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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 November 2020 (26.11.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/237219 A1

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

(21) International Application Number:
PCT/US2020/034394

(22) International Filing Date:
22 May 2020 (22.05.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/852,216 23 May 2019 (23.05.2019) US

(71) Applicant: **ROCKET PHARMACEUTICALS, LTD.** [US/US]; The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US).

(72) Inventors: **BEARD, Brian**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US). **RICKS, David**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US). **PRABHAKAR, Raj**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US).

(74) Agent: **CHRISTIANSEN, William** et al.; Cooley LLP, 1299 Pennsylvania Aveenue, NW, Suite 700, Washington, District of Columbia 20004 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

(54) Title: GENE THERAPY VECTORS FOR INFANTILE MALIGNANT OSTEOPETROSIS

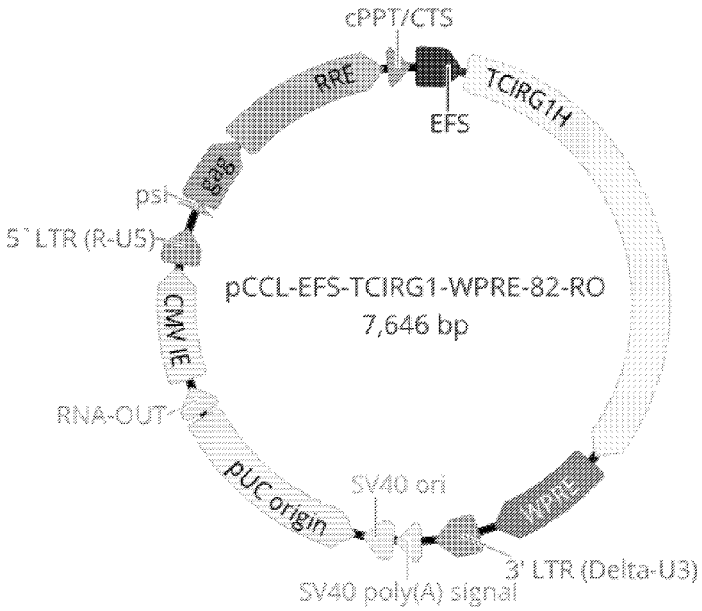


FIG. 1

(57) Abstract: The present disclosure provides improved gene therapy vectors comprising a polynucleotide sequence encoding a TCIRG1 polypeptide or functional variant thereof, methods of use thereof, pharmaceutical compositions, and more. In particular, the disclosure provides lentiviral vectors for treatment of infantile malignant osteopetrosis (IMO).

WO 2020/237219 A1

[Continued on next page]

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

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

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

Published:

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

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

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

GENE THERAPY VECTORS FOR INFANTILE MALIGNANT OSTEOPETROSIS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the priority benefit of U.S. Provisional Application Serial No. 62/852,216, filed on May 23, 2019, the contents of this application are hereby incorporated by reference herein in their entirety.

STATEMENT REGARDING THE SEQUENCE LISTING

[0002] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is ROPA_003_01WO_ST25.txt. The text file is about 62 KB, created on May 22, 2020, and is being submitted electronically via EFS-Web.

FIELD

[0003] The disclosure relates generally to gene therapy for diseases associated with mutations in a T cell immune regulator 1, ATPase H+ transporting V0 subunit a3 gene (*TCIRG1*). In particular, the disclosure provides gene therapy vectors and plasmids comprising expression cassettes that encode TCIRG1 protein (TCIRG1).

BACKGROUND

[0004] Infantile malignant osteopetrosis (IMO) is a rare, recessive disorder characterized by increased bone mass caused by dysfunctional osteoclasts. The disease is most often caused by mutations in T cell immune regulator 1, ATPase H+ transporting V0 subunit a3 (*TCIRG1*). TCIRG1 is involved in osteoclasts' capacity to resorb bone.

[0005] Osteoclast function can be restored by lentiviral vector-mediated expression of TCIRG1. Moscatelli et al. *Bone* 57:1-9 (2013). Further studies show that lentiviral-mediated expression of TCIRG1 is regulated in the same manner as the endogenous gene product despite being expressed by a lentiviral vector with a constitutive physiologic promoter. Thudium et al. *Calcif Tissue Int.* 99:638-648 (2016). In addition, they established that the natural *TCIRG1* gene sequence leads to higher level of protein expression and functional

rescue in osteoclasts than a codon-optimized cDNA of the gene, even though mRNA levels from the latter were considerably higher. Furthermore, the data show that only a low fraction of human pre-osteoclasts with a functional TCIRG1 is needed to significantly increase resorptive function *in vitro*, likely due to the fusion of resorbing and non-resorbing osteoclasts, in line with previous results from the oc/oc mouse model of osteopetrosis. From both an efficacy and a safety perspective, the findings are encouraging for the further development of gene therapy for osteopetrosis.

[0006] There remains a need in the art for gene therapy vectors for TCIRG1 and method of treatment using such vectors. Furthermore, there is a need for reliable methods of producing such gene therapy vectors. The present disclosure provides such gene therapy vectors, methods of manufacture thereof, methods of use thereof, pharmaceutical compositions, and more.

SUMMARY OF THE INVENTION

[0007] The present disclosure provides improved gene therapy vectors comprising a polynucleotide sequence encoding a TCIRG1 polypeptide or functional variant thereof, methods of use thereof, pharmaceutical compositions, and more.

[0008] In one aspect, the disclosure provides a transfer plasmid comprising an expression cassette comprising a coding polynucleotide encoding an isoform of T cell immune regulator 1 (TCIRG1) or a functional variant thereof, and a promoter, wherein the polynucleotide is operatively linked to the promoter, and wherein the transfer plasmid comprises an RNA-OUT repressor and a CMV IE promoter.

[0009] In some embodiments, the RNA-OUT repressor shares at least 95% identity or at least 99% identity to SEQ ID NO: 32.

[0010] In some embodiments, the CMV IE promoter shares at least 95% identity or at least 99% identity to SEQ ID NO: 33.

[0011] In some embodiments, the transfer plasmid comprises a pCCL backbone

[0012] In some embodiments, the pCCL backbone comprises the RNA-OUT repressor.

[0013] In some embodiments, the transfer plasmid shares at least 95% identity to SEQ ID NO: 39.

[0014] In some embodiments, the transfer plasmid comprises SEQ ID NO: 39.

[0015] In some embodiments, the promoter is an EFS promoter.

[0016] In some embodiments, the EFS promoter shares at least 95% identity with SEQ ID NO: 2.

[0017] In some embodiments, the EFS promoter is SEQ ID NO: 2.

[0018] In some embodiments, the coding polynucleotide shares at least 95% identity with SEQ ID NO: 3.

[0019] In some embodiments, the coding polynucleotide shares at least 99% identity with SEQ ID NO: 3.

[0020] In some embodiments, the coding polynucleotide is SEQ ID NO: 3.

[0021] In some embodiments, the expression cassette comprises a Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE).

[0022] In some embodiments, the WPRE is SEQ ID NO: 4.

[0023] In some embodiments, the expression cassette shares at least 95% identity with SEQ ID NO: 1.

[0024] In some embodiments, the expression cassette is flanked by a 5′ long terminal repeat (LTR) and a 3′ LTR.

[0025] In some embodiments, the 5′ LTR is SEQ ID NO: 34 and/or the 3′ LTR is SEQ ID NO: 28.

[0026] In some embodiments, expression cassette shares at least 95% identity to SEQ ID NO: 1.

[0027] In some embodiments, expression cassette is SEQ ID NO: 1.

[0028] In another aspect, the disclosure provides an expression cassette comprising a polynucleotide encoding an isoform of T cell immune regulator 1 (TCIRG1), or a functional variant thereof, and EFS promoter, wherein optionally the polynucleotide is operatively linked to the EFS promoter.

[0029] In some embodiments, the coding polynucleotide shares at least 95% identity with SEQ ID NO: 3. In some embodiments, the coding polynucleotide shares at least 99% identity with SEQ ID NO: 3. In some embodiments, the coding polynucleotide is SEQ ID NO: 3. In some embodiments, the EFS promoter shares at least 95% identity with SEQ ID NO: 2. In some embodiments, the EFS promoter is SEQ ID NO: 2.

[0030] In some embodiments, the expression cassette comprises a Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE). In some embodiments, the WPRE is SEQ ID NO: 4.

[0031] In some embodiments, the expression cassette shares at least 95% identity with SEQ ID NO: 1. In some embodiments, the expression cassette is SEQ ID NO: 1.

[0032] In another aspect, the disclosure provides a recombinant lentiviral genome, comprising in 5' to 3' order a lentiviral 5' long terminal repeat (LTR); an expression cassette disclosed herein; and a lentiviral 3'LTR, wherein the recombinant lentiviral genome is replication incompetent.

[0033] In another aspect, the disclosure provides a lentiviral particle, comprising such a recombinant lentiviral genome.

[0034] In another aspect, the disclosure provides a transfer plasmid comprising such a recombinant lentiviral genome. In certain embodiments, the transfer plasmid comprises an RNA-OUT sequence. In some embodiments, the RNA-OUT sequence is SEQ ID NO: 22. In some embodiments, the RNA-OUT sequence is configured such that the transfer plasmid is capable of stable propagation in a packaging cell line.

[0035] In particular embodiments, the transfer plasmid does not comprise an antibiotic resistance gene or does not comprise an ampicillin resistance gene, such as AmpR.

[0036] In particular embodiments, the transfer plasmid comprises the sequence set forth in SEQ ID NO: 23.

[0037] In another aspect, the disclosure provides a method of generating a lentiviral particle, comprising transfecting a packaging cell line with any transfer plasmid of the disclosure, and optionally one or more additional plasmid, and culturing said packaging cell line. In some embodiments, the transfer plasmid is stably propagated in a bacterial host at 30-37°C using shake flasks or fermentation for at least 1, 2, 3, 4, 5, 6, or 7 days.

[0038] In a related aspect, the disclosure provides a lentiviral particle produced using a transfer plasmid disclosed herein.

[0039] In another aspect, the disclosure provides a pharmaceutical composition comprising any lentiviral particle of the disclosure

[0040] In another aspect, the disclosure provides a modified cell comprising any expression cassette of the disclosure.

[0041] In another aspect, the disclosure provides a modified cell comprising any recombinant lentiviral genome of the disclosure.

[0042] In some embodiments, the modified cell lacks an endogenous functional *TCIRG1* gene.

[0043] In some embodiments, the modified cell is derived from a subject having or suspected of having infantile malignant osteopetrosis (IMO).

[0044] In some embodiments, the modified cell expresses TCIRG1 or a functional variant thereof at a level similar to the level of expression of TCIRG1 observed in an osteoclast having a functional *TCIRG1* gene.

[0045] In some embodiments, the modified cell expresses TCIRG1 or a functional variant thereof at a level similar to the level of expression of TCIRG1 observed in an osteoclast derived from a subject not having or suspected of having IMO.

[0046] In some embodiments, the modified cell is a hematopoietic stem cell (HSC).

[0047] In some embodiments, the modified cell is a CD34+ progenitor cell.

[0048] In some embodiments, the modified cell is derived a HSC isolated from a subject having or suspected of having IMO by apheresis.

[0049] In some embodiments, the modified cell is derived a HSC isolated from a subject having or suspected of having IMO by apheresis after mobilization of HSCs by administration of G-CSF, plerifaxor, or a combination of G-CSF and plerifaxor.

[0050] In some embodiments, the modified cell is derived from a population of cells enriched for CD34+ cells by magnetic capture.

[0051] In another aspect, the disclosure provides a pharmaceutical composition comprising any modified cell of the disclosure.

[0052] In another aspect, the disclosure provides an *in vitro* method of modifying one or more cells of a subject having or suspected of having IMO, comprising providing peripheral blood mononuclear cells (PBMCs) mobilized from the subject by administering to the subject a composition comprising G-CSF, plerifaxor, or a combination of G-CSF and plerifaxor; enriching the PBMCs for CD34+ cells by magnetic separation to generate a population of CD34-enriched cells; and contacting the CD34-enriched cells with a lentiviral particle comprising a recombinant lentiviral genome, comprising in 5' to 3' order: a lentiviral 5' long terminal repeat (LTR); any expression cassette of the disclosure; and a lentiviral 3'LTR, wherein the recombinant lentiviral genome is replication incompetent.

[0053] In another aspect, the disclosure provides a method of treating infantile malignant osteopetrosis (IMO) in a subject having or suspected of having IMO, comprising administering any modified cell of the disclosure or any pharmaceutical composition of the disclosure to the subject.

[0054] In some embodiments, the method repopulates the HSC niche with modified cells expressing TCIRG1 or a functional variant thereof.

[0055] In some embodiments, the method repopulates the osteoclast niche with modified cells expressing TCIRG1 or a functional variant thereof.

[0056] In some embodiments, the method treats, ameliorates, prevents, reduces, inhibits, or relieves IMO.

[0057] In some embodiments, the method extends the mean overall survival of treated subjects by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more years.

[0058] In some embodiments, the method prevents the death of the subject from IMO.

[0059] In some embodiments, the subject is a human.

[0060] In some embodiments, the subject exhibited symptoms of IMO before treatment.

[0061] In some embodiments, the subject was identified as having reduced or non-detectable expression of TCIRG1 before treatment.

[0062] In some embodiments, the subject was identified as having a mutated TCIRG1 gene.

[0063] In some embodiments, the subject is an infant.

[0064] In some embodiments, the method comprises autologous treatment.

[0065] In some embodiments, the administration is performed via a intravenous infusion.

[0066] In another aspect, the disclosure provides a recombinant lentiviral genome for use in the preparation of a medicament for treating or preventing infantile malignant osteopetrosis (IMO), wherein the lentiviral genome comprises in 5' to 3' order a lentiviral 5' long terminal repeat (LTR), any expression cassette of the disclosure, and a lentiviral 3'LTR; and wherein the recombinant lentiviral genome is replication incompetent.

[0067] In another aspect, the disclosure provides a lentiviral particle for use in the preparation of a medicament for treating or preventing infantile malignant osteopetrosis (IMO), comprising a recombinant lentiviral genome, wherein the lentiviral genome comprises in 5' to 3' order: a lentiviral 5' long terminal repeat (LTR), any expression cassette of the disclosure, and a lentiviral 3'LTR; and wherein the recombinant lentiviral genome is replication incompetent.

[0068] Other features and advantages of the invention will be apparent from and encompassed by the following detailed description and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0069] **FIG. 1** provides a diagram of a transfer plasmid for producing a lentiviral gene therapy vector encoding TCIRG1 (pCCL.PPT.EFS.tcirg1h.wpre).

[0070] **FIG. 2** shows the gene sequence of an expression cassette (SEQ ID NO: 1), including in 5' to 3' order an elongation factor 1-α short (EFS) promoter (underlined; SEQ ID NO: 2), a polynucleotide encoding TCIRG1 (white letters on black; SEQ ID NO: 3), and a Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE) (underlined and bold; SEQ ID NO: 4).

[0071] **FIGS. 3A-3B** provides a comparison of the stability of two different lentiviral plasmids. **FIG. 3A** shows a photograph of an agarose gel stained with ethidium bromide showing plasmid pRRL.PPT.EFS.tcirg1h.wpre, either not digested with a restriction enzyme ("Uncut") or digested with AflIII ("AflIII") or AflIII and NarI ("AflIII/NarI"). **FIG. 3B** shows a photograph of an agarose gel stained with ethidium bromide showing the plasmid pCCL.PPT.EFS.tcirg1h.wpre, either not digested with a restriction enzyme ("Uncut") or digested with AflIII ("AflIII") or AflIII and NarI ("AflIII/NarI"). **FIG. 3C** shows schematic diagrams of the pRRL.PPT.EFS.tcirg1h.wpre and pCCL.PPT.EFS.tcirg1h.wpre plasmids.

[0072] **FIG. 4** depicts an illustrative process for lentiviral particle manufacturing.

[0073] **FIGS. 5A-5B** show (**FIG. 5A**) vector copy number (VCN) in bulk CD34+ cells liquid culture 6 and (**FIG. 5B**) 12 days after transduction. VCN was assessed by qPCR of extracted gDNA after culturing transduced CD34+ cells in SCGM complete media. VCN for each donor and the mean is represented for each transduction condition.

DETAILED DESCRIPTION

[0074] The present inventors have shown that transplantation of autologous cells transduced with a lentiviral vector encoding TCIRG1 is effective in treating infantile malignant osteopetrosis (IMO). In addition, the inclusion of specific sequence elements in the expression cassette sequences of gene therapy vectors encoding TCIRG1 result in a safe and effective gene therapy for IMO. The present disclosure provides lentiviral vectors and plasmids encoding TCIRG1, including stable transfer plasmids advantageous for producing the lentiviral vectors.

Vectors and Plasmids

[0075] The inventors have surprisingly discovered that production of lentivirus vector for TCIRG1 gene therapy at large scale is improved by modifying a pRRL plasmid containing

the desired expression cassette in two ways: (i) replacing the pRRL vector backbone with a pCCL vector backbone, and (ii) replacing a conventional antibiotic resistance cassette in the pCCL backbone with the RNA-OUT selectable marker. The improved plasmid is then transfected into a lentiviral particle production system, along with helper plasmids, to produce the desired lentiviral vector.

[0076] The resulting pCCL/RNA-OUT vector for TCIRG1 gene therapy (*e.g.*, the vector depicted FIG. 1) has improved stability, reflected in higher plasmid yields from *E. coli*-based plasmid production and reduced levels of undesirable recombination products in the purified plasmid (shown in Example 1 and FIG. 3A-3C). This improvement to the transfer plasmid enables manufacture of lentiviral particles comprising the TCIRG1 expression cassette in yields sufficient for clinical testing and use. Further data provided herein demonstrate that lentiviral particles produced using the methods and compositions disclosed herein transduces $CD34^+$ cells efficiently enough to reach clinically relevant vector copy number (VCN) levels.

[0077] In some embodiments, the disclosure provides a transfer plasmid that is a lentiviral vector based on the pCCL transfer plasmid used in third-generation lentiviral vector systems. The pCCL transfer plasmid contains the chimeric cytomegalovirus (CMV)-HIV 5′ LTR and vector backbones in which the simian virus 40 polyadenylation and (enhancerless) origin of replication sequences have been included downstream of the HIV 3′ LTR, replacing most of the human sequence remaining from the HIV integration site. The CCL 5′ hybrid long terminal repeat (LTR) is the enhancer and promoter (nucleotides -673 to -1 relative to the transcriptional start site; GenBank accession no. K03104) of cytomegalovirus (CMV) joined to the R region of HIV-1 LTR. In some embodiments, the transfer plasmid comprises an EFS promoter linked to the TCIRG1 gene with upstream RRE and cPPT/CTS elements and a downstream WPRE element (Figure 1). In some embodiments, the transfer plasmid comprises a PGK promoter (SEQ ID NO: 24) linked to the TCIRG1 gene with upstream RRE and cPPT/CTS elements and a downstream WPRE element. In some embodiments, the transfer plasmid comprises an RNA-OUT element. Advantageously, the RNA-OUT sequence contributes to stable propagation of the transfer plasmid in a packing cell line. In some embodiments, the transfer plasmid does not comprise an antibiotic resistance gene, e.g., AmpR.

[0078] In some embodiments, the PGK promoter comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%,

89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 24. In some embodiments, the PGK promoter comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 24. In some embodiments, the PGK promoter has the sequence SEQ ID NO: 24.

GGGGTTGGGGTTGCGCCTTTTCCAAGGCAGCCCTGGGTTTGCGCAGGGACGCGGCTGCTCTGGGCGTGGTTCCGG
GAAACGCAGCGGCGCCGACCCTGGGTCTCGCACATTCTTCACGTCCGTTCGCAGCGTCACCCGGATCTTCGCCGC
TACCCTTGTGGGCCCCCCGGCGACGCTTCCTGCTCCGCCCCTAAGTCGGGAAGGTTCCTTGCGGTTCGCGGCGTG
CCGGACGTGACAAACGGAAGCCGCACGTCTCACTAGTACCCTCGCAGACGGACAGCGCCAGGGAGCAATGGCAGC
GCGCCGACCGCGATGGGCTGTGGCCAATAGCGGCTGCTCAGCAGGGCGCGCCGAGAGCAGCGGCCGGGAAGGGGC
GGTGCGGGAGGCGGGGTGTGGGGCGGTAGTGTGGGCCCTGTTCCTGCCCGCGCGGTGTTCCGCATTCTGCAAGCC
TCCGGAGCGCACGTCGGCAGTCGGCTCCCTCGTTGACCGAATCACCGACCTCTCTCCCCAG

(SEQ ID NO: 24)

[0079] In certain embodiments, the transfer plasmid is more stable than another plasmid comprising the same expression cassette when cultured or propagated in E. coli, thus resulting in a higher yield of plasmid, which is advantageous for use in producing vector. In some embodiments, the transfer plasmid is more stable that the pRRL.PPT.EFS.tcirg1h.wpre transfer plasmid. In particular embodiments, at least 2-fold, at least 5-fold, or at least 10-fold more of the transfer plasmid is produced as compared to the amount of pRRL.PPT.EFS.tcirg1h.wpre transfer plasmid produced under the same culture conditions.

[0080] In particular embodiments, the transfer plasmid is pCCL.PPT.EFS.tcirg1h.wpre or functional variants thereof, e.g., those disclosed herein. The disclosure provides in particular embodiments, the transfer plasmid pCCL.PPT.EFS.tcirg1h.wrpe or functional variants thereof. The transfer plasmid pCCL.PPT.EFS.tcirg1h.wrpe may have the sequence SEQ ID NO: 23.

[0081] Alternatively, the transfer plasmid pCCL.PPT.EFS.tcirg1h.wrpe may have the sequence SEQ ID NO: 25, in which the sequence GATCACGAGACTAGCCTCGAGAAGCTTGATCGATTGGCTCCGGTGCC (SEQ ID NO: 26) is deleted.

[0082] The sequence SEQ ID NO: 25 represents a circular plasmid. The same sequence permuted to start with the EFS promoter at base pair 1 is provided as SEQ ID NO: 27.

[0083] In some embodiments, the transfer plasmid comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%,

89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 23. In some embodiments, the transfer plasmid comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 23. In some embodiments, the transfer plasmid has the sequence SEQ ID NO: 23.

[0084] In some embodiments, the transfer plasmid comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 25. In some embodiments, the transfer plasmid comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 25. In some embodiments, the transfer plasmid has the sequence SEQ ID NO: 25.

[0085] In some embodiments, the transfer plasmid comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 27. In some embodiments, the transfer plasmid comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 27. In some embodiments, the transfer plasmid has the sequence SEQ ID NO: 27.

[0086] In some embodiments, the transfer plasmid is comprises one or more of the vector elements listed in **Table 1**.

Table 1: pCCL.PPT.EFS.tcirg1h.wpre Vector Elements

Name	Type	Min. – Max. Position in Reference Sequence (nucleotide)	Length (base pairs)	SEQ ID NO:
EFS	Promoter	1-243	243	SEQ ID NO: 2
TCIRG1	CDS	257-2,749	2,493	SEQ ID NO: 3
WPRE	Regulatory	2,782-3,384	603	SEQ ID NO: 4
3`LTR	LTR	3,471-3,704	234	SEQ ID NO: 28
SV40 poly(A)	polyA signal	3,776-3,907	132	SEQ ID NO: 29
SV40 ori	Origin of replication	3,917-4,076	160	SEQ ID NO: 30
pUC origin	Origin of replication	4,115-5,129	1,015	SEQ ID NO: 31
RNA-OUT	Repressor	5,146-5,284	139	SEQ ID NO: 32
CMV IE	Promoter	5,334-5,910	577	SEQ ID NO: 33
5-LTR	LTR	5,933-6,120	188	SEQ ID NO: 34

psi	Packaging	6,222-6,266	45	SEQ ID NO: 35
gag	CDS	6,267-6,628	362	SEQ ID NO: 36
RRE	Regulatory	6,629-7,486	858	SEQ ID NO: 37
cPPT/CTS	Poly purine tract	7,505-7,622	118	SEQ ID NO: 38
-	Backbone	3,776-7,622	3,847	SEQ ID NO: 39

[0087] In some embodiments, lentiviral particles are generated by transient transfection of a third-generation lentiviral vector system that includes pCCL.PPT.EFS.tcirg1h.wpre.

[0088] In some embodiments, the expression cassette comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 1. In some embodiments, the expression cassette comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 1. In some embodiments, the expression cassette has the sequence SEQ ID NO: 1.

[0089] In some embodiments, the expression cassette comprises, in 5' to 3' order, an EFS promoter, a polynucleotide encoding T cell immune regulator 1, ATPase H+ transporting V0 subunit a3 (TCIRG1) or a functional variant thereof, and a Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE). In some embodiments, the EFS promoter is operatively linked to the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof. Related embodiments comprise a transfer plasmid comprising the expression cassette, and a vector produced using the transfer plasmid.

[0090] In some embodiments, the EFS promoter comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 2. In some embodiments, the EFS promoter comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 2. In some embodiments, the EFS promoter has the sequence SEQ ID NO: 2.

GGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAG
TTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTA
AACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAG
AACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGC
CGCCAGAACACAGGTGTCGTGACGC

(SEQ ID NO: 2)

[0091] In some embodiments, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 3. In some embodiments, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 3. In some embodiments, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof has the sequence SEQ ID NO: 3.

[0092] In some embodiments, the WPRE comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 4. In some embodiments, the WPRE comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 4. In some embodiments, the WPRE has the sequence SEQ ID NO: 4.

ATTCGAGCATCTTACCGCCATTTATACCCATATTTGTTCTGTTTTTCTTGATTTGGG
TATACATTTAAATGTTAATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAG
GGATATGTAATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAAC
TTTCCCGTTATTTACGCTCTGTTCCTGTTAATCAACCTCTGGATTACAAAATTTGT
GAAAGATTGACTGATATTCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATATGC
TGCTTTAATGCCTCTGTATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTC
CTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTC
AACGTGGCGTGGTGTGCTCTGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCAT
TGCCACCACCTGTCAACTCCTTTCTGGGACTTTCGCTTTCCCCCTCCCGATCGCCA
CGGCAGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTAGGTTGCT
GGGCACTGATAATTCCGTGGTGTTGTCGGGGAAGCTGACGTCCTTTCG

(SEQ ID NO: 4)

[0093] In some embodiments, the isoform of TCIRG1 or a functional variant thereof comprises a polypeptide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 5. In some embodiments, the isoform of TCIRG1 or a functional variant thereof comprises a polypeptide that shares at least 80%, 85%, 90%,

95%, 99%, or 100% identity with SEQ ID NO: 5. In some embodiments, the isoform of TCIRG1 or a functional variant thereof has the sequence SEQ ID NO: 5.

MGSMFRSEEVALVQLFLPTAAAYTCVSRLGELGLVEFRDLNASVSAFQRRFVVDVR
RCEELEKTFTFLQEEVRRAGLVLPPPKGRLPAPPPRDLLRIQEETERLAQELRDVRGN
QQALRAQLHQLQLHAAVLRQGHEPQLAAAHTDGASERTPLLQAPGGPHQDLRVNF
VAGAVEPHKAPALERLLWRACRGFLIASFRELEQPLEHPVTGEPATWMTFLISYWGE
QIGQKIRKITDCFHCHVFPFLQQEEARLGALQQLQQQSQELQEVLGETERFLSQVLGR
VLQLLPPGQVQVHKMKAVYLALNQCSVSTTHKCLIAEAWCSVRDLPALQEALRDSS
MEEGVSAVAHRIPCRDMPPTLIRTNRFTASFQGIVDAYGVGRYQEVNPAPYTIITFPFL
FAVMFGDVGHGLLMFLFALAMVLAENRPAVKAAQNEIWQTFFRGRYLLLLMGLFSI
YTGFIYNECFSRATSIFPSGWSVAAMANQSGWSDAFLAQHTMLTLDPNVTGVFLGP
YPFGIDPIWSLAANHLSFLNSFKMKMSVILGVVHMAFGVVLGVFNHVHFGQRHRLL
LETLPELTFLLGLFGYLVFLVIYKWLCVWAARAASAPSILIHFINMFLFSHSPSNRLLY
PRQEVVQATLVVLALAMVPILLLGTPLHLLHRHRRRLRRRPADRQEENKAGLLDLP
DASVNGWSSDEEKAGGLDDEEEAELVPSEVLMHQAIHTIEFCLGCVSNTASYLRLW
ALSLAHAQLSEVLWAMVMRIGLGLGREVGVAAVVLVPIFAAFAVMTVAILLVMEG
LSAFLHALRLHWVEFQNKFYSGTGYKLSPFTFAATDD

(SEQ ID NO: 5)

[0094] In an embodiment, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof is codon-optimized for expression in a human host cell. In an embodiment, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof is modified, or "codon optimized" to enhance expression by replacing infrequently represented codons with more frequently represented codons. In an embodiment, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof is not codon-optimized. In an embodiment, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof is not modified. In an embodiment, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof is not codon-optimized. In an embodiment, the polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof is a native polynucleotide sequence.

[0095] As used herein the term "transgene" refers to a polynucleotide encoding an isoform of TCIRG1 or a functional variant thereof.

[0096] The coding sequence is the portion of the mRNA sequence that encodes the amino acids for translation. During translation, each of 61 trinucleotide codons are translated to one of 20 amino acids, leading to a degeneracy, or redundancy, in the genetic code. However, different cell types, and different animal species, utilize tRNAs (each bearing an anticodon) coding for the same amino acids at different frequencies. When a gene sequence contains codons that are infrequently represented by the corresponding tRNA, the ribosome translation machinery may slow, impeding efficient translation. Expression can be improved via "codon optimization" for a particular species, where the coding sequence is altered to encode the same protein sequence, but utilizing codons that are highly represented, and/or utilized by highly expressed human proteins (Cid-Arregui et al., 2003; J. Virol. 77:4928).

[0097] In some embodiments, the coding sequence of the transgene is modified to replace codons infrequently expressed in mammal or in primates with codons frequently expressed in primates. For example, in some embodiments, the transgene encodes a polypeptide having at least 85% sequence identity to a reference polypeptide (*e.g.* wild-type TCIRG1; SEQ ID NO: 3)—for example, at least 90% sequence identity, at least 95% sequence identity, at least 98% identity, or at least 99% identity to the reference polypeptide—wherein at least one codon of the coding sequence has a higher tRNA frequency in humans than the corresponding codon in the sequence disclosed above or herein.

[0098] In an embodiment, the transgene comprises fewer alternative open reading frames than SEQ ID: 3. In an embodiment, the transgene is modified to enhance expression by termination or removal of open reading frames (ORFs) that do not encode the desired transgene. An open reading frame (ORF) is the nucleic acid sequence that follows a start codon and does not contain a stop codon. ORFs may be in the forward or reverse orientation, and may be "in frame" or "out of frame" compared with the gene of interest. Such open reading frames have the potential to be expressed in an expression cassette alongside the gene of interest, and could lead to undesired adverse effects. In some embodiments the transgene has been modified to remove open reading frames by further altering codon usage. This is done by eliminating one or more start codons (ATG) and/or introducing one or more stop codons (TAG, TAA, or TGA) in reverse orientation or out-of-frame to the desired ORF, while preserving the encoded amino acid sequence and, optionally, maintaining highly utilized codons in the gene of interest (*i.e.*, avoiding codons with frequency < 20%).

[0099] In variations of the present disclosure, the transgene coding sequence may be optimized by either of codon optimization and removal of non-transgene ORFs or using both techniques. In some cases, one removes or minimizes non-transgene ORFs after codon optimization in order to remove ORFs introduced during codon optimization.

[0100] In an embodiment, the transgene contains fewer CpG sites than SEQ ID: 3. Without being bound by theory, it is believed that the presence of CpG sites in a polynucleotide sequence is associated with the undesirable immunological responses of the host against a viral vector comprising the polynucleotide sequence. In some embodiments, the transgene is designed to reduce the number of CpG sites. Exemplary methods are provides in U.S. Patent Application Publication No. US20020065236A1.

[0101] In an embodiment, the transgene contains fewer cryptic splice sites than SEQ ID: 3. For the optimization, GeneArt® software may be used, *e.g.*, to increase the GC content and/or remove cryptic splice sites in order to avoid transcriptional silencing and, therefore, increase transgene expression. Alternatively, any optimization method known in the art may be used. Removal of cryptic splice sites is described, for example, in International Patent Application Publication No. WO2004015106A1.

[0102] Also disclosed herein are expression cassettes and gene therapy vectors encoding TCIRG1, *e.g.*, a TCIRG1 sequence disclosed herein, comprising: a consensus optimal Kozak sequence, a full-length polyadenylation (polyA) sequence (or substitution of full-length polyA for a truncated polyA), and minimal or no upstream (*i.e.* 5') start codons (*i.e.* ATG sites).

[0103] In some embodiments, the expression cassette contains two or more of a 5' long terminal repeat (LTR), an enhancer/promoter region, a consensus optimal Kozak sequence, a transgene (*e.g.*, a transgene encoding a TCIRG1 disclosed herein), a 3' untranslated region including a full-length polyA sequence, and a 3' LTR.

[0104] In an embodiment, the expression cassette comprises a Kozak sequence operatively linked to the transgene. In an embodiment, the Kozak sequence is a consensus optimal Kozak sequence comprising or consisting of SEQ ID NO: 6.

GCCGCCACCATGG (SEQ ID NO: 6)

[0105] In various embodiments, the expression cassette comprises an alternative Kozak sequence operatively linked to the transgene. In an embodiment, the Kozak sequence is an alternative Kozak sequence comprising or consisting of any one of SEQ ID NOs. 14-18.

(gcc)gccRccAUGG (SEQ ID NO: 14)

AGNNAUGN (SEQ ID NO: 15)

ANNAUGG (SEQ ID NO: 16)

ACCAUGG (SEQ ID NO: 17)

GACACCAUGG (SEQ ID NO: 18)

[0106] In SEQ ID NO: 14, a lower-case letter denotes the most common base at a position where the base can nevertheless vary; an upper-case letter indicate a highly conserved base; indicates adenine or guanine. In SEQ ID NO: 14, the sequence in parentheses (gcc) is optional. IN SEQ ID NOs: 15-17, 'N' denotes any base.

[0107] A variety of sequences can be used in place of this consensus optimal Kozak sequence as the translation-initiation site and it is within the skill of those in the art to identify and test other sequences. *See* Kozak M. An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* 115 (4): 887–903 (1991).

[0108] In an embodiment, the expression cassette comprises a full-length polyA sequence operatively linked to the transgene. In an embodiment, the full-length polyA sequence comprises SEQ ID NO: 7.

TGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAAG
GACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGC
CCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGC
TGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGTGTT
ATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTG
ACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGATC

(SEQ ID NO: 7)

[0109] Various alternative polyA sequences may be used in expression cassettes of the present disclosure, including without limitation, bovine growth hormone polyadenylation signal (bGHpA) (SEQ ID NO: 19), the SV40 early/late polyadenylation signal (SEQ ID NO: 20), and human growth hormone (HGH) polyadenylation signal (SEQ ID NO: 21).

TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAA
GGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAT
TCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAGGACAATAGCAGGCATGCT
GGGGATGCGGTGGGCTCTATGGCTTCTG

(SEQ ID NO: 19)

CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTA
TTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACA
ACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAACC
TCTACAAATGTGGTA

(SEQ ID NO: 20)

CTGCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTG
CCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAATATT
ATGGGGTGGAGGGGGGTGGTATGGAGCAAGGGGCCCAAGTTGGGAAGAAACCTGTAGGGCCTGC

(SEQ ID NO: 21)

[0110] In some embodiments, the expression cassette comprises an active fragment of a polyA sequence. In particular embodiments, the active fragment of the polA sequence comprises or consists of less than 20 base pair (bp), less than 50 bp, less than 100 bp, or less than 150 bp, e.g., of any of the polA sequences disclosed herein.

[0111] In some cases, expression of the transgene is increased by ensuring that the expression cassette does not contain competing ORFs. In an embodiment, the expression cassette comprises no start codon within 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, or 500 basepairs 5' of the start codon of the transgene. In an embodiment, the expression cassette comprises no start codon 5' of the start codon of the transgene.

[0112] In an embodiment, the expression cassette comprises operatively linked, in the 5' to 3' direction, a first inverse terminal repeat, an enhancer/promoter region, introns, a consensus optimal Kozak sequence, the transgene, a 3' untranslated region including a full-length polyA sequence, and a second inverse terminal repeat, where the expression cassette comprises no start codon 5' to the start codon of the transgene.

[0113] In an embodiment, the enhancer/promoter region comprises, in the 5' to 3' direction: a CMV IE Enhancer and a Chicken Beta-Actin Promoter. In an embodiment, the enhancer/promoter region comprises a CAG promoter. As used herein "CAG promoter" refers to a polynucleotide sequence comprising a CMV early enhancer element, a chicken

beta-actin promoter, the first exon and first intron of the chicken beta-actin gene, and a splice acceptor from the rabbit beta-globin gene.

[0114] In an embodiment, the enhancer/promoter region comprises an elongation factor 1α short promoter (EFS promoter) and is a shorter intron-less version of elongation factor 1α promoter. As used herein "EFS promoter" refers to a polynucleotide sequence comprising a short, intron-less form of EF1alpha. The EFS promoter has been recently used in many clinical trials. It is a cellular-derived enhancer/promoter with decreased cross-activation of nearby promoters, therefore hypothetically decreasing the risk of genotoxicity.

[0115] In an embodiment, the expression cassette shares at least 95% identity to a sequence selected from SEQ ID NOs: 1. In an embodiment, the expression cassette shares complete identity to a sequence selected from SEQ ID NOs: 1, or shares at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to a sequence selected from SEQ ID NOs: 1. In certain embodiments, the expression cassette comprises one or more modifications as compare to a sequence selected from SEQ ID NOs: 1. In particular embodiments, the one or more modifications comprises one or more of: removal of one or more (e.g., all) upstream ATG sequences, replacement of the Kozak sequence with an optimized consensus Kozak sequence or another Kozak sequence, including but not limited to any of those disclosed herein, and/or replacement of the polyadenylation sequence with a full-length polyadenylation sequence or another polyadenylation sequence, including but not limited to any of those disclosed herein. An illustrative configuration of genetic elements within these exemplary expression cassettes is depicted in **FIG. 1**.

[0116] In related embodiments, the disclosure provides gene therapy vectors comprising an expression cassette disclosed herein. Generally, the gene therapy vectors described herein comprise an expression cassette comprising a polynucleotide encoding one or more isoforms of TCIRG1, that allows for the expression of TCIRG1 to partially or wholly rectify deficient TCIRG1 protein expression levels and/or a defect in osteoclast formation in a subject in need thereof (*e.g.*, a subject having Infantile Malignant Osteopetrosis or another disorder characterized by deficient osteoclast formation at least in part due to deficient TCIRG1 expression). In particular embodiments, the expression cassette comprises a polynucleotide sequence encoding TCIRG1 disclosed herein, *e.g.*, SEQ ID NOs:3 or a sequence having at least 90%, at least 95%, at least 98%, or at least 99% identity to any of SEQ ID NO:s: 3. The gene therapy vectors can be viral or non-viral vectors. Illustrative non-viral vectors include,

e.g., naked DNA, cationic liposome complexes, cationic polymer complexes, cationic liposome-polymer complexes, and exosomes. Examples of viral vector include, but are not limited, to adenoviral, retroviral, lentiviral, herpesvirus and adeno-associated virus (AAV) vectors.

[0117] Gene delivery viral vectors useful in the practice of the present invention can be constructed utilizing methodologies well known in the art of molecular biology. Typically, viral vectors carrying transgenes are assembled from polynucleotides encoding the transgene, suitable regulatory elements and elements necessary for production of viral proteins, which mediate cell transduction. Such recombinant viruses may be produced by techniques known in the art, *e.g.*, by transfecting packaging cells or by transient transfection with helper plasmids or viruses. Typical examples of virus packaging cells include but are not limited to HeLa cells, SF9 cells (optionally with a baculovirus helper vector), 293 cells, etc. A Herpesvirus-based system can be used to produce AAV vectors, as described in US20170218395A1. Detailed protocols for producing such replication-defective recombinant viruses may be found for instance in W095/14785, W096/22378, U.S. Pat. No. 5,882,877, U.S. Pat. No. 6,013,516, U.S. Pat. No. 4,861,719, U.S. Pat. No. 5,278,056 and W094/19478, the complete contents of each of which is hereby incorporated by reference.

[0118] In some embodiments, the vector is a retroviral vector, or more specifically, a lentiviral vector. As used herein, the term "retrovirus" or "retroviral" refers an RNA virus that reverse transcribes its genomic RNA into a linear double-stranded DNA copy and subsequently covalently integrates its genomic DNA into a host genome. Retrovirus vectors are a common tool for gene delivery (Miller, 2000, *Nature*. 357: 455-460). Once the virus is integrated into the host genome, it is referred to as a "provirus." The provirus serves as a template for RNA polymerase II and directs the expression of RNA molecules encoded by the virus.

[0119] Illustrative retroviruses (family Retroviridae) include, but are not limited to: (1) genus gammaretrovirus, such as, Moloney murine leukemia virus (M-MuLV), Moloney murine sarcoma virus (MoMSV), murine mammary tumor virus (MuMTV), gibbon ape leukemia virus (GaLV), and feline leukemia virus (FLV), (2) genus spumavirus, such as, simian foamy virus, (3) genus lentivirus, such as, human immunodeficiency virus-1 and simian immunodeficiency virus.

[0120] As used herein, the term "lentiviral" or "lentivirus" refers to a group (or genus) of complex retroviruses. Illustrative lentiviruses include, but are not limited to: HIV (human immunodeficiency virus; including HIV type 1, and HIV type 2; visna-maedi virus (VMV) virus; the caprine arthritis-encephalitis virus (CAEV); equine infectious anemia virus (EIAV); feline immunodeficiency virus (FIV); bovine immune deficiency virus (BIV); and simian immunodeficiency virus (SIV). In one embodiment, HIV-based vector backbones (*i.e.*, HIV cis-acting sequence elements) are preferred.

[0121] Retroviral vectors, and more particularly, lentiviral vectors, may be used in practicing the present invention. Accordingly, the term "retroviral vector," as used herein is meant to include "lentiviral vector"; and the term "retrovirus" as used herein is meant to include "lentivirus."

[0122] The term viral vector may refer either to a vector or viral particle capable of transferring a nucleic acid into a cell or to the transferred nucleic acid itself. Viral vectors contain structural and/or functional genetic elements that are primarily derived from a virus. The term "retroviral vector" refers to a viral vector containing structural and functional genetic elements, or portions thereof, that are primarily derived from a retrovirus. The term "lentiviral vector" refers to a viral vector containing structural and functional genetic elements, or portions thereof, including LTRs that are primarily derived from a lentivirus. The term "hybrid" refers to a vector, LTR or other nucleic acid containing both retroviral, *e.g.*, lentiviral, sequences and non-lentiviral viral sequences. In one embodiment, a hybrid vector refers to a vector or transfer plasmid comprising retroviral, *e.g.*, lentiviral, sequences for reverse transcription, replication, integration and/or packaging.

[0123] In particular embodiments, the terms "lentiviral vector" and "lentiviral expression vector" may be used to refer to lentiviral transfer plasmids and/or infectious lentiviral particles. Where reference is made herein to elements such as cloning sites, promoters, regulatory elements, heterologous nucleic acids, *etc.*, it is to be understood that the sequences of these elements are present in RNA form in the lentiviral particles of the invention and are present in DNA form in the DNA plasmids of the invention.

[0124] According to certain specific embodiments, most or all of the viral vector backbone sequences are derived from a lentivirus, *e.g.*, HIV-1. However, it is to be understood that many different sources of lentiviral sequences can be used, and numerous

substitutions and alterations in certain of the lentiviral sequences may be accommodated without impairing the ability of a transfer vector to perform the functions described herein. Moreover, a variety of lentiviral vectors are known in the art, *see* Naldini *et al.*, (1996a, 1996b, and 1998); Zufferey *et al.,* (1997); Dull *et al.*, 1998, U.S. Pat. Nos. 6,013,516; and 5,994,136, many of which may be adapted to produce a viral vector or transfer plasmid of the present invention.

[0125] In preparing lentiviral vector, any host cells for producing lentiviral vectors may be employed, including, for example, mammalian cells (e.g. HEK 293T cells). Host cells can also be packaging cells in which the lentiviral *gag/pol* and *rev* genes are stably maintained in the host cell or producer cells in which the lentiviral vector genome is stably maintained and packaged. Lentiviral vectors are purified and formulated using standard techniques known in the art.

[0126] In certain embodiments, the present invention includes a cell comprising a gene expression cassette, gene transfer cassette, or recombinant lentiviral vector disclosed herein. In related embodiments, the cell is transduced with a recombinant lentiviral vector comprising an expression cassette disclosed herein or has an expression cassette disclosed herein integrated into the cell's genome. In certain embodiments, the cell is a cell used to produce a recombinant retroviral vector, *e.g.*, a packaging cell.

[0127] In some embodiments, the lentiviral vector is pseudotyped. For example, a plasmid comprising a heterologous *env* gene can be used for pseudotyping. Suitable *env* genes include, without limitation, VSV-G.

[0128] In some embodiments, the backbone of the transfer plasmid comprises an RNA-OUT sequence. RNA-OUT is a selectable marker system that facilitates selection of cells harboring the transfer plasmid within the use of antibiotics, as described, *e.g.*, in U.S. Patent Nos. 9,109,012 and 9,737,620, which are incorporated by reference herein. In some embodiments, the RNA-OUT sequence is:

GTAGAATTGGTAAAGAGAGTCGTGTAAAATATCGAGTTCGCACATCTTGTTGTCTGATTATTGATTTTT
GGCGAAACCATTTGATCATATGACAAGATGTGTATCTACCTTAACTTAATGATTTTGATAAAAATCATT
AGG

(SEQ ID NO: 22)

[0129] Advantageously, the RNA-OUT sequence contributes to stable propagation of the transfer plasmid in a packing cell line.

[0130] In some embodiments, the disclosure provides a transfer the expression cassette comprises a polynucleotide that shares at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity with SEQ ID NO: 1. In some embodiments, the expression cassette comprises a polynucleotide that shares at least 80%, 85%, 90%, 95%, 99%, or 100% identity with SEQ ID NO: 1. In some embodiments, the expression cassette has the sequence SEQ ID NO: 1.

[0131] AAV is a 4.7 kb, single stranded DNA virus. Recombinant vectors based on AAV are associated with excellent clinical safety, since wild-type AAV is nonpathogenic and has no etiologic association with any known diseases. In addition, AAV offers the capability for highly efficient gene delivery and sustained transgene expression in numerous tissues. By an "AAV vector" is meant a vector derived from an adeno-associated virus serotype, including without limitation, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAVrh.10, AAVrh.74, *etc*. AAV vectors can have one or more of the AAV wild-type genes deleted in whole or part, *e.g.*, the rep and/or cap genes, but retain functional flanking inverted terminal repeat (ITR) sequences. Functional ITR sequences are necessary for the rescue, replication and packaging of the AAV virion. Thus, an AAV vector is defined herein to include at least those sequences required in cis for replication and packaging (*e.g.*, functional ITRs) of the virus. The ITRs need not be the wild-type nucleotide sequences, and may be altered, *e.g.* by the insertion, deletion or substitution of nucleotides, as long as the sequences provide for functional rescue, replication and packaging. AAV vectors may comprise other modifications, including but not limited to one or more modified capsid protein (e.g., VP1, VP2 and/or VP3). For example, a capsid protein may be modified to alter tropism and/or reduce immunogenicity. AAV expression vectors are constructed using known techniques to at least provide as operatively linked components in the direction of transcription, control elements including a transcriptional initiation region, the DNA of interest (*i.e.* the TCIRG1 gene) and a transcriptional termination region.

Pharmaceutical Compositions and Methods of Use

[0132] The present disclosure also provides pharmaceutical compositions comprising an expression cassette or vector (*e.g.*, gene therapy vector) disclosed herein and one or more pharmaceutically acceptable carriers, diluents or excipients. In some embodiments, the pharmaceutical composition comprises a lentiviral particle comprising an expression cassette disclosed herein, *e.g.*, wherein the expression cassette comprises a codon-transgene encoding TCIRG1, *e.g.*, SEQ ID NOs: 3. Provided are pharmaceutical compositions, *e.g.*, for use in preventing or treating a disorder characterized by deficient osteoclast formation (*e.g.*, Infantile Malignant Osteopetrosis) which comprises a therapeutically effective amount of a lentiviral particle that comprises a nucleic acid sequence of a polynucleotide that encodes one or more isoforms of TCIRG1.

[0133] In particular embodiments, the lentiviral particles disclosed herein are used to transduce autologous CD34+ hematopoietic stem cells (HSCs) derived from a subject, thus complementing the genetic defect. Transduction may occur *in vivo* or *ex vivo*. The CD34+ enriched cell population is cultured, in some embodiments, in CellGenix Stem Cell Growth Media (SCGM) with recombinant human cytokines and incubated in 5% CO_2 and 5% O_2 at 37°C. The CD34+ enriched cell population is, in some embodiments, incubated with the same additives as used for the pre-stimulation, optionally with the addition of transduction enhancers, and lentiviral particles comprising the expression cassette EFS-TCIRG1-WPRE (at, for example, MOI 50). Following transduction, the cell suspension is, in some embodiments, washed a portion of cells and supernatant are removed for release testing and the drug product is frozen in preparation for infusion.In some embodiments, HSC are mobilized by treating the patient with G-CSF, plerifaxor, or a combination of G-CSF and plerifaxor. The HSCs are then collected from peripheral blood of the patient by apheresis. CD34+ cells are enriched, e.g., using magnetic capture (*e.g.*, on the Miltenyi Biotec CliniMACs system), and the CD34+ enriched cells are transduced *ex vivo* with the lentiviral particles. In some embodiments, the transduction process incorporates the use of transductions enhancers, such as, without limitation, polyaxamers and Prostaglandin E2 (PGE2).

[0134] In some embodiments, the transduced HSCs are then transplanted into a subject, *e.g.*, a human subject, by infusion with at least $2.0x10^6$ CD34+ cells/kg. In some embodiments, they repopulate the HSC niche with TCIRG1-expressing cells. In some embodiments, they repopulate the osteoclast niche with TCIRG1-expressing cells.

[0135] Provided also are pharmaceutical compositions, *e.g.*, for use in preventing or treating a disorder characterized by deficient osteoclast formation (*e.g.*, Infantile Malignant Osteopetrosis) which comprises a therapeutically effective amount of a modified cell that comprises a nucleic acid sequence of a polynucleotide that encodes one or more isoforms of TCIRG1. In some embodiments, the modified cell expresses TCIRG1 or a functional variant thereof at a level similar to the level of expression of TCIRG1 observed in an osteoclast having a functional TCIRG1 gene. In some embodiments, the modified cell expresses TCIRG1 or a functional variant thereof at a level similar to the level of expression of TCIRG1 observed in an osteoclast derived from a subject not having or suspected of having IMO. In some embodiments, the modified cell is a hematopoietic stem cell (HSC). In some embodiments, the modified cell is a CD34+ progenitor cell. In some embodiments, the modified cell is derived a HSC isolated from a subject having or suspected of having IMO by apheresis. In some embodiments, the modified cell is autologous to the subject. In some embodiments, the modified cell is derived a HSC isolated from a subject having or suspected of having IMO by apheresis after mobilization of HSCs by administration of G-CSF, plerifaxor, or a combination of G-CSF and plerifaxor. In some embodiments, the modified cell is derived from a population of cells enriched for CD34+ cells by magnetic capture. In some embodiments, the modified cell was transduced using a vector disclosed herein, e.g., a lentiviral vector produced using a transfer plasmid disclosed herein.

[0136] The pharmaceutical compositions that contain the expression cassette or lentiviral particle or modified cell may be in any form that is suitable for the selected mode of administration, for example, for intraventricular, intramyocardial, intracoronary, intravenous, intra-arterial, intra-renal, intraurethral, epidural or intramuscular administration. The gene modified cell comprising a polynucleotide encoding one or more TCIRG1 isoforms can be administered as sole active agent, or in combination with other active agents, in a unit administration form, as a mixture with conventional pharmaceutical supports, to animals and human beings. In some embodiments, the pharmaceutical composition comprises cells transduced *ex vivo* with any of the gene therapy vectors of the disclosure.

[0137] In various embodiments, the pharmaceutical compositions contain vehicles (*e.g.*, carriers, diluents and excipients) that are pharmaceutically acceptable for a formulation capable of being injected. These may be in particular isotonic, sterile, saline solutions (monosodium or disodium phosphate, sodium, potassium, calcium or magnesium chloride

and the like or mixtures of such salts), or dry, especially freeze-dried compositions which upon addition, depending on the case, of sterilized water or physiological saline, permit the constitution of injectable solutions. Illustrative pharmaceutical forms suitable for injectable use include, *e.g.*, sterile aqueous solutions or dispersions; formulations including sesame oil, peanut oil or aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions.

[0138] In another aspect, the disclosure provides methods of preventing, mitigating, ameliorating, reducing, inhibiting, eliminating and/or reversing one or more symptoms of Infantile Malignant Osteopetrosis (IMO) or another disorder in a subject in need thereof, comprising administering to the subject a gene therapy vector of the disclosure. The term "Infantile Malignant Osteopetrosis" or "malignant infantile osteopetrosis" or "infantile autosomal recessive osteopetrosis" or "infantile osteopetrosis" or "IMO" refers to a rare osteosclerosis type of skeletal dysplasia that typically presents in infancy and is characterized by a unique radiographic appearance of generalized hyperostosis – excessive growth of bone. The generalized increase in bone density has a special predilection to involve the medullary portion with relative sparing of the cortices. Obliteration of bone marrow spaces and subsequent depression of the cellular function can result in serious hematologic complications. Optic atrophy and cranial nerve damage secondary to bony expansion can result in marked morbidity. The prognosis is extremely poor in untreated cases Plain radiography provides the key information to the diagnosis. Clinical and radiologic correlations are also fundamental to the diagnostic process, with additional gene testing being confirmatory.

[0139] In an embodiment, the modified cell, *e.g.* an autologous cell transduced with a lentiviral particle of the disclosure, is administered via a route selected from the group consisting of parenteral, intravenous, intra-arterial, intracardiac, intracoronary, intramyocardial, intrarenal, intraurethral, epidural, and intramuscular. In some embodiments, the modified cells is administered by infusion, *e.g.* intravenous infusion. In an embodiment, modified cells are administered multiple times. In an embodiment, modified cells are administered by infusion.

[0140] In an embodiment, the disclosure provides a method of treating a disease or disorder, optionally IMO, in a subject in need thereof, comprising contacting cells with a gene therapy vector according to the present disclosure and administering the cells to the

subject. In an embodiment, the cells are stem cells, optionally pluripotent stem cells. In an embodiment, the stem cells are capable of differentiation into bone cells. In an embodiment, the stem cells are capable of differentiation into osteoclasts. In an embodiment, the stem cells are autologous. In an embodiment, the stem cells are CD34+ stem cells.

[0141] In an embodiment, the subject is exhibiting symptoms of IMO or another disorder. In an embodiment, the subject has been identified as having reduced or non-detectable TCIRG1 expression. In an embodiment, the subject has been identified as having a mutated TCIRG1 gene.

[0142] Subjects/patients amenable to treatment using the methods described herein include individuals at risk of a disease or disorder characterized by insufficient osteoclasts (*e.g.*, IMO as well as other known disorders of osteoclast formation. In some embodiments, the subject is not showing symptoms. In some embodiments, subjects is presently showing symptoms. Such subject may have been identified as having a mutated TCIRG1 gene or as having reduced or non-detectable levels of TCIRG1 expression. The symptoms may be actively manifesting, or may be suppressed or controlled (*e.g.*, by medication) or in remission. The subject may or may not have been diagnosed with the disorder, *e.g.*, by a qualified physician.

Definitions

[0143] The terms "T cell immune regulator 1, ATPase H+ transporting V0 subunit a3" and "TCIRG1" interchangeably refer to nucleic acids and polypeptide polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have an amino acid sequence that has greater than about 90% amino acid sequence identity, for example, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or greater amino acid sequence identity, preferably over a region of at least about 25, 50, 100, 200, 300, 400, or more amino acids, or over the full-length, to an amino acid sequence encoded by a TCIRG1 nucleic acid (*see, e.g.*, GenBank Accession Nos. NM_006019.4 (variant 1). NM_006053.3 (variant 2), NM_001351059.1 (variant 3)) or to an amino acid sequence of a TCIRG1 polypeptide (*see e.g.*, GenBank Accession Nos. NP_006044.1 (isoform A), NP_006044.1 (isoform B), NP_001337988.1 (isoform C)); (2) bind to antibodies, *e.g.*, polyclonal antibodies, raised against an immunogen comprising an amino acid sequence of a TCIRG1 polypeptide (*e.g.*, TCIRG1 polypeptides described herein); or an amino acid sequence encoded by a TCIRG1 nucleic acid (*e.g.*,

TCIRG1 polynucleotides described herein), and conservatively modified variants thereof; (3) specifically hybridize under stringent hybridization conditions to an anti-sense strand corresponding to a nucleic acid sequence encoding a TCIRG1 protein, and conservatively modified variants thereof; (4) have a nucleic acid sequence that has greater than about 90%, preferably greater than about 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher nucleotide sequence identity, preferably over a region of at least about 25, 50, 100, 200, 500, 1000, 2000 or more nucleotides, or over the full-length, to a TCIRG1 nucleic acid (*e.g.*, TCIRG1 polynucleotides, as described herein, and TCIRG1 polynucleotides that encode TCIRG1 polypeptides, as described herein).

[0144] The *TCIRG1* gene encodes several protein isoforms, with 2 main isoforms. The full-length isoform a (OC116) encodes the A3 subunit of vacuolar H(+)-ATPase, which is involved in regulation of the pH of intracellular compartments and organelles of eukaryotic cells, including the pH of intracellular compartments and organelles of osteoclasts. The shorter isoform b (TIRC7) encodes a T-cell-specific membrane protein that plays an essential role in T-lymphocyte activation and immune response.

[0145] The terms "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 (*i.e.* , share at least about 80% identity, for example, at least about 85%, 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identity over a specified region to a reference sequence, *e.g.*, TCIRG1 polynucleotide or polypeptide sequence as described herein, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the compliment of a test sequence. Preferably, the identity exists over a region that is at least about 25 amino acids or nucleotides in length, for example, over a region that is 50, 100, 200, 300, 400 amino acids or nucleotides in length, or over the full-length of a reference sequence.

[0146] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. 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. 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. For sequence comparison of nucleic acids and proteins to TCIRG1 nucleic acids and proteins, the BLAST and BLAST 2.0 algorithms and the default parameters are used.

[0147] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 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. 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, P*roc. 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, Wl), or by manual alignment and visual inspection (see, *e.g.*, Ausubel et al., eds., Current Protocols in Molecular Biology (1995 supplement)). 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., *Nucleic Acids Res.* 25:3389-3402 (1977) and Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (on the worldwide web at ncbi.nlm.nih.gov).

[0148] An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions. Yet

another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

[0149] As used herein, "administering" refers to local and systemic administration, *e.g.*, including enteral, parenteral, pulmonary, and topical/transdermal administration. Routes of administration for compounds (*e.g.*, polynucleotide encoding one or more TCIRG1 isoforms) that find use in the methods described herein include, *e.g.*, oral (per os (P.O.)) administration, nasal or inhalation administration, administration as a suppository, topical contact, transdermal delivery (*e.g.*, via a transdermal patch), intrathecal (IT) administration, intravenous ("iv") administration, intraperitoneal ("ip") administration, intramuscular ("im") administration, intralesional administration, or subcutaneous ("sc") administration, or the implantation of a slow-release device *e.g.*, a mini-osmotic pump, a depot formulation, etc. , to a subject. Administration can be by any route including parenteral and transmucosal (*e.g.*, oral, nasal, vaginal, rectal, or transdermal). Parenteral administration includes, *e.g.*, intravenous, intramuscular, intraarterial, intrarenal, intraurethral, intracardiac, intracoronary, intramyocardial, intradermal, epidural, subcutaneous, intraperitoneal, intraventricular, ionophoretic and intracranial. Other modes of delivery include, but are not limited to, the use of liposomal formulations, intravenous infusion, transdermal patches, etc.

[0150] The terms "systemic administration" and "systemically administered" refer to a method of administering a compound or composition to a mammal so that the compound or composition is delivered to sites in the body, including the targeted site of pharmaceutical action, via the circulatory system. Systemic administration includes, but is not limited to, oral, intranasal, rectal and parenteral (*e.g.*, other than through the alimentary tract, such as intramuscular, intravenous, intra-arterial, transdermal and subcutaneous) administration.

[0151] The term "co-administering" or "concurrent administration", when used, for example with respect to the compounds (*e.g.*, TCIRG1 polynucleotides) and/or analogs thereof and another active agent, refers to administration of the compound and/or analogs and the active agent such that both can simultaneously achieve a physiological effect. The two agents, however, need not be administered together. In certain embodiments, administration of one agent can precede administration of the other. Simultaneous physiological effect need not necessarily require presence of both agents in the circulation at the same time. However, in certain embodiments, co-administering typically results in both agents being simultaneously present in the body (*e.g.*, in the plasma) at a significant fraction (*e.g.*, 20% or

greater, *e.g.*, 30% or 40% or greater, *e.g.*, 50% or 60% or greater, *e.g.*, 70% or 80% or 90% or greater) of their maximum serum concentration for any given dose.

[0152] The term "effective amount" or "pharmaceutically effective amount" refer to the amount and/or dosage, and/or dosage regime of one or more compositions (*e.g.*, gene therapy vectors, modified cells) necessary to bring about the desired result *e.g.*, increased expression of one or more TCIRG1 isoforms in an amount sufficient to reduce the ultimate severity of a disease characterized by impaired or deficient autophagy (*e.g.*, IMO).

[0153] The phrase "cause to be administered" refers to the actions taken by a medical professional (*e.g.*, a physician), or a person controlling medical care of a subject, that control and/or permit the administration of the agent(s)/compound(s) at issue to the subject. Causing to be administered can involve diagnosis and/or determination of an appropriate therapeutic or prophylactic regimen, and/or prescribing particular agent(s)/compounds for a subject. Such prescribing can include, for example, drafting a prescription form, annotating a medical record, and the like.

[0154] As used herein, the terms "treating" and "treatment" refer to delaying the onset of, retarding or reversing the progress of, reducing the severity of, or alleviating or preventing either the disease or condition to which the term applies, or one or more symptoms of such disease or condition. The terms "treating" and "treatment" also include preventing, mitigating, ameliorating, reducing, inhibiting, eliminating and/or reversing one or more symptoms of the disease or condition.

[0155] The term "mitigating" refers to reduction or elimination of one or more symptoms of that pathology or disease, and/or a reduction in the rate or delay of onset or severity of one or more symptoms of that pathology or disease, and/or the prevention of that pathology or disease. In certain embodiments, the reduction or elimination of one or more symptoms of pathology or disease can include, *e.g.*, measurable and sustained increase in the expression levels of one or more isoforms of TCIRG1.

[0156] As used herein, the phrase "consisting essentially of refers to the genera or species of active pharmaceutical agents recited in a method or composition, and further can include other agents that, on their own do not have substantial activity for the recited indication or purpose.

[0157] The terms "subject," "individual," and "patient" interchangeably refer to a mammal, preferably a human or a non-human primate, but also domesticated mammals (*e.g.*, canine or feline), laboratory mammals (*e.g.*, mouse, rat, rabbit, hamster, guinea pig) and agricultural mammals (*e.g.*, equine, bovine, porcine, ovine). In various embodiments, the subject can be a human (*e.g.*, adult male, adult female, adolescent male, adolescent female, male child, female child).

[0158] The terms "gene transfer" or "gene delivery" refer to methods or systems for reliably inserting foreign DNA into host cells. Such methods can result in transient expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred replicons (*e.g.* episomes), or integration of transferred genetic material into the genomic DNA of host cells.

[0159] The term "vector" is used herein to refer to a nucleic acid molecule capable transferring or transporting another nucleic acid molecule. The transferred nucleic acid is generally linked to, *e.g.*, inserted into, the vector nucleic acid molecule. A vector may include sequences that direct autonomous replication or reverse transcription in a cell, or may include sequences sufficient to allow integration into host cell DNA. "vectors" include gene therapy vectors. As used herein, the term "gene therapy vector" refers to a vector capable of use in performing gene therapy, e.g., delivering a polynucleotide sequence encoding a therapeutic polypeptide to a subject. Gene therapy vectors may comprise a nucleic acid molecule ("transgene") encoding a therapeutically active polypeptide, *e.g.*, a TCIRG1 or other gene useful for replacement gene therapy when introduced into a subject. Useful vectors include, but are not limited to, viral vectors.

[0160] As used herein, the term "expression cassette" refers to a DNA segment that is capable in an appropriate setting of driving the expression of a polynucleotide (*e.g.*, a transgene) encoding a therapeutically active polypeptide (*e.g.*, TCIRG1) that is incorporated in said expression cassette. When introduced into a host cell, an expression cassette *inter alia* is capable of directing the cell's machinery to transcribe the transgene into RNA, which is then usually further processed and finally translated into the therapeutically active polypeptide. The expression cassette can be comprised in a gene therapy vector. Generally, the term expression cassette excludes polynucleotide sequences 5' to the 5' LTR and 3' to the 3' LTR.

[0161] All patents, patent publications, and other publications referenced and identified in the present specification are individually and expressly incorporated herein by reference in their entirety for all purposes.

EXAMPLES

EXAMPLE 1: Stable Propagation of Transfer Plasmids

[0162] The stability of different plasmids comprising the minimal TCIRG1 expression cassette, EFS-TCIRG1-WPRE (SEQ ID NO: 1) was examined. The plasmid construct pRRL.PPT.EFS.tcirg1h.wpre (FIG. 3C) with ampicillin resistance (AmpR) showed unexpectedly poor growth and instability during culture in *E. coli* cells used to propagate the plasmid prior to transfection into a packaging cell line, as shown by low yield of plasmid and a general smear ofdegraded DNA indicative of instability (FIG. 3A). Various other plasmid backbones also showed instability (data not shown). However, when the minimal expression cassette EFS-TCIRG1-WPRE (SEQ ID NO: 1) was cloned from the pRRL vector into a pCCL vector with an RNA-OUT sequence (FIG. 3C), the resulting plasmid construct, pCCL.PPT.EFS.tcirg1h.wpre (SEQ ID NO: 27), exhibited unexpectedly good growth and stability when propagated in *E. coli*. This was shown by high yield of plasmid and the restriction digest pattern observed (FIG. 3B). The full vector sequence is provided as SEQ ID NO: 27, with the positions of each vector element provided in Table 2.

Table 2: pCCL.PPT.EFS.tcirg1h.wpre Vector Elements

Name	Type	Min. – Max. Position in Reference Sequence (nucleotide)	Length (base pairs)
EFS	Promoter	1-243	243
TCIRG1	CDS	257-2,749	2,493
WPRE	Regulatory	2,782-3,384	603
3`LTR	LTR	3,471-3,704	234
SV40 poly(A)	polyA signal	3,776-3,907	132
SV40 ori	Origin of replication	3,917-4,076	160
pUC origin	Origin of replication	4,115-5,129	1,015
RNA-OUT	Repressor	5,146-5,284	139
CMV IE	Promoter	5,334-5,910	577
5-LTR	LTR	5,933-6,120	188
psi	Packaging	6,222-6,266	45
gag	CDS	6,267-6,628	362

RRE	Regulatory	6,629-7,486	858
cPPT/CTS	Poly purine tract	7,505-7,622	118

[0163] Lentiviral vectors are produced by transient transfection of the pCCL/RNA-OUT vector into 293T cells along with packaging plasmid (pCMV ΔR8.91), and envelope plasmid (VSV-G pMDG) and produced according to the protocol depicted in FIG. 4.

EXAMPLE 2: Restoration of Resorptive Function of Osteoclasts from IMO Patients with pCCL.PPT.EFS.tcirg1h.wpre

[0164] This example demonstrates use of the pCCL.PPT.EFS.tcirg1h.wpre for lentiviral-mediated TCIRG1 gene transfer in patient-derived HSCs. HSCs are obtained, expanded and transduced with lentiviral particles carrying the pCCL.PPT.EFS.tcirg1h.wpre described in Example 1 to obtain gene-modified HSCs. Following infusion, the gene-modified HSCs will differentiate into osteoclasts. Methods used are essentially as described in Moscatelli et al. *Hum. Gene Therap.* 29:938-949 (2017).

[0165] Samples of peripheral blood from IMO patients or umbilical cord blood (CB) from normal deliveries are obtained. Mononuclear cells from these sources are isolated using density gradient centrifugation with Ficoll and CD34+ cells are separated from the mononuclear cell fraction using magnet-activated cell sorting (MACS) columns (Miltenyi Biotec, Bergisch Gladbach, Germany). For expansion, cells are cultured in SFEM StemSpan medium (StemCell Technologies, Vancouver, BC) with the human recombinant cytokines M-CSF (50 ng/ml), GM-CSF (30 ng/ml), SCF (200 ng/ml), IL-6 (10 ng/ml) and Flt3L (50 ng/ml) (R&D Systems, Minneapolis MN). CD34+ cells are plated at a density of $5x10^4$ cells in 1 ml medium using 24-well bacteriological plates and incubated for a week at 37°C before collection and replating at a density of $1x10^5$/well. From day 7 the medium is exchanged every 2-3 days by demi-depletion. For transplantation, CD34+ cells are cultured for 30 hours in SFEM StemSpan medium (StemCell Technologies, Vancouver, BC) with the human recombinant cytokines: SCF (100 ng/ml), Flt3L (100 ng/ml) and TPO (100 ng/ml) (R&D Systems, Minneapolis MN).

[0166] Transductions are carried out in 24-well plates coated with RetroNectin (Takara Bio, Otsu, Japan). For the *in vitro* experiments CD34+ cells were transduced with a first hit at a multiplicity of infection (MOI) of 30 for 6 hours on day 3 and a second hit at MOI 30 for 6 hours on day 7 followed by a week of culture with a myeloid cytokine cocktail and

subsequent differentiation to osteoclasts. For the *in vivo* experiments, a shorter transduction protocol is used to allow efficient transduction while maintaining the stem/progenitor nature of the CD34+ population. Mononuclear cells are isolated and transduced with the first hit (MOI 30 or 100) overnight, followed by transduction on the following day with a second hit (MOI 30 or 100) for six hours, after which the cells are ready for transplanted into subjects (mice or human patients).

[0167] Osteoclastogenesis can be assessed by differentiation for about ten days in the presence of 50 ng/ml M-CSF and 50 ng/ml RANKL, followed by fixation of the cells with 4% formaldehyde for further analyzes or lysis of cells for western blot analysis. Resorption is assessed by assaying for release c-terminal type I collagen fragments (CTX-I) and for release of Ca2+ into the media, and by visualization of the formation of resorption pits using hematoxylin staining of fixed cells.

[0168] For animal studies, transduced osteoclasts are transplanted into NSG mice. NSG mice, 8 to 15 week old, are sublethally irradiated with 300 cGy and transplanted six hours later with $1x10^5$ untransduced CB CD34+ cells or IMO CD34+ cells transduced with lentiviral particles derived from pCCL.PPT.EFS.tcirg1h.wpre, by tail vein injection. The mice are administered ciprofloxacin via their drinking water for two weeks to avoid post-transplantation infections. Peripheral blood was harvested at different time points and bone marrow cells are harvested by crushing the femora with a mortar after termination of the mice.

[0169] Vector copy number analysis is performed on whole bone marrow genomic DNA from samples harvested from mice 9-19 weeks after transplantation. Peripheral blood and bone marrow of transplanted NSG mice are analyzed for human reconstitution by determining the percentage of cells positive for huCD45-APC. For lineage analysis, the cells were stained with antibodies directed against CD33-PeCy7, CD15-PeCy7, CD19-BV605 and CD3-PE.

[0170] The methods described above are used to confirm restoration of resorptive function of osteoclasts from IMO patients after lentiviral-mediated TCIRG1 gene transfer and long-term engraftment of transduced CD34+ cells.

Example 3: In Vitro Transduction of Human CD34$^+$ Enriched Cells Using a Clinically Established Transduction Protocol

[0171] This example demonstrates suitability of a pre-GMP batch of EFS-TCIRG1-WPRE (SEQ ID NO: 25 or 27) made using the plasmid described in Example 1 by: (1) Phenotypic characterization of transduced $CD34^+$ cells by means of % viable $CD34^+$ cells and multilineage differentiation capacity, and (2) Transduction efficiency by means of vector copy number (VCN determination both in liquid culture and colonies.

Preserved Phenotype and Multilineage Capacity of mPB CD34+ Cells After Transduction with EFS-TCIRG1-WPRE

[0172] The performance of a pre-GMP EFS-TCIRG1-WPRE batch was compared with the LVs produced for the treatment of other disorders (four batches). Mobilized PB $CD34^+$ cells were used as target cells similarly as in the envisioned IMO clinical trial. Various MOIs were tested. High cell viability 20 hours after transduction (>95%) was obtained across all vectors and MOIs tested, indicating no short-term toxicity. The percentage of $CD34^+$ cells was very high for all conditions tested (>97%) shortly after transduction, and after 2 days in liquid culture was progressively lost over time at comparable levels in all conditions, as expected for this type of culture.

[0173] Multilineage capacity of transduced $CD34^+$ cells was evaluated by means of quantification of differentiated CFUs in semisolid methylcellulose medium cultures. Total CFUs as well as BFU-E, CFU-GM, and CFU-GEMM, accounting for the erythroid and myeloid lineages, were evaluated. The presence of the EFS-TCIRG1-WPRE did not affect CFU growth as no significant differences in comparison with a Mock control were observed in their total numbers among experimental conditions. No differences in comparison with the Mock were observed in any colony type thus confirming the maintenance of multilineage capacity *in vitro* after transduction with the EFS-TCIRG1-WPRE even at high MOI values.

EFS-TCIRG1-WPRE Shows High Transduction Efficiency of mPB $CD34^+$ Cells

[0174] To determine the vector dose of EFS-TCIRG1-WPRE required to provide suitable transduction efficiency for therapeutic use, lentiviral vectors were tested using the established clinical transduction protocol. Transduced $CD34^+$ cells were maintained in liquid culture for up to 12 days to allow for cellular clearance of episomal LV genome copies prior to VCN assessment. High VCN/cell values were obtained with the IMO vector that were dose-dependent. The effects of increasing dose were consistent across all vectors tested, and transduction with the same MOI resulted in higher VCN/cell values for the IMO vector than

for LAD-I and FA vectors (**FIGS. 5A-5B**), indicating high transduction efficiency of EFS-TCIRG1-WPRE.

[0175] VCN/cell was also evaluated in isolated CFUs depending on their phenotype: BFU-E, CFU-GM, or CFU-GEMM, to confirm transduction in different progenitors. EFS-TCIRG1-WPRE showed higher colony VCN values in cells from both donors, Similar to results in liquid culture, transduction with IMO vector resulted in high transduction efficiency and VCN/cell of colonies cultured in methylcellulose medium.

[0176] VCN pattern in the different CFU types (BFU-E, CFU-GM, and CFU-GEMM) was found to be similar to the other vectors and with the highest values usually found in the erythroid colonies as previously described in Charrier et al. *Gene Therapy* 18:479–487(2011).

[0177] The IMO vector EFS-TCIRG1-WPRE transduces human $CD34^+$ cells at levels comparable to clinical lots of lentivirus. The phenotype of $CD34^+$ cells and multilineage capacity were preserved while high transduction efficiency was achieved. IMO vector performed successfully at lower MOI than control vectors, demonstrating its suitability for use in the gene therapy of patients with IMO.

[0178] These studies demonstrated that a VCN and transduction efficiency was achieved that parallels corrective levels of *in vivo* gene-modified hematopoietic cells, enabling use of this gene therapy in the treatment of infantile malignant osteopetrosis due to mutations in *TCIRG1*.

CLAIMS

What is claimed is:

1. A transfer plasmid, comprising an expression cassette comprising a coding polynucleotide encoding an isoform of T cell immune regulator 1 (TCIRG1) or a functional variant thereof, and a promoter, wherein the polynucleotide is operatively linked to the promoter, and wherein the transfer plasmid comprises an RNA-OUT repressor and a CMV IE promoter.

2. The transfer plasmid of claim 1, wherein the RNA-OUT repressor shares at least 95% identity or at least 99% identity to SEQ ID NO: 32.

3. The transfer plasmid of claim 1 or claim 2, wherein the CMV IE promoter shares at least 95% identity or at least 99% identity to SEQ ID NO: 33.

4. The transfer plasmid of any one of claims 1 to 3, wherein the transfer plasmid comprises a pCCL backbone

5. The transfer plasmid of claim 4, wherein the pCCL backbone comprises the RNA-OUT repressor.

6. The transfer plasmid of claim 5, wherein the transfer plasmid shares at least 95% or 100% identity to SEQ ID NO: 39.

7. The transfer plasmid of any one of claims 1 to 7, wherein the promoter is an EFS promoter.

8. The transfer plasmid of claim 7, wherein the EFS promoter shares at least 95% identity with SEQ ID NO: 2.

9. The transfer plasmid of claim 8, wherein the EFS promoter is SEQ ID NO: 2.

10. The transfer plasmid of any one of claims 1 to 9, wherein the coding polynucleotide shares at least 95% identity with SEQ ID NO: 3.

11. The transfer plasmid of claim 10, wherein the coding polynucleotide shares at least 99% identity with SEQ ID NO: 3.

12. The transfer plasmid of claim 11, wherein the coding polynucleotide is SEQ ID NO: 3.

13. The transfer plasmid of any one of claims 1 to 12, wherein the expression cassette comprises a Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE).

14. The transfer plasmid of claim 13, wherein the WPRE is SEQ ID NO: 4.

15. The transfer plasmid of any one of claims 1 to 14, wherein the expression cassette shares at least 95% identity with SEQ ID NO: 1.

16. The transfer plasmid of any one of claims 1 to 15, wherein the expression cassette is flanked by a 5′ long terminal repeat (LTR) and a 3′ LTR.

17. The transfer plasmid of claim 16, wherein the 5′ LTR is SEQ ID NO: 34 and/or the 3′ LTR is SEQ ID NO: 28.

18. The transfer plasmid of any one of claims 1 to 17, wherein expression cassette shares at least 95% identity to SEQ ID NO: 1.

19. The transfer plasmid of any one of claims 1 to 17, wherein expression cassette is SEQ ID NO: 1.

20. A lentiviral particle produced by transfecting a host cell with the transfer plasmid of any one of claims 1 to 20.

21. An expression cassette comprising a coding polynucleotide encoding an isoform of T cell immune regulator 1 (TCIRG1) or a functional variant thereof, and an EFS promoter, wherein the polynucleotide is operatively linked to the EFS promoter.

22. The expression cassette of claim 1, wherein the coding polynucleotide shares at least 95% identity with SEQ ID NO: 3.

23. The expression cassette of claim 2, wherein the coding polynucleotide shares at least 99% identity with SEQ ID NO: 3.

24. The expression cassette of claim 3, wherein the coding polynucleotide is SEQ ID NO: 3.

25. The expression cassette of any one of claims 1 to 4, wherein the EFS promoter shares at least 95% identity with SEQ ID NO: 2.

26. The expression cassette of claim 25, wherein the EFS promoter is SEQ ID NO: 2.

27. The expression cassette of any one of claims 21 to 26, comprising a Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE).

28. The expression cassette of claim 27, wherein the WPRE is SEQ ID NO: 4.

29. The expression cassette of any one of claims 21 to 28, wherein the expression cassette shares at least 95% identity with SEQ ID NO: 1.

30. The expression cassette of claim 29, wherein the expression cassette is SEQ ID NO: 1.

31. A recombinant lentiviral genome, comprising in 5' to 3' order:

(a) a lentiviral 5' long terminal repeat (LTR);

(b) the expression cassette of any one of claims 21 to 30; and

(c) a lentiviral 3'LTR,

wherein the recombinant lentiviral genome is replication incompetent.

32. A transfer plasmid, comprising the recombinant lentiviral genome of claim 31.

33. A lentiviral particle, comprising the recombinant lentiviral genome of claim 31.

34. A pharmaceutical composition, comprising the lentiviral particle of claim 33.

35. A modified cell, comprising the expression cassette of any one of claims 21 to 30.

36. A modified cell, comprising the recombinant lentiviral genome of claim 31.

37. The modified cell of claim 36, wherein the modified cell lacks an endogenous functional *TCIRG1* gene.

38. The modified cell of claim 36 or 37, wherein the modified cell is derived from a subject having or suspected of having infantile malignant osteopetrosis (IMO).

39. The modified cell of any of one claims 36 to 38, wherein the modified cell expresses TCIRG1 or a functional variant thereof at a level similar to the level of expression of TCIRG1 observed in an osteoclast having a functional *TCIRG1* gene.

40. The modified cell of any of one claims 36 to 39, wherein the modified cell expresses TCIRG1 or a functional variant thereof at a level similar to the level of expression of TCIRG1 observed in an osteoclast derived from a subject not having or suspected of having IMO.

41. The modified cell of any of one claims 36 to 40, wherein the modified cell is a hematopoietic stem cell (HSC).

42. The modified cell of any of one claims 36 to 41, wherein the modified cell is a CD34+ progenitor cell.

43. The modified cell of any of on claim 41 or 42, wherein the modified cell is derived from a HSC isolated from a subject having or suspected of having IMO by apheresis, optionally after mobilization of HSCs by administration of G-CSF, plerifaxor, or a combination of G-CSF and plerifaxor.

44. The modified cell of any one of claims 35 to 43, wherein the modified cell is derived from a population of cells enriched for CD34+ cells by magnetic capture.

45. A pharmaceutical composition comprising the modified cell of any one of claims 35 to 44.

46. An *in vitro* method of modifying one or more cells of a subject having or suspected of having IMO, comprising:

(a) providing peripheral blood mononuclear cells (PBMCs) mobilized from the subject by administering to the subject a composition comprising G-CSF, plerifaxor, or a combination of G-CSF and plerifaxor;

(b) enriching the PBMCs for CD34+ cells by magnetic separation to generate a population of CD34-enriched cells; and

(c) contacting the CD34-enriched cells with a lentiviral particle comprising a recombinant lentiviral genome, comprising in 5' to 3' order:

(i) a lentiviral 5' long terminal repeat (LTR);

(ii) the expression cassette of any one of claims 21 to 30; and

(iii) a lentiviral 3'LTR,

wherein the recombinant lentiviral genome is replication incompetent.

47. A method of treating infantile malignant osteopetrosis (IMO) in a subject having or suspected of having IMO, comprising administering the modified cell of any one of claims 35 to 44 or the pharmaceutical composition of claim 45 to the subject.

48. The method of claim 47, wherein the method repopulates the HSC niche with modified cells expressing TCIRG1 or a functional variant thereof.

49. The method of claim 47 or 48, wherein the method repopulates the osteoclast niche with modified cells expressing TCIRG1 or a functional variant thereof.

50. The method of any one of claims 47 to 49, wherein the method treats, ameliorates, prevents, reduces, inhibits, or relieves IMO.

51. The method of any one of claims 47 to 50, wherein the method extends the mean overall survival of treated subjects by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more years.

52. The method of any one of claims 47 to 51, wherein the method prevents the death of the subject from IMO.

53. The method of any one of claims 47 to 52, wherein the subject is a human.

54. The method of any one of claims 47 to 53, wherein the subject exhibited symptoms of IMO before treatment.

55. The method of any one of claims 47 to 54, wherein the subject was identified as having reduced or non-detectable expression of TCIRG1 before treatment.

56. The method of any one of claims 47 to 55, wherein the subject was identified as having a mutated TCIRG1 gene.

57. The method of any one of claims 47 to 56, wherein the subject is an infant.

58. The method of any one of claims 47 to 57, wherein the method comprises autologous treatment.

59. The method of any one of claims 47 to 58, wherein administration is performed via a intravenous infusion.

60. A recombinant lentiviral genome for use in the preparation of a medicament for treating or preventing infantile malignant osteopetrosis (IMO), wherein the lentiviral genome comprises in 5' to 3' order:

(i) a lentiviral 5' long terminal repeat (LTR),

(ii) the expression cassette of any one of claims 21 to 30, and

(iii) a lentiviral 3'LTR; and

wherein the recombinant lentiviral genome is replication incompetent.

61. A lentiviral particle for use in the preparation of a medicament for treating or preventing infantile malignant osteopetrosis (IMO), comprising a recombinant lentiviral genome,

wherein the lentiviral genome comprises in 5' to 3' order:

(i) a lentiviral 5' long terminal repeat (LTR),

(ii) the expression cassette of any one of claims 21 to 50, and

(iii) a lentiviral 3'LTR; and

wherein the recombinant lentiviral genome is replication incompetent.

62. A transfer plasmid comprising the expression cassette of any one of claims 24-30.

63. The transfer plasmid of claim 62, further comprising an RNA-OUT sequence.

64. The transfer plasmid of claim 63, wherein the RNA-OUT sequence is SEQ ID NO: 22.

65. The transfer plasmid of claim 62 or claim 63, wherein the RNA-OUT sequence is configured such that the transfer plasmid is capable of stable propagation in a packaging cell line.

66. A method of producing a lentiviral particle, comprising transforming a bacterial cell with the transfer plasmid of any one of claims 1-19 or claims 62-65, such that the transfer plasmid is replicated, isolating the replicated transfer plasmid, and transducing a packaging cell line with the replicated transfer plasmid, and optionally one or more additional plasmids, thereby producing the lentiviral particle.

67. A method of producing a lentiviral particle, comprising transfecting a packaging cell line with the transfer plasmid of any one of claims 1-19 or claims 62-65, and optionally one or more additional plasmids, and culturing said packaging cell line.

68. The method of claim 67, wherein the transfer plasmid is stably propagated.

69. The method of claim 68, wherein the transfer plasmid is stably propagated in a bacterial host at 30-37°C using shake flasks or fermentation for at least 1, 2, 3, 4, 5, 6, or 7 days.

70. A lentiviral particle produced according to a method of any one of claims 66-69.

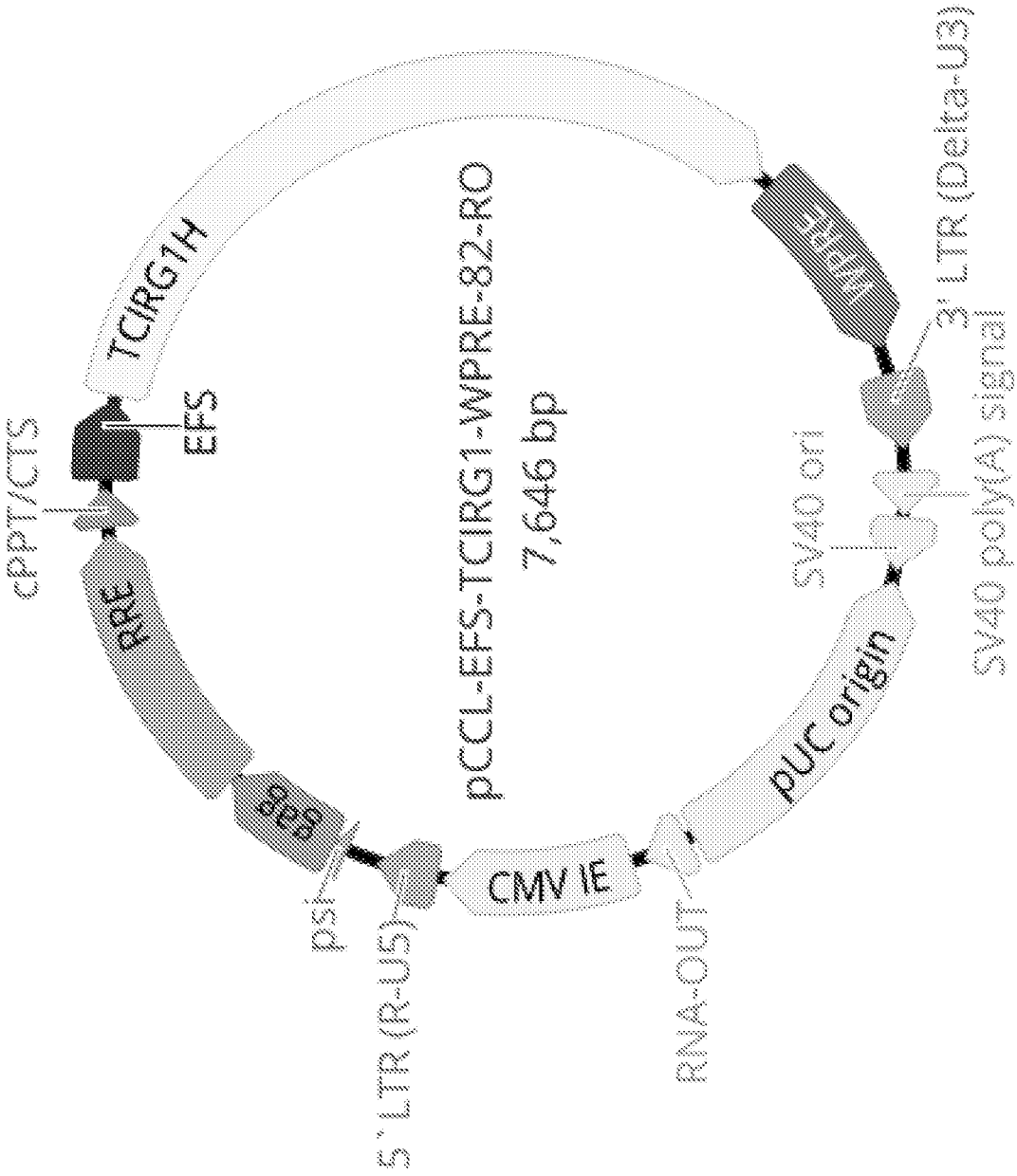

pCCL-EFS-TCIRG1-WPRE-82-RO
7,646 bp
TCIRG1H
EFS
cPPT/CTS
RRE
gag
psi
5' LTR (R-U5)
CMV IE
RNA-OUT
pUC origin
SV40 ori
SV40 poly(A) signal
3' LTR (Delta-U3)

FIG. 1

GGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCG

GGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCG

CAACGGGTTTGCCGCCAGAACACAGGTGTCGTGACGCGGGATCCGCCACC**ATGGGCTCCATGTTTCGGAGCGAGGAGGTGGCCCTGGTCCAGCTCTTTCTGCC**

CACAGCGGCTGCCTACACCTGCGTGAGTCGGGCTGGGCGAGCTGGGGCTCGTGGAGTTCAGAGACCTCAACGCCTCGGTGAGCGCCTTCCAGAGACGCTTTGTG

GTTTGATGTTCGGCGCTGTGAGGAGCTGGAAGACCTTCACCTTCCTGCAGGAGGATGCGGCGGGCTGGCGTCCTGCCCCGCCAAAGGGAGGAGGCTGCC

CGGCACCCCCCACCCCGGGACCTGCTGCGCATCCAGGAGGAGACGGAGCGCCTGGCCCAGGAGCTGCGGGATGTGCGGGGCGAACCAGCAGGCCCTGCGGGCCCA

GCTGCACCAGCTGCAGCTCCACGCCGCCGTGCTACGCCAGGGCCATGAACCTCAGCTGGCAGCCGCCCACACAGATGGGGCCTCAGAGAGGACGCCCCTGCTC

CAGGCCCCCGGGGGGCCGCACCAGGACCTGAGGTCCAAACTTTGTGGCAGGTGCCGTGGAGCCCCACAAGGCCCCTGCCCTAGAGCGCCTGCTCTGGAGGGCCT

GCAGAGGCTTCCTCATTGCCAGCTTCAGGGAGCTGGAGCAGCCGCTGGAGCACCCCGTGACGCAGCCACGTGGATGACCTTCCTCATCTCCTACTG

GGGTGAGCAGATCGGACAGAAGATCACGGACTGCTTCCACTGCCACGTCTTCCGTTTCTGCAGCAGGAGGAGGCCCGCCTCGGGGCCCTGCAG

CAGCTGCAACAGCAGGAGCCAGGAGCTGCAGGAGGTTCCTGAGCCAGGTGCTAGGCCGGGTGCTGCAGCTGCTGCCGCCAGGGC

AGGTGCAGGTCCACACAAGATGAAGGCCGTGTACCTGGCCCTGAACCACGTGAGCACCACGCACAAGTGCCTCATTGCCGAGGCCTGGTGTGCTCTGTGCG

AGACCTGCCCGCCCTGCAGGAGGCCCTGCGGGACAGCTCGATGGAGGAGTGAGTGCCGTGGCTCACCGCATCCCCTGCCGGGACATGCCCCCCCACACTC

ATCCGCACCACCAACCGCTTCACGGCCAGGCATCGTGGATGCCTACGGCGTGGGCCGCTACGCCAGGAGGTTCAACCCCGCTCCCTACACCATCATCACCT

TCCCCTTCCTGTTTGCTGTGATGTTCGGGGATGTGCTCATGTTCCTCTTCGCCCTGGCCATGGTCCTTGCGGAGAACCGGCTGTGAA

GGCCGCGCAGAACGAGATCTGGCAGACTTTCTTCAGGGGCCGCTACCTGCTCCTGTTATGGGCCTGTTTCTCCATCTACACCGGCTTCATCTACAACGAGTGC

TTCAGTCGCGCCACCAGCATCTTCCCCTCGGGCTGGAGTGTGGCCGCCAACCAGTCTGGCTGGAGTGATGCATTCCTGGCCCAGCACACGATGCTTTA

CCCTGGACCCCAACGTCACCGGTGTCTTCCTGGGACCCTACCCCTTTGGCATCGATCCTATTTGGAGCCTGGCTGCCAACCACTTGAGCTTCCTCAACTCCTTT

CAAGATGAAGATGTCCGTCATCCTGGGCGTCGTGCACATGGCCTTTTGGGGTGGTCCTCGGAGTCTTCAACCACGTGCACTTTGGCCAGAGGCACCGGCTGCTG

CTGGAGACGCTGCCGGAGCTCACCTTCCTGCTGGGACTCTTCGGTTACCTCGTTCCTAGTCATCTACAAGTGGCTGTGTGTCTGGGCTGCCAGGGCCGCCT

CGGCCCCCAGCATCCTCATCCACTTCATCAACATGTTCCTCTTCTCCCACAGCCCCAGCAACAGGCTGCTCTACCCCCGGCAGGAGGTGGTCCAGGCCACGCT

GGTGGTCCTGGCCTTGGCCATGGTGCTCCATCCTGCTGTTGGCACACCCCTGCACCTGCTGCACCGCCGCTGCGGAGGAGGCCCGCTGACCGA

CAGGAGGAAAACAAGGCCGGGTTGCTGGACCTGCCTGAACGCATCTGTGAATGGCTCCGATGAGGAAAAGGCAGGGGGCCTGGATGATGAAGAGGAGG

CCGAGCTCGTCCCCTCCGAGGTGCTCATGCACCAGGCCATCCACACCATCGAGTTCTGCCTGGGCTGCGTCTCCAACACCGCCTCCTACCTGCGCCTGTGGGC

CCTGAGCCTGGCCCACGCCCAGCTGTCCGAGGTTCTGTGGGCCATAGGCCTGGGCCTGGGAGGTGGCGTGGCGGCTGTGGTGCTG

GTCCCCATCTTTGCCGCCTTTGCCGTGATGACCGTTCTGGTGATGGAGGGACTCTCAGCCTTCCTGCACGCCCTGCGGCTGCACTGGGTGGAAT

TCCAGAACAAGTTCTACTCAGGCACGGGCTACAAGCTGAGTCCCTTCACCTTCGCTGCCACAGATGACTAGTAAGTCGACGGATCCCCCGGGCTGCAGGA**ATT**

CGAGCATCTTACCGCCATTTATACCCATATTTGTTCTGTTTTTCTTGATTTGGGTATACATTTAAATGTTAATAAAACAAAATGGTGGGGCAATCATTTACAT

TTTTAGGGATATGTAATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAACTTTCCCGTTATTTACGCTCTGTTCCTGTTAATCAACCTCTGG

ATTACAAAATTTGTGAAAGATTGACTGATATTCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATATGCTGCTTTATAGCCTCTGTATCATGCTATTGCTTC

CCGTACGGCTTTCGTTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTCAACGTGGCGTGGTGTGCTCTGTG

TTTGCTGACGCAACCCCCACTGGCTGGGGCATTGCCACCACCTGTCAACTCCTTTCCTCTTCCCCTCCCGATCGCCACGGCAGAACTCATCG

CCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTAGGTTGCTGGGCACTGATAATTCCGTGGTGTTGTCGGGGAAGCTGACGTCCTTTCG

FIG. 2

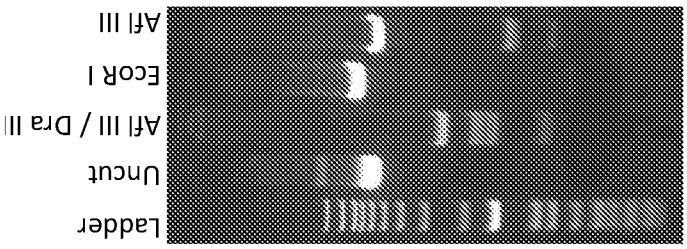


FIG. 3B

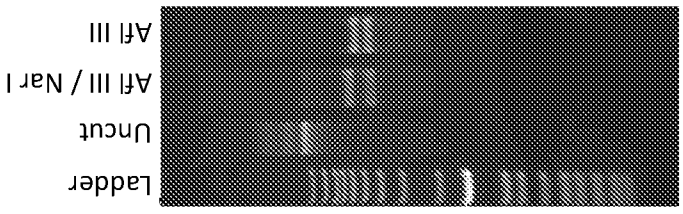
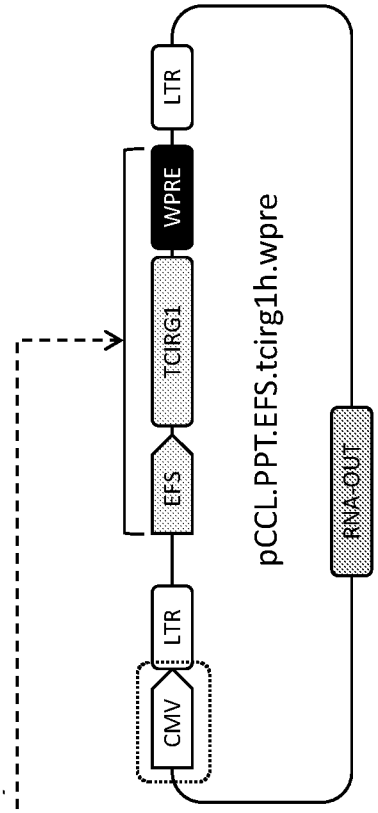


FIG. 3A



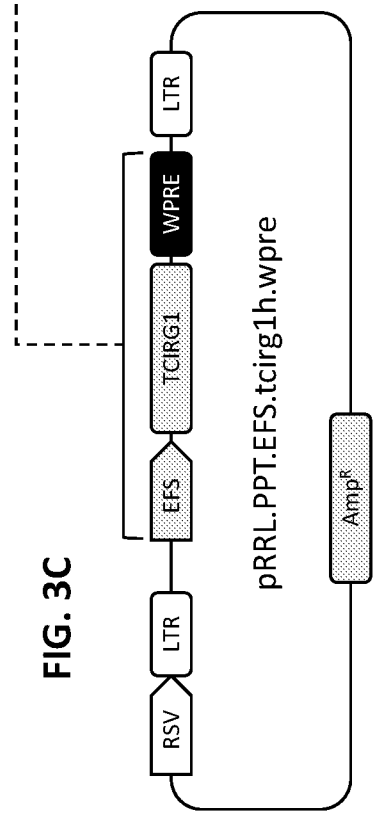


FIG. 3C

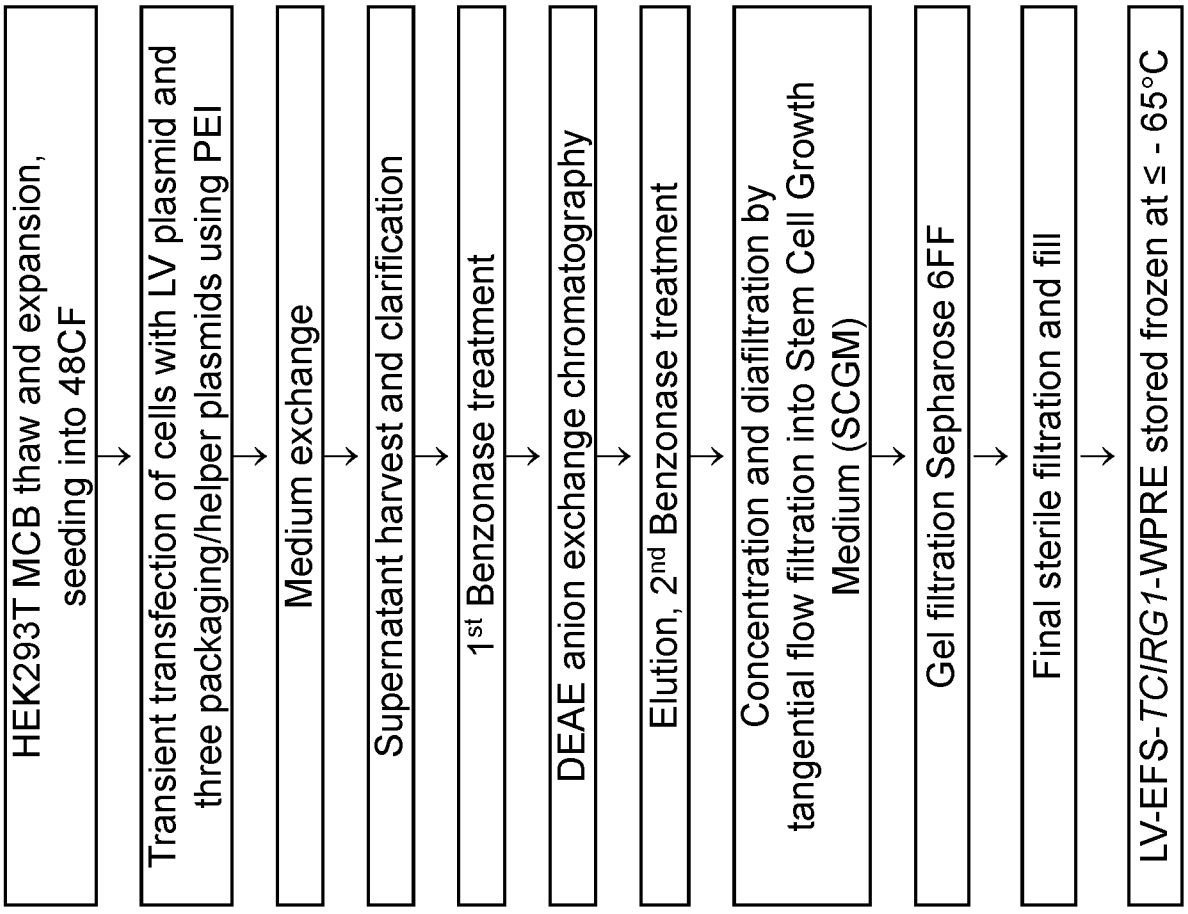
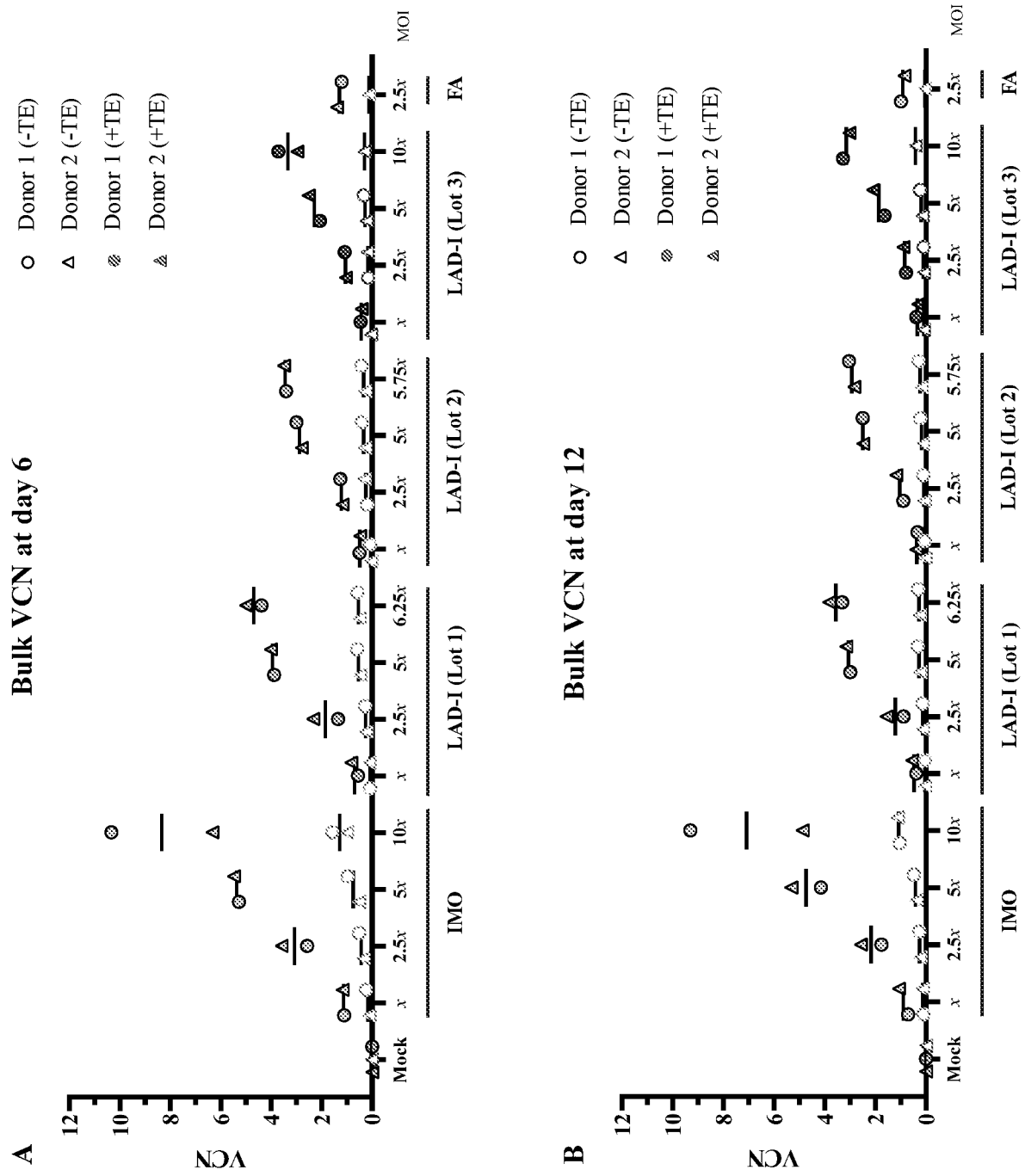

HEK293T MCB thaw and expansion, seeding into 48CF
Transient transfection of cells with LV plasmid and three packaging/helper plasmids using PEI
Medium exchange
Supernatant harvest and clarification
1st Benzonase treatment
DEAE anion exchange chromatography
Elution, 2nd Benzonase treatment
Concentration and diafiltration by tangential flow filtration into Stem Cell Growth Medium (SCGM)
Gel filtration Sepharose 6FF
Final sterile filtration and fill
LV-EFS-TCIRG1-WPRE stored frozen at ≤ - 65°C

FIG. 4


A
Bulk VCN at day 6
Donor 1 (-TE)
Donor 2 (-TE)
Donor 1 (+TE)
Donor 2 (+TE)
VCN
0
2
4
6
8
10
12
Mock
x
2.5x
5x
10x
IMO
x
2.5x
5x
6.25x
LAD-I (Lot 1)
x
2.5x
5x
5.75x
LAD-I (Lot 2)
x
2.5x
5x
10x
LAD-I (Lot 3)
2.5x
FA
MOI
B
Bulk VCN at day 12
Donor 1 (-TE)
Donor 2 (-TE)
Donor 1 (+TE)
Donor 2 (+TE)
VCN
0
2
4
6
8
10
12
Mock
x
2.5x
5x
10x
IMO
x
2.5x
5x
6.25x
LAD-I (Lot 1)
x
2.5x
5x
5.75x
LAD-I (Lot 2)
x
2.5x
5x
10x
LAD-I (Lot 3)
2.5x
FA
MOI

FIGS. 5A-5B

SEQUENCE LISTING

<110> Rocket Pharmaceuticals, LTD
Beard, Brian
Ricks, David
Prabhakar, Raj

<120> GENE THERAPY VECTORS FOR INFANTILE MALIGNANT OSTEOPETROSIS

<130> ROPA-003/01WO 329592-2098

<150> US 62/852,216
<151> 2019-05-23

<160> 39

<170> PatentIn version 3.5

<210> 1
<211> 3384
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - Expression cassette

<400> 1
ggctccggtg cccgtcagtg ggcagagcgc acatcgccca cagtccccga gaagttgggg 60

ggaggggtcg gcaattgaac cggtgcctag agaaggtggc gcggggtaaa ctgggaaagt 120

gatgtcgtgt actggctccg cctttttccc gagggtgggg gagaaccgta tataagtgca 180

gtagtcgccg tgaacgttct ttttcgcaac gggtttgccg ccagaacaca ggtgtcgtga 240

cgcgggatcc gccaccatgg gctccatgtt tcggagcgag gaggtggccc tggtccagct 300

ctttctgccc acagcggctg cctacacctg cgtgagtcgg ctgggcgagc tgggcctcgt 360

ggagttcaga gacctcaacg cctcggtgag cgccttccag agacgctttg tggttgatgt 420

tcggcgctgt gaggagctgg agaagacctt caccttcctg caggaggagg tgcggcgggc 480

tgggctggtc ctgcccccgc caaaggggag gctgccggca cccccacccc gggacctgct 540

gcgcatccag gaggagacgg agcgcctggc ccaggagctg cgggatgtgc ggggcaacca 600

gcaggccctg cgggcccagc tgcaccagct gcagctccac gccgccgtgc tacgccaggg 660

ccatgaacct cagctggcag ccgcccacac agatggggcc tcagagagga cgcccctgct 720

ccaggccccc ggggggccgc accaggacct gagggtcaac tttgtggcag gtgccgtgga 780

gccccacaag gcccctgccc tagagcgcct gctctggagg gcctgcagag gcttcctcat 840

tgccagcttc agggagctgg agcagccgct ggagcacccc gtgacgggcg agccagccac 900

gtggatgacc ttcctcatct cctactgggg tgagcagatc ggacagaaga tccgcaagat 960

cacggactgc ttccactgcc acgtcttccc gtttctgcag caggaggagg cccgcctcgg 1020

ggccctgcag cagctgcaac agcagagcca ggagctgcag gaggtcctcg gggagacaga 1080

gcggttcctg agccaggtgc taggccgggt gctgcagctg ctgccgccag ggcaggtgca 1140

ggtccacaag atgaaggccg tgtacctggc cctgaaccag tgcagcgtga gcaccacgca 1200

caagtgcctc attgccgagg cctggtgctc tgtgcgagac ctgcccgccc tgcaggaggc 1260

cctgcgggac agctcgatgg aggagggagt gagtgccgtg gctcaccgca tcccctgccg 1320

ggacatgccc cccacactca tccgcaccaa ccgcttcacg gccagcttcc aggggcatcgt 1380

ggatgcctac ggcgtgggcc gctaccagga ggtcaacccc gctccctaca ccatcatcac 1440

cttccccttc ctgtttgctg tgatgttcgg ggatgtgggc cacgggctgc tcatgttcct 1500

cttcgccctg gccatggtcc ttgcggagaa ccgaccggct gtgaaggccg cgcagaacga 1560

gatctggcag actttcttca ggggccgcta cctgctcctg cttatgggcc tgttctccat 1620

ctacaccggc ttcatctaca acgagtgctt cagtcgcgcc accagcatct tcccctcggg 1680

ctggagtgtg gccgccatgg ccaaccagtc tggctggagt gatgcattcc tggcccagca 1740

cacgatgctt accctggacc ccaacgtcac cggtgtcttc ctgggaccct acccctttgg 1800

catcgatcct atttggagcc tggctgccaa ccacttgagc ttcctcaact ccttcaagat 1860

gaagatgtcc gtcatcctgg gcgtcgtgca catggccttt ggggtggtcc tcggagtctt 1920

caaccacgtg cactttggcc agaggcaccg gctgctgctg gagacgctgc cggagctcac 1980

cttcctgctg ggactcttcg gttacctcgt gttcctagtc atctacaagt ggctgtgtgt 2040

ctgggctgcc agggccgcct cggcccccag catcctcatc cacttcatca acatgttcct 2100

cttctcccac agccccagca acaggctgct ctacccccgg caggaggtgg tccaggccac 2160

gctggtggtc ctggccttgg ccatggtgcc catcctgctg cttggcacac ccctgcacct 2220

gctgcaccgc caccgccgcc gcctgcggag gaggcccgct gaccgacagg aggaaaacaa 2280

ggccgggttg ctggacctgc ctgacgcatc tgtgaatggc tggagctccg atgaggaaaa 2340

ggcaggggggc ctggatgatg aagaggaggc cgagctcgtc ccctccgagg tgctcatgca 2400

ccaggccatc cacaccatcg agttctgcct gggctgcgtc tccaacaccg cctcctacct 2460

gcgcctgtgg gccctgagcc tggcccacgc ccagctgtcc gaggttctgt gggccatggt 2520

gatgcgcata ggcctgggcc tgggccggga ggtgggcgtg gcggctgtgg tgctggtccc 2580

catctttgcc gcctttgccg tgatgaccgt ggctatcctg ctggtgatgg agggactctc 2640

agccttcctg cacgccctgc ggctgcactg ggtggaattc cagaacaagt tctactcagg 2700

cacgggctac aagctgagtc ccttcacctt cgctgccaca gatgactagt aagtcgacgg 2760

atcccccggg ctgcaggaat tcgagcatct taccgccatt tatacccata tttgttctgt 2820

ttttcttgat ttgggtatac atttaaatgt taataaaaca aaatggtggg gcaatcattt 2880

acatttttag ggatatgtaa ttactagttc aggtgtattg ccacaagaca aacatgttaa 2940

gaaactttcc cgttatttac gctctgttcc tgttaatcaa cctctggatt acaaaatttg 3000

tgaaagattg actgatattc ttaactatgt tgctcctttt acgctgtgtg gatatgctgc 3060

tttaatgcct ctgtatcatg ctattgcttc ccgtacggct ttcgttttct cctccttgta 3120

taaatcctgg ttgctgtctc tttatgagga gttgtggccc gttgtccgtc aacgtggcgt 3180

ggtgtgctct gtgtttgctg acgcaacccc cactggctgg ggcattgcca ccacctgtca 3240

actcctttct gggactttcg ctttcccccct cccgatcgcc acggcagaac tcatcgccgc 3300

ctgccttgcc cgctgctgga caggggctag gttgctgggc actgataatt ccgtggtgtt 3360

gtcggggaag ctgacgtcct ttcg 3384

<210> 2
<211> 243
<212> DNA
<213> Homo sapiens

<400> 2
ggctccggtg cccgtcagtg ggcagagcgc acatcgccca cagtccccga gaagttgggg 60

ggaggggtcg gcaattgaac cggtgcctag agaaggtggc gcggggtaaa ctgggaaagt 120

gatgtcgtgt actggctccg cctttttccc gagggtgggg gagaaccgta tataagtgca 180

gtagtcgccg tgaacgttct ttttcgcaac gggtttgccg ccagaacaca ggtgtcgtga 240

cgc 243

<210> 3
<211> 2493
<212> DNA
<213> Homo sapiens

<400> 3
atgggctcca tgtttcggag cgaggaggtg gccctggtcc agctctttct gcccacagcg 60

gctgcctaca cctgcgtgag tcggctgggc gagctgggcc tcgtggagtt cagagacctc 120

aacgcctcgg tgagcgcctt ccagagacgc tttgtggttg atgttcggcg ctgtgaggag 180

ctggagaaga ccttcacctt cctgcaggag gaggtgcggc gggctgggct ggtcctgccc 240

ccgccaaagg ggaggctgcc ggcaccccca ccccgggacc tgctgcgcat ccaggaggag 300

acggagcgcc tggcccagga gctgcgggat gtgcggggca accagcaggc cctgcgggcc 360

cagctgcacc agctgcagct ccacgccgcc gtgctacgcc agggccatga acctcagctg 420

gcagccgccc acacagatgg ggcctcagag aggacgcccc tgctccaggc ccccgggggg 480

ccgcaccagg acctgagggt caactttgtg gcaggtgccg tggagcccca caaggcccct 540

gccctagagc gcctgctctg gagggcctgc agaggcttcc tcattgccag cttcagggag 600

ctggagcagc cgctggagca ccccgtgacg ggcgagccag ccacgtggat gaccttcctc 660

atctcctact ggggtgagca gatcggacag aagatccgca agatcacgga ctgcttccac 720

tgccacgtct tcccgtttct gcagcaggag gaggcccgcc tcggggccct gcagcagctg 780

caacagcaga gccaggagct gcaggaggtc ctcggggaga cagagcggtt cctgagccag 840

gtgctaggcc gggtgctgca gctgctgccg ccagggcagg tgcaggtcca caagatgaag 900

gccgtgtacc tggccctgaa ccagtgcagc gtgagcacca cgcacaagtg cctcattgcc 960

gaggcctggt gctctgtgcg agacctgccc gccctgcagg aggccctgcg ggacagctcg 1020

atggaggagg gagtgagtgc cgtggctcac cgcatcccct gccgggacat gccccccaca 1080

ctcatccgca ccaaccgctt cacggccagc ttccagggca tcgtggatgc ctacggcgtg 1140

ggccgctacc aggaggtcaa ccccgctccc tacaccatca tcaccttccc cttcctgttt 1200

gctgtgatgt tcggggatgt gggccacggg ctgctcatgt tcctcttcgc cctggccatg 1260

gtccttgcgg agaaccgacc ggctgtgaag gccgcgcaga acgagatctg gcagactttc 1320

ttcaggggcc gctacctgct cctgcttatg ggcctgttct ccatctacac cggcttcatc 1380

tacaacgagt gcttcagtcg cgccaccagc atcttcccct cgggctggag tgtggccgcc 1440

atggccaacc agtctggctg gagtgatgca ttcctggccc agcacacgat gcttaccctg 1500

gaccccaacg tcaccggtgt cttcctggga ccctacccct ttggcatcga tcctatttgg 1560

agcctggctg ccaaccactt gagcttcctc aactccttca agatgaagat gtccgtcatc 1620

ctgggcgtcg tgcacatggc ctttggggtg gtcctcggag tcttcaacca cgtgcacttt 1680

ggccagaggc accggctgct gctggagacg ctgccggagc tcaccttcct gctgggactc 1740

ttcggttacc tcgtgttcct agtcatctac aagtggctgt gtgtctgggc tgccagggcc 1800

gcctcggccc ccagcatcct catccacttc atcaacatgt tcctcttctc ccacagcccc 1860

agcaacaggc tgctctaccc ccggcaggag gtggtccagg ccacgctggt ggtcctggcc 1920

ttggccatgg tgcccatcct gctgcttggc acacccctgc acctgctgca ccgccaccgc 1980

cgccgcctgc ggaggaggcc cgctgaccga caggaggaaa acaaggccgg gttgctggac 2040

ctgcctgacg catctgtgaa tggctggagc tccgatgagg aaaaggcagg gggcctggat 2100

gatgaagagg aggccgagct cgtcccctcc gaggtgctca tgcaccaggc catccacacc 2160

atcgagttct gcctgggctg cgtctccaac accgcctcct acctgcgcct gtgggccctg 2220

agcctggccc acgcccagct gtccgaggtt ctgtgggcca tggtgatgcg cataggcctg 2280

ggcctgggcc gggaggtggg cgtggcggct gtggtgctgg tccccatctt tgccgccttt 2340

gccgtgatga ccgtggctat cctgctggtg atggagggac tctcagcctt cctgcacgcc 2400

ctgcggctgc actgggtgga attccagaac aagttctact caggcacggg ctacaagctg 2460

agtcccttca ccttcgctgc cacagatgac tag 2493

<210> 4
<211> 606
<212> DNA
<213> Woodchuck hepatitis virus

<400> 4
attcgagcat cttaccgcca tttataccca tatttgttct gtttttcttg atttgggtat 60

acatttaaat gttaataaaa caaaatggtg gggcaatcat ttacattttt agggatatgt 120

aattactagt tcaggtgtat tgccacaaga caaacatgtt aagaaacttt cccgttattt 180

acgctctgtt cctgttaatc aacctctgga ttacaaaatt tgtgaaagat tgactgatat 240

tcttaactat gttgctcctt ttacgctgtg tggatatgct gctttaatgc ctctgtatca 300

tgctattgct tcccgtacgg ctttcgtttt ctcctccttg tataaatcct ggttgctgtc 360

tctttatgag gagttgtggc ccgttgtccg tcaacgtggc gtggtgtgct ctgtgtttgc 420

tgacgcaacc cccactggct ggggcattgc caccacctgt caactccttt ctgggacttt 480

cgctttcccc ctcccgatcg ccacggcaga actcatcgcc gcctgccttg cccgctgctg 540

gacaggggct aggttgctgg gcactgataa ttccgtggtg ttgtcgggga agctgacgtc 600

<210> 5
<211> 830
<212> PRT
<213> Homo sapiens

<400> 5

Met Gly Ser Met Phe Arg Ser Glu Glu Val Ala Leu Val Gln Leu Phe
1 5 10 15

Leu Pro Thr Ala Ala Ala Tyr Thr Cys Val Ser Arg Leu Gly Glu Leu
20 25 30

Gly Leu Val Glu Phe Arg Asp Leu Asn Ala Ser Val Ser Ala Phe Gln
35 40 45

Arg Arg Phe Val Val Asp Val Arg Arg Cys Glu Glu Leu Glu Lys Thr
50 55 60

Phe Thr Phe Leu Gln Glu Glu Val Arg Arg Ala Gly Leu Val Leu Pro
65 70 75 80

Pro Pro Lys Gly Arg Leu Pro Ala Pro Pro Pro Arg Asp Leu Leu Arg
85 90 95

Ile Gln Glu Glu Thr Glu Arg Leu Ala Gln Glu Leu Arg Asp Val Arg
100 105 110

Gly Asn Gln Gln Ala Leu Arg Ala Gln Leu His Gln Leu Gln Leu His
115 120 125

Ala Ala Val Leu Arg Gln Gly His Glu Pro Gln Leu Ala Ala Ala His
130 135 140

Thr Asp Gly Ala Ser Glu Arg Thr Pro Leu Leu Gln Ala Pro Gly Gly
145 150 155 160

Pro His Gln Asp Leu Arg Val Asn Phe Val Ala Gly Ala Val Glu Pro
165 170 175

His Lys Ala Pro Ala Leu Glu Arg Leu Leu Trp Arg Ala Cys Arg Gly
180 185 190

Phe Leu Ile Ala Ser Phe Arg Glu Leu Glu Gln Pro Leu Glu His Pro
195 200 205

Val Thr Gly Glu Pro Ala Thr Trp Met Thr Phe Leu Ile Ser Tyr Trp
210 215 220

Gly Glu Gln Ile Gly Gln Lys Ile Arg Lys Ile Thr Asp Cys Phe His
225 230 235 240

```
Cys His Val Phe Pro Phe Leu Gln Gln Glu Glu Ala Arg Leu Gly Ala
                245                 250                 255

Leu Gln Gln Leu Gln Gln Gln Ser Gln Glu Leu Gln Glu Val Leu Gly
            260                 265                 270

Glu Thr Glu Arg Phe Leu Ser Gln Val Leu Gly Arg Val Leu Gln Leu
        275                 280                 285

Leu Pro Pro Gly Gln Val Gln Val His Lys Met Lys Ala Val Tyr Leu
    290                 295                 300

Ala Leu Asn Gln Cys Ser Val Ser Thr Thr His Lys Cys Leu Ile Ala
305                 310                 315                 320

Glu Ala Trp Cys Ser Val Arg Asp Leu Pro Ala Leu Gln Glu Ala Leu
                325                 330                 335

Arg Asp Ser Ser Met Glu Glu Gly Val Ser Ala Val Ala His Arg Ile
            340                 345                 350

Pro Cys Arg Asp Met Pro Pro Thr Leu Ile Arg Thr Asn Arg Phe Thr
        355                 360                 365

Ala Ser Phe Gln Gly Ile Val Asp Ala Tyr Gly Val Gly Arg Tyr Gln
    370                 375                 380

Glu Val Asn Pro Ala Pro Tyr Thr Ile Ile Thr Phe Pro Phe Leu Phe
385                 390                 395                 400

Ala Val Met Phe Gly Asp Val Gly His Gly Leu Leu Met Phe Leu Phe
                405                 410                 415

Ala Leu Ala Met Val Leu Ala Glu Asn Arg Pro Ala Val Lys Ala Ala
            420                 425                 430

Gln Asn Glu Ile Trp Gln Thr Phe Phe Arg Gly Arg Tyr Leu Leu Leu
        435                 440                 445

Leu Met Gly Leu Phe Ser Ile Tyr Thr Gly Phe Ile Tyr Asn Glu Cys
    450                 455                 460

Phe Ser Arg Ala Thr Ser Ile Phe Pro Ser Gly Trp Ser Val Ala Ala
465                 470                 475                 480

Met Ala Asn Gln Ser Gly Trp Ser Asp Ala Phe Leu Ala Gln His Thr
                485                 490                 495

Met Leu Thr Leu Asp Pro Asn Val Thr Gly Val Phe Leu Gly Pro Tyr
            500                 505                 510
```

```
Pro Phe Gly Ile Asp Pro Ile Trp Ser Leu Ala Ala Asn His Leu Ser
        515                 520                 525

Phe Leu Asn Ser Phe Lys Met Lys Met Ser Val Ile Leu Gly Val Val
    530                 535                 540

His Met Ala Phe Gly Val Val Leu Gly Val Phe Asn His Val His Phe
545                 550                 555                 560

Gly Gln Arg His Arg Leu Leu Leu Glu Thr Leu Pro Glu Leu Thr Phe
                565                 570                 575

Leu Leu Gly Leu Phe Gly Tyr Leu Val Phe Leu Val Ile Tyr Lys Trp
            580                 585                 590

Leu Cys Val Trp Ala Ala Arg Ala Ala Ser Ala Pro Ser Ile Leu Ile
        595                 600                 605

His Phe Ile Asn Met Phe Leu Phe Ser His Ser Pro Ser Asn Arg Leu
    610                 615                 620

Leu Tyr Pro Arg Gln Glu Val Val Gln Ala Thr Leu Val Val Leu Ala
625                 630                 635                 640

Leu Ala Met Val Pro Ile Leu Leu Leu Gly Thr Pro Leu His Leu Leu
                645                 650                 655

His Arg His Arg Arg Arg Leu Arg Arg Arg Pro Ala Asp Arg Gln Glu
            660                 665                 670

Glu Asn Lys Ala Gly Leu Leu Asp Leu Pro Asp Ala Ser Val Asn Gly
        675                 680                 685

Trp Ser Ser Asp Glu Glu Lys Ala Gly Gly Leu Asp Asp Glu Glu Glu
    690                 695                 700

Ala Glu Leu Val Pro Ser Glu Val Leu Met His Gln Ala Ile His Thr
705                 710                 715                 720

Ile Glu Phe Cys Leu Gly Cys Val Ser Asn Thr Ala Ser Tyr Leu Arg
                725                 730                 735

Leu Trp Ala Leu Ser Leu Ala His Ala Gln Leu Ser Glu Val Leu Trp
            740                 745                 750

Ala Met Val Met Arg Ile Gly Leu Gly Leu Gly Arg Glu Val Gly Val
        755                 760                 765

Ala Ala Val Val Leu Val Pro Ile Phe Ala Ala Phe Ala Val Met Thr
    770                 775                 780
```

```
Val Ala Ile Leu Leu Val Met Glu Gly Leu Ser Ala Phe Leu His Ala
785                 790                 795                 800


Leu Arg Leu His Trp Val Glu Phe Gln Asn Lys Phe Tyr Ser Gly Thr
                805                 810                 815


Gly Tyr Lys Leu Ser Pro Phe Thr Phe Ala Ala Thr Asp Asp
            820                 825                 830
```

<210> 6
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus Kozak sequence

<400> 6
gccgccacca tgg 13

<210> 7
<211> 387
<212> DNA
<213> Mus musculus

<400> 7
tggctaataa aggaaattta ttttcattgc aatagtgtgt tggaattttt tgtgtctctc 60

actcggaagg acatatggga gggcaaatca tttaaaacat cagaatgagt atttggttta 120

gagtttggca acatatgccc atatgctggc tgccatgaac aaaggttggc tataaagagg 180

tcatcagtat atgaaacagc cccctgctgt ccattcctta ttccatagaa aagccttgac 240

ttgaggttag attttttttta tattttgttt tgtgttattt ttttctttaa catccctaaa 300

attttcctta catgttttac tagccagatt tttcctcctc tcctgactac tcccagtcat 360

agctgtccct cttctcttat ggagatc 387

<210> 8

<400> 8
000

<210> 9

<400> 9
000

<210> 10

<400> 10
000

<210> 11

<400> 11
000

<210> 12

<400> 12
000

<210> 13

<400> 13
000

<210> 14
<211> 13
<212> RNA
<213> Artificial Sequence

<220>
<223> Consensus Kozak sequence

<400> 14
gccgccrcca ugg 13

<210> 15
<211> 8
<212> RNA
<213> Artificial Sequence

<220>
<223> Kozak sequence

<220>
<221> misc_feature
<222> (2)..(3)
<223> n is A, C, T, G or U

<220>
<221> misc_feature
<222> (4)..(4)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (8)..(8)
<223> n is A, C, T, G or U

<400> 15
agnnaugn 8

<210> 16
<211> 7
<212> RNA
<213> Artificial Sequence

<220>
<223> Kozak sequence

<220>
<221> misc_feature
<222> (2)..(3)
<223> n is A, C, T, G or U

<400> 16
annaugg 7

<210> 17
<211> 7
<212> RNA
<213> Artificial Sequence

<220>
<223> Kozak sequence

<400> 17
accaugg 7

<210> 18
<211> 10
<212> RNA
<213> Artificial Sequence

<220>
<223> n is A, C, T, G or U

<400> 18
gacaccaugg 10

<210> 19
<211> 235
<212> DNA
<213> Bos taurus

<400> 19
tcgactgtgc cttctagttg ccagccatct gttgtttgcc cctcccccgt gccttccttg 60

accctggaag gtgccactcc cactgtcctt tcctaataaa atgaggaaat tgcatcgcat 120

tgtctgagta ggtgtcattc tattctgggg ggtggggtgg ggcaggacag caaggggggag 180

gattgggagg acaatagcag gcatgctggg gatgcggtgg gctctatggc ttctg 235

<210> 20
<211> 222
<212> DNA
<213> simian virus 40

<400> 20
cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa 60

aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 120

ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggagatgt 180

gggaggtttt ttaaagcaag taaaacctct acaaatgtgg ta 222

<210> 21
<211> 202
<212> DNA
<213> Homo sapiens

<400> 21
ctgcccgggt ggcatccctg tgacccctcc ccagtgcctc tcctggccct ggaagttgcc 60

actccagtgc ccaccagcct tgtcctaata aaattaagtt gcatcatttt gtctgactag 120

gtgtccttct ataatattat ggggtggagg ggggtggtat ggagcaaggg gcccaagttg 180

ggaagaaacc tgtagggcct gc 202

<210> 22
<211> 141
<212> DNA
<213> Escherichia coli

<400> 22
gtagaattgg taaagagagt cgtgtaaaat atcgagttcg cacatcttgt tgtctgatta 60

ttgatttttg gcgaaaccat ttgatcatat gacaagatgt gtatctacct taacttaatg 120

attttgataa aaatcattag g 141

<210> 23
<211> 7660
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - plasmid construct

<400> 23
acattgatta ttgactagtt attaatagta atcaattacg gggtcattag ttcatagccc 60

atatatggag ttccgcgtta cataacttac ggtaaatggc ccgcctggct gaccgcccaa 120

cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac 180

tttccattga cgtcaatggg tggagtattt acggtaaact gcccacttgg cagtacatca 240

agtgtatcat atgccaagta cgcccccctat tgacgtcaat gacggtaaat ggcccgcctg 300

gcattatgcc cagtacatga ccttatggga ctttcctact tggcagtaca tctacgtatt 360

agtcatcgct attaccatgg tgatgcggtt ttggcagtac atcaatgggc gtggatagcg 420

gtttgactca cggggatttc caagtctcca ccccattgac gtcaatggga gtttgttttg 480

gcaccaaaat caacgggact ttccaaaatg tcgtaacaac tccgccccat tgacgcaaat 540

gggcggtagg cgtgtacggt gggaggtcta tataagcaga gctcgtttag tgaaccgggg 600

tctctctggt tagaccagat ctgagcctgg gagctctctg gctaactagg gaacccactg 660

cttaagcctc aataaagctt gccttgagtg cttcaagtag tgtgtgcccg tctgttgtgt 720

gactctggta actagagatc cctcagaccc ttttagtcag tgtggaaaat ctctagcagt 780

ggcgcccgaa cagggacttg aaagcgaaag ggaaaccaga ggagctctct cgacgcagga 840

ctcggcttgc tgaagcgcgc acggcaagag gcgaggggcg gcgactggtg agtacgccaa 900

aaattttgac tagcggaggc tagaaggaga gagatgggtg cgagagcgtc agtattaagc 960

gggggagaat tagatcgcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat 1020

ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg 1080

gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa ccatcccttc 1140

agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc 1200

atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa 1260

acaaaagtaa gaccaccgca cagcaagcgg ccgctgatct tcagacctgg aggaggagat 1320

atgagggaca attggagaag tgaattatat aaatataaag tagtaaaaat tgaaccatta 1380

ggagtagcac ccaccaaggc aaagagaaga gtggtgcaga gagaaaaaag agcagtggga 1440

ataggagctt tgttccttgg gttcttggga gcagcaggaa gcactatggg cgcagcgtca 1500

atgacgctga cggtacaggc cagacaatta ttgtctggta tagtgcagca gcagaacaat 1560

ttgctgaggg ctattgaggc gcaacagcat ctgttgcaac tcacagtctg gggcatcaag 1620

cagctccagg caagaatcct ggctgtggaa agatacctaa aggatcaaca gctcctgggg 1680

atttggggtt gctctggaaa actcatttgc accactgctg tgccttggaa tgctagttgg 1740

agtaataaat ctctggaaca gatttggaat cacacgacct ggatggagtg ggacagagaa 1800

attaacaatt acacaagctt aatacactcc ttaattgaag aatcgcaaaa ccagcaagaa 1860

aagaatgaac aagaattatt ggaattagat aaatgggcaa gtttgtggaa ttggtttaac 1920

ataacaaatt ggctgtggta tataaaatta ttcataatga tagtaggagg cttggtaggt 1980

ttaagaatag tttttgctgt actttctata gtgaatagag ttaggcaggg atattcacca 2040

ttatcgtttc agacccacct cccaaccccg aggggacccg acaggcccga aggaatagaa 2100

gaagaaggtg gagagagaga cagagacaga tccattcgat tagtgaacgg atctcgacgg 2160

tatcggttaa cttttaaaag aaaaggggggg attgggggggt acagtgcagg ggaaagaata 2220

gtagacataa tagcaacaga catacaaact aaagaattac aaaaacaaat tacaaaaatt 2280

caaaatttta tcgatcacga gactagcctc gagaagcttg atcgattggc tccggtgccc 2340

gtcagtgggc agagcgcaca tcgcccacag tccccgagaa gttgggggga ggggtcggca 2400

attgaaccgg tgcctagaga aggtggcgcg gggtaaactg ggaaagtgat gtcgtgtact 2460

ggctccgcct ttttcccgag ggtgggggag aaccgtatat aagtgcagta gtcgccgtga 2520

acgttctttt tcgcaacggg tttgccgcca gaacacaggt gtcgtgacgc gggatccgcc 2580

accatgggct ccatgtttcg gagcgaggag gtggccctgg tccagctctt tctgcccaca 2640

gcggctgcct acacctgcgt gagtcggctg ggcgagctgg gcctcgtgga gttcagagac 2700

ctcaacgcct cggtgagcgc cttccagaga cgctttgtgg ttgatgttcg gcgctgtgag 2760

gagctggaga agaccttcac cttcctgcag gaggaggtgc ggcgggctgg gctggtcctg 2820

cccccgccaa aggggaggct gccggcaccc ccaccccggg acctgctgcg catccaggag 2880

gagacggagc gcctggccca ggagctgcgg gatgtgcggg gcaaccagca ggccctgcgg 2940

gcccagctgc accagctgca gctccacgcc gccgtgctac gccagggcca tgaacctcag 3000

ctggcagccg cccacacaga tggggcctca gagaggacgc ccctgctcca ggcccccggg 3060

gggccgcacc aggacctgag ggtcaacttt gtggcaggtg ccgtggagcc ccacaaggcc 3120

cctgccctag agcgcctgct ctggagggcc tgcagaggct tcctcattgc cagcttcagg 3180

gagctggagc agccgctgga gcaccccgtg acgggcgagc cagccacgtg gatgaccttc 3240

ctcatctcct actggggtga gcagatcgga cagaagatcc gcaagatcac ggactgcttc 3300

cactgccacg tcttcccgtt tctgcagcag gaggaggccc gcctcggggc cctgcagcag 3360

ctgcaacagc agagccagga gctgcaggag gtcctcgggg agacagagcg gttcctgagc 3420

caggtgctag gccgggtgct gcagctgctg ccgccagggc aggtgcaggt ccacaagatg 3480

aaggccgtgt acctggccct gaaccagtgc agcgtgagca ccacgcacaa gtgcctcatt 3540

gccgaggcct ggtgctctgt gcgagacctg cccgccctgc aggaggccct gcgggacagc 3600

tcgatggagg agggagtgag tgccgtggct caccgcatcc cctgccggga catgcccccc 3660

acactcatcc gcaccaaccg cttcacggcc agcttccagg gcatcgtgga tgcctacggc 3720

gtgggccgct accaggaggt caaccccgct ccctacacca tcatcacctt ccccttcctg 3780

tttgctgtga tgttcgggga tgtgggccac gggctgctca tgttcctctt cgccctggcc 3840

atggtccttg cggagaaccg accggctgtg aaggccgcgc agaacgagat ctggcagact 3900

ttcttcaggg gccgctacct gctcctgctt atgggcctgt tctccatcta caccggcttc 3960

atctacaacg agtgcttcag tcgcgccacc agcatcttcc cctcgggctg gagtgtggcc 4020

gccatggcca accagtctgg ctggagtgat gcattcctgg cccagcacac gatgcttacc 4080

ctggacccca acgtcaccgg tgtcttcctg ggaccctacc cctttggcat cgatcctatt 4140

tggagcctgg ctgccaacca cttgagcttc ctcaactcct tcaagatgaa gatgtccgtc 4200

atcctgggcg tcgtgcacat ggcctttggg gtggtcctcg gagtcttcaa ccacgtgcac 4260

tttggccaga ggcaccggct gctgctggag acgctgccgg agctcacctt cctgctggga 4320

ctcttcggtt acctcgtgtt cctagtcatc tacaagtggc tgtgtgtctg ggctgccagg 4380

gccgcctcgg cccccagcat cctcatccac ttcatcaaca tgttcctctt ctcccacagc 4440

cccagcaaca ggctgctcta cccccggcag gaggtggtcc aggccacgct ggtggtcctg 4500

gccttggcca tggtgcccat cctgctgctt ggcacacccc tgcacctgct gcaccgccac 4560

cgccgccgcc tgcggaggag gcccgctgac cgacaggagg aaaacaaggc cgggttgctg 4620

gacctgcctg acgcatctgt gaatggctgg agctccgatg aggaaaaggc agggggcctg 4680

gatgatgaag aggaggccga gctcgtcccc tccgaggtgc tcatgcacca ggccatccac 4740

accatcgagt tctgcctggg ctgcgtctcc aacaccgcct cctacctgcg cctgtgggcc 4800

ctgagcctgg cccacgccca gctgtccgag gttctgtggg ccatggtgat gcgcataggc 4860

ctgggcctgg gccgggaggt gggcgtggcg gctgtggtgc tggtccccat ctttgccgcc 4920

tttgccgtga tgaccgtggc tatcctgctg gtgatggagg gactctcagc cttcctgcac 4980

gccctgcggc tgcactgggt ggaattccag aacaagttct actcaggcac gggctacaag 5040

ctgagtccct tcaccttcgc tgccacagat gactagtaag tcgacggatc ccccgggctg 5100

caggaattcg agcatcttac cgccatttat acccatattt gttctgtttt tcttgatttg 5160

ggtatacatt taaatgttaa taaaacaaaa tggtggggca atcatttaca tttttaggga 5220

tatgtaatta ctagttcagg tgtattgcca caagacaaac atgttaagaa actttcccgt 5280

tatttacgct ctgttcctgt taatcaacct ctggattaca aaatttgtga aagattgact 5340

gatattctta actatgttgc tccttttacg ctgtgtggat atgctgcttt aatgcctctg 5400

tatcatgcta ttgcttcccg tacggctttc gttttctcct ccttgtataa atcctggttg 5460

ctgtctcttt atgaggagtt gtggcccgtt gtccgtcaac gtggcgtggt gtgctctgtg 5520

tttgctgacg caacccccac tggctggggc attgccacca cctgtcaact cctttctggg 5580

actttcgctt tccccctccc gatcgccacg gcagaactca tcgccgcctg ccttgcccgc 5640

tgctggacag gggctaggtt gctgggcact gataattccg tggtgttgtc ggggaagctg 5700

acgtcctttc gaattcgata tcaagctgta cctttaagac caatgactta caaggcagct 5760

gtagatctta gccacttttt aaaagaaaag gggggactgg aagggctaat tcactcccaa 5820

cgaagacaag atctgctttt tgcttgtact gggtctctct ggttagacca gatctgagcc 5880

tgggagctct ctggctaact agggaaccca ctgcttaagc ctcaataaag cttgccttga 5940

gtgcttcaag tagtgtgtgc ccgtctgttg tgtgactctg gtaactagag atccctcaga 6000

cccttttagt cagtgtggaa aatctctagc agtagtagtt catgtcatct tattattcag 6060

tatttataac ttgcaaagaa atgaatatca gagagtgaga ggaacttgtt tattgcagct 6120

tataatggtt acaaataaag caatagcatc acaaatttca caaataaagc attttttttca 6180

ctgcattcta gttgtggttt gtccaaactc atcaatgtat cttatcatgt ctggctctag 6240

ctatcccgcc cctaactccg cccatcccgc ccctaactcc gcccagttcc gcccattctc 6300

cgccccatgg ctgactaatt ttttttattt atgcagaggc cgaggccgcc tcggcctctg 6360

agctattcca gaagtagtga ggaggctttt ttggaggcct aggtagcccg cctaatgagc 6420

gggctttttt ttcttaggcc ttcttccgct tcctcgctca ctgactcgct gcgctcggtc 6480

gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg taatacggtt atccacagaa 6540

tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt 6600

aaaaaggccg cgttgctggc gtttttccat aggctccgcc cccctgacga gcatcacaaa 6660

aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcgttt 6720

ccccctggaa gctccctcgt gcgctctcct gttccgaccc tgccgcttac cggatacctg 6780

tccgcctttc tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc 6840

agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc 6900

gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccggtaag acacgactta 6960

tcgccactgg cagcagccac tggtaacagg attagcagag cgaggtatgt aggcggtgct 7020

acagagttct tgaagtggtg gcctaactac ggctacacta gaagaacagt atttggtatc 7080

tgcgctctgc tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa 7140

caaaccaccg ctggtagcgg tggtttttt gtttgcaagc agcagattac gcgcagaaaa 7200

aaaggatctc aagaagatcc tttgatcttt tctacggggt ctgacgctca gtggaacgaa 7260

aactcacgtt aagggatttt ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt 7320

ttaaattaaa aatgaagttt taaatcaatc taaagtatat atgagtaaac ttggtctgac 7380

agttaccaat gcttaatcag tgaggcacct atctcagcga tctgtctatt tcgttcatcc 7440

atagttgcct gactcctgca aaccacgttg tggtagaatt ggtaaagaga gtcgtgtaaa 7500

atatcgagtt cgcacatctt gttgtctgat tattgatttt tggcgaaacc atttgatcat 7560

atgacaagat gtgtatctac cttaacttaa tgattttgat aaaaatcatt aggtacctgt 7620

acatttatat tggctcatgt ccaacattac cgccatgttg 7660

<210> 24
<211> 511

<212> DNA
<213> Homo sapiens

<400> 24
ggggttgggg ttgcgccttt tccaaggcag ccctgggttt gcgcagggac gcggctgctc 60

tgggcgtggt tccgggaaac gcagcggcgc cgaccctggg tctcgcacat tcttcacgtc 120

cgttcgcagc gtcacccgga tcttcgccgc taccttgtg ggccccccgg cgacgcttcc 180

tgctccgccc ctaagtcggg aaggttcctt gcggttcgcg gcgtgccgga cgtgacaaac 240

ggaagccgca cgtctcacta gtaccctcgc agacggacag cgccagggag caatggcagc 300

gcgccgaccg cgatgggctg tggccaatag cggctgctca gcagggcgcg ccgagagcag 360

cggccgggaa ggggcggtgc gggaggcggg gtgtggggcg gtagtgtggg ccctgttcct 420

gcccgcgcgg tgttccgcat tctgcaagcc tccggagcgc acgtcggcag tcggctccct 480

cgttgaccga atcaccgacc tctctcccca g 511

<210> 25
<211> 7613
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - plasmid construct

<400> 25
acattgatta ttgactagtt attaatagta atcaattacg gggtcattag ttcatagccc 60

atatatggag ttccgcgtta cataacttac ggtaaatggc ccgcctggct gaccgcccaa 120

cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac 180

tttccattga cgtcaatggg tggagtattt acggtaaact gcccacttgg cagtacatca 240

agtgtatcat atgccaagta cgcccccctat tgacgtcaat gacggtaaat ggcccgcctg 300

gcattatgcc cagtacatga ccttatggga ctttcctact tggcagtaca tctacgtatt 360

agtcatcgct attaccatgg tgatgcggtt ttggcagtac atcaatgggc gtggatagcg 420

gtttgactca cggggatttc caagtctcca ccccattgac gtcaatggga gtttgttttg 480

gcaccaaaat caacgggact ttccaaaatg tcgtaacaac tccgccccat tgacgcaaat 540

gggcggtagg cgtgtacggt gggaggtcta tataagcaga gctcgtttag tgaaccgggg 600

tctctctggt tagaccagat ctgagcctgg gagctctctg gctaactagg gaacccactg 660

cttaagcctc aataaagctt gccttgagtg cttcaagtag tgtgtgcccg tctgttgtgt 720

gactctggta actagagatc cctcagaccc ttttagtcag tgtggaaaat ctctagcagt 780

ggcgcccgaa cagggacttg aaagcgaaag ggaaaccaga ggagctctct cgacgcagga 840

ctcggcttgc tgaagcgcgc acggcaagag gcgaggggcg gcgactggtg agtacgccaa 900

aaattttgac tagcggaggc tagaaggaga gagatgggtg cgagagcgtc agtattaagc 960

gggggagaat tagatcgcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat 1020

ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg 1080

gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa ccatcccttc 1140

agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc 1200

atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa 1260

acaaaagtaa gaccaccgca cagcaagcgg ccgctgatct tcagacctgg aggaggagat 1320

atgagggaca attggagaag tgaattatat aaatataaag tagtaaaaat tgaaccatta 1380

ggagtagcac ccaccaaggc aaagagaaga gtggtgcaga gagaaaaaag agcagtggga 1440

ataggagctt tgttccttgg gttcttggga gcagcaggaa gcactatggg cgcagcgtca 1500

atgacgctga cggtacaggc cagacaatta ttgtctggta tagtgcagca gcagaacaat 1560

ttgctgaggg ctattgaggc gcaacagcat ctgttgcaac tcacagtctg gggcatcaag 1620

cagctccagg caagaatcct ggctgtggaa agatacctaa aggatcaaca gctcctgggg 1680

atttggggtt gctctggaaa actcatttgc accactgctg tgccttggaa tgctagttgg 1740

agtaataaat ctctggaaca gatttggaat cacacgacct ggatggagtg ggacagagaa 1800

attaacaatt acacaagctt aatacactcc ttaattgaag aatcgcaaaa ccagcaagaa 1860

aagaatgaac aagaattatt ggaattagat aaatgggcaa gtttgtggaa ttggtttaac 1920

ataacaaatt ggctgtggta tataaaatta ttcataatga tagtaggagg cttggtaggt 1980

ttaagaatag tttttgctgt actttctata gtgaatagag ttaggcaggg atattcacca 2040

ttatcgtttc agacccacct cccaaccccg aggggacccg acaggcccga aggaatagaa 2100

gaagaaggtg gagagagaga cagagacaga tccattcgat tagtgaacgg atctcgacgg 2160

tatcggttaa cttttaaaag aaaaggggggg attgggggggt acagtgcagg ggaaagaata 2220

gtagacataa tagcaacaga catacaaact aaagaattac aaaaacaaat tacaaaaatt 2280

caaaatttta tccgtcagtg ggcagagcgc acatcgccca cagtccccga gaagttgggg 2340

ggaggggtcg gcaattgaac cggtgcctag agaaggtggc gcggggtaaa ctgggaaagt 2400

gatgtcgtgt actggctccg cctttttccc gagggtgggg gagaaccgta tataagtgca 2460

gtagtcgccg tgaacgttct ttttcgcaac gggtttgccg ccagaacaca ggtgtcgtga 2520

cgcgggatcc gccaccatgg gctccatgtt tcggagcgag gaggtggccc tggtccagct 2580

ctttctgccc acagcggctg cctacacctg cgtgagtcgg ctgggcgagc tgggcctcgt 2640

ggagttcaga gacctcaacg cctcggtgag cgccttccag agacgctttg tggttgatgt 2700

tcggcgctgt gaggagctgg agaagacctt caccttcctg caggaggagg tgcggcgggc 2760

tgggctggtc ctgcccccgc caaaggggag gctgccggca cccccacccc gggacctgct 2820

gcgcatccag gaggagacgg agcgcctggc ccaggagctg cgggatgtgc ggggcaacca 2880

gcaggccctg cgggcccagc tgcaccagct gcagctccac gccgccgtgc tacgccaggg 2940

ccatgaacct cagctggcag ccgcccacac agatggggcc tcagagagga cgcccctgct 3000

ccaggccccc ggggggccgc accaggacct gagggtcaac tttgtggcag gtgccgtgga 3060

gccccacaag gcccctgccc tagagcgcct gctctggagg gcctgcagag gcttcctcat 3120

tgccagcttc agggagctgg agcagccgct ggagcacccc gtgacgggcg agccagccac 3180

gtggatgacc ttcctcatct cctactgggg tgagcagatc ggacagaaga tccgcaagat 3240

cacggactgc ttccactgcc acgtcttccc gtttctgcag caggaggagg cccgcctcgg 3300

ggccctgcag cagctgcaac agcagagcca ggagctgcag gaggtcctcg gggagacaga 3360

gcggttcctg agccaggtgc taggccgggt gctgcagctg ctgccgccag ggcaggtgca 3420

ggtccacaag atgaaggccg tgtacctggc cctgaaccag tgcagcgtga gcaccacgca 3480

caagtgcctc attgccgagg cctggtgctc tgtgcgagac ctgcccgccc tgcaggaggc 3540

cctgcgggac agctcgatgg aggagggagt gagtgccgtg gctcaccgca tcccctgccg 3600

ggacatgccc cccacactca tccgcaccaa ccgcttcacg gccagcttcc agggcatcgt 3660

ggatgcctac ggcgtgggcc gctaccagga ggtcaacccc gctccctaca ccatcatcac 3720

cttccccttc ctgtttgctg tgatgttcgg ggatgtgggc cacgggctgc tcatgttcct 3780

cttcgccctg gccatggtcc ttgcggagaa ccgaccggct gtgaaggccg cgcagaacga 3840

gatctggcag actttcttca ggggccgcta cctgctcctg cttatgggcc tgttctccat 3900

ctacaccggc ttcatctaca acgagtgctt cagtcgcgcc accagcatct tcccctcggg 3960

ctggagtgtg gccgccatgg ccaaccagtc tggctggagt gatgcattcc tggcccagca 4020

cacgatgctt accctggacc ccaacgtcac cggtgtcttc ctgggaccct acccctttgg 4080

catcgatcct atttggagcc tggctgccaa ccacttgagc ttcctcaact ccttcaagat 4140

gaagatgtcc gtcatcctgg gcgtcgtgca catggccttt ggggtggtcc tcggagtctt 4200

caaccacgtg cactttggcc agaggcaccg gctgctgctg gagacgctgc cggagctcac 4260

cttcctgctg ggactcttcg gttacctcgt gttcctagtc atctacaagt ggctgtgtgt 4320

ctgggctgcc agggccgcct cggcccccag catcctcatc cacttcatca acatgttcct 4380

cttctcccac agccccagca acaggctgct ctacccccgg caggaggtgg tccaggccac 4440

gctggtggtc ctggccttgg ccatggtgcc catcctgctg cttggcacac ccctgcacct 4500

gctgcaccgc caccgccgcc gcctgcggag gaggcccgct gaccgacagg aggaaaacaa 4560

ggccgggttg ctggacctgc ctgacgcatc tgtgaatggc tggagctccg atgaggaaaa 4620

ggcaggggggc ctggatgatg aagaggaggc cgagctcgtc ccctccgagg tgctcatgca 4680

ccaggccatc cacaccatcg agttctgcct gggctgcgtc tccaacaccg cctcctacct 4740

gcgcctgtgg gccctgagcc tggcccacgc ccagctgtcc gaggttctgt gggccatggt 4800

gatgcgcata ggcctgggcc tgggccggga ggtgggcgtg gcggctgtgg tgctggtccc 4860

catctttgcc gcctttgccg tgatgaccgt ggctatcctg ctggtgatgg agggactctc 4920

agccttcctg cacgccctgc ggctgcactg ggtggaattc cagaacaagt tctactcagg 4980

cacgggctac aagctgagtc ccttcacctt cgctgccaca gatgactagt aagtcgacgg 5040

atcccccggg ctgcaggaat tcgagcatct taccgccatt tatacccata tttgttctgt 5100

ttttcttgat ttgggtatac atttaaatgt taataaaaca aaatggtggg gcaatcattt 5160

acatttttag ggatatgtaa ttactagttc aggtgtattg ccacaagaca aacatgttaa 5220

gaaactttcc cgttatttac gctctgttcc tgttaatcaa cctctggatt acaaaatttg 5280

tgaaagattg actgatattc ttaactatgt tgctccttt acgctgtgtg gatatgctgc 5340

tttaatgcct ctgtatcatg ctattgcttc ccgtacggct ttcgttttct cctccttgta 5400

taaatcctgg ttgctgtctc tttatgagga gttgtggccc gttgtccgtc aacgtggcgt 5460

ggtgtgctct gtgtttgctg acgcaacccc cactggctgg ggcattgcca ccacctgtca 5520

actcctttct gggactttcg ctttcccccc cccgatcgcc acggcagaac tcatcgccgc 5580

ctgccttgcc cgctgctgga caggggctag gttgctgggc actgataatt ccgtggtgtt 5640

gtcggggaag ctgacgtcct ttcgaattcg atatcaagct gtacctttaa gaccaatgac 5700

ttacaaggca gctgtagatc ttagccactt tttaaaagaa aaggggggac tggaagggct 5760

aattcactcc caacgaagac aagatctgct ttttgcttgt actgggtctc tctggttaga 5820

ccagatctga gcctgggagc tctctggcta actagggaac ccactgctta agcctcaata 5880

aagcttgcct tgagtgcttc aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta 5940

gagatccctc agacccttttt agtcagtgtg gaaaatctct agcagtagta gttcatgtca 6000

tcttattatt cagtatttat aacttgcaaa gaaatgaata tcagagagtg agaggaactt 6060

gtttattgca gcttataatg gttacaaata aagcaatagc atcacaaatt tcacaaataa 6120

agcatttttt tcactgcatt ctagttgtgg tttgtccaaa ctcatcaatg tatcttatca 6180

tgtctggctc tagctatccc gcccctaact ccgcccatcc cgcccctaac tccgcccagt 6240

tccgcccatt ctccgcccca tggctgacta attttttttа tttatgcaga ggccgaggcc 6300

gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg cctaggtagc 6360

ccgcctaatg agcgggcttt tttttcttag gccttcttcc gcttcctcgc tcactgactc 6420

gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg 6480

gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa 6540

ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga 6600

cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag 6660

ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct 6720

taccggatac ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc atagctcacg 6780

ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc 6840

ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt 6900

aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta 6960

tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 7020

agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc 7080

ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat 7140

tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc 7200

tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt 7260

cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta 7320

aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct 7380

atttcgttca tccatagttg cctgactcct gcaaaccacg ttgtggtaga attggtaaag 7440

agagtcgtgt aaaatatcga gttcgcacat cttgttgtct gattattgat ttttggcgaa 7500

accatttgat catatgacaa gatgtgtatc taccttaact taatgatttt gataaaaatc 7560

attaggtacc tgtacattta tattggctca tgtccaacat taccgccatg ttg 7613

<210> 26
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - section of plasmid construct

<400> 26
gatcacgaga ctagcctcga gaagcttgat cgattggctc cggtgcc 47

<210> 27
<211> 7646
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in lab - plasmid construct

<400> 27
ggctccggtg cccgtcagtg ggcagagcgc acatcgccca cagtccccga gaagttgggg 60

ggaggggtcg gcaattgaac cggtgcctag agaaggtggc gcggggtaaa ctgggaaagt 120

gatgtcgtgt actggctccg cctttttccc gagggtgggg gagaaccgta tataagtgca 180

gtagtcgccg tgaacgttct ttttcgcaac gggtttgccg ccagaacaca ggtgtcgtga 240

cgcgggatcc gccaccatgg gctccatgtt tcggagcgag gaggtggccc tggtccagct 300

ctttctgccc acagcggctg cctacacctg cgtgagtcgg ctgggcgagc tgggcctcgt 360

ggagttcaga gacctcaacg cctcggtgag cgccttccag agacgctttg tggttgatgt 420

tcggcgctgt gaggagctgg agaagacctt caccttcctg caggaggagg tgcggcgggc 480

tgggctggtc ctgcccccgc caaaggggag gctgccggca cccccacccc gggacctgct 540

gcgcatccag gaggagacgg agcgcctggc ccaggagctg cgggatgtgc ggggcaacca 600

gcaggccctg cgggcccagc tgcaccagct gcagctccac gccgccgtgc tacgccaggg 660

ccatgaacct cagctggcag ccgcccacac agatggggcc tcagagagga cgcccctgct 720

ccaggccccc ggggggccgc accaggacct gagggtcaac tttgtggcag gtgccgtgga 780

gccccacaag gcccctgccc tagagcgcct gctctggagg gcctgcagag gcttcctcat 840

tgccagcttc agggagctgg agcagccgct ggagcacccc gtgacgggcg agccagccac 900

gtggatgacc ttcctcatct cctactgggg tgagcagatc ggacagaaga tccgcaagat 960

cacggactgc ttccactgcc acgtcttccc gtttctgcag caggaggagg cccgcctcgg 1020

ggccctgcag cagctgcaac agcagagcca ggagctgcag gaggtcctcg gggagacaga 1080

gcggttcctg agccaggtgc taggccgggt gctgcagctg ctgccgccag ggcaggtgca 1140

ggtccacaag atgaaggccg tgtacctggc cctgaaccag tgcagcgtga gcaccacgca 1200

caagtgcctc attgccgagg cctggtgctc tgtgcgagac ctgcccgccc tgcaggaggc 1260

cctgcgggac agctcgatgg aggagggagt gagtgccgtg gctcaccgca tcccctgccg 1320

ggacatgccc cccacactca tccgcaccaa ccgcttcacg gccagcttcc agggcatcgt 1380

ggatgcctac ggcgtgggcc gctaccagga ggtcaacccc gctccctaca ccatcatcac 1440

cttccccttc ctgtttgctg tgatgttcgg ggatgtgggc cacgggctgc tcatgttcct 1500

cttcgccctg gccatggtcc ttgcggagaa ccgaccggct gtgaaggccg cgcagaacga 1560

gatctggcag actttcttca ggggccgcta cctgctcctg cttatgggcc tgttctccat 1620

ctacaccggc ttcatctaca acgagtgctt cagtcgcgcc accagcatct tcccctcggg 1680

ctggagtgtg gccgccatgg ccaaccagtc tggctggagt gatgcattcc tggcccagca 1740

cacgatgctt accctggacc ccaacgtcac cggtgtcttc ctgggaccct acccctttgg 1800

catcgatcct atttggagcc tggctgccaa ccacttgagc ttcctcaact ccttcaagat 1860

gaagatgtcc gtcatcctgg gcgtcgtgca catggccttt ggggtggtcc tcggagtctt 1920

caaccacgtg cactttggcc agaggcaccg gctgctgctg gagacgctgc cggagctcac 1980

cttcctgctg ggactcttcg gttacctcgt gttcctagtc atctacaagt ggctgtgtgt 2040

ctgggctgcc agggccgcct cggcccccag catcctcatc cacttcatca acatgttcct 2100

cttctcccac agccccagca acaggctgct ctacccccgg caggaggtgg tccaggccac 2160

gctggtggtc ctggccttgg ccatggtgcc catcctgctg cttggcacac ccctgcacct 2220

gctgcaccgc caccgccgcc gcctgcggag gaggcccgct gaccgacagg aggaaaacaa 2280

ggccgggttg ctggacctgc ctgacgcatc tgtgaatggc tggagctccg atgaggaaaa 2340

ggcaggggggc ctggatgatg aagaggaggc cgagctcgtc ccctccgagg tgctcatgca 2400

ccaggccatc cacaccatcg agttctgcct gggctgcgtc tccaacaccg cctcctacct 2460

gcgcctgtgg gccctgagcc tggcccacgc ccagctgtcc gaggttctgt gggccatggt 2520

gatgcgcata ggcctgggcc tgggccggga ggtgggcgtg gcggctgtgg tgctggtccc 2580

catctttgcc gcctttgccg tgatgaccgt ggctatcctg ctggtgatgg agggactctc 2640

agccttcctg cacgccctgc ggctgcactg ggtggaattc cagaacaagt tctactcagg 2700

cacgggctac aagctgagtc ccttcacctt cgctgccaca gatgactagt aagtcgacgg 2760

atcccccggg ctgcaggaat tcgagcatct taccgccatt tatacccata tttgttctgt 2820

ttttcttgat ttgggtatac atttaaatgt taataaaaca aaatggtggg gcaatcattt 2880

acatttttag ggatatgtaa ttactagttc aggtgtattg ccacaagaca aacatgttaa 2940

gaaactttcc cgttatttac gctctgttcc tgttaatcaa cctctggatt acaaaatttg 3000

tgaaagattg actgatattc ttaactatgt tgctcctttt acgctgtgtg gatatgctgc 3060

tttaatgcct ctgtatcatg ctattgcttc ccgtacggct ttcgttttct cctccttgta 3120

taaatcctgg ttgctgtctc tttatgagga gttgtggccc gttgtccgtc aacgtggcgt 3180

ggtgtgctct gtgtttgctg acgcaacccc cactggctgg ggcattgcca ccacctgtca 3240

actccttttct gggactttcg ctttcccccct cccgatcgcc acggcagaac tcatcgccgc 3300

ctgccttgcc cgctgctgga caggggctag gttgctgggc actgataatt ccgtggtgtt 3360

gtcggggaag ctgacgtcct ttcgaattcg atatcaagct gtacctttaa gaccaatgac 3420

ttacaaggca gctgtagatc ttagccactt tttaaaagaa aaggggggac tggaagggct 3480

aattcactcc caacgaagac aagatctgct ttttgcttgt actgggtctc tctggttaga 3540

ccagatctga gcctgggagc tctctggcta actagggaac ccactgctta agcctcaata 3600

aagcttgcct tgagtgcttc aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta 3660

gagatccctc agacccttttt agtcagtgtg gaaaatctct agcagtagta gttcatgtca 3720

tcttattatt cagtatttat aacttgcaaa gaaatgaata tcagagagtg agaggaactt 3780

gtttattgca gcttataatg gttacaaata aagcaatagc atcacaaatt tcacaaataa 3840

agcatttttt tcactgcatt ctagttgtgg tttgtccaaa ctcatcaatg tatcttatca 3900

tgtctggctc tagctatccc gcccctaact ccgcccatcc cgcccctaac tccgcccagt 3960

tccgcccatt ctccgcccca tggctgacta attttttttta tttatgcaga ggccgaggcc 4020

gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg cctaggtagc 4080

ccgcctaatg agcgggcttt tttttcttag gccttcttcc gcttcctcgc tcactgactc 4140

gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg 4200

gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa 4260

ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccccctga 4320

cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag 4380

ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct 4440

taccggatac ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc atagctcacg 4500

ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc 4560

ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt 4620

aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta 4680

tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 4740

agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc 4800

ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat 4860

tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc 4920

tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt 4980

cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta 5040

aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct 5100

atttcgttca tccatagttg cctgactcct gcaaaccacg ttgtggtaga attggtaaag 5160

agagtcgtgt aaaatatcga gttcgcacat cttgttgtct gattattgat ttttggcgaa 5220

accatttgat catatgacaa gatgtgtatc taccttaact taatgatttt gataaaaatc 5280

attaggtacc tgtacattta tattggctca tgtccaacat taccgccatg ttgacattga 5340

ttattgacta gttattaata gtaatcaatt acggggtcat tagttcatag cccatatatg 5400

gagttccgcg ttacataact tacggtaaat ggcccgcctg gctgaccgcc caacgacccc 5460

cgcccattga cgtcaataat gacgtatgtt cccatagtaa cgccaatagg gactttccat 5520

tgacgtcaat gggtggagta tttacggtaa actgcccact tggcagtaca tcaagtgtat 5580

catatgccaa gtacgccccc tattgacgtc aatgacggta aatggcccgc ctggcattat 5640

gcccagtaca tgaccttatg ggactttcct acttggcagt acatctacgt attagtcatc 5700

gctattacca tggtgatgcg gttttggcag tacatcaatg ggcgtggata gcggtttgac 5760

tcacggggat ttccaagtct ccaccccatt gacgtcaatg ggagtttgtt ttggcaccaa 5820

aatcaacggg actttccaaa atgtcgtaac aactccgccc cattgacgca aatgggcggt 5880

aggcgtgtac ggtgggaggt ctatataagc agagctcgtt tagtgaaccg gggtctctct 5940

ggttagacca gatctgagcc tgggagctct ctggctaact agggaaccca ctgcttaagc 6000

ctcaataaag cttgccttga gtgcttcaag tagtgtgtgc ccgtctgttg tgtgactctg 6060

gtaactagag atccctcaga cccttttagt cagtgtggaa aatctctagc agtggcgccc 6120

gaacagggac ttgaaagcga aagggaaacc agaggagctc tctcgacgca ggactcggct 6180

tgctgaagcg cgcacggcaa gaggcgaggg gcggcgactg gtgagtacgc caaaaatttt 6240

gactagcgga ggctagaagg agagagatgg gtgcgagagc gtcagtatta agcgggggag 6300

aattagatcg cgatgggaaa aaattcggtt aaggccaggg ggaaagaaaa aatataaatt 6360

aaaacatata gtatgggcaa gcagggagct agaacgattc gcagttaatc ctggcctgtt 6420

agaaacatca gaaggctgta gacaaatact gggacagcta caaccatccc ttcagacagg 6480

atcagaagaa cttagatcat tatataatac agtagcaacc ctctattgtg tgcatcaaag 6540

gatagagata aaagacacca aggaagcttt agacaagata gaggaagagc aaaacaaaag 6600

taagaccacc gcacagcaag cggccgctga tcttcagacc tggaggagga gatatgaggg 6660

acaattggag aagtgaatta tataaatata aagtagtaaa aattgaacca ttaggagtag 6720

cacccaccaa ggcaaagaga agagtggtgc agagagaaaa aagagcagtg ggaataggag 6780

ctttgttcct tgggttcttg ggagcagcag gaagcactat gggcgcagcg tcaatgacgc 6840

tgacggtaca ggccagacaa ttattgtctg gtatagtgca gcagcagaac aatttgctga 6900

gggctattga ggcgcaacag catctgttgc aactcacagt ctggggcatc aagcagctcc 6960

aggcaagaat cctggctgtg gaaagatacc taaaggatca acagctcctg gggatttggg 7020

gttgctctgg aaaactcatt tgcaccactg ctgtgccttg gaatgctagt tggagtaata 7080

aatctctgga acagatttgg aatcacacga cctggatgga gtgggacaga gaaattaaca 7140

attacacaag cttaatacac tccttaattg aagaatcgca aaaccagcaa gaaaagaatg 7200

aacaagaatt attggaatta gataaatggg caagtttgtg gaattggttt aacataacaa 7260

attggctgtg gtatataaaa ttattcataa tgatagtagg aggcttggta ggtttaagaa 7320

tagtttttgc tgtactttct atagtgaata gagttaggca gggatattca ccattatcgt 7380

ttcagaccca cctcccaacc ccgaggggac ccgacaggcc cgaaggaata gaagaagaag 7440

gtggagagag agacagagac agatccattc gattagtgaa cggatctcga cggtatcggt 7500

taacttttaa aagaaaaggg gggattgggg ggtacagtgc aggggaaaga atagtagaca 7560

taatagcaac agacatacaa actaaagaat tacaaaaaca aattacaaaa attcaaaatt 7620

ttatcgatca cgagactagc ctcgag 7646

<210> 28
<211> 234
<212> DNA
<213> human immunodeficiency virus

<400> 28
tggaagggct aattcactcc caacgaagac aagatctgct ttttgcttgt actgggtctc 60

tctggttaga ccagatctga gcctgggagc tctctggcta actagggaac ccactgctta 120

agcctcaata aagcttgcct tgagtgcttc aagtagtgtg tgcccgtctg ttgtgtgact 180

ctggtaacta gagatccctc agacccttttt agtcagtgtg gaaaatctct agca 234

<210> 29
<211> 132
<212> DNA
<213> Simian virus 40

<400> 29
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 60

aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 120

tatcatgtct gg 132

<210> 30
<211> 160
<212> DNA
<213> Simian virus 40

<400> 30
tcccgcccct aactccgccc atcccgcccc taactccgcc cagttccgcc cattctccgc 60

cccatggctg actaattttt tttatttatg cagaggccga ggccgcctcg gcctctgagc 120

tattccagaa gtagtgagga ggcttttttg gaggcctagg 160

<210> 31
<211> 1015
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab

<400> 31
tcttccgctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg gcgagcggta 60

tcagctcact caaaggcggt aatacggtta tccacagaat caggggataa cgcaggaaag 120

aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg 180

tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg 240

tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg 300

cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct cccttcggga 360

agcgtggcgc tttctcatag ctcacgctgt aggtatctca gttcggtgta ggtcgttcgc 420

tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgcgc cttatccggt 480

aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc agcagccact 540

ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtggtgg 600

cctaactacg gctacactag aagaacagta tttggtatct gcgctctgct gaagccagtt 660

accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggt 720

ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct 780

ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta agggattttg 840

gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt 900

aaatcaatct aaagtatata tgagtaaact tggtctgaca gttaccaatg cttaatcagt 960

gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg actcc 1015

<210> 32
<211> 139
<212> DNA
<213> Escherichia coli

<400> 32
gtagaattgg taaagagagt cgtgtaaaat atcgagttcg cacatcttgt tgtctgatta 60

ttgatttttg gcgaaaccat ttgatcatat gacaagatgt gtatctacct taacttaatg 120

attttgataa aaatcatta 139

<210> 33
<211> 577
<212> DNA
<213> Human betaherpesvirus 5

<400> 33
acattgatta ttgactagtt attaatagta atcaattacg gggtcattag ttcatagccc 60

atatatggag ttccgcgtta cataacttac ggtaaatggc ccgcctggct gaccgcccaa 120

cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac 180

tttccattga cgtcaatggg tggagtattt acggtaaact gcccacttgg cagtacatca 240

agtgtatcat atgccaagta cgcccccctat tgacgtcaat gacggtaaat ggcccgcctg 300

gcattatgcc cagtacatga ccttatggga ctttcctact tggcagtaca tctacgtatt 360

agtcatcgct attaccatgg tgatgcggtt ttggcagtac atcaatgggc gtggatagcg 420

gtttgactca cggggatttc caagtctcca ccccattgac gtcaatggga gtttgttttg 480

gcaccaaaat caacgggact ttccaaaatg tcgtaacaac tccgccccat tgacgcaaat 540

gggcggtagg cgtgtacggt gggaggtcta tataagc 577

<210> 34
<211> 188
<212> DNA
<213> human immunodeficiency virus

<400> 34
gtctctctgg ttagaccaga tctgagcctg ggagctctct ggctaactag ggaacccact 60

gcttaagcct caataaagct tgccttgagt gcttcaagta gtgtgtgccc gtctgttgtg 120

tgactctggt aactagagat ccctcagacc cttttagtca gtgtggaaaa tctctagcag 180

tggcgccc 188

<210> 35
<211> 45
<212> DNA
<213> Human immunodeficiency virus 1

<400> 35
tgagtacgcc aaaaattttg actagcggag gctagaagga gagag 45

<210> 36
<211> 362
<212> DNA
<213> human immunodeficiency virus

<400> 36
atgggtgcga gagcgtcagt attaagcggg ggagaattag atcgcgatgg gaaaaaattc 60

ggttaaggcc agggggaaag aaaaaatata aattaaaaca tatagtatgg gcaagcaggg 120

agctagaacg attcgcagtt aatcctggcc tgttagaaac atcagaaggc tgtagacaaa 180

tactgggaca gctacaacca tcccttcaga caggatcaga agaacttaga tcattatata 240

atacagtagc aaccctctat tgtgtgcatc aaaggataga gataaaagac accaaggaag 300

ctttagacaa gatagaggaa gagcaaaaca aaagtaagac caccgcacag caagcggccg 360

ct 362

<210> 37
<211> 858
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - plasmid element

<400> 37
gatcttcaga cctggaggag gagatatgag ggacaattgg agaagtgaat tatataaata 60

taaagtagta aaaattgaac cattaggagt agcacccacc aaggcaaaga gaagagtggt 120

gcagagagaa aaaagagcag tgggaatagg agctttgttc cttgggttct tgggagcagc 180

aggaagcact atgggcgcag cgtcaatgac gctgacggta caggccagac aattattgtc 240

tggtatagtg cagcagcaga acaatttgct gagggctatt gaggcgcaac agcatctgtt 300

gcaactcaca gtctggggca tcaagcagct ccaggcaaga atcctggctg tggaaagata 360

cctaaaggat caacagctcc tggggatttg gggttgctct ggaaaactca tttgcaccac 420

tgctgtgcct tggaatgcta gttggagtaa taaatctctg gaacagattt ggaatcacac 480

gacctggatg gagtgggaca gagaaattaa caattacaca agcttaatac actccttaat 540

tgaagaatcg caaaaccagc aagaaaagaa tgaacaagaa ttattggaat tagataaatg 600

ggcaagtttg tggaattggt ttaacataac aaattggctg tggtatataa aattattcat 660

aatgatagta ggaggcttgg taggtttaag aatagttttt gctgtacttt ctatagtgaa 720

tagagttagg cagggatatt caccattatc gtttcagacc cacctcccaa ccccgagggg 780

acccgacagg cccgaaggaa tagaagaaga aggtggagag agagacagag acagatccat 840

tcgattagtg aacggatc 858

<210> 38
<211> 118
<212> DNA
<213> human immunodeficiency virus

<400> 38
ttttaaaaga aaaggggga ttgggggta cagtgcaggg gaaagaatag tagacataat 60

agcaacagac atacaaacta aagaattaca aaaacaaatt acaaaaattc aaaatttt 118

<210> 39
<211> 3847
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in lab - plasmid backbone construct

<400> 39
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 60

aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 120

tatcatgtct ggctctagct atcccgcccc taactccgcc catcccgccc ctaactccgc 180

ccagttccgc ccattctccg ccccatggct gactaatttt ttttatttat gcagaggccg 240

aggccgcctc ggcctctgag ctattccaga agtagtgagg aggcttttt ggaggcctag 300

gtagcccgcc taatgagcgg gctttttttt cttaggcctt cttccgcttc ctcgctcact 360

gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat cagctcactc aaaggcggta 420

atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag 480

caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc 540

cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta 600

taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg 660

ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc 720

tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac 780

gaaccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac 840

ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg 900

aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga 960

agaacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt 1020

agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag 1080

cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct 1140

gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg 1200

atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat 1260

gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc 1320

tgtctatttc gttcatccat agttgcctga ctcctgcaaa ccacgttgtg gtagaattgg 1380

taaagagagt cgtgtaaaat atcgagttcg cacatcttgt tgtctgatta ttgatttttg 1440

gcgaaaccat ttgatcatat gacaagatgt gtatctacct taacttaatg attttgataa 1500

aaatcattag gtacctgtac atttatattg gctcatgtcc aacattaccg ccatgttgac 1560

attgattatt gactagttat taatagtaat caattacggg gtcattagtt catagcccat 1620

atatggagtt ccgcgttaca taacttacgg taaatggccc gcctggctga ccgcccaacg 1680

acccccgccc attgacgtca ataatgacgt atgttcccat agtaacgcca atagggactt 1740

tccattgacg tcaatgggtg gagtatttac ggtaaactgc ccacttggca gtacatcaag 1800

tgtatcatat gccaagtacg ccccctattg acgtcaatga cggtaaatgg cccgcctggc 1860

attatgccca gtacatgacc ttatgggact ttcctacttg gcagtacatc tacgtattag 1920

tcatcgctat taccatggtg atgcggtttt ggcagtacat caatgggcgt ggatagcggt 1980

ttgactcacg gggatttcca agtctccacc ccattgacgt caatgggagt ttgttttggc 2040

accaaaatca acgggacttt ccaaaatgtc gtaacaactc cgccccattg acgcaaatgg 2100

gcggtaggcg tgtacggtgg gaggtctata taagcagagc tcgtttagtg aaccggggtc 2160

tctctggtta gaccagatct gagcctggga gctctctggc taactaggga acccactgct 2220

taagcctcaa taaagcttgc cttgagtgct tcaagtagtg tgtgcccgtc tgttgtgtga 2280

ctctggtaac tagagatccc tcagaccctt ttagtcagtg tggaaaatct ctagcagtgg 2340

cgcccgaaca gggacttgaa agcgaaaggg aaaccagagg agctctctcg acgcaggact 2400

cggcttgctg aagcgcgcac ggcaagaggc gaggggcggc gactggtgag tacgccaaaa 2460

attttgacta gcggaggcta gaaggagaga gatgggtgcg agagcgtcag tattaagcgg 2520

gggagaatta gatcgcgatg ggaaaaaatt cggttaaggc caggggggaaa gaaaaaatat 2580

aaattaaaac atatagtatg ggcaagcagg gagctagaac gattcgcagt taatcctggc 2640

ctgttagaaa catcagaagg ctgtagacaa atactgggac agctacaacc atcccttcag 2700

acaggatcag aagaacttag atcattatat aatacagtag caaccctcta ttgtgtgcat 2760

caaaggatag agataaaaga caccaaggaa gctttagaca agatagagga agagcaaaac 2820

aaaagtaaga ccaccgcaca gcaagcggcc gctgatcttc agacctggag gaggagatat 2880

gagggacaat tggagaagtg aattatataa atataaagta gtaaaaattg aaccattagg 2940

agtagcaccc accaaggcaa agagaagagt ggtgcagaga gaaaaaagag cagtgggaat 3000

aggagctttg ttccttgggt tcttgggagc agcaggaagc actatgggcg cagcgtcaat 3060

gacgctgacg gtacaggcca gacaattatt gtctggtata gtgcagcagc agaacaattt 3120

gctgagggct attgaggcgc aacagcatct gttgcaactc acagtctggg gcatcaagca 3180

gctccaggca agaatcctgg ctgtggaaag atacctaaag gatcaacagc tcctggggat 3240

ttggggttgc tctggaaaac tcatttgcac cactgctgtg ccttggaatg ctagttggag 3300

taataaatct ctggaacaga tttggaatca cacgacctgg atggagtggg acagagaaat 3360

taacaattac acaagcttaa tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa 3420

gaatgaacaa gaattattgg aattagataa atgggcaagt ttgtggaatt ggtttaacat 3480

aacaaattgg ctgtggtata taaaattatt cataatgata gtaggaggct tggtaggttt 3540

aagaatagtt tttgctgtac tttctatagt gaatagagtt aggcagggat attcaccatt 3600

atcgtttcag acccacctcc caaccccgag gggacccgac aggcccgaag gaatagaaga 3660

agaaggtgga gagagagaca gagacagatc cattcgatta gtgaacggat ctcgacggta 3720

tcggttaact tttaaaagaa aaggggggat tgggggggtac agtgcagggg aaagaatagt 3780

agacataata gcaacagaca tacaaactaa agaattacaa aaacaaatta caaaaattca 3840

aaatttt 3847