioorg. Med. Chem. Lett.* 1994, 4, 1053), a thioether, *e.g.*, hexyl-S-tritylthiol (Manoharan *et al. Ann. N.Y. Acad. Sci.* 1992, 660, 306; Manoharan *et al. Bioorg. Med. Chem. Lett.* 1993, 3, 2765), a thiocholesterol (Oberhauser *et al.*, *Nucl. Acids Res.* 1992, 20, 533), an aliphatic chain, *e.g.*, dodecandiol or undecyl residues (Saison-Behmoaras *et al. EMBO J.* 1991, 10, 111; Kabanov *et al. FEBS Lett.* 1990, 259, 327; Svinarchuk *et al. Biochimie* 1993, 75, 49), a phospholipid, *e.g.*, di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate (Manoharan *et al. Tetrahedron Lett.* 1995, 36, 3651; Shea *et al. Nucl. Acids Res.* 1990, 18, 3777), a polyamine or a polyethylene glycol chain (Manoharan *et al. Nucleosides & Nucleotides* 1995, 14, 969), or adamantane acetic acid (Manoharan *et al. Tetrahedron Lett.* 1995, 36, 3651). It is not necessary for all positions in a given nucleic acid sequence to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single nucleic acid sequence or even at within a single nucleoside within a nucleic acid sequence.

[00192] In some embodiments, the RNA molecules *e.g.* crRNA, tracrRNA, gRNA are engineered to comprise one or more modified nucleobases. For example, known modifications of RNA molecules can be found, for example, in Genes VI, Chapter 9 (“Interpreting the Genetic Code”), Lewis, ed. (1997, Oxford University Press, New York), and Modification and Editing of RNA, Grosjean and Benne, eds. (1998, ASM Press, Washington DC). Modified RNA components include the following: 2'-O-methylcytidine; N⁴-methylcytidine; N⁴-2'-O-dimethylcytidine; N⁴-acetylcytidine; 5-methylcytidine; 5,2'-O-dimethylcytidine; 5-

hydroxymethylcytidine; 5- formylcytidine; 2'-O-methyl-5-formylcytidine; 3-methylcytidine; 2-thiocytidine; lysidine; 2'-O- methyluridine; 2-thiouridine; 2-thio-2'-O-methyluridine; 3,2'-O-dimethyluridine; 3-(3-amino-3- carboxypropyl)uridine; 4-thiouridine; ribosylthymine; 5,2'-O-dimethyluridine; 5-methyl-2- thiouridine; 5-hydroxyuridine; 5-methoxyuridine; uridine 5-oxyacetic acid; uridine 5-oxyacetic acid methyl ester; 5-carboxymethyluridine; 5-methoxycarbonylmethyluridine; 5- methoxycarbonylmethyl-2'-O-methyluridine; 5-methoxycarbonylmethyl-2'-thiouridine; 5- carbamoylmethyluridine; 5-carbamoylmethyl-2'-O-methyluridine; 5- (carboxyhydroxymethyl)uridine; 5-(carboxyhydroxymethyl) uridinemethyl ester; 5- aminomethyl-2-thiouridine; 5-methylaminomethyluridine; 5-methylaminomethyl-2-thiouridine; 5-methylaminomethyl-2-selenouridine; 5-carboxymethylaminomethyluridine; 5-carboxymethylaminomethyl-2'-O-methyl- uridine; 5-carboxymethylaminomethyl-2-thiouridine; dihydrouridine; dihydroribosylthymine; 2'-methyladenosine; 2-methyladenosine; N⁶Methyladenosine; N⁶, N⁶-dimethyladenosine; N⁶,2'-O-trimethyladenosine; 2 methylthio-N⁶Nisopentenyladenosine; N⁶-(cis-hydroxyisopentenyl)-adenosine; 2-methylthio-N⁶-(cis-hydroxyisopentenyl)-adenosine; N⁶-glycinylicarbamoyl)adenosine; N⁶ threonylicarbamoyl adenosine; N⁶-methyl-N⁶-threonylicarbamoyl adenosine; 2-methylthio-N⁶-methyl-N⁶-threonylicarbamoyl adenosine; N⁶-hydroxynorvalylcarbamoyl adenosine; 2-methylthio-N⁶-hydroxynorvalylcarbamoyl adenosine; 2'-O-ribosyladenosine (phosphate); inosine; 2'O-methyl inosine; 1-methyl inosine; 1,2'-O-dimethyl inosine; 2'-O-methyl guanosine; 1-methyl guanosine; N²-methyl guanosine; N², N²-dimethyl guanosine; N², 2'-O-dimethyl guanosine; N², N², 2'-O-trimethyl guanosine; 2'-O-ribosyl guanosine (phosphate); 7-methyl guanosine; N², 7-dimethyl guanosine; N², N²,7-trimethyl guanosine; wyosine; methylwyosine; under-modified hydroxywybutosine; wybutosine; hydroxywybutosine; peroxywybutosine; queuosine; epoxyqueuosine; galactosyl-queuosine; mannosyl-queuosine; 7-cyano-7-deazaguanosine; arachaeosine [also called 7-formamido-7-deazaguanosine]; and 7-aminomethyl-7-deazaguanosine.

[00193] The isolated nucleic acid molecules of the present disclosure can be produced by standard techniques. For example, polymerase chain reaction (PCR) techniques can be used to obtain an isolated nucleic acid containing a nucleotide sequence described herein. Various PCR methods are described in, for example, *PCR Primer: A Laboratory Manual*, Dieffenbach and Dveksler, eds., Cold Spring Harbor Laboratory Press, 1995. Generally, sequence information from the ends

of the region of interest or beyond is employed to design oligonucleotide primers that are identical or similar in sequence to opposite strands of the template to be amplified. Various PCR strategies also are available by which site-specific nucleotide sequence modifications can be introduced into a template nucleic acid. Isolated nucleic acids also can be chemically synthesized, either as a single nucleic acid molecule (*e.g.*, using automated DNA synthesis in the 3' to 5' direction using phosphoramidite technology) or as a series of oligonucleotides. For example, one or more pairs of long oligonucleotides (*e.g.*, >50-100 nucleotides) can be synthesized that contain the desired sequence, with each pair containing a short segment of complementarity (*e.g.*, about 15 nucleotides) such that a duplex is formed when the oligonucleotide pair is annealed. DNA polymerase is used to extend the oligonucleotides, resulting in a single, double-stranded nucleic acid molecule per oligonucleotide pair, which then can be ligated into a vector.

[00194] METHODS OF TREATMENT

[00195] The method represents a solution to the problem of latent HIV reservoirs, a solution which is essential to the treatment and prevention of AIDS and other retroviral diseases. HIV-1 replication in the CNS is initiated from invading monocytes and CD4⁺ T cells, and then spreads to microglial cells and, arguably, astrocytes within the brain parenchyma. Definitive evidence that HIV replicates in myeloid lineage cells within the CNS comes from the observation that HAD patients harbor macrophage-tropic HIV-1 variants that grow selectively in the CNS (Alvarez-Carbonell D, *et al.*, Cross-talk between microglia and neurons regulates HIV latency. *PLoS Pathog.* 2019 Dec 30;15(12): e1008249. doi: 10.1371/journal.ppat.1008249. PMID: 31887215; PMCID: PMC6953890). A consequence of HIV-1 replication in longer-lived cell types in the brain, including microglia, is that virus is depleted more slowly in the cerebrospinal fluid (CSF) than virus in the peripheral circulation after the initiation of therapy. Minimal viral replication still persists in the CNS, especially in microglia and perivascular macrophages, in part because not all anti-HIV drugs are able to cross the blood-brain barrier with high efficiency. However, latent infections of microglial cells and astrocytes also appear to contribute to the long-term persistence of HIV in the CNS Garcia-Mesa Y, *et al.* Immortalization of primary microglia: a new platform to study HIV regulation in the central nervous system. *J Neurovirol.* 2017;23(1):47–66. Epub 2016/11/23. 10.1007/s13365-016-0499-3).

[00196] Microglia constitute the first barrier of the innate immune response in the CNS. They constantly survey the brain parenchyma to detect physiological changes and then migrate to regions of damage, where they become activated. Normally, the activation of microglia in response to inflammatory stimuli is characterized by a transition from a resting state (M0 cells) to an activated proinflammatory phenotype (M1 cells). It is generally believed that over-activated microglia exacerbate neuronal injury through the synthesis and secretion of proinflammatory and cytotoxic factors. Therefore, in NeuroHIV, microglia-mediated neuronal injury appears to result from excitotoxicity, which disrupts the intrinsic molecular mechanisms that control ion homeostasis and energy production in neurons (Alvarez-Carbonell D, *et al.*, 2019).

[00197] HIV-1 neuropathology arises because of the combined neurotoxic effects of the viral proteins and exaggerated inflammatory responses by microglial cells. The HIV proteins Tat, gp120, Vpr, and Nef can directly induce neuronal damage. Additionally, microglia can contribute to neurodegeneration through the release of cytokines and toxins that damage neurons and astrocytes. The natural control mechanisms preventing over-activation of microglial cells are likely to be impaired as a consequence of HIV infection. HIV-infected microglia respond vigorously to proinflammatory signals and produce an excess of cytokines. Importantly, neuronal dysfunction does not correlate with the number of HIV-infected cells or viral antigens in CNS, but rather with elevated inflammatory cytokine levels. Elevated TNF- α mRNA levels in microglia and astrocytes and high levels of IL-1 β and TNF- α are seen in the CNS of patients with HAD. Similarly, increased IL-6 and IL-8 levels are found in the brains of HIV-1 infected patients. Immune-activated HIV-infected, brain-infiltrating macrophages and resident microglia release high levels of neurotoxic cytokines such as TNF- α and IL-1 β . The compositions of the present disclosure, when stably expressed in potential host cells, e.g., microglia reduce or prevent new infection by HIV.

[00198] Accordingly, the present disclosure also provides a method of treatment to reduce the risk of HIV infection in a mammalian subject at risk for infection. The method includes the steps of determining that a mammalian subject is at risk of HIV infection, administering an effective amount of the previously described pharmaceutical composition, and reducing the risk of HIV infection in the mammalian subject. Preferably, the pharmaceutical composition includes a vector that provides stable and/or inducible expression of at least one of the previously enumerated.

[00199] Pharmaceutical compositions according to the present disclosure can be prepared in a variety of ways known to one of ordinary skill in the art. For example, the nucleic acids and vectors described above can be formulated in compositions for application to cells in tissue culture or for administration to a patient or subject. These compositions can be prepared in a manner well known in the pharmaceutical art, and can be administered by a variety of routes, depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including intranasal, vaginal and rectal delivery), pulmonary (*e.g.*, by inhalation or insufflation of powders or aerosols, including by nebulizer; intratracheal, intranasal, epidermal and transdermal), ocular, oral or parenteral. Methods for ocular delivery can include topical administration (eye drops), subconjunctival, periocular or intravitreal injection or introduction by balloon catheter or ophthalmic inserts surgically placed in the conjunctival sac. Parenteral administration includes intravenous, intraarterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, *e.g.*, intrathecal or intraventricular administration. Parenteral administration can be in the form of a single bolus dose, or may be, for example, by a continuous perfusion pump. Pharmaceutical compositions and formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids, powders, and the like. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

[00200] This disclosure also includes pharmaceutical compositions which contain, as the active ingredient, nucleic acids and vectors described herein, in combination with one or more pharmaceutically acceptable carriers. The terms “pharmaceutically acceptable” (or “pharmacologically acceptable”) refer to molecular entities and compositions that do not produce an adverse, allergic or other untoward reaction when administered to an animal or a human, as appropriate. The term “pharmaceutically acceptable carrier,” as used herein, includes any and all solvents, dispersion media, coatings, antibacterial, isotonic and absorption delaying agents, buffers, excipients, binders, lubricants, gels, surfactants and the like, that may be used as media for a pharmaceutically acceptable substance. In making the compositions of the disclosure, the active ingredient is typically mixed with an excipient, diluted by an excipient or enclosed within such a carrier in the form of, for example, a capsule, tablet, sachet, paper, or other container. When the excipient serves as a diluent, it can be a solid, semisolid, or liquid material (*e.g.*, normal saline),

which acts as a vehicle, carrier or medium for the active ingredient. Thus, the compositions can be in the form of tablets, pills, powders, lozenges, sachets, cachets, elixirs, suspensions, emulsions, solutions, syrups, aerosols (as a solid or in a liquid medium), lotions, creams, ointments, gels, soft and hard gelatin capsules, suppositories, sterile injectable solutions, and sterile packaged powders. As is known in the art, the type of diluent can vary depending upon the intended route of administration. The resulting compositions can include additional agents, such as preservatives. In some embodiments, the carrier can be, or can include a lipid-based or polymer-based colloid. In some embodiments, the carrier material can be a colloid formulated as a liposome, a hydrogel, a microparticle, a nanoparticle, or a block copolymer micelle. As noted, the carrier material can form a capsule, and that material may be a polymer-based colloid.

[00201] The nucleic acid sequences of the disclosure can be delivered to an appropriate cell of a subject. This can be achieved by, for example, the use of a polymeric, biodegradable microparticle or microcapsule delivery vehicle, sized to optimize phagocytosis by phagocytic cells such as macrophages. For example, PLGA (poly-lacto-co-glycolide) microparticles approximately 1-10 μm in diameter can be used. The polynucleotide is encapsulated in these microparticles, which are taken up by macrophages and gradually biodegraded within the cell, thereby releasing the polynucleotide. Once released, the DNA is expressed within the cell. A second type of microparticle is intended not to be taken up directly by cells, but rather to serve primarily as a slow-release reservoir of nucleic acid that is taken up by cells only upon release from the microparticle through biodegradation. These polymeric particles should therefore be large enough to preclude phagocytosis (*i.e.*, larger than 5 μm and preferably larger than 20 μm). Another way to achieve uptake of the nucleic acid is using liposomes, prepared by standard methods. The nucleic acids can be incorporated alone into these delivery vehicles or co-incorporated with tissue-specific antibodies, for example antibodies that target cell types that are common latently infected reservoirs of HIV infection, for example, brain macrophages, microglia, astrocytes, and gut-associated lymphoid cells. Alternatively, one can prepare a molecular complex composed of a plasmid or other vector attached to poly-L-lysine by electrostatic or covalent forces. Poly-L-lysine binds to a ligand that can bind to a receptor on target cells. Delivery of "naked DNA" (*i.e.*, without a delivery vehicle) to an intramuscular, intradermal, or subcutaneous site, is another means to achieve *in vivo* expression. In the relevant polynucleotides (*e.g.*, expression vectors) the nucleic acid sequence encoding an isolated nucleic acid sequence comprising a sequence encoding a

CRISPR-associated endonuclease and a guide RNA is operatively linked to a promoter or enhancer-promoter combination. Promoters and enhancers are described above.

[00202] In some embodiments, the compositions of the disclosure can be formulated as a nanoparticle, for example, nanoparticles comprised of a core of high molecular weight linear polyethylenimine (LPEI) complexed with DNA and surrounded by a shell of polyethyleneglycol-modified (PEGylated) low molecular weight LPEI.

[00203] The nucleic acids and vectors may also be applied to a surface of a device (*e.g.*, a catheter) or contained within a pump, patch, or other drug delivery device. The nucleic acids and vectors of the disclosure can be administered alone, or in a mixture, in the presence of a pharmaceutically acceptable excipient or carrier (*e.g.*, physiological saline). The excipient or carrier is selected on the basis of the mode and route of administration. Suitable pharmaceutical carriers, as well as pharmaceutical necessities for use in pharmaceutical formulations, are described in *Remington's Pharmaceutical Sciences* (E. W. Martin), a well-known reference text in this field, and in the USP/NF (United States Pharmacopeia and the National Formulary).

[00204] In some embodiments, the compositions can be formulated as a nanoparticle encapsulating a nucleic acid encoding Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4, mutants, variants, high-fidelity variants, orthologs, analogs, fragments, or combinations thereof, and at least one gRNA sequence complementary to a target HIV and/or to a receptor target sequence, such as CCR5 ; or it can include a vector encoding these components. Alternatively, the compositions can be formulated as a nanoparticle encapsulating the CRISPR-associated endonuclease the polypeptides encoded by one or more of the nucleic acid compositions of the present disclosure.

[00205] In methods of treatment of HIV-1 infection, a subject can be identified using standard clinical tests, for example, immunoassays to detect the presence of HIV antibodies or the HIV polypeptide p24 in the subject's serum, or through HIV nucleic acid amplification assays. An amount of such a composition provided to the subject that results in a complete resolution of the symptoms of the infection, a decrease in the severity of the symptoms of the infection, or a slowing of the infection's progression is considered a therapeutically effective amount. The present methods may also include a monitoring step to help optimize dosing and scheduling as well as

predict outcome. In some methods of the present disclosure, one can first determine whether a patient has a latent HIV infection, and then make a determination as to whether or not to treat the patient with one or more of the compositions described herein. In some embodiments, the methods can further include the step of determining the nucleic acid sequence of the particular HIV harbored by the patient and then designing the guide RNA to be complementary to those particular sequences. For example, one can determine the nucleic acid sequence of a subject's LTR U3, R or U5 region, or pol, gag, or env genes, region and then design or select one or more gRNAs to be precisely complementary to the patient's sequences. The novel gRNAs provided by the present disclosure greatly enhance the chances of formulating an effective treatment. The gRNAs targeted to nucleic acid sequences encoding a receptor used by a virus to infect a cell would prevent further infection.

[00206] In methods of reducing the risk of HIV infection, a subject at risk for having an HIV infection can be, for example, any sexually active individual engaging in unprotected sex, i.e., engaging in sexual activity without the use of a condom; a sexually active individual having another sexually transmitted infection; an intravenous drug user; or an uncircumcised man. A subject at risk for having an HIV infection can be, for example, an individual whose occupation may bring him or her into contact with HIV-infected populations, e.g., healthcare workers or first responders. A subject at risk for having an HIV infection can be, for example, an inmate in a correctional setting or a sex worker, that is, an individual who uses sexual activity for income employment or nonmonetary items such as food, drugs, or shelter.

[00207] COMBINATION THERAPIES

[00208] Accordingly, the disclosure features compositions which include therapeutically effective amounts of at least one antiretroviral agent administered sequentially or alternately or in conjunction with a composition for inactivating a proviral DNA integrated into a host cell. This composition comprises an isolated nucleic acid sequence encoding a CRISPR-associated endonuclease and one or more isolated nucleic acid sequences encoding one or more gRNAs complementary to a target sequence in HIV or another retrovirus.

[00209] In one embodiment, the antiretroviral agent comprises viral entry inhibitors, reverse transcriptase inhibitors, protease inhibitors, and immune-based therapeutic agents.

[00210] For example, when used to treat or prevent HIV infection, the antiretroviral agent or its prodrug or pharmaceutically acceptable salt can be administered in combination or alternation with another anti-HIV agent and/or a gene-editing agent embodied herein. In general, in combination therapy, effective dosages of two or more agents are administered together, whereas during alternation therapy, an effective dosage of each agent is administered serially. The dosage will depend on absorption, inactivation and excretion rates of the drug, as well as other factors known to those of skill in the art. It is to be noted that dosage values will also vary with the severity of the condition to be alleviated. It is to be further understood that for any particular subject, specific dosage regimens and schedules should be adjusted over time according to the individual need and the professional judgment of the person administering or supervising the administration of the compositions.

[00211] In certain embodiments, the gene-editing compositions embodied herein are administered to a patient in combination with one or more other anti-viral agents or therapeutics. Examples include any molecules that are used for the treatment of a virus and include agents which alleviate any symptoms associated with the virus, for example, anti-pyretic agents, anti-inflammatory agents, chemotherapeutic agents, and the like. An antiviral agent includes, without limitation: antibodies, aptamers, adjuvants, anti-sense oligonucleotides, chemokines, cytokines, immune stimulating agents, immune modulating agents, B-cell modulators, T-cell modulators, NK cell modulators, antigen presenting cell modulators, enzymes, siRNA's, ribavirin, protease inhibitors, helicase inhibitors, polymerase inhibitors, helicase inhibitors, neuraminidase inhibitors, nucleoside reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors, purine nucleosides, chemokine receptor antagonists, interleukins, or combinations thereof.

[00212] In certain embodiments, the gene-editing compositions embodied herein are administered with one or more compositions comprising a therapeutically effective amount of a non-nucleoside reverse transcriptase inhibitor (NNRTI) and/or a nucleoside reverse transcriptase inhibitor (NRTI), analogs, variants or combinations thereof. In certain embodiments, an NNRTI comprises: etravirine, efavirenz, nevirapine, rilpivirine, delavirdine, or nevirapine. In embodiments, an NRTI comprises: lamivudine, zidovudine, emtricitabine, abacavir, zalcitabine, dideoxycytidine, azidothymidine, tenofovir disoproxil fumarate, didanosine (ddI EC), dideoxyinosine, stavudine, abacavir sulfate or combinations thereof. In certain embodiments, a composition comprises a therapeutically effective amount of at least one NNRTI or a combination of NNRTI's, analogs,

variants or combinations thereof. In certain embodiments, the NNRTI is rilpivirine. In certain embodiments, an NRTI comprises: lamivudine, zidovudine, emtricitabine, abacavir, zalcitabine, dideoxycytidine, azidothymidine, tenofovir disoproxil fumarate, didanosine (ddI EC), dideoxyinosine, stavudine, abacavir sulfate or combinations thereof. In certain embodiments, the composition comprises a therapeutically effective amount of at least one or a combination of NRTI's, analogs, variants or combinations thereof.

[00213] Kits

[00214] The present disclosure also includes a kit including an isolated nucleic acid sequence encoding a CRISPR-associated endonuclease, for example, a Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4 endonucleases, and at least one isolated nucleic acid sequence encoding a gRNA complementary to a target sequence in an HIV provirus and at least one isolated nucleic acid sequence encoding a gRNA complementary to a target sequence in a gene or nucleic acid sequence encoding a receptor that is used by a virus to infect a cell. Alternatively, at least one of the isolated nucleic acid sequences can be encoded in a vector, such as an expression vector. Possible uses of the kit include the treatment or prophylaxis of HIV infection. Preferably, the kit includes instructions for use, syringes, delivery devices, buffers sterile containers and diluents, or other reagents for required for treatment or prophylaxis. The kit can also include a suitable stabilizer, a carrier molecule, a flavoring, or the like, as appropriate for the intended use.

EXAMPLES

[00215] The studies herein, focused on eradication of HIV latent reservoir in the CNS, which requires the widespread gene delivery via systemic i.v. injection because extensive distribution of HIV provirus exists not only throughout the brain but also in the peripheral organs and tissues. The topical gene delivery via stereotactic, intranasal, retrograde and intrathecal/intracerebroventricular injection is invasive and has limited intraparenchymal spread, although it is promising for some applications. Thus, extensive efforts have identified AAV9 and its derivatives such as AAV-PhP.B, AAV-PhP-eB, AAV-F, AAV9-Retro, AAV9-HR, AAV-Cap-MAC or C2, AAV9P31, AAV-B10 and series, AAV-cc47, AAV-Pal, AAV-MDV1A, AAV.CPP.16 or 21, AAV-BI28 and series, as well as AAV8, AAV10 and their derivatives,

exhibiting potent capability of crossing BBB in both neonatal and adult animals. These AAV-B serotypes have high efficiency to transduce neurons, astrocytes, oligodendrocytes and/or blood vessels. Unfortunately, they failed to transduce brain myeloid cells (microglia and macrophages). On the other hand, several AAV serotypes have been identified for their potential tropism to myeloid cells (AAV-M). For example, early studies showed that AAV2/2 and AAV2/5 can transduce mouse microglia both *in vitro* and *in vivo* (local striatum injection) but the transduction efficiency was <20%. By screening several AAV serotypes (AAV1, 2, 5, 6, 8, 9, etc.), AAV2/6 was found to be the most effective while AAV9 and the others with less efficiency in transducing microglia *in vitro*. The triply mutated AAV2/6 exhibits 95% transduction efficiency in cultured primary microglia but achieves 10-80% transduction efficiency *in vivo* (local injection) in a promoter-dependent manner. Intraspinal injection of 5 AAV serotypes (1, 5, 9, rh10 and hu11) in rats showed 23% transduction efficiency of AAV-rh10 in microglia/macrophages. In non-human primates, AAV9 can transduce microglia/macrophages despite of low efficiency. Therefore, a need exists for an ideal AAV serotype (AAV-BM) with high levels of both BBB crossing and myeloid cell targeting *in vivo*. The primary one AAV-BM1 (=R2-MAC) may transduce not only microglia but also non-microglia neural cells such as neurons, astrocytes, oligodendrocytes and endothelial cells. This non-specific targeting may be beneficial in some situations but detrimental in other cases such as the delivery of gene editors. To increase microglia specificity, this disclosure also relates to the inclusion of microglia-specific promoter such as hCD68, HexB, CD11b, hIBA1, TMEM119, CX3CR1, as well as miR9T for microglia illumination in the transfer vector. Based on our preliminary data, similar or even better transduction efficiency is expected as compared with the parent AAV serotype AAV2 and AAV-Cap-MAC.

[00216] To discover a novel AAV serotype (AAV-BM) that can cross BBB (AAV-B) and transduce microglia (AAV-M), epitope insertional screening was performed to generate infectious AAV-BM virus that carries a 9-mer epitope to cross BBB and infect microglia. In the preliminary studies, the failure of AAV9 and AAV-PhP.eB to transduce human and monkey primary microglia was confirmed. However, 8 AAV-M serotypes were discovered having >90% transduction efficiency at multiplicities of infection (MOI) of 1×10^6 viral genome (vg)/cell (FIGS. 1A, 1B). AAV2 was selected as the initial screening. The AAV2 mutants such as AAV2.7m8 and AAV2-QuadYF showed very high transduction efficiency (99%). AAV-CAP-MAC, AAV-9P31 and

AAV-PhP.eB were selected because they were highly efficient at crossing the blood-brain barrier (BBB) at relatively low dosage (1×10^{13} vg/kg) in multiple species such as monkeys, mice.

[00217] A BBB-penetrating 9-mer epitopes of these AAV-Bs were inserted into AAV2 capsid VP1 protein at position N587 to generate R2-MAC, R2-9P31 and R2-eB (FIGS. 2A-2C). On the other hand, a microglia-specific peptide MG1 was inserted into AAV-PhP.eB at position N452-Q458 (eB-MG1) to explore if MG1 can direct the BBB-crossing AAV-PhP.eB to transduce brain microglia (FIGS. 2A-2C). A small scale of AAV packaging (in 96-well plate) was performed using self-complementary AAV transfer vector scAAV-CB-eGFP in a three-plasmid transfection system and found that 4 AAV-BMs had various degrees of reduction in the packaging efficiency compared to the original serotype AAV2 (for R2-MAC, R2-9P31 and R2.eB) or AAV-PhP.eB (for eB-MG1). For the transduction in HEK293T cells, R2-MAC showed around 1.5-fold reduction while R2-9P31 and R2-eB exhibited about 10-fold reduction as compared with AAV2, while eB-MG1 lost 10-fold transduction compared with AAV-PhP.eB (FIGS. 2A-2C, 3). Intriguingly, R2-MAC showed very strong transduction (+++++) in mouse neural stem cells (NSC), apparently stronger than AAV2 (+++), while R2-9P31, R2-eB and eB-MG1 had almost no infection (FIG. 4). Similarly, R2-MAC showed the highest transduction in mouse microglia, but all the tested AAV-BMs showed moderate transduction in mouse fibroblast with the original AAV-PhP.eB highest (FIGS. 5, 6). Then, a small scale of AAV packaging (1x10 cm dish) was performed and the transduction efficiency was tested in human primary microglia (FIGS. 7A-7E, 9A-9D). Using universal promoter for the transfer reporter vector (scAAV- CB-eGFP), it was found that R2-MAC showed stronger transduction efficiency in human primary microglia than the original AAV2 serotype but the other three AAV-BMs lost the transduction (FIGS. 7A-7E, 8A-8E). Finally, a preliminary study was performed for *in vivo* systemic or local injection in C57BL/6 mice. Using microglia-specific promoter hCD68 for the transfer reporter vector (scAAV-hCD68-135-UTR-eGFP) in a small scale of packaging (1x10 cm dish), a titer of 3.82×10^{11} GC/ml for R2-MAC and 2.67×10^{11} GC/ml for AAD-DJ8 (as the positive control for brain transduction) in a final volume of 150 μ l (FIGS. 9A-9D) and a high transduction efficiency was confirmed in mouse NSCs (FIGS. 9A-9D). 2×10^{10} GC was injected per animal (23 days old, 9-12 g, = about 2×10^{12} GC/kg) via jugular vein (direct viewing of successful i.v. injection) under light anesthesia. Neonatal mice were selected due to potential higher efficiency to cross BBB. At 2 weeks after injection, the animals were perfused and all the organs/tissues were collected. Some degree of eGFP expression was

observed in brain and other organs in both R2-MAC and AAV-DJ8 groups (FIGS. 10, 11), but the transduction efficiency was very low perhaps due to the insufficient dosage.

[00218] The Cre-mediated recombination in the target cells would dramatically increase the resolution of detecting AAV transduction because the reporter tdTomato expression is driven by universal Rosa26 promoter after Cre recombination even at a very low level of gene delivery. Thus, the CMV-Cre-GFP transfer vector was utilized in a parallel study using a small scale of packaging (1x10 cm dish) as described above. A titer of 2.11×10^{11} GC/ml for R2-MAC and 18.3×10^{11} GC/ml was obtained for AAV-DJ8 in final volume of 150 μ l (FIGS. 12A-12D) and found moderate efficiency of Cre-mediated LoxP recombination in NSCs from AI14 transgenic mice (FIGS. 12A-12D). By local injection (brain striatum and hippocampus) with 2 μ l per site (4.22×10^8 GC for R2-MAC and 3.66×10^9 GC for AAV-DJ8) in AI14 mice, the Cre-mediated tdT expression was validated around the injection sites with AAV-DJ8 mainly in neurons but R2-MAC in both microglia-like cells and neurons (FIGS. 13A-13F). 100 μ l (2.11×10^{10} GC for R2-MAC or 1.83×10^{11} GC for AAV-DJ8) was injected per animal (3.5 months old, 26-28 g, = $2\text{--}6 \times 10^{12}$ GC/kg) via jugular vein under light anesthesia. At 2 weeks after injection, the animals were perfused and all the organs/tissues were collected. Some degree of tdT expression was observed in the brain and other organs in both R2-MAC and AAV-DJ8 groups (FIGS. 14A-14D). Microglia-like cells expressed tdT mainly in R2-MAC group, supporting the main hypothesis that R2-MAC can cross BBB and transduce microglia (FIGS. 14A-14D). However, the transduction efficiency was very low due to the insufficient dosage.

[00219] A large scale of packaging and testing of R2-MAC after i.v. injection for its ability to cross the BBB and transduce microglia was conducted. HEK293T cells in 3x10-cm dishes were transfected using another AAV-CMV-Cre transfer vector. At 5 days after transfection, the R2-MAC-CMV-Cre virus was purified via PEG/Chloroform precipitation and ultrafiltration. As shown in FIG. 15A, 9 μ l of purified virus (1.5×10^7 MOI) showed very high transduction efficiency in AI14 mouse NSC. Local injection (3.7×10^9 vg) in brain induced more transduction of microglia-like cells and neurons (FIGS. 15B-15D). Around 70 μ l of purified AAV-R2-MAC-CMV-Cre virus was injected per animal (2.1 months old, 21-23 g, = 5.6×10^{13} vg/kg) via jugular vein. At 1 week after injection, the animals were perfused and all the organs/tissues were collected. Extensive tdT expression was observed throughout the brain, mainly in microglia-like cells, neurons and astrocytes based on the cellular morphology (FIGS. 16, 18). Scattered tdT-positive cells were

observed in lung, liver, kidney and other organs (Fig. 17). However, these tdT-positive cells were not colocalized with IBA1-positive cells that are assumed microglia and macrophages (FIG. 19). Further characterization of tdT-positive cells remains on the way.

[00220] Higher dosage of R2-MAC carrying CMV-Cre induced extensive tdT expression in many cell types. To determine if R2-MAC can transduce microglia *in vivo* after systemic IV injection, we prepared a large scale of purification for AAV-R2-MAC carrying scAAV-hHexB-134-UTR-eGFP transfer vector (TP1986). IV injection at a dose of 9.6×10^{13} vg/kg induced extensive expression of EGFP, validated by immunostaining with two anti-GFP antibodies (FIG. 20). Further characterization of the cell types for these EGFP expressions remains on the way using multilabeled immunostaining and HCR RNA ISH.

[00221] As described above, both *in vitro* and *in vivo* (local and systematic injection) studies indicate that R2-MAC can cross BBB and transduce not only microglia but also many non-microglia cells such as neurons, neural stem cells, astrocytes and others. In the past years, extensive efforts using various types of AAV library screening have identified an increasing number of BBB-crossing serotypes (AAV-B) that can efficiently transduce neurons and astrocytes but not microglia, perhaps due to the weak expression of transgene in microglia. To illuminate microglia expression but conceal transgene expression in non-microglia cells such as neurons, astrocytes, neural stem cells, and oligodendrocytes, the miRNA-9, 124, 129-2-3p target site (miR9T, miR124T and miR129-2-3pT) have been used in lentiviral or AAV gene delivery in mouse models. To determine if miR9T inclusion can improve the specificity of R2-MAC to transduce microglia, we inserted 4x miR9T in TP1986 (scAAV-hHexB-134-UTR-eGFP) transfer vector. As shown in FIG 21, inclusion of miR9T induced the loss of EGFP expression in mouse NSCs as predicted because non-microglia cells like NSCs endogenously express miR9 that can bind to MiR9T resulting in mRNA degradation. However, microglia retain the expression of transgene EGFP due to the absence of miR9 (FIG. 21). Similar result was observed with AAV-PhP.eB packaging (FIG. 22). Optimization of the combined miRNA target sites remain on the way to improve the microglia-targeting specificity and AAV packaging and transduction efficiency.

[00222] In different AAV capsid, the mutations of tyrosine-to-phenylalanine (YF) and threonine-to-valine (TV) have been extensively tested for substantial boosting of AAV transduction efficiency due to loss of phosphorylation and ubiquitin-mediated proteolysis. To determine if these

mutations can improve the transduction efficiency of R2-MAC in microglia, we first validate the boosting activity of 4YF substitutions at specific sites throughout the shared C-terminal domain (Y272F, Y444F, T491V, Y500F, Y730F) and 4YF plus TV mutation of the original AAV2 (FIG. 21). In accord with previous report, AAV2-4YF/TV (R2e) mutant robustly increased AAV2 transduction efficiency in both mouse NSCs and human microglia, and inclusion of miR9T concealed transgene expression in NSCs but illuminated transgene expression in microglia (FIG. 23). Then we assessed 4YF/TV mutations in AAV-R2-MAC vector and found similar boosting activity to these 4YF/TV mutations (Re2-MAC) in the original AAV2 vector (FIGS. 24A-24C). In both serotypes, combined 4YF/TV mutation exhibited stronger transduction efficiency than 4YF alone (FIGS. 23, 24A-22C).

[00223] To further improve the transduction efficiency and specificity of R2-MAC, we developed a new AAV transfer vector with microglia-specific transgene expression (pscAAV-HexB-UTR-iCre-WPRE3-miR9T-BGH, 2.418 kb including 2 ITRs). This novel vector incorporates several previous technologies: 1) The self-complementary double-stranded AAV transfer vector (scAAV with packaging capacity around 2.4 kb) was selected to increase the transduction and transgene expression; 2) The minimal/essential promoter of microglia-specific gene HexB (135 bp) or hCD68 (135 bp) was used due to its small size, high specificity and pathological upregulation; 3) The SARS-CoV-2 5'-UTR and the shortest WPRE3 were included to boost the mRNA expression of transgene; 4) The improved Cre (iCre) was chosen to increase the sensitivity of detecting AAV transduction efficiency because single copy of iCre delivered by AAV could remove STOP terminator for strong tdT expression; and 5) Most importantly, the 4x repeated miRNA-9 target site (miR9T) was added downstream of WPRE3 to illuminate microglia expression (due to miR9 absence) but conceal transgene expression in non-microglia cells that endogenously express miR9. The functions of HexB promoter and miR9T were validated using pscAAV-HexB-GFP-miR9T vector packaged with AAV2, AAV-R2-MAC and AAV-PhP.eB in human microglia and mouse neural stem cells (FIGS. 21, 22, 23 and 24A-24C). The packaging efficiency of pscAAV-HexB-UTR-iCre-WPRE3-miR9T-BGH was assessed by standard 3-plasmid packaging in HEK293T cells with the titer at 1.84e13 GC/ml, comparable to the standard pscAAV-Cb-GFP. In cultured mouse primary microglia (FIG. 25A), pscAAV-HexB-GFP-miR9T with AAV-R2-MAC-4YFTV (R2e-MAC) showed high transduction (FIG. 25B). In cultured neural stem cells from Ai14 mice, AAV-iCre-miR9T showed very few positive cells even at 1e6 MOI, while AAV-iCre without

miR9T exhibited high transduction efficiency in a dose-dependent manner (FIGS. 25C-25D). In cultured mouse microglia, AAV-iCre-miR9T induced dose-dependent transduction as determined by tdT positivity (FIG. 25E). Mouse brain stereotactic injection of purified AAV-iCre viruses (6.4×10^9 GC/mouse, 2 μ l/site) induced expression of tdT mainly in neurons but rarely in astrocytes and microglia (FIG. 24). Addition of miR9T to barcoded AAV library induces specific transgene expression in microglia *in vivo* (FIGS. 25A, 25B).

[00224] Utilizing the optimized microglia-specific AAV transfer vector (FIGS. 21, 22, 23 and 24A-24C), we constructed a bar-coded (bc) AAV-B library by encapsulating four bar-coded AAV transfer vectors (utilizing CD68 or CD11b promoters, each duplicated) within the existing AAV-B serotypes (FIG. 26). This innovative design aims to pinpoint potential AAV-B serotypes capable of efficiently transducing microglia both *in vitro* and *in vivo*. Initial experiments conducted on primary microglia cultured from Ai14 mice revealed that the majority of AAV-B serotypes displayed either minimal or negligible transduction efficiency, consistent with previous reports (FIG. 26). However, notably, seven AAV-B serotypes demonstrated significant transduction, yielding 50-250 tdT-positive cells per well (FIGS. 26, 27), which is comparable to our new R2e-MAC (FIG. 27). These promising results indicate that the novel transfer vector has facilitated the identification of AAV-B serotypes with enhanced microglia targeting capabilities.

[00225] To ascertain the potential of AAV-B in transducing microglia *in vivo*, we initiated brain stereotactic injections of a mixed AAV-B formulation encompassing eight serotypes (1.34×10^{10} GC each, including AAV1/2, AAV-DJ8, AAV6TM, AAV9, AAVrh.10, AAV-HSC16, AAV-Pal2, and AAV-MDV1A) in adult LSL-tdT mice (Male, 3-month-old, 2 μ l/site, 5 sites). Remarkably, extensive tdT⁺ expression was evident solely in microglia-like cells on the injected side, contrasting with the contralateral side, thus providing evidence of successful transduction by at least one of these eight AAV serotypes, and the miR9T addition to these barcoded transfer vectors facilitated highly specific transgene (iCre) expression in microglia *in vivo* (FIG. 28A). The transgene iCre was validated by immunostaining with anti-Cre antibody (FIGS. 28B-28D). Further verification of tdT expression was achieved through immunostaining with anti-RFP or anti-mCherry antibodies (FIGS. 29A-29D). Validation of microglial transduction *in vivo* was confirmed through brain injection of selected individual AAV-B serotypes (FIGS. 30A-30B) and confocal colocalization analysis (FIG. 31). These findings collectively suggest that certain

previously reported AAV-B serotypes indeed possess the capability to transduce microglia, albeit with potential variations across serotypes and various titers.

[00226] Following studies utilizing our novel AAV transfer vector, which identified varying degrees of microglial transduction by most AAV-B serotypes both *in vitro* and *in vivo*, we hypothesized their potential as a novel AAV-BM to cross BBB and transduce microglia in animals. To investigate this, we administered intravenous injections of each AAV-B serotype at 50 μ l (FIG. 32) or 150 μ l (FIG. 33) per adult LSL-tdT mouse. Our results indicate that the majority of the tested AAV-B serotypes can indeed cross the BBB and transduce microglia to varying extents, with efficacy dependent on titers and serotypes. In summary, the utilization of our novel microglia-specific AAV transfer vector, featuring a double-lock mechanism via specific promoter and miRNA target site, demonstrated the capability of previously reported but unidentified AAV-B serotypes to cross the BBB and transduce microglia, albeit with variations in transduction efficiency across different serotypes and titers.

[00227] As an additional control, brain stereotactic injection of purified AAV-CPP21-iCre (without miRT) viruses into mouse brain (6.4×10^9 GC/mouse, 2 μ l/site) induced expression of tdT mainly in neurons but rarely in astrocytes and microglia (FIG. 34).

[00228] The BBB-crossing peptide within R2e-MAC was originally derived from AAV-CAP-MAC that can cross BBB in adult marmosets and newborn macaques. To test if our new R2e-MAC can cross BBB and transduce microglia in macaques, we have initiated a large-scale package and purification using two AAV transfer reporter vectors under the commercial service from two CDMO companies. The pilot trial demonstrated feasible yield for R2e-MAC (FIG. 35) despite a few folds lower than the standard AAV2 system.

[00229] To increase the packaging yield of R2e-MAC, we cloned it into AAVone vector recently developed by AAVngene. This all-in-one packaging system can increase 2-4 fold packaging efficiency for most AAV serotypes. We utilized scAAV-CMV-EGFP transfer vector to endure if R2e-MAC can cross BBB in mice before we use TP2018. After i.v. injection (2.5×10^{14} GC/kg, equal to 5.4×10^{12} GC per animal) of R2e-MAC-CMV-EGFP in C57BL/6 mice, extensive transduction in mouse brain was observed (FIGS. 36A, 36B), clearly demonstrating that R2e-MAC can cross BBB in mice, although the original AAV-CAP-MAC does not work with mouse model. Multilabeled immunohistochemistry and confocal image analysis identified extensive microglia-like cells

(labeled by anti-IBA1 antibody) in different regions of mouse brain (FIGS. 36A, 36B, and 37-39). As designed, Re2-MAC also transduced neurons and astrocytes but the transduction efficiency is much lower than the original AAV-CAP-MAC (compared to the original publication). Interestingly, not all eGFP-labeled microglia-like cells are colocalized with IBA1-labeled cells, indicating that other types of microglia-like cells are present, warranting further characterization with different markers for microglia. Testing if R2e-MAC can cross BBB and transduce microglia using microglia-specific promoter and miR9T system (TP2018) in mice and other species (monkey, pig, rat, etc.) remains on the way.

[00230] In conclusion, a novel hybrid R2e-MAC AAV-BM serotype is identified for BBB penetration and microglia transduction. A novel and optimal AAV transfer vector is maximized for its high specificity and sensitivity in transducing microglia both *in vitro* and *in vivo*. This vector validated the capability of several previously reported AAV-B to transduce microglia both *in vitro* and *in vivo*. It is also useful in screening novel AAV-BM serotypes with BBB-penetration and microglia transduction.

OTHER EMBODIMENTS

[00231] From the foregoing description, it will be apparent that variations and modifications may be made to the disclosure described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

[00232] All citations to sequences, patents and publications in this specification are herein incorporated by reference to the same extent as if each independent patent and publication was specifically and individually indicated to be incorporated by reference.

What is claimed:

1. An adeno-associated virus (AAV) comprising a nucleic acid sequence encoding: (i) a capsid protein comprising a blood brain barrier (BBB) penetrating peptide and (ii) a microglial specific peptide.
2. The AAV of claim 1, wherein the blood brain barrier (BBB) penetrating peptide is derived from an AAV serotype (AAV-B) that efficiently crosses the BBB relative to an AAV serotype (AAV-M) that lacks ability to cross the blood brain barrier transduces microglia.
3. The AAV of claim 2, wherein the AAV serotype that efficiently crosses the BBB comprises AAV2, AAV8, AAV9, AAV10 or derivatives thereof.
4. The AAV of claim 3, wherein AAV9 derivatives comprise AAV-PhP.B, AAV-PhP-eB, AAV-F, AAV9-Retro, AAV-Cap-MAC, AAV-C2, AAV9P31, AAV-B10, AAV-B22, AAV-DJ8, or combinations thereof.
5. The AAV of any one of claims 1-4, wherein the BBB penetrating peptide is inserted into capsid VP1 protein of an AAV (AAV-M) that can transduce microglia.
6. The AAV of claim 5, wherein the AAV VP1 capsid protein is an AAV2 VP1 capsid protein.
7. The AAV of claim 6, wherein the BBB penetrating peptide is inserted at amino acid position N587.
8. The AAV of any one of claims 1-7, wherein the transgene encodes a therapeutic agent, comprising CRISPR/Cas gene editors.
9. The AAV of any one of claims 1-7, wherein the transgene encodes a gene-editing complex.
10. The AAV of claim 9, wherein the gene editing complex comprises at least two isolated nucleic acid sequences comprising: (i) a first isolated nucleic acid sequence encoding a first Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a first target sequence in a virus genome, and (ii) a second isolated nucleic acid sequence encoding a second Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least

one guide RNA (gRNA), the gRNA being complementary to a second target sequence in a virus genome.

11. The AAV of claim 10, wherein the first and second target sequences are different.
12. The AAV of claim 10, wherein the virus comprises a retrovirus.
13. The AAV of claim 11, wherein the virus is a human immunodeficiency virus (HIV).
14. The AAV of claim 13, wherein the first and second target sequences comprise one or more nucleic acid sequences in a human immunodeficiency virus (HIV) comprising: long terminal repeat (LTR) nucleic acid sequences, nucleic acid sequences encoding structural proteins, non-structural proteins, or combinations thereof.
15. The AAV of claim 14, wherein the sequences encoding structural proteins comprise nucleic acid sequences encoding: Gag, Gag-Pol precursor, Pro (protease), Reverse Transcriptase (RT), integrase (In), Env or combinations thereof, and wherein the sequences encoding non-structural proteins comprise nucleic acid sequences encoding: regulatory proteins, accessory proteins or combinations thereof.
16. The AAV of claim 15, wherein regulatory proteins comprise: Tat, Rev or combinations thereof, and wherein accessory proteins comprise Nef, Vpr, Vpu, Vif or combinations thereof.
17. The AAV of claim 10, wherein the endonuclease comprises Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4, mutants, variants, high-fidelity variants, orthologs, analogs, fragments, or combinations thereof.
18. The AAV of claim 1, wherein the nucleic acid sequence encoding the blood brain barrier (BBB) penetrating peptide comprises at least an 80% sequence identity to SEQ ID NO: 1.
19. The AAV of claim 18, wherein the nucleic acid sequence encoding the blood brain barrier (BBB) penetrating peptide comprises SEQ ID NO: 1.
20. The AAV of claim 1, wherein the blood brain barrier (BBB) penetrating peptide comprises poly-L-arginine, a transferrin or lactoferrin ligand, transactivator of transcription (tat), penetratin, insulin peptide or lipoprotein.

21. The AAV of claim 8, wherein the therapeutic agent is a treatment for neurological disorders.

22. The AAV of claim 21, wherein the neurological disorders comprise: AIDS dementia complex, Alzheimer's disease, amyotrophic lateral sclerosis, adrenoleukodystrophy, Alexander disease, Alper's disease, ataxia telangiectasia, Batten disease, bovine spongiform encephalopathy (BSE), Canavan disease, corticobasal degeneration, Creutzfeldt-Jakob disease, dementia with Lewy bodies, fatal familial insomnia, frontotemporal lobar degeneration, Huntington's disease, Kennedy's disease, Krabbe disease, Lyme disease, Machado-Joseph disease, multiple sclerosis, multiple system atrophy, neuroacanthocytosis, Niemann-Pick disease, Parkinson's disease, Pick's disease, primary lateral sclerosis, progressive supranuclear palsy, Refsum disease, Sandhoff disease, diffuse myelinoclastic sclerosis, spinocerebellar ataxia, subacute combined degeneration of spinal cord, tabes dorsalis, Tay-Sachs disease, toxic encephalopathy, transmissible spongiform encephalopathy, or wobbly hedgehog syndrome.

23. The AAV of claim 21, wherein the therapeutic agent comprises antibodies, aptamers, peptides, proteins, natural ligands of one or more CNS target(s), modified versions of natural ligands of one or more CNS target(s), aptamers, inhibitory nucleic acids, small inhibitory RNAs (siRNA), short hairpin RNAs (shRNA)), miRNA, dsDNA, lncRNA, antisense oligonucleotides ribozymes, and small molecules, or active fragments of thereof.

24. The AAV of claim 1, wherein the microglia specific peptide comprises M1, M2 or the combination thereof.

25. The AAV of claim 1, wherein the AAV further comprises a woodchuck hepatitis virus post-transcriptional response element (WPRE).

26. A method of treating a subject suffering from a virus mediated neurological disorder or a neurological disorder comprising: administering to the subject a therapeutically effective amount of an adeno-associated virus (AAV) comprising a nucleic acid sequence encoding: (i) a capsid protein comprising a blood brain barrier penetrating peptide, (ii) a microglial specific peptide, (iii) a microglial tropic promoter and (iv) a transgene.

27. The method of claim 26, wherein the blood brain barrier (BBB) penetrating peptide is derived from an AAV serotype that efficiently crosses the BBB relative to an AAV serotype that lacks ability to cross the blood brain barrier.
28. The method of claim 27, wherein the AAV serotype that efficiently crosses the BBB comprises AAV2, AAV8, AAV9, AAV10 or derivatives thereof.
29. The method of claim 28, wherein AAV9 derivatives comprise AAV-PhP.B, AAV-PhP-eB, AAV-F, AAV9-Retro, AAV-Cap-MAC, AAV-C2, AAV9P31, AAV-B10, AAV-B22, AAV-DJ8, or combinations thereof.
30. The method of claim 26, wherein the BBB penetrating peptide is inserted into an AAV2, capsid VP1 protein.
31. The method of claim 26, wherein the BBB penetrating peptide is inserted at amino acid position Q587.
32. The method of claim 26, wherein the transgene encodes a therapeutic agent.
33. The method of claim 32, wherein the transgene encodes a gene-editing complex.
34. The method of claim 33, wherein the gene editing complex comprises one or more isolated nucleic acid sequences encoding a Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a target sequence in a virus genome.
35. The method of claim 33, wherein the gene editing complex comprises at least two isolated nucleic acid sequences comprising: (i) a first isolated nucleic acid sequence encoding a first Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a first target sequence in a virus genome, and (ii) a second isolated nucleic acid sequence encoding a second Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated endonuclease and at least one guide RNA (gRNA), the gRNA being complementary to a second target sequence in a virus genome.
36. The method of claim 35, wherein the first and second target sequences are different.
37. The method of claim 33, wherein the virus comprises a retrovirus.

38. The method of claim 37, wherein the virus is a human immunodeficiency virus (HIV).

39. The method of claim 35, wherein the first and second target sequences comprise one or more nucleic acid sequences in a human immunodeficiency virus (HIV) comprising: long terminal repeat (LTR) nucleic acid sequences, nucleic acid sequences encoding structural proteins, non-structural proteins, or combinations thereof.

40. The method of claim 39, wherein the sequences encoding structural proteins comprise nucleic acid sequences encoding: Gag, Gag-Pol precursor, Pro (protease), Reverse Transcriptase (RT), integrase (In), Env or combinations thereof, and wherein the sequences encoding non-structural proteins comprise nucleic acid sequences encoding: regulatory proteins, accessory proteins or combinations thereof.

41. The method of claim 40, wherein regulatory proteins comprise: Tat, Rev or combinations thereof, and wherein accessory proteins comprise Nef, Vpr, Vpu, Vif or combinations thereof.

42. The method of claim 33, wherein the endonuclease comprises Cas9, CasX, CasY.1, CasY.2, CasY.3, CasY.4, CasY.5, CasY.6, spCas, eSpCas, SpCas9-HF1, SpCas9-HF2, SpCas9-HF3, SpCas9-HF4, saCas9, cjCas9, Cas12a, Cas12f, ARMAN 1, ARMAN 4, mutants, variants, high-fidelity variants, orthologs, analogs, fragments, or combinations thereof.

43. The method of claim 26, wherein the nucleic acid sequence encoding the blood brain barrier (BBB) penetrating peptide comprises at least an 80% sequence identity to SEQ ID NO: 1.

44. The method of claim 43, wherein the nucleic acid sequence encoding the blood brain barrier (BBB) penetrating peptide comprises SEQ ID NO: 1.

45. The method of claim 26, wherein the blood brain barrier (BBB) penetrating peptide comprises poly-L-arginine, a transferrin or lactoferrin ligand, transactivator of transcription (tat), penetratin, insulin peptide or lipoprotein.

46. The method of claim 26, wherein the therapeutic agent is a treatment for neurological disorders.

47. The method of claim 46, wherein the neurological disorders comprise: AIDS dementia complex, Alzheimer's disease, amyotrophic lateral sclerosis, adrenoleukodystrophy, Alexander disease, Alper's disease, ataxia telangiectasia, Batten disease, bovine spongiform encephalopathy (BSE), Canavan disease, corticobasal degeneration, Creutzfeldt-Jakob disease, dementia with

Lewy bodies, fatal familial insomnia, frontotemporal lobar degeneration, Huntington's disease, Kennedy's disease, Krabbe disease, Lyme disease, Machado-Joseph disease, multiple sclerosis, multiple system atrophy, neuroacanthocytosis, Niemann-Pick disease, Parkinson's disease, Pick's disease, primary lateral sclerosis, progressive supranuclear palsy, Refsum disease, Sandhoff disease, diffuse myelinoclastic sclerosis, spinocerebellar ataxia, subacute combined degeneration of spinal cord, tabes dorsalis, Tay-Sachs disease, toxic encephalopathy, transmissible spongiform encephalopathy, or wobbly hedgehog syndrome.

48. The method of claim 46, wherein the therapeutic agent comprises antibodies, aptamers, peptides, proteins, natural ligands of one or more CNS target(s), modified versions of natural ligands of one or more CNS target(s), aptamers, inhibitory nucleic acids, small inhibitory RNAs (siRNA), short hairpin RNAs (shRNA), ribozymes, and small molecules, or active fragments of thereof.

49. The method of claim 26, wherein the microglia specific peptide comprises M1, M2 or the combination thereof.

50. The method of claim 26, further comprising administering to the subject one or more secondary treatments.

51. The method of claim 50, wherein the secondary treatments comprise anti-virus agents, chemotherapeutic agents, immunotherapy, growth inhibitory agents, cytotoxic agents, radiation therapy, surgery, anti-angiogenesis agents, apoptotic agents, anti-tubulin agents and combinations thereof.

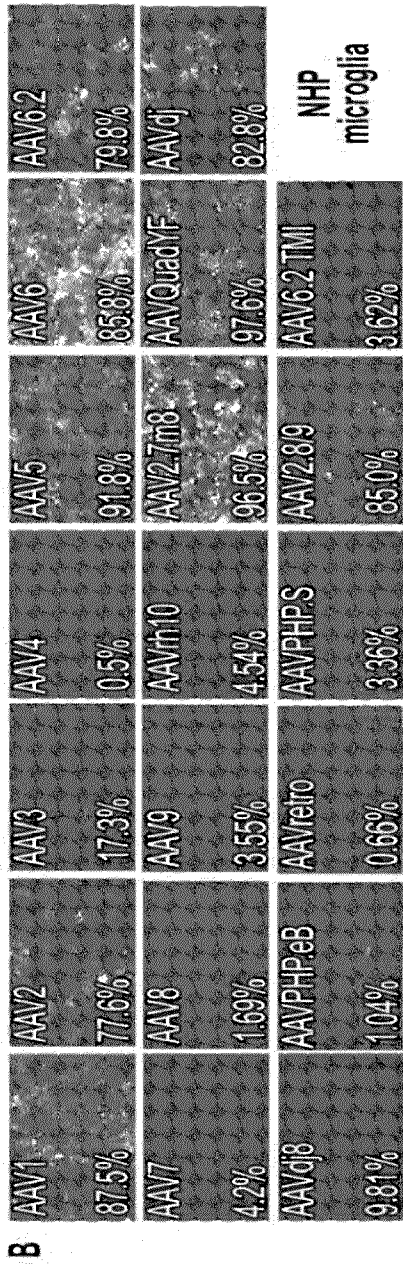
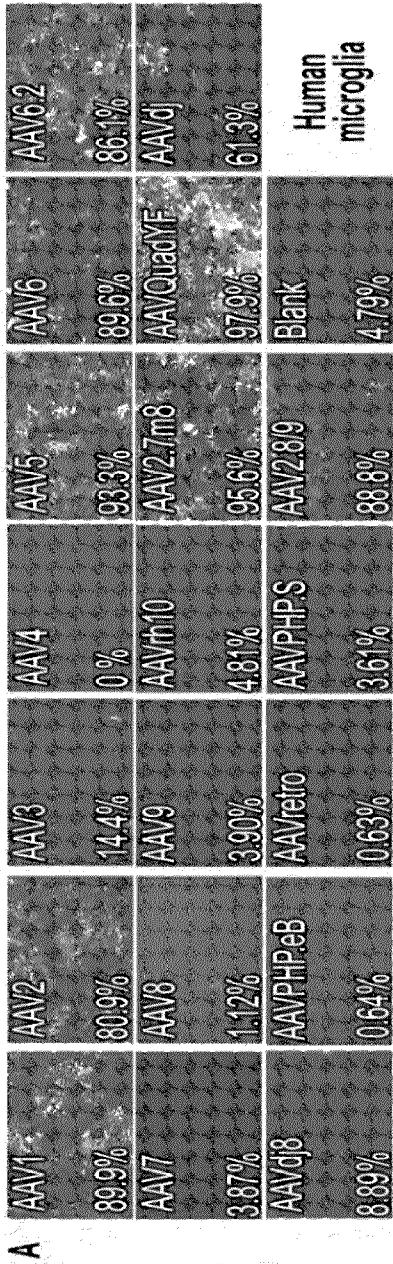
52. The method of claim 51, wherein the anti-virus agent comprises antibodies, aptamers, adjuvants, anti-sense oligonucleotides, chemokines, cytokines, immune stimulating agents, immune modulating agents, B-cell modulators, T-cell modulators, NK cell modulators, antigen presenting cell modulators, enzymes, siRNA's, ribavirin, protease inhibitors, helicase inhibitors, polymerase inhibitors, helicase inhibitors, neuraminidase inhibitors, nucleoside reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors, purine nucleosides, chemokine receptor antagonists, interleukins, or combinations thereof.

53. The method of claim 26, wherein the AAV further comprises a woodchuck hepatitis virus post-transcriptional response element 3 (WPRE3).

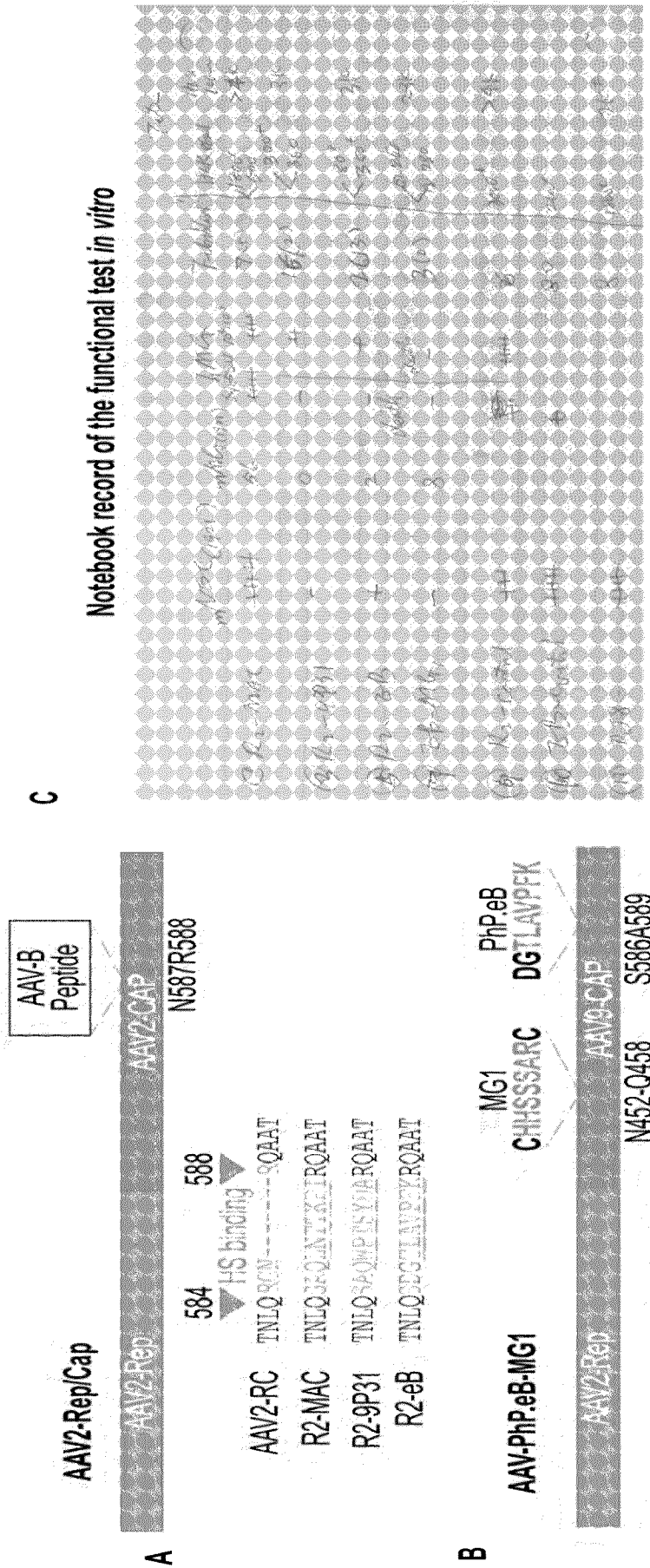
54. An adeno-associated virus (AAV) transfer vector for microglia-specific transgene expression comprising a nucleic acid including: (i) a microglial tropic minimal/essential promoter, (ii) a mRNA expression-boosting UTR, (iii) a transgene, (iv) a mRNA stabilizing minimal Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element 3 (WPRE3) and (v) a microglia-illuminating miR9T and (V) a poly(A) sequence.
55. A recombinant adeno-associated virus (rAAV) comprising a nucleic acid sequence encoding: (i) a capsid protein comprising a BBB penetrating peptide, (ii) at least two microglial specific peptides, (iii) a microglial tropic promoter and (iv) a transgene.
56. The rAAV of claim 55, further comprising a nucleic acid encoding: (i) a microglial tropic minimal/essential promoter, (ii) a mRNA expression-boosting UTR, (iii) a mRNA stabilizing minimal Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element 3 (WPRE3) and (iv) a poly(A) sequence.
57. A bar-coded (bc) adeno- associated virus (AAV) library comprising bar-coded AAV transfer vectors wherein each bar coded AAV transfer vector comprises two CD68 or two CD11b promoters one or a plurality of microglial specific peptides.
58. The bar coded AAV library of claim 57, wherein each microglial specific peptide specifically targets the bc AAV to microglia.
59. The bar coded AAV library of claim 57, wherein each microglial specific peptide specifically binds to different microglial targets.
60. The bar coded AAV library of claim 57, wherein the bc AAV comprises adeno-associated virus serotype B (AAV-B).
61. A method of screening for adeno-associated virus target specificity comprising, screening the bar coded AAV library of claim 57 against microglial specific markers.
62. The method of claim 61, wherein the microglial specific markers comprise Tmem119, P2ry12, HexB, Fcrls, Sall1, C1q, Gpr34, Olfrml3, Mertk, Pros1, Tyro3, Tgfbr1 or combinations thereof.
63. A synthetic sequence comprising any one of SEQ ID NOs: 1-10.

64. A pharmaceutical composition comprising any one of SEQ ID NOs: 1-9, 10 or combinations thereof.

Figs. 1A-1B. AAV serotype screening for infectivity in human ad monkey primary microglia identified 8 AAV serotypes with >80% transduction efficiency at MOI of 10⁵ vg/cell (>90% efficiency at MOI 10⁶ vg/cell, data not shown). Cells in 96-well plate (5x10³/well) were infected with indicated AAV carrying CMV-EGFP (10⁸ vg per well for 3 wells. At 3 days postinfection (dpi), fluorescent microscopic imaging (10x objective) and flow cytometry were performed. Data represent mean $\pm$ SD (n=3).



FIGS. 1A-1B



Figs. 2A-2C. Diagram of AAV-BM generation and summary of functional test in vitro. 2A, Reported peptides from BBB-crossing serotypes (AAV-B) were inserted into the heparan sulfate (HS) binding sites (R585-R588) of AAV2 CAP VP1 and underlined within the colored inserted amino acid sequences. **2B,** Microglia-specific peptide (highlighted red) was inserted into the 455-loop of AAV-PhP.eB. **2C,** Transduction efficiency of AAV-BM in various cell types in vitro

FIGS. 2A-2C

Fig. 3. AAV2-Rep-Cap-MAC (R2-MAC) shows highest packaging titer but all tested AAV-BM lost various degrees of transduction efficiency in HEK293T cells. Cells in 96-well were transfected with helper vector, transfer vector TP1152 (scAAV-CB-EGFP) and indicated serotype vectors. At 3 days, 10 μ l of crude supernatant was used to infect HEK293T cells (1x10⁴/well of 96-well). At 3 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m

Note: Two clones of each AAV-BM were sequenced and tested for the functional packaging. The number (red) indicates transduction unit (TU/ml)

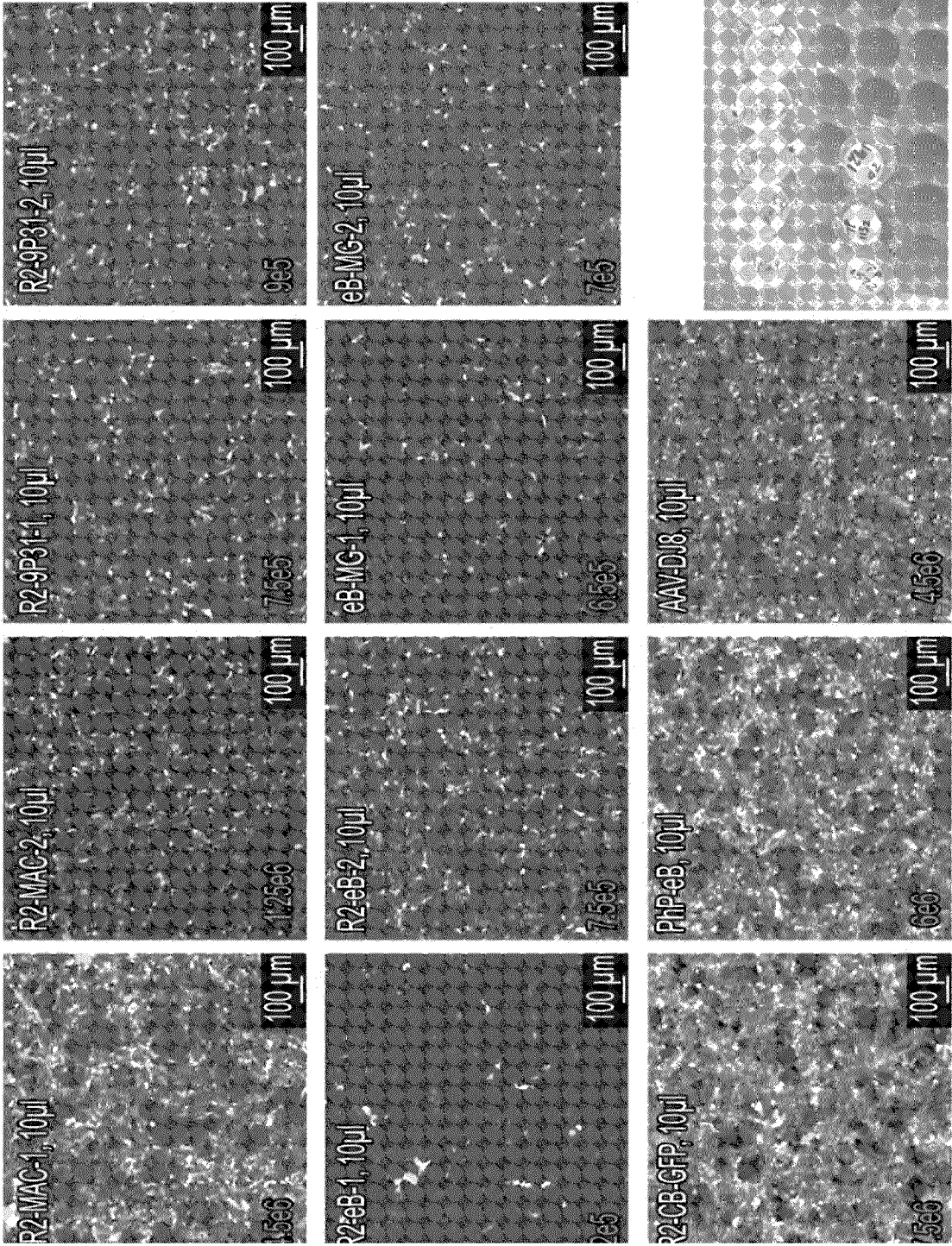


FIG. 3

Fig. 4. R2-MAC shows almost similar efficiency in transducing mouse neural stem cells (NSC) to the original AAV2-RC serotype but the other three AAV-BM lost the transduction compared with the original AAV serotype (AAV2 and PhP-eB). HEK293T cells in 96-well were transfected with helper vector, transfer vector TP1921 (AAV-CMV-Cre-CD63-MG1MG2-MCP) and indicated serotype vector. At 3 days, 10 or 100 μ l of crude supernatant was used to infect cultured NSCs (1×10^4 /well of 96-well) from Ai14 LoxP-STOP-LoxP (LSL)-tdTomato (tdT) mice. at 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 μ m.

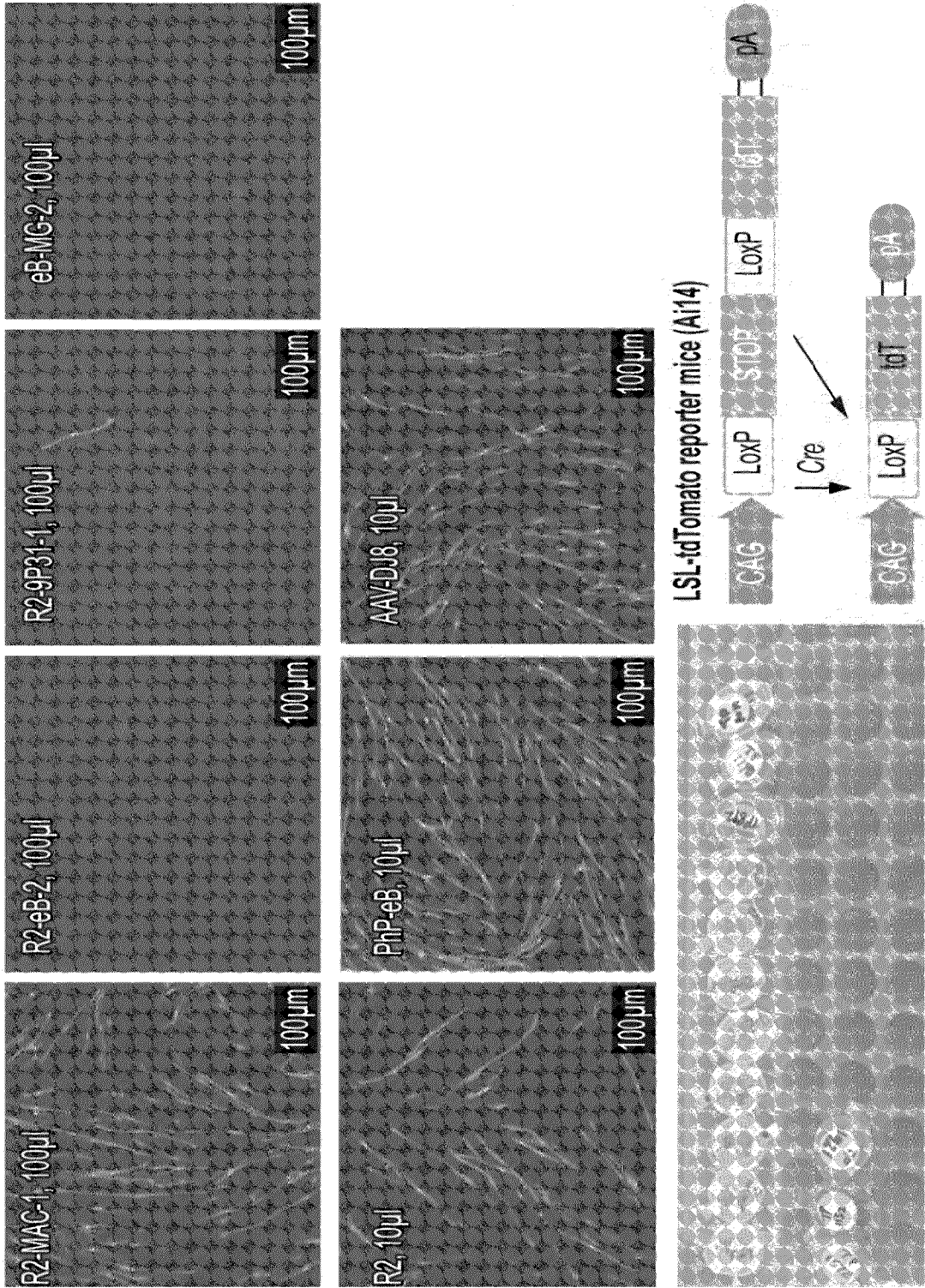


FIG. 4

Fig. 5. R2-MAC transduces mouse primary microglia but the other three AAV-BM does not. HEK293T cells in 96-well were transfected with helper vector, transfer vector TP1921 (AAV-CMV-Cre-CD63-MG1MG2-MCP) and indicated serotype vector. At 3 days, 10 μ l of crude supernatant was used to infect primary microglia from Ai14 dpi. confocal images were taken at 10x objective. Scale bar = 100 μ m

The number in the paranthesis indicated the positive tdT cells in the entire well.

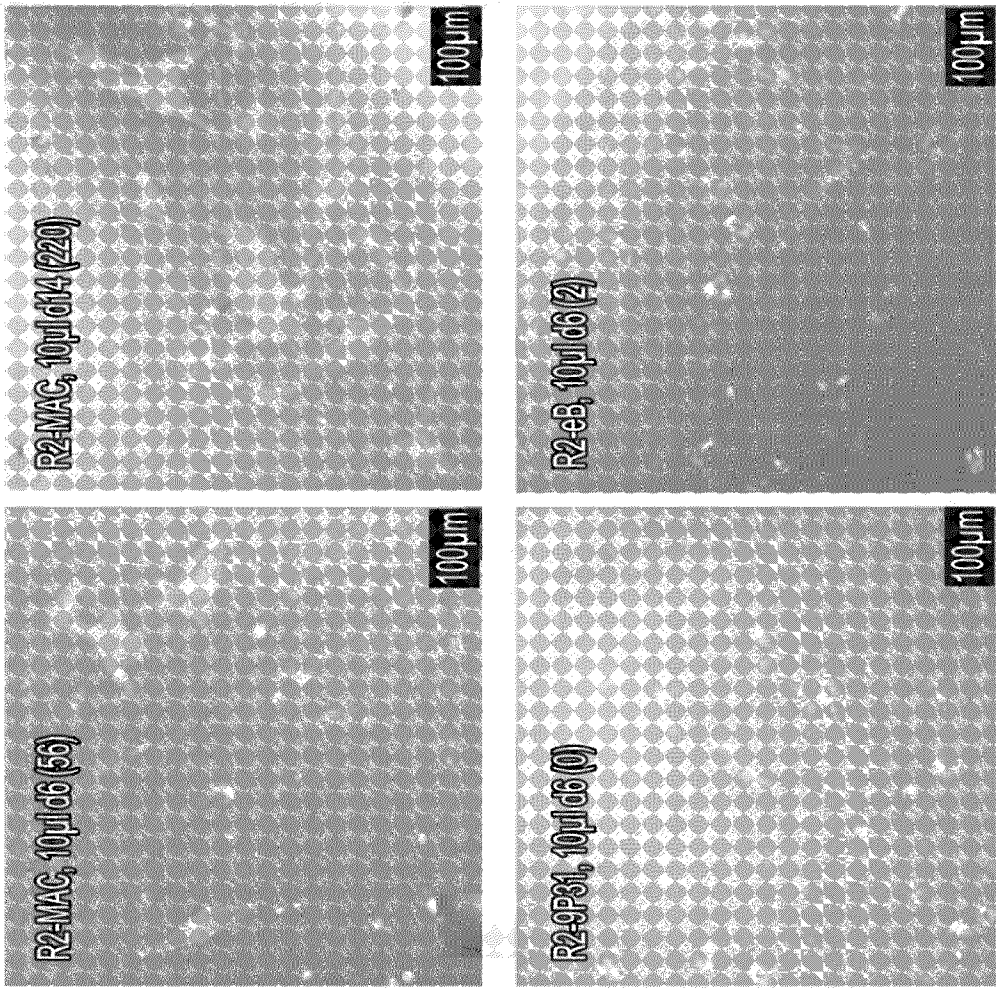


FIG. 5

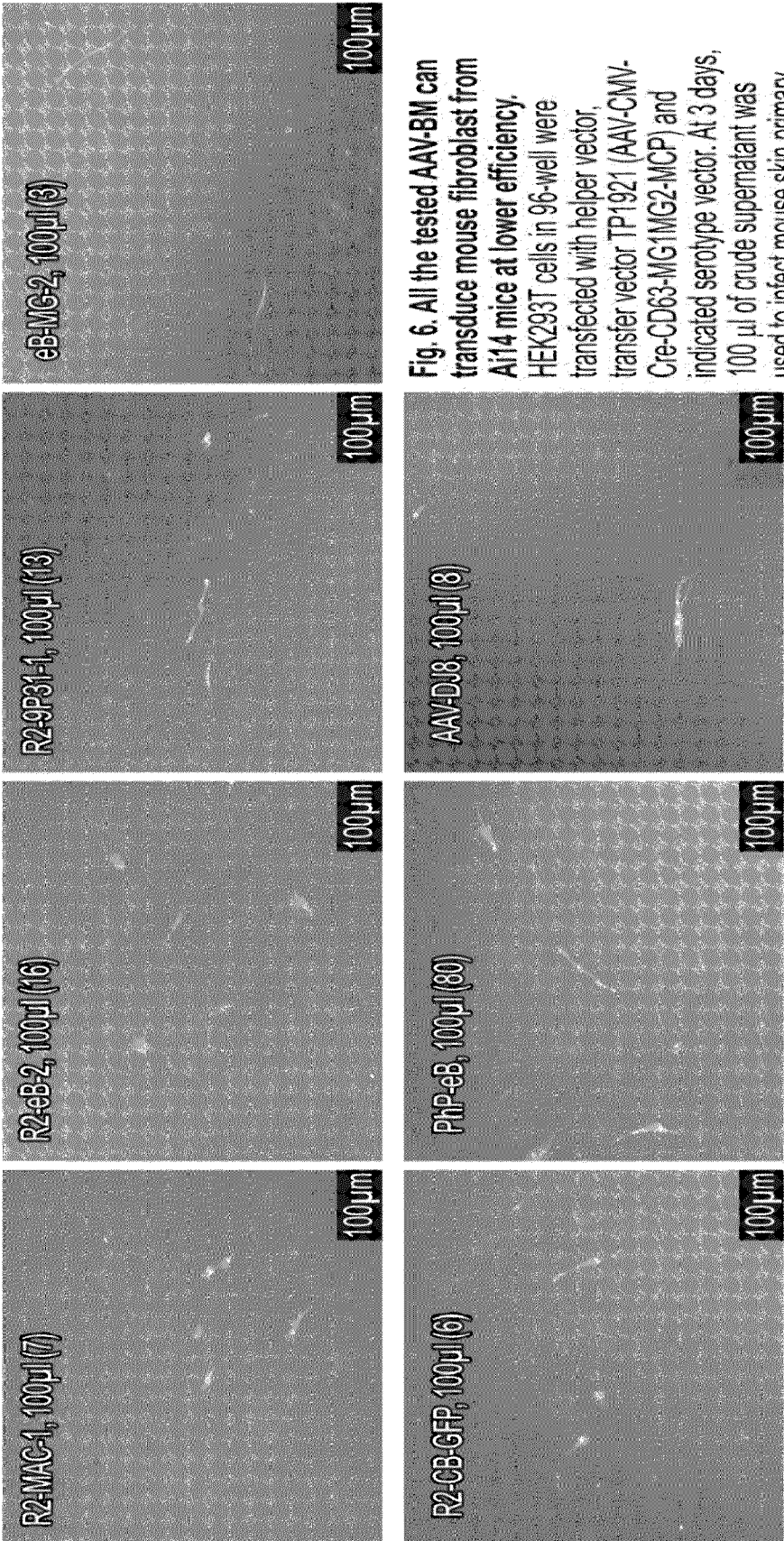
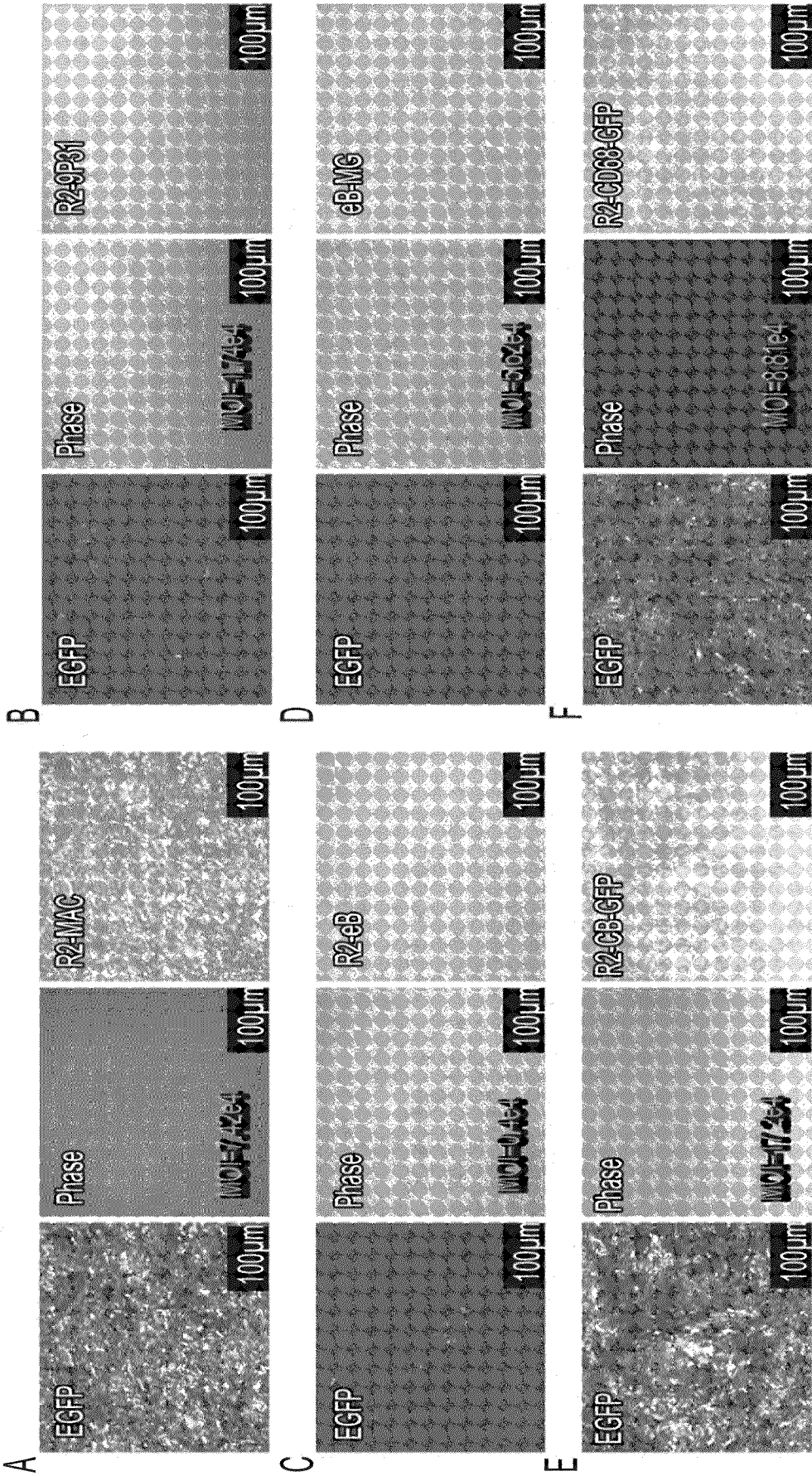


Fig. 6. All the tested AAV-BM can transduce mouse fibroblast from Ai14 mice at lower efficiency. HEK293T cells in 96-well were transfected with helper vector, transfer vector TP1921 (AAV-CMV-Cre-CD63-MG1MG2-MCP) and indicated serotype vector. At 3 days, 100 µl of crude supernatant was used to infect mouse skin primary fibroblast (1x10⁴/well of 96-well) from Ai14 LSL-tdT mice. At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 µm

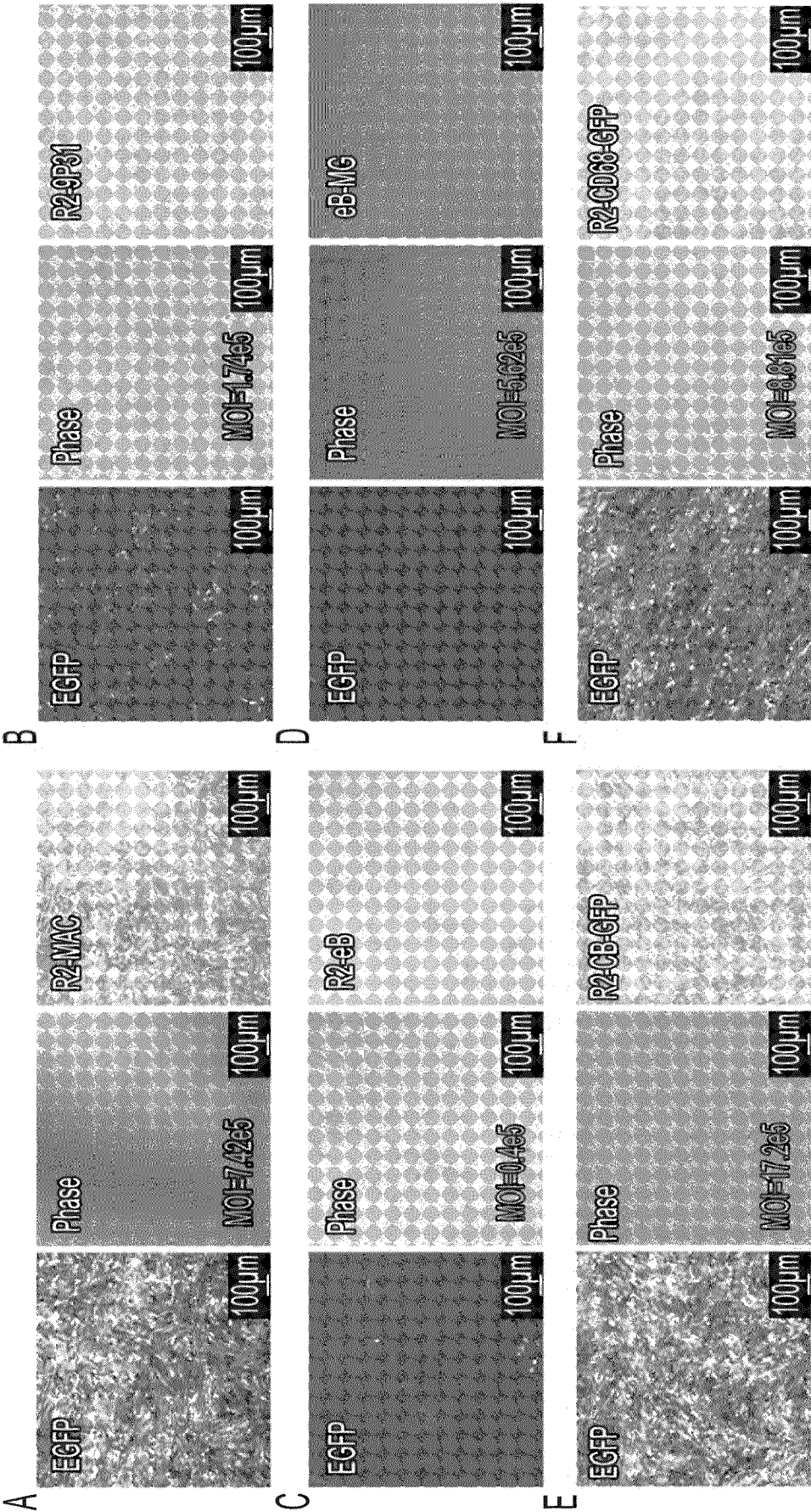
FIG. 6

The number in parathesis indicated the positive tdT cells in the entire well at 4 dpi



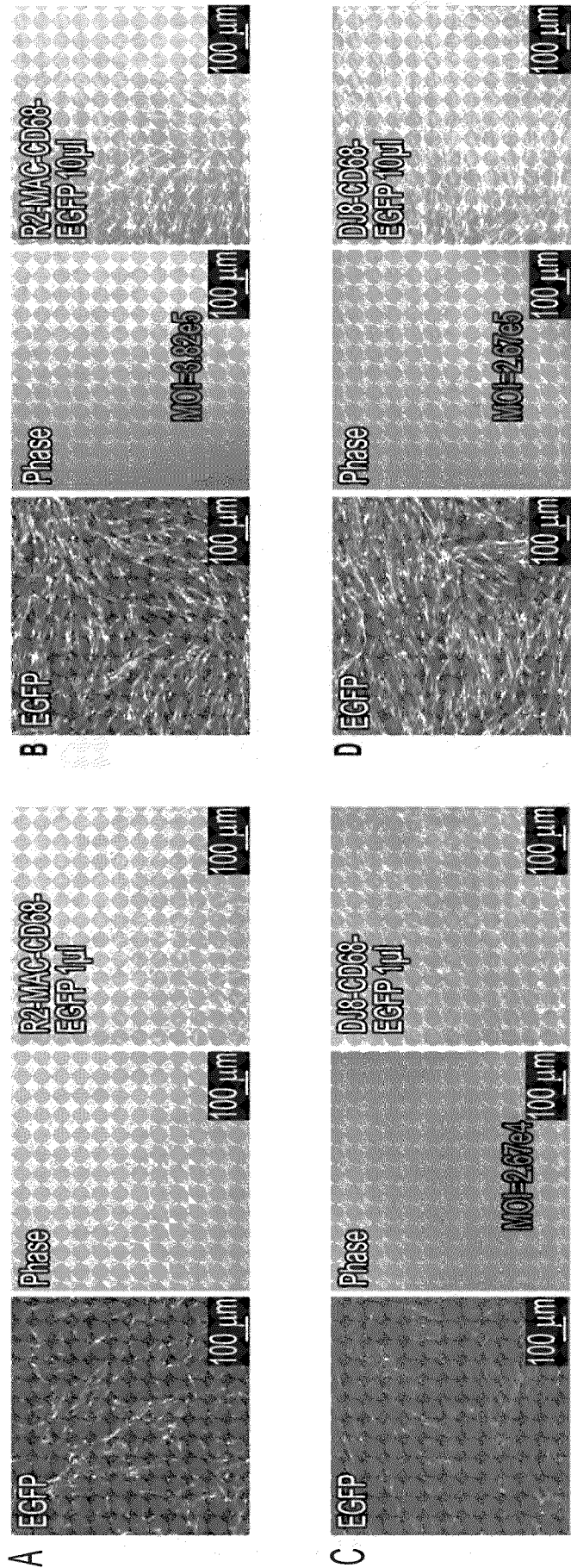
Figs. 7A-7F. R2-MAC shows stronger transduction efficient in human primary microglia than the original AAV2-RC serotype but the other three AAV-BM lost the transduction. HEK293T cells in 10-cm dish were transferred with helper vector, transger vector TP1152 (scAAV-CB-GFP, A-E) or TP1987 (scAAV-CD68-135-UTR-EGFP, F) and indicated serotype vector. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 1 µl was used to infect human primary microglia (1x10⁴/well of 96-well). At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 µm. No toxicity present in all groups.

FIGS. 7A-7F



Figs. 8A-8F. R2-MAC shows stronger transduction efficiency than AAV2 (normalized by moi) in human primary microglia, but the other three AAV-BM lost the transduction. HEK293T cells in 10-cm dish were transfected with helper vector, transfer vector TP1152 (scAAV-CB-GFP, A-E) or TP1987 (scAAV-CD68-135-UTR-EGFP, F) and indicated serotype vector. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 10 µl was used to infect human primary microglia (1x10⁴/well of 96-well). At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 µm. No toxicity present in all groups.

FIGS. 8A-8F



Figs. 9A-9D. R2-MAC shows higher transduction efficiency in mouse NSC than AAV-DJ8. HEK293T cells in 10-cm dish were transfected with helper vector, transfer vector TP1987 (scAAV-hCD68-135-UTR-EGFP) and indicated serotype vectors AAV2-MAC and AAV-DJ8. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 1 and 10 µl of purified virus (in 1x DPBS) was used to infect mouse NSC (1x10⁴/well of 96-well). At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 µm. Note: MG-specific promoter hCD68 still has activity in NSC.

FIGS. 9A-9D

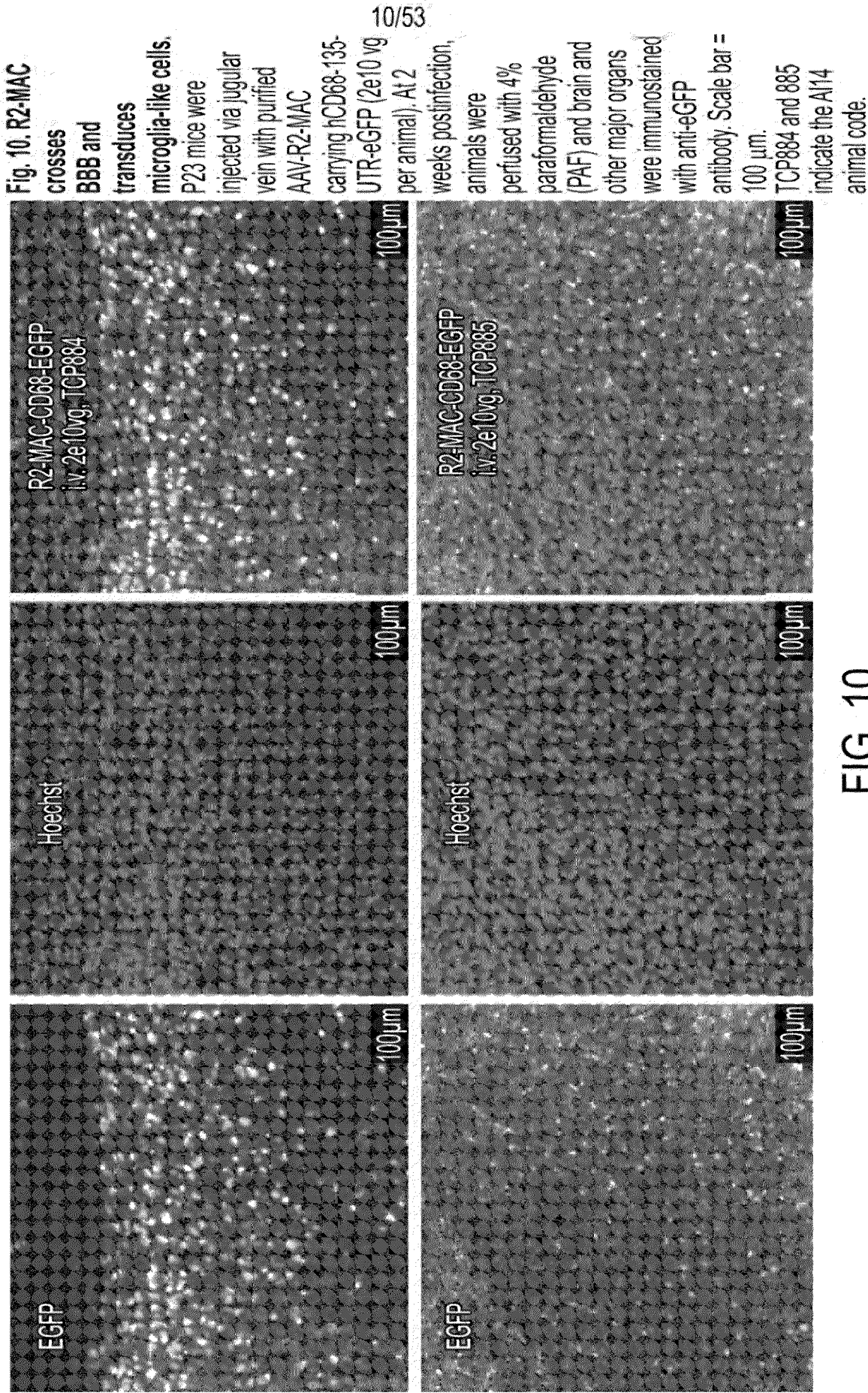


FIG. 10

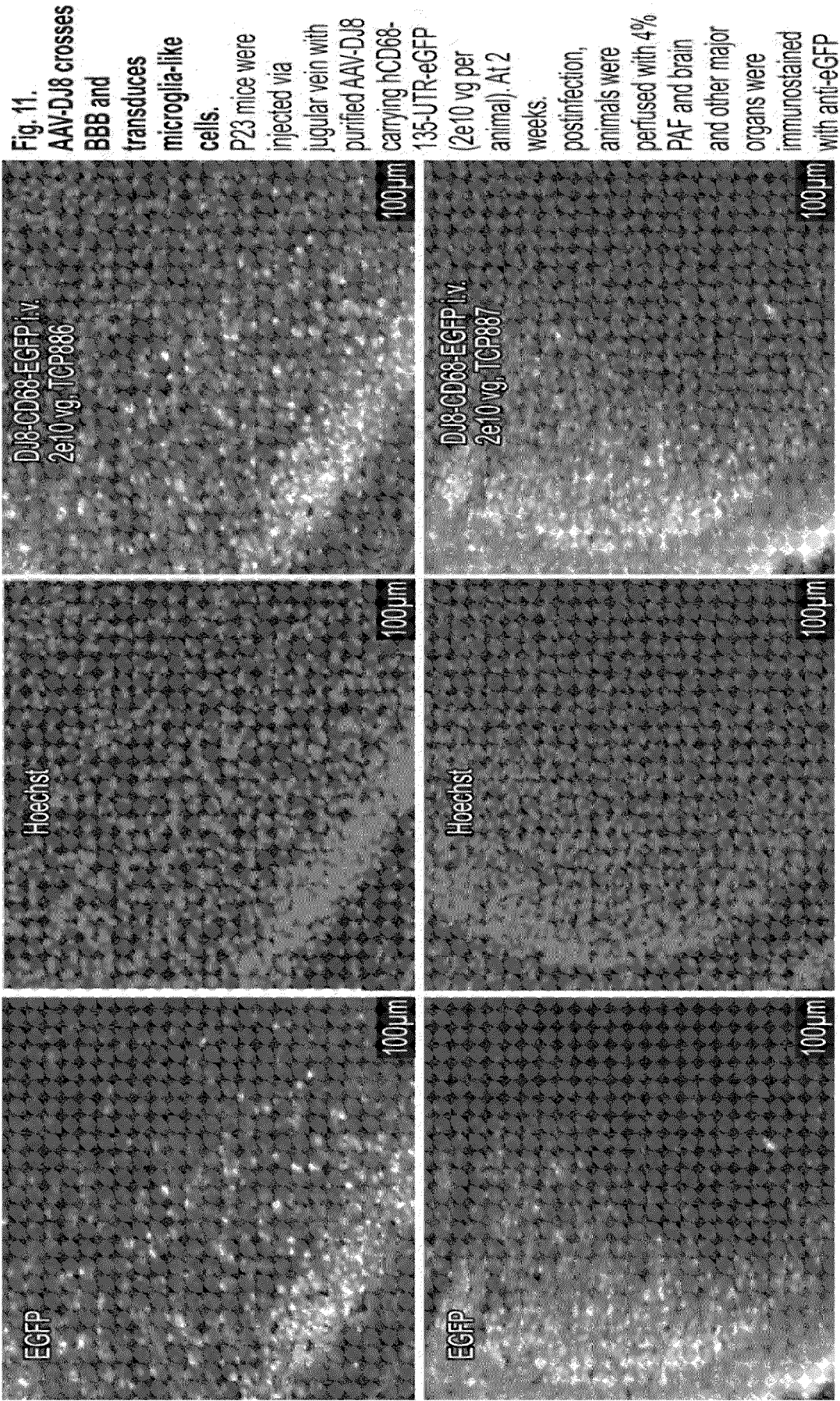
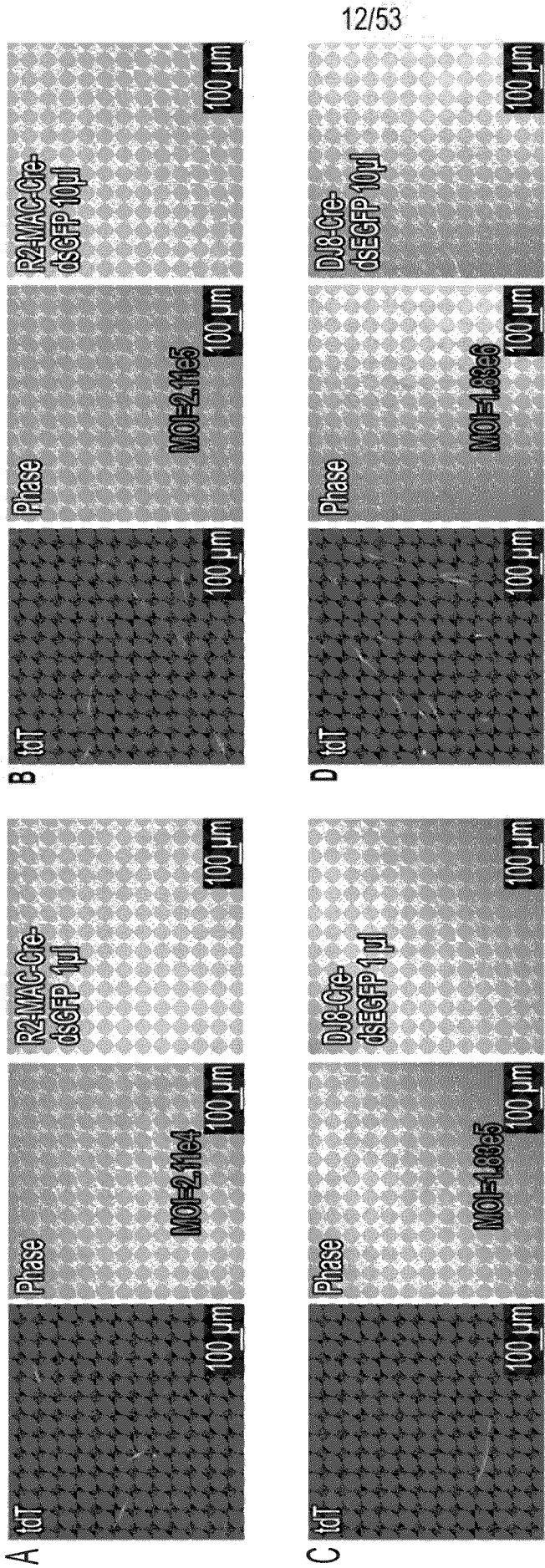


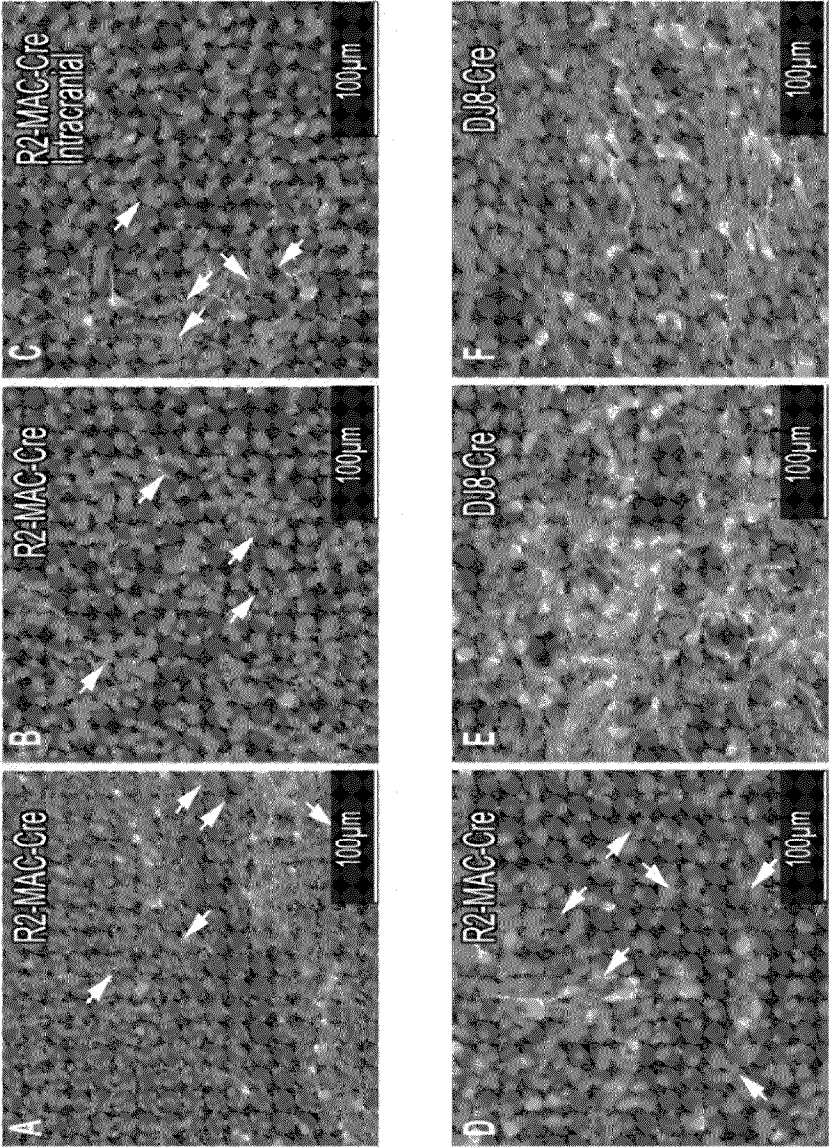
FIG. 11



Figs. 12A-12D. R2-MAC shows moderate transduction efficiency in mouse neural stem cells (NSC) at MOI of 2x10⁵ GC. HEK293T cells in 1x10-cm dish were transfected with helper vector, transfer vector TP1960 (AAV-CMV-Cre-GFP-CD63-MG1MG2-MCP) and indicated serotype vectors. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 1 and 10 µl of purified virus (in 1x DPBS) was used to infect mouse NSC (1x10⁴/well of 96-well). At 4 dpi, confocal images were taken at 10x objective. Scale bar = 100 µm. More positivity in AAV-DJ8 resulted from higher MOI.

FIGS. 12A-12D

Figs. 13A-13F. microinjection of AAV-R2-MAC induce effective delivery of Cre to microglia (arrows) and neurons but DJ8 transduced only neurons. R2-MAC (A-D) or AAV-DJ8 (E, F) carrying AAV-CMV-Cre (TP1960) were injected stereotactically into striatum and hippocampus of adult A14 mice (3-8 months old). At 2 weeks postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μ m. **Dose=2 μ l/point:** 4.2e8 vg for R2-MAC and 3.6eg vg for DJ8 TP1960 AAV-CMV-Cre-GFP-CD63-MG1MG2-MCP



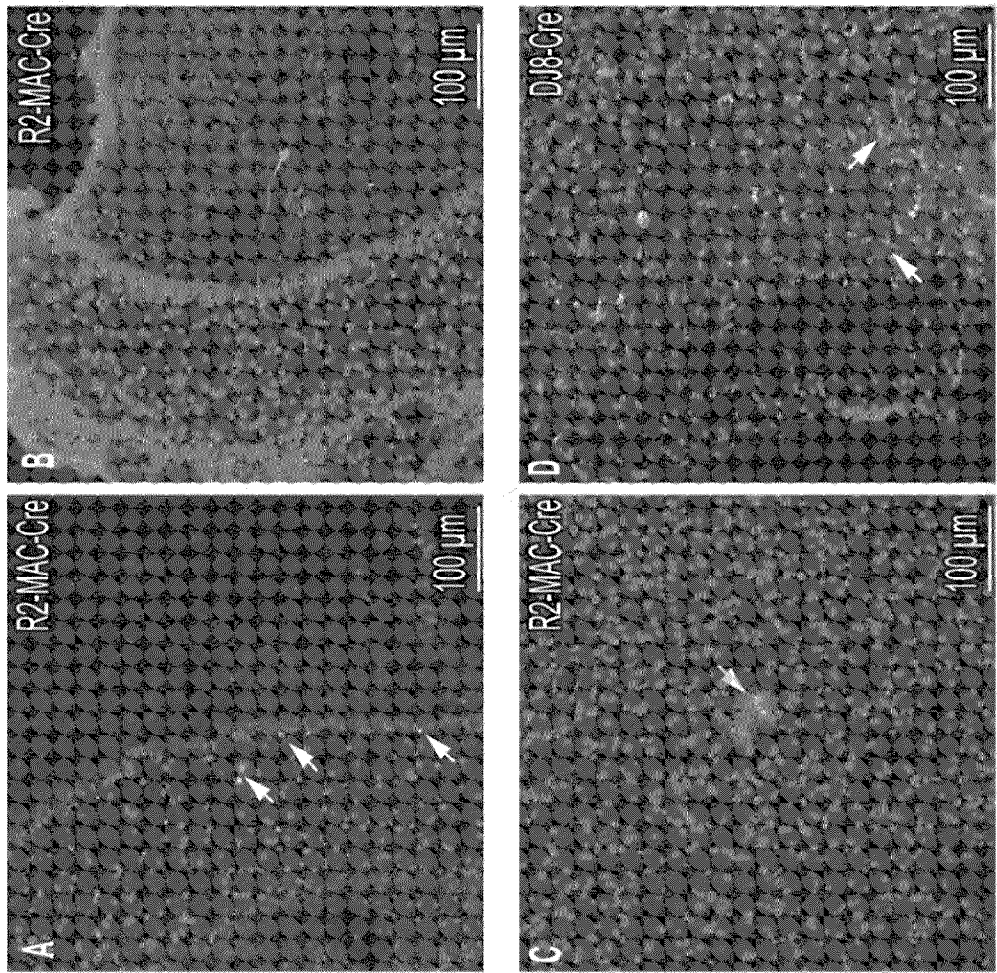
FIGS. 13A-13F

Figs. 14A-14D. I.V. injection of AAV-R2-MAC can cross BBB and transduce microglia (white arrow), neurons and microglia-like cells (red arrow). R2-MAC (A-C) or AAV-DJ8 (D) carrying AAV-CMV-Cre (TP1960) were injected via jugular vein of adult A14 mice (3-4 months old). At 2 weeks postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μ m. Note: AAV-DJ8 has more positive cells due to higher dosage.

R2-Mac: Dose=1.1e12 vg/kg

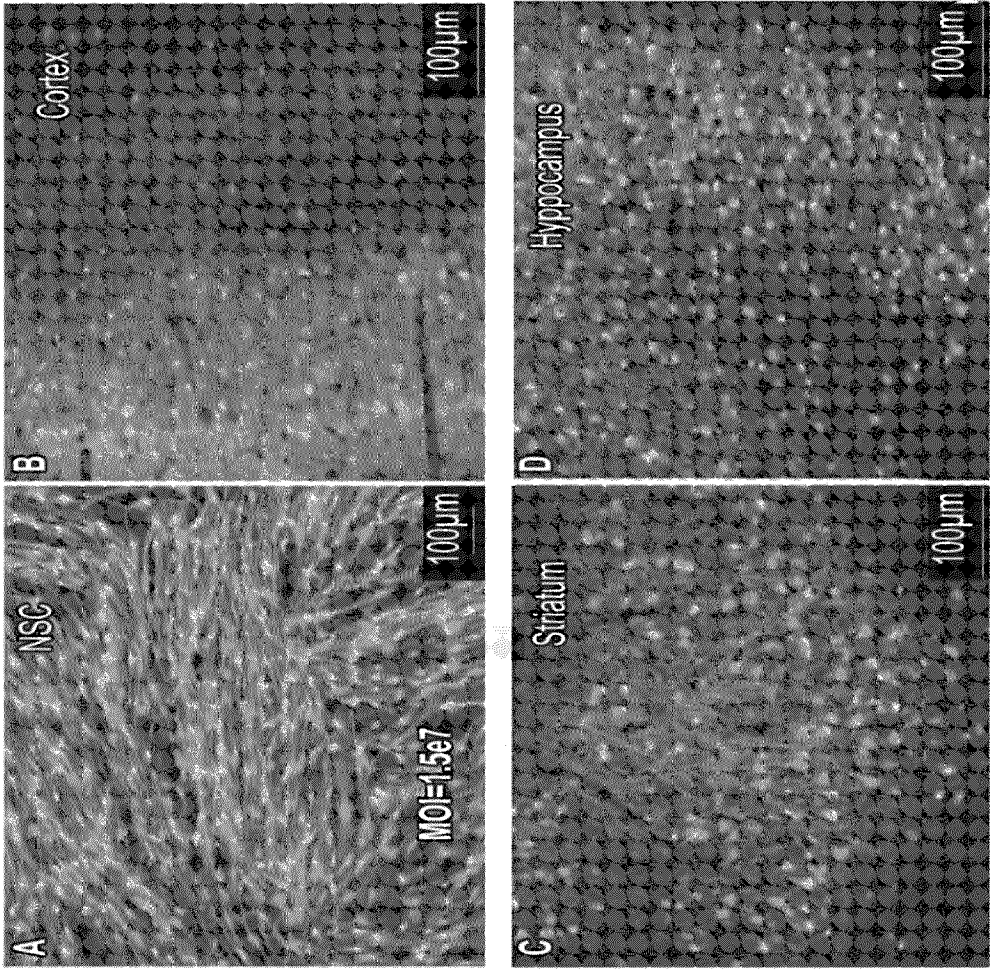
DJ8: Dose=9.1e12 vg/kg

TP1960: AAV-CMV-Cre-GFP-CD63-MG1MG2-MCP



FIGS. 14A-14D

Figs. 15A-15D. Higher titer of AAV-R2-MAC-CMV-Cre induce effective delivery of Cre to cultured NSC (A), and in vivo microglia-like cells and neurons (B-D). HEK293T cells in 3x10-cm dishes were transfected with helper vector, TP1989 (R2-MAC) and TP1921 transfer vector (AAV-CMV-Cre-CD63-MG1MG2-MCP). At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. 9 μ l of purified virus (in 1x DPBS) was used to infect AI14 mouse NSC (1x10⁴/well of 96-well). At 6 dpi, confocal images were taken at 10x objective (A). The same virus was injected stereotactically into cortex (B), striatum (C) and hippocampus (D) of adult AI14 mice (TCP75, Female, 7 months old) at 2 μ l/point ($\approx$ 3.7e9 vg). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μ m.



FIGS. 15A-15D

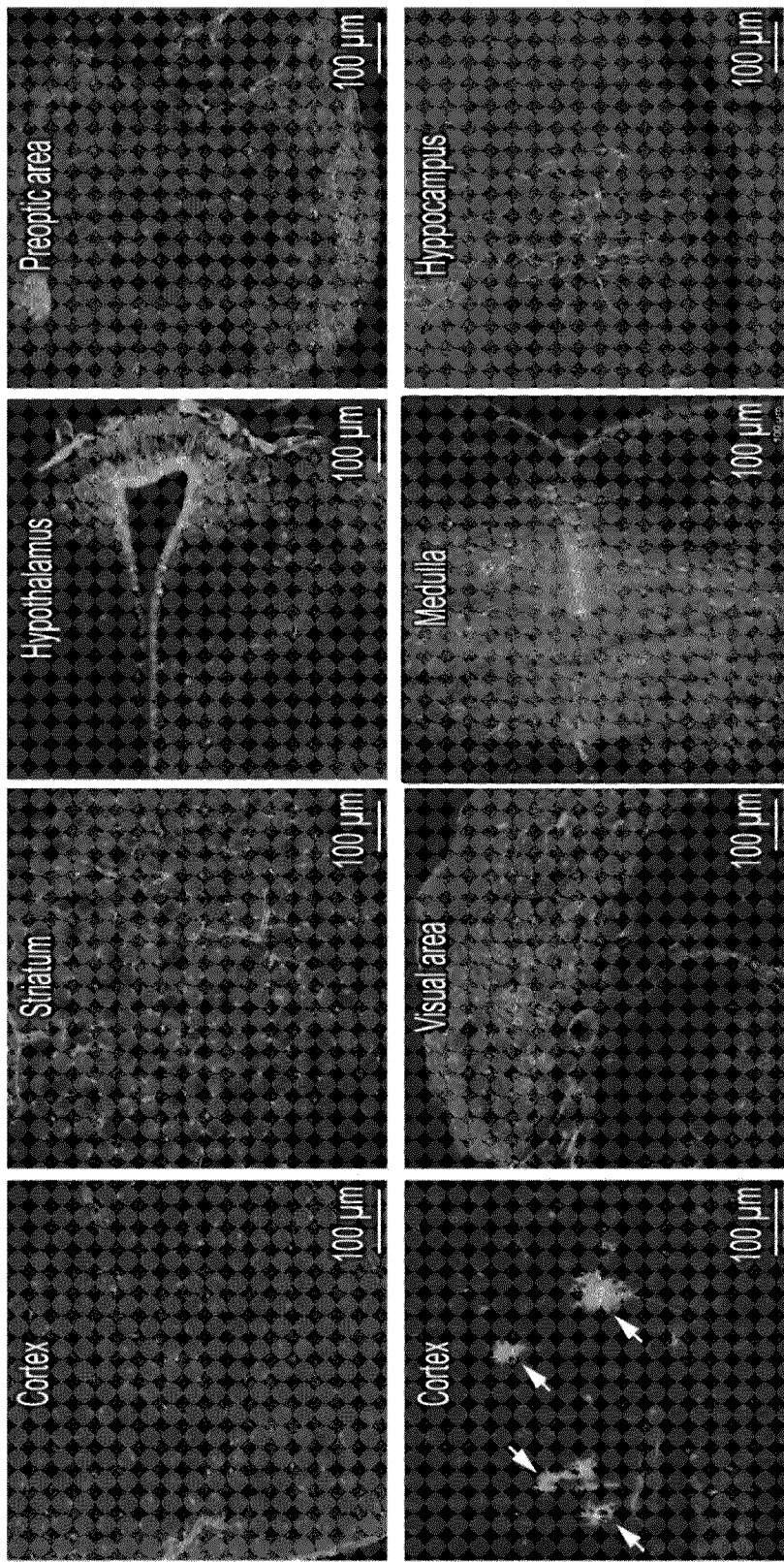


Fig. 16. Higher titer of AAV-R2-MAC-CMV-Cre induce extensive transduction throughout the brain, mainly in microglia-like cells (white arrows), neurons, astrocytes based on morphology. HEK293T cells in 3x10-cm dishes were transfected with helper vector, TP1989 (R2-MAC) and TP1921 transfer vector (AAV-CMV-Cre-CD63-MG1MG2-MCP). At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. Around 70 ul of purified virus (5.6e13 vg/kg) was injected intravenously in adult A14 mice (TCP90, male, 21g, 2 months old). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μm.

FIG. 16

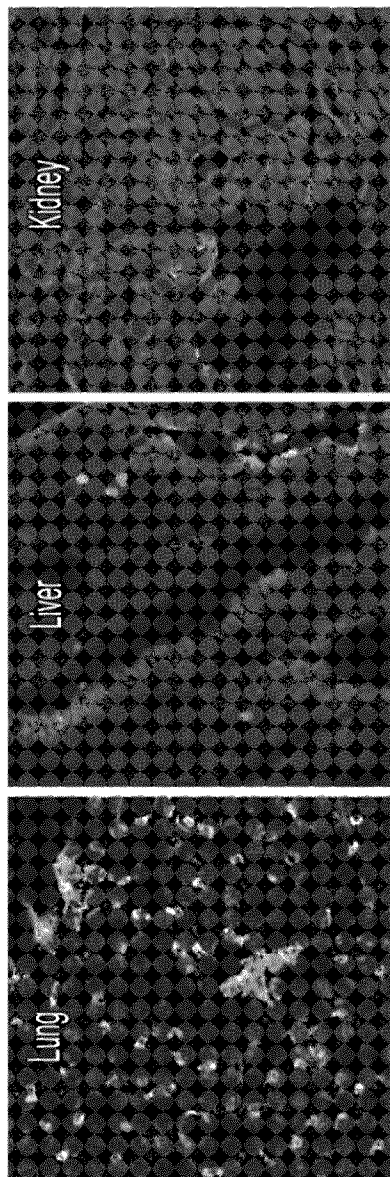


Fig. 17. AAV-R2-MAC-CMV-Cre also induces transduction of macrophage in various organs. Various organs of the same animals as described in Fig. 16 were collected for cryosection and tdTomato positive cells without immunostaining were imaged under confocal microscope.

FIG. 17

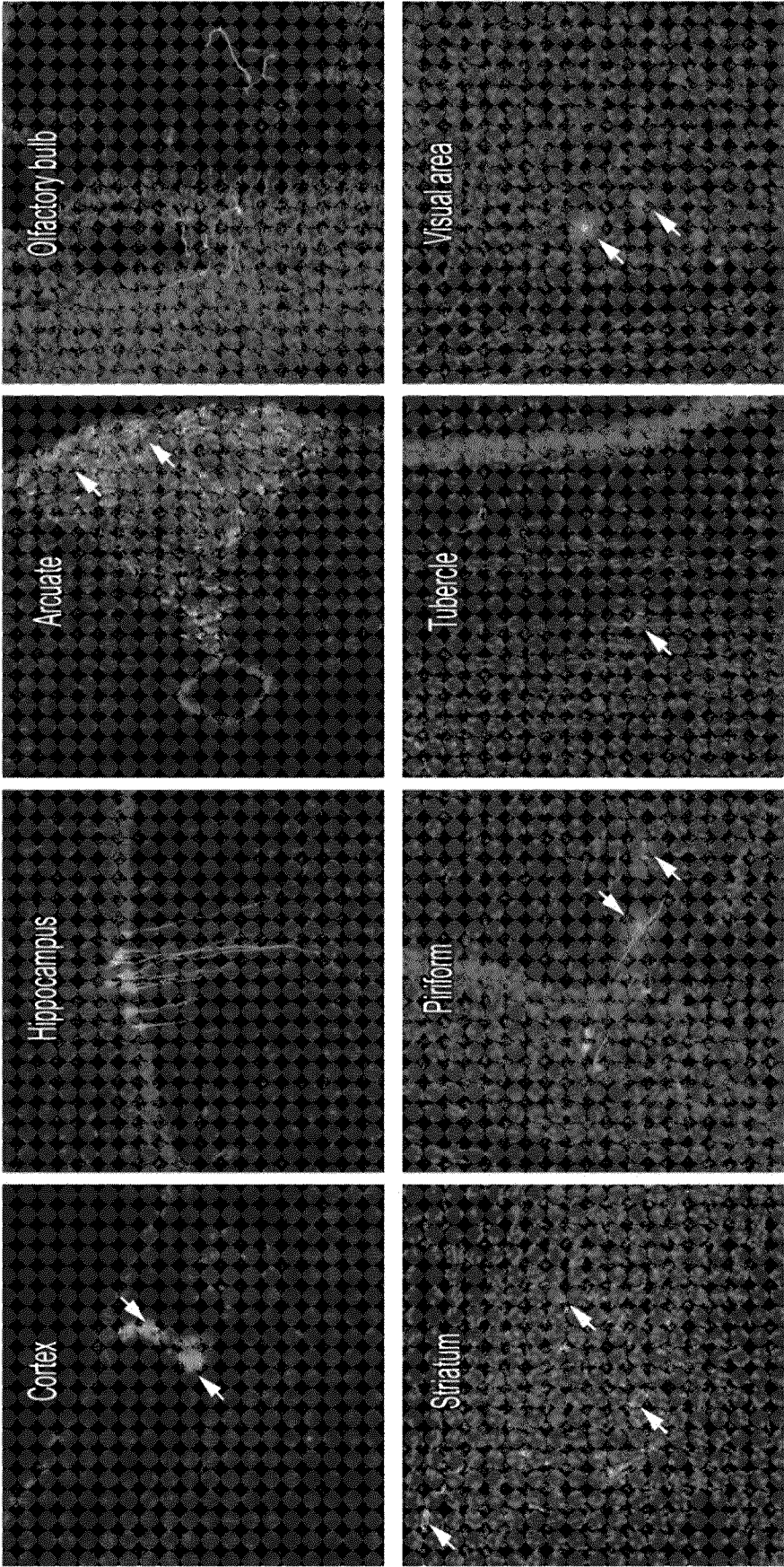


Fig. 18. AAV-R2-MAC-CMV-Cre induce transduction throughout the brain, in microglia-like cells (white arrows), neurons, astrocytes based on morphology. The same amount of purified virus as described in Fig. 16 was injected intravenously in adult A14 mice (TCP91, male, 22g, 2 months old). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were imaged under confocal microscope. Scale bar = 100 μ m.

FIG. 18

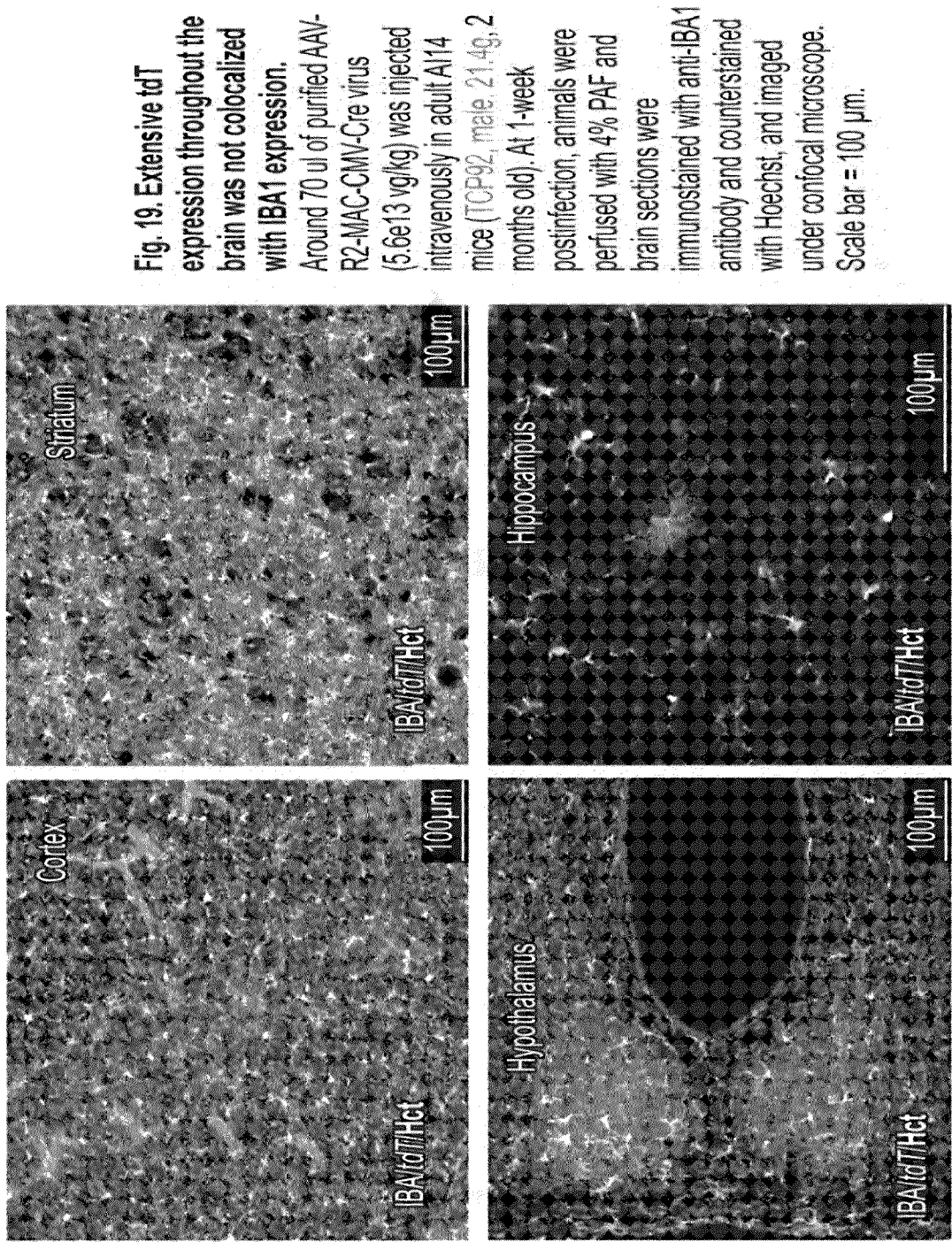


FIG. 19

Fig. 19. Extensive tdT expression throughout the brain was not colocalized with IBA1 expression. Around 70 ul of purified AAV-R2-MAC-CMV-Cre virus (5.6e13 vg/kg) was injected intravenously in adult A14 mice (TCP92, male, 21.4g, 2 months old). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were immunostained with anti-IBA1 antibody and counterstained with Hoechst, and imaged under confocal microscope. Scale bar = 100 μm.

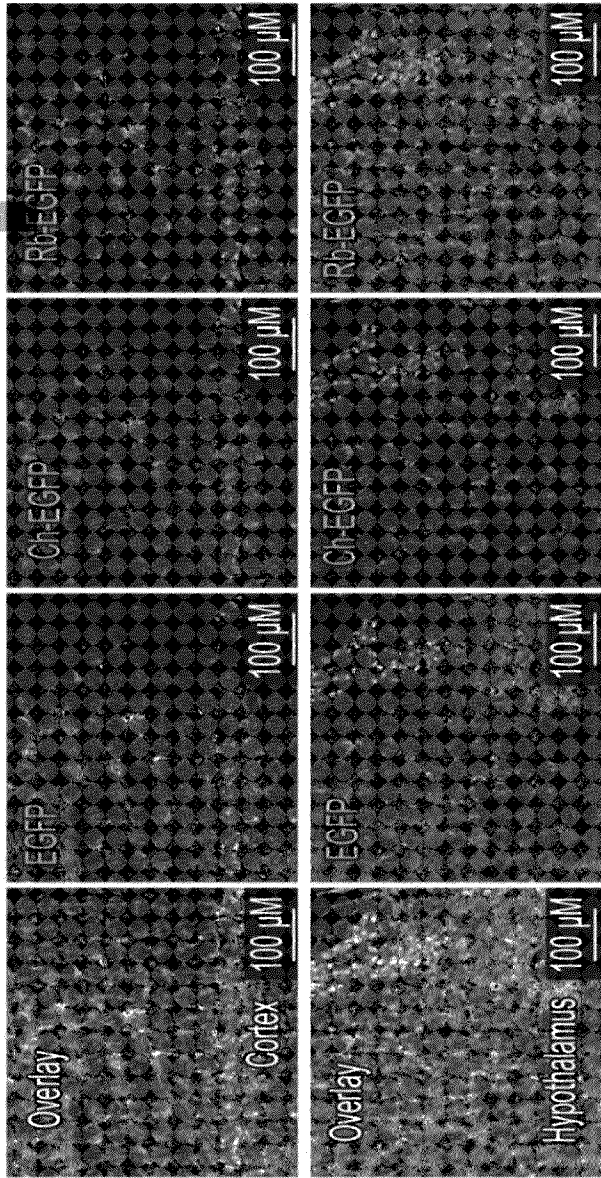


Fig. 20. Higher titer of AAV-R2-MAC-hHexB-134-UTR-eGFP induces effective transduction in microglia-like cells. HEK293T cells in 3x10-cm dishes were transfected with helper vector, TP1989 (R2-MAC) and TP1986 transfer vector (scAAV-hHexB-134-UTR-eGFP). At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30K) were performed. Around 50 ul of purified virus (9.6e13 vg/kg) was injected intravenously of adult wildtype mice (Tg26Q115, 25.9g, male, 3 months old). At 1-week postinfection, animals were perfused with 4% PAF and brain sections were immunostained with Chicken anti-GFP and Rabbit anti-GFP antibodies. Scale bar = 100 μm.

FIG. 20

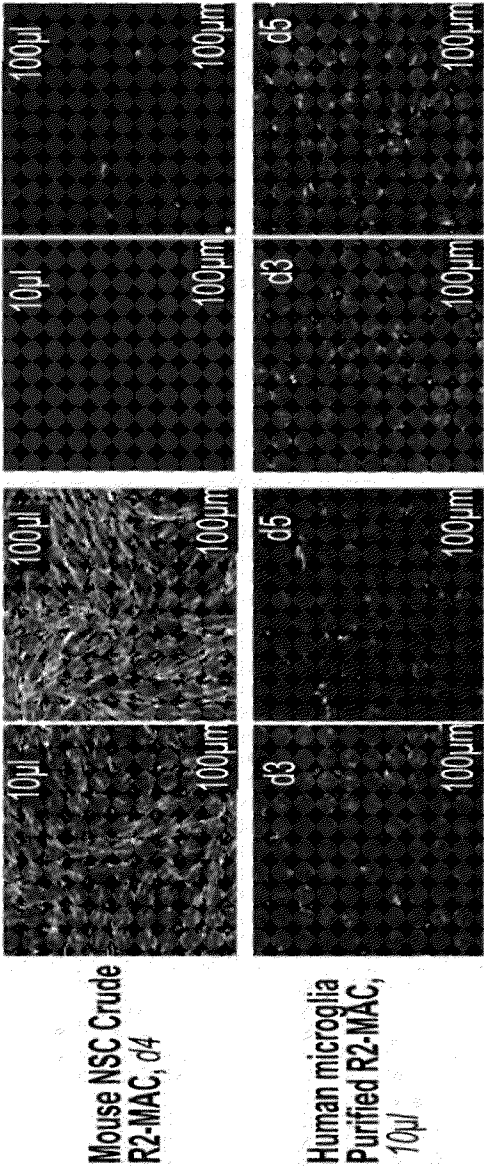
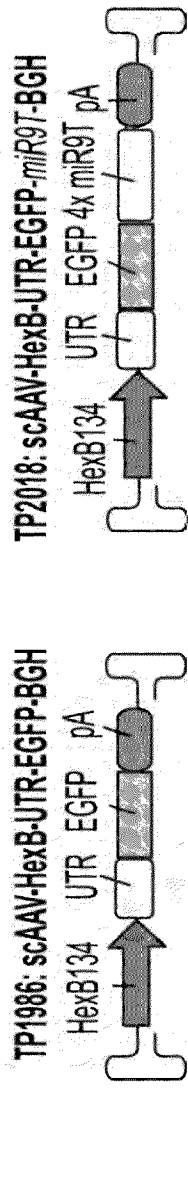


Fig. 21. Inclusion of miRNA 9 target (miR9T) in AAV-R2-MAC-hHexB-134-UTR-eGFP virus induces specific transgene expression in microglia but not neural stem cells (NSC). HEK293T cells in 6-well plate were transfected with helper vector, TP1989 (R2-MAC) and indicated transfer vector. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30K) were performed. Crude or purified viruses were used to treat mouse NSCs or human primary microglia, and at indicated days, confocal images were taken. Scale bar = 100µm.

FIG. 21

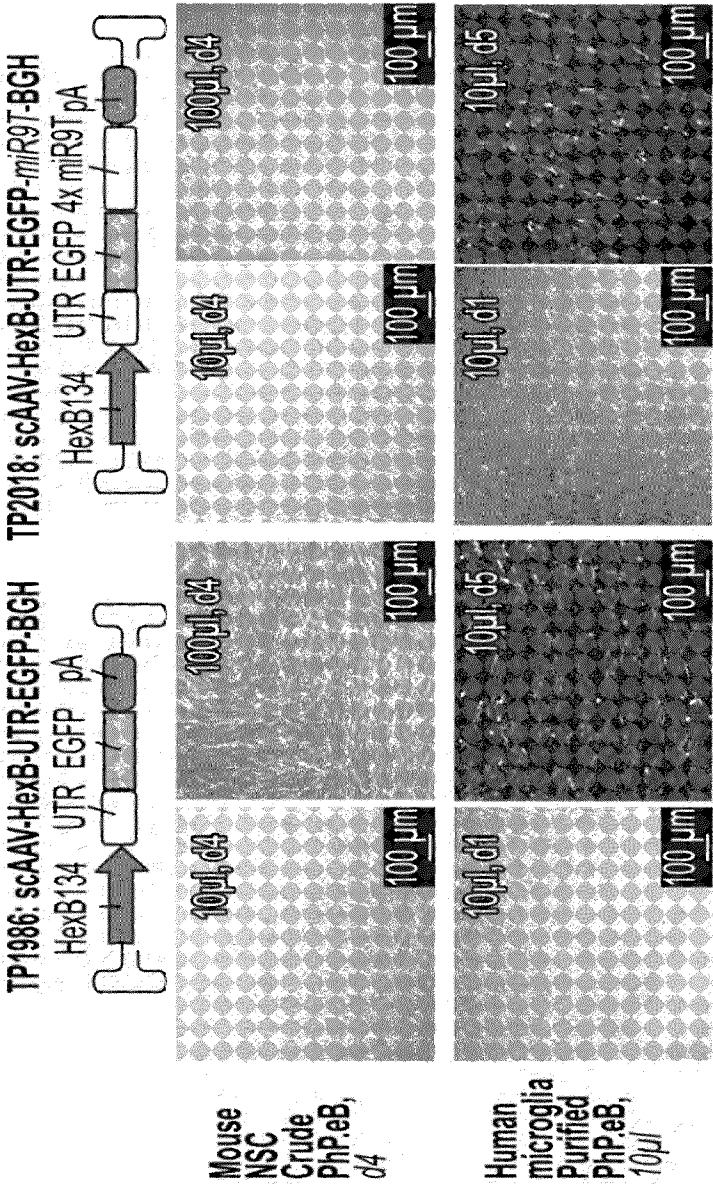


Fig 22. Inclusion of miRNA 9 target (miR9T) in AAV-PhP.eB-hHexB-134-UTR-eGFP virus induces specific transgene expression in microglia but not neural stem cells (NSC). HEK293T cells in 6-well plate were transfected with helper vector, TP1190 (AAV-PhP.eB) and indicated transfer vector. At 5 days, PEG/Chloroform purification and buffer exchange via ultrafiltration (30KD) were performed. Crude or purified viruses were used to treat mouse NSCs or human primary microglia, and at indicated days, confocal images were taken. Scale bar = 100µm.

FIG. 22

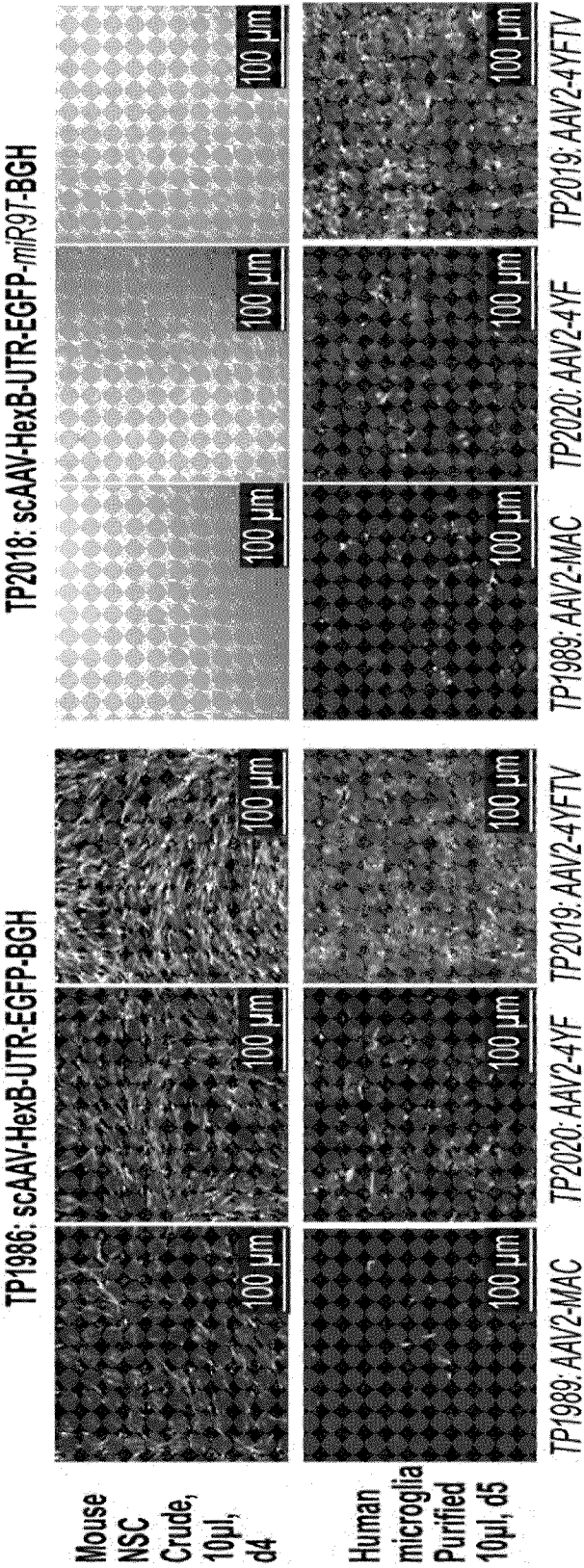


Fig. 23. AAV2 mutations at 4YF (Tyrosine to Phenylalanine) plus TV (Threonine to Valine) robustly increase transgene expression in mouse NSC and human microglia but miR97 addition illuminates microglia expression. HEK293T cells in a 6-well plate were transfected with helper vector, indicated capsids and transfer vectors. On day 5, PEG/Chloroform purification and ultrafiltration (30KD) were performed. Crude or purified viruses were used to treat mouse NSCs or human primary microglia, and at indicated days, confocal images were taken. Scale bar = 100µm.

FIG. 23

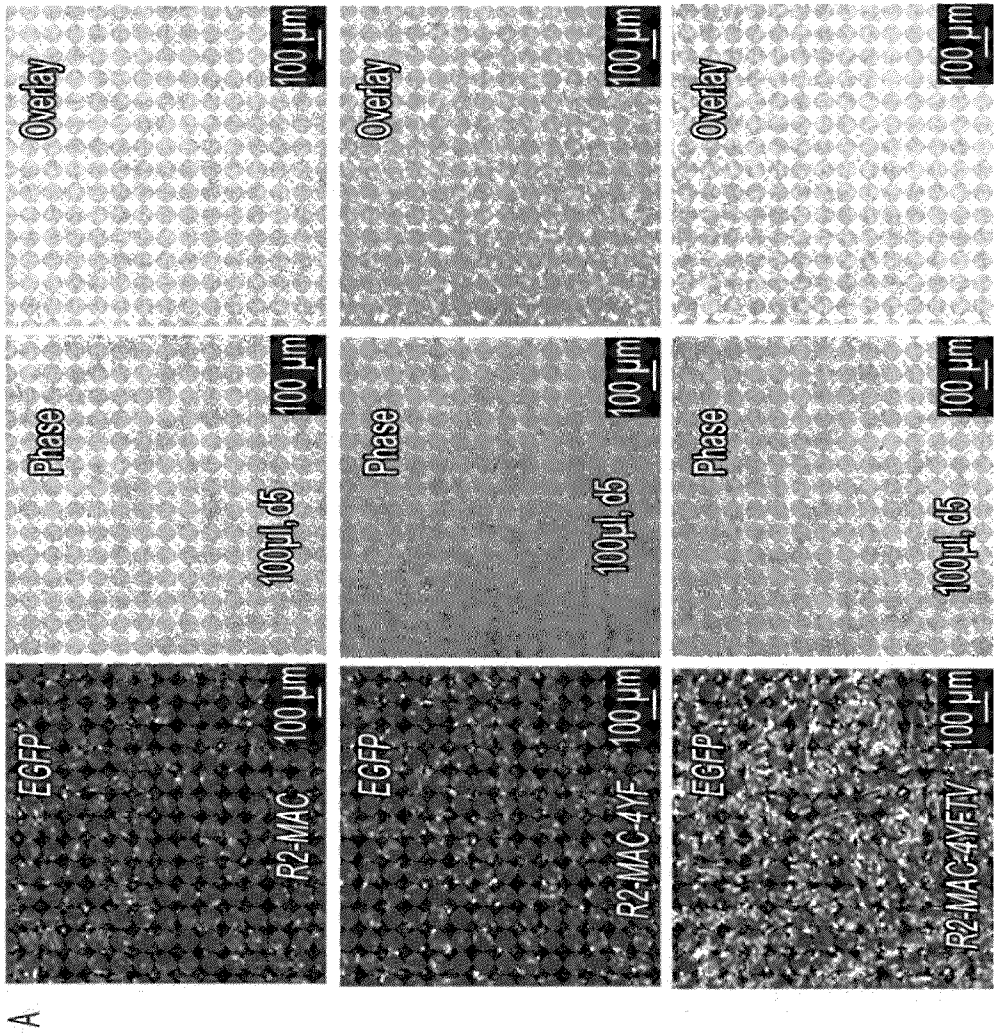
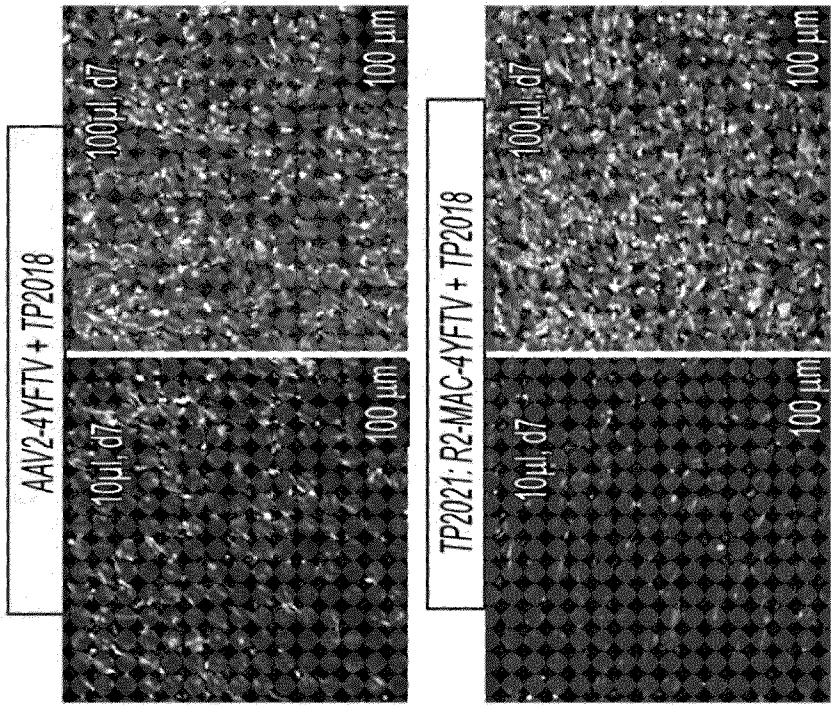


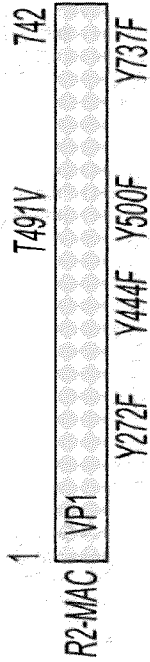
Fig. 24A. R2-MAC mutations at 4YF plus TV robustly increase transgene expression with miR9T in human primary microglia. (A) Combined mutation (4YF/TV) significantly improved the transduction efficiency. (B) R2-MAC-4YFV (R2e-MAC) mutants showed similar transduction efficiency to AAV2-4YFV (R2e) mutant. (C) Diagram of mutation residues. HEK293T cells in 6-well plate were transfected with three plasmids. At day 5, PEG/Chloroform purification and ultrafiltration (30KD) were performed. Crude viruses were used to treat human microglia, and at day 7, confocal images were taken. Scale bar = 100 μm.

FIG. 24A

B

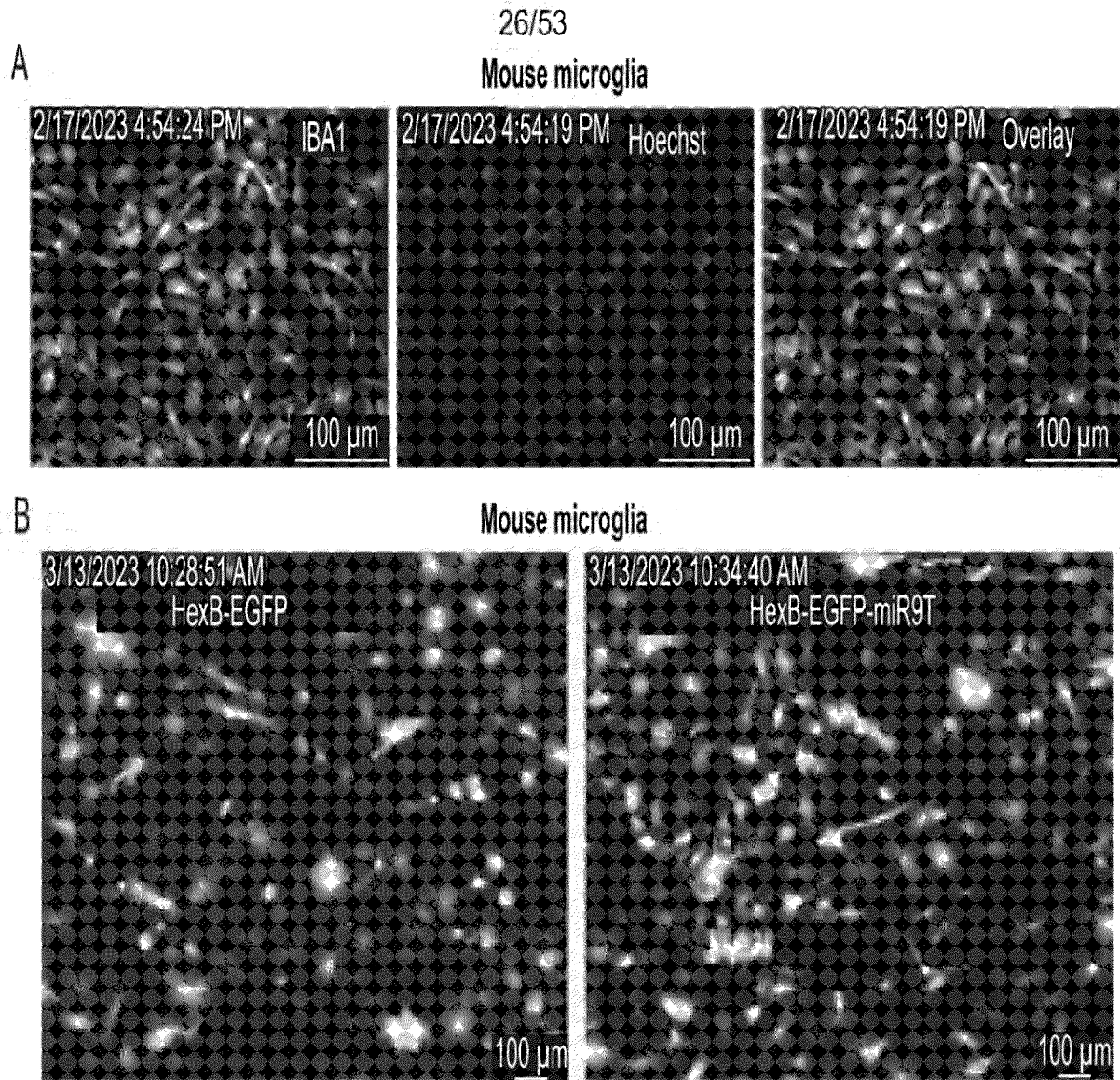


C



Figs. 24B-24C. R2-MAC mutations at 4YF plus TV robustly increase transgene expression with miR9T in human primary microglia. (A) Combined mutation (4YF/TV) significantly improved the transduction efficiency. **(B)** R2-MAC-4YFTV (R2e-MAC) mutants showed similar transduction efficiency to AAV2-4YFTV (R2e) mutant. **(C)** Diagram of mutation residues. HEK293T cells in 6-well plate were transfected with three plasmids. At day 5, PEG/Chloroform purification and ultrafiltration (30KD) were performed. Crude viruses were used to treat human microglia, and at day 7, confocal images were taken. Scale bar = 100 µm.

FIGS. 24B-24C

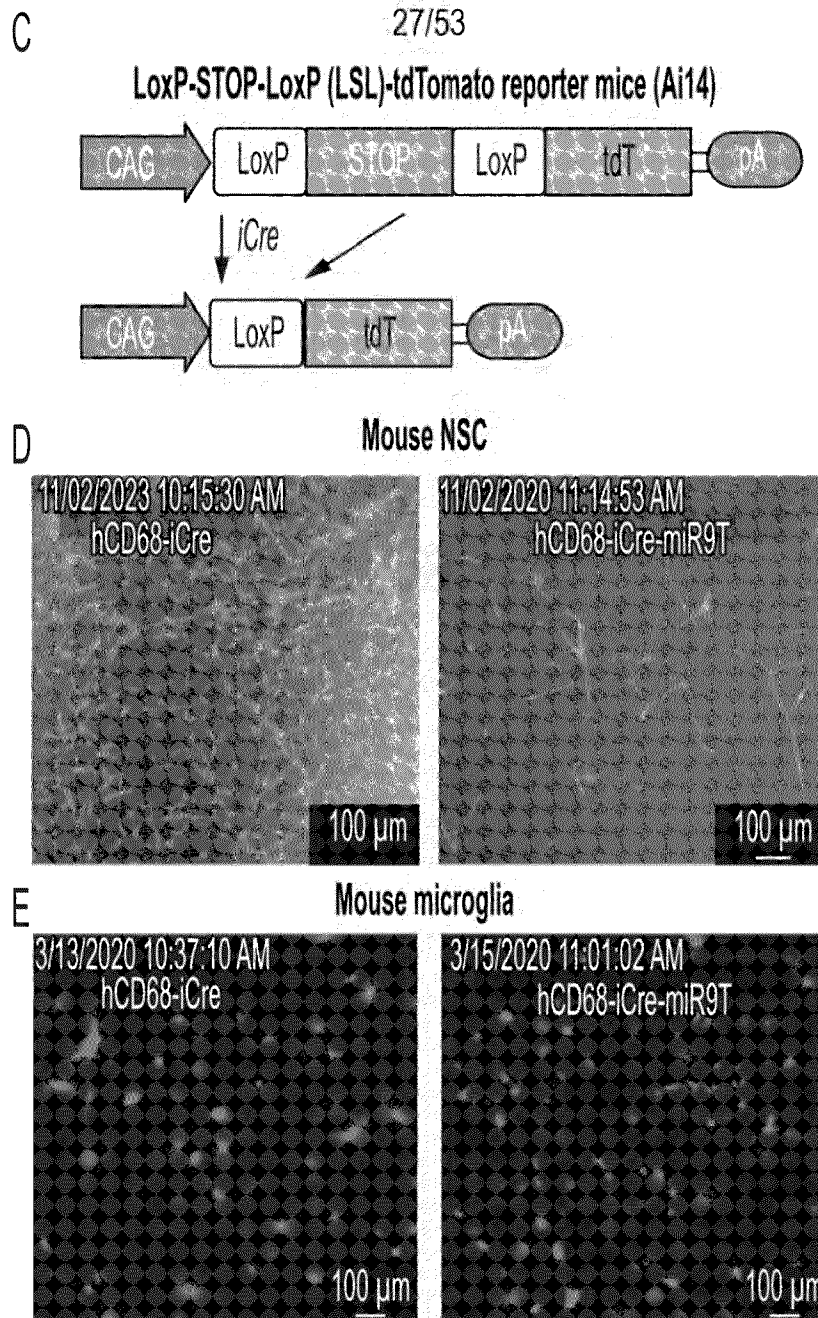


Figs. 25A-25B, The 4x repeated miRNA-9 target site(miR9T) was added downstream of WPRE3 to illuminate microglia expression.

(A) Mouse

primary microglia validated by immunocytochemistry with anti-IBA1 antibody. **(B)** R2-MAC-4YF/TV efficient delivery of scAAV-HexB-UTR-EGFP-miR9T-BGH(TP2018) to mouse microglia with or without miR9T. **(C)** Diagram of the improved Cre (iCre)-mediated LoxP recombination to induce tdTomato (tdT) expression in Ai14 transgenic mice. **(D, E)** R2-MAC-4YF/TV increases transgene expression in mouse NSC and human microglia without miR9T (TP2057: scAAV-hCD68-iCre-3xHA-WPRE3-BGH) but miR9T addition illuminates microglia expression (TP2058: scAAV-hCD68-iCre-3xHA-WPRE3-miR9T-BGH). Equal MOI (1.8×10^6) of the purified virus was added to indicated mouse NSC **(D)** or microglia **(E)**. On day 8, confocal images were taken. Scale bar = 100 µm.

FIGS. 25A-25B

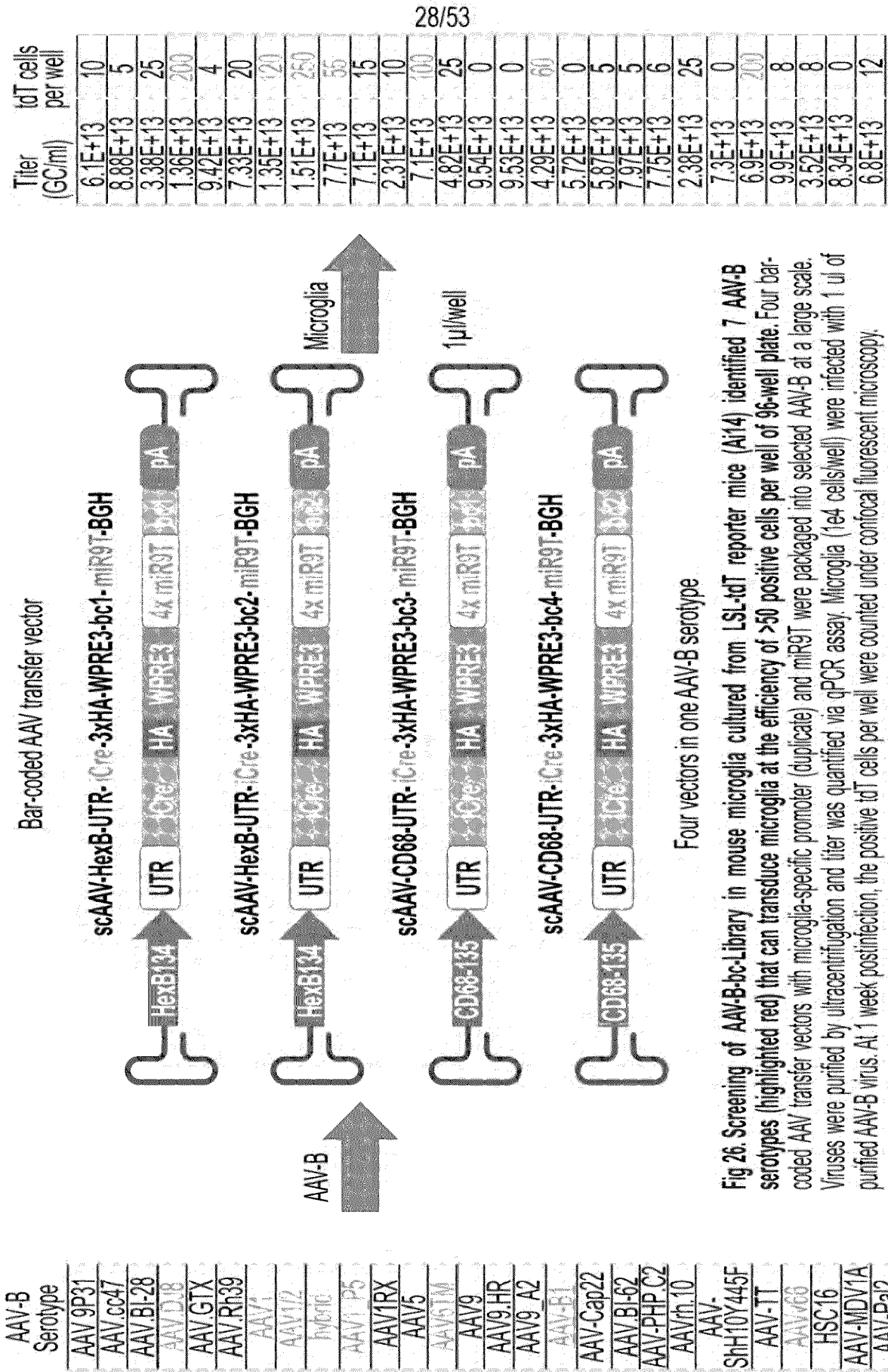


Figs. 25C-25E. The 4x repeated miRNA-9 target site(miR9T) was added downstream of WPRE3 to illuminate microglia expression.

(A) Mouse

primary microglia validated by immunocytochemistry with anti-IBA1 antibody. (B) R2-MAC-4YF/TV efficient delivery of scAAV-HexB-UTR-EGFP-miR9T-BGH(TP2018) to mouse microglia with or without miR9T. (C) Diagram of the improved Cre (iCre)-mediated LoxP recombination to induce tdTomato (tdT) expression in Ai14 transgenic mice. (D, E) R2-MAC-4YF/TV increases transgene expression in mouse NSC and human microglia without miR9T (TP2057: scAAV-hCD68-iCre-3xHA-WPRE3-BGH) but miR9T addition illuminates microglia expression (TP2058: scAAV-hCD68-iCre-3xHA-WPRE3-miR9T-BGH). Equal MOI (1.8e+6) of the purified virus was added to indicated mouse NSC (D) or microglia (E). On day 8, confocal images were taken. Scale bar = 100 μm.

FIGS. 25C-25E



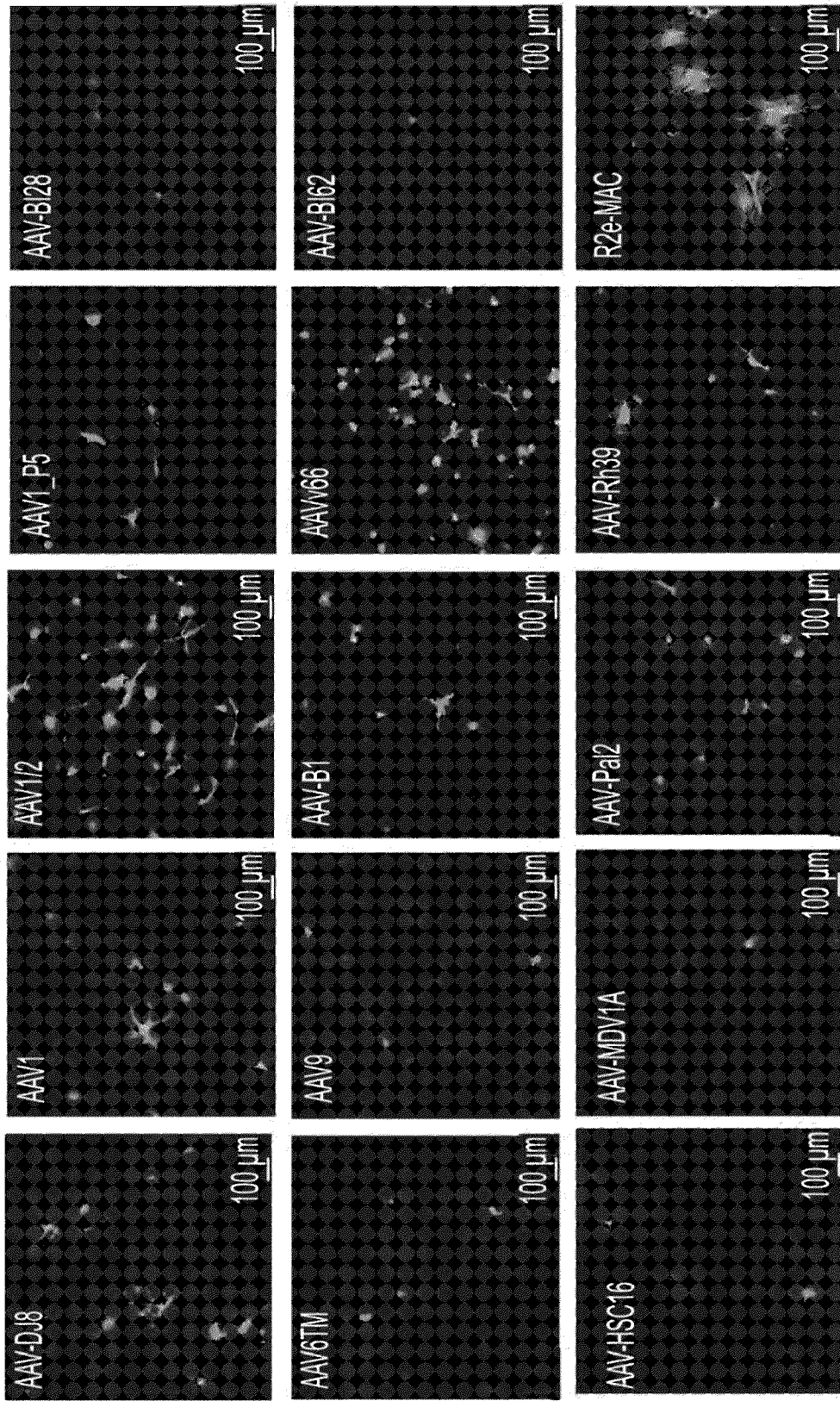
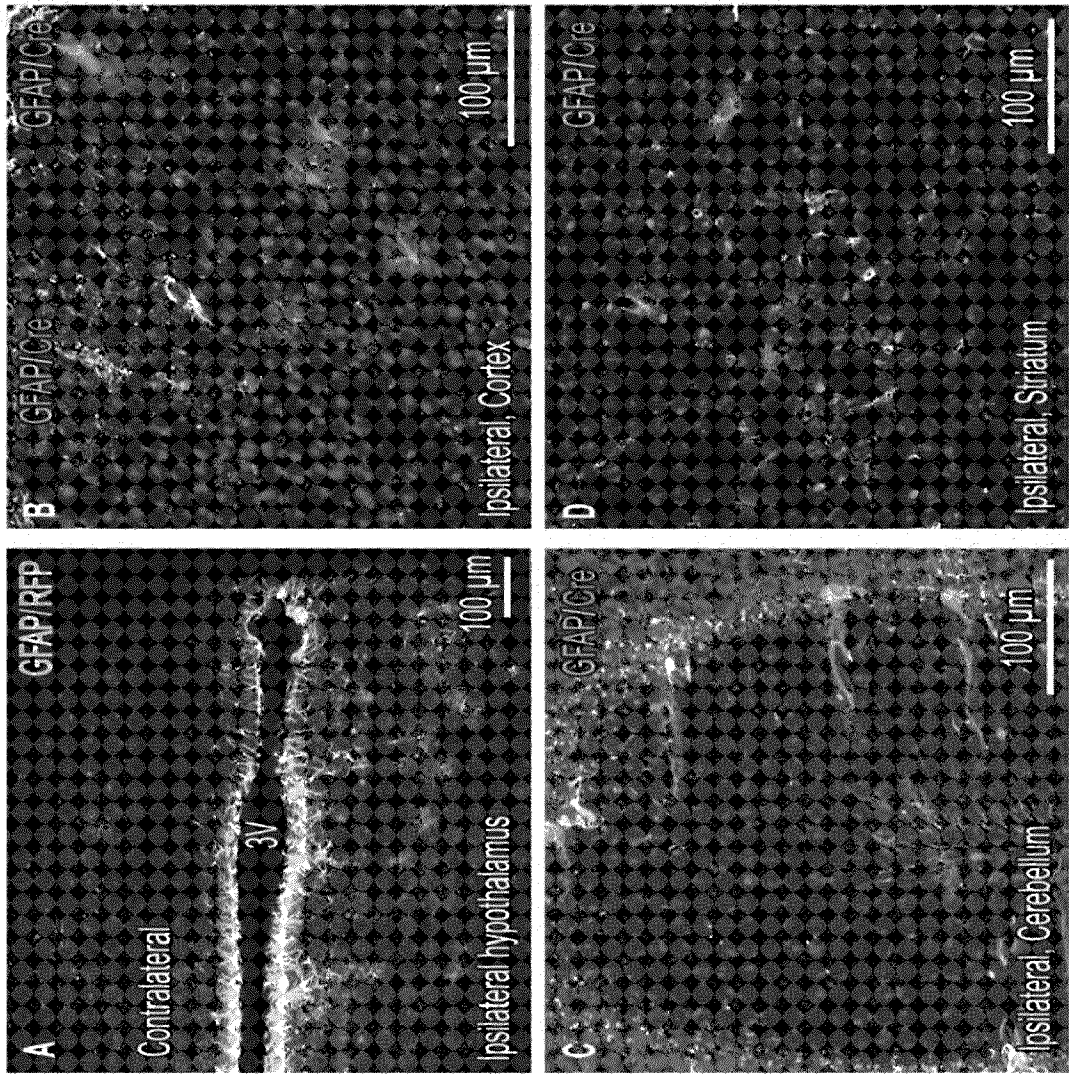


Fig. 27. Representative images for AAV-B-bc-library transduction in mouse microglia cultured from LSL-tdT reporter mice (Ai14). Four bar-coded AAV transfer vectors with microglia-specific promoter (duplicate) and miR9T were packaged into selected AAV-Bata large scale. Viruses were purified by ultracentrifugation and titer was quantified via qPCR assay. Microglia (1e4 cells/well) were infected with 1 ul of purified AAV-B virus. At 1 week postinfection, the positive tdT cells per well were imaged and counted under confocal fluorescent microscopy. Scale bar=100μm

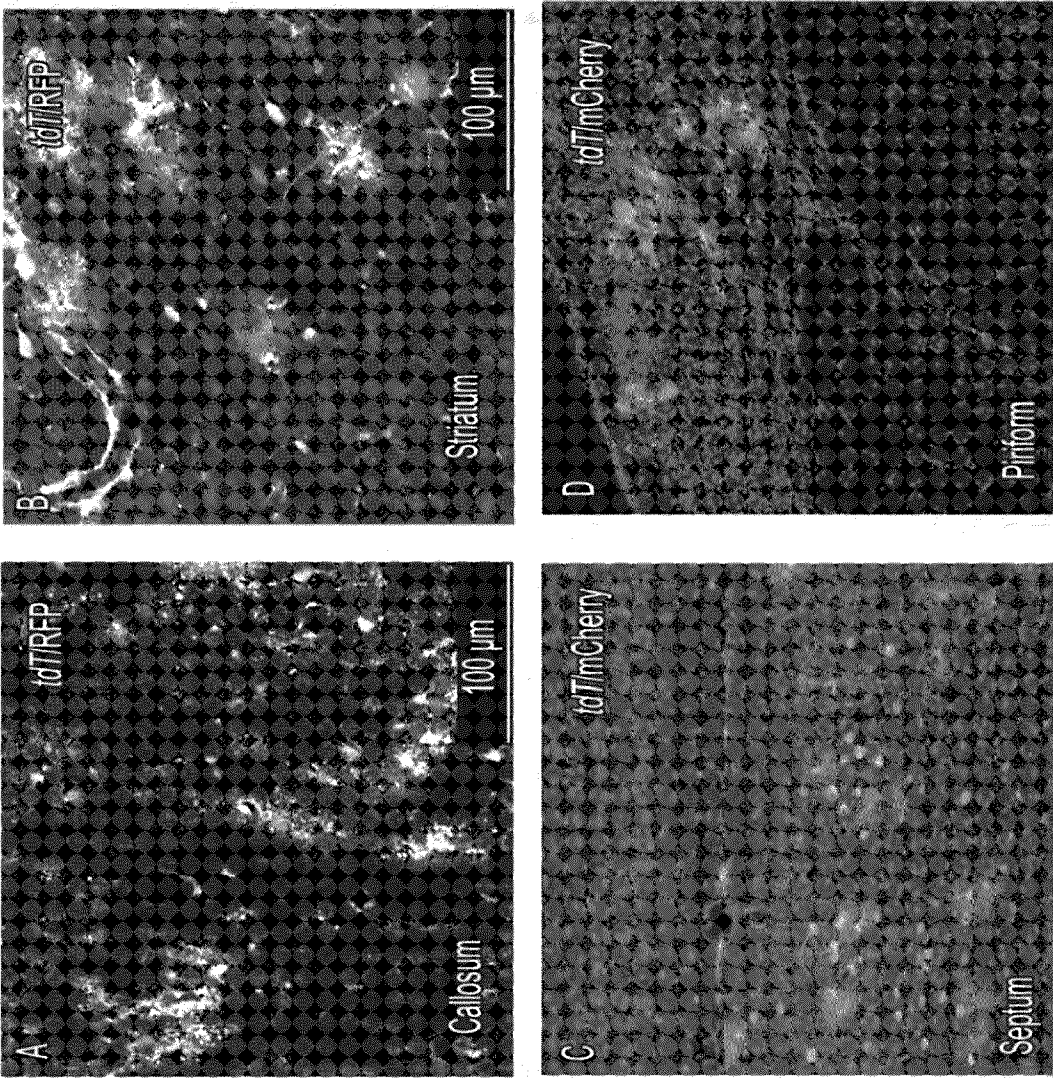
FIG. 27

Figs. 28A-28D.
(A) Brain injection of mixed 8 AAV serotype (0.3-2.5e+10 GC, 2 μ l/site) carrying 4 bc iCre-mR9T vectors in LSL-tdT mice (M, 3 months old) induces microglia tdT expression in the injection sites but not contralateral sites. (B-D) Representative images in other brain regions showing transgenes mainly in microglia but also some neuron-like cells. On day 7 after injection, brain sections were immunostained with anti-GFAP and anti-RFP(A) or anti-Cre (B-D) antibodies, and confocal images were taken. 3V, third ventricles. Scale bar=100 μ m

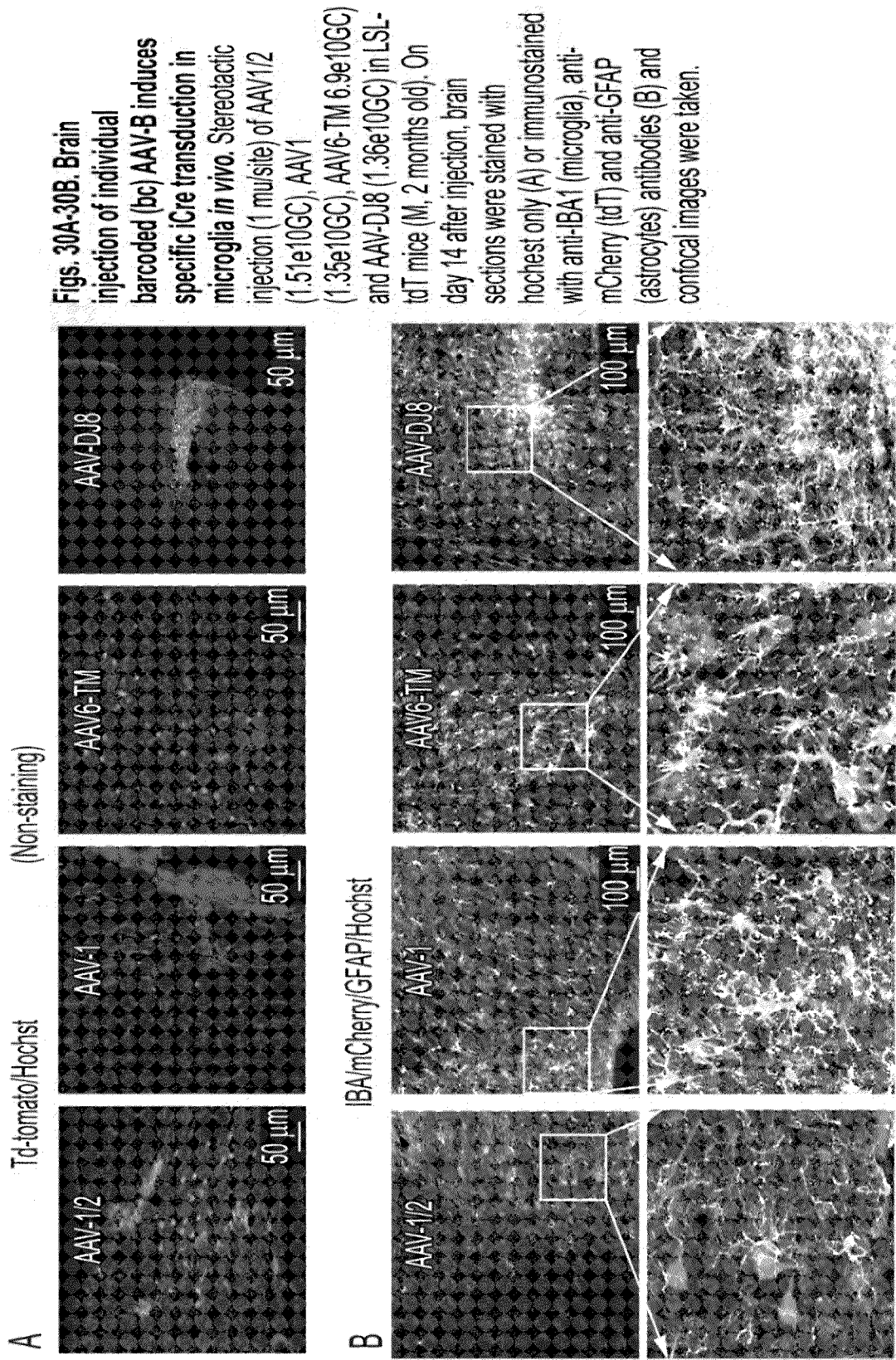


FIGS. 28A-28D

Figs. 29A-29D. Fluorescent immunohistochemistry verifies the specificity of tdT positive cells. Brain sections as described in Fig. 28 were immunostained with rat anti-RFP antibody and Alexa-680 donkey anti-Raf secondary antibody (A, B) or chicken anti-mCherry antibody and Alexa-594 donkey anti-chicken secondary antibody (C, D). Confocal images were taken. Scale bar = 100 μ m.



FIGS. 29A-29D



FIGS. 30A-30B

Fig. 31. Further illustration of microglia transduction by AAV-DJ8 as described i
Fig. 30. Ten of confocal Z-stack images were maximized by projection and overlay of individual channel shows clear colocalization (white arrows) of tdT expression with microglia marker IBA1 but no astrocyte marker GFAP.

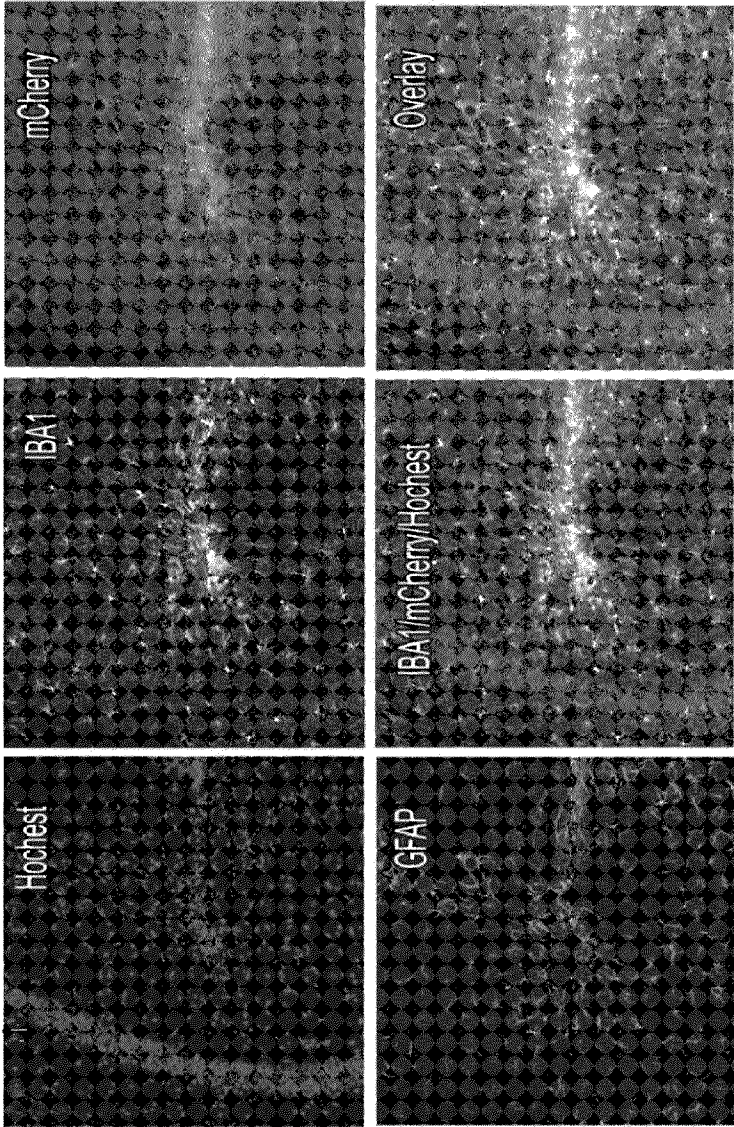


FIG. 31

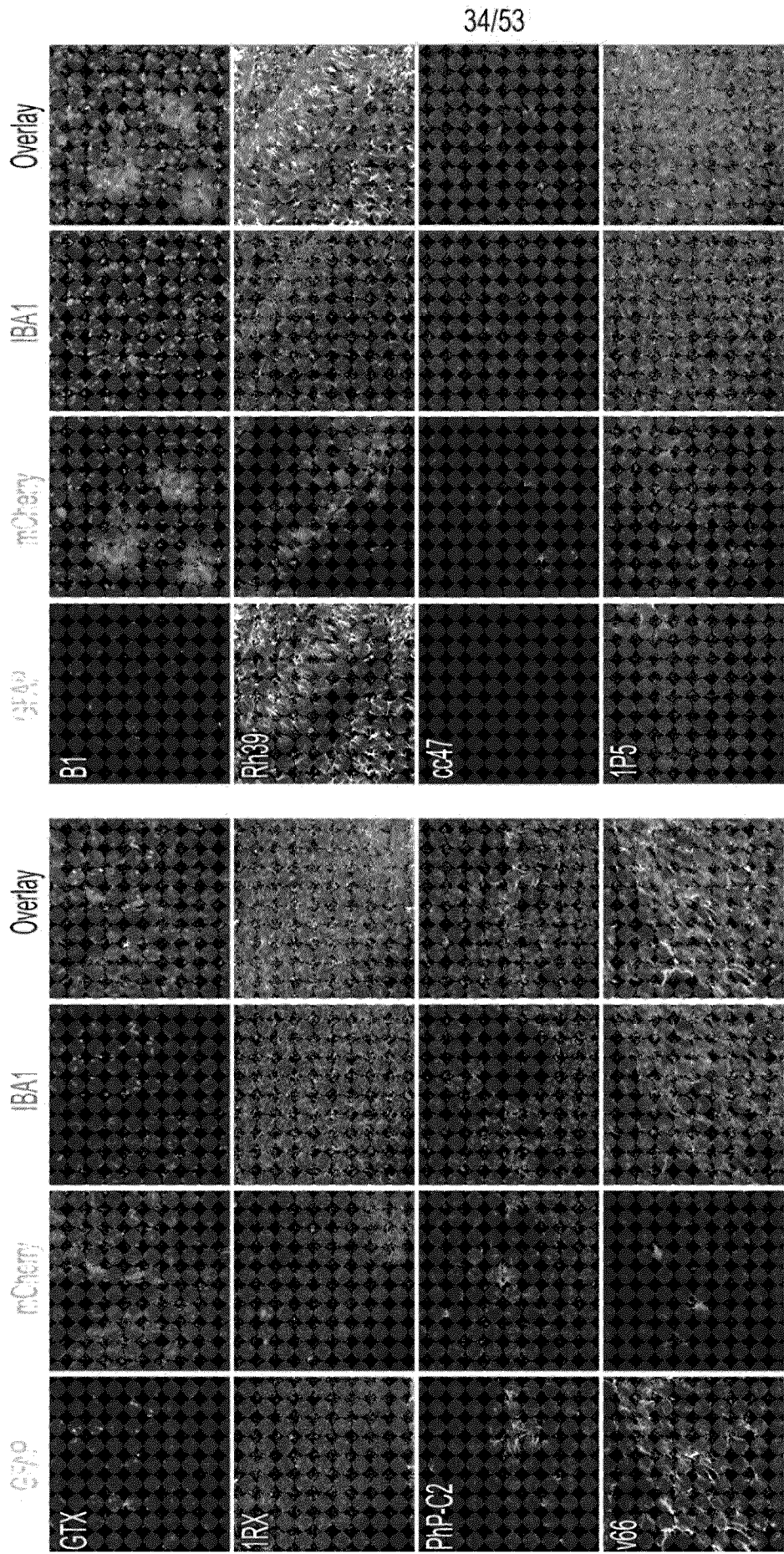
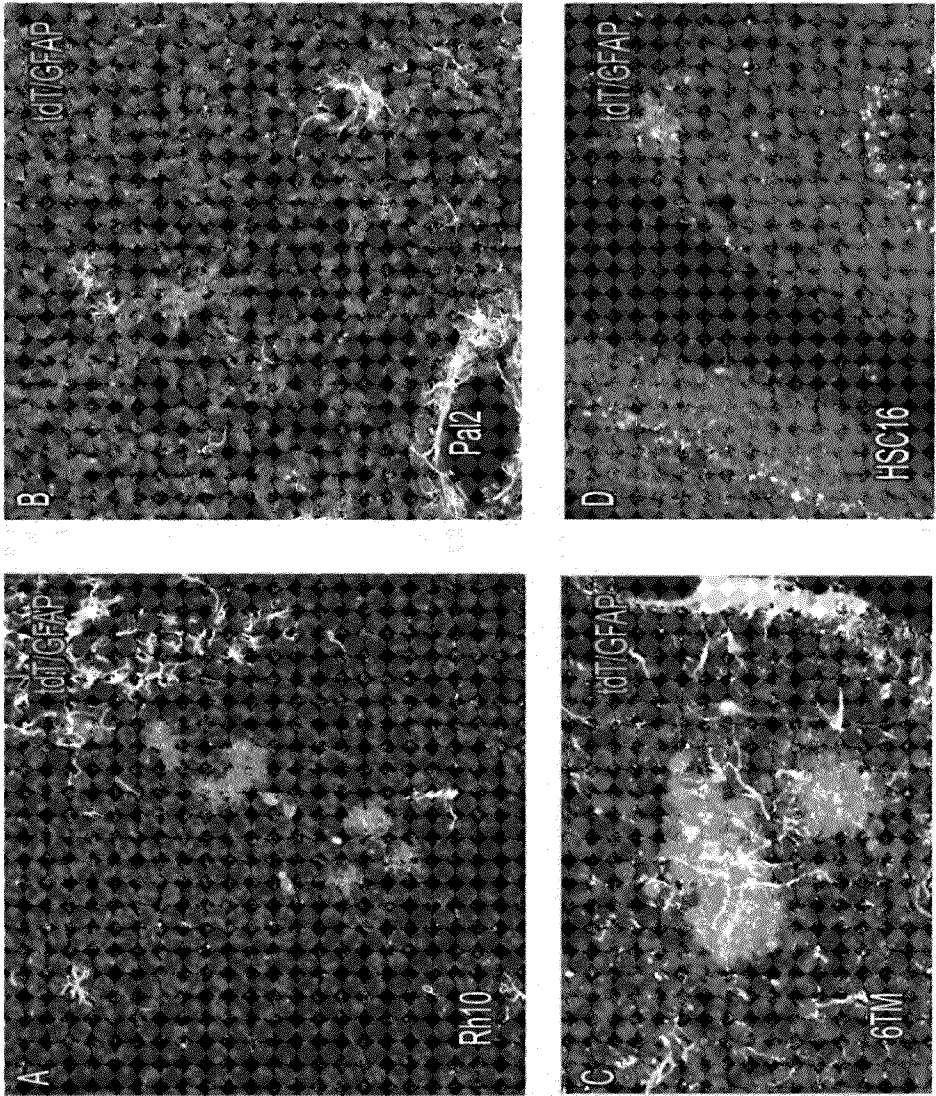


Fig. 32. Systemic (i.v., 50 μ l) injection of individual barcoded (bc) AAV-B induces specific iCre transduction in microglia in vivo. Jugular vein injection of AAV-GTX (TCP101, F, 9 months old, 25.7g, 1.71e12 GC), AAV-1RX (TCP100, F, 9-month old, 34.8g, 3.55e12GC), AAV-PHP-C2 (TCP102, F, 9 month old, 29g, 3.98e12GC), AAVv66 (TCP105, F, 9 months old, 32.5g, 3.45e12GC), AAV-B1 (TCP109, F, 9 months old, 25.4g, 2.145e12GC), AAV-Rh39 (TCP115, F, 8 month old, 20.5g, 3.65e12GC), AAV-cc47 (TCP122, M, 4 months old, 32.2g, 4.44e12GC) and AAV1_P5 (TCP127, F, 4 months old, 23.2g, 1.95e12GC) in LSL-tdT mice. On day 10 after injection, brain sections were immunostained with anti-GFAP (astrocytes), anti-mCherry(tdT) and anti-IBA1 (microglia) antibodies and confocal images were taken. Ten of confocal Z-stack images were maximized by projection and overlay of individual channel shows clear colocalization of tdT expression with microglia marker IBA1 but no astrocyte marker GFAP.

FIG. 32

Figs. 33A-33D. Systemic (i.v., 150 μ l) injection of individual barcoded (bc) AAV-B induces specific iCre transduction in microglia *in vivo*. Jugular vein injection of AAVRh10 (TCP107, F, 7 months old, 23.2g, 6.435e12 GC), AAV-Pal2 (TCP114, F, 6 months old, 23g, 1.251e13 GC), AAV6-TM (TCP104, F, 7-month old, 26g, 1.065e13GC) and AAV-HSC16 (TCP108, F, 7-month old, 24.2g, 1.485e13GC) in LSL-tdT mice. On day 14 after injection, brain sections were immunostained with anti-RFP (tdT) and anti-GFAP (astrocytes) antibodies and confocal images were taken.



FIGS. 33A-33D

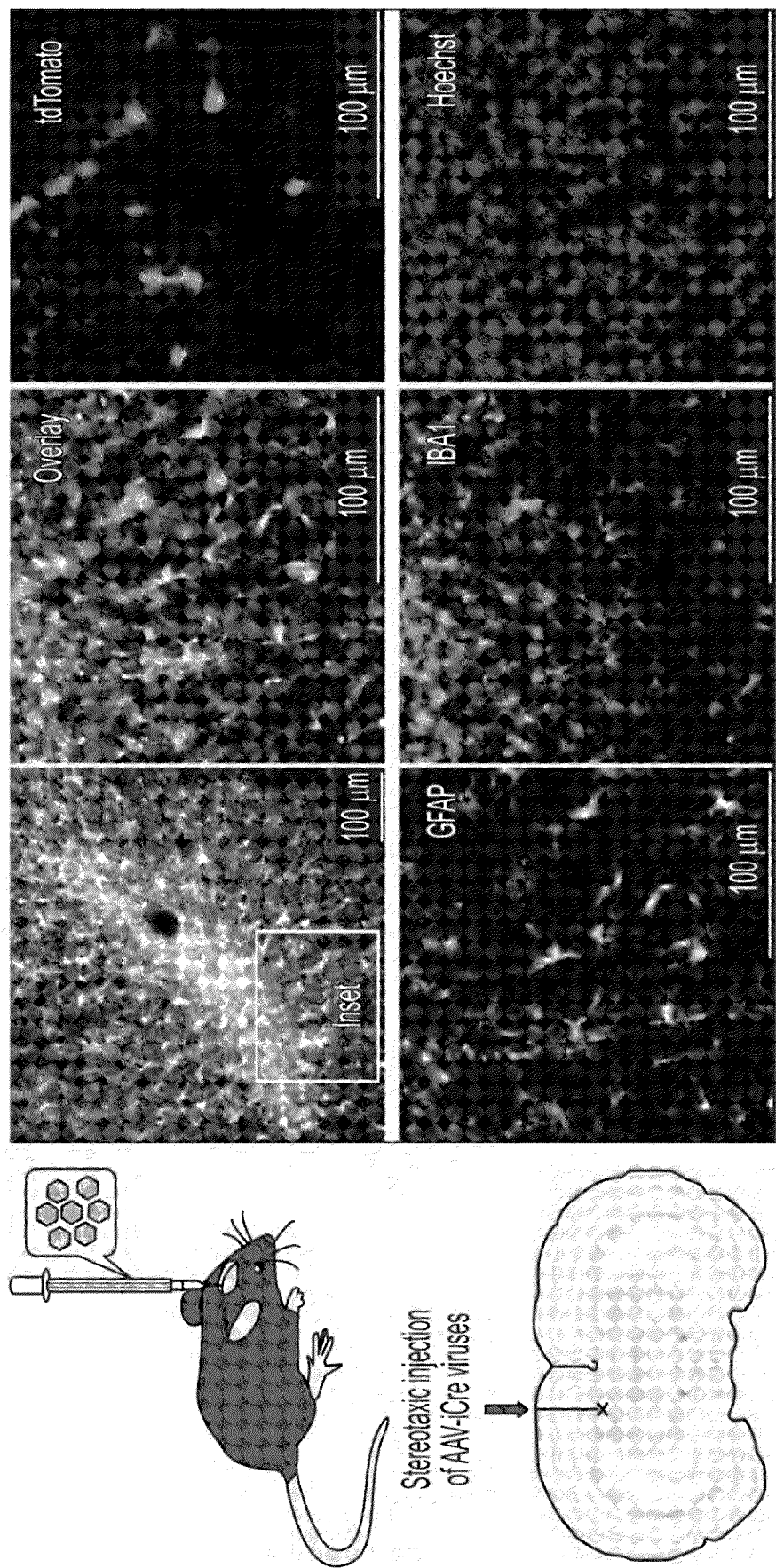


Fig. 34. AAV-CPP21-iCre brain injection induces expression of tdT mainly in neurons but rarely in astrocytes and microglia. Mouse brain stereotaxic injection of AAV-CPP21-hCD68-iCre-3xHA-WPRE3 viruses (6.4e+9 GC, 2 µl/site). On day 7, brain sections were immunostained with anti-GFAP (astrocytes) and anti-IBA1 (microglia) antibodies, and confocal images were taken.

FIG. 34

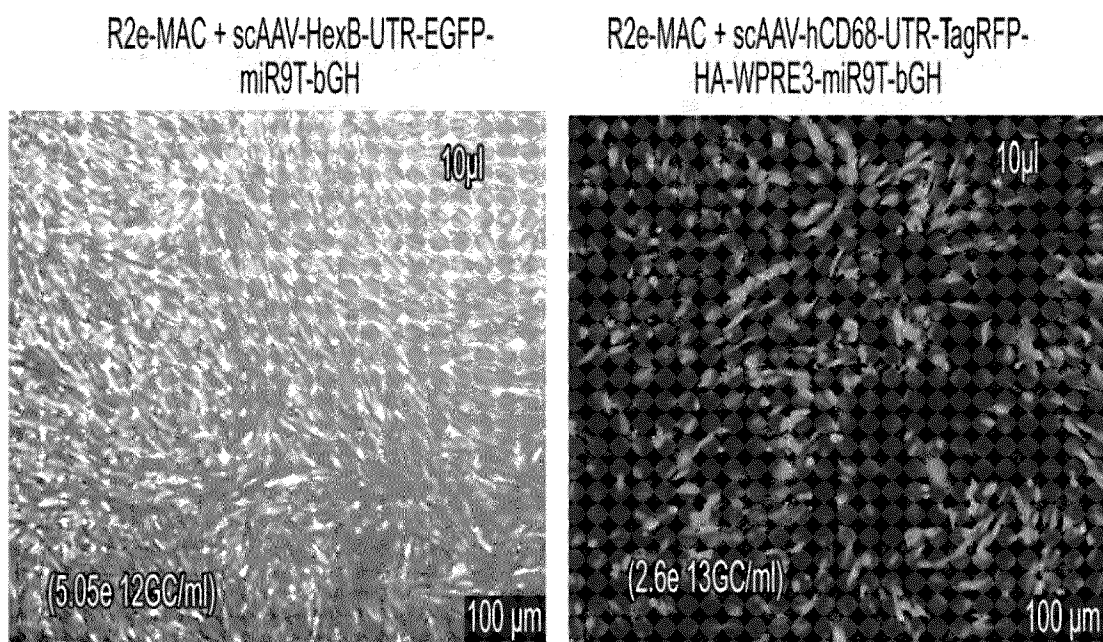
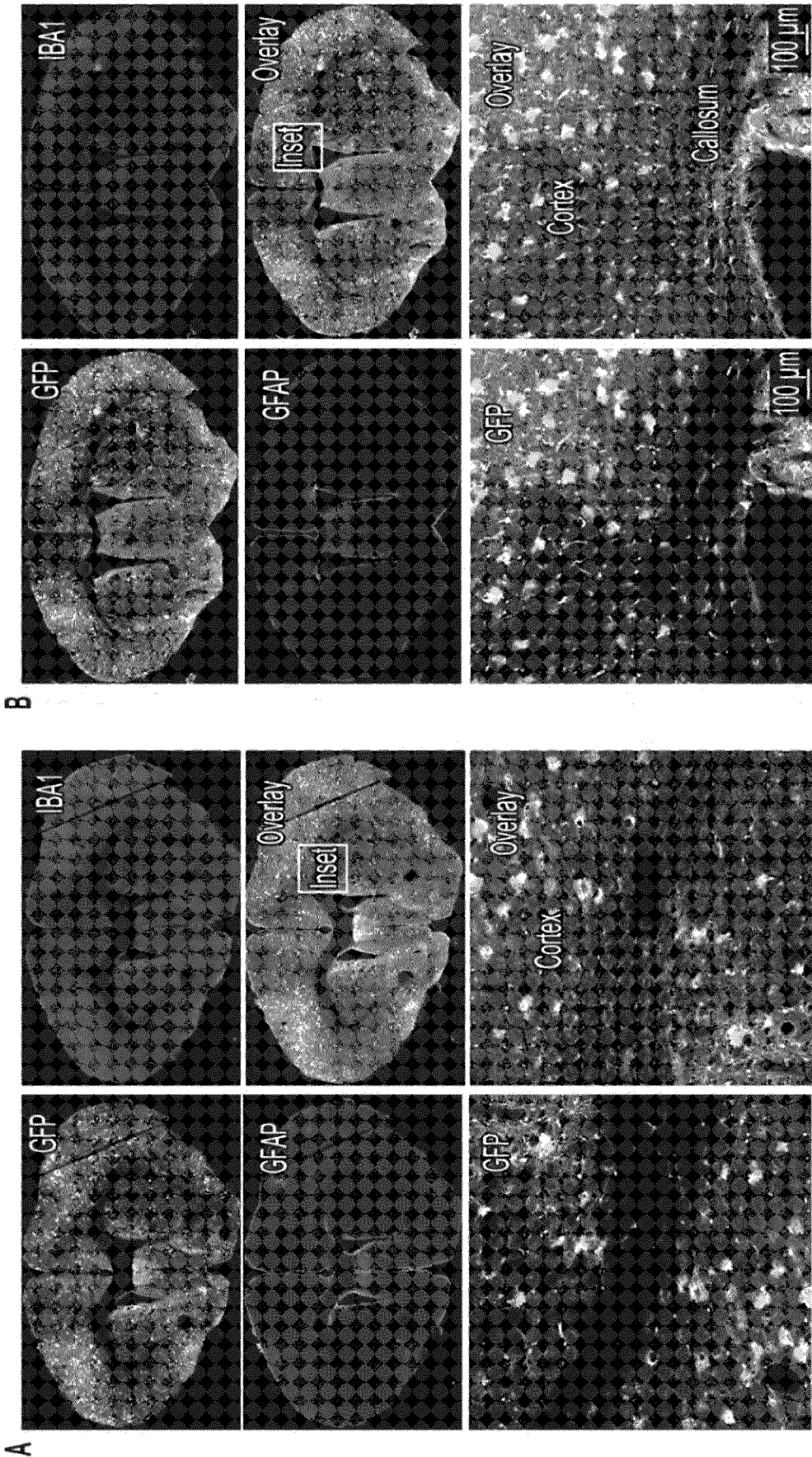


Fig. 35. Large scale of packaging and purification of R2e-MAC with different transfer vector retain high transduction in microglia. Two CDMO companies performed large-scale purification and obtained similar yield although lower than the standard AAV2 system (2-5x). SIM-A9 microglia (4e4cells/well) were treated with 10 μ l of purified R2e-MAC viruses and fluorescent images were taken at day 3 after infection. Scale bar = 100 μ m

FIG. 35



Figs. 36A-36B. Systemic (i.v., 150 μl) injection of AAVone-R2-MAC-4YFTV-scAAV-CMV-EGFP (3.66e13GC/ml) in C57BL6 mice (22g, male) induced extensive transduction mainly in microglia-like cells and rarely in astrocytes and neurons. On day 7 after injection, brain sections were immunostained with anti-GFP, anti-IBA (microglia) and anti-GFAP (astrocytes) antibodies and representative confocal tiling images at 10x objective are shown. Each inset is one of the tiled individual images.

FIGS. 36A-36B

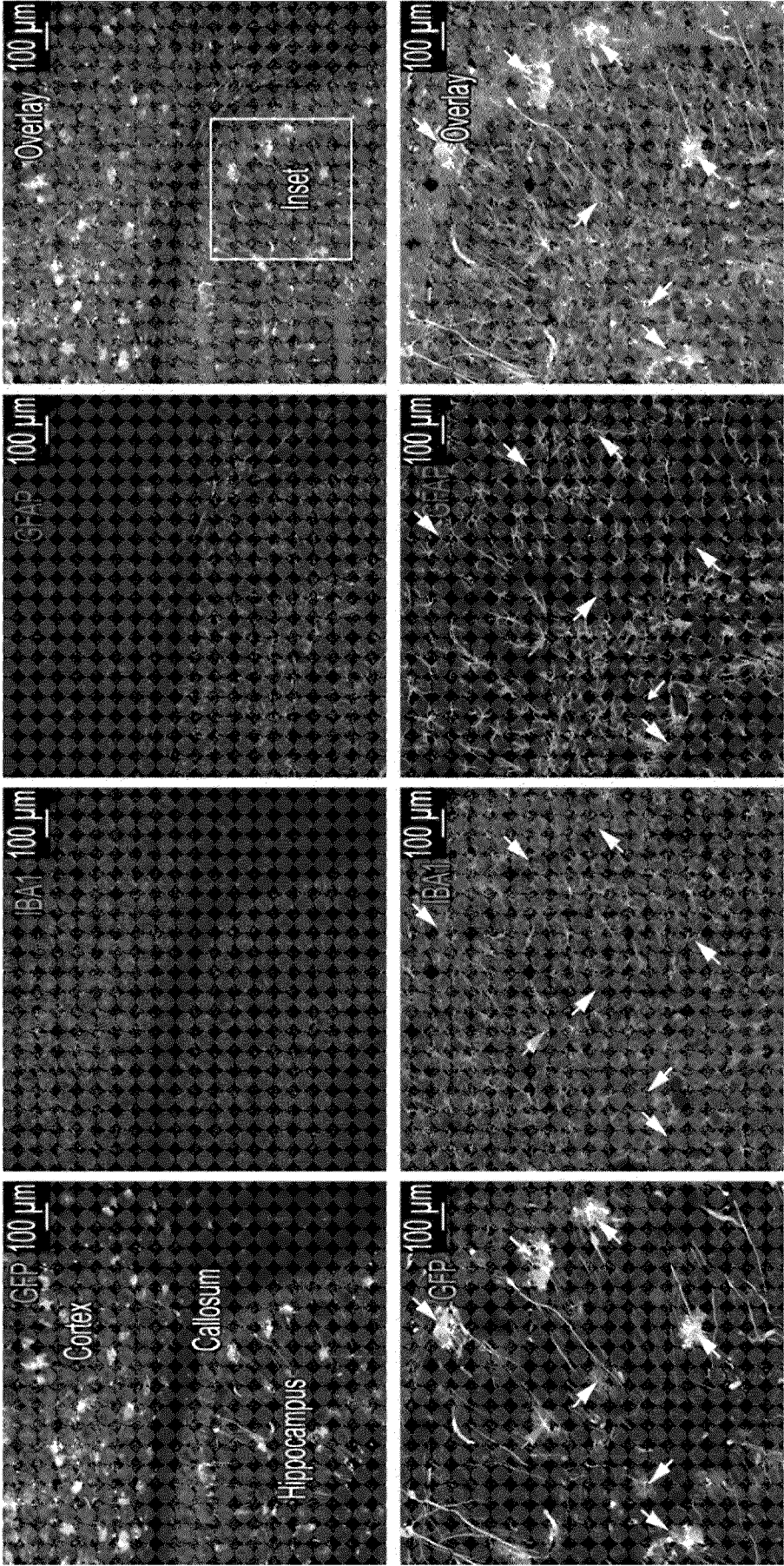


Fig. 37. Representative confocal Z-stack images show extensive transduction in microglia-like cells as well as some neurons and astrocytes after i.v. injection as described in Fig. 36. The inset of hippocampus shows colocalization of microglia-like cells with IBA1+ cells in most cases (white arrows). Red arrow shows astrocyte colocalization. Maximized images from 12 Z-stacks at interval thickness of 1 μm.

FIG. 37

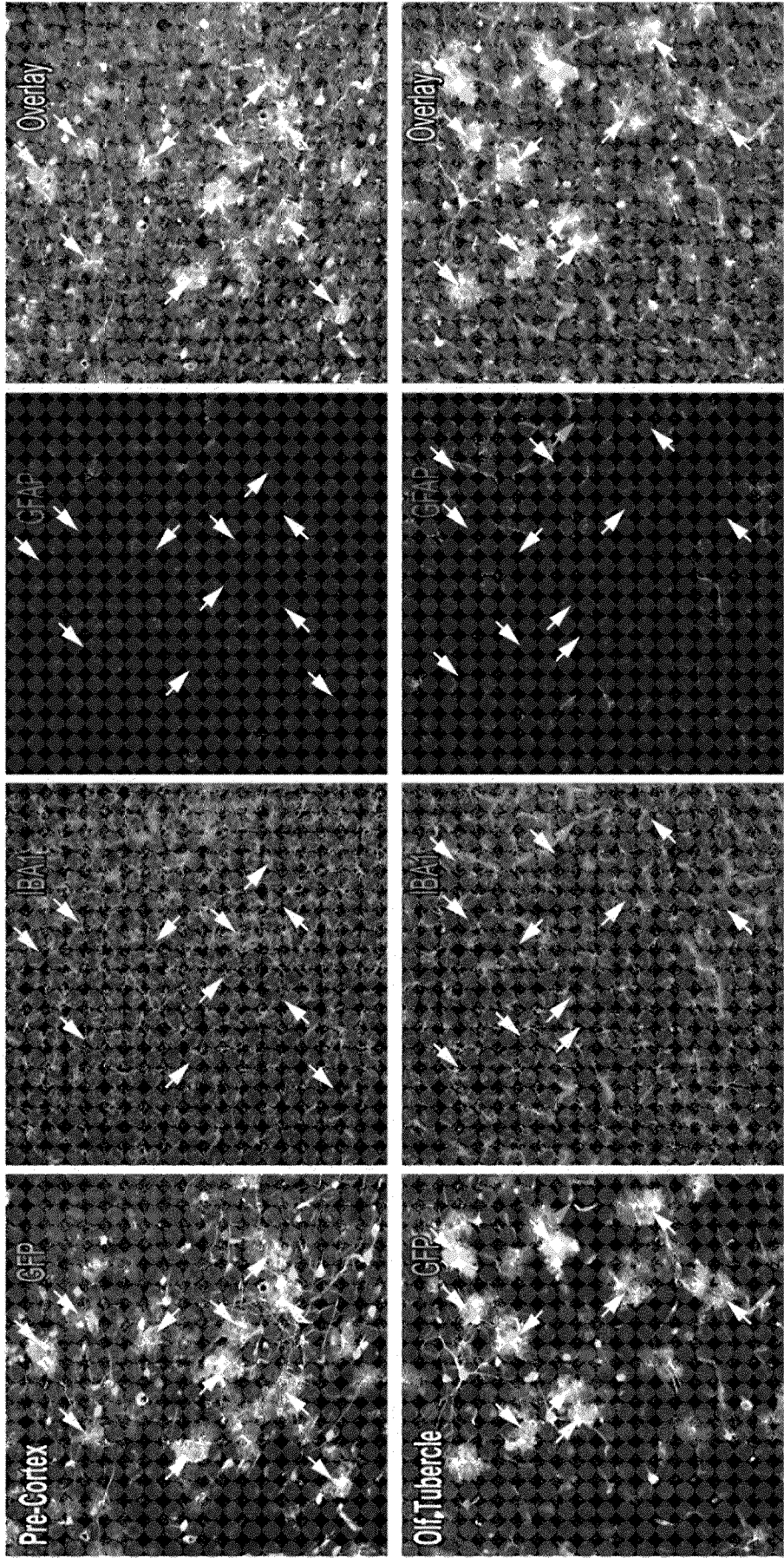


Fig. 38. Representative confocal Z-stack images show extensive transduction in microglia-like cells as well as some neurons and astrocytes in pre-cortex and olfactory tubercle after i.v. injection as described in Fig. 36. White arrows show the colocalization of GFP-labeled microglia-like cells with IBA1+ cells in most cases. Red arrow shows astrocyte colocalization. Maximized images from 12-14 Z-stacks at interval thickness of 1 μm.

FIG. 38

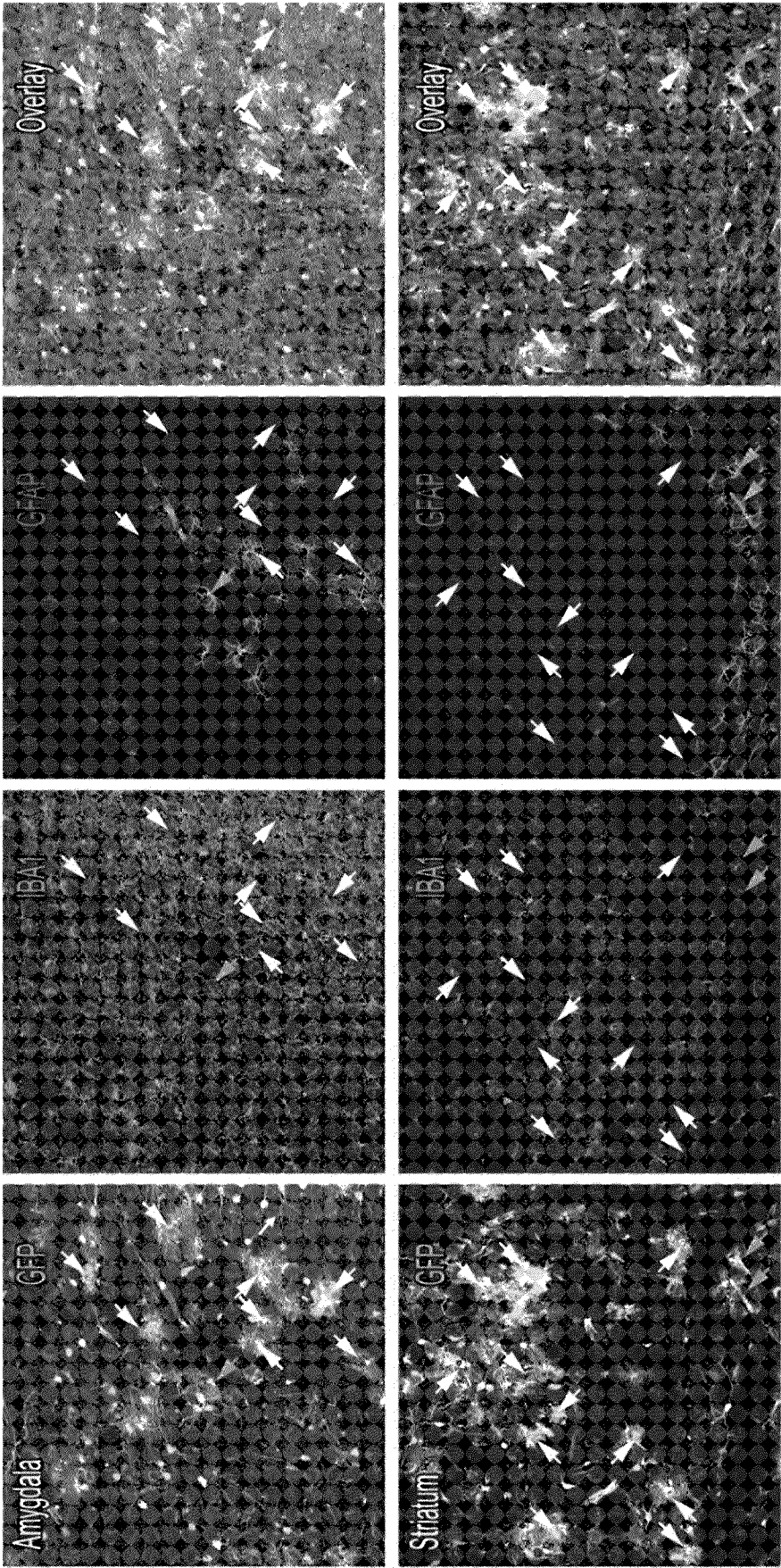


Fig. 39. Representative confocal Z-stack images show extensive transduction in microglia-like cells as well as some neurons and astrocytes in amygdala and lateral striatum after i.v. injection as described in Fig. 36. White arrows show the colocalization of GFP-labeled microglia-like cells with IBA1+ cells in most cases. Red arrow shows astrocyte colocalization. Maximized images from 12-14 Z-stacks at interval thickness of 1 µm.

FIG. 39

AAV2-REF-link-VP1(SAO-MAC)

FIG. 40

43/53

CTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGC
 AGAACAAACACTCCAAGTGGAACCACCACCGAGTCAAGGCTTCAGTTTCTCAGGCCGGAGCGAGTGACA
 TTCGGGACCAGTCTAGGAACTGGCTTCCTGGACCCCTGTTACCGCCAGCAGCCAGTATCAAAGACATCTGC
 GGATAACAACAACAGTGAATACTCGTGGACTGGAGCTACCAACTACCACCTCAATGGCAGAGACTCTCTG
 GTGAATCCGGGCCCCGGCCATGGCAAGCCACAAGGACGATGAAGAAAAGTTTTTCTCAGAGCGGGGTTT
 TCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGGACATTGAAAAGGTGATGATTACAGACGAAGA
 GGAAATCAGGACAACCAATCCCGTGGCTACGGAGCAGTATGGTTCTGTATCTACCAACCTCCAGagtgcc
 caaCTCAACACCACGAAGCCGATAAGACAAGCAGCTACCGCAGATGTCAACACACAAGCGTTCTTCAG
 GCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGGCCATCTGGGCAAGATTCCACACACGGACGG
 ACATTTTCACCCCTCTCCCTCATGGGTGGATTTCGGACTTAAACACCCTCCTCCACAGATTCTCATCAAG
 AACACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGGCGCAAAGTTTGCTTCCTTCATCACACAGT
 ACTCCACGGGACAGGTGAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAACGCTGGAATCC
 CGAATTCAGTACACTTCCAACATAACAAGTCTGTTAATGTGGACTTTACTGTGGACACTAATGGCGTG
 TATTCACAGCCTCGCCCATTTGGCACCAGATACCTGACTCGTAATCTGTAA

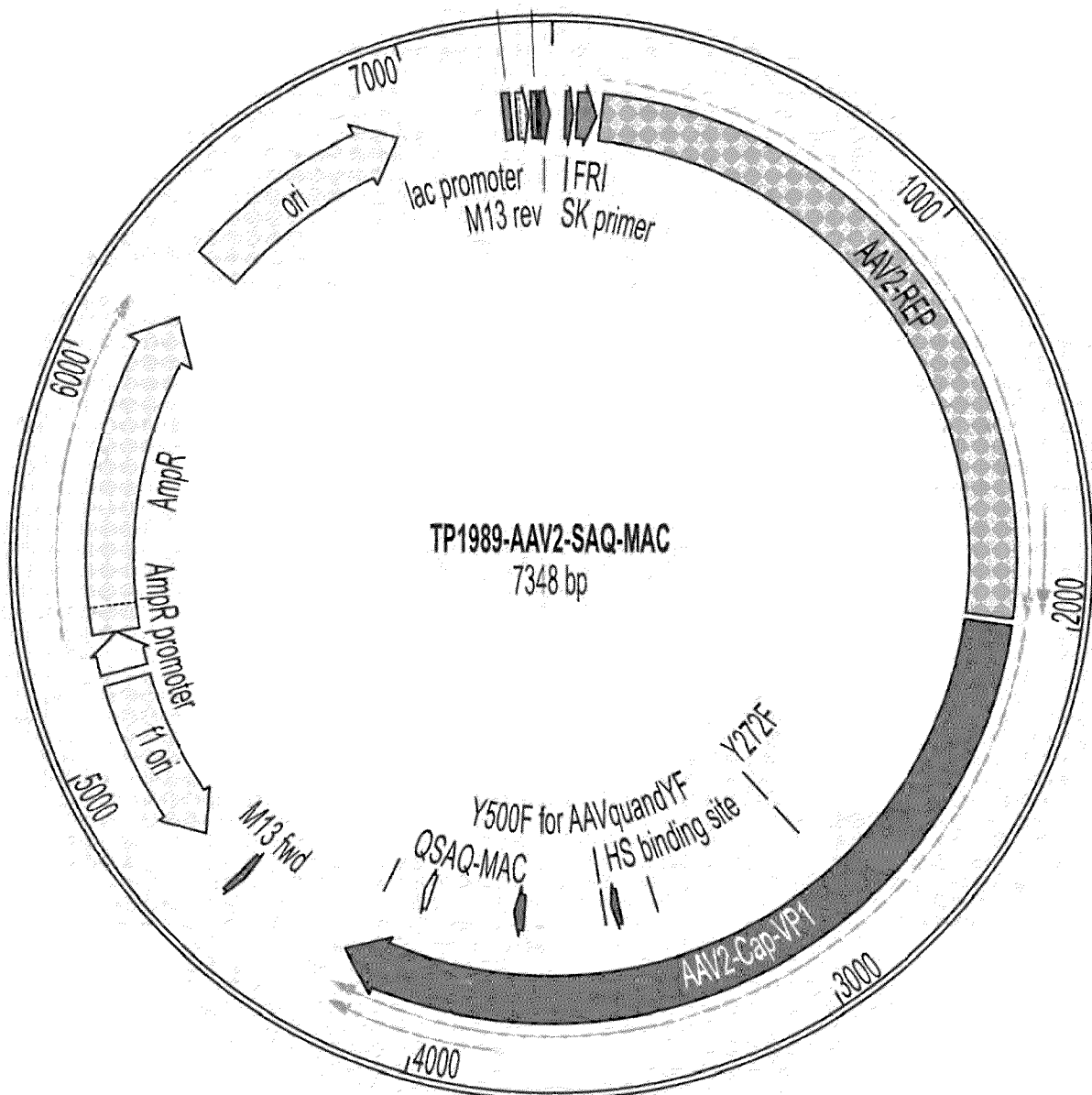


FIG. 40
CONTINUED

R2e-MAC, REP protein

MPGFYEIVIKVPSDDLDEHLPGISDSFVNWVAEKEWELPPDSMDLNLIEQAPLTVAEKLQRFLEWRRV
SKAPEALFFVQFEKGEYFHMHLVETTGVKSMVLGRFLSQIREKLIQRIYRGIEPTLPNWFAVTKTRNG
AGGKNKVVDECYIPNYLLPKTQPELOWAWTNMEQYLSACLNLTERKRLVAQHLTHVSQTQEQNKENQNP
SDAPVIRSKTSARYMELVGLVDKGITSEKQWIOEDQASYISFNAASNSRSQIKAALDNAGKIMSLTKTA
PDYLVGCQPVEDISSNRIYKILELNGYDPOYAASVFLGWATKKFGKRNTIWLFGPATTGKTNI AEAIHT
VPFYGCVNWTNENFPFNDKVIWEEGKMTAKVVESAKAILGGSKVRVDQCKSSAQIDPTPVIVTS
NTNMCAVIDGNSTTFEHQQPLQDRMFKEELTRRLDHDFGKVTQEVKDFFRWAKDHVVEVEHEFYVKKGG
AKKRPAFSDADISEPKRVRESVAQPSTSDAEASINYADRYQNKCSRHVGMNLMFLPCROCERMNQNSNIC
FTHGQKCLECFPVSESQPVSVVKKAYQKLCYIHHIMGKVPDACTACDLNVNVDLDDCIFEQ*

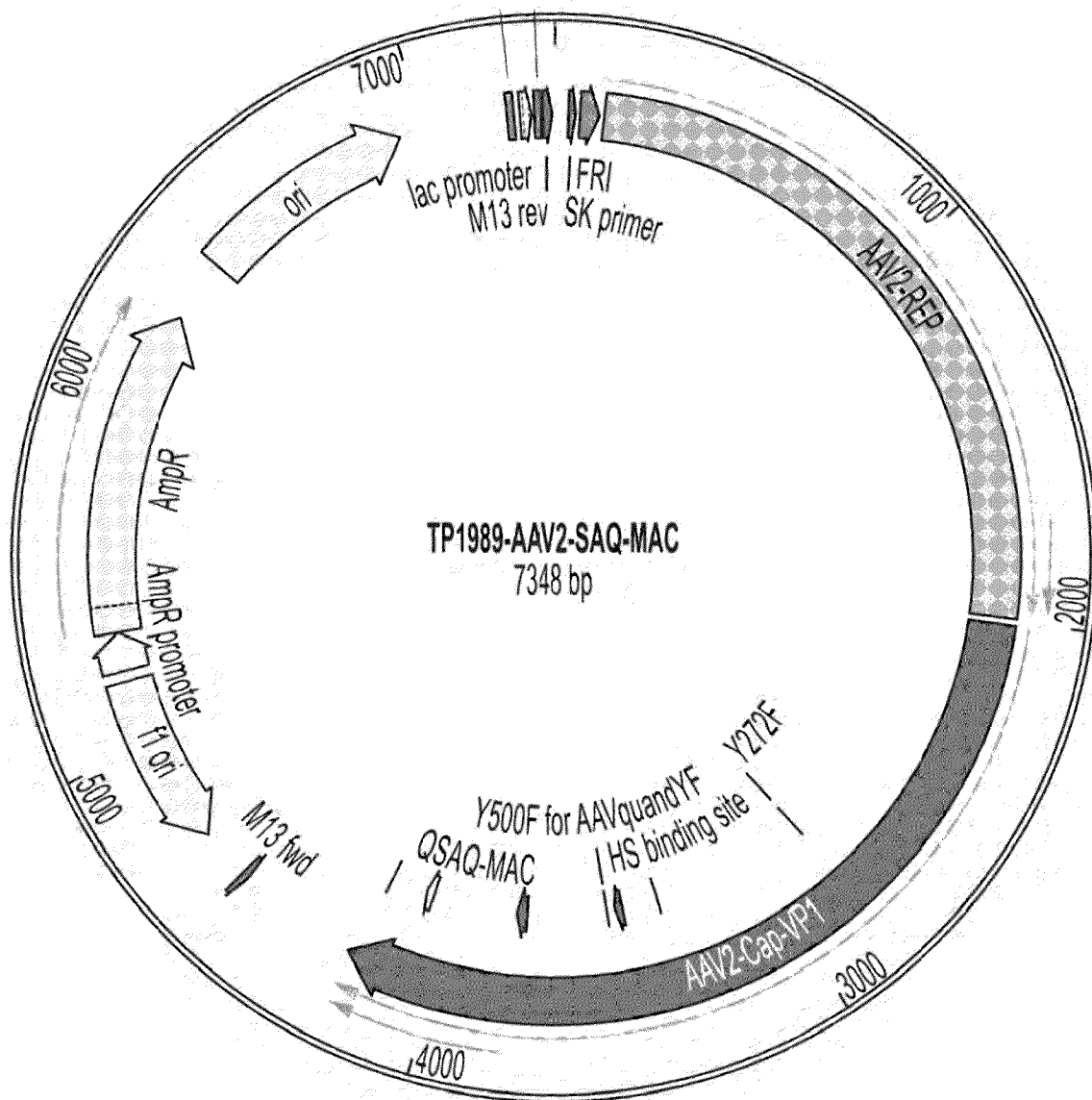
R2-MAC, VP1 (SAQ-MAC) protein

MAADGYLPDWLEDTLSEGIQWWKLKPGPPPKPAERHKDDSRGLVLPGYKYLGPENGLDKGEPVNEADA
AALEHDKAYDRQLDSCDNPYLKYNHADAEEQERLKEDTSFGGNLGRAVFQAKKRVLEPLGLVEEPVKTA
GKKRPVEHSPVEPDSSSGTGKAGQOPARKRLNFGQTDADSVDPQPLGQPPAAPSGLGTNTMATGSGAP
MADNNECADGVGNSSGNWHCDSTWMGDRVITTTSTRTWALPTYNNHLYKQISSQSGASNDNHFGYSTPWG
YFDENRFHCHFSPRDWQRLINNNWGRPKRLNFKLENIQVKEVTQNDGTTTIANNLTSTVQVFTDSEYQL
PYVLGSAHQGCLPPFPADVFMVPQYGYLTLLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFTFSYTFEDVPF
HSSYAHSQSLDRLMNPLIDQYLYLSRTNTPSGTTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQORVSK
TSADNNNSEYSWTGATKYHLNGRDSLVPNGPAMASHKDDEEKFFPQSGVLIFGKQSEKTNVDIEKVMIT
DEEEIRITNPVATEQYGSVSTNLQSAQINNTKPIRQAATADVNTQGVLPGMVWQDRDVYLOGPIWAKIPH
TDGFHFPSPLMGGFGLKHPPPQILIKNTPVPANPSTTFSAAKFASFITQYSTQVSVIEIWELOKENS
KRWNPEIQYTSNYSVNVDFTVDTNGVYSEPRPIGTRYLTRNL*

FIG. 41

FIG. 42

CTATGCTCAGACACCAAGATTTGTAAGCTGCTCATGAATCCCTCCATACGACCAAGTACCTGTTATTTCTTGTGAGC
AGAGCAATACAGCTCCAAATGTGAACCAACCAAGTAGTCATGGGATTCAGATTTCTCCAGGCGGAGAGCAATGACA
TTTTCGACCAAGCTATCGAATACCTTCCTCCCAACCTTTTTCACCTCAGCACTATGCAAGCTATTTCG
CGATACCAACCAACAGTAAATCTTCGGGACCTGGAGCTACCAAGTACCACTCCAAATGGCAGCACTCTCTG
CTGATTCCTGGCTCCGCGCAATCTGTAACCTCAAGCAACCTTCACAAATATTTTTCCTACCAACCGCGCTC
TCATCTTTGGGAAGCAAGGTCACAGAAAAAATAATGTGGACATTTAAAGAGTCTATGTTACAGACGAGGA
CGATATACAGCAATACCAATCTCTGCTACAGCAATATCTTTCTGTTTATACCAACCTCCAGCTCTGCG
CAACTTACACCAACCAAGCCGATACAGACCAAGCTACCGGAGATGTCAACAAACAGAGCGCTTCTTCCAG
CCATCTTCCTGGAGCAACAGATCTGTATCCCTTCAGGCGCACTCTGCGCAAAAGATTTCACCAACCGGAGG
ACATTTTCACCCCTCTTCCCTCTATGGGTGGATTCGGCTTAAACACCCCTCTTCACAGATTTCTATCAG
AAATCCCTCGCTACCTGCAATCTTCAACCACTTCAGTCCCAAAATTCCTCTCTTATTCACCAACCT
ACTCCAGGGCAAGGTTGAGCTGGAGATCGAGTGGAGCTGGAGAGGAAACAGCAATCTGGAACTC
CCATATTCAGTACACTTCCACCTACCAATCTCTGTTTAACTTCAGCTTCACTCTCCACCAATATTCCTT
TATTCAGAGCTTCGCCCAATGGCACCAGATTCCTGACTCGTAATCTGTA

FIG. 42
CONTINUED

R2e-MAC, REP protein

MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEWELPPDSMDLNLIEQAPLTVAEKLQDRDFTLEWRRV
 SKAPEALFFVQFEKGESYFHMVHVETTGVKSMVLGRFLSQIREKLIQRIYRGIEPTLPNWEAVTKTRNG
 AGGGNKVVDECIYPNYLLPKTQPELOWAWTNMEQYLSACLNLERKRLVAQHLTHVSQTQEQNKENQNP
 SDAPVIRSKTSARYMELVGWLVDKGITSEKQWIOEDQASYISFNAASNSRSQIKAALDNAGKIMSLTKTA
 PDYLVGQQPVEDISSNRIYKILELNGYDPOYAASVFLGWATKKFGKRNTIWLFGPATTGKTNIAEAIHT
 VPFYGCNVNWTNENFPFNDKVDKMWIWEWGKMTAKVVESAKAILGGSKVRVDQCKSSAQIDPTPVIVTS
 NTNMCVIDGNSTTFEHQQPLQDRMFKFELTRRLDHDGKVTQEVKDFRWAKDHVVEVEHEFYVKKGG
 AKKRPA PSDADISEPKRVRESVAQPSTSDAEASINYADRYQNKCSRHVGMNLMFLPCRQCEFMNQNSNIC
 FTHGQKDCLECFPVSESQPVSVVKKAYQKLCYIHHIMGKVPDACTACDLNVNVDLDDCIFEQ*

R2e-MAC, VP1 (4YF / TV and SAQ-MAC) protein

MAADGYLPDWLEDTLSEGIQWWKLKPGPPPKPAERHKDDSRGLVLPGYKYLGPFGNLDKGEVNEADA
 AALEHDKAYDRQLDSDGNPYLKYNHADADEFQERLKEDTSFGGNLGRAVFQAKKRVLEPLGLVEEPVKTA
 GKKRPVEHSPVEPDSSSGTGKAGQQPARKRLNFGQTDADSVDPQPLGQPPAAPSGLGTNTMATGSGAP
 MADNNEGADGVCNSSGNWHCDSTWMGDRVITSTRTWALPTYNNHLYKQISSQSGASNDNHFFGYSTPWG
 YFDENRFHCHFSRPDWQRLINNNWGFRPKRLNFKLENIQVKEVTQNDGTTTIANNLTSTVQVFTDSEYQL
 PYVLGSAHQGCLPPFPADVFMVPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNTFTSYTFEDVFP
 HSSYAHSQSLDRLMNPLIDQYLYELSRNTNTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQORVSK
 SADNNNSESWTGATKYHLNGRDSLVPNGPAMASHKDDEEKFFPQSGVLI FGKQGSEKTNVDIEKVMIT
 DEEEIIRTNPVATEQYGSVSTNLQSAQLNTTKPIRQAATADVNTQGVLPGMVWQDRDVYLCQPIWAKIPH
 TDGHFHPSPLMGGFGLKHPPPQILIKNTVPANPSTTFSAAKFAFITQYSTGQVSVEIEWELQKENSKR
 WNPEIQYTSNYSVNVDFTVDTNGVYSEPRPIGTRELTRNL*

FIG. 43

TP2018: ITR-pdsAAV-hHexB-134-UTR-EGFP-4xmiR9T-BGH-ITR

ctgcgcgcctcgctcgctcaactgaggccgcccgggcaaaagccgggctcgggcgaccttgggtcgccgg
cctcagtgagcgagcgagcgcgagagagggagtggaattcacgcgtggatctgaattcaattcacgcgt
ggtaccGTGGGGGCGGGCGCGCAGTCATCTGACTCGGTGACTCACCCGCGGCCGCGCTTCCTCTGAT
CCGGGCGGGCGGGAAGTCGGGTCCCGAGGCTCCGGCTCGGCAGACCGGGCGGAAAGCAGCCGAGCGGCC
aATTAAAGGTTTATACCTTCCAGGTAACAAACCAACCAACTTTTCGATCTCTTGTAGATCTGTTCCTAA
ACGAACTTTAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACCTCACGCAGTATAATTAATAA
GTAATTACTGTGTTGACAGGACACGAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGT
GTTGCAGCCGATCATCAGCACATCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGaccggtcgccacca
tggtgagcaagggcgaggagctgttcacgggggtggtgcccacatcctggtcgagctggaagggcgacgtaa
cgccacaaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttc
atctgcaccaccggcaagctgcccgtgccctggcccacctcgtgaccacctgacctacggcgtgagt
gettacagcgtaccccgaccacatgaagcagcagacttcttcaagtcggccatgccogaaggctacgt
ccaggagcgaccatcttcttcaaggacgacggcaactacaagaaccggcgccgaggtgaagttcgaggcc
gacacctggtgaaccgcatcgagctgaagggcatcgacttcaaggagggaeggcaacatcctggggcaca
agctggagtacaaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggt
gaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgacctaccagcagaacacc
cccacggcgacgggcccgtgctgctgcccgaacacctacctgagcaccagtcggccctgagcaag
accccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccggcgccgggatactctcggc
ggacgagctgtacaagtaaaagcggccatcaagctTCATACAGCTAGATAACCAAAGCGATTTCATACAGCT
AGATAACCAAAGACCGGTTTCATACAGCTAGATAACCAAAGTCACTCATAACAGCTAGATAACCAAAGgtcg
actagagctcgctgatcagcctcgactgtgccttctagtgtccagccatctgttgtttgcccctccccg
tgcttccttgaccctggaaggtgccactcccactgtcctttcctaataaaaatgaggaaattgcatcgca
ttgtctgagtaggtgtcattctattctggggggtgggggtggggcaggacagcaagggggaggattgggaa
gacaatagcaggcatgctggggagagatTgatctGAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCT
CTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGAACAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCC

FIG. 44

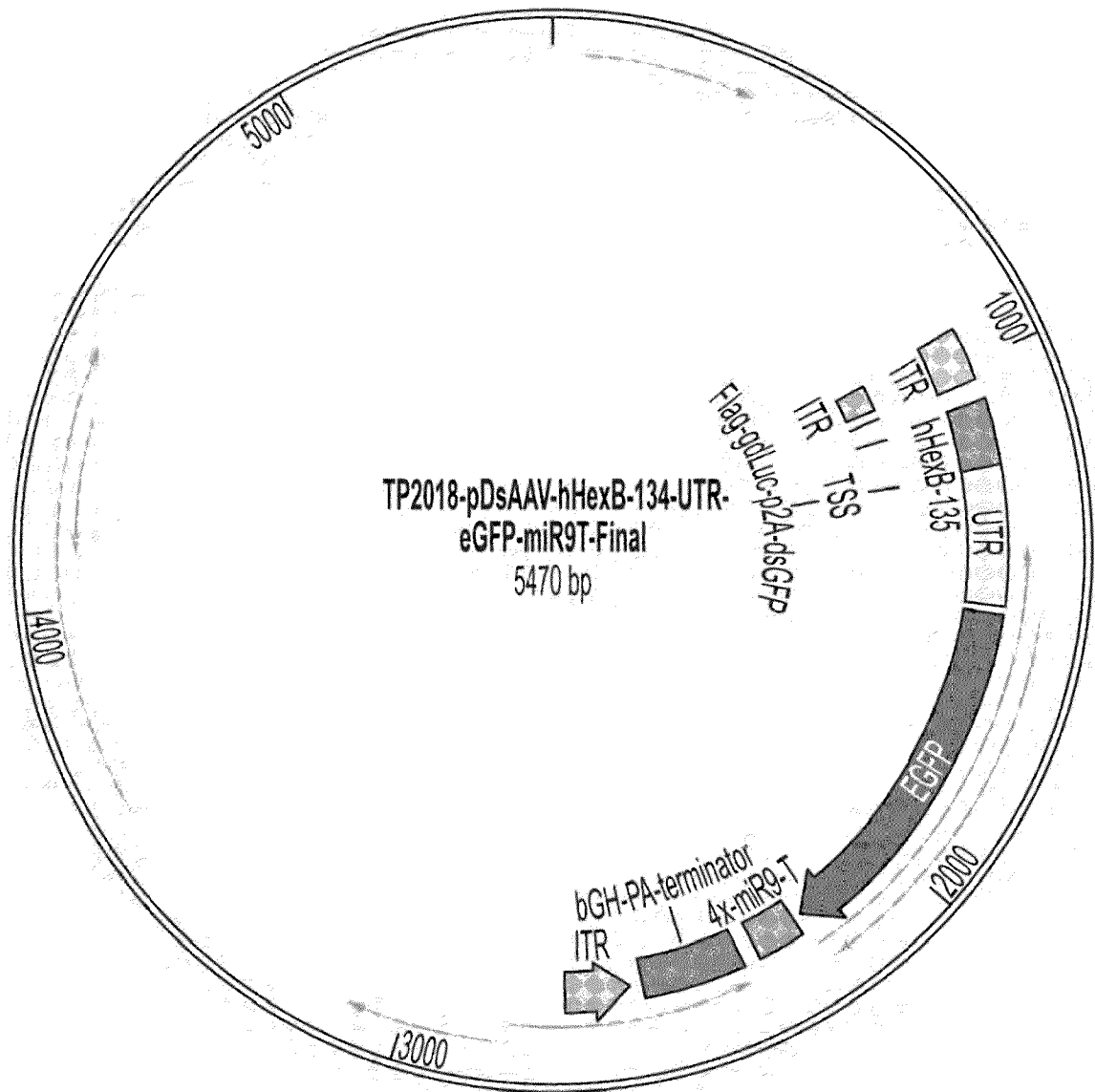


FIG. 44
CONTINUED

AAV barcode reporter vector

ITR-pDsAAV-hHexB134-UTR-NLS-iCre-HA-WPRE3-bc25-miR9T-bGH-ITR

ctgcgcgcctcgctcgctcaactgagggcgcccgggcaaaagcccgggcgctcgggcgaccttttggtcgcccg
cctcagtgagcgagcgagcgcgagagagggagtggaattcacgcgtggatctgaattcaattcacgcgt
ggtaccGTCGGGGCGGGCGCGCAGTCATCTGACTCGGTGACTCACCCGCGGCCGCGCTTCCTCTGAT
CCGGGCGGGCGGGGAAGTCGGGTCCCGAGGCTCCGGCTCGGCAGACCAGGGCGGAAAGCAGCCGAGCGGCC
aATTAAAGGTTTATACCTTCCAGGTAAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAA
ACGAACCTTAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACCTCACGCAGTATAATTAATAA
GTAATTACTGTCTGACAGGACACGAGTAACCTGCTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGT
GTTGCAGCCGATCATCAGCACATCTAGGTTTCGTCCGGGTTGACCGAAAGGTAAGaccgggtcgccacca
tgCCcaagaagaagCggaagtggaagtctgcaactgctgactgtgcacaaaacctgacctgacctcc
tgtggatgcaacctctgatgaagtccaggaagaacctgatggacatgttcagggacaggcagggccttctct
gaacacacctggaagatgctcctgtctgtgtgcagatcctgggctgctggtgcaagctgaacaacagga
aatggttccctgctgaacctgaggatgtgagggactacctcctgtacctgcaagccagaggcctggtgt
gaagaccatccaacagcacctgggcccagctcaacatgctgcacaggagatctggcctgacctgacctct
gactccaatgctgtgtccctggtgatgaggagaatcagaaaggagaatgtggatgctggggagagagcca
agcaggccctgacctttgaacgcactgactttgaaccaagtcagatccctgatggagaactctgacagatg
ccaggacatcaggaacctggccttccctgggcattgacctacaacacctgctgcgcattgcccgaattgccc
agaatcagagtgaaggacatctccgcaccgatggtgggagaatgctgatccacattggcaggaccaaga
ccctggtgtccacagctggtgtggagaaggccctgtccctgggggttaccaagctggtggagagatggat
ctctgtgtctggtgtggctgatgaccccaacaactacctgttctgccgggtcagaaagaatggtgtggct
gacctcttctgccacctcccaactgtccacctgggcccctggaaggatctttgaggccaccaccgctga
tctatggtgccaaggatgactctgggcagagataacctggcctggtctggccactctgccagagtgggtgc
tgccagggaatggccagggtggtgtgtccatccctgaaatcatgcaggctggtggctggaccaatgtg
aacattgtgatgaactacatcagaaacctggactctgagactggggccatggtgaggtgctcgaggatg
gtgactaccataacgatgttccagattaacgtttaagataatcaacctctggattacaaaatttgtgaaag
attgaactggtattcttaactatgttgctccttttaagctatgtggatacgtgctttaatgcctttgtat
catgctattgcttccgctatggctttcattttctcctccttgtataaatcctggtTAGTTCTtgccacgg
cggaactcctcgccgctgcttgcctgctggacaggggctcggtctgtgggactgacaattccgt
ggNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAagtaaagcgccatcaagctTCATACAGCTAGATAACCAAAG
CGATTTCATACAGCTAGATAACCAAAGACGCGTTCATACAGCTAGATAACCAAAGTCACTCATAACAGCTAG
ATAACCAAAGgtcgactagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttgt
ttgccccctcccccgctgcttcccttgacctggaagggtgccactcccactgtcctttcctaataaaaatgag
gaaattgcatcgcatgtctgagttaggtgtcattctattctggggggtgggggtggggcaggacagcaagg
gggaggattgggaagacaatagcaggcatgctggggagagatcgatctaggaacocctagtgatggagtt
ggcactccctctctgcgcgctcgctcgctcaactgagggcgcccgggcaaaagcccgggcgctcgggcgac
tttggtcgcccgccctcagtgagcgagcgagcgagagagggagtgccaa

FIG. 45

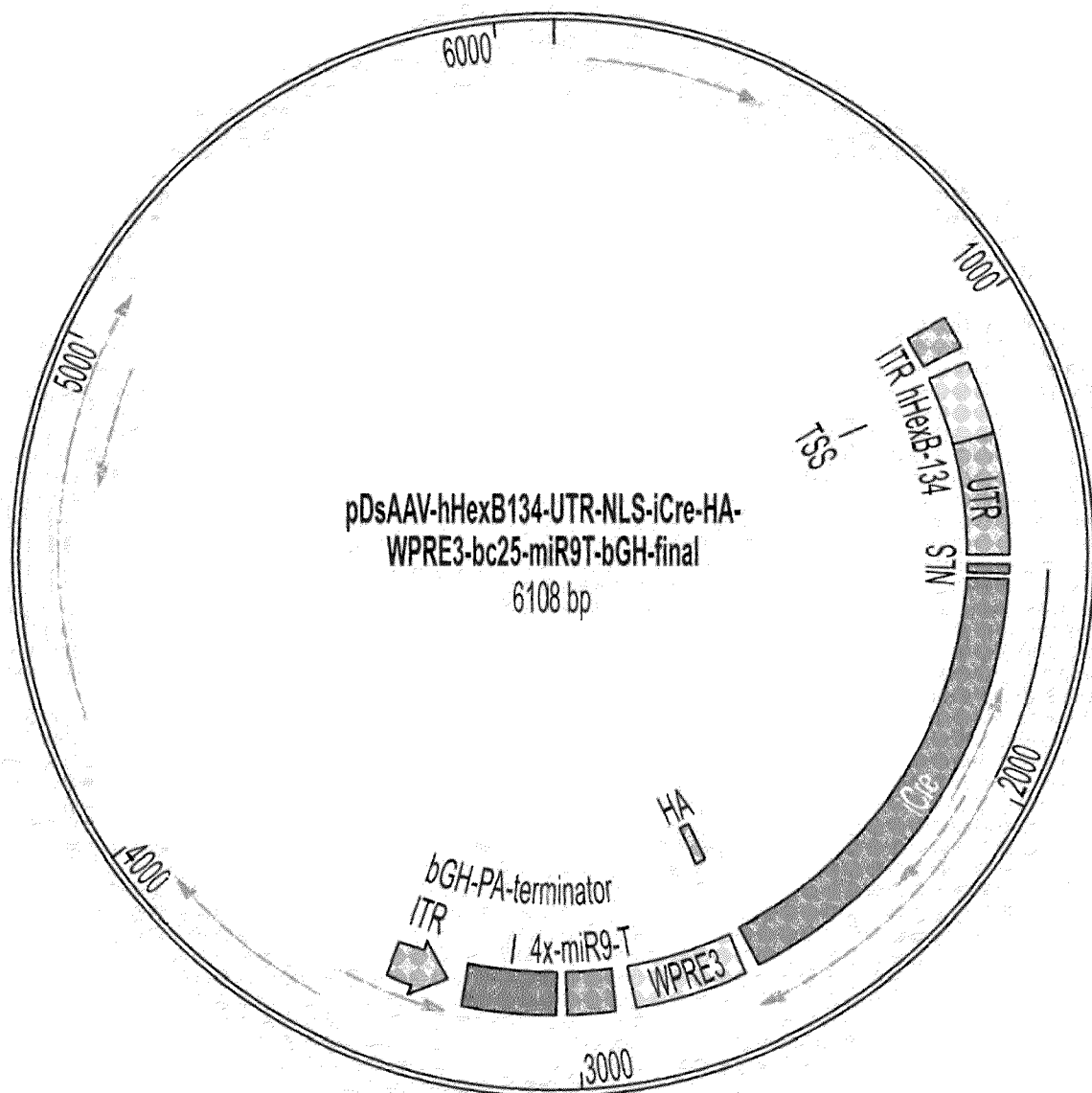


FIG. 45
CONTINUED

AAV barcode reporter vectorpDsAAV-hCD68-135-**UTR**-**NLS**-iCre-**HA**-WPRES-bc25-miR9T-bGH-**ITR**

ctgcgcgctcgctcgctcactgagggcgccgggcaaaagccgggctcgggcgacctttgggtcgccggg
cctcagtgagcgagcgagcgcgagagagggagtggaattcacgcgtggatctgaattcaattcacgcgt
ggtaccgccccctgagtcaggctgtgggtgggatcatctccagtcaggaagtgcagactttcatttccctcc
tttccaagagagggctgagggagcaggggtgagcaactggtgcagacagcctagctggactttgggtgag
gcggttcagccatatcgctcgacgCCACCGTTAACGCTAGCATTAAAGGTTTATACCTTCCCAGGTAACAA
ACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAAACGAACCTTAAAAATCTGTGTGGCTGTCACTC
GGCTGCATGCTTAGTGCACCTCAGGCAGTATAATTAATAACTAATTACTGTGTTGACAGGACACGAGTAA
CTCGTCTATCTTCTGCGGGCTGCTTACGGTTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGGTTT
CGTCCGGGTGTGACCGAAAGGTAAGaccgggtcgccaccatgCCaagaagaagCggaaagtggagtcctg
caacctgctgactgtgcacaaaacctgacctgacctgctggatgccacctctgatgaagtcaggaag
aacctgatggacatgttcagggacaggcagggccttctctgaacacacctggaagatgctcctgtctgtgt
gcagatcctgggtgctggtgcaagctgaacaacaggaatgggttccctgctgaacctgaggatgtgag
ggactacctcctgtacctgcaagccagagggcctggctgtgaagaccatccaacagcacctgggccagctc
aacatgctgcacaggagatctggcctgcctcgcccttctgactccaatgctgtgtcctgggtgatgagga
gaatcagaaaggagaatgtggatgctggggagagagccaagcagggcctggcctttgaacgcactgactt
tgaccaagtgcagatcctgatggagaactctgacagatgccaggacatcaggaacctggccttccctgggc
attgctacacacacctgctgcgcattgcccgaatggcagaatcagagtgaggacatctcccgacccg
atgggtgggagaatgctgatccacattggcaggacccaagacctgggtgtccacagctgggtgtggagaaggc
cctgtccctgggggttaccaagctgggtggagagatggatctctgtgtctgggtgtggctgatgaccccaac
aactacctgtttctgcgggtcagaaagaatgggtgtggctgccccttctgcccacctcccaactgtccaccc
gggccttggaagggatctttgaggccacccacgcctgatctatgggtgccaaggatgactctgggcagag
atacctggcctgggtctggcactctgcccagagtgggtgctgccaggacatggccagggctgggtgtgtcc
atccctgaaatcatgcaggctgggtggctggaccaatgtgaacattgtgatgaactacatcagaaacctgg
actctgagactggggccatgggtgaggtgctcgaggatgggtgactaccatacagatgltccagattaccc
Etaagaattogatatcaagcttatcgataatcaacctctggattacaaaatttgtaagagattgactgggt
attcttaactatgttgctccttttacgctatgtggatacctgctttaatgcctttgtatcatgtattg
cttcccgatggctttcattttctcctccttgatataaatcctgggtAGTTCttgccacggcggaactcat
cgccgctgcttgccegetgctggacaggggctcggtgttgggactgacaattccgtggNNNNNNNN
NNNNNNNNNNNNNNNAagtaaagcgccatcaagctTCATACAGCTAGATAACCAAAGCGATTTCATAC
AGCTAGATAACCAAAGACGCGTTCATACAGCTAGATAACCAAAGTCACTCATAACAGCTAGATAACCAAAG
gtcgactagagctcgctgatcagcctcgactgtgccttctagtggccagccatctgttgtttgccctcc
cccgtgccttcccttgacctggaagggtgccactcccactgtcctttcctaataaaaatgaggaaattgcat
cgcattgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacagcaagggggaggattg
ggaagacaatagcaggcatgctggggagagatcgatctaggaacctctagtgtgaggttggcactcc
Tctctgcgcgctcgctcgctcactgagggcgccgggcaaaagccgggctcgggcgacctttgggtcgcc
cgccctcagtcagcgagcgagcgcgagagagggagtgccaa

FIG. 46

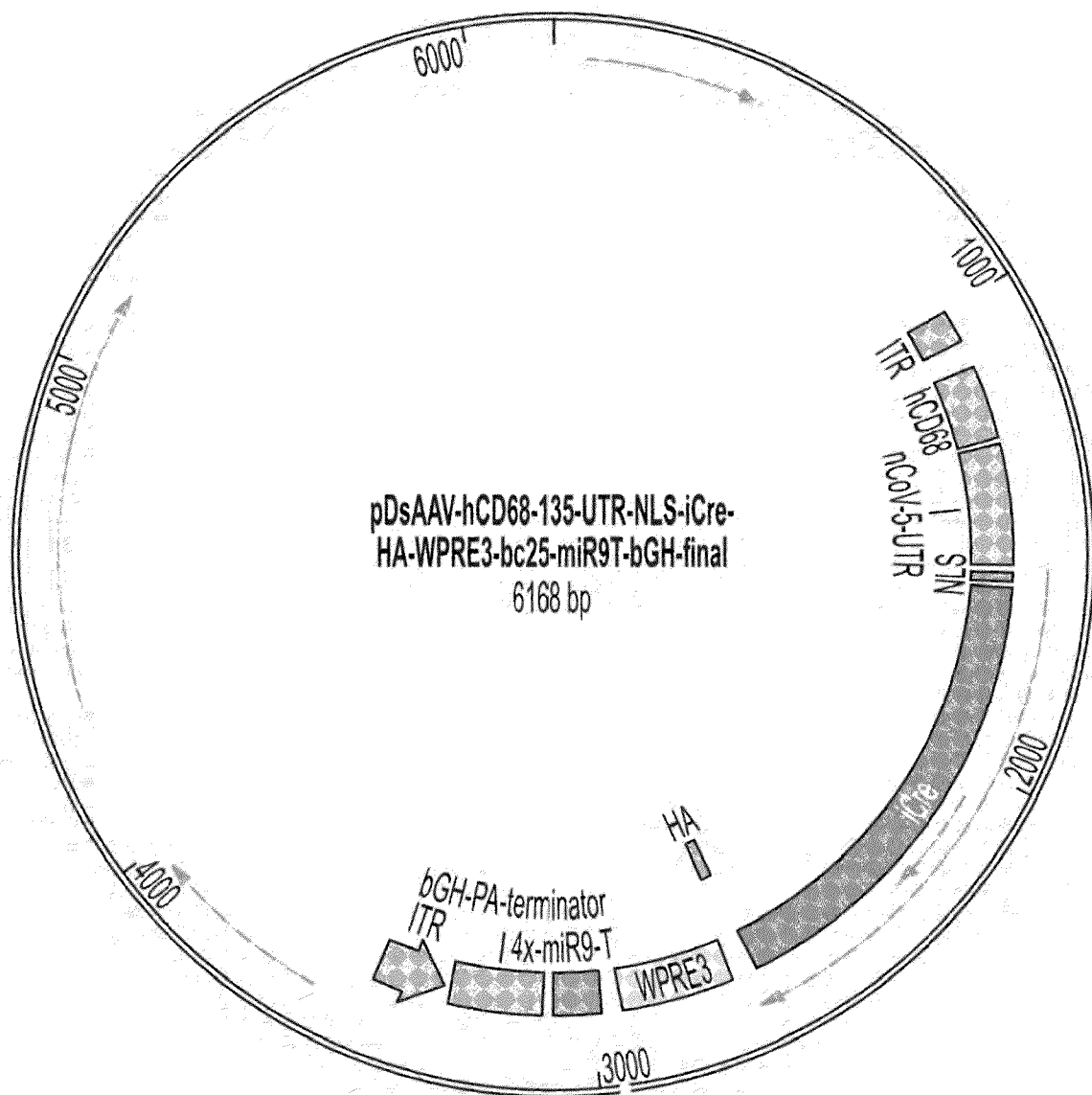


FIG. 46
CONTINUED