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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 December 2017 (21.12.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/218519 A1

(51) International Patent Classification:
A61K 39/21 (2006.01) *A61P 25/00* (2006.01)
A61K 48/00 (2006.01) *A61P 31/00* (2006.01)
A61P 3/00 (2006.01) *C12N 7/01* (2006.01)

(21) International Application Number:
PCT/US2017/037230

(22) International Filing Date:
13 June 2017 (13.06.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/349,505 13 June 2016 (13.06.2016) US
62/457,498 10 February 2017 (10.02.2017) US

(71) Applicant: **BLUEBIRD BIO, INC.** [US/US]; 60 Binney Street, Cambridge, Massachusetts 02142 (US).

(72) Inventors: **GOSS, Kendrick A.**; 75 Winter Street, Lexington, Massachusetts 02420 (US). **PARSONS, Geoffrey B.**; 17 Castleton Street, Jamaica Plain, Massachusetts 02130 (US). **GINIATULLINA, Asiya**; 85 Prescott Street, Apt 8, Cambridge, Massachusetts 02138 (US).

(74) Agent: **MASHRUWALA, Mary Anne** et al.; Cooley LLP, 1299 Pennsylvania Avenue, NW, Suite 700, Washington, District of Columbia 20004 (US).

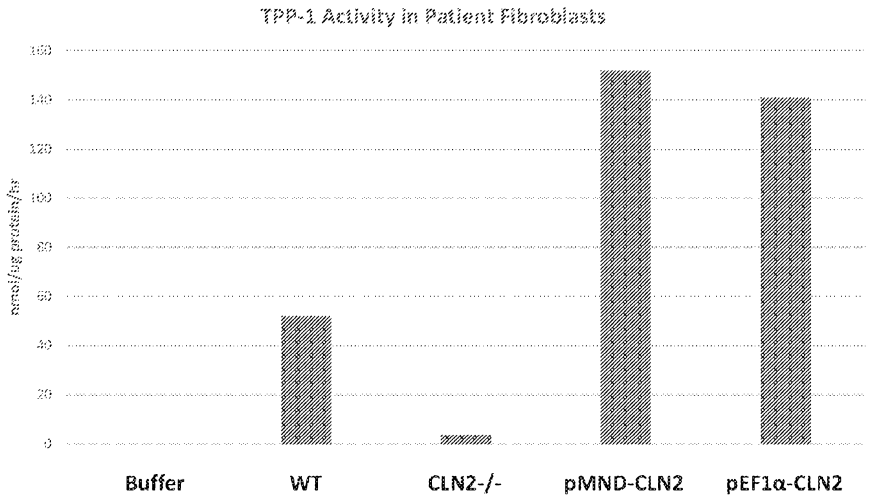
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

WO 2017/218519 A1

(54) Title: GENE THERAPY OF NEURONAL CEROID LIPOFUSCINOSES

Figure 3A



(57) Abstract: The invention provides, in part, compositions and methods for treating neuronal ceroid lipofuscinoses (NCL). The invention further provides, in part, gene therapy compositions and methods for the treatment, prevention, or amelioration of at least one symptom of NCL. Particular embodiments provide a pharmaceutical composition comprising a pharmaceutically acceptable carrier and a lentiviral vector, or a mammalian cell transduced with a lentiviral vector. Further embodiments provide a method of treating NCL comprising administering a lentiviral vector or a mammalian cell transduced with a lentiviral vector.

[Continued on next page]

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

Published:

— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

GENE THERAPY OF NEURONAL CEROID LIPOFUSCINOSES

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 62/457,498, filed February 10, 2017, and 62/349,505, filed June 13, 2016, each of which is incorporated by reference herein in its entirety.

STATEMENT REGARDING SEQUENCE LISTING

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 BLBD_070_02WO_ST25.txt. The text file is 33 KB, was created on June 13, 2017, and is being submitted electronically via EFS-Web, concurrent with the filing of the specification.

BACKGROUND

Technical Field

The present invention relates to gene therapy. More particularly, the invention relates to gene therapy compositions and methods of using the same to treat neuronal ceroid lipofuscinoses.

Description of the Related Art

Neuronal ceroid lipofuscinoses (NCLs) are group of fatal, inherited disorders of the nervous system. Most NCLs are autosomal recessive diseases. NCLs typically begin in early childhood and affect an estimated 2 to 4 of every 100,000 individuals. These disorders appear to be more common in Finland, Sweden, other parts of northern Europe, and Newfoundland, Canada. Early symptoms of these disorders include subtle personality and behavior changes, slow learning, clumsiness, or stumbling. As the disease progresses, affected children suffer cognitive impairment, worsening seizures, and progressive loss of

sight and motor skills. Eventually, children with NCL become blind, bedridden, and demented.

Batten disease (Spielmeyer-Vogt-Sjogren-Batten disease) is the most common NCL. Although Batten disease originally referred specifically to the juvenile form of NCL (JNCL), the term Batten disease is increasingly used by pediatricians to describe all forms of NCL. There are four other main types of NCL, including three forms that begin earlier in childhood and a very rare form that strikes adults. The symptoms of these childhood types are similar to those caused by Batten disease, but they become apparent at different ages and progress at different rates.

Congenital NCL is a very rare and severe form of NCL; babies have microcephaly and seizures, and die soon after birth. Infantile NCL (INCL or Santavuori-Haltia disease) begins between about ages 6 months and 2 years and is marked by failure to thrive, microcephaly, and myoclonic contractions. These children usually die before age 5. Late infantile NCL (LINCL, or Jansky-Bielschowsky disease) begins between ages 2 and 4 and results in loss of muscle coordination (ataxia), seizures, and eventually death between ages 8 and 12. Adult NCL (Kufs disease, Parry's disease, and ANCL) usually manifests before age 40, causes milder symptoms that progress slowly, and does not cause blindness. Although age of death varies among affected individuals, this form does shorten life expectancy.

To date, no specific treatment is known that can halt or reverse the symptoms of NCLs. While enzyme replacement therapies have recently provided somewhat encouraging results in improving the neurological manifestations of some forms of Batten disease, they do not address the full spectrum of Batten disease manifestations, and their administration is associated with significant inconvenience. Stem cell therapies have yielded disappointing results. Most available treatments, only treat the symptoms of NCLs: seizures may be controlled with anticonvulsant drugs, physical and occupational therapy may help patients retain function as long as possible, and support and encouragement can help patients and families cope with the profound disability and dementia caused by NCLs.

Ultimately, due to the lack of effective treatments, NCL diseases are fatal.

BRIEF SUMMARY

The invention generally relates, in part, to gene therapy compositions and methods for the treatment, prevention, or amelioration of at least one symptom of neuronal ceroid lipofuscinoses.

In various embodiments, a polynucleotide is provided comprising: a left (5′) lentiviral LTR; a Psi (ψ) packaging signal; a retroviral export element; a central polypurine tract/DNA flap (cPPT/FLAP); a promoter operably linked to a polynucleotide encoding a tripeptidyl peptidase 1(TPP1) polypeptide; and a right (3′) lentiviral LTR.

In particular embodiments, the lentivirus is selected from the group consisting of: HIV (human immunodeficiency virus; including HIV type 1, and HIV type 2); visna-maedi virus (VMV) virus; 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 certain embodiments, the lentivirus is HIV-1 or HIV-2.

In some embodiments, the lentivirus is HIV-1.

In additional embodiments, the promoter of the 5′ LTR is replaced with a heterologous promoter selected from the group consisting of: a cytomegalovirus (CMV) promoter, a Rous Sarcoma Virus (RSV) promoter, and a Simian Virus 40 (SV40) promoter.

In further embodiments, the 3′ LTR comprises one or more modifications.

In some embodiments, the 3′ LTR comprises one or more deletions that prevent viral transcription beyond the first round of viral replication.

In particular embodiments, the 3′ LTR comprises a deletion of the TATA box and Sp1 and NF-κB transcription factor binding sites in the U3 region of the 3′ LTR.

In some embodiments, the 3′ LTR is a self-inactivating (SIN) LTR.

In certain embodiments, the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide comprises a myeloproliferative sarcoma virus enhancer, negative control region deleted, dl587rev primer-binding site substituted (MND) promoter or transcriptionally active fragment thereof.

In particular embodiments, the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide is selected from the group consisting of: integrin subunit alpha M (ITGAM; CD11b), CD68, C-X3-C motif chemokine receptor 1 (CX3CR1), ionized calcium binding adaptor molecule 1 (IBA1), transmembrane protein 119

(TMEM119), spalt like transcription factor 1 (SALL1) and adhesion G protein-coupled receptor E1 (F4/80).

In additional embodiments, the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide comprises an elongation factor 1 alpha (EF1α) promoter or transcriptionally active fragment thereof.

In particular embodiments, the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide is a short EF1α promoter.

In some embodiments, the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide is a long EF1α promoter.

In further embodiments, the polynucleotide encoding the TPP1 polypeptide is a cDNA.

In particular embodiments, the polynucleotide encoding the TPP1 polypeptide is codon optimized for expression.

In particular embodiments, a polynucleotide is provided, comprising: a left (5′) HIV-1 LTR; a Psi (ψ) packaging signal; an RRE retroviral export element; a cPPT/FLAP; an MND promoter or EF1α promoter operably linked to a polynucleotide encoding a TPP1 polypeptide; and a right (3′) HIV-1 LTR.

In particular embodiments, a polynucleotide is provided, comprising: a left (5′) CMV promoter/HIV-1 chimeric LTR; a Psi (ψ) packaging signal; an RRE retroviral export element; a cPPT/FLAP; an MND promoter or EF1α promoter operably linked to a polynucleotide encoding a TPP1 polypeptide; and a right (3′) SIN HIV-1 LTR.

In particular embodiments, the polynucleotide further comprise a bovine growth hormone polyadenylation signal or a rabbit β-globin polyadenylation signal.

In various embodiments, a mammalian cell transduced with a lentiviral vector is provided, comprising a polynucleotide contemplated herein.

In some embodiments, the cell is a hematopoietic cell.

In certain embodiments, the cell is a CD34+ cell.

In particular embodiments, the cell is a stem cell or progenitor cell.

In various embodiments, a producer cell comprising: a first polynucleotide encoding gag, a second polynucleotide encoding pol, a third polynucleotide encoding env, and a polynucleotide contemplated herein.

In various particular embodiments, a lentiviral vector produced by the producer cell contemplated herein is provided.

In various certain embodiments, a composition comprising a lentiviral vector comprising a polynucleotide or a mammalian cell contemplated herein is provided.

In various further embodiments, a pharmaceutical composition comprising a pharmaceutically acceptable carrier and a lentiviral vector comprising a polynucleotide or a mammalian cell contemplated herein is provided.

In various additional embodiments, a method of treating neuronal ceroid lipofuscinoses (NCL), comprising administering to a subject a lentiviral vector comprising a polynucleotide, a cell transduced with a lentiviral vector comprising a polynucleotide, or a mammalian cell contemplated herein is provided.

In various some embodiments, a method of treating neuronal ceroid lipofuscinoses, comprising administering to a subject a pharmaceutical composition contemplated herein is provided.

In various particular embodiments, a method of decreasing at least one symptom associated with neuronal ceroid lipofuscinoses in a subject comprising administering to a subject a lentiviral vector comprising a polynucleotide, a cell transduced with a lentiviral vector comprising a polynucleotide, or a mammalian cell contemplated herein is provided.

In various embodiments, a method of decreasing at least one symptom associated with neuronal ceroid lipofuscinoses in a subject is provided, comprising administering to a subject a pharmaceutical composition contemplated herein.

In some embodiments, at least one symptom is selected from the group consisting of: seizures, loss of vision, cognitive function decline, and motor function decline.

In particular embodiments, the subject has been diagnosed with late-infantile NCL (LINCL).

In certain embodiments, the subject has been diagnosed with juvenile Batten Disease (JNCL).

BRIEF DESCRIPTION OF SEVERAL VIEWS OF THE DRAWINGS

Figure 1 shows exemplary architectures of lentiviral vectors encoding TPP1.

Figure 2 shows a representative Western blot of TPP1 expression in wild type fibroblasts, $TPP1^{-/-}$ fibroblasts, and $TPP1^{-/-}$ fibroblasts transduced with pMND-TPP1 or pEF1α-TPP1. The lower arrow indicates the β-actin control protein at its expected size of ~42kDa in all samples. TPP-1 protein (upper arrow) was detected at low levels in the wild

type fibroblasts and was overexpressed in transduced $TPP1^{-/-}$ fibroblasts (right-most two lanes).

Figure 3A shows the data from a representative experiment assaying TPP1 enzymatic activity in wild type control cells, $TPP1^{-/-}$ cells, and $TPP1^{-/-}$ cells transduced with the lentiviral vectors encoding TPP1 (pMND-TPP1 and pEF1α-TPP1).

Figure 3B shows that $TPP1^{-/-}$ fibroblasts transduced with lentiviral vectors encoding TPP1 secrete about 10-fold more active TPP1 into cell culture supernatant compared to background levels assayed in wild type cells and untransduced $TPP1^{-/-}$ fibroblasts.

Figure 4 shows that F108 and PGE_2 increase transduction with lentiviral vectors encoding tripeptidyl peptidase 1 (TPP1). Human $CD34^+$ cells were transduced with LVVs encoding either EF1α-TPP1 or MND-TPP1 at an MOI of 5 or 15 with and without F108 and PGE_2. Day 14 pooled methylcellulose colony VCNs are shown in upper left panel. TPP-1 enzyme activity from day 7 cytokine cultures are shown in upper middle panel (cell pellet) and upper right panel (cell supernatant). Mouse Lineage negative (Lin-) bone marrow cells were transduced with an MND-TPP1 LVV at an MOI of 5, 15, or 30 with and without F108 and PGE_2. D7 liquid culture VCNs are shown in the lower left panel; individual colony VCN by qPCR is shown in the lower middle panel; and %LVV+ colonies are shown in the lower right panel.

Figure 5 shows that $hCD34^+$ cell growth was not adversely affected by transducing the cells with lentiviral vector in the presence of F108 and PGE_2.

Figure 6 shows that $hCD34^+$ cells transduced with a pMND-CLN2 LVV or pEF1αCLN2 LVV in the presence of F108 and PGE_2 and cultured for 14 days had increased VCNs across all MOIs for both lentiviral vectors.

Figure 7 shows that $hCD34^+$ cells transduced with a pMND-CLN2 LVV or pEF1αCLN2 LVV in the presence of F108 and PGE_2 and cultured for 14 days in methylcellulose had increased VCNs and %LVV+ cells when measured from individual colonies (left panel). Transduction did not significantly affect colony formation (right panel).

Figure 8 shows that $hCD34^+$ cells transduced with a pMND-CLN2 LVV or pEF1αCLN2 LVV in the presence of F108 and PGE_2 show increased TPP-1 activity.

Figure 9 shows confocal microscopy for TPP1 (to detect TPP1 expression) and ATP synthase subunit c (to detect TPP1 activity) in untranduced wild type induced pluripotent stem cells (iPSCs), untransduced patient-derived TPP1 $^{-/-}$ iPSCs, or patient-derived TPP $^{-/-}$ iPSCs transduced with lentiviral vectors (LVVs) comprising an MND or EF1α-short promoter operably linked to a polynucleotide encoding TPP1. TPP1 expression is present and subunit c expression is absent in patient-derived TPP1 $^{-/-}$ iPSCs transduced with LVVs encoding TPP1.

BRIEF DESCRIPTION OF THE SEQUENCE IDENTIFIERS

SEQ ID NO: 1 sets forth the sequence of an exemplary lentiviral vector encoding TTP1.

SEQ ID NO: 2 sets forth the sequence of an exemplary lentiviral vector encoding TTP1.

SEQ ID NO: 3 sets forth a polynucleotide sequence encoding a human tripeptidyl peptidase 1 (TPP1) polypeptide.

SEQ ID NO: 4 sets forth a codon-optimized polynucleotide sequence encoding a human tripeptidyl peptidase 1 (TPP1) polypeptide.

SEQ ID NO: 5 sets forth an amino acid sequence encoding a human tripeptidyl peptidase 1 (TPP1) polypeptide.

SEQ ID NOs: 6-16 set forth the amino acid sequences of various linkers.

SEQ ID NOs: 17-19 set forth the amino acid sequences of protease cleavage sites and self-cleaving polypeptide cleavage sites.

DETAILED DESCRIPTION

A. OVERVIEW

The invention generally relates, in part, to improved gene therapy compositions and methods for treating, preventing, or ameliorating at least one symptom of a neuronal ceroid lipofuscinoses (NCL).

NCLs are a group of inherited disorders of the nervous system for which there is no clinically approved curative treatment and for which palliative care is the only option. NCLs are lysosomal storage diseases, marked by a buildup of substances called lipofuscins

or lipopigments in membrane bound organelles called lysosomes. Lysosomes are compartments in the cell that normally digest and recycle different types of molecules. The accumulation of lipofuscins in lysosomes occurs in cells throughout the body, but neurons in the brain seem to be particularly vulnerable to the damage caused by lipofuscins. The progressive death of cells, especially in the central nervous system, leads to vision loss, seizures, and decline in motor function and cognitive ability in people with NCLs.

Late-infantile NCL (LINCL or CLN2) is a devastating childhood neuronal ceroid lipofuscinosis disease caused by deficiency in tripeptidyl peptidase 1 (TTP1). At least 100 mutations in the TPP1 gene have been found to cause LINCL. Most mutations in the TPP1 gene alter single amino acids in tripeptidyl peptidase 1, resulting in a severe decrease in enzymatic activity. The mutations IV5-1G>C and R508X, cause nearly 90 percent of cases of LINCL worldwide. Loss of TPP1 function leads to accumulation of lipofuscin deposits in the lysosomes causing cell damage and atrophy with cells of the nervous system most acutely affected. LINCL patients first experience symptoms between the ages of 2 and 4 years old, progressively loosing neurological function beginning with blindness, ataxia, dementia, and then death, usually before the second decade of life (Anderson *et al.*, 2013. *Biochimica et Biophysica Acta.* 1832:1807–1826).

TPP1 deficiency also causes a small percentage of juvenile Batten disease (JNCL). Children with JNCL develop progressive vision loss, intellectual and motor disability, speech difficulties, and seizures. Vision impairment is often the first noticeable sign of JNCL, beginning between the ages of 4 and 8 years. Vision loss tends to progress rapidly, eventually resulting in blindness. After vision impairment has begun, children with juvenile Batten disease experience the loss of previously acquired skills (developmental regression), usually beginning with the ability to speak in complete sentences. Affected children also have difficulty learning new information. In addition to the intellectual decline, affected children lose motor skills such as the ability to walk or sit. They also develop movement abnormalities that include rigidity or stiffness, slow or diminished movements (hypokinesia), and stooped posture. Affected children may have recurrent seizures (epilepsy), heart problems, behavioral problems, difficulty sleeping, and problems with attention that begin in mid- to late childhood. Most people with JNCL live into their twenties or thirties.

In various embodiments, a gene therapy vector encoding a tripeptidyl peptidase 1 (TPP1) polypeptide is contemplated. The gene therapy preferentially includes a promoter

operably linked to the polynucleotide encoding the TPP1 polypeptide. The gene therapy vector may be a viral vector, including but not limited to a gammaretroviral vector, a lentiviral vector, an adeno-associated viral (AAV) vector, an adenoviral vector, or a herpes virus vector.

Cells transduced with the gene therapy vectors contemplated herein are also provided in various embodiments. In some preferred embodiments, the transduced cells are hematopoietic cells, including, but not limited to $CD34^+$ cells.

In various other embodiments, gene therapy compositions contemplated herein are preferably administered to a subject that has a neuronal ceroid lipofuscinosis, more preferably a subject that has late-infantile NCL (LINCL) or juvenile Batten Disease (JNCL) or even more preferably, a subject that has one or more mutations in a TPP1 gene.

The practice of the particular embodiments will employ, unless indicated specifically to the contrary, conventional methods of chemistry, biochemistry, organic chemistry, molecular biology, microbiology, recombinant DNA techniques, genetics, immunology, and cell biology that are within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. *See e.g.,* Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (3rd Edition, 2001); Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); Maniatis *et al.*, *Molecular Cloning: A Laboratory Manual* (1982); Ausubel *et al.*, *Current Protocols in Molecular Biology* (John Wiley and Sons, updated July 2008); *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, Greene Pub. Associates and Wiley-Interscience; Glover, *DNA Cloning: A Practical Approach*, vol. I & II (IRL Press, Oxford, 1985); Anand, T*echniques for the Analysis of Complex Genomes*, (Academic Press, New York, 1992); T*ranscription and Translation* (B. Hames & S. Higgins, Eds., 1984); Perbal, *A Practical Guide to Molecular Cloning* (1984); Harlow and Lane, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998) *Current Protocols in Immunology* Q. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach and W. Strober, eds., 1991); *Annual Review of Immunology*; as well as monographs in journals such as *Advances in Immunology*.

B. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the

invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of particular embodiments, preferred embodiments of compositions, methods and materials are described herein. For the purposes of the present disclosure, the following terms are defined below.

The articles "a," "an," and "the" are used herein to refer to one or to more than one (*i.e.*, to at least one, or to one or more) of the grammatical object of the article. By way of example, "an element" means one element or one or more elements.

The use of the alternative (*e.g.*, "or") should be understood to mean either one, both, or any combination thereof of the alternatives.

The term "and/or" should be understood to mean either one, or both of the alternatives.

As used herein, the term "about" or "approximately" refers to a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that varies by as much as 15%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% or 1% to a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length. In one embodiment, the term "about" or "approximately" refers a range of quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length ± 15%, ± 10%, ± 9%, ± 8%, ± 7%, ± 6%, ± 5%, ± 4%, ± 3%, ± 2%, or ± 1% about a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length.

In one embodiment, a range, *e.g.*, 1 to 5, about 1 to 5, or about 1 to about 5, refers to each numerical value encompassed by the range. For example, in one non-limiting and merely illustrative embodiment, the range "1 to 5" is equivalent to the expression 1, 2, 3, 4, 5; or 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, or 5.0; or 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, or 5.0.

As used herein, the term "substantially" refers to a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that is 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or higher compared to a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length. In one embodiment, "substantially the same" refers to a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that produces an

effect, *e.g.*, a physiological effect, that is approximately the same as a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length.

Throughout this specification, unless the context requires otherwise, the words "comprise", "comprises" and "comprising" will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements. By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of." Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present. By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that no other elements are present that materially affect the activity or action of the listed elements.

Reference throughout this specification to "one embodiment," "an embodiment," "a particular embodiment," "a related embodiment," "a certain embodiment," "an additional embodiment," or "a further embodiment" or combinations thereof means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment. Thus, the appearances of the foregoing phrases in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments. It is also understood that the positive recitation of a feature in one embodiment, serves as a basis for excluding the feature in a particular embodiment.

By "enhance" or "promote," or "increase" or "expand" refers generally to the ability of the compositions and/or methods contemplated herein to elicit, cause, or produce higher numbers of transduced cells compared to the number of cells transduced by either vehicle or a control molecule/composition. In one embodiment, a hematopoietic stem or progenitor cell transduced with compositions and methods contemplated herein comprises an increase in the number of transduced cells compared to existing transduction compositions and methods. Increases in cell transduction, can be ascertained using methods known in the art, such as reporter assays, RT-PCR, and cell surface protein expression, among others. An "increased" or "enhanced" amount of transduction is typically a "statistically significant" amount, and may include an increase that is 1.1, 1.2,

1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30 or more times (*e.g.*, 500, 1000 times) (including all integers and decimal points in between and above 1, *e.g.*, 1.5, 1.6, 1.7. 1.8, *etc.*) the number of cells transduced by vehicle, a control composition, or other transduction method.

By "decrease" or "lower," or "lessen," or "reduce," or "abate" refers generally to compositions or methods that result in comparably fewer transduced cells compared to cells transduced with compositions and/or methods according to the present invention. A "decrease" or "reduced" amount of transduced cells is typically a "statistically significant" amount, and may include an decrease that is 1.1, 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30 or more times (*e.g.*, 500, 1000 times) (including all integers and decimal points in between and above 1, *e.g.*, 1.5, 1.6, 1.7. 1.8, *etc.*) the number of transduced cells (reference response) produced by compositions and/or methods according to the present invention.

By "maintain," or "preserve," or "maintenance," or "no change," or "no substantial change," or "no substantial decrease" refers generally to a physiological response that is comparable to a response caused by either vehicle, a control molecule/composition, or the response in a particular cell. A comparable response is one that is not significantly different or measurable different from the reference response.

In the following description, certain specific details are set forth in order to provide a thorough understanding of various illustrative embodiments of the invention contemplated herein. However, one skilled in the art will understand that particular illustrative embodiments may be practiced without these details. In addition, it should be understood that the individual vectors, or groups of vectors, derived from the various combinations of the structures and substituents described herein, are disclosed by the present application to the same extent as if each vector or group of vectors was set forth individually. Thus, selection of particular vector structures or particular substituents is within the scope of the present disclosure.

C. POLYPEPTIDES

"Polypeptide," "polypeptide fragment," "peptide" and "protein" are used interchangeably, unless specified to the contrary, and according to conventional meaning, *i.e.*, as a sequence of amino acids. In one embodiment, a "polypeptide" includes fusion polypeptides and other variants. Polypeptides can be prepared using any of a variety of well-known recombinant and/or synthetic techniques. Polypeptides are not limited to a specific length, *e.g.*, they may comprise a full length protein sequence, a fragment of a full

length protein, or a fusion protein, and may include post-translational modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

In various embodiments, polypeptides are contemplated herein, including, but not limited to, TPP1 polypeptides, *e.g.,* SEQ ID NO: 5.

An "isolated peptide" or an "isolated polypeptide" and the like, as used herein, refer to *in vitro* isolation and/or purification of a peptide or polypeptide molecule from a cellular environment, and from association with other components of the cell, *i.e.*, it is not significantly associated with *in vivo* substances.

Polypeptides include "polypeptide variants." Polypeptide variants may differ from a naturally occurring polypeptide in one or more amino acid substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more amino acids of the above polypeptide sequences. For example, in particular embodiments, it may be desirable to modulate the biological properties of a polypeptide by introducing one or more substitutions, deletions, additions and/or insertions into the polypeptide. In particular embodiments, polypeptides include polypeptide variants having at least about 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%,85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% amino acid identity to any of the reference sequences contemplated herein, typically where the variant maintains at least one biological activity of the reference sequence.

Polypeptides variants include biologically active "polypeptide fragments." As used herein, the term "biologically active fragment" or "minimal biologically active fragment" refers to a polypeptide fragment that retains at least 100%, at least 90%, at least 80%, at least 70%, at least 60%, at least 50%, at least 40%, at least 30%, at least 20%, at least 10%, or at least 5% of the naturally occurring polypeptide activity. Polypeptide fragments refer to a polypeptide, which can be monomeric or multimeric that has an amino-terminal deletion, a carboxyl-terminal deletion, and/or an internal deletion or substitution of one or more amino acids of a naturally-occurring or recombinantly-produced polypeptide. In certain embodiments, a polypeptide fragment can comprise an amino acid chain at least 5 to about 1700 amino acids long. It will be appreciated that in certain embodiments, fragments are at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,

28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700 or more amino acids long.

Illustrative examples of polypeptide fragments include catalytic domains and the like.

As noted above, polypeptides may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of a reference polypeptide can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. *See*, for example, Kunkel (1985, *Proc. Natl. Acad. Sci. USA.* 82: 488-492), Kunkel *et al.*, (1987, *Methods in Enzymol*, 154: 367-382), U.S. Pat. No. 4,873,192, Watson, J. D. *et al.*, (*Molecular Biology of the Gene*, Fourth Edition, Benjamin/Cummings, Menlo Park, Calif., 1987) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff *et al.*, (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.).

In certain embodiments, a variant will contain one or more conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. Modifications may be made in the structure of the polynucleotides and polypeptides contemplated in particular embodiments, polypeptides include polypeptides having at least about and still obtain a functional molecule that encodes a variant or derivative polypeptide with desirable characteristics. When it is desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, variant polypeptide, one skilled in the art, for example, can change one or more of the codons of the encoding DNA sequence.

Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological activity can be found using computer programs well known in the art, such as DNASTAR, DNA Strider, Geneious, Mac Vector, or Vector NTI software. Preferably, amino acid changes in the protein variants disclosed herein are conservative amino acid changes, *i.e.*, substitutions of similarly charged or uncharged

amino acids. A conservative amino acid change involves substitution of one of a family of amino acids which are related in their side chains. Naturally occurring amino acids are generally divided into four families: acidic (aspartate, glutamate), basic (lysine, arginine, histidine), non-polar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), and uncharged polar (glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine) amino acids. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In a peptide or protein, suitable conservative substitutions of amino acids are known to those of skill in this art and generally can be made without altering a biological activity of a resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (*see*, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Benjamin/Cummings Pub. Co., p.224).

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cysteine (+2.5); methionine (+1.9); alanine (+1.8); glycine (0.4); threonine (0.7); serine (0.8); tryptophan (0.9); tyrosine (1.3); proline (1.6); histidine (3.2); glutamate (3.5); glutamine (3.5); aspartate (3.5); asparagine (3.5); lysine (3.9); and arginine (4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.*, still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity.

As detailed in U.S. Patent No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0);

threonine (–0.4); proline (–0.5 ± 1); alanine (–0.5); histidine (–0.5); cysteine (–1.0); methionine (–1.3); valine (–1.5); leucine (–1.8); isoleucine (–1.8); tyrosine (–2.3); phenylalanine (–2.5); tryptophan (–3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions may be based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like.

Polypeptide variants further include glycosylated forms, aggregative conjugates with other molecules, and covalent conjugates with unrelated chemical moieties (*e.g.*, pegylated molecules). Covalent variants can be prepared by linking functionalities to groups which are found in the amino acid chain or at the N- or C-terminal residue, as is known in the art. Variants also include allelic variants, species variants, and muteins. Truncations or deletions of regions which do not affect functional activity of the proteins are also variants.

Polypeptides contemplated in particular embodiments include fusion polypeptides. In particular embodiments, fusion polypeptides and polynucleotides encoding fusion polypeptides are provided. Fusion polypeptides and fusion proteins refer to a polypeptide having at least two, three, four, five, six, seven, eight, nine, or ten polypeptide segments.

In another embodiment, two or more polypeptides can be expressed as a fusion protein that comprises one or more self-cleaving polypeptide sequences as disclosed elsewhere herein.

Fusion polypeptides can comprise one or more polypeptide domains or segments including, but are not limited to signal peptides, cell permeable peptide domains (CPP), DNA binding domains, nuclease domains, chromatin remodeling domains, histone modifying domains, epigenetic modifying domains, exodomains, extracellular ligand binding domains, antigen binding domains, transmembrane domains, intracellular signaling domains, multimerization domains, epitope tags (*e.g.*, maltose binding protein ("MBP"), glutathione S transferase (GST), HIS6, MYC, FLAG, V5, VSV-G, and HA), polypeptide

linkers, and polypeptide cleavage signals. Fusion polypeptides are typically linked C-terminus to N-terminus, although they can also be linked C-terminus to C-terminus, N-terminus to N-terminus, or N-terminus to C-terminus. In particular embodiments, the polypeptides of the fusion protein can be in any order. Fusion polypeptides or fusion proteins can also include conservatively modified variants, polymorphic variants, alleles, mutants, subsequences, and interspecies homologs, so long as the desired activity of the fusion polypeptide is preserved. Fusion polypeptides may be produced by chemical synthetic methods or by chemical linkage between the two moieties or may generally be prepared using other standard techniques. Ligated DNA sequences comprising the fusion polypeptide are operably linked to suitable transcriptional or translational control elements as disclosed elsewhere herein.

Fusion polypeptides may optionally comprises a linker that can be used to link the one or more polypeptides or domains within a polypeptide. A peptide linker sequence may be employed to separate any two or more polypeptide components by a distance sufficient to ensure that each polypeptide folds into its appropriate secondary and tertiary structures so as to allow the polypeptide domains to exert their desired functions

Exemplary linkers include, but are not limited to the following amino acid sequences: glycine polymers (G)n; glycine-serine polymers (G1-5S1-5)n, where n is an integer of at least one, two, three, four, or five; glycine-alanine polymers; alanine-serine polymers; GGG (SEQ ID NO: 6); DGGGS (SEQ ID NO: 7); TGEKP (SEQ ID NO: 8) (see *e.g.*, Liu *et al.*, PNAS 5525-5530 (1997)); GGRR (SEQ ID NO: 9) (Pomerantz *et al.* 1995, supra); (GGGGS)n wherein n = 1, 2, 3, 4 or 5 (SEQ ID NO: 10) (Kim *et al.*, PNAS 93, 1156-1160 (1996.); EGKSSGSGSESKVD (SEQ ID NO: 11) (Chaudhary *et al.*, 1990, Proc. Natl. Acad. Sci. U.S.A. 87:1066-1070); KESGSVSSEQLAQFRSLD (SEQ ID NO: 12) (Bird *et al.*, 1988, Science 242:423-426), GGRRGGGS (SEQ ID NO: 13); LRQRDGERP (SEQ ID NO: 14); LRQKDGGGSERP (SEQ ID NO: 15); LRQKd(GGGS)2ERP (SEQ ID NO: 16). Alternatively, flexible linkers can be rationally designed using a computer program capable of modeling both DNA-binding sites and the peptides themselves (Desjarlais & Berg, *PNAS* 90:2256-2260 (1993)) or by phage display methods.

Fusion polypeptides may further comprise a polypeptide cleavage signal between each of the polypeptide domains described herein or between an endogenous open reading frame and a polypeptide encoded by a donor repair template. In addition, a polypeptide

cleavage site can be put into any linker peptide sequence. Exemplary polypeptide cleavage signals include polypeptide cleavage recognition sites such as protease cleavage sites, nuclease cleavage sites (*e.g.*, rare restriction enzyme recognition sites, self-cleaving ribozyme recognition sites), and self-cleaving viral oligopeptides (*see* deFelipe and Ryan, 2004. *Traffic*, 5(8); 616-26).

Suitable protease cleavages sites and self-cleaving peptides are known to the skilled person (*see, e.g.*, in Ryan *et al.*, 1997. *J. Gener. Virol.* 78, 699-722; Scymczak *et al.* (2004) *Nature Biotech.* 5, 589-594). Exemplary protease cleavage sites include, but are not limited to the cleavage sites of potyvirus NIa proteases (*e.g.*, tobacco etch virus protease), potyvirus HC proteases, potyvirus P1 (P35) proteases, byovirus NIa proteases, byovirus RNA-2-encoded proteases, aphthovirus L proteases, enterovirus 2A proteases, rhinovirus 2A proteases, picorna 3C proteases, comovirus 24K proteases, nepovirus 24K proteases, RTSV (rice tungro spherical virus) 3C-like protease, PYVF (parsnip yellow fleck virus) 3C-like protease, heparin, thrombin, factor Xa and enterokinase. Due to its high cleavage stringency, TEV (tobacco etch virus) protease cleavage sites are preferred in one embodiment, *e.g.*, EXXYXQ(G/S) (SEQ ID NO: 17), for example, ENLYFQG (SEQ ID NO: 18) and ENLYFQS (SEQ ID NO: 19), wherein X represents any amino acid (cleavage by TEV occurs between Q and G or Q and S).

In certain embodiments, the self-cleaving polypeptide site comprises a 2A or 2A-like site, sequence or domain (Donnelly *et al.*, 2001. *J. Gen. Virol.* 82:1027-1041). In a particular embodiment, the viral 2A peptide is an aphthovirus 2A peptide, a potyvirus 2A peptide, or a cardiovirus 2A peptide.

In various embodiments, the expression or stability of polypeptides or fusion polypeptides contemplated herein is regulated by one or more protein destabilization sequences or protein degradation sequences (degrons). Several strategies to destabilize proteins to enforce their rapid proteasomal turnover are contemplated herein.

Illustrative examples of protein destabilization sequences include, but are not limited to: the destabilization box (D box), a nine amino acid is present in cell cycle-dependent proteins that must undergo rapid and complete ubiquitin-mediated proteolysis to achieve cycling within the cell cycle (see *e.g.*, Yamano *et al.* 1998. *Embo J* 17:5670-8); the KEN box, an APC recognition signal targeted by Cdh1 (see *e.g.*, Pfleger *et al.* 2000. *Genes Dev* 14:655-65); the O box, a motif present in origin recognition complex protein 1 (ORC1), which is degraded at the end of M phase and throughout much of G1 by anaphase-

promoting complexes (APC) activated by Fzr/Cdh1 (see *e.g.*, Araki *et al.* 2005. *Genes Dev* 19(20):2458-2465); the A-box, a motif present in Aurora-A, which is degraded during mitotic exit by Cdh1 (see *e.g.*, Littlepage *et al.* 2002. *Genes Dev* 16:2274-2285); PEST domains, motifs enriched in proline (P), glutamic acid (E), serine (S) and threonine (T) residues and that target proteins for rapid proteasomal destruction (Rechsteiner *et al.* 1996. *Trens Biochem Sci.* 21(7):267-271); N-end rule motifs, N-degron motifs, and ubiquitin-fusion degradation (UFD) motifs, which are rapidly processed for proteasomal destruction (*see e.g.*, Dantuma *et al.* 2000. *Nat Biotechnol* 18:538-4).

Further illustrative examples of degrons suitable for use in particular embodiments include, but are not limited to, ligand controllable degrons and temperature regulatable degrons. Non-limiting examples of ligand controllable degrons include those stabilized by Shield 1 (see *e.g.*, Bonger *et al.* 2011. *Nat Chem Viol.* 7(8):531-537), destabilized by auxin (see *e.g.*, Nishimura *et al.* 2009. *Nat Methods* 6(12):917-922), and stabilized by trimethoprim (see *e.g.*, Iwamoto *et al.*, 2010. *Chem Biol.* 17(9):981-8).

Non-limiting examples of temperature regulatable degrons include, but are not limited to DHFRTS degrons (see *e.g.*, Dohmen *et al.*, 1994. *Science* 263(5151):1273-1276).

In particular embodiments, a polypeptide contemplated herein comprises one or more degradation sequences selected from the group consisting of: a D box, an O box, an A box, a KEN motif, a PEST motifs, Cyclin A and UFD domain/substrates, ligand controllable degrons, and temperature regulatable degrons.

D. POLYNUCLEOTIDES

As used herein, the terms "polynucleotide" or "nucleic acid" refer to deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and DNA/RNA hybrids. Polynucleotides may be single-stranded or double-stranded and either recombinant, synthetic, or isolated. Polynucleotides include, but are not limited to: pre-messenger RNA (pre-mRNA), messenger RNA (mRNA), RNA, short interfering RNA (siRNA), short hairpin RNA (shRNA), microRNA (miRNA), ribozymes, synthetic RNA, genomic RNA (gRNA), plus strand RNA (RNA(+)), minus strand RNA (RNA(-)), tracrRNA, crRNA, single guide RNA (sgRNA), synthetic RNA, genomic DNA (gDNA), PCR amplified DNA, complementary DNA (cDNA), synthetic DNA, or recombinant DNA. Polynucleotides refer to a polymeric form of nucleotides of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 100, at least 200, at least

300, at least 400, at least 500, at least 1000, at least 5000, at least 10000, or at least 15000 or more nucleotides in length, either ribonucleotides or deoxyribonucleotides or a modified form of either type of nucleotide, as well as all intermediate lengths. It will be readily understood that "intermediate lengths," in this context, means any length between the quoted values, such as 6, 7, 8, 9, *etc.*, 101, 102, 103, *etc.*; 151, 152, 153, *etc.*; 201, 202, 203, *etc.* In particular embodiments, polynucleotides or variants have at least or about 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 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% sequence identity to a reference sequence.

In particular embodiments, polynucleotides may be codon-optimized. As used herein, the term "codon-optimized" refers to substituting codons in a polynucleotide encoding a polypeptide in order to increase the expression, stability and/or activity of the polypeptide. Factors that influence codon optimization include, but are not limited to one or more of: (i) variation of codon biases between two or more organisms or genes or synthetically constructed bias tables, (ii) variation in the degree of codon bias within an organism, gene, or set of genes, (iii) systematic variation of codons including context, (iv) variation of codons according to their decoding tRNAs, (v) variation of codons according to GC %, either overall or in one position of the triplet, (vi) variation in degree of similarity to a reference sequence for example a naturally occurring sequence, (vii) variation in the codon frequency cutoff, (viii) structural properties of mRNAs transcribed from the DNA sequence, (ix) prior knowledge about the function of the DNA sequences upon which design of the codon substitution set is to be based, and/or (x) systematic variation of codon sets for each amino acid.

Illustrative examples of polynucleotides include, but are not limited to polynucleotides sequences set forth in SEQ ID NOs: 1-4.

In various illustrative embodiments, polynucleotides contemplated herein include, but are not limited to polynucleotides comprising expression vectors, viral vectors, transfer plasmids, expression cassettes and polynucleotides encoding a tripeptidyl peptidase 1 (TPP1) polypeptide.

The tripeptidyl peptidase 1 (TPP1) gene encodes TPP1 (also referred to as CLN2, GIG1, LPIC, SCAR7, and TPP-1), a member of the sedolisin family of serine proteases. TPP1 encodes a 563-amino acid preproenzyme with a 19-amino acid signal peptide and a 176-amino acid prodomain that are removed during maturation, yielding a 368-amino acid

mature enzyme. TPP1 also contains 5 N-glycosylation sites. The proenzyme has an apparent molecular mass of 66 kD, and the mature enzyme has an apparent molecular mass of 46 to 48 kD. TPP1 is a lysosomal exopeptidase that sequentially cleaves N-terminal tripeptides from substrates, and has weaker endopeptidase activity. TPP1 is synthesized as a catalytically-inactive enzyme which is activated and auto-proteolyzed upon acidification. Mutations in the TPP1 gene result in late-infantile neuronal ceroid lipofuscinosis and a rare form of juvenile Batten disease, which are associated with the failure to degrade specific neuropeptides and a subunit of ATP synthase in the lysosome.

As used herein, the terms "polynucleotide variant" and "variant" and the like refer to polynucleotides displaying substantial sequence identity with a reference polynucleotide sequence or polynucleotides that hybridize with a reference sequence under stringent conditions. These terms also encompass polynucleotides that are distinguished from a reference polynucleotide by the addition, deletion, substitution, or modification of at least one nucleotide. Accordingly, the terms "polynucleotide variant" and "variant" include polynucleotides in which one or more nucleotides have been added or deleted, or modified, or replaced with different nucleotides. In this regard, it is well understood in the art that certain alterations inclusive of mutations, additions, deletions and substitutions can be made to a reference polynucleotide whereby the altered polynucleotide retains the biological function or activity of the reference polynucleotide.

The recitations "sequence identity" or, for example, comprising a "sequence 50% identical to," as used herein, refer to the extent that sequences are identical on a nucleotide-by-nucleotide basis or an amino acid-by-amino acid basis over a window of comparison. Thus, a "percentage of sequence identity" may be calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, I) or the identical amino acid residue (*e.g.*, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys and Met) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. Included are nucleotides and polypeptides having at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to any of the reference sequences described herein, typically where

the polypeptide variant maintains at least one biological activity of the reference polypeptide.

An "isolated polynucleotide," as used herein, refers to a polynucleotide that has been purified from the sequences which flank it in a naturally-occurring state, *e.g.*, a DNA fragment that has been removed from the sequences that are normally adjacent to the fragment. In particular embodiments, an "isolated polynucleotide" refers to a complementary DNA (cDNA), a recombinant polynucleotide, a synthetic polynucleotide, or other polynucleotide that does not exist in nature and that has been made by the hand of man.

Terms that describe the orientation of polynucleotides include: 5' (normally the end of the polynucleotide having a free phosphate group) and 3' (normally the end of the polynucleotide having a free hydroxyl (OH) group). Polynucleotide sequences can be annotated in the 5' to 3' orientation or the 3' to 5' orientation. For DNA and mRNA, the 5' to 3' strand is designated the "sense," "plus," or "coding" strand because its sequence is identical to the sequence of the pre-messenger (pre-mRNA) [except for uracil (U) in RNA, instead of thymine (T) in DNA]. For DNA and mRNA, the complementary 3' to 5' strand which is the strand transcribed by the RNA polymerase is designated as "template," "antisense," "minus," or "non-coding" strand. As used herein, the term "reverse orientation" refers to a 5' to 3' sequence written in the 3' to 5' orientation or a 3' to 5' sequence written in the 5' to 3' orientation.

The terms "complementary" and "complementarity" refer to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, the complementary strand of the DNA sequence 5' A G T C A T G 3' is 3' T C A G T A C 5'. The latter sequence is often written as the reverse complement with the 5' end on the left and the 3' end on the right, 5' C A T G A C T 3'. A sequence that is equal to its reverse complement is said to be a palindromic sequence. Complementarity can be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there can be "complete" or "total" complementarity between the nucleic acids.

The term "nucleic acid cassette" or "expression cassette" as used herein refers to genetic sequences within the vector which can express an RNA, and subsequently a polypeptide. In one embodiment, the nucleic acid cassette contains a gene(s)-of-interest, *e.g.*, a polynucleotide(s)-of-interest. In another embodiment, the nucleic acid cassette contains one or more expression control sequences, *e.g.*, a promoter, enhancer, poly(A)

sequence, and a gene(s)-of-interest, *e.g.*, a polynucleotide(s)-of-interest. Vectors may comprise one, two, three, four, five or more nucleic acid cassettes. The nucleic acid cassette is positionally and sequentially oriented within the vector such that the nucleic acid in the cassette can be transcribed into RNA, and when necessary, translated into a protein or a polypeptide, undergo appropriate post-translational modifications required for activity in the transformed cell, and be translocated to the appropriate compartment for biological activity by targeting to appropriate intracellular compartments or secretion into extracellular compartments. Preferably, the cassette has its 3' and 5' ends adapted for ready insertion into a vector, *e.g.*, it has restriction endonuclease sites at each end. In a preferred embodiment, the nucleic acid cassette contains the sequence of a therapeutic gene used to treat, prevent, or ameliorate a genetic disorder. The cassette can be removed and inserted into a plasmid or viral vector as a single unit.

As used herein, the term "polynucleotide(s)-of-interest" refers to one or more polynucleotides, *e.g.*, a polynucleotide encoding a polypeptide (*i.e.*, a polypeptide-of-interest), inserted into an expression vector that is desired to be expressed. In preferred embodiments, vectors and/or plasmids of the present invention comprise one or more polynucleotides-of-interest, *e.g.*, a polynucleotide encoding a TPP1 polypeptide. In certain embodiments, a polynucleotide-of-interest encodes a polypeptide that provides a therapeutic effect in the treatment, prevention, or amelioration of a neuronal ceroid lipofuscinoses, which may be referred to as a "therapeutic polypeptide," *e.g.*, a polynucleotide encoding a TPP1 polypeptide.

In a certain embodiment, a polynucleotide-of-interest comprises an inhibitory polynucleotide including, but not limited to, a crRNA, a tracrRNA, a single guide RNA (sgRNA), an siRNA, an miRNA, an shRNA, a ribozyme or another inhibitory RNA.

Polynucleotides, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters and/or enhancers, untranslated regions (UTRs), Kozak sequences, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, internal ribosomal entry sites (IRES), recombinase recognition sites (*e.g.*, LoxP, FRT, and Att sites), termination codons, transcriptional termination signals, post-transcription response elements, and polynucleotides encoding self-cleaving polypeptides, epitope tags, as disclosed elsewhere herein or as known in the art, such that their overall length may vary considerably. It is therefore contemplated that a polynucleotide fragment of almost any length may be employed, with the total length

preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

Polynucleotides can be prepared, manipulated, expressed and/or delivered using any of a variety of well-established techniques known and available in the art. In order to express a desired polypeptide, a nucleotide sequence encoding the polypeptide, can be inserted into appropriate vector.

Illustrative examples of vectors include, but are not limited to plasmid, autonomously replicating sequences, and transposable elements, *e.g.*, Sleeping Beauty, PiggyBac.

Additional illustrative examples of vectors include, without limitation, plasmids, phagemids, cosmids, artificial chromosomes such as yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC), or P1-derived artificial chromosome (PAC), bacteriophages such as lambda phage or M13 phage, and animal viruses.

Illustrative examples of viruses useful as vectors include, without limitation, retrovirus (including lentivirus), adenovirus, adeno-associated virus, herpesvirus (*e.g.*, herpes simplex virus), poxvirus, baculovirus, papillomavirus, and papovavirus (*e.g.*, SV40).

Illustrative examples of expression vectors include, but are not limited to pClneo vectors (Promega) for expression in mammalian cells; pLenti4/V5-DEST™, pLenti6/V5-DEST™, and pLenti6.2/V5-GW/lacZ (Invitrogen) for lentivirus-mediated gene transfer and expression in mammalian cells. In particular embodiments, coding sequences of polypeptides disclosed herein can be ligated into such expression vectors for the expression of the polypeptides in mammalian cells.

In particular embodiments, the vector is an episomal vector or a vector that is maintained extrachromosomally. As used herein, the term "episomal" refers to a vector that is able to replicate without integration into host's chromosomal DNA and without gradual loss from a dividing host cell also meaning that said vector replicates extrachromosomally or episomally.

"Expression control sequences," "control elements," or "regulatory sequences" present in an expression vector are those non-translated regions of the vector—origin of replication, selection cassettes, promoters, enhancers, translation initiation signals (Shine Dalgarno sequence or Kozak sequence) introns, post-transcriptional regulatory elements, a polyadenylation sequence, 5' and 3' untranslated regions—which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength

and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including ubiquitous promoters and inducible promoters may be used.

In particular embodiments, a polynucleotide is a vector, including but not limited to expression vectors and viral vectors, and includes exogenous, endogenous, or heterologous control sequences such as promoters and/or enhancers. An "endogenous" control sequence is one which is naturally linked to a given gene in the genome. An "exogenous" control sequence is one which is placed in juxtaposition to a gene by means of genetic manipulation (*i.e.*, molecular biological techniques) such that transcription of that gene is directed by the linked enhancer/promoter. A "heterologous" control sequence is an exogenous sequence that is from a different species than the cell being genetically manipulated. A "synthetic" control sequence may comprise elements of one more endogenous and/or exogenous sequences, and/or sequences determined *in vitro* or *in silico* that provide optimal promoter and/or enhancer activity for the particular gene therapy.

The term "promoter" as used herein refers to a recognition site of a polynucleotide (DNA or RNA) to which an RNA polymerase binds. An RNA polymerase initiates and transcribes polynucleotides operably linked to the promoter. In particular embodiments, promoters operative in mammalian cells comprise an AT-rich region located approximately 25 to 30 bases upstream from the site where transcription is initiated and/or another sequence found 70 to 80 bases upstream from the start of transcription, a CNCAAT region where N may be any nucleotide.

The term "enhancer" refers to a segment of DNA which contains sequences capable of providing enhanced transcription and in some instances can function independent of their orientation relative to another control sequence. An enhancer can function cooperatively or additively with promoters and/or other enhancer elements. The term "promoter/enhancer" refers to a segment of DNA which contains sequences capable of providing both promoter and enhancer functions.

The term "operably linked", refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. In one embodiment, the term refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, and/or enhancer) and a second polynucleotide sequence, *e.g.*, a polynucleotide-of-interest, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

As used herein, the term "constitutive expression control sequence" refers to a promoter, enhancer, or promoter/enhancer that continually or continuously allows for transcription of an operably linked sequence. A constitutive expression control sequence may be a "ubiquitous" promoter, enhancer, or promoter/enhancer that allows expression in a wide variety of cell and tissue types or a "cell specific," "cell type specific," "cell lineage specific," or "tissue specific" promoter, enhancer, or promoter/enhancer that allows expression in a restricted variety of cell and tissue types, respectively.

Illustrative ubiquitous expression control sequences suitable for use in particular embodiments include, but are not limited to, a cytomegalovirus (CMV) immediate early promoter, a viral simian virus 40 (SV40) (*e.g.*, early or late), a Moloney murine leukemia virus (MoMLV) LTR promoter, a Rous sarcoma virus (RSV) LTR, a herpes simplex virus (HSV) (thymidine kinase) promoter, H5, P7.5, and P11 promoters from vaccinia virus, a short elongation factor 1-alpha (EF1a-short) promoter, a long elongation factor 1-alpha (EF1a-long) promoter, early growth response 1 (EGR1), ferritin H (FerH), ferritin L (FerL), Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), eukaryotic translation initiation factor 4A1 (EIF4A1), heat shock 70kDa protein 5 (HSPA5), heat shock protein 90kDa beta, member 1 (HSP90B1), heat shock protein 70kDa (HSP70), β-kinesin (β-KIN), the human ROSA 26 locus (Irions *et al.*, Nature Biotechnology 25, 1477 - 1482 (2007)), a Ubiquitin C promoter (UBC), a phosphoglycerate kinase-1 (PGK) promoter, a cytomegalovirus enhancer/chicken β-actin (CAG) promoter, a β-actin promoter and a myeloproliferative sarcoma virus enhancer, negative control region deleted, dl587rev primer-binding site substituted (MND) promoter (Challita *et al.*, J Virol. 69(2):748-55 (1995)).

In a particular embodiment, it may be desirable to use a cell, cell type, cell lineage or tissue specific expression control sequence to achieve cell type specific, lineage specific, or tissue specific expression of a desired polynucleotide sequence (*e.g.*, to express a particular nucleic acid encoding a polypeptide in only a subset of cell types, cell lineages, or tissues or during specific stages of development).

Illustrative examples of tissue specific promoters include, but are not limited to: an B29 promoter (B cell expression), a runt transcription factor (CBFa2) promoter (stem cell specific expression), an CD14 promoter (monocytic cell expression), an CD43 promoter (leukocyte and platelet expression), an CD45 promoter (hematopoietic cell expression), an CD68 promoter (macrophage expression), a CYP450 3A4 promoter (hepatocyte

expression), an desmin promoter (muscle expression), an elastase 1 promoter (pancreatic acinar cell expression, an endoglin promoter (endothelial cell expression), a fibroblast specific protein 1 promoter (FSP1) promoter (fibroblast cell expression), a fibronectin promoter (fibroblast cell expression), a fms-related tyrosine kinase 1 (FLT1) promoter (endothelial cell expression), a glial fibrillary acidic protein (GFAP) promoter (astrocyte expression), an insulin promoter (pancreatic beta cell expression), an integrin, alpha 2b (ITGA2B) promoter (megakaryocytes), an intracellular adhesion molecule 2 (ICAM-2) promoter (endothelial cells), an interferon beta (IFN-β) promoter (hematopoietic cells), a keratin 5 promoter (keratinocyte expression), a myoglobin (MB) promoter (muscle expression), a myogenic differentiation 1 (MYOD1) promoter (muscle expression), a nephrin promoter (podocyte expression), a bone gamma-carboxyglutamate protein 2 (OG-2) promoter (osteoblast expression), an 3-oxoacid CoA transferase 2B (Oxct2B) promoter, (haploid-spermatid expression), a surfactant protein B (SP-B) promoter (lung expression), a synapsin promoter (neuron expression), a Wiskott-Aldrich syndrome protein (WASP) promoter (hematopoietic cell expression).

Other illustrative examples of expression control sequences suitable for use in particular embodiments contemplated herein include, but are not limited to expression control sequences isolated from: integrin subunit alpha M (ITGAM; CD11b), CD68, C-X3-C motif chemokine receptor 1 (CX3CR1), ionized calcium binding adaptor molecule 1 (IBA1), transmembrane protein 119 (TMEM119), spalt like transcription factor 1 (SALL1) and adhesion G protein-coupled receptor E1 (F4/80).

As used herein, "conditional expression" may refer to any type of conditional expression including, but not limited to, inducible expression; repressible expression; expression in cells or tissues having a particular physiological, biological, or disease state, *etc.* This definition is not intended to exclude cell type or tissue specific expression. Certain embodiments provide conditional expression of a polynucleotide-of-interest, *e.g.*, expression is controlled by subjecting a cell, tissue, organism, *etc.*, to a treatment or condition that causes the polynucleotide to be expressed or that causes an increase or decrease in expression of the polynucleotide encoded by the polynucleotide-of-interest.

Illustrative examples of inducible promoters/systems include, but are not limited to, steroid-inducible promoters such as promoters for genes encoding glucocorticoid or estrogen receptors (inducible by treatment with the corresponding hormone), metallothionine promoter (inducible by treatment with various heavy metals), MX-1

promoter (inducible by interferon), the "GeneSwitch" mifepristone-regulatable system (Sirin *et al.*, 2003, *Gene*, 323:67), the cumate inducible gene switch (WO 2002/088346), tetracycline-dependent regulatory systems, *etc.*

Conditional expression can also be achieved by using a site specific DNA recombinase. According to certain embodiments, polynucleotides comprises at least one (typically two) site(s) for recombination mediated by a site specific recombinase. As used herein, the terms "recombinase" or "site specific recombinase" include excisive or integrative proteins, enzymes, co-factors or associated proteins that are involved in recombination reactions involving one or more recombination sites (*e.g.*, two, three, four, five, six, seven, eight, nine, ten or more.), which may be wild-type proteins (*see* Landy, *Current Opinion in Biotechnology* 3:699-707 (1993)), or mutants, derivatives (*e.g.*, fusion proteins containing the recombination protein sequences or fragments thereof), fragments, and variants thereof. Illustrative examples of recombinases suitable for use in particular embodiments include, but are not limited to: Cre, Int, IHF, Xis, Flp, Fis, Hin, Gin, ΦC31, Cin, Tn3 resolvase, TndX, XerC, XerD, TnpX, Hjc, Gin, SpCCE1, and ParA.

The polynucleotides may comprise one or more recombination sites for any of a wide variety of site specific recombinases. It is to be understood that the target site for a site specific recombinase is in addition to any site(s) required for integration of a vector, *e.g.*, a retroviral vector or lentiviral vector. As used herein, the terms "recombination sequence," "recombination site," or "site specific recombination site" refer to a particular nucleic acid sequence to which a recombinase recognizes and binds.

For example, one recombination site for Cre recombinase is loxP which is a 34 base pair sequence comprising two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8 base pair core sequence (*see* FIG. 1 of Sauer, B., *Current Opinion in Biotechnology* 5:521-527 (1994)). Other exemplary loxP sites include, but are not limited to: lox511 (Hoess *et al.*, 1996; Bethke and Sauer, 1997), lox5171 (Lee and Saito, 1998), lox2272 (Lee and Saito, 1998), m2 (Langer *et al.*, 2002), lox71 (Albert *et al.*, 1995), and lox66 (Albert *et al.*, 1995).

Suitable recognition sites for the FLP recombinase include, but are not limited to: FRT (McLeod, *et al.*, 1996), F1, F2, F3 (Schlake and Bode, 1994), F4, F5 (Schlake and Bode, 1994), FRT(LE) (Senecoff *et al.*, 1988), FRT(RE) (Senecoff *et al.*, 1988).

Other examples of recognition sequences are the attB, attP, attL, and attR sequences, which are recognized by the recombinase enzyme λ Integrase, *e.g.*, phi-c31.

The φC31 SSR mediates recombination only between the heterotypic sites attB (34 bp in length) and attP (39 bp in length) (Groth *et al.*, 2000). attB and attP, named for the attachment sites for the phage integrase on the bacterial and phage genomes, respectively, both contain imperfect inverted repeats that are likely bound by φC31 homodimers (Groth *et al.*, 2000). The product sites, attL and attR, are effectively inert to further φC31-mediated recombination (Belteki *et al.*, 2003), making the reaction irreversible. For catalyzing insertions, it has been found that attB-bearing DNA inserts into a genomic attP site more readily than an attP site into a genomic attB site (Thyagarajan *et al.*, 2001; Belteki *et al.*, 2003). Thus, typical strategies position by homologous recombination an attP-bearing "docking site" into a defined locus, which is then partnered with an attB-bearing incoming sequence for insertion.

In particular embodiments, to achieve efficient translation of each of the plurality of polypeptides, the polynucleotide sequences can be separated by one or more IRES sequences or polynucleotide sequences encoding self-cleaving polypeptides.

As used herein, an "internal ribosome entry site" or "IRES" refers to an element that promotes direct internal ribosome entry to the initiation codon, such as ATG, of a cistron (a protein encoding region), thereby leading to the cap-independent translation of the gene. See, *e.g.*, Jackson *et al.*, 1990. *Trends Biochem Sci* 15(12):477-83) and Jackson and Kaminski. 1995. *RNA* 1(10):985-1000. Examples of IRES generally employed by those of skill in the art include those described in U.S. Pat. No. 6,692,736. Further examples of "IRES" known in the art include, but are not limited to IRES obtainable from picornavirus (Jackson *et al.*, 1990) and IRES obtainable from viral or cellular mRNA sources, such as for example, immunoglobulin heavy-chain binding protein (BiP), the vascular endothelial growth factor (VEGF) (Huez *et al.* 1998. *Mol. Cell. Biol.* 18(11):6178-6190), the fibroblast growth factor 2 (FGF-2), and insulin-like growth factor (IGFII), the translational initiation factor eIF4G and yeast transcription factors TFIID and HAP4, the encephelomycarditis virus (EMCV) which is commercially available from Novagen (Duke *et al.*, 1992. *J. Virol* 66(3):1602-9) and the VEGF IRES (Huez *et al.*, 1998. *Mol Cell Biol* 18(11):6178-90). IRES have also been reported in viral genomes of Picornaviridae, Dicistroviridae and Flaviviridae species and in HCV, Friend murine leukemia virus (FrMLV) and Moloney murine leukemia virus (MoMLV).

In one embodiment, the IRES used in polynucleotides contemplated herein is an EMCV IRES.

In particular embodiments, the polynucleotides comprise polynucleotides that have a consensus Kozak sequence and that encode a desired polypeptide. As used herein, the term "Kozak sequence" refers to a short nucleotide sequence that greatly facilitates the initial binding of mRNA to the small subunit of the ribosome and increases translation. The consensus Kozak sequence is (GCC)RCCATGG [SEQ ID NO:20], where R is a purine (A or G) (Kozak, 1986. *Cell.* 44(2):283-92, and Kozak, 1987. *Nucleic Acids Res.* 15(20):8125-48).

Elements directing the efficient termination and polyadenylation of the heterologous nucleic acid transcripts increases heterologous gene expression. Transcription termination signals are generally found downstream of the polyadenylation signal. In particular embodiments, vectors comprise a polyadenylation sequence 3′ of a polynucleotide encoding a polypeptide to be expressed. The term "polyA site" or "polyA sequence" as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript by RNA polymerase II. Polyadenylation sequences can promote mRNA stability by addition of a polyA tail to the 3′ end of the coding sequence and thus, contribute to increased translational efficiency. Efficient polyadenylation of the recombinant transcript is desirable as transcripts lacking a polyA tail are unstable and are rapidly degraded. Illustrative examples of polyA signals that can be used in a vector, includes an ideal polyA sequence (*e.g.*, AATAAA, ATTAAA, AGTAAA), a bovine growth hormone polyA sequence (BGHpA), a rabbit β-globin polyA sequence (rβgpA), or another suitable heterologous or endogenous polyA sequence known in the art.

In some embodiments, a polynucleotide or cell harboring the polynucleotide utilizes a suicide gene, including an inducible suicide gene to reduce the risk of direct toxicity and/or uncontrolled proliferation. In specific embodiments, the suicide gene is not immunogenic to the host harboring the polynucleotide or cell. A certain example of a suicide gene that may be used is caspase-9 or caspase-8 or cytosine deaminase. Caspase-9 can be activated using a specific chemical inducer of dimerization (CID).

In certain embodiments, polynucleotides comprise gene segments that cause the genetically modified cells contemplated herein to be susceptible to negative selection *in vivo*. "Negative selection" refers to an infused cell that can be eliminated as a result of a change in the *in vivo* condition of the individual. The negative selectable phenotype may result from the insertion of a gene that confers sensitivity to an administered agent, for

example, a compound. Negative selection genes are known in the art, and include, but are not limited to: the Herpes simplex virus type I thymidine kinase (HSV-I TK) gene which confers ganciclovir sensitivity; the cellular hypoxanthine phosphribosyltransferase (HPRT) gene, the cellular adenine phosphoribosyltransferase (APRT) gene, and bacterial cytosine deaminase.

In some embodiments, genetically modified cells comprise a polynucleotide further comprising a positive marker that enables the selection of cells of the negative selectable phenotype *in vitro*. The positive selectable marker may be a gene, which upon being introduced into the host cell, expresses a dominant phenotype permitting positive selection of cells carrying the gene. Genes of this type are known in the art, and include, but are not limited to hygromycin-B phosphotransferase gene (hph) which confers resistance to hygromycin B, the amino glycoside phosphotransferase gene (neo or aph) from Tn5 which codes for resistance to the antibiotic G418, the dihydrofolate reductase (DHFR) gene, the adenosine deaminase gene (ADA), and the multi-drug resistance (MDR) gene.

In one embodiment, the positive selectable marker and the negative selectable element are linked such that loss of the negative selectable element necessarily also is accompanied by loss of the positive selectable marker. In a particular embodiment, the positive and negative selectable markers are fused so that loss of one obligatorily leads to loss of the other. An example of a fused polynucleotide that yields as an expression product a polypeptide that confers both the desired positive and negative selection features described above is a hygromycin phosphotransferase thymidine kinase fusion gene (HyTK). Expression of this gene yields a polypeptide that confers hygromycin B resistance for positive selection *in vitro*, and ganciclovir sensitivity for negative selection *in vivo*. See also the publications of PCT US91/08442 and PCT/US94/05601, by S. D. Lupton, describing the use of bifunctional selectable fusion genes derived from fusing a dominant positive selectable markers with negative selectable markers.

Preferred positive selectable markers are derived from genes selected from the group consisting of hph, nco, and gpt, and preferred negative selectable markers are derived from genes selected from the group consisting of cytosine deaminase, HSV-I TK, VZV TK, HPRT, APRT and gpt. Exemplary bifunctional selectable fusion genes contemplated in particular embodiments include, but are not limited to genes wherein the positive selectable marker is derived from hph or neo, and the negative selectable marker is derived from cytosine deaminase or a TK gene or selectable marker.

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 in a cell, or may include sequences sufficient to allow integration into host cell DNA. Illustrative examples of vectors include, but are not limited to plasmids (*e.g.*, DNA plasmids or RNA plasmids), transposons, cosmids, bacterial artificial chromosomes, and viral vectors.

Illustrative methods of delivering polynucleotides contemplated in particular embodiments include, but are not limited to: electroporation, sonoporation, lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, nanoparticles, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, DEAE-dextran-mediated transfer, gene gun, and heat-shock.

Illustrative examples of polynucleotide delivery systems suitable for use in particular embodiments contemplated in particular embodiments include, but are not limited to those provided by Amaxa Biosystems, Maxcyte, Inc., BTX Molecular Delivery Systems, and Copernicus Therapeutics Inc. Lipofection reagents are sold commercially (*e.g.,* Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides have been described in the literature. See *e.g.,* Liu *et al.* (2003) *Gene Therapy.* 10:180–187; and Balazs *et al.* (2011) *Journal of Drug Delivery.* 2011:1-12. Antibody-targeted, bacterially derived, non-living nanocell-based delivery is also contemplated in particular embodiments.

In preferred embodiments, polynucleotides encoding one or more therapeutic polypeptides, or fusion polypeptides may be introduced into a target cell by viral methods.

E. VIRAL VECTORS

Polynucleotides encoding one or more therapeutic polypeptides, or fusion polypeptides may be introduced into a target cell by non-viral or viral methods. In particular embodiments, polynucleotides encoding a TPP1 polypeptide are introduced into a target cell using a vector, preferably a viral vector, more preferably a retroviral vector, and even more preferably, a lentiviral vector.

As will be evident to one of skill in the art, the term "viral vector" is widely used to refer either to a nucleic acid molecule (*e.g.,* a transfer plasmid) that includes virus-derived

nucleic acid elements that typically facilitate transfer of the nucleic acid molecule or integration into the genome of a cell or to a virus or viral particle that mediates nucleic acid transfer. Viral particles will typically include various viral components and sometimes also host cell components in addition to nucleic acid(s).

Illustrative examples of viral vector systems suitable for use in particular embodiments contemplated in particular embodiments include, but are not limited to adeno-associated virus (AAV), retrovirus, herpes simplex virus, adenovirus, vaccinia virus vectors for gene transfer.

Retroviruses are a common tool for gene delivery (Miller, 2000, *Nature*. 357: 455-460). As used herein, the term "retrovirus" refers to an RNA virus that reverse transcribes its genomic RNA into a linear double-stranded DNA copy and subsequently covalently integrates its genomic DNA into a host genome. Once the virus is integrated into the host genome, it is referred to as a "provirus." The provirus serves as a template for RNA polymerase II and directs the expression of RNA molecules which encode the structural proteins and enzymes needed to produce new viral particles.

Illustrative retroviruses suitable for use in particular embodiments, include, but are not limited to: Moloney murine leukemia virus (M-MuLV), Moloney murine sarcoma virus (MoMSV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), gibbon ape leukemia virus (GaLV), feline leukemia virus (FLV), spumavirus, Friend murine leukemia virus, Murine Stem Cell Virus (MSCV) and Rous Sarcoma Virus (RSV)) and lentivirus.

As used herein, the term "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. In particular embodiments, a lentivirus is used to deliver a polynucleotide encoding a TPP1 polypeptide to a cell.

The term viral vector may refer either to a virus or viral particle capable of transferring a nucleic acid into a cell or to the transferred nucleic acid itself. Viral vectors and transfer plasmids contain structural and/or functional genetic elements that are primarily derived from a virus. The term "retroviral vector" refers to a viral vector or

plasmid containing structural and functional genetic elements, or portions thereof, that are primarily derived from a retrovirus. The term "lentiviral vector" refers to a viral vector or plasmid containing structural and functional genetic elements, or portions thereof, including LTRs that are primarily derived from a lentivirus. The term "hybrid vector" 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.

In particular embodiments, the terms "lentiviral vector," "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 and are present in DNA form in the DNA plasmids.

In various embodiments, a lentiviral vector contemplated herein comprises one or more LTRs, and one or more, or all, of the following accessory elements: a cPPT/FLAP, a Psi (Ψ) packaging signal, an export element, a promoter operably linked to a polynucleotide encoding a TPP1 polypeptide, a poly (A) sequence, and may optionally comprise a WPRE or HPRE, an insulator element, a selectable marker, and a cell suicide gene, as discussed elsewhere herein.

In particular embodiments, lentiviral vectors contemplated herein may be integrative or non-integrating or integration defective lentivirus. As used herein, the term "integration defective lentivirus" or "refers to a lentivirus having an integrase that lacks the capacity to integrate the viral genome into the genome of the host cells. Integration-incompetent viral vectors have been described in patent application WO 2006/010834, which is herein incorporated by reference in its entirety.

Illustrative mutations in the HIV-1 pol gene suitable to reduce integrase activity include, but are not limited to: H12N, H12C, H16C, H16V, S81 R, D41A, K42A, H51A, Q53C, D55V, D64E, D64V, E69A, K71A, E85A, E87A, D116N, D1161, D116A, N120G, N1201, N120E, E152G, E152A, D35E, K156E, K156A, E157A, K159E, K159A, K160A, R166A, D167A, E170A, H171A, K173A, K186Q, K186T, K188T, E198A, R199c, R199T, R199A, D202A, K211A, Q214L, Q216L, Q221 L, W235F, W235E, K236S, K236A, K246A, G247W, D253A, R262A, R263A and K264H.

The term "long terminal repeat (LTR)" refers to domains of base pairs located at the ends of retroviral DNAs which, in their natural sequence context, are direct repeats and contain U3, R and U5 regions. The LTR contains numerous regulatory signals including transcriptional control elements, polyadenylation signals and sequences needed for replication and integration of the viral genome. Adjacent to the 5' LTR are sequences necessary for reverse transcription of the genome (the tRNA primer binding site) and for efficient packaging of viral RNA into particles (the Psi site).

As used herein, the term "packaging signal" or "packaging sequence," "psi" and the symbol "Ψ," refers to non-coding sequences located within the retroviral genome which are required for encapsidation of retroviral RNA strands during viral particle formation, *see e.g.,* Clever *et al.,* 1995. *J. of Virology*, Vol. 69, No. 4; pp. 2101–2109.

Lentiviral vectors preferably contain several safety enhancements as a result of modifying the LTRs. "Self-inactivating" (SIN) vectors refers to replication-defective vectors, *e.g.,* in which the right (3') LTR enhancer-promoter region, known as the U3 region, has been modified (*e.g.,* by deletion or substitution) to prevent viral transcription beyond the first round of viral replication. In a further embodiment, the 3' LTR is modified such that the U5 region is replaced, for example, with an ideal poly(A) sequence. An additional safety enhancement is provided by replacing the U3 region of the 5' LTR with a heterologous promoter to drive transcription of the viral genome during production of viral particles. Examples of heterologous promoters which can be used include, for example, viral simian virus 40 (SV40) (*e.g.,* early or late), cytomegalovirus (CMV) (*e.g.,* immediate early), Moloney murine leukemia virus (MoMLV), Rous sarcoma virus (RSV), and herpes simplex virus (HSV) (thymidine kinase) promoters. Typical promoters are able to drive high levels of transcription in a Tat-independent manner. This replacement reduces the possibility of recombination to generate replication-competent virus because there is no complete U3 sequence in the virus production system. It should be noted that modifications to the LTRs such as modifications to the 3' LTR, the 5' LTR, or both 3' and 5' LTRs, are also included.

As used herein, the term "FLAP element" or "cPPT/FLAP" refers to a nucleic acid whose sequence includes the central polypurine tract and central termination sequences (cPPT and CTS) of a retrovirus, *e.g.,* HIV-1 or HIV-2. Suitable FLAP elements are described in U.S. Pat. No. 6,682,907 and in Zennou, *et al.,* 2000, *Cell*, 101:173. During HIV-1 reverse transcription, central initiation of the plus-strand DNA at the central

polypurine tract (cPPT) and central termination at the central termination sequence (CTS) lead to the formation of a three-stranded DNA structure: the HIV-1 central DNA flap. While not wishing to be bound by any theory, the DNA flap may act as a cis-active determinant of lentiviral genome nuclear import and/or may increase the titer of the virus. In particular embodiments, the retroviral or lentiviral vector backbones comprise one or more FLAP elements upstream or downstream of the heterologous genes of interest in the vectors. For example, in particular embodiments a transfer plasmid includes a FLAP element. In one embodiment, a vector comprises a FLAP element isolated from HIV-1. In another embodiment, a lentiviral vector contains a FLAP element with one or more mutations in the cPPT and/or CTS elements. In yet another embodiment, a lentiviral vector comprises either a cPPT or CTS element. In yet another embodiment, a lentiviral vector does not comprise a cPPT or CTS element.

The term "export element" refers to a cis-acting post-transcriptional regulatory element which regulates the transport of an RNA transcript from the nucleus to the cytoplasm of a cell. Examples of RNA export elements include, but are not limited to, the human immunodeficiency virus (HIV) rev response element (RRE) (*see e.g.,* Cullen *et al.*, 1991. *J. Virol.* 65: 1053; and Cullen *et al.,* 1991. *Cell* 58: 423), and the hepatitis B virus post-transcriptional regulatory element (HPRE).

In particular embodiments, expression of heterologous sequences in viral vectors is increased by incorporating posttranscriptional regulatory elements, efficient polyadenylation sites, and optionally, transcription termination signals into the vectors. A variety of posttranscriptional regulatory elements can increase expression of a heterologous nucleic acid at the protein, *e.g.,* woodchuck hepatitis virus posttranscriptional regulatory element (WPRE; Zufferey *et al.*, 1999, *J. Virol.*, 73:2886); the posttranscriptional regulatory element present in hepatitis B virus (HPRE) (Huang *et al., Mol. Cell. Biol.*, 5:3864); and the like (Liu *et al.,* 1995, *Genes Dev.*, 9:1766). In particular embodiments, vectors comprise a posttranscriptional regulatory element such as a WPRE or HPRE. In particular embodiments, vectors lack or do not comprise a posttranscriptional regulatory element such as a WPRE or HPRE.

Elements directing the efficient termination and polyadenylation of the heterologous nucleic acid transcripts increases heterologous gene expression. Illustrative examples of polyA signals that can be used in a vector, includes an ideal polyA sequence (*e.g.,* AATAAA, ATTAAA, AGTAAA), a bovine growth hormone polyA sequence

(BGHpA), a rabbit β-globin polyA sequence (rβgpA), or another suitable heterologous or endogenous polyA sequence known in the art.

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 retroviral and/or lentiviral sequences can be used, or combined 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.

In particular embodiments, a retroviral vector comprises a left (5′) lentiviral LTR; a Psi (ψ) packaging signal; a retroviral export element; a cPPT/FLAP; a promoter operably linked to a polynucleotide encoding a tripeptidyl peptidase 1(TPP1) polypeptide; and a right (3′) lentiviral LTR. In certain embodiments, the retroviral vector is preferably a lentiviral vector, more preferably an HIV lentiviral vector, and even preferably, an HIV-1 lentiviral vector.

In particular embodiments, a lentiviral vector comprises a left (5′) lentiviral LTR wherein the promoter region of the LTR is replaced with a heterologous promoter; a Psi (ψ) packaging signal; a retroviral export element; a cPPT/FLAP; a promoter operably linked to a polynucleotide encoding a tripeptidyl peptidase 1(TPP1) polypeptide; and a right (3′) lentiviral LTR. In certain embodiments, the heterologous promoter is a cytomegalovirus (CMV) promoter, a Rous Sarcoma Virus (RSV) promoter, or a Simian Virus 40 (SV40) promoter.

In particular embodiments, a lentiviral vector comprises a left (5′) lentiviral LTR; a Psi (ψ) packaging signal; a retroviral export element; a cPPT/FLAP; a promoter operably linked to a polynucleotide encoding a tripeptidyl peptidase 1(TPP1) polypeptide; and a right (3′) lentiviral LTR that comprises one or more modification compared to an unmodified LTR. In certain embodiments, the 3′ LTR preferably comprises one or more deletions that prevent viral transcription beyond the first round of viral replication, more preferably comprises a deletion of the TATA box and Sp1 and NF-κB transcription factor binding sites in the U3 region of the 3′ LTR, and even more preferably is a self-inactivating (SIN) LTR.

In particular embodiments, a lentiviral vector comprises a left (5′) lentiviral LTR wherein the promoter region of the LTR is replaced with a heterologous promoter; a Psi (ψ) packaging signal; a retroviral export element; a cPPT/FLAP; a promoter operably linked to a polynucleotide encoding a tripeptidyl peptidase 1(TPP1) polypeptide; and a right (3′) lentiviral SIN LTR.

In particular embodiments, a lentiviral vector comprises a left (5′) lentiviral LTR wherein the promoter region of the LTR is replaced with a heterologous promoter; a Psi (ψ) packaging signal; a retroviral export element; a cPPT/FLAP; a myeloproliferative sarcoma virus enhancer, negative control region deleted, dl587rev primer-binding site substituted (MND) promoter or transcriptionally active fragment thereof operably linked to a polynucleotide encoding a human tripeptidyl peptidase 1(TPP1) polypeptide; and a right (3′) lentiviral SIN LTR.

In particular embodiments, a lentiviral vector comprises a left (5′) lentiviral LTR wherein the promoter region of the LTR is replaced with a heterologous promoter; a Psi (ψ) packaging signal; a retroviral export element; a cPPT/FLAP; an elongation factor 1 alpha (EF1α) promoter or transcriptionally active fragment thereof operably linked to a polynucleotide encoding a human tripeptidyl peptidase 1(TPP1) polypeptide; and a right (3′) lentiviral SIN LTR. In preferred embodiments, the EF1α promoter lacks the first intron of the human EF1a gene and is referred to as an "EF1α short promoter." In other embodiments, the EF1α promoter comprises the first intron of the human EF1a gene and is referred to as an "EF1α long promoter."

In particular embodiments, a lentiviral vector comprises a left (5′) CMV promoter/HIV-1 chimeric LTR; a Psi (ψ) packaging signal; an RRE retroviral export element; a cPPT/FLAP; an MND promoter or EF1α-short promoter operably linked to a polynucleotide encoding a human tripeptidyl peptidase 1(TPP1) polypeptide; and a right (3′) lentiviral SIN LTR.

In particular embodiments, a lentiviral vector comprises a left (5′) CMV promoter/HIV-1 chimeric LTR; a Psi (ψ) packaging signal; an RRE retroviral export element; a cPPT/FLAP; an MND promoter or EF1α-short promoter operably linked to a polynucleotide encoding a human tripeptidyl peptidase 1(TPP1) polypeptide; a right (3′) lentiviral SIN LTR; and a heterologous polyadenylation signal. In certain embodiments, the polyadenylation signal is an artificial polyadenylation signal, a bovine growth hormone polyadenylation signal or a rabbit β-globin polyadenylation signal.

Large scale viral particle production is often necessary to achieve a reasonable viral titer. Viral particles are produced by transfecting a transfer vector into a packaging cell that comprises viral structural and/or accessory genes, *e.g.*, gag, pol, env, tat, rev, vif, vpr, vpu, vpx, or nef genes or other retroviral genes.

As used herein, the term "packaging vector" refers to an expression vector or viral vector that lacks a packaging signal and comprises a polynucleotide encoding one, two, three, four or more viral structural and/or accessory genes. Typically, the packaging vectors are included in a packaging cell, and are introduced into the cell via transfection, transduction or infection. Methods for transfection, transduction or infection are well known by those of skill in the art. A retroviral/lentiviral transfer vector can be introduced into a packaging cell line, via transfection, transduction or infection, to generate a producer cell or cell line. The packaging vectors can be introduced into human cells or cell lines by standard methods including, *e.g.*, calcium phosphate transfection, lipofection or electroporation. In some embodiments, the packaging vectors are introduced into the cells together with a dominant selectable marker, such as neomycin, hygromycin, puromycin, blastocidin, zeocin, thymidine kinase, DHFR, Gln synthetase or ADA, followed by selection in the presence of the appropriate drug and isolation of clones. A selectable marker gene can be linked physically to genes encoding by the packaging vector, *e.g.*, by IRES or self-cleaving viral peptides.

Viral envelope proteins (env) determine the range of host cells which can ultimately be infected and transformed by recombinant retroviruses generated from the cell lines. In the case of lentiviruses, such as HIV-1, HIV-2, SIV, FIV and EIV, the env proteins include gp41 and gp120. Preferably, the viral env proteins expressed by packaging cells are encoded on a separate vector from the viral gag and pol genes, as has been previously described.

Illustrative examples of retroviral-derived env genes which can be employed in particular embodiments include, but are not limited to: MLV envelopes, 10A1 envelope, BAEV, FeLV-B, RD114, SSAV, Ebola, Sendai, FPV (Fowl plague virus), and influenza virus envelopes. Similarly, genes encoding envelopes from RNA viruses (*e.g.*, RNA virus families of *Picornaviridae, Calciviridae, Astroviridae, Togaviridae, Flaviviridae, Coronaviridae, Paramyxoviridae, Rhabdoviridae, Filoviridae, Orthomyxoviridae, Bunyaviridae, Arenaviridae, Reoviridae, Birnaviridae, Retroviridae*) as well as from the DNA viruses (families of *Hepadnaviridae, Circoviridae, Parvoviridae, Papovaviridae,*

Adenoviridae, Herpesviridae, Poxyiridae, and Iridoviridae) may be utilized. Representative examples of these viruses include, but are not limited to, FeLV, VEE, HFVW, WDSV, SFV, Rabies, ALV, BIV, BLV, EBV, CAEV, SNV, ChTLV, STLV, MPMV, SMRV, RAV, FuSV, MH2, AEV, AMV, CT10, and EIAV.

In other embodiments, envelope proteins for pseudotyping a virus include, but are not limited to any of the following virus: Influenza A such as H1N1, H1N2, H3N2 and H5N1 (bird flu), Influenza B, Influenza C virus, Hepatitis A virus, Hepatitis B virus, Hepatitis C virus, Hepatitis D virus, Hepatitis E virus, Rotavirus, any virus of the Norwalk virus group, enteric adenoviruses, parvovirus, Dengue fever virus, Monkey pox, Mononegavirales, Lyssavirus such as rabies virus, Lagos bat virus, Mokola virus, Duvenhage virus, European bat virus 1 & 2 and Australian bat virus, Ephemerovirus, Vesiculovirus, Vesicular Stomatitis Virus (VSV), Herpesviruses such as Herpes simplex virus types 1 and 2, varicella zoster, cytomegalovirus, Epstein-Bar virus (EBV), human herpesviruses (HHV), human herpesvirus type 6 and 8, Human immunodeficiency virus (HIV), papilloma virus, murine gammaherpesvirus, Arenaviruses such as Argentine hemorrhagic fever virus, Bolivian hemorrhagic fever virus, Sabia-associated hemorrhagic fever virus, Venezuelan hemorrhagic fever virus, Lassa fever virus, Machupo virus, Lymphocytic choriomeningitis virus (LCMV), Bunyaviridiae such as Crimean-Congo hemorrhagic fever virus, Hantavirus, hemorrhagic fever with renal syndrome causing virus, Rift Valley fever virus, Filoviridae (filovirus) including Ebola hemorrhagic fever and Marburg hemorrhagic fever, Flaviviridae including Kaysanur Forest disease virus, Omsk hemorrhagic fever virus, Tick-borne encephalitis causing virus and Paramyxoviridae such as Hendra virus and Nipah virus, variola major and variola minor (smallpox), alphaviruses such as Venezuelan equine encephalitis virus, eastern equine encephalitis virus, western equine encephalitis virus, SARS-associated coronavirus (SARS-CoV), West Nile virus, any encephaliltis causing virus.

In one embodiment, packaging cells are provided, which produce recombinant retrovirus, *e.g.,* lentivirus, pseudotyped with the VSV-G glycoprotein.

The terms "pseudotype" or "pseudotyping" as used herein, refer to a virus whose viral envelope proteins have been substituted with those of another virus possessing preferable characteristics. For example, HIV can be pseudotyped with vesicular stomatitis virus G-protein (VSV-G) envelope proteins, which allows HIV to infect a wider range of cells because HIV envelope proteins (encoded by the env gene) normally target the virus to

CD4+ presenting cells. In a preferred embodiment, lentiviral envelope proteins are pseudotyped with VSV-G. In one embodiment, packaging cells are provided which produce recombinant retrovirus, *e.g.,* lentivirus, pseudotyped with the VSV-G envelope glycoprotein.

As used herein, the term "packaging cell lines" is used in reference to cell lines that do not contain a packaging signal, but do stably or transiently express viral structural proteins and replication enzymes (*e.g.,* gag, pol and env) which are necessary for the correct packaging of viral particles. Any suitable cell line can be employed to prepare packaging cells. Generally, the cells are mammalian cells. In a particular embodiment, the cells used to produce the packaging cell line are human cells. Suitable cell lines which can be used include, for example, CHO cells, BHK cells, MDCK cells, C3H 10T1/2 cells, FLY cells, Psi-2 cells, BOSC 23 cells, PA317 cells, WEHI cells, COS cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, W138 cells, MRC5 cells, A549 cells, HT1080 cells, 293 cells, 293T cells, B-50 cells, 3T3 cells, NIH3T3 cells, HepG2 cells, Saos-2 cells, Huh7 cells, HeLa cells, W163 cells, 211 cells, and 211A cells. In preferred embodiments, the packaging cells are 293 cells, 293T cells, or A549 cells. In another preferred embodiment, the cells are A549 cells.

As used herein, the term "producer cell line" refers to a cell line which is capable of producing recombinant retroviral particles, comprising a packaging cell line and a transfer vector construct comprising a packaging signal. The production of infectious viral particles and viral stock solutions may be carried out using conventional techniques. Methods of preparing viral stock solutions are known in the art and are illustrated by, *e.g.,* Y. Soneoka *et al.* (1995) *Nucl. Acids Res.* 23:628-633, and N. R. Landau *et al.* (1992) *J. Virol.* 66:5110-5113. Infectious virus particles may be collected from the packaging cells using conventional techniques. For example, the infectious particles can be collected by cell lysis, or collection of the supernatant of the cell culture, as is known in the art. Optionally, the collected virus particles may be purified if desired. Suitable purification techniques are well known to those skilled in the art.

In particular embodiments, host cells transduced with viral vector that expresses one or more polypeptides to generate genetically modified cells that are administered to a subject to treat and/or prevent and/or ameliorate at least one symptom of a neuronal ceroid lipofuscinoses. Other methods relating to the use of viral vectors in gene therapy, which may be utilized according to certain embodiments, can be found in, *e.g.,* Kay, M. A. (1997)

Chest 111(6 Supp.):138S-142S; Ferry, N. and Heard, J. M. (1998) *Hum. Gene Ther.* 9:1975-81; Shiratory, Y. *et al.* (1999) Liver 19:265-74; Oka, K. *et al.* (2000) *Curr. Opin. Lipidol.* 11:179-86; Thule, P. M. and Liu, J. M. (2000) *Gene Ther.* 7:1744-52; Yang, N. S. (1992) *Crit. Rev. Biotechnol.* 12:335-56; Alt, M. (1995) *J. Hepatol.* 23:746-58; Brody, S. L. and Crystal, R. G. (1994) *Ann. N.Y. Acad. Sci.* 716:90-101; Strayer, D. S. (1999) *Expert Opin. Investig. Drugs* 8:2159-2172; Smith-Arica, J. R. and Bartlett, J. S. (2001) *Curr. Cardiol. Rep.* 3:43-49; and Lee, H. C. *et al.* (2000) *Nature* 408:483-8.

A "host cell" includes cells transfected, infected, or transduced *in vivo*, *ex vivo*, or *in vitro* with a recombinant vector or a polynucleotide contemplated herein. Host cells may include packaging cells, producer cells, and cells infected with viral vectors. In particular embodiments, host cells infected with viral vector of the invention are administered to a subject in need of therapy. In certain embodiments, the term "target cell" is used interchangeably with host cell and refers to transfected, infected, or transduced cells of a desired cell type. In preferred embodiments, the target cell is a stem cell or progenitor cell. In certain preferred embodiments, the target cell is a somatic cell, *e.g.*, adult stem cell, progenitor cell, or differentiated cell. In particular preferred embodiments, the target cell is a hematopoietic cell, *e.g.*, a hematopoietic stem or progenitor cell, or $CD34^+$ cell. Further therapeutic target cells are discussed, herein.

F. GENETICALLY MODIFIED CELLS

In various embodiments, cells are genetically modified to express a TPP1 polypeptide, and the genetically modified cells are used to treat neuronal ceroid lipofuscinoses. The cells may be genetically modified *ex vivo*, *in vitro*, or *ex vivo*. As used herein, the term "genetically engineered" or "genetically modified" refers to the addition of extra genetic material in the form of DNA or RNA into the total genetic material in a cell. The terms, "genetically modified cells," "modified cells," and, "genetically engineered cells," are used interchangeably. As used herein, the term "gene therapy" refers to the introduction of extra genetic material in the form of DNA or RNA into the total genetic material in a cell that restores, corrects, or modifies expression of a gene, or for the purpose of expressing a therapeutic polypeptide, *e.g.*, TPP1.

The cells can be autologous/autogeneic ("self") or non-autologous ("non-self," *e.g.*, allogeneic, syngeneic or xenogeneic). "Autologous," as used herein, refers to cells from the same subject. "Allogeneic," as used herein, refers to cells of the same species that differ

genetically to the cell in comparison. "Syngeneic," as used herein, refers to cells of a different subject that are genetically identical to the cell in comparison. "Xenogeneic," as used herein, refers to cells of a different species to the cell in comparison. In preferred embodiments, the cells are allogeneic.

In particular embodiments, vectors encoding TPP1 are introduced into one or more animal cells, preferably a mammal, *e.g.*, a non-human primate or human, and more preferably a human.

In certain embodiments, a population of cells is transduced with a vector contemplated herein. As used herein, the term "population of cells" refers to a plurality of cells that may be made up of any number and/or combination of homogenous or heterogeneous cell types, as described elsewhere herein. For example, for transduction of hematopoietic stem or progenitor cells, a population of cells may be isolated or obtained from umbilical cord blood, placental blood, bone marrow, or peripheral blood. A population of cells may comprise about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, or about 100% of the target cell type to be transduced. In certain embodiments, hematopoietic stem or progenitor cells may be isolated or purified from a population of heterogeneous cells using methods known in the art.

In particular embodiments, the cell is a primary cell. The term "primary cell" as used herein is known in the art to refer to a cell that has been isolated from a tissue and has been established for growth *in vitro* or *ex vivo*. Corresponding cells have undergone very few, if any, population doublings and are therefore more representative of the main functional component of the tissue from which they are derived in comparison to continuous cell lines, thus representing a more representative model to the *in vivo* state. Methods to obtain samples from various tissues and methods to establish primary cell lines are well-known in the art (see, *e.g.*, Jones and Wise, Methods Mol Biol. 1997). Primary cells for use in the method of the invention are derived from, *e.g.*, blood, lymphoma and epithelial tumors. In one embodiment, the primary cell is a hematopoietic stem or progenitor cell.

The term "stem cell" refers to a cell which is an undifferentiated cell capable of (1) long term self-renewal, or the ability to generate at least one identical copy of the original cell, (2) differentiation at the single cell level into multiple, and in some instance only one, specialized cell type and (3) of *in vivo* functional regeneration of tissues. Stem cells are

subclassified according to their developmental potential as totipotent, pluripotent, multipotent and oligo/unipotent. "Self-renewal" refers a cell with a unique capacity to produce unaltered daughter cells and to generate specialized cell types (potency). Self-renewal can be achieved in two ways. Asymmetric cell division produces one daughter cell that is identical to the parental cell and one daughter cell that is different from the parental cell and is a progenitor or differentiated cell. Symmetric cell division produces two identical daughter cells. "Proliferation" or "expansion" of cells refers to symmetrically dividing cells.

As used herein, the term "progenitor" or "progenitor cells" refers to cells have the capacity to self-renew and to differentiate into more mature cells. Many progenitor cells differentiate along a single lineage, but may have quite extensive proliferative capacity.

Hematopoietic stem cells (HSCs) give rise to committed hematopoietic progenitor cells (HPCs) that are capable of generating the entire repertoire of mature blood cells over the lifetime of an organism. The term "hematopoietic stem cell" or "HSC" refers to multipotent stem cells that give rise to the all the blood cell types of an organism, including myeloid (*e.g.*, monocytes and macrophages, neutrophils, basophils, eosinophils, erythrocytes, megakaryocytes/platelets, dendritic cells), and lymphoid lineages (*e.g.*, T-cells, B-cells, NK-cells), and others known in the art (See Fei, R., *et al.*, U.S. Patent No. 5,635,387; McGlave, *et al.*, U.S. Patent No. 5,460,964; Simmons, P., *et al.*, U.S. Patent No. 5,677,136; Tsukamoto, *et al.*, U.S. Patent No. 5,750,397; Schwartz, *et al.*, U.S. Patent No. 5,759,793; DiGuisto, *et al.*, U.S. Patent No. 5,681,599; Tsukamoto, *et al.*, U.S. Patent No. 5,716,827). In one embodiment, the HSC is a $CD34^+$ cell. When transplanted into lethally irradiated animals or humans, hematopoietic stem and progenitor cells can repopulate the erythroid, neutrophil-macrophage, megakaryocyte and lymphoid hematopoietic cell pool.

Preferred target cell types transduced with the compositions and methods contemplated herein include, hematopoietic cells, preferably human hematopoietic cells, more preferably human hematopoietic stem and progenitor cells, and even more preferably $CD34^+$ human hematopoietic stem cells.

Illustrative sources to obtain hematopoietic cells transduced with the methods and compositions contemplated herein include, but are not limited to: cord blood, bone marrow or mobilized peripheral blood.

In particular embodiments, hematopoietic cells transduced with viral vectors encoding TPP1 contemplated herein include $CD34^+$ cells. The term "$CD34^+$ cell," as used

herein refers to a cell expressing the CD34 protein on its cell surface. "CD34," as used herein refers to a cell surface glycoprotein (*e.g.*, sialomucin protein) that often acts as a cell-cell adhesion factor. $CD34^+$ is a cell surface marker of both hematopoietic stem and progenitor cells.

Additional illustrative examples of hematopoietic stem or progenitor cells suitable for transduction with the methods and compositions contemplated herein include hematopoietic cells that are $CD34^+CD38^{Lo}CD90^+CD45^{RA-}$, hematopoietic cells that are $CD34^+$, $CD59^+$, $Thy1/CD90^+$, $CD38^{Lo/-}$, $C\text{-}kit/CD117^+$, and $Lin^{(-)}$, and hematopoietic cells that are $CD133^+$.

In one embodiment, hematopoietic cells transduced with viral vectors encoding TPP1 contemplated herein include $CD34^+CD133^+$ cells.

Various methods exist to characterize hematopoietic hierarchy. One method of characterization is the SLAM code. The SLAM (Signaling lymphocyte activation molecule) family is a group of >10 molecules whose genes are located mostly tandemly in a single locus on chromosome 1 (mouse), all belonging to a subset of immunoglobulin gene superfamily, and originally thought to be involved in T-cell stimulation. This family includes CD48, CD150, CD244, *etc.*, CD150 being the founding member, and, thus, also called slamF1, *i.e.*, SLAM family member 1. The signature SLAM code for the hematopoietic hierarchy is hematopoietic stem cells (HSC) - $CD150^+CD48^-CD244^-$; multipotent progenitor cells (MPPs) - $CD150^-CD48^-CD244^+$; lineage-restricted progenitor cells (LRPs) - $CD150^-CD48^+CD244^+$; common myeloid progenitor (CMP) - $lin\text{-}SCA\text{-}1\text{-}c\text{-}kit^+CD34^+CD16/32^{mid}$; granulocyte-macrophage progenitor (GMP) - $lin^-SCA\text{-}1\text{-}c\text{-}kit^+CD34^+CD16/32^{hi}$; and megakaryocyte-erythroid progenitor (MEP) - $lin^-SCA\text{-}1\text{-}c\text{-}kit^+CD34^-CD16/32^{low}$.

In one embodiment, hematopoietic cells transduced with viral vectors encoding TPP1 contemplated herein include $CD150^+CD48^-CD244^-$ cells.

In various embodiments, a population of hematopoietic cells comprising hematopoietic stem and progenitor cells (HSPCs) transduced with a viral vector encoding TPP1 as contemplated herein is provided. In preferred embodiments, the HSPCs are $CD34^+$ hematopoietic cells.

G. COMPOSITIONS AND FORMULATIONS

The compositions and formulations contemplated herein may comprise a combination of any number of transduced or non-transduced cells or a combination thereof, viral vectors, polypeptides, and polynucleotides contemplated herein. Compositions include, but are not limited to pharmaceutical compositions. A "pharmaceutical composition" refers to a composition formulated with a pharmaceutically-acceptable carrier for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy. It will also be understood that, if desired, the compositions may be administered in combination with other agents as well, such as, *e.g.,* cytokines, growth factors, hormones, small molecules, pro-drugs, drugs, antibodies, or other various pharmaceutically-active agents. In particular embodiments, there is virtually no limit to other components that may also be included in the compositions, provided that the additional agents do not adversely affect the ability of the composition to deliver the intended therapy.

Particular *ex vivo* and *in vitro* formulations and compositions contemplated herein may comprise a combination of transduced or non-transduced cells or a combination thereof, and viral vectors formulated with a pharmaceutically-acceptable carrier for administration to a cell, tissue, organ, or an animal, either alone, or in combination with one or more other modalities of therapy.

Particular *in vivo* formulations and compositions contemplated herein may comprise a combination of viral vectors formulated with a pharmaceutically-acceptable carrier for administration to a cell, tissue, organ, or an animal, either alone, or in combination with one or more other modalities of therapy.

In certain embodiments, compositions contemplated herein comprise a population of cells, comprising a therapeutically-effective amount of transduced cells, *e.g.,* hematopoietic cells, hematopoietic stem cells, hematopoietic progenitor cells, $CD34^+$ cells, $CD133^+$ cells, *etc.,* formulated with one or more pharmaceutically acceptable carriers.

In certain other embodiments, the present invention provides compositions comprising a retroviral vector, *e.g.,* a lentiviral vector formulated with one or more pharmaceutically acceptable carriers.

Pharmaceutical compositions contemplated herein comprise transduced cells comprising a vector or provirus encoding TPP1 as contemplated herein and a pharmaceutically acceptable carrier.

The phrase "pharmaceutically acceptable" is employed herein to refer to those compounds, materials, compositions, and/or dosage forms which are, within the scope of sound medical judgment, suitable for use in contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio.

The term "pharmaceutically acceptable carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the therapeutic cells are administered. Illustrative examples of pharmaceutical carriers can be sterile liquids, such as cell culture media, water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Suitable pharmaceutical excipients in particular embodiments, include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions.

In one embodiment, a composition comprising a pharmaceutically acceptable carrier is suitable for administration to a subject. In particular embodiments, a composition comprising a carrier is suitable for parenteral administration, *e.g.*, intravascular (intravenous or intraarterial), intraperitoneal or intramuscular administration. In particular embodiments, a composition comprising a pharmaceutically acceptable carrier is suitable for intraventricular, intraspinal, or intrathecal administration. Pharmaceutically acceptable carriers include sterile aqueous solutions, cell culture media, or dispersions. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the transduced cells, use thereof in the pharmaceutical compositions is contemplated.

In particular embodiments, compositions contemplated herein comprise genetically modified hematopoietic stem and/or progenitor cells and a pharmaceutically acceptable carrier. A composition comprising a cell-based composition contemplated herein can be administered separately by enteral or parenteral administration methods or in combination with other suitable compounds to effect the desired treatment goals

The pharmaceutically acceptable carrier must be of sufficiently high purity and of sufficiently low toxicity to render it suitable for administration to the human subject being treated. It further should maintain or increase the stability of the composition. The pharmaceutically acceptable carrier can be liquid or solid and is selected, with the planned manner of administration in mind, to provide for the desired bulk, consistency, *etc.*, when combined with other components of the composition. For example, the pharmaceutically acceptable carrier can be, without limitation, a binding agent (*e.g.*, pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose, *etc.*), a filler (*e.g.*, lactose and other sugars, microcrystalline cellulose, pectin, gelatin, calcium sulfate, ethyl cellulose, polyacrylates, calcium hydrogen phosphate, *etc.*), a lubricant (*e.g.,* magnesium stearate, talc, silica, colloidal silicon dioxide, stearic acid, metallic stearates, hydrogenated vegetable oils, corn starch, polyethylene glycols, sodium benzoate, sodium acetate, *etc.*), a disintegrant (*e.g.*, starch, sodium starch glycolate, *etc.*), or a wetting agent (*e.g.*, sodium lauryl sulfate, *etc.*). Other suitable pharmaceutically acceptable carriers for the compositions contemplated herein include, but are not limited to, water, salt solutions, alcohols, polyethylene glycols, gelatins, amyloses, magnesium stearates, talcs, silicic acids, viscous paraffins, hydroxymethylcelluloses, polyvinylpyrrolidones and the like.

Such carrier solutions also can contain buffers, diluents and other suitable additives. The term “buffer” as used herein refers to a solution or liquid whose chemical makeup neutralizes acids or bases without a significant change in pH. Examples of buffers contemplated herein include, but are not limited to, Dulbecco’s phosphate buffered saline (PBS), Ringer's solution, 5% dextrose in water (D5W), normal/physiologic saline (0.9% NaCl).

The pharmaceutically acceptable carriers may be present in amounts sufficient to maintain a pH of the composition of about 7. Alternatively, the composition has a

pH in a range from about 6.8 to about 7.4, *e.g.*, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, and 7.4. In still another embodiment, the composition has a pH of about 7.4.

Compositions contemplated herein may comprise a nontoxic pharmaceutically acceptable medium. The compositions may be a suspension. The term "suspension" as used herein refers to non-adherent conditions in which cells are not attached to a solid support. For example, cells maintained as a suspension may be stirred or agitated and are not adhered to a support, such as a culture dish.

In particular embodiments, compositions contemplated herein are formulated in a suspension, where the hematopoietic stem and/or progenitor cells are dispersed within an acceptable liquid medium or solution, *e.g.*, saline or serum-free medium, in an intravenous (IV) bag or the like. Acceptable diluents include, but are not limited to water, PlasmaLyte, Ringer's solution, isotonic sodium chloride (saline) solution, serum-free cell culture medium, and medium suitable for cryogenic storage, *e.g.*, Cryostor® medium.

In certain embodiments, a pharmaceutically acceptable carrier is substantially free of natural proteins of human or animal origin, and suitable for storing a composition comprising a population of cells, *e.g.,* hematopoietic stem and progenitor cells. The therapeutic composition is intended to be administered into a human patient, and thus is substantially free of cell culture components such as bovine serum albumin, horse serum, and fetal bovine serum.

In some embodiments, compositions are formulated in a pharmaceutically acceptable cell culture medium. Such compositions are suitable for administration to human subjects. In particular embodiments, the pharmaceutically acceptable cell culture medium is a serum free medium.

Serum-free medium has several advantages over serum containing medium, including a simplified and better defined composition, a reduced degree of contaminants, elimination of a potential source of infectious agents, and lower cost. In various embodiments, the serum-free medium is animal-free, and may optionally be protein-free. Optionally, the medium may contain biopharmaceutically acceptable recombinant proteins. "Animal-free" medium refers to medium wherein the components are derived from non-animal sources. Recombinant proteins replace native animal proteins in animal-free medium and the nutrients are obtained from

synthetic, plant or microbial sources. "Protein-free" medium, in contrast, is defined as substantially free of protein.

Illustrative examples of serum-free media used in particular compositions includes, but is not limited to QBSF-60 (Quality Biological, Inc.), StemPro-34 (Life Technologies), and X-VIVO 10.

In a preferred embodiment, the compositions comprising hematopoietic stem and/or progenitor cells are formulated in PlasmaLyte.

In various embodiments, compositions comprising hematopoietic stem and/or progenitor cells are formulated in a cryopreservation medium. For example, cryopreservation media with cryopreservation agents may be used to maintain a high cell viability outcome post-thaw. Illustrative examples of cryopreservation media used in particular compositions includes, but is not limited to, CryoStor CS10, CryoStor CS5, and CryoStor CS2.

In one embodiment, the compositions are formulated in a solution comprising 50:50 PlasmaLyte A to CryoStor CS10.

In particular embodiments, the composition is substantially free of mycoplasma, endotoxin, and microbial contamination. By "substantially free" with respect to endotoxin is meant that there is less endotoxin per dose of cells than is allowed by the FDA for a biologic, which is a total endotoxin of 5 EU/kg body weight per day, which for an average 70 kg person is 350 EU per total dose of cells. In particular embodiments, compositions comprising hematopoietic stem or progenitor cells transduced with a retroviral vector contemplated herein contains about 0.5 EU/mL to about 5.0 EU/mL, or about 0.5 EU/mL, 1.0 EU/mL, 1.5 EU/mL, 2.0 EU/mL, 2.5 EU/mL, 3.0 EU/mL, 3.5 EU/mL, 4.0 EU/mL, 4.5 EU/mL, or 5.0 EU/mL.

In certain embodiments, compositions and formulations suitable for the delivery of viral vector systems (*i.e.*, viral-mediated transduction) are contemplated including, but not limited to, retroviral (*e.g.*, lentiviral) vectors.

Exemplary formulations for *ex vivo* delivery may also include the use of various transfection agents known in the art, such as calcium phosphate, electroporation, heat shock and various liposome formulations (*i.e.*, lipid-mediated transfection). Liposomes, as described in greater detail below, are lipid bilayers entrapping a fraction of aqueous fluid. DNA spontaneously associates to the external

surface of cationic liposomes (by virtue of its charge) and these liposomes will interact with the cell membrane.

In particular embodiments, formulation of pharmaceutically-acceptable carrier solutions is well-known to those of skill in the art, as is the development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.*, enteral and parenteral, *e.g.*, intravascular, intravenous, intrarterial, intraosseously, intraventricular, intracerebral, intracranial, intraspinal, intrathecal, and intramedullary administration and formulation. It would be understood by the skilled artisan that particular embodiments contemplated herein may comprise other formulations, such as those that are well known in the pharmaceutical art, and are described, for example, in Remington: The Science and Practice of Pharmacy, 20th Edition. Baltimore, MD: Lippincott Williams & Wilkins, 2005, which is incorporated by reference herein, in its entirety.

H. GENE THERAPY METHODS

The genetically modified cells contemplated herein provide improved drug products for use in the prevention, treatment, and amelioration of a neuronal ceroid fuscinoses or for preventing, treating, or ameliorating at least one symptom associated with a neuronal ceroid fuscinoses or a subject having a mutation in a TPP1 gene that decreases or abolishes TPP1 expression. In preferred embodiments, the neuronal ceroid fuscinoses is late-infantile neuronal ceroid fuscinoses (LINCL). In another preferred embodiment, the neuronal ceroid fuscinoses is juvenile Batten disease (JNCL).

As used herein, the term "drug product" refers to genetically modified cells produced using the compositions and methods contemplated herein. In particular embodiments, the drug product comprises genetically modified hematopoietic stem or progenitor cells, *e.g.*, $CD34^+$ cells. Without wishing to be bound to any particular theory, increasing the amount of a therapeutic gene in a drug product may allow treatment of subjects having no or minimal expression of the corresponding gene *in vivo*, thereby significantly expanding the opportunity to bring gene therapy to subjects for which gene therapy was not previously a viable treatment option.

The transduced cells and corresponding retroviral vectors contemplated herein provide improved methods of gene therapy. As used herein, the term "gene therapy" refers to the introduction of a gene into a cell's genome. In various embodiments, a

viral vector of the invention comprises an expression control sequence that expresses a therapeutic transgene encoding a polypeptide that provides curative, preventative, or ameliorative benefits to a subject diagnosed with or that is suspected of having an NCL, LINCL, JNCL, or a subject having TPP1 gene comprising one or more mutations that decrease TPP1 expression.

In various embodiments, the retroviral vectors are administered by direct injection to a cell, tissue, or organ of a subject in need of gene therapy, *in vivo*. In various other embodiments, cells are transduced *in vitro* or *ex vivo* with vectors contemplated herein, and optionally expanded *ex vivo*. The transduced cells are then administered to a subject in need of gene therapy.

Cells suitable for transduction and administration in the gene therapy methods contemplated herein include, but are not limited to stem cells, progenitor cells, and differentiated cells as described elsewhere herein. In certain embodiments, the transduced cells are hematopoietic stem or progenitor cells as described elsewhere herein.

Preferred cells for use in the gene therapy compositions and methods contemplated herein include autologous/autogeneic ("self") cells.

As used herein, the terms "individual" and "subject" are often used interchangeably and refer to any animal that exhibits a symptom of a disease, disorder, or condition that can be treated with the gene therapy vectors, cell-based therapeutics, and methods contemplated elsewhere herein. In preferred embodiments, a subject includes any animal that exhibits symptoms of a neuronal ceroid lipofuscinoses that can be treated with the gene therapy vectors, cell-based therapeutics, and methods contemplated elsewhere herein. Suitable subjects (*e.g.,* patients) include laboratory animals (such as mouse, rat, rabbit, or guinea pig), farm animals, and domestic animals or pets (such as a cat or dog). Non-human primates and, preferably, human patients, are included. Typical subjects include human patients that have an NCL, have been diagnosed with an NCL, or are at risk or having an NCL.

As used herein, the term "patient" refers to a subject that has been diagnosed with a particular disease, disorder, or condition that can be treated with the gene therapy vectors, cell-based therapeutics, and methods disclosed elsewhere herein.

As used herein "treatment" or "treating," includes any beneficial or desirable effect on the symptoms or pathology of a disease or pathological condition, and may include even

minimal reductions in one or more measurable markers of the disease or condition being treated. Treatment can involve optionally either the reduction the disease or condition, or the delaying of the progression of the disease or condition. "Treatment" does not necessarily indicate complete eradication or cure of the disease or condition, or associated symptoms thereof.

As used herein, "prevent," and similar words such as "prevented," "preventing" *etc.*, indicate an approach for preventing, inhibiting, or reducing the likelihood of the occurrence or recurrence of, a disease or condition. It also refers to delaying the onset or recurrence of a disease or condition or delaying the occurrence or recurrence of the symptoms of a disease or condition. As used herein, "prevention" and similar words also includes reducing the intensity, effect, symptoms and/or burden of a disease or condition prior to onset or recurrence of the disease or condition.

As used herein, the phrase "ameliorating at least one symptom of" refers to decreasing one or more symptoms of the disease or condition for which the subject is being treated. In particular embodiments, the disease or condition being treated is an NCL, wherein the at least one symptom is selected from the group consisting of: progressive loss of motor function, progressive loss of cognitive function, visual impairment, blindness, speech difficulties, ataxia, dementia, heart problems, behavioral problems, difficulty sleeping, and problems with attention, and seizures.

In particular embodiments, a subject is administered an amount of genetically modified cell or gene therapy vector sufficient to treat, prevent, or ameliorate at least one symptom of an NCL.

As used herein, the term "amount" refers to "an amount effective" or "an effective amount" of a virus or transduced therapeutic cell to achieve a beneficial or desired prophylactic or therapeutic result, including clinical results.

A "prophylactically effective amount" refers to an amount of a virus or transduced therapeutic cell effective to achieve the desired prophylactic result. Typically but not necessarily, since a prophylactic dose is used in subjects prior to or at an earlier stage of disease, the prophylactically effective amount is less than the therapeutically effective amount.

A "therapeutically effective amount" of a virus or transduced therapeutic cell may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of the stem and progenitor cells to elicit a desired response in

the individual. A therapeutically effective amount is also one in which any toxic or detrimental effects of the virus or transduced therapeutic cells are outweighed by the therapeutically beneficial effects. The term "therapeutically effective amount" includes an amount that is effective to "treat" a subject (*e.g.*, a patient).

Without wishing to be bound to any particular theory, an important advantage provided by the vectors, compositions, and methods of the present invention is the high efficacy of gene therapy that can be achieved by administering populations of cells comprising high percentages of transduced cells compared to existing methods.

The transduced cells may be administered as part of a bone marrow or cord blood transplant in an individual that has or has not undergone bone marrow ablative therapy. In one embodiment, transduced cells of the invention are administered in a bone marrow transplant to an individual that has undergone chemoablative or radioablative bone marrow therapy.

In one embodiment, a dose of transduced cells is delivered to a subject intravenously. In preferred embodiments, transduced hematopoietic stem cells are intravenously administered to a subject.

In one illustrative embodiment, the effective amount of transduced cells provided to a subject is at least 2 x 10^6cells/kg, at least 3 x 10^6cells/kg, at least 4 x 10^6cells/kg, at least 5 x 10^6cells/kg, at least 6 x 10^6cells/kg, at least 7 x 10^6cells/kg, at least 8 x 10^6cells/kg, at least 9 x 10^6cells/kg, or at least 10 x 10^6cells/kg, or more cells/kg, including all intervening doses of cells.

In another illustrative embodiment, the effective amount of transduced cells provided to a subject is about 2 x 10^6cells/kg, about 3 x 10^6cells/kg, about 4 x 10^6cells/kg, about 5 x 10^6cells/kg, about 6 x 10^6cells/kg, about 7 x 10^6cells/kg, about 8 x 10^6cells/kg, about 9 x 10^6cells/kg, or about 10 x 10^6cells/kg, or more cells/kg, including all intervening doses of cells.

In another illustrative embodiment, the effective amount of transduced cells provided to a subject is from about 2 x 10^6cells/kg to about 10 x 10^6cells/kg, about 3 x 10^6cells/kg to about 10 x 10^6cells/kg, about 4 x 10^6cells/kg to about 10 x 10^6cells/kg, about 5 x 10^6cells/kg to about 10 x 10^6cells/kg, 2 x 10^6cells/kg to about 6 x 10^6cells/kg, 2 x 10^6cells/kg to about 7 x 10^6cells/kg, 2 x 10^6cells/kg to about 8 x 10^6cells/kg, 3 x 10^6cells/kg to about 6 x 10^6cells/kg, 3 x 10^6cells/kg to about 7 x 10^6cells/kg, 3 x 10^6cells/kg to about 8 x 10^6cells/kg, 4 x 10^6cells/kg to about 6 x 10^6cells/kg, 4 x

10^6cells/kg to about 7 x 10^6cells/kg, 4 x 10^6cells/kg to about 8 x 10^6cells/kg, 5 x 10^6cells/kg to about 6 x 10^6cells/kg, 5 x 10^6cells/kg to about 7 x 10^6cells/kg, 5 x 10^6cells/kg to about 8 x 10^6cells/kg, or 6 x 10^6cells/kg to about 8 x 10^6cells/kg, including all intervening doses of cells.

In certain embodiments, it can generally be stated that a pharmaceutical composition comprising the genetically modified cells described herein may be administered at a dosage of 10^2 to 10^{10} cells/kg body weight, preferably 10^5 to 10^7 cells/kg body weight, including but not limited to 1 x 10^6 cells/mL, 2 x 10^6 cells/mL, 3 x 10^6 cells/mL, 4 x 10^6 cells/mL, 5 x 10^6 cells/mL, 6 x 10^6 cells/mL, 7 x 10^6 cells/mL, 8 x 10^6 cells/mL, 9 x 10^6 cells/mL, 10 x 10^6 cells/mL, and all integer values within those ranges. The number of cells will depend upon the ultimate use for which the composition is intended as will the type of cells included therein. For uses provided in some embodiments, the cells are generally in a volume of a liter or less, can be 500 mLs or less, even 250 mLs or 100 mLs or less. Hence the density of the desired cells in particular embodiments is typically greater than 10^6 cells/mL, 10^7 cells/mL, or 10^8 cells/mL. The clinically relevant number of cells can be apportioned into multiple infusions that cumulatively equal or exceed 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, or 10^{12} cells. Cell-based compositions may be administered multiple times at dosages within these ranges. The cells may be allogeneic, syngeneic, xenogeneic, or autologous to the patient undergoing therapy.

Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject.

One of ordinary skill in the art would be able to use routine methods in order to determine the appropriate route of administration and the correct dosage of an effective amount of a composition comprising transduced cells or gene therapy vectors contemplated herein.

In particular embodiments, multiple administrations of pharmaceutical compositions contemplated herein may be required to effect therapy. In particular embodiments, the drug product is administered once. In certain embodiments, the drug product is administered 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more times over a span of 1 year, 2 years, 5, years, 10 years, or more.

All publications, patent applications, and issued patents cited in this specification are herein incorporated by reference as if each individual publication, patent application, or issued patent were specifically and individually indicated to be incorporated by reference.

Although the foregoing embodiments have been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings contemplated herein that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims. The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

EXAMPLES

EXAMPLE 1

CONSTRUCTION OF TPP1 VECTORS

Third generation lentiviral vectors containing a chimeric 5′ LTR; a myeloproliferative sarcoma virus enhancer, negative control region deleted, dl587rev primer-binding site substituted (MND) promoter or a short elongation factor 1 alpha (EF1α) promoter; a polynucleotide encoding a tripeptidyl peptidase 1(TPP1) polypeptide; and a self-inactivating (SIN) 3′ LTR were constructed. *See e.g.,* Figure 1 and SEQ ID NOs: 1 and 2. Tables 1 and 2 show the Identity, Genbank Reference, Source Name and Citation for the various nucleotide segments of exemplary lentiviral vectors encoding TPP1.

Table 1: pMND-TPP1 LVV

Nucleotides	Identity	GenBank Reference	Source Name	Citation
1-185	pUC19 plasmid backbone	Accession #L09137.2 nt 1 – 185	pUC19	New England Biolabs (Attachment 1)
185-222	Linker	Not applicable	Synthetic	Not applicable[1]
223-800	CMV	Not Applicable	pHCMV	(1994) PNAS 91: 9564-68
801-1136	R, U5, PBS, and packaging sequences	Accession #M19921.2 nt 454-789	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
1137-1139	Gag start codon (ATG) changed to stop codon (TAG)	Not Applicable[1]	Synthetic	Not applicable
1244-1595	HIV-1 gag sequence	Accession #M19921.2 nt 897-1248	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
1596-1992	HIV-1 pol cPPT/CTS	Accession #M19921.2 nt 4745-5125	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
1993-2517	HIV-1, isolate HXB3 env region (RRE)	Accession #M14100.1 nt 1875-2399	PgTAT-CMV	Malim, M. H. Nature (1988) 335:181-183
2518-2693	HIV-1 env sequences S/A	Accession #M19921.2 nt 8290-8470	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
2694-2708	Linker	Not applicable	Synthetic	Not applicable[1]

2709-3096	MND	Not applicable	pccl-c-MNDU3c-x2	Challita *et al.* (1995) J.Virol. 69: 748-755
3097-3124	Linker	Not applicable	Synthetic	Not applicable
3125-4822	TPP1, human	Not applicable	Synthetic	Not applicable
4823-4933	HIV-1 ppt and part of U3	Accession #M19921.2 nt 9005-9110	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
4934-5050	HIV-1 part of U3 (399bp deletion) and R	Accession #M19921.2 nt 9511-9627	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
5051-5074	Synthetic polyA	Not applicable	Synthetic	Levitt, N. Genes & Dev (1989) 3:1019-1025
5075-5093	Linker	Not applicable	Synthetic	Not Applicable
5094-7566	pUC19 backbone	Accession #L09137.2 nt 2636-2686	pUC19	New England Biolabs (Attachment 1)

Table 2: pEF1α-TPP1 LVV

Nucleotides	Identity	GenBank Reference	Source Name	Citation
1-185	pUC19 plasmid backbone	Accession #L09137.2 nt 1 – 185	pUC19	New England Biolabs (Attachment 1)
185-222	Linker	Not applicable	Synthetic	Not applicable
223-800	CMV	Not Applicable	pHCMV	(1994) PNAS 91: 9564-68
801-1136	R, U5, PBS, and packaging sequences	Accession #M19921.2 nt 454-789	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
1137-1139	Gag start codon (ATG) changed to stop codon (TAG)	Not Applicable	Synthetic	Not applicable
1140-1240	HIV-1 gag sequence	Accession #M19921.2 nt 793-893	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
1241-1243	HIV-1 gag sequence changed to a second stop codon	Not Applicable	Synthetic	Not applicable
1244-1595	HIV-1 gag sequence	Accession #M19921.2 nt 897-1248	pNL4-3	Maldarelli, et.al. (1991) J Virol:

				65(11):5732-43
1596-1992	HIV-1 pol cPPT/CTS	Accession #M19921.2 nt 4745-5125	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
1993-2517	HIV-1, isolate HXB3 env region (RRE)	Accession #M14100.1 nt 1875-2399	PgTAT-CMV	Malim, M. H. Nature (1988) 335:181-183
2518-2693	HIV-1 env sequences S/A	Accession #M19921.2 nt 8290-8470	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
2694-2698	Linker	Not applicable	Synthetic	Not applicable
2699-3242	EF1alpha/HTLV promoter	Not applicable	Synthetic	Takebe *et al.* (1988) MCB: 8(1): 466-472 Kim DW *et al.* (1990), Gene: 91(2): 217-223
3243-3258	Linker	Not applicable	Synthetic	Not applicable
3259-4959	TPP1, human	Not applicable	Synthetic	Not applicable
4960-5067	HIV-1 ppt and part of U3	Accession #M19921.2 nt 9005-9110	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
5068-5184	HIV-1 part of U3 (399bp deletion) and R	Accession #M19921.2 nt 9511-9627	pNL4-3	Maldarelli, et.al. (1991) J Virol: 65(11):5732-43
5185-5208	Synthetic polyA	Not applicable	Synthetic	Levitt, N. Genes & Dev (1989) 3:1019-1025
5209-5227	Linker	Not applicable	Synthetic	Not Applicable
5228-7700	pUC19 backbone	Accession #L09137.2 nt 2636-2686	pUC19	New England Biolabs (Attachment 1)

EXAMPLE 2

FIBROBLASTS TRANSDUCED WITH LENTIVIRAL VECTORS ENCODING TPP1

Human fibroblasts deficient in TPP1 activity because of homozygous mutations in the TPP1 gene (R127X/R208X; $TPP1^{-/-}$ cells) were acquired from the Coriell Institute Cell Repository (cell line GM16485). $TPP1^{-/-}$ cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) plus 10% fetal bovine serum (FBS) for twenty-four hours prior to transduction. Cultured $TPP1^{-/-}$ cells were resuspended at 5.0E4 cells/mL of DMEM plus

10% FBS and two mL of this cell suspension were plated per well in a 6-well tissue culture plate and placed at 37°C. Twenty-four hours post cell seeding, cells were transduced with one mL of either unpurified lentiviral vector (titers of 1.45E8 TU/mL and 8.80E7 TU/mL). One mL of DMEM plus 10% FBS was added to a control well and the cells were replaced in a 37°C incubator. Twenty-four hours post transduction, a complete media exchange was performed. Forty-eight hours post transduction, 250uL of supernatant from each well was removed to a sterile Eppendorf tube and frozen at -80°C. Cells were washed with one mL phosphate buffered saline and lifted using 0.5mL of 1X TryplE Express Enzyme (Thermo Fisher). Cells were removed to two sterile Eppendorf tubes per sample and pelleted for five minutes at 1500rpm. The supernatant was aspirated and cell pellets were frozen at -80°C.

EXAMPLE 3

PROTEIN EXPRESSION IN CELLS TRANSDUCED WITH LENTIVIRAL VECTORS ENCODING TPP1

Frozen cell pellets from wild type control cells, $TPP1^{-/-}$ cells, and $TPP1^{-/-}$ cells transduced with the lentiviral vectors encoding TPP1 (pMND-TPP1 and pEF1α-TPP1) were thawed on ice for Western blotting. 300µL of mammalian protein extraction reagent and 3µL of 100X HALT protease inhibitor cocktail (ThermoFisher) were added to each cell pellet. Pellets were resuspended by pipetting gently up and down and cells were incubated for 10 minutes at room temperature on a plate rocker. Cells were centrifuged for fifteen minutes at 4^0C at 14,000 rpm and supernatants were removed to sterile Eppendorf tubes. Loading dye was prepared by adding 25µL β-mercaptoethanol to 475µL 4X Laemmli sample buffer (Bio-Rad). Samples were mixed in a 3:1 sample to loading dye ratio with 30µL prepared loading dye to 90µL sample. 20µL of each sample and 8µL Precision Plus Protein Kaleidoscope ladder were loaded into the wells of a NuPage 4-12 Bis-Tris protein gel. Gels were run in 1X MES SDS running buffer for 40 minutes at 200V.

Gels were transferred using an iBlot transfer stack on the iBlot 7 minute transfer system. Membranes were rinsed in 1X Tris-buffered saline for five minutes at room temperature. Membranes were incubated in Odyssey blocking buffer plus a 1:500 dilution of rabbit anti-TPP1 antibody (Abcam ab96498) and a 1:1000 dilution of mouse anti-β-actin antibody (Abcam ab3280) at 4^0C. The next morning, membranes were rinsed three times in Tris-buffered saline for five minutes at room temperature. A secondary antibody

cocktail containing a 1:1000 dilution of 800RD donkey anti-mouse IgG (Licor 926-32212) and a 1:1000 dilution of 680RD donkey anti-rabbit IgG (Licor 926-68073) in Odyssey blocking buffer. Membranes were incubated for one hour at room temperature in secondary antibody cocktail and rinsed three times with Tris-buffered saline for five minutes at room temperature. Blots were imaged on a Licor Odyssey CLX imaging system.

A representative Western blot of TPP1 expression in wild type control cells, $TPP1^{-/-}$ cells, and $TPP1^{-/-}$ cells transduced with the lentiviral vectors encoding TPP1 (pMND-TPP1 and pEF1α-TPP1) is shown in Figure 2.

EXAMPLE 4

RESTORATION OF TPP1 ACTIVITY IN $TPP1^{-/-}$ CELLS TRANSDUCED WITH LENTIVIRAL VECTORS ENCODING TPP1

Cell pellets from wild type control cells, TPP1-/- cells, and TPP1-/- cells transduced with the lentiviral vectors encoding TPP1 (pMND-TPP1 and pEF1α-TPP1) were resuspended in 150µL of acetate buffer (0.1 M Sodium Acetate (NaAc), 0.15 M Sodium Chloride(NaCl) (pH 4.0), 10uM each Pepstatin A and trans-Epoxysuccinyl-L-leucylamido-(4-guanidino)butane (E64)). Fluorometric measurement of TPP-1 activity was calculated based on cleavage of the Ala-Ala-Phe-7-Amido-4-methylcoumarin substrate (AAF-MCA) as described previously, with some modifications (Page *et al.*, 1993. *Arch Biochem Biophys.* 306(2):354-9; Lukacs *et al.*, 2003. *Clin Chem.* 49(3):509-11; Ezaki *et al.*, 2000. *Biochem Biophys Res Commun.* 268(3):904-8.). 15 to 25µg total protein of cell lysate or cell supernatant was incubated in 150µL acetate buffer with a final concentration of 62.5µM AAF-MCA for 20 hours at 37°C. The assay was stopped by the addition of 100µL 0.5M EDTA (pH 12.0). Fluorescence was measured using a Molecular Devices SpectraMax M2 spectrofluorimeter (Ex. 355, Em. 460).

Figure 3A shows the results from a representative experiment assaying TPP1 enzymatic activity in wild type control cells, $TPP1^{-/-}$ cells, and $TPP1^{-/-}$ cells transduced with the lentiviral vectors encoding TPP1 (pMND-TPP1 and pEF1α-TPP1).

TPP1 overexpression in transduced cells also led to secretion of enzymatically active TPP1 in the cell culture supernatant. TPP1 activity in both patient and wild type fibroblasts remained at background levels; whereas overexpression of TPP1 in transduced

TPP1$^{-/-}$ fibroblasts increased TPP1 activity in the supernatant 10-fold. Figure 3B. Thus, TPP1 gene therapy not only corrects transduced cells, but also has the potential to correct TPP1 deficiency in neighboring cells.

EXAMPLE 5

HUMAN TPP1 -/- NEURONS TRANSDUCED WITH LVV ENCODING TPP1

TPP1 deficient patient neurons progressively accumulate storage material with classical NCL features, and have additional morphological changes in lysosomal and endoplasmic reticulum compartments (Lojewski *et al.*, HMG, 2014, v23, pp2005-2022). To assess the impact of lentiviral-mediated TPP1 replacement in the human neuronal cell based system on these known phenotypes recombinant TPP1 protein levels will be determined, TPP1 enzyme activity levels will be determined, and neuronal morphology will be assessed, *e.g.*, to determine the size and number of storage deposits (using established autofluorescence and subunit c immunostaining procedures).

Patient-derived TPP1 $^{-/-}$ induced pluripotent stem cells (iPSC) were differentiated into neurons. The neurons from patients with confirmed CLN2 mutations were transduced at an MOI of 5 with lentiviral vectors comprising an MND or EF1α-short promoter operably linked to a polynucleotide encoding TPP1. Transduced cells were fixed and imaged using confocal microscopy at 17 days post-transduction/post-differentiation, as neural progenitors and as differentiated neurons. Antibodies specific to the LLV expressed protein, TPP1, and the CLN2 lipofuscin storage material, ATP synthase subunit c, were used as co-stains to visualize the two simultaneously in the same cells.

No expression of TPP1 was detected in the untransduced control cells and subunit c was abundant. In transduced cells, staining of TPP1 is visible, but the staining for subunit c is substantially reduced or undetectable. Figure 9. The absence of subunit c in the transduced patient derived cells shows that the LVV expressed TPP1 is catalytically active against the pathologic lipofuscin.

EXAMPLE 6

IN VIVO TPP1 GENE THERAPY MODEL

Mice with TPP1 mutations will be administered HSCs transduced with lentiviral vectors encoding TPP1 and phenotypically characterized. TPP1 mutant mice will undergo treatment to ablate bone marrow hematopoietic stem cells and administered HSCs transduced with lentiviral vectors encoding TPP1 at no more than 2 weeks of age.

Clinical assessment will be performed beginning the first day after initial treatment, and, at ~4 weeks of age, mice will undergo clinical assessment, which includes observation for tremors, general body condition, weight gain (weekly, starting at ~4 weeks of age), grip strength (biweekly, beginning at ~8 weeks of age), rotarod (at ~13, 18 weeks of age), and gait analysis (at ~16 and ~24 weeks of age).

In addition to the behavioral assays, mice will be tested post-transplant for other parameters to assess their general health and immune system reconstitution after hematopoietic stem cell therapy including full clinical blood chemistry panels, CNS gross morphology and histological analysis to assess storage material, neuronal and glial cell numbers, and morphology (*e.g.*, axonal degeneration) in sagittal sections (to capture multiple brain regions in each section), evidence of cross-correction (expression) in tissues affected by TPP1 deficiency, TPP1 enzyme activity in blood/brain/tissue lysates, bone marrow morphology, measurement of vector copy number in mouse bone marrow at the end of all experiments; and identification of engrafted cells.

EXAMPLE 7

F108 AND PGE_2 INCREASE TRANSDUCTION EFFICIENCY AND VCN IN HSCS TRANSDUCED WITH A CLN2 LENTIVIRAL VECTOR

Human CD34+ cells were transduced with a lentiviral vector (LVV) comprising an EF1α promoter or MND promoter operably linked to a polynucleotide encoding CLN2. Cells were prestimulated in cytokine containing media for 48 hours and transduced for 24 hours at an MOI of either 5 or 15 using either standard transduction conditions of 8 μg/mL protamine sulfate (PS) or 200 μg/mL F108 (poloxamer 338, BASF) and 10 μM PGE_2 (Cayman). After transduction, cells were plated in

methylcellulose and cultured for approximately 14 days to allow for hematopoietic progenitor colony formation. Colonies were pooled for VCN analysis.

Transduction in protamine sulfate yielded VCNs below 1.5 and VCN increased with increasing MOI. Substantial increases in VCN were observed in cells transduced in the presence of F108 and PGE_2 at either MOI (Figure 4, top left panel). After transduced cells were cultured in cytokines for 7 days; cell pellets and supernatants were assayed for TPP-1 activity. TPP-1 activity was higher in cells with higher VCNs, *i.e.*, cells transduced in the presence of F108 and PGE_2 (Figure 4, top middle and top right panels).

Mouse lineage-depleted bone marrow was transduced with the MND-TPP1 LVV at MOIs of 5, 15, or 30, in the presence of PS or F108 and PGE_2. VCN was increased in cytokine cultured cells at day 7 at all MOIs of 5, 15, and 30 (Figure 4, bottom left panel). Individual colonies were plucked and analyzed by qPCR for VCN from mouse hematopoietic progenitors; F108 and PGE_2 increased VCN in these cells (Figure 4, lower middle panel). F108 and PGE_2 also substantially increased transduction efficiency (Figure 4, lower right panel).

EXAMPLE 8

F108 AND PGE_2 INCREASE TRANSDUCTION EFFICIENCY AND VCN IN HUMAN CD34+ CELLS TRANSDUCED WITH A CLN2 LENTIVIRAL VECTOR

Human CD34+ cells were transduced with a lentiviral vector (LVV) comprising an EF1α promoter or MND promoter operably linked to a polynucleotide encoding CLN2. Cells were prestimulated in cytokine containing media for 48 hours and transduced for 24 hours at an MOI of either 5, 15, or 30 using either standard transduction conditions of 8 μg/mL protamine sulfate (PS) or 200 μg/mL F108 (poloxamer 338, BASF) and 10 μM PGE_2 (Cayman). After transduction, cells were plated in methylcellulose and cultured for approximately 14 days to allow for hematopoietic progenitor colony formation or cultured in cytokine containing media for 14 days. Cell growth, VCN from liquid cultures, individual colony VCN and %LVV+ cells from Day 14 methylcellulose cultures, and TPP1 activity were measured.

Cell growth was not adversely affected by transducing the $CD34^+$ cells with lentiviral vector in the presence of F108 and PGE_2. Figure 5.

VCN was measured in transduced cells cultured in cytokines at 7 days and 14 days. F108 and PGE_2 increased transduction across all MOIs for both lentiviral vectors. Figure 6.

Individual colonies were plucked from day 14 methylcellulose cultures and analyzed by qPCR for VCN, %LVV+ cells, and colony formation. F108 and PGE_2 increased VCN and %LVV+ cells (Figure 7, left panel). F108 and PGE_2 did not significantly affect colony formation (Figure 7, right panel).

Day 14 cell pellets and day 7 supernatants from liquid cultures were assayed for TPP-1 activity. TPP-1 activity was higher in cells with higher VCNs. Figure 8

In general, in the following claims, the terms used should not be construed to limit the claims to the specific embodiments disclosed in the specification and the claims, but should be construed to include all possible embodiments along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

CLAIMS

What is claimed is:

1. A polynucleotide comprising:

(a) a left (5′) lentiviral LTR;

(b) a Psi (ψ) packaging signal;

(c) a retroviral export element;

(d) a central polypurine tract/DNA flap (cPPT/FLAP);

(e) a promoter operably linked to a polynucleotide encoding a tripeptidyl peptidase 1(TPP1) polypeptide; and

(f) a right (3′) lentiviral LTR.

2. The polynucleotide of claim 1, wherein the lentivirus is selected from the group consisting of: HIV (human immunodeficiency virus; including HIV type 1, and HIV type 2); visna-maedi virus (VMV) virus; caprine arthritis-encephalitis virus (CAEV); equine infectious anemia virus (EIAV); feline immunodeficiency virus (FIV); bovine immune deficiency virus (BIV); and simian immunodeficiency virus (SIV).

3. The polynucleotide of claim 1 or claim 2, wherein the lentivirus is HIV-1 or HIV-2.

4. The polynucleotide of any one of claims 1-3, wherein the lentivirus is HIV-1.

5. The polynucleotide of any one of claims 1-4, wherein the promoter of the 5′ LTR is replaced with a heterologous promoter selected from the group consisting of: a cytomegalovirus (CMV) promoter, a Rous Sarcoma Virus (RSV) promoter, and a Simian Virus 40 (SV40) promoter.

6. The polynucleotide of any one of claims 1-5, wherein the 3′ LTR comprises one or more modifications.

7. The polynucleotide of any one of claims 1-6, wherein the 3′ LTR comprises one or more deletions that prevent viral transcription beyond the first round of viral replication.

8. The polynucleotide of any one of claims 1-6, wherein the 3′ LTR comprises a deletion of the TATA box and Sp1 and NF-κB transcription factor binding sites in the U3 region of the 3′ LTR.

9. The polynucleotide of any one of claims 1-6, wherein the 3′ LTR is a self-inactivating (SIN) LTR.

10. The polynucleotide of any one of claims 1-9, wherein the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide is selected from the group consisting of: integrin subunit alpha M (ITGAM; CD11b), CD68, C-X3-C motif chemokine receptor 1 (CX3CR1), ionized calcium binding adaptor molecule 1 (IBA1), transmembrane protein 119 (TMEM119), spalt like transcription factor 1 (SALL1) and adhesion G protein-coupled receptor E1 (F4/80).

11. The polynucleotide of any one of claims 1-9, wherein the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide comprises a myeloproliferative sarcoma virus enhancer, negative control region deleted, dl587rev primer-binding site substituted (MND) promoter or transcriptionally active fragment thereof.

12. The polynucleotide of any one of claims 1-9, wherein the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide comprises an elongation factor 1 alpha (EF1α) promoter or transcriptionally active fragment thereof.

13. The polynucleotide of any one of claims 1-9, wherein the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide is a short EF1α promoter.

14. The polynucleotide of any one of claims 1-9, wherein the promoter operably linked to a polynucleotide encoding a TPP1 polypeptide is a long EF1α promoter.

15. The polynucleotide of any one of claims 1-14, wherein the polynucleotide encoding the TPP1 polypeptide is a cDNA.

16. The polynucleotide of any one of claims 1-15, wherein the polynucleotide encoding the TPP1 polypeptide is codon optimized for expression.

17. A polynucleotide comprising:

(a) a left (5′) HIV-1 LTR;

(b) a Psi (ψ) packaging signal;

(c) an RRE retroviral export element;

(d) a cPPT/FLAP;

(e) an MND promoter or EF1α promoter operably linked to a polynucleotide encoding a TPP1 polypeptide; and

(f) a right (3′) HIV-1 LTR.

18. A polynucleotide comprising:

(a) a left (5′) CMV promoter/HIV-1 chimeric LTR;

(b) a Psi (ψ) packaging signal;

(c) an RRE retroviral export element;

(d) a cPPT/FLAP;

(e) an MND promoter or EF1α promoter operably linked to a polynucleotide encoding a TPP1 polypeptide; and

(f) a right (3′) SIN HIV-1 LTR.

19. The polynucleotide of any one of claims 1-18, further comprising a bovine growth hormone polyadenylation signal or a rabbit β-globin polyadenylation signal.

20. A mammalian cell transduced with a lentiviral vector comprising a polynucleotide according to any one of claims 1-19.

21. The mammalian cell of claim 20, wherein the cell is a hematopoietic cell.

22. The mammalian cell of claim 20 or claim 21, wherein the cell is a $CD34^+$ cell.

23. The mammalian cell of any one of claims 20-22, wherein the cell is a stem cell or progenitor cell.

24. A producer cell comprising: a first polynucleotide encoding gag, a second polynucleotide encoding pol, a third polynucleotide encoding env, and a polynucleotide according to any one of claims 1-19.

25. A lentiviral vector produced by the producer cell of claim 24.

26. A composition comprising a lentiviral vector comprising a polynucleotide according to any one of claims 1-19 or a mammalian cell according to any one of claims 20-23.

27. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and a lentiviral vector comprising a polynucleotide according to any one of claims 1-19 or a mammalian cell according to any one of claims 20-23.

28. A method of treating neuronal ceroid lipofuscinoses (NCL), comprising administering to a subject a lentiviral vector comprising a polynucleotide according to any one of claims 1-19; a cell transduced with a lentiviral vector comprising a polynucleotide according to any one of claims 1-19; or a mammalian cell according to any one of claims 20-23.

29. A method of treating neuronal ceroid lipofuscinoses, comprising administering to a subject a pharmaceutical composition of claim 27.

30. A method of decreasing at least one symptom associated with neuronal ceroid lipofuscinoses in a subject comprising administering to a subject a lentiviral vector comprising a polynucleotide according to any one of claims 1-19; a cell transduced with a lentiviral vector comprising a polynucleotide according to any one of claims 1-19; or a mammalian cell according to any one of claims 20-23.

31. A method of decreasing at least one symptom associated with neuronal ceroid lipofuscinoses in a subject comprising administering to a subject a pharmaceutical composition of claim 27.

32. The method of claim 30 or claim 31, wherein the at least one symptom is selected from the group consisting of: seizures, loss of vision, cognitive function decline, and motor function decline.

33. The method of any one of claims 28-32, wherein the subject has been diagnosed with late-infantile NCL (LINCL).

34. The method of any one of claims 28-32, wherein the subject has been diagnosed with juvenile Batten Disease (JNCL).

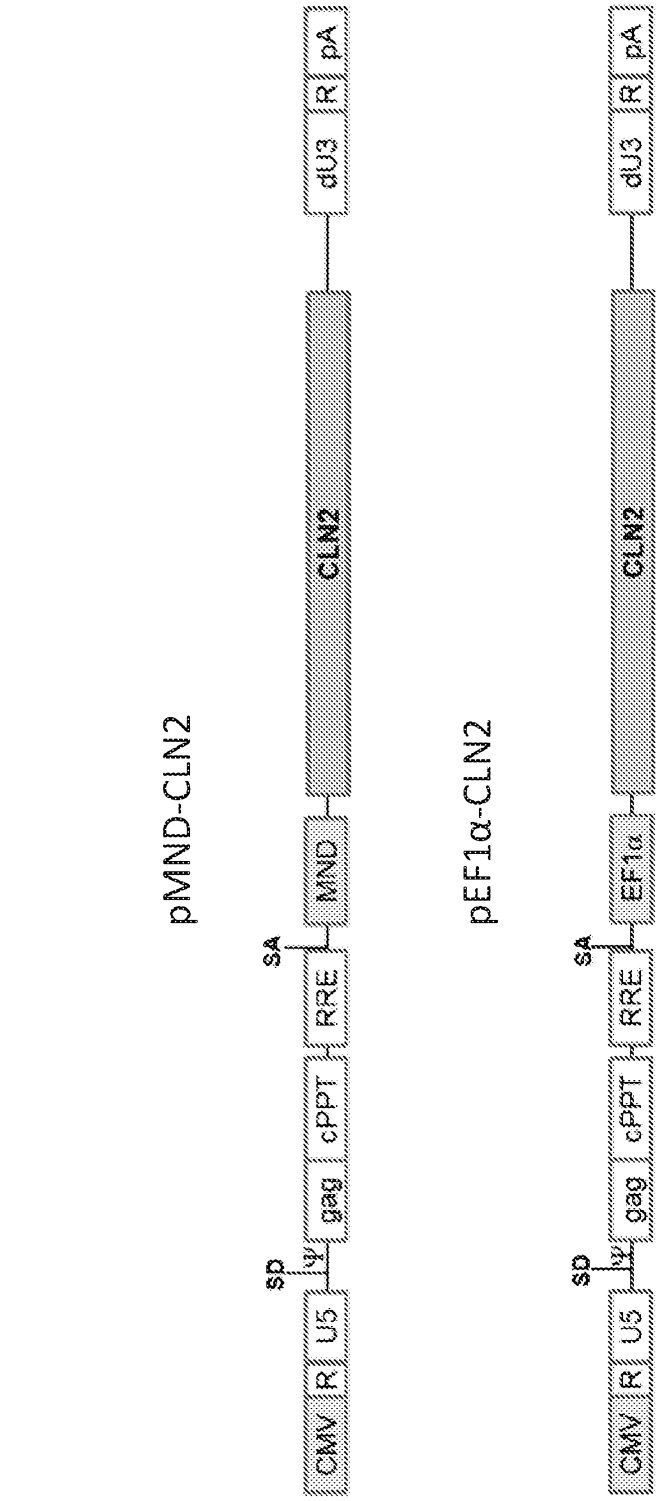

pMND-CLN2
CMV
R
U5
SD
ψ
gag
cPPT
RRE
SA
MND
CLN2
dU3
R
pA
pEF1α-CLN2
CMV
R
U5
SD
ψ
gag
cPPT
RRE
SA
EF1α
CLN2
dU3
R
pA

Figure 1

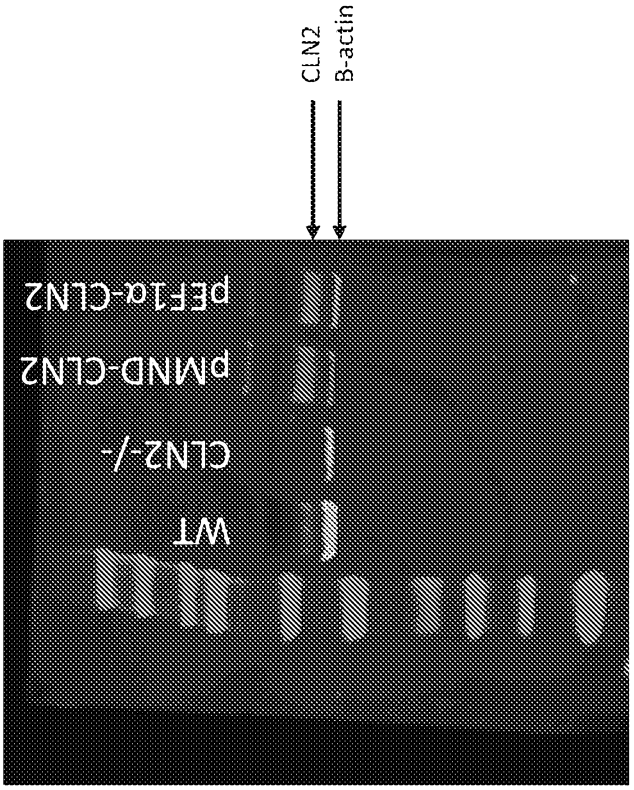


Figure 2

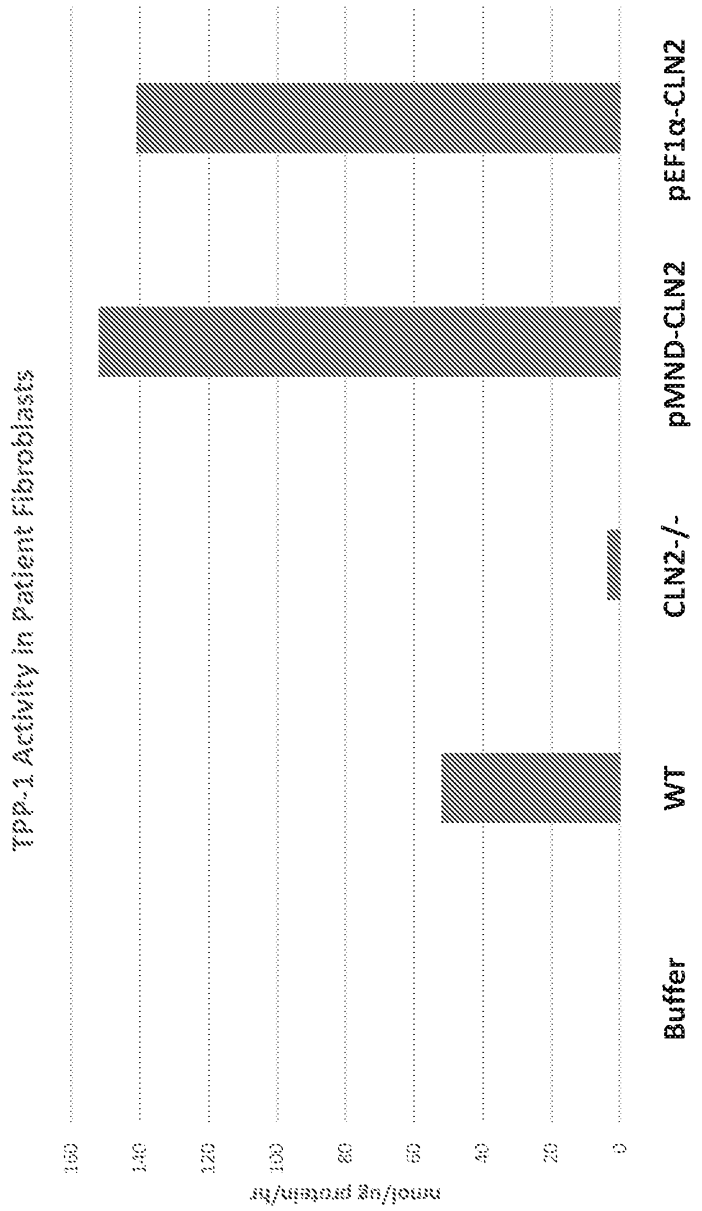

TPP-1 Activity in Patient Fibroblasts
nmol/μL protein/hr
160
140
120
100
80
60
40
20
0
Buffer
WT
CLN2-/-
pMND-CLN2
pEF1α-CLN2

Figure 3A

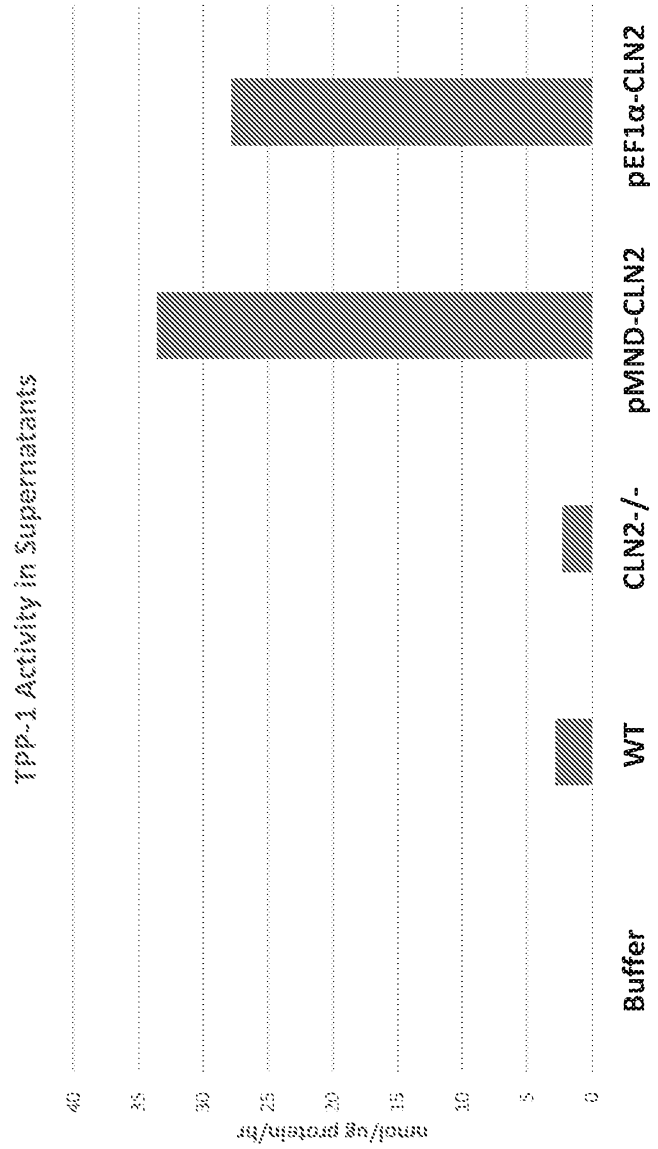
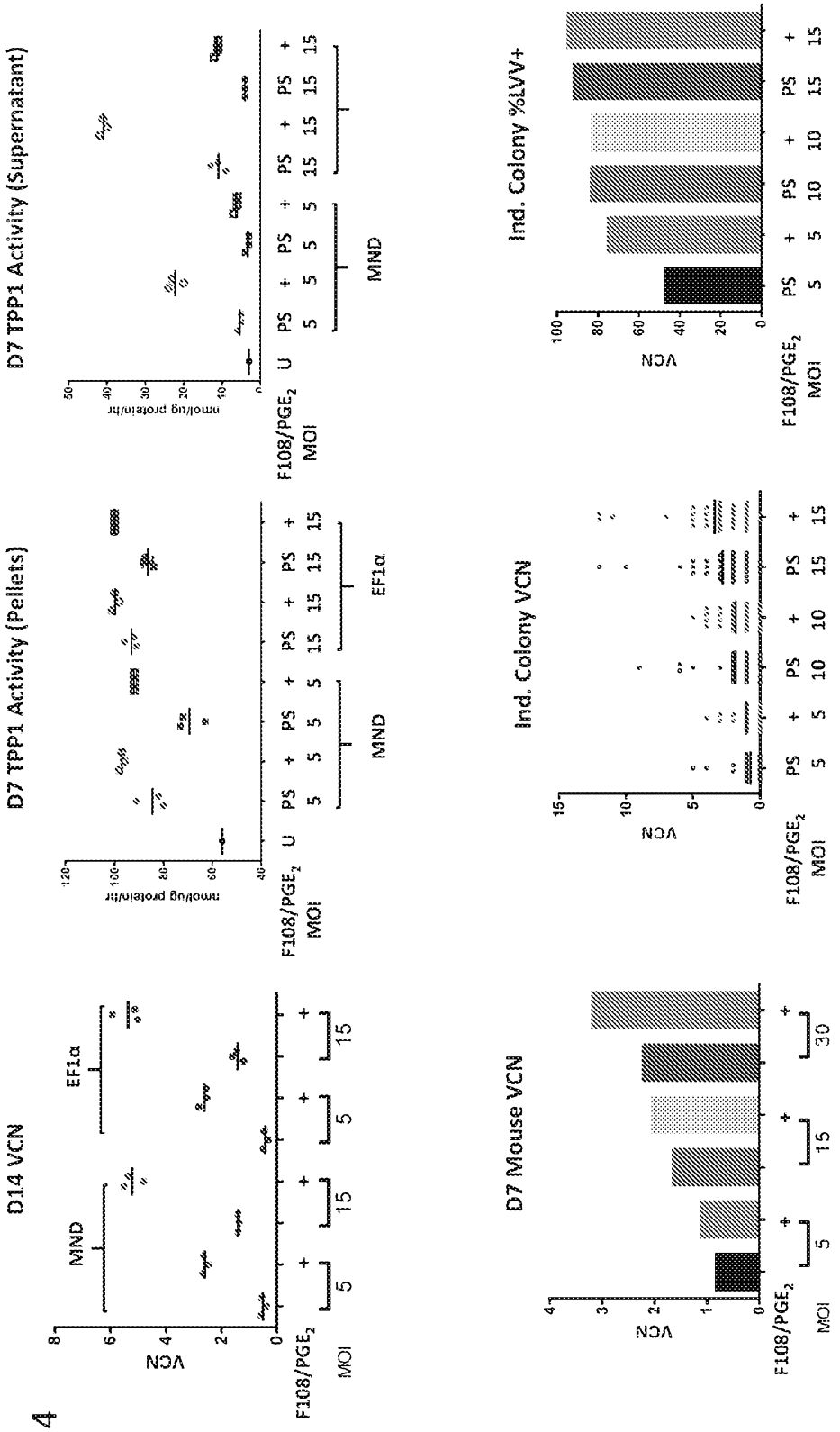

TPP-1 Activity in Supernatants
nmol/ug protein/hr
40
35
30
25
20
15
10
5
0
Buffer
WT
CLN2-/-
pMND-CLN2
pEF1α-CLN2

Figure 3B

Figure 4


D14 VCN
MND
EF1α
VCN
F108/PGE2
MOI
D7 TPP1 Activity (Pellets)
nmol/ug protein/hr
D7 TPP1 Activity (Supernatant)
D7 Mouse VCN
Ind. Colony VCN
Ind. Colony %LV+

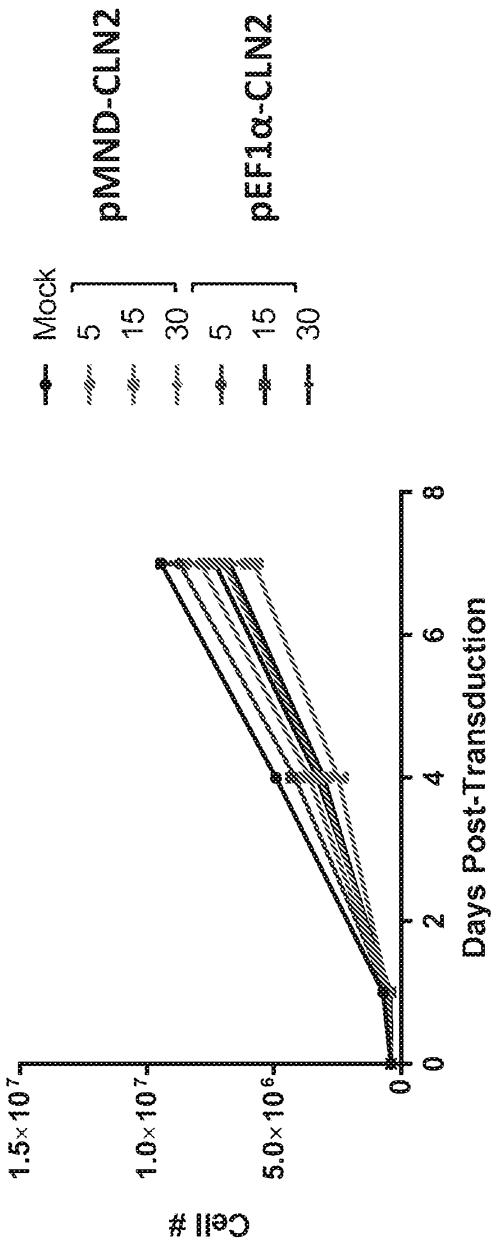

pMND-CLN2
pEF1α-CLN2
Mock
5
15
30
5
15
30
Cell #
1.5×10^7
1.0×10^7
5.0×10^6
0
0
2
4
6
8
Days Post-Transduction

Figure 5

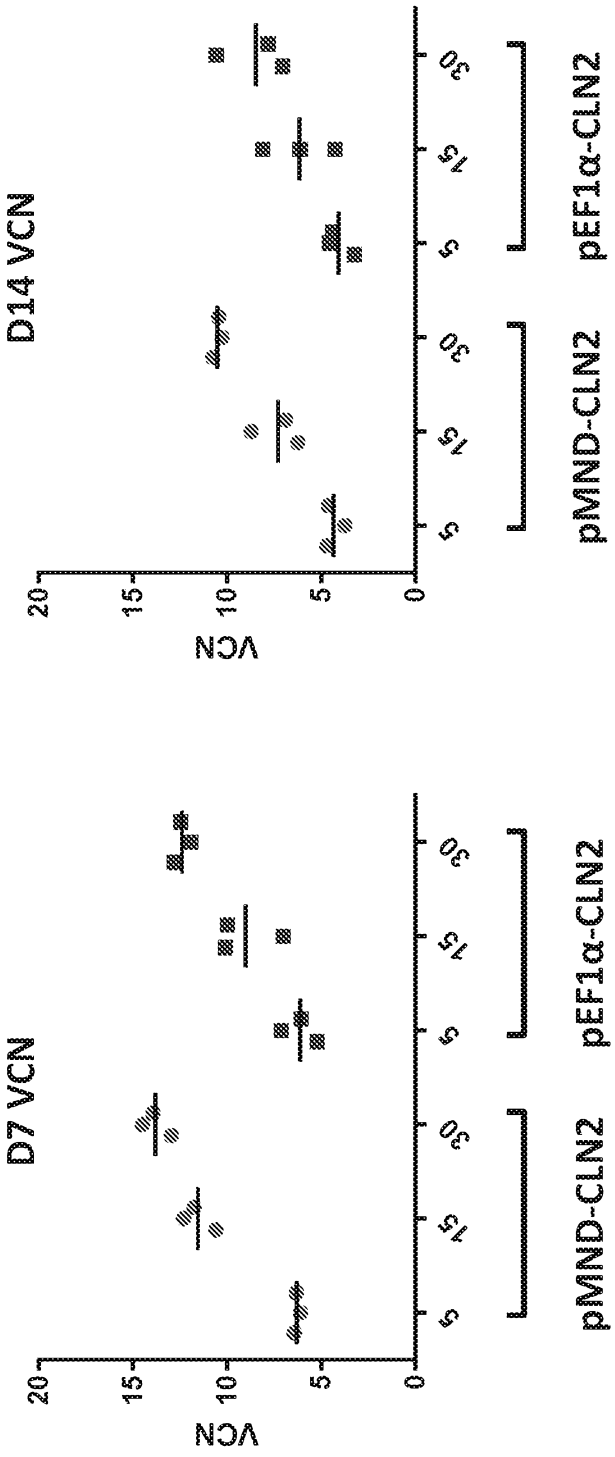
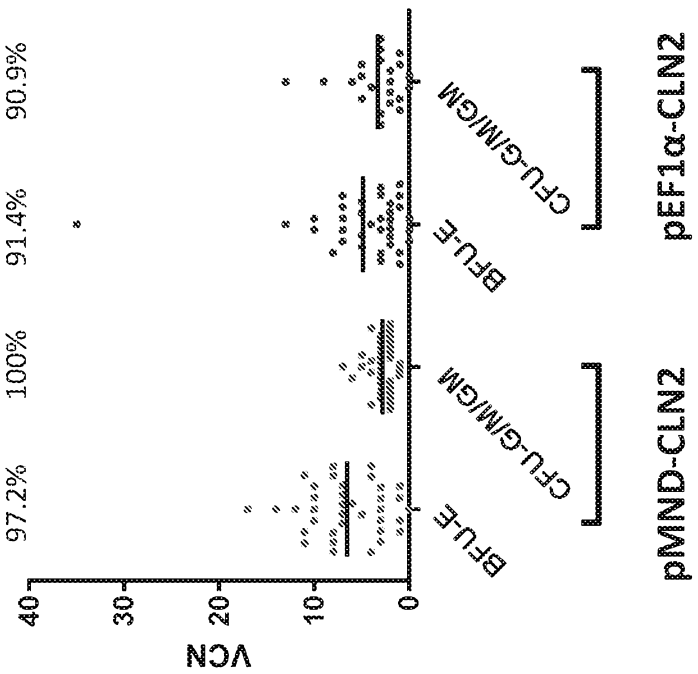

D14 VCN
VCN
20
15
10
5
0
5
15
30
5
15
30
pMND-CLN2
pEF1α-CLN2
D7 VCN
VCN
20
15
10
5
0
5
15
30
5
15
30
pMND-CLN2
pEF1α-CLN2

Figure 6


VCN
40
30
20
10
0
97.2%
100%
91.4%
90.9%
BFU-E
CFU-GM/GM
BFU-E
CFU-GM/GM
pMND-CLN2
pEF1α-CLN2

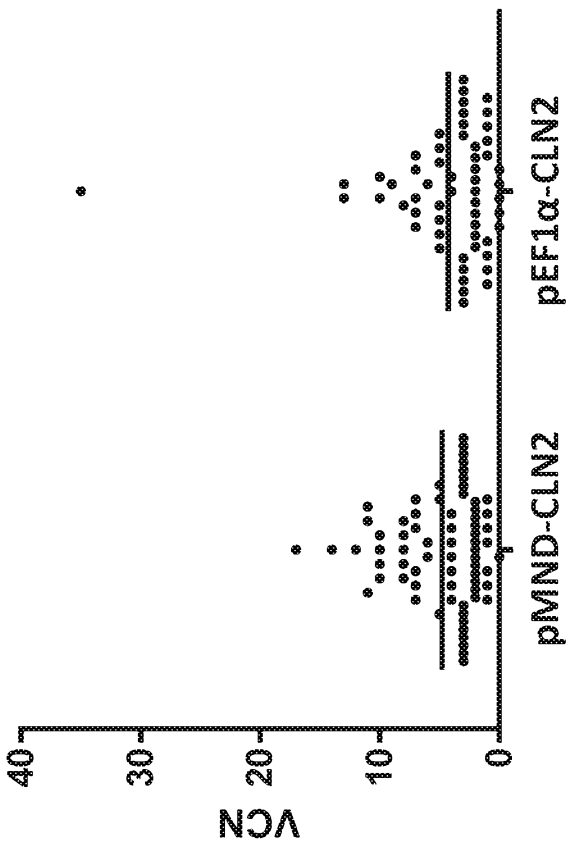

%LVV+
98.6%
91.2%
VCN
40
30
20
10
0
pMND-CLN2
pEF1α-CLN2

Figure 7

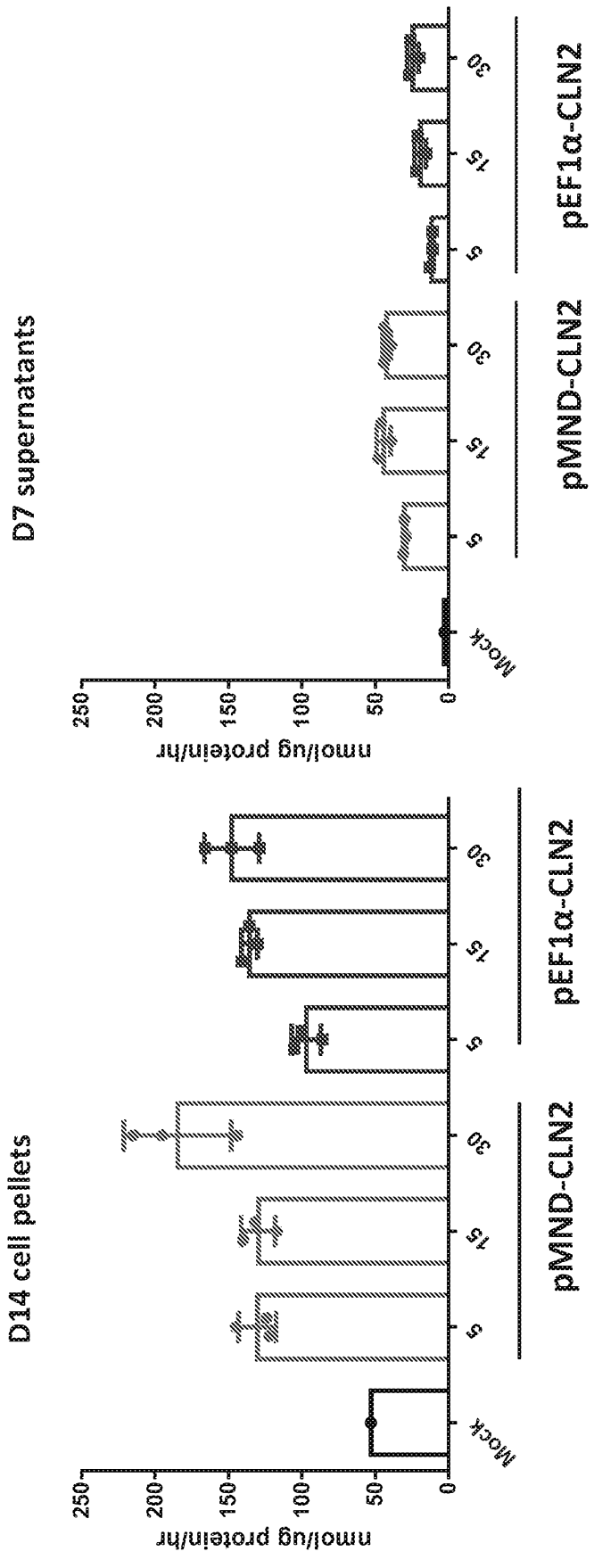

D14 cell pellets
nmol/ug protein/hr
250
200
150
100
50
0
Mock
5
15
30
pMND-CLN2
pEF1α-CLN2
D7 supernatants
nmol/ug protein/hr
250
200
150
100
50
0
Mock
5
15
30
pMND-CLN2
pEF1α-CLN2

Figure 8

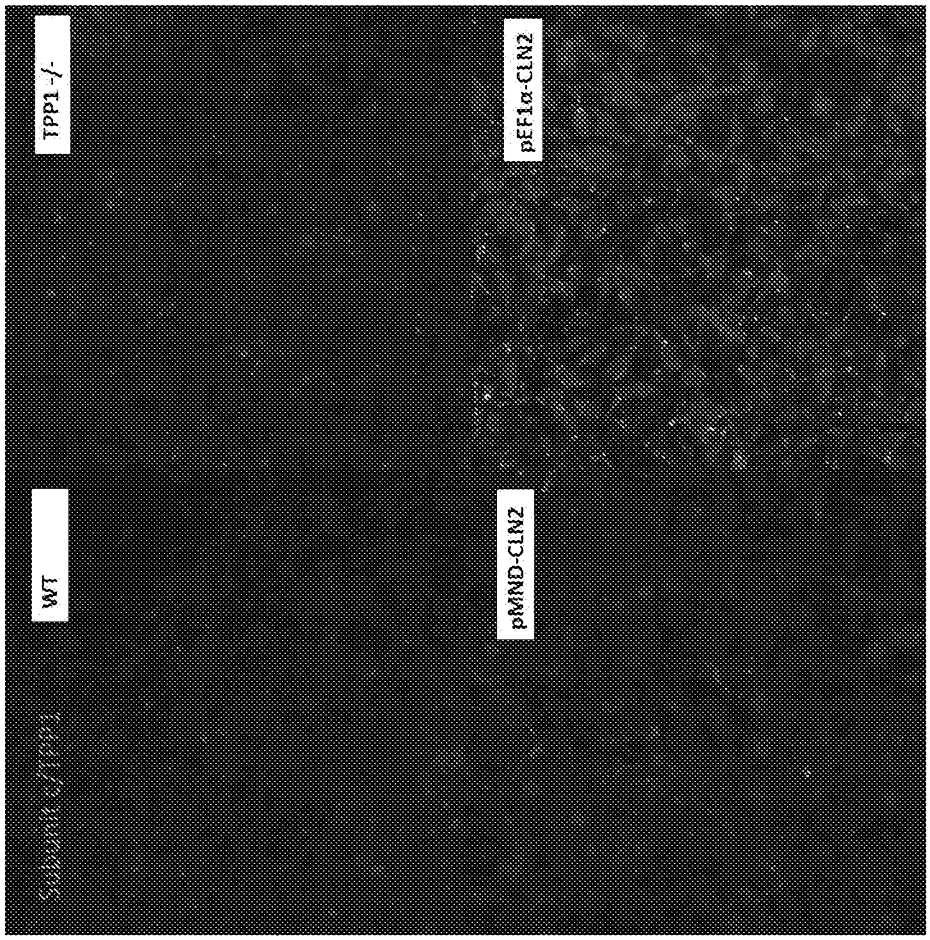


Figure 9

SEQUENCE LISTING

<110> bluebird bio, Inc.
Goss, Kendrick
Parsons, Geoffrey
Giniatullina, Asiya

<120> GENE THERAPY OF NEURONAL CEROID LIPOFUSCINOSES

<130> BLBD-070/02WO

<150> US 62/457,498
<151> 2017-02-10

<150> US 62/349,505
<151> 2016-06-13

<160> 20

<170> PatentIn version 3.5

<210> 1
<211> 7566
<212> DNA
<213> Artificial Sequence

<220>
<223> Nucleic acid sequence of a lentiviral vector encoding TTP1 (pMND-CLN2)

<400> 1
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca 60

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120

ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180

accatcatat gccagcctat ggtgacattg attattgact agttattaat agtaatcaat 240

tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300

tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360

tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420

aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480

caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540

tacttggcag tacatctacg tattagtcat cgctattacc atggtgatgc ggttttggca 600

gtacatcaat gggcgtggat agcggtttga ctcacgggga tttccaagtc tccaccccat 660

tgacgtcaat gggagtttgt tttggcacca aaatcaacgg gactttccaa aatgtcgtaa 720

caactccgcc ccattgacgc aaatgggcgg taggcgtgta cggtgggagg tctatataag 780

cagagctcgt ttagtgaacc gggtctctct ggttagacca gatctgagcc tgggagctct 840

ctggctaact agggaaccca ctgcttaagc ctcaataaag cttgccttga gtgctcaaag 900

tagtgtgtgc ccgtctgttg tgtgactctg gtaactagag atccctcaga ccctttagt 960

cagtgtggaa aatctctagc agtggcgccc gaacagggac ttgaaagcga aagtaaagcc 1020

agaggagatc tctcgacgca ggactcggct tgctgaagcg cgcacggcaa gaggcgaggg 1080

gcggcgactg gtgagtacgc caaaaatttt gactagcgga ggctagaagg agagagtagg 1140

gtgcgagagc gtcggtatta agcgggggag aattagataa atgggaaaaa attcggttaa 1200

ggccaggggg aaagaaacaa tataaactaa aacatatagt tagggcaagc agggagctag 1260

aacgattcgc agttaatcct ggccttttag agacatcaga aggctgtaga caaatactgg 1320

gacagctaca accatccctt cagacaggat cagaagaact tagatcatta tataatacaa 1380

tagcagtcct ctattgtgtg catcaaagga tagatgtaaa agacaccaag gaagccttag 1440

ataagataga ggaagagcaa aacaaaagta agaaaaaggc acagcaagca gcagctgaca 1500

caggaaacaa cagccaggtc agccaaaatt accctatagt gcagaacctc caggggcaaa 1560

tggtacatca ggccatatca cctagaactt taaattaaga cagcagtaca aatggcagta 1620

ttcatccaca attttaaaag aaaaggggg attgggggt acagtgcagg ggaaagaata 1680

gtagacataa tagcaacaga catacaaact aaagaattac aaaaacaaat tacaaaaatt 1740

caaaattttc gggtttatta cagggacagc agagatccag tttggaaagg accagcaaag 1800

ctcctctgga aaggtgaagg ggcagtagta atacaagata atagtgacat aaaagtagtg 1860

ccaagaagaa aagcaaagat catcagggat tatggaaaac agatggcagg tgatgattgt 1920

gtggcaagta gacaggatga ggattaacac atggaaaaga ttagtaaaac accatagctc 1980

tagagcgatc ccgatcttca gacctggagg aggagatatg agggacaatt ggagaagtga 2040

attatataaa tataaagtag taaaaattga accattagga gtagcaccca ccaaggcaaa 2100

gagaagagtg gtgcagagag aaaaaagagc agtgggaata ggagctttgt tccttgggtt 2160

cttgggagca gcaggaagca ctatgggcgc agcgtcaatg acgctgacgg tacaggccag 2220

acaattattg tctggtatag tgcagcagca gaacaatttg ctgagggcta ttgaggcgca 2280

acagcatctg ttgcaactca cagtctgggg catcaagcag ctccaggcaa gaatcctggc 2340

tgtggaaaga tacctaaagg atcaacagct cctggggatt tggggttgct ctggaaaact 2400

catttgcacc actgctgtgc cttggaatgc tagttggagt aataaatctc tggaacagat 2460

ttggaatcac acgacctgga tggagtggga cagagaaatt aacaattaca caagcttggt 2520

aggtttaaga atagtttttg ctgtactttc tatagtgaat agagttaggc agggatattc 2580

accattatcg tttcagaccc acctcccaac cccgagggga cccgacaggc ccgaaggaat 2640

agaagaagaa ggtggagaga gagacagaga cagatccatt cgattagtga acggatccat 2700

ctcgacggaa tgaaagaccc cacctgtagg tttggcaagc taggatcaag gttaggaaca 2760

gagagacagc agaatatggg ccaaacagga tatctgtggt aagcagttcc tgccccggct 2820

cagggccaag aacagttgga acagcagaat atgggccaaa caggatatct gtggtaagca 2880

gttcctgccc cggctcaggg ccaagaacag atggtcccca gatgcggtcc cgccctcagc 2940

agtttctaga gaaccatcag atgtttccag ggtgccccaa ggacctgaaa tgaccctgtg 3000

ccttatttga actaaccaat cagttcgctt ctcgcttctg ttcgcgcgct tctgctcccc 3060

gagctcaata aaagagccca caacccctca ctcggcgcga ttcacctgac gcgtctacgc 3120

caccatgggg ttgcaggctt gcctgctggg acttttcgct ctgatcctga gcggaaagtg 3180

ctcctactca cctgagccag accagaggag aactctgccc cccggatggg tgtccctggg 3240

aagggccgac cctgaggagg aactctcgct caccttcgca ctgcggcagc agaacgtgga 3300

aagactgtcc gaactggtgc aggcagtgtc cgacccctcg agcccgcagt acggaaagta 3360

cctgaccctc gaaaacgtgg cagacttggt ccggccctcc cctctcaccc tgcacaccgt 3420

gcaaaaatgg ctgctggccg ccggagctca gaagtgccat tccgtgatta cacaggactt 3480

ccttacctgt tggcttagca tccgccaagc ggagctgctg ctgcctggtg ccgagttcca 3540

ccactacgtg ggcgggccaa ctgaaaccca cgtcgtgcgc agcccgcacc cgtatcagct 3600

gccccaggcg ctggctcctc atgtggactt cgtgggaggt ctgcaccggt tcccaccgac 3660

ttcaagcctc cggcagcgcc ccgaacctca agtcaccgga actgtggggc tccacctcgg 3720

cgtcacccct tccgtgatcc ggaagcggta caatctgacc tcgcaagacg tgggctcggg 3780

aacctcaaac aacagccagg cctgcgccca atttctggaa cagtacttcc acgatagcga 3840

tctggcccag ttcatgcgac ttttcggggg gaatttcgcc caccaagcca gcgtggcccg 3900

cgtggtcggg caacaggggc gcggaagggc gggcatcgag gcttccctgg atgtccagta 3960

cctcatgtcc gccggggcca acatctccac ttgggtgtac tcctcacctg gccgccacga 4020

ggggcaggaa ccgtttctgc aatggctgat gctgctgagc aacgaatccg cactcccgca 4080

cgtgcatact gtctcgtacg gcgacgatga ggactcactg tcctccgcgt acatccagag 4140

agtgaacact gagctcatga aggccgccgc gcggggcctg actttgttgt tcgcaagcgg 4200

cgattcggga gcgggatgtt ggtcggtgtc cggacgccat cagttccgcc cgaccttccc 4260

tgcctcaagc ccctacgtga caaccgtggg aggcaccagc tttcaggagc cgtttctgat 4320

taccaacgaa atcgtcgact acatttcggg cggcggtttc tccaacgtgt tcccacgccc 4380

ctcgtaccaa gaagaggccg tcaccaagtt cctgtcctcc tcccctcatc tcccgccatc 4440

ctcctacttt aacgcctccg gtcgggccta tcccgatgtg gccgccctgt cggacggcta 4500

ctgggtggtg tcgaataggg tgccgatccc ctgggtcagc ggaacttccg cgtccactcc 4560

tgtgtttggc ggcattcttt ccttgatcaa cgagcaccgg attctgtcgg gtagaccgcc 4620

gctgggattc ctcaacccgc ggctgtacca gcagcacggt gccggactgt tcgacgtgac 4680

gagagggtgc cacgagtcct gcctggacga ggaagtggaa ggacagggat tctgctctgg 4740

acccggatgg gatccggtca ccggctgggg caccccgaac ttccctgcgc tgctcaagac 4800

cctcctgaac ccctgatagt aatgacaggt acctttaaga ccaatgactt acaaggcagc 4860

tgtagatctt agccactttt taaaagaaaa gggggggactg gaagggctaa ttcactccca 4920

aagaagacaa gatctgcttt ttgcctgtac tgggtctctc tggttagacc agatctgagc 4980

ctgggagctc tctggctaac tagggaaccc actgcttaag cctcaataaa gcttgccttg 5040

agtgcttcaa tgtgtgtgtt ggttttttgt gtgtcgaaat tctagcgatt ctagcttggc 5100

gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa 5160

catacgagcc ggaagcataa agtgtaaagc ctggggtgcc taatgagtga gctaactcac 5220

attaattgcg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt gccagctgca 5280

ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgct cttccgcttc 5340

ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat cagctcactc 5400

aaaggcggta atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc 5460

aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag 5520

gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc 5580

gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt 5640

tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct 5700

ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg 5760

ctgtgtgcac gaaccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct 5820

tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat 5880

tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg 5940

ctacactaga agaacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa 6000

aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gttttttgt 6060

ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc 6120

tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt 6180

atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta 6240

aagtatatat gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat 6300

ctcagcgatc tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac 6360

tacgatacgg gagggcttac catctggccc cagtgctgca atgataccgc gagacccacg 6420

ctcaccggct ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag 6480

tggtcctgca actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt 6540

aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt 6600

gtcacgctcg tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt 6660

tacatgatcc cccatgttgt gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt 6720

cagaagtaag ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct 6780

tactgtcatg ccatccgtaa gatgcttttc tgtgactggt gagtactcaa ccaagtcatt 6840

ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac gggataatac 6900

cgcgccacat agcagaactt taaaagtgct catcattgga aaacgttctt cggggcgaaa 6960

actctcaagg atcttaccgc tgttgagatc cagttcgatg taacccactc gtgcacccaa 7020

ctgatcttca gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca 7080

aaatgccgca aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttcct 7140

ttttcaatat tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga 7200

atgtatttag aaaaataaac aaataggggt tccgcgcaca tttccccgaa aagtgccacc 7260

tgggactagc tttttgcaaa agcctaggcc tccaaaaaag cctcctcact acttctggaa 7320

tagctcagag gccgaggcgg cctcggcctc tgcataaata aaaaaaatta gtcagccatg 7380

gggcggagaa tgggcggaac tgggcggagt taggggcggg atgggcggag ttaggggcgg 7440

gactatggtt gctgactaat tgagatgagc ttgcatgccg acattgatta ttgactagtc 7500

cctaagaaac cattcttatc atgacattaa cctataaaaa taggcgtatc acgaggccct 7560

ttcgtc 7566

<210> 2
<211> 7700
<212> DNA
<213> Artificial Sequence

<220>
<223> Nucleic acid sequence of a lentiviral vector encoding TTP1 (pEF1alpha-CLN2)

<400> 2
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca 60

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120

ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180

accatcatat gccagcctat ggtgacattg attattgact agttattaat agtaatcaat 240

tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300

tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360

tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420

aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480

caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540

tacttggcag tacatctacg tattagtcat cgctattacc atggtgatgc ggttttggca 600

gtacatcaat gggcgtggat agcggtttga ctcacgggga tttccaagtc tccaccccat 660

tgacgtcaat gggagtttgt tttggcacca aaatcaacgg gactttccaa aatgtcgtaa 720

caactccgcc ccattgacgc aaatgggcgg taggcgtgta cggtgggagg tctatataag 780

cagagctcgt ttagtgaacc gggtctctct ggttagacca gatctgagcc tgggagctct 840

ctggctaact agggaaccca ctgcttaagc ctcaataaag cttgccttga gtgctcaaag 900

tagtgtgtgc ccgtctgttg tgtgactctg gtaactagag atccctcaga ccctttttagt 960

cagtgtggaa aatctctagc agtggcgccc gaacagggac ttgaaagcga aagtaaagcc 1020

agaggagatc tctcgacgca ggactcggct tgctgaagcg cgcacggcaa gaggcgaggg 1080

gcggcgactg gtgagtacgc caaaaatttt gactagcgga ggctagaagg agagagtagg 1140

gtgcgagagc gtcggtatta agcgggggag aattagataa atgggaaaaa attcggttaa 1200

ggccaggggg aaagaaacaa tataaactaa aacatatagt tagggcaagc agggagctag 1260

aacgattcgc agttaatcct ggccttttag agacatcaga aggctgtaga caaatactgg 1320

gacagctaca accatccctt cagacaggat cagaagaact tagatcatta tataatacaa 1380

tagcagtcct ctattgtgtg catcaaagga tagatgtaaa agacaccaag gaagccttag 1440

ataagataga ggaagagcaa aacaaaagta agaaaaaggc acagcaagca gcagctgaca 1500

caggaaacaa cagccaggtc agccaaaatt accctatagt gcagaacctc caggggcaaa 1560

tggtacatca ggccatatca cctagaactt taaattaaga cagcagtaca aatggcagta 1620

ttcatccaca attttaaaag aaaaggggg attgggggt acagtgcagg ggaaagaata 1680

gtagacataa tagcaacaga catacaaact aaagaattac aaaaacaaat tacaaaaatt 1740

caaaattttc gggtttatta cagggacagc agagatccag tttggaaagg accagcaaag 1800

ctcctctgga aaggtgaagg ggcagtagta atacaagata atagtgacat aaaagtagtg 1860

ccaagaagaa aagcaaagat catcagggat tatggaaaac agatggcagg tgatgattgt 1920

gtggcaagta gacaggatga ggattaacac atggaaaaga ttagtaaaac accatagctc 1980

tagagcgatc ccgatcttca gacctggagg aggagatatg agggacaatt ggagaagtga 2040

attatataaa tataaagtag taaaaattga accattagga gtagcaccca ccaaggcaaa 2100

gagaagagtg gtgcagagag aaaaaagagc agtgggaata ggagctttgt tccttgggtt 2160

cttgggagca gcaggaagca ctatgggcgc agcgtcaatg acgctgacgg tacaggccag 2220

acaattattg tctggtatag tgcagcagca gaacaatttg ctgagggcta ttgaggcgca 2280

acagcatctg ttgcaactca cagtctgggg catcaagcag ctccaggcaa gaatcctggc 2340

tgtggaaaga tacctaaagg atcaacagct cctggggatt tggggttgct ctggaaaact 2400

catttgcacc actgctgtgc cttggaatgc tagttggagt aataaatctc tggaacagat 2460

ttggaatcac acgacctgga tggagtggga cagagaaatt aacaattaca caagcttggt 2520

aggtttaaga atagtttttg ctgtactttc tatagtgaat agagttaggc agggatattc 2580

accattatcg tttcagaccc acctcccaac cccgagggga cccgacaggc ccgaaggaat 2640

agaagaagaa ggtggagaga gagacagaga cagatccatt cgattagtga acggatccaa 2700

ggatctgcga tcgctccggt gcccgtcagt gggcagagcg cacatcgccc acagtccccg 2760

agaagttggg gggaggggtc ggcaattgaa cgggtgccta gagaaggtgg cgcggggtaa 2820

actgggaaag tgatgtcgtg tactggctcc gccttttcc cgagggtggg ggagaaccgt 2880

atataagtgc agtagtcgcc gtgaacgttc tttttcgcaa cgggtttgcc gccagaacac 2940

agctgaagct tcgaggggct cgcatctctc cttcacgcgc ccgccgccct acctgaggcc 3000

gccatccacg ccggttgagt cgcgttctgc cgcctcccgc ctgtggtgcc tcctgaactg 3060

cgtccgccgt ctaggtaagt ttaaagctca ggtcgagacc gggcctttgt ccggcgctcc 3120

cttggagcct acctagactc agccggctct ccacgctttg cctgaccctg cttgctcaac 3180

tctacgtctt tgtttcgttt tctgttctgc gccgttacag atccaagctg tgaccggcgc 3240

ctacgcgtct acgccaccat ggggttgcag gcttgcctgc tgggactttt cgctctgatc 3300

ctgagcggaa agtgctccta ctcacctgag ccagaccaga ggagaactct gccccccgga 3360

tgggtgtccc tgggaagggc cgaccctgag gaggaactct cgctcacctt cgcactgcgg 3420

cagcagaacg tggaaagact gtccgaactg gtgcaggcag tgtccgaccc ctcgagcccg 3480

cagtacggaa agtacctgac cctcgaaaac gtggcagact tggtccggcc ctcccctctc 3540

accctgcaca ccgtgcaaaa atggctgctg gccgccggag ctcagaagtg ccattccgtg 3600

attacacagg acttccttac ctgttggctt agcatccgcc aagcggagct gctgctgcct 3660

ggtgccgagt tccaccacta cgtgggcggg ccaactgaaa cccacgtcgt gcgcagcccg 3720

cacccgtatc agctgcccca ggcgctggct cctcatgtgg acttcgtggg aggtctgcac 3780

cggttcccac cgacttcaag cctccggcag cgccccgaac ctcaagtcac cggaactgtg 3840

gggctccacc tcggcgtcac cccttccgtg atccggaagc ggtacaatct gacctcgcaa 3900

gacgtgggct cgggaacctc aaacaacagc caggcctgcg cccaatttct ggaacagtac 3960

ttccacgata gcgatctggc ccagttcatg cgacttttcg gggggaattt cgcccaccaa 4020

gccagcgtgg cccgcgtggt cgggcaacag gggcgcggaa gggcgggcat cgaggcttcc 4080

ctggatgtcc agtacctcat gtccgccggg gccaacatct ccacttgggt gtactcctca 4140

cctggccgcc acgaggggca ggaaccgttt ctgcaatggc tgatgctgct gagcaacgaa 4200

tccgcactcc cgcacgtgca tactgtctcg tacggcgacg atgaggactc actgtcctcc 4260

gcgtacatcc agagagtgaa cactgagctc atgaaggccg ccgcgcgggg cctgactttg 4320

ttgttcgcaa gcggcgattc gggagcggga tgttggtcgg tgtccggacg ccatcagttc 4380

cgcccgacct tccctgcctc aagcccctac gtgacaaccg tgggaggcac cagctttcag 4440

gagccgtttc tgattaccaa cgaaatcgtc gactacattt cgggcggcgg tttctccaac 4500

gtgttcccac gcccctcgta ccaagaagag gccgtcacca agttcctgtc ctcctcccct 4560

catctcccgc catcctccta ctttaacgcc tccggtcggg cctatcccga tgtggccgcc 4620

ctgtcggacg gctactgggt ggtgtcgaat agggtgccga tcccctgggt cagcggaact 4680

tccgcgtcca ctcctgtgtt tggcggcatt ctttccttga tcaacgagca ccggattctg 4740

tcgggtagac cgccgctggg attcctcaac ccgcggctgt accagcagca cggtgccgga 4800

ctgttcgacg tgacgagagg gtgccacgag tcctgcctgg acgaggaagt ggaaggacag 4860

ggattctgct ctggacccgg atgggatccg gtcaccggct ggggcacccc gaacttccct 4920

gcgctgctca agaccctcct gaacccctga tagtaatgac aggtaccttt aagaccaatg 4980

acttacaagg cagctgtaga tcttagccac tttttaaaag aaaaggggg actggaaggg 5040

ctaattcact cccaaagaag acaagatctg ctttttgcct gtactgggtc tctctggtta 5100

gaccagatct gagcctggga gctctctggc taactaggga acccactgct taagcctcaa 5160

taaagcttgc cttgagtgct tcaatgtgtg tgttggtttt ttgtgtgtcg aaattctagc 5220

gattctagct tggcgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc 5280

acaattccac acaacatacg agccggaagc ataaagtgta aagcctgggg tgcctaatga 5340

gtgagctaac tcacattaat tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg 5400

tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg 5460

cgctcttccg cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg 5520

gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga 5580

aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg 5640

gcgtttttcc ataggctccg ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag 5700

aggtggcgaa acccgacagg actataaaga taccaggcgt ttcccccctgg aagctccctc 5760

gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg 5820

ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt 5880

cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc 5940

ggtaactatc gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc 6000

actggtaaca ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg 6060

tggcctaact acggctacac tagaagaaca gtatttggta tctgcgctct gctgaagcca 6120

gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc 6180

ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat 6240

cctttgatct tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt 6300

ttggtcatga gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt 6360

tttaaatcaa tctaaagtat atatgagtaa acttggtctg acagttacca atgcttaatc 6420

agtgaggcac ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc 6480

gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata 6540

ccgcgagacc cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg 6600

gccgagcgca gaagtggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc 6660

cgggaagcta gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct 6720

acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa 6780

cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcggttag ctccttcggt 6840

cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca 6900

ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtac 6960

tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca 7020

atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt 7080

tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc 7140

actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca 7200

aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata 7260

ctcatactct tccttttca atattattga agcatttatc agggttattg tctcatgagc 7320

ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc 7380

cgaaaagtgc cacctgggac tagcttttg caaaagccta ggcctccaaa aaagcctcct 7440

cactacttct ggaatagctc agaggccgag gcggcctcgg cctctgcata aataaaaaaa 7500

attagtcagc catggggcgg agaatgggcg gaactgggcg gagttagggg cgggatgggc 7560

ggagttaggg gcgggactat ggttgctgac taattgagat gagcttgcat gccgacattg 7620

attattgact agtccctaag aaaccattct tatcatgaca ttaacctata aaaataggcg 7680

tatcacgagg ccctttcgtc 7700

<210> 3
<211> 1719
<212> DNA
<213> Homo sapiens

<400> 3
atgacagcag atccgcggaa gggcagaatg ggactccaag cctgcctcct agggctcttt 60

gccctcatcc tctctggcaa atgcagttac agcccggagc ccgaccagcg gaggacgctg 120

cccccaggct gggtgtccct gggccgtgcg gaccctgagg aagagctgag tctcaccttt 180

gccctgagac agcagaatgt ggaaagactc tcggagctgg tgcaggctgt gtcggatccc 240

agctctcctc aatacggaaa atacctgacc ctagagaatg tggctgatct ggtgaggcca 300

tccccactga ccctccacac ggtgcaaaaa tggctcttgg cagccggagc ccagaagtgc 360

cattctgtga tcacacagga ctttctgact tgctggctga gcatccgaca agcagagctg 420

ctgctccctg gggctgagtt tcatcactat gtgggaggac ctacggaaac ccatgttgta 480

aggtccccac atccctacca gcttccacag gccttggccc cccatgtgga ctttgtgggg 540

ggactgcacc gttttccccc aacatcatcc ctgaggcaac gtcctgagcc gcaggtgaca 600

gggactgtag gcctgcatct gggggtaacc ccctctgtga tccgtaagcg atacaacttg 660

acctcacaag acgtgggctc tggcaccagc aataacagcc aagcctgtgc ccagttcctg 720

gagcagtatt tccatgactc agacctggct cagttcatgc gcctcttcgg tggcaacttt 780

gcacatcagg catcagtagc ccgtgtggtt ggacaacagg gccggggccg ggccgggatt 840

gaggccagtc tagatgtgca gtacctgatg agtgctggtg ccaacatctc cacctgggtc 900

tacagtagcc ctggccggca tgagggacag gagcccttcc tgcagtggct catgctgctc 960

agtaatgagt cagccctgcc acatgtgcat actgtgagct atggagatga tgaggactcc 1020

ctcagcagcg cctacatcca gcgggtcaac actgagctca tgaaggctgc cgctcggggt 1080

ctcaccctgc tcttcgcctc aggtgacagt ggggccgggt gttggtctgt ctctggaaga 1140

caccagttcc gccctacctt ccctgcctcc agcccctatg tcaccacagt gggaggcaca 1200

tccttccagg aacctttcct catcacaaat gaaattgttg actatatcag tggtggtggc 1260

ttcagcaatg tgttcccacg gccttcatac caggaggaag ctgtaacgaa gttcctgagc 1320

tctagccccc acctgccacc atccagttac ttcaatgcca gtggccgtgc ctacccagat 1380

gtggctgcac tttctgatgg ctactgggtg gtcagcaaca gagtgcccat tccatgggtg 1440

tccggaacct cggcctctac tccagtgttt gggggggatcc tatccttgat caatgagcac 1500

aggatcctta gtggccgccc ccctcttggc tttctcaacc caaggctcta ccagcagcat 1560

ggggcaggac tctttgatgt aacccgtggc tgccatgagt cctgtctgga tgaagaggta 1620

gagggccagg gtttctgctc tggtcctggc tgggatcctg taacaggctg gggaacaccc 1680

aacttcccag ctttgctgaa gactctactc aacccctga 1719

<210> 4
<211> 1692
<212> DNA
<213> Artificial Sequence

<220>
<223> Codon-optimized polynucleotide sequence encoding a human tripeptidyl peptidase 1 (TPP1)

<400> 4

atggggttgc aggcttgcct gctgggactt ttcgctctga tcctgagcgg aaagtgctcc 60

tactcacctg agccagacca gaggagaact ctgccccccg gatgggtgtc cctgggaagg 120

gccgaccctg aggaggaact ctcgctcacc ttcgcactgc ggcagcagaa cgtggaaaga 180

ctgtccgaac tggtgcaggc agtgtccgac ccctcgagcc cgcagtacgg aaagtacctg 240

accctcgaaa acgtggcaga cttggtccgg ccctcccctc tcaccctgca caccgtgcaa 300

aaatggctgc tggccgccgg agctcagaag tgccattccg tgattacaca ggacttcctt 360

acctgttggc ttagcatccg ccaagcggag ctgctgctgc ctggtgccga gttccaccac 420

tacgtgggcg ggccaactga aacccacgtc gtgcgcagcc cgcacccgta tcagctgccc 480

caggcgctgg ctcctcatgt ggacttcgtg ggaggtctgc accggttccc accgacttca 540

agcctccggc agcgccccga acctcaagtc accggaactg tggggctcca cctcggcgtc 600

acccccttccg tgatccggaa gcggtacaat ctgacctcgc aagacgtggg ctcgggaacc 660

tcaaacaaca gccaggcctg cgcccaattt ctggaacagt acttccacga tagcgatctg 720

gcccagttca tgcgactttt cgggggggaat ttcgcccacc aagccagcgt ggcccgcgtg 780

gtcgggcaac aggggcgcgg aagggcgggc atcgaggctt ccctggatgt ccagtacctc 840

atgtccgccg gggccaacat ctccacttgg gtgtactcct cacctggccg ccacgagggg 900

caggaaccgt ttctgcaatg gctgatgctg ctgagcaacg aatccgcact cccgcacgtg 960

catactgtct cgtacggcga cgatgaggac tcactgtcct ccgcgtacat ccagagagtg 1020

aacactgagc tcatgaaggc cgccgcgcgg ggcctgactt tgttgttcgc aagcggcgat 1080

tcgggagcgg gatgttggtc ggtgtccgga cgccatcagt tccgcccgac cttccctgcc 1140

tcaagcccct acgtgacaac cgtgggaggc accagctttc aggagccgtt tctgattacc 1200

aacgaaatcg tcgactacat ttcgggcggc ggtttctcca acgtgttccc acgcccctcg 1260

taccaagaag aggccgtcac caagttcctg tcctcctccc ctcatctccc gccatcctcc 1320

tactttaacg cctccggtcg ggcctatccc gatgtggccg ccctgtcgga cggctactgg 1380

gtggtgtcga atagggtgcc gatcccctgg gtcagcggaa cttccgcgtc cactcctgtg 1440

tttggcggca ttctttcctt gatcaacgag caccggattc tgtcgggtag accgccgctg 1500

ggattcctca acccgcggct gtaccagcag cacggtgccg gactgttcga cgtgacgaga 1560

gggtgccacg agtcctgcct ggacgaggaa gtggaaggac agggattctg ctctggaccc 1620

ggatgggatc cggtcaccgg ctggggcacc ccgaacttcc ctgcgctgct caagaccctc 1680

ctgaacccct ga 1692

<210> 5
<211> 563
<212> PRT
<213> Homo sapiens

<400> 5

```
Met Gly Leu Gln Ala Cys Leu Leu Gly Leu Phe Ala Leu Ile Leu Ser
1               5                  10                  15

Gly Lys Cys Ser Tyr Ser Pro Glu Pro Asp Gln Arg Arg Thr Leu Pro
            20                  25                  30

Pro Gly Trp Val Ser Leu Gly Arg Ala Asp Pro Glu Glu Glu Leu Ser
        35                  40                  45

Leu Thr Phe Ala Leu Arg Gln Gln Asn Val Glu Arg Leu Ser Glu Leu
    50                  55                  60

Val Gln Ala Val Ser Asp Pro Ser Ser Pro Gln Tyr Gly Lys Tyr Leu
65                  70                  75                  80

Thr Leu Glu Asn Val Ala Asp Leu Val Arg Pro Ser Pro Leu Thr Leu
                85                  90                  95

His Thr Val Gln Lys Trp Leu Leu Ala Ala Gly Ala Gln Lys Cys His
            100                 105                 110

Ser Val Ile Thr Gln Asp Phe Leu Thr Cys Trp Leu Ser Ile Arg Gln
        115                 120                 125

Ala Glu Leu Leu Leu Pro Gly Ala Glu Phe His His Tyr Val Gly Gly
    130                 135                 140

Pro Thr Glu Thr His Val Val Arg Ser Pro His Pro Tyr Gln Leu Pro
145                 150                 155                 160

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

Pro Pro Thr Ser Ser Leu Arg Gln Arg Pro Glu Pro Gln Val Thr Gly
            180                 185                 190

Thr Val Gly Leu His Leu Gly Val Thr Pro Ser Val Ile Arg Lys Arg
        195                 200                 205

Tyr Asn Leu Thr Ser Gln Asp Val Gly Ser Gly Thr Ser Asn Asn Ser
    210                 215                 220

Gln Ala Cys Ala Gln Phe Leu Glu Gln Tyr Phe His Asp Ser Asp Leu
225                 230                 235                 240
```

```
Ala Gln Phe Met Arg Leu Phe Gly Gly Asn Phe Ala His Gln Ala Ser
                245                 250                 255

Val Ala Arg Val Val Gly Gln Gln Gly Arg Gly Arg Ala Gly Ile Glu
            260                 265                 270

Ala Ser Leu Asp Val Gln Tyr Leu Met Ser Ala Gly Ala Asn Ile Ser
        275                 280                 285

Thr Trp Val Tyr Ser Ser Pro Gly Arg His Glu Gly Gln Glu Pro Phe
    290                 295                 300

Leu Gln Trp Leu Met Leu Leu Ser Asn Glu Ser Ala Leu Pro His Val
305                 310                 315                 320

His Thr Val Ser Tyr Gly Asp Asp Glu Asp Ser Leu Ser Ser Ala Tyr
                325                 330                 335

Ile Gln Arg Val Asn Thr Glu Leu Met Lys Ala Ala Ala Arg Gly Leu
            340                 345                 350

Thr Leu Leu Phe Ala Ser Gly Asp Ser Gly Ala Gly Cys Trp Ser Val
        355                 360                 365

Ser Gly Arg His Gln Phe Arg Pro Thr Phe Pro Ala Ser Ser Pro Tyr
    370                 375                 380

Val Thr Thr Val Gly Gly Thr Ser Phe Gln Glu Pro Phe Leu Ile Thr
385                 390                 395                 400

Asn Glu Ile Val Asp Tyr Ile Ser Gly Gly Gly Phe Ser Asn Val Phe
                405                 410                 415

Pro Arg Pro Ser Tyr Gln Glu Glu Ala Val Thr Lys Phe Leu Ser Ser
            420                 425                 430

Ser Pro His Leu Pro Pro Ser Ser Tyr Phe Asn Ala Ser Gly Arg Ala
        435                 440                 445

Tyr Pro Asp Val Ala Ala Leu Ser Asp Gly Tyr Trp Val Val Ser Asn
    450                 455                 460

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

Phe Gly Gly Ile Leu Ser Leu Ile Asn Glu His Arg Ile Leu Ser Gly
                485                 490                 495
```

```
Arg Pro Pro Leu Gly Phe Leu Asn Pro Arg Leu Tyr Gln Gln His Gly
            500                 505                 510

Ala Gly Leu Phe Asp Val Thr Arg Gly Cys His Glu Ser Cys Leu Asp
        515                 520                 525

Glu Glu Val Glu Gly Gln Gly Phe Cys Ser Gly Pro Gly Trp Asp Pro
    530                 535                 540

Val Thr Gly Trp Gly Thr Pro Asn Phe Pro Ala Leu Leu Lys Thr Leu
545                 550                 555                 560

Leu Asn Pro
```

<210> 6
<211> 3
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 6

```
Gly Gly Gly
1
```

<210> 7
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 7

```
Asp Gly Gly Gly Ser
1               5
```

<210> 8
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 8

```
Thr Gly Glu Lys Pro
1               5
```

<210> 9
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 9

Gly Gly Arg Arg
1

<210> 10
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 10

Gly Gly Gly Gly Ser
1 5

<210> 11
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 11

Glu Gly Lys Ser Ser Gly Ser Gly Ser Glu Ser Lys Val Asp
1 5 10

<210> 12
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 12

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

Leu Asp

<210> 13
<211> 8
<212> PRT

<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 13

Gly Gly Arg Arg Gly Gly Gly Ser
1 5

<210> 14
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 14

Leu Arg Gln Arg Asp Gly Glu Arg Pro
1 5

<210> 15
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 15

Leu Arg Gln Lys Asp Gly Gly Gly Ser Glu Arg Pro
1 5 10

<210> 16
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Exemplary linker sequence

<400> 16

Leu Arg Gln Lys Asp Gly Gly Gly Ser Gly Gly Gly Ser Glu Arg Pro
1 5 10 15

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

<220>
<223> TEV protease cleavage sequence motif

<220>

<221> MISC_FEATURE
<222> (2)..(3)
<223> Xaa is any amino acid

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa is any amino acid

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> Xaa is Gly or Ser

<400> 17

Glu Xaa Xaa Tyr Xaa Gln Xaa
1 5

<210> 18
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Cleavage sequence by TEV protease

<400> 18

Glu Asn Leu Tyr Phe Gln Gly
1 5

<210> 19
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Cleavage sequence by TEV protease

<400> 19

Glu Asn Leu Tyr Phe Gln Ser
1 5

<210> 20
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Consensus Kozak sequence

<400> 20
gccrccatgg 10