



US 20250262325A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2025/0262325 A1  
Barbon et al. (43) Pub. Date: Aug. 21, 2025

## (54) LENTIVIRAL VECTOR

- (71) Applicant: **Genespire S.r.l.**, Milan (IT)
- (72) Inventors: **Elena Barbon**, Cernusco sul Naviglio (Milano) (IT); **Alessio Cantore**, Milano (IT)
- (21) Appl. No.: **18/856,469**
- (22) PCT Filed: **Apr. 14, 2023**
- (86) PCT No.: **PCT/EP2023/059850**  
§ 371 (c)(1),  
(2) Date: **Oct. 11, 2024**

## (30) Foreign Application Priority Data

Apr. 14, 2022 (GB) ..... 2205615.4

## Publication Classification

- (51) Int. Cl.  
**A61K 48/00** (2006.01)  
**C12N 9/12** (2006.01)  
**C12N 15/86** (2006.01)
- (52) U.S. Cl.  
CPC ..... **A61K 48/0058** (2013.01); **C12N 9/12** (2013.01); **C12N 15/86** (2013.01); **C12Y 207/11004** (2013.01); **C12N 2740/15043** (2013.01)

## (57) ABSTRACT

The present invention relates to lentiviral vectors encoding a branched-chain alpha-ketoacid dehydrogenase complex (BCKDC) subunit. The present invention also relates to cells and pharmaceutical compositions comprising said lentiviral vectors and to uses of said lentiviral vectors in treating maple syrup urine disease (MSUD).

Specification includes a Sequence Listing.

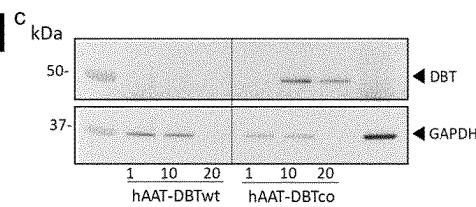
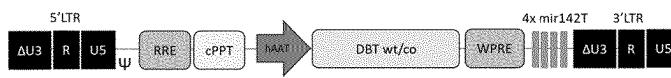
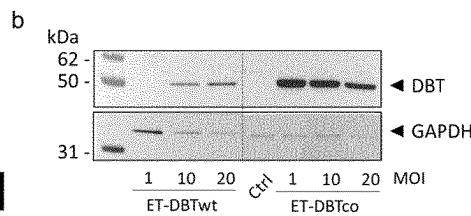
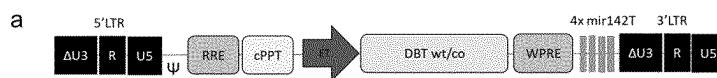


FIGURE 1

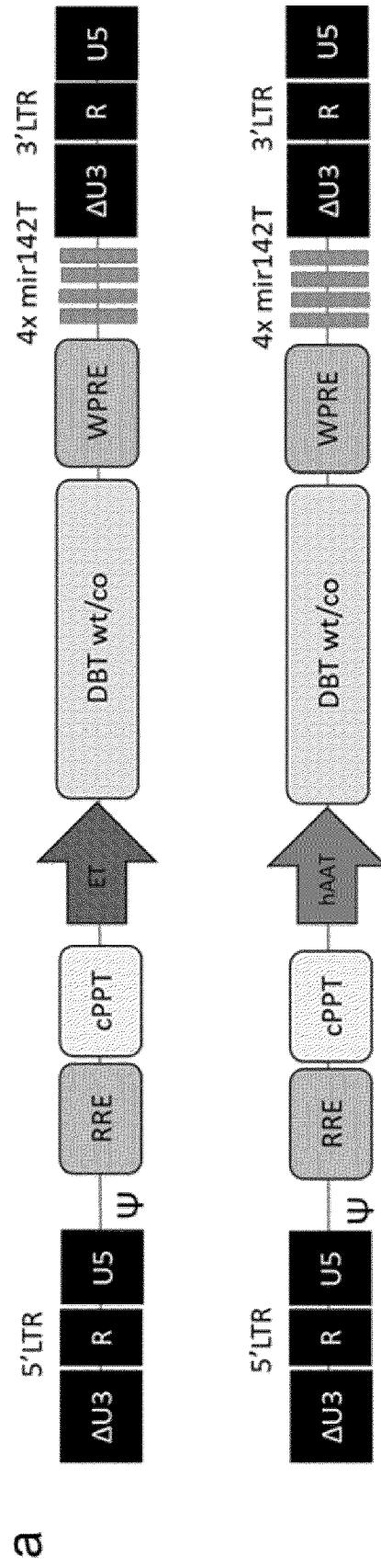


FIGURE 1 (CONTINUED)

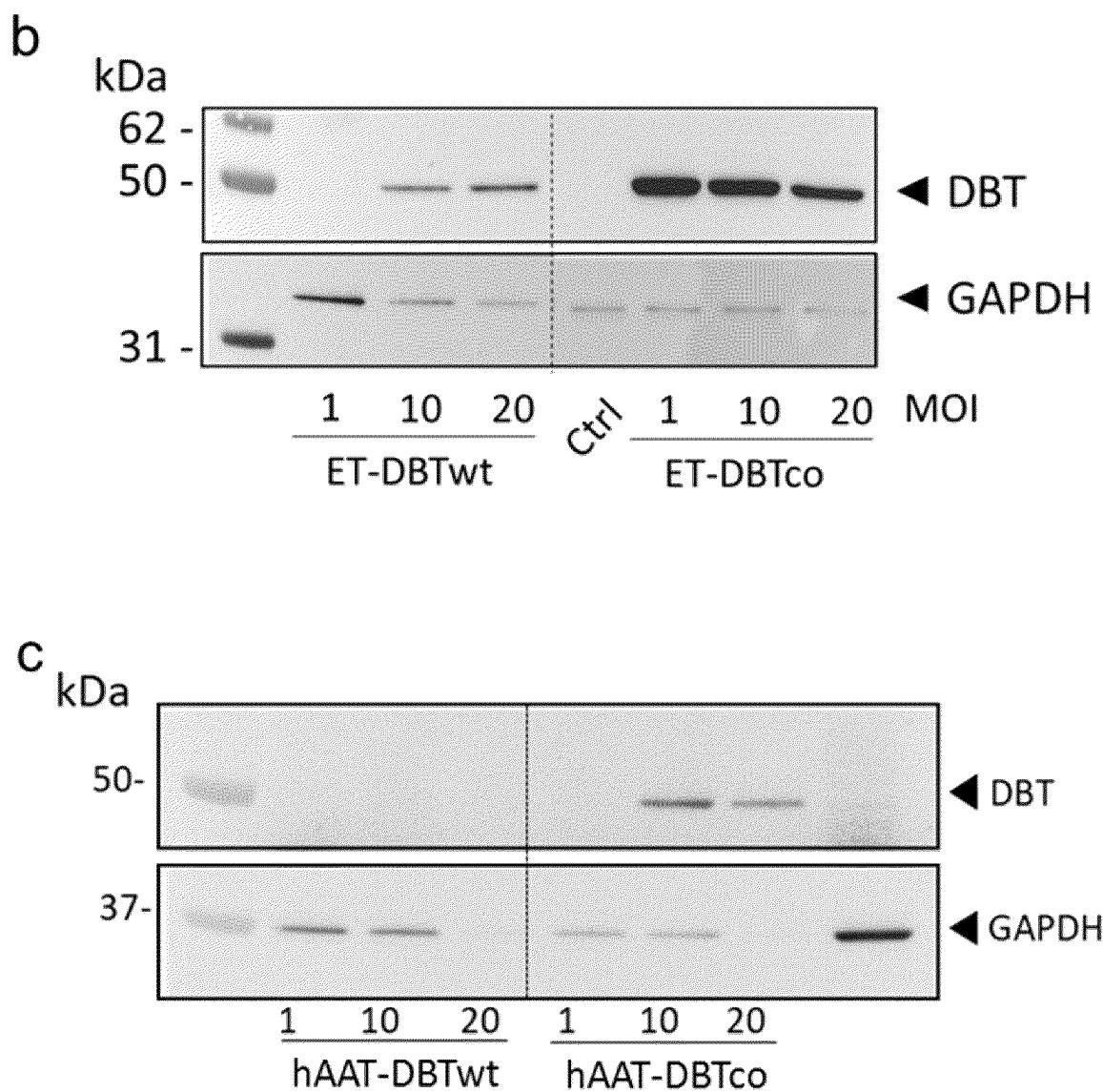
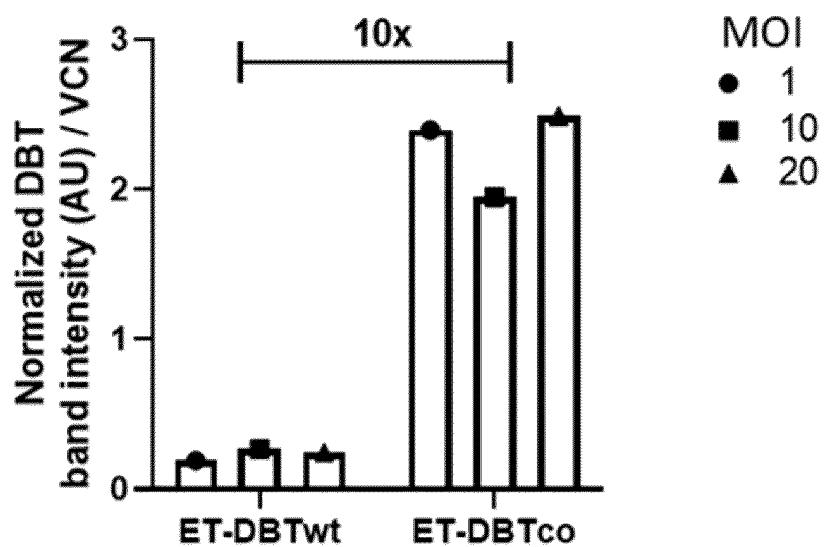


FIGURE 1 (CONTINUED)

d



e

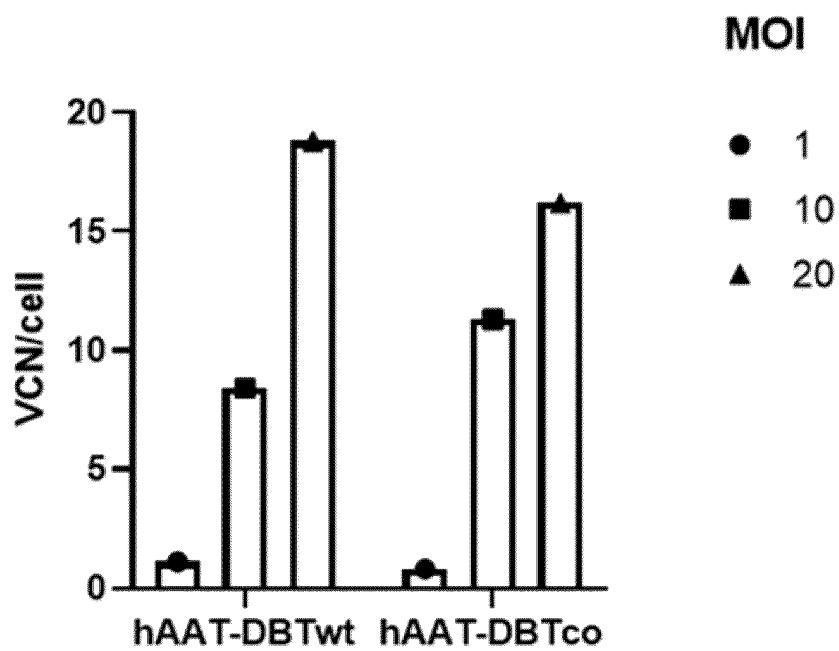


FIGURE 1 (CONTINUED)

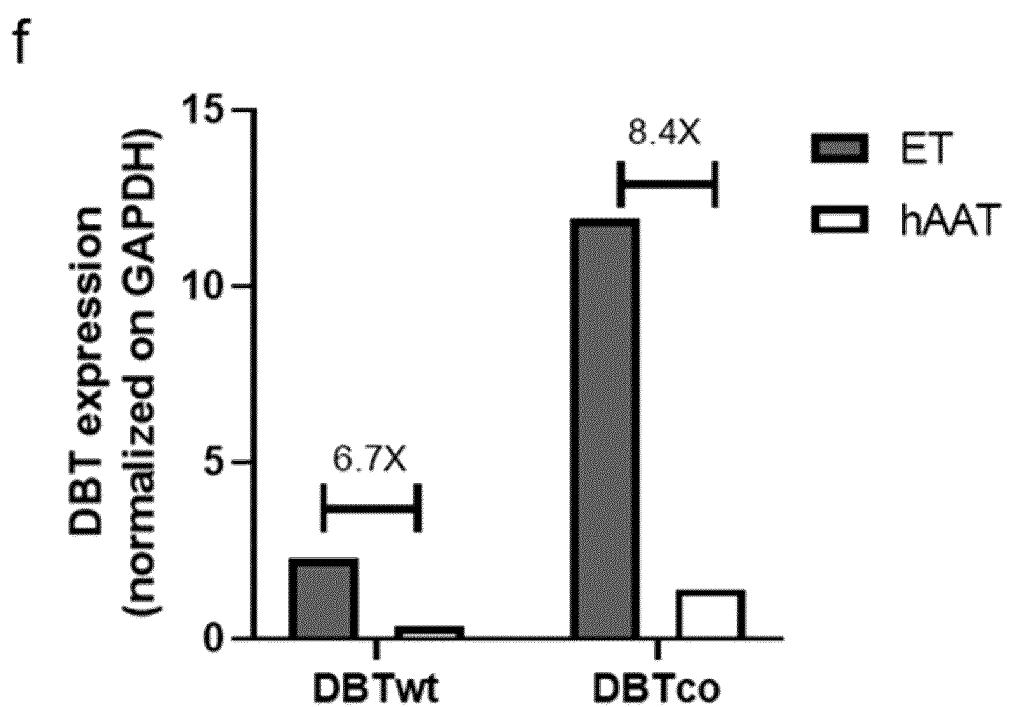


FIGURE 2

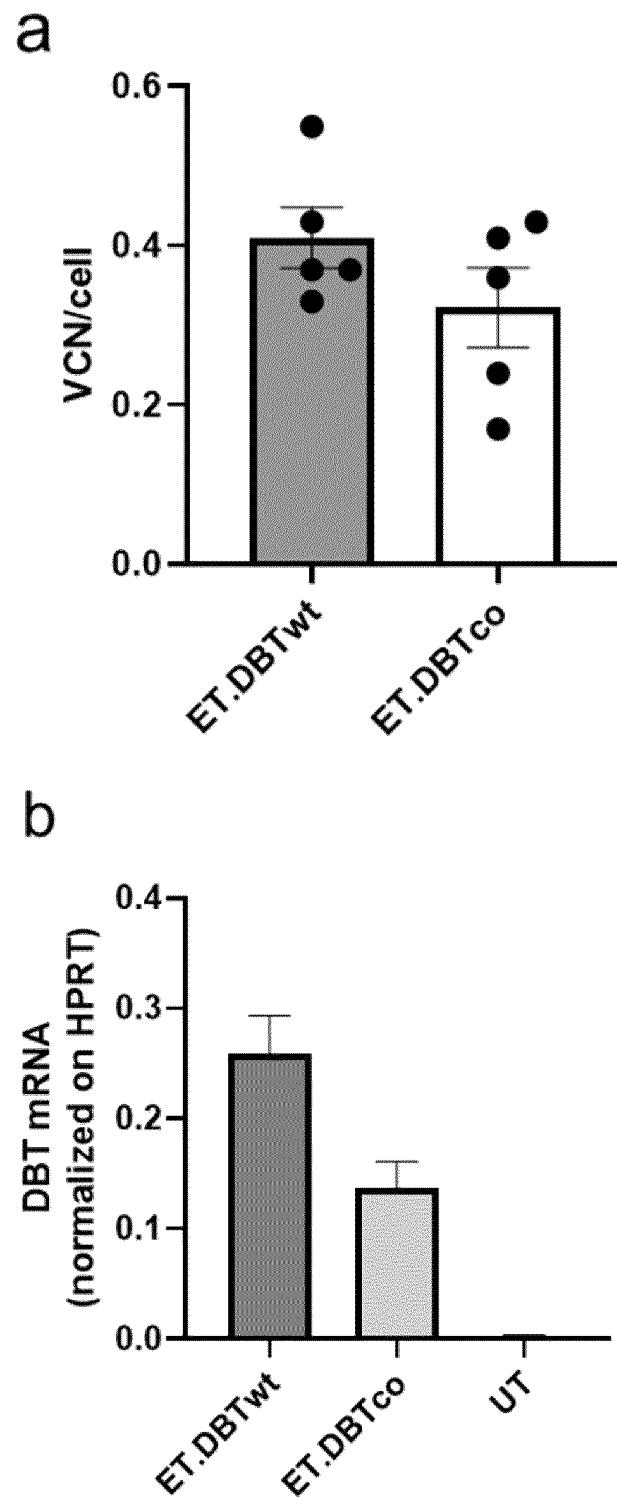


FIGURE 2 (CONTINUED)

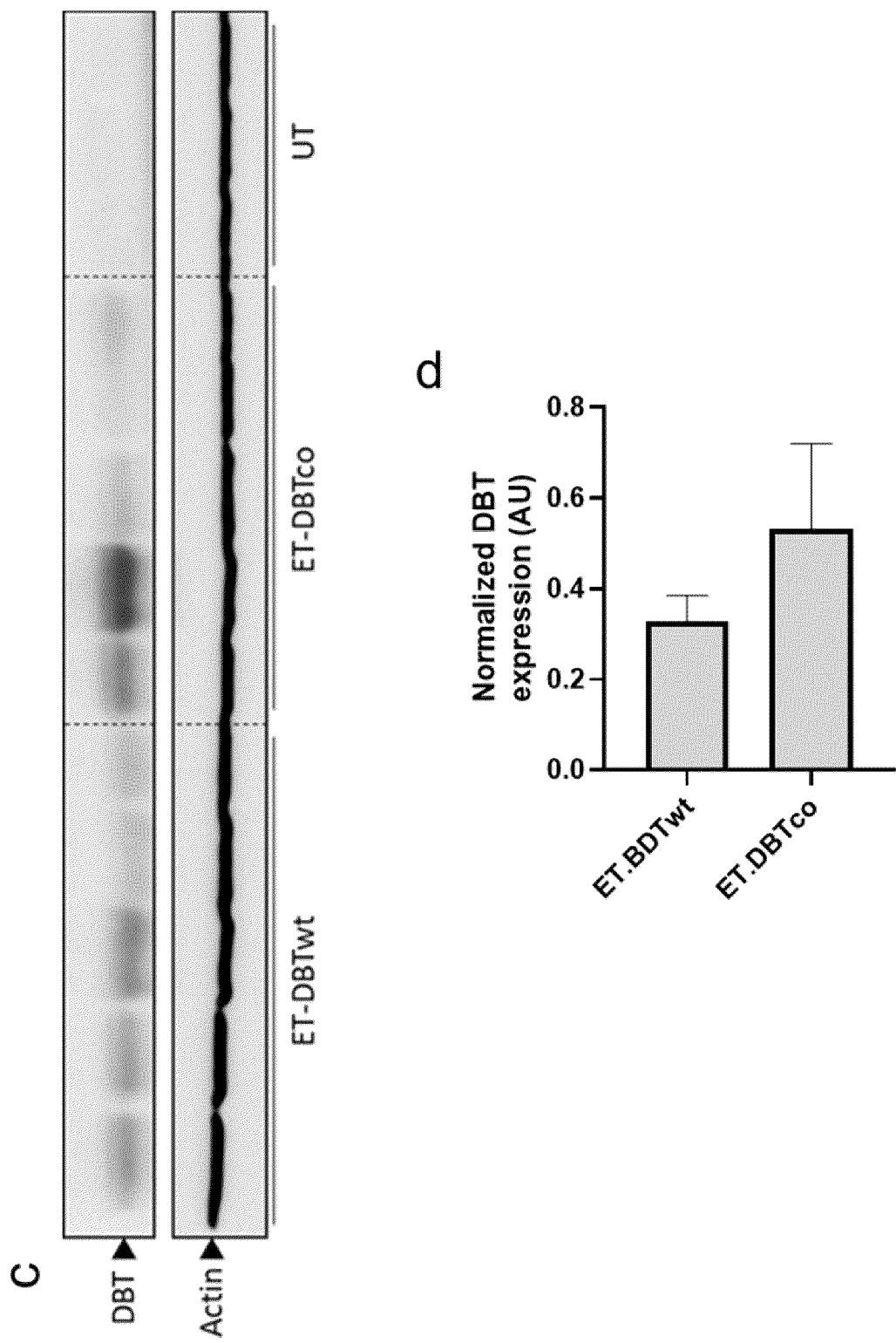
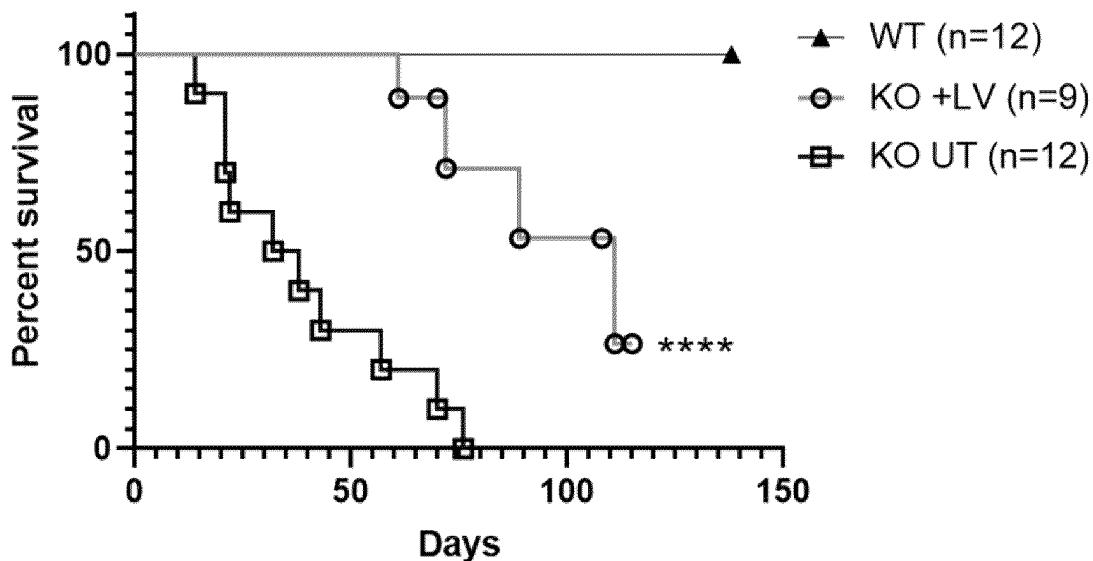


FIGURE 3

a



b

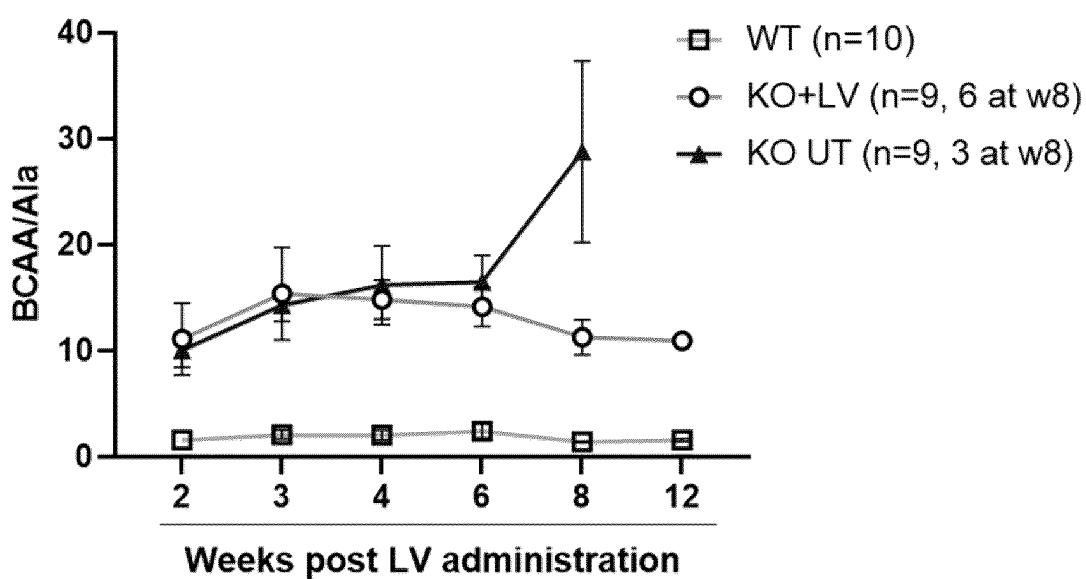
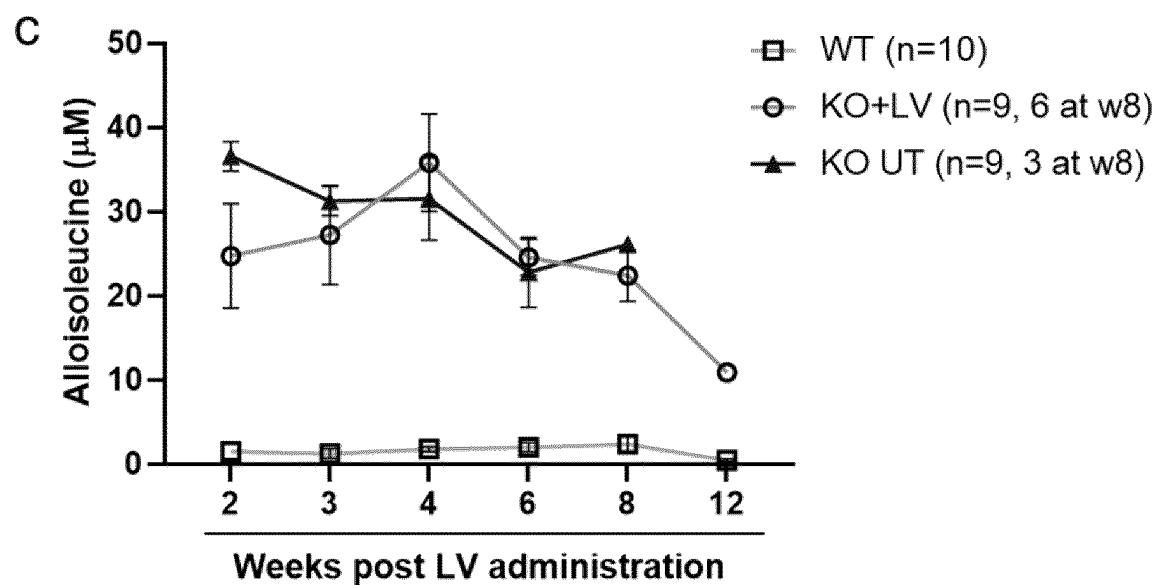


FIGURE 3 (CONTINUED)



**LENTIVIRAL VECTOR****FIELD OF THE INVENTION**

**[0001]** The present invention relates to lentiviral vectors encoding a branched-chain alpha-ketoacid dehydrogenase complex (BCKDC) subunit. The present invention also relates to cells and pharmaceutical compositions comprising said lentiviral vectors and to uses of said lentiviral vectors in treating maple syrup urine disease (MSUD).

**BACKGROUND TO THE INVENTION**

**[0002]** Maple syrup urine disease is a monogenic inborn error of metabolism caused by the deficiency of the mitochondrial enzymatic complex branched-chain alpha-ketoacid dehydrogenase complex (BCKDC), which catalyses the second step in the degradative pathway of the branched-chain amino acids (BCAAs), which include leucine, isoleucine, and valine (see e.g. Strauss, K. A. et al., 2006. In Gene Reviews).

**[0003]** The dysfunction of BCKDC leads to an accumulation in the blood and urine of BCAAs, in particular leucine, and of the respective alpha-keto acids. These metabolic intermediates, if present at high concentrations, cause metabolic crises and neurotoxicity (see e.g. Zinnanti, W. J., et al., 2009. Brain, 132(4), pp. 903-918).

**[0004]** There are five distinct clinical phenotypes of MSUD. Classic and E3-deficient MSUD typically present in the neonatal period, while the intermediate, intermittent, and thiamine-responsive forms may present at any time of life. The severity of MSUD can depend on the amount of residual BCKD enzyme activity (see e.g. Blackburn, P. R., et al., 2017. The application of clinical genetics, 10, pp. 57-66).

**[0005]** The classic form of the disease manifests in early childhood with difficulty in feeding, lethargy, vomiting, and earwax, noticed immediately after birth (and later the urine), which smells like maple syrup. If not treated, MSUD patients develop a progressive encephalopathy and central respiratory failure. In patients with intermittent MSUD, symptoms may appear later, with episodes of acute metabolic decompensation and the development of neurological problems during childhood (see e.g. Blackburn, P. R., et al., 2017. The application of clinical genetics, 10, pp. 57-66).

**[0006]** Currently, the treatment of patients with MSUD is based on a dietary regimen aimed at maintaining the BCAAs levels in a range considered normal with respect to age and individual tolerance, to ensure normal growth and to avoid the deficiency of essential nutrients and other amino acids. When dietary treatment does not allow the rapid normalization of clinical and biochemical parameters, exogenous purification with dialysis treatment (peritoneal dialysis, hemofiltration) is performed to remove toxic metabolites (see e.g. Blackburn, P. R., et al., 2017. The application of clinical genetics, 10, pp. 57-66). Liver transplantation represents an important therapeutic option for MSUD patients. However, liver transplantation is associated with predictable morbidities and does not reverse pre-existing neurological complications (see e.g. Strauss, K. A., et al., 2020. Molecular genetics and metabolism, 129(3), pp. 193-206).

**[0007]** Thus, there is a critical unmet need for safe and effective disease-modifying therapies for MSUD which can be implemented early in life.

**SUMMARY OF THE INVENTION**

**[0008]** The present inventors have developed a gene therapy for treating and/or preventing MSUD. The lentiviral vector mediated gene therapy described herein may allow for a stable gene transfer even in paediatric (e.g. neonatal) patients at the first disease stages by virtue of lentiviral vector genomic integration and may therefore alleviate symptoms and limit progressive damage to the hepatocytes.

**[0009]** The present inventors have surprisingly shown that the lentiviral vector mediated gene therapy described herein is safe and efficacious in a model of MSUD. The present inventors produced a lentiviral vector encoding a human codon optimized DBT-coding sequence under the control of a hepatocyte-specific cassette. Administration of the lentiviral vector resulted in a decrease of plasma branched-chain amino acids (BCAA) and alloisoleucine, and improved survival.

**[0010]** In one aspect, the present invention provides a lentiviral vector comprising a nucleotide sequence encoding a branched-chain alpha-ketoacid dehydrogenase complex (BCKDC) subunit. The lentiviral vector may be an immune-shielded lentiviral vector.

**[0011]** The BCKDC subunit may selected from BCKD E1 alpha subunit (BCKDE1A), or a fragment thereof; BCKD E1 beta subunit (BCKDE1B), or a fragment thereof; BCKD E2 subunit (DBT), or a fragment thereof; and BCKD E3 subunit (DLD), or a fragment thereof. In preferred embodiments, the BCKDC subunit is DBT, or a fragment thereof.

**[0012]** Suitably, the BCKDC subunit comprises or consists of an amino acid sequence which is at least 70% identical to one of SEQ ID NOS: 37, 38, 40, 41, 43, 46, 47 or 48, or a fragment thereof. In some embodiments, the BCKDC subunit comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 43, or a fragment thereof.

**[0013]** Suitably, the nucleotide sequence encoding a BCKDC subunit comprises or consists of a nucleotide sequence which is at least 70% identical to one of SEQ ID NOS: 39, 42, 44, 45 or 49, or a fragment thereof. In some embodiments, the nucleotide sequence encoding a BCKDC subunit comprises or consists of a nucleotide sequence which is at least 70% identical to one of SEQ ID NOS: 44 or 45, or a fragment thereof. Suitably, the nucleotide sequence encoding a BCKDC subunit is codon-optimised. In some embodiments, the nucleotide sequence encoding a BCKDC subunit comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 45, or a fragment thereof.

**[0014]** The lentiviral vector may be a CD47<sup>high</sup> lentiviral vector. Suitably, the lentiviral vector is obtained from a CD47<sup>high</sup> host cell, preferably wherein the host cell is genetically engineered to increase expression of CD47 on the cell surface. Suitably, the lentiviral vector has at least about 2-fold more CD47 on its surface than a lentiviral vector obtained from an unmodified host cell.

**[0015]** The lentiviral vector may be a MHC-I<sup>free</sup> lentiviral vector. Suitably, the lentiviral vector is obtained from a MHC-I<sup>free</sup> host cell, preferably wherein the host cell is genetically engineered to disrupt expression of MHC-I on the cell surface. Suitably, the lentiviral vector has less than about 50% of the number of surface-exposed MHC-I molecules that are displayed on a lentiviral vector obtained from an unmodified host cell. Suitably, MHC-I is not detectable on the surface of the lentiviral vector.

[0016] The lentiviral vector may be a CD47<sup>high</sup>/MHC-I<sup>free</sup> lentiviral vector. Suitably, the lentiviral vector is obtained from a CD47<sup>high</sup>/MHC-I<sup>free</sup> host cell. Suitably, the lentiviral vector has at least about 2-fold more CD47 on its surface than a lentiviral vector obtained from an unmodified host cell and MHC-I is not detectable on the surface of the lentiviral vector.

[0017] The nucleotide sequence encoding a BCKDC subunit may be operably linked to one or more miRNA target sequences. In some embodiments, the one or more miRNA target sequences suppress transgene expression in one or more cells other than hepatocytes. In some embodiments, the one or more miRNA target sequence suppress transgene expression in hematopoietic-lineage cells and/or antigen-presenting cells.

[0018] In some embodiments, the one or more miRNA target sequences are selected from miR-181, miR-142, miR-223, and miR-155 target sequences. In some embodiments, the nucleotide sequence encoding a BCKDC subunit is operably linked to one or more mir-142 target sequence, two or more mir-142 target sequences, three or more mir-142 target sequences, or four or more mir-142 target sequences. In some embodiments, the nucleotide sequence encoding a BCKDC subunit is operably linked to four mir-142 target sequences. In some embodiments, the one or more miRNA target sequences comprise or consist of a nucleotide sequence which is at least 90% identical to SEQ ID NO: 17. In some embodiments, the one or more miRNA target sequences comprise or consist of a nucleotide sequence which is at least 90% identical to SEQ ID NO: 18.

[0019] The nucleotide sequence encoding a BCKDC subunit may be operably linked to a liver-specific promoter. In some embodiments, the nucleotide sequence encoding a BCKDC subunit is operably linked to a hepatocyte-specific promoter. Suitably, the nucleotide sequence encoding a BCKDC subunit is operably linked to a transthyretin (TTR) promoter, an alpha-1-antitrypsin (AAT) promoter, a thyroxine-binding globulin (TBG) promoter, an APoE/hAAT promoter, a HCR-hAAT promoter, a LP1 promoter, or a HLP promoter. In some embodiments, the nucleotide sequence encoding a BCKDC subunit is operably linked to a transthyretin (TTR) promoter. In preferred embodiments, the nucleotide sequence encoding a BCKDC subunit is operably linked to an Enh1mTTR (ET) promoter. In some embodiments, the nucleotide sequence encoding a BCKDC subunit is operably linked to a promoter which comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 19.

[0020] The lentiviral vector may be pseudotyped. Suitably, the lentiviral vector is VSV.G-pseudotyped. The lentiviral vector may be a self-inactivating (SIN) lentiviral vector. Suitably, the lentiviral vector comprises self-inactivating (SIN) LTRs which comprise or consist of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 23, or a fragment thereof. The lentiviral vector may be an integrating lentiviral vector and/or a replication-defective lentiviral vector. The lentiviral vector may be HIV-derived.

[0021] In some embodiments, the lentiviral vector comprises a nucleotide sequence which is at least 70% identical to SEQ ID NO: 36.

[0022] In one aspect, the present invention provides a method of producing a lentiviral vector according to the present invention.

[0023] In one aspect, the present invention provides a kit or system for producing a lentiviral vector of the present invention.

[0024] In one aspect, the present invention provides a cell comprising a lentiviral vector according to the present invention. The cell may be an isolated cell.

[0025] In one aspect, the present invention provides a pharmaceutical composition comprising a lentiviral vector or a cell according to the present invention, in combination with a pharmaceutically acceptable carrier, diluent or excipient.

[0026] In one aspect, the present invention provides a lentiviral vector, a cell, or a pharmaceutical composition according to the present invention, for use as a medicament.

[0027] In one aspect, the present invention provides use of a lentiviral vector, a cell, or a pharmaceutical composition according to the present invention, for the manufacture of a medicament.

[0028] In one aspect, the present invention provides a method comprising administering a therapeutically effective amount of a lentiviral vector, a cell, or a pharmaceutical composition according to the present invention, to a subject in need thereof.

[0029] In one aspect, the present invention provides a lentiviral vector, a cell, or a pharmaceutical composition according to the present invention, for use in preventing or treating maple syrup urine disease (MSUD).

[0030] In one aspect, the present invention provides use of a lentiviral vector, a cell, or a pharmaceutical composition according to the present invention, for the manufacture of a medicament for preventing or treating maple syrup urine disease (MSUD).

[0031] In one aspect, the present invention provides a method of preventing or treating maple syrup urine disease (MSUD), comprising administering a therapeutically effective amount of a lentiviral vector, a cell, or a pharmaceutical composition according to the present invention, to a subject in need thereof.

[0032] Suitably, the BCKDC subunit is BCKDE1A, or a fragment thereof, and the MSUD is MSUD type 1A; the BCKDC subunit is BCKDE1B, or a fragment thereof, and the MSUD is MSUD type 1B; the BCKDC subunit is DBT, or a fragment thereof, and the MSUD is MSUD type 2; or the BCKDC subunit is DLD, or a fragment thereof, and the MSUD is MSUD type 3. In preferred embodiments, the BCKDC subunit is DBT, or a fragment thereof, and the MSUD is MSUD type 2.

[0033] The lentiviral vector, cell, or pharmaceutical composition may be administered to any subject in need thereof. In some embodiments, the subject is a human subject. In some embodiments, the subject is a juvenile. In some embodiments, the subject is a paediatric patient. In some embodiments, the subject is a neonatal patient or an infantile patient.

[0034] The lentiviral vector, cell, or pharmaceutical composition may be administered by any suitable route. In some embodiments, the lentiviral vector, cell, or pharmaceutical composition is administered systemically (e.g. by intravenous injection or intraperitoneal injection). In some embodiments, the lentiviral vector, cell, or pharmaceutical composition is administered locally (e.g. by direct injection, intraarterial injection, or intraportal injection). In some embodiments, the lentiviral vector, cell, or pharmaceutical

composition is administered locally to the liver (e.g. by intrahepatic injection, intrahepatic arterial injection, or intra-portal injection).

[0035] The lentiviral vector, cell, or pharmaceutical composition may be administered in any suitable dose. In some embodiments, the lentiviral vector is administered at a dose of at least about 10<sup>8</sup> TU/kg, at least about 10<sup>9</sup> TU/kg, or at least about 10<sup>10</sup> TU/kg. In some embodiments, the lentiviral vector is administered in a dose of from about 10<sup>8</sup> to about 10<sup>11</sup> TU/kg, from about 10<sup>8</sup> to about 10<sup>10</sup> TU/kg, or from about 10<sup>9</sup> to about 10<sup>10</sup> TU/kg.

[0036] The lentiviral vector may integrate into the genome of liver cells and be maintained as the liver cells duplicate. Suitably, the lentiviral vector integrates into the genome of hepatocytes and is maintained as the hepatocytes duplicate.

[0037] Following administration, the subject's serum levels of branched-chain amino acids (BCAA) and/or alloisoleucine may be reduced and/or normalised; and/or the subject's branched-chain alpha-ketoacid dehydrogenase (BCKD) activity may be increased and/or normalised in the liver.

[0038] In one aspect, the present invention provides an immune-shielded lentiviral vector for use in a method of therapy, wherein the method comprises administration of the immune-shielded lentiviral vector to a juvenile or paediatric subject. In one aspect, the present invention provides a cell for use in a method of therapy, wherein the cell comprises an immune-shielded lentiviral vector, and wherein the method comprises administration of the cell to a juvenile or paediatric subject.

[0039] In one aspect, the present invention provides use of an immune-shielded lentiviral vector for the manufacture of a medicament for treatment or prevention of a disease, wherein the treatment or prevention comprises administration of the immune-shielded lentiviral vector to a juvenile or paediatric subject. In one aspect, the present invention provides use of a cell for the manufacture of a medicament for treatment or prevention of a disease, wherein the cell comprises an immune-shielded lentiviral vector, and wherein the treatment or prevention comprises administration of the cell to a juvenile or paediatric subject.

[0040] In one aspect, the present invention provides a method of treatment, wherein the method comprises administration of an immune-shielded lentiviral vector to a juvenile or paediatric subject. In one aspect, the present invention provides a method of treatment, wherein the method comprises administration of a cell to a juvenile or paediatric subject, wherein the cell comprises an immune-shielded lentiviral vector.

[0041] In one aspect, the present invention provides an immune-shielded lentiviral vector for use in a method of treatment of a juvenile or paediatric subject. In one aspect, the present invention provides a cell for use in a method of treatment of a juvenile or paediatric subject, wherein the cell comprises an immune-shielded lentiviral vector.

[0042] In one aspect, the present invention provides use of an immune-shielded lentiviral vector for the manufacture of a medicament for treatment or prevention of a disease in a juvenile or paediatric subject. In one aspect, the present invention provides use of a cell for the manufacture of a medicament for treatment or prevention of a disease in a juvenile or paediatric subject, wherein the cell comprises an immune-shielded lentiviral vector.

[0043] In one aspect, the present invention provides a method of treatment of a juvenile or paediatric subject, wherein the method comprises administration of an immune-shielded lentiviral vector to the juvenile or paediatric subject in need thereof. In one aspect, the present invention provides a method of treatment of a juvenile or paediatric subject, wherein the method comprises administration of a cell to the juvenile or paediatric subject in need thereof, wherein the cell comprises an immune-shielded lentiviral vector.

[0044] In some embodiments, the subject is a neonatal subject or an infantile subject.

[0045] In some embodiments, the method comprises transduction of a liver cell with the immune-shielded lentiviral vector.

[0046] In some embodiments, the immune-shielded lentiviral vector comprises a nucleotide sequence encoding a branched-chain alpha-ketoacid dehydrogenase complex (BCKDC) subunit.

#### DESCRIPTION OF DRAWINGS

[0047] FIG. 1—Design and testing of a lentiviral vector (LV) encoding a BSEP transgene

[0048] (a) Schematic of the third-generation lentiviral vectors (provirus). LTRs, long terminal repeats: the 5' LTR and the 3' LTR each have an almost complete deletion of the U3 region ( $\Delta$ U3);  $\Psi$ , the psi packaging sequence; RRE, Rev response element; cPPT, central polypurine tract; ET, enhanced transtiretin promoter; DBT wt/co, DBT coding sequences (wild-type or codon optimised); WPRE, posttranscriptional element from the genome of the woodchuck hepatitis virus; 4xmir142T, four copies of miR-142 target sequence in tandem.

[0049] (b) Western blot analysis of DBT in Huh7 cells transduced with LV-ET-DBTw or LV-ET-DBTco. kDa, molecular weight marker; Ctrl, untransduced cells. (c) Western blot analysis of DBT in Huh7 cells transduced with LV-hAAT-DBTw or LV-hAAT-DBTco. kDa, molecular weight marker; Ctrl, untransduced cells. (d) DBT protein expression as determined by WB densitometry analysis and normalized on vector copy number (VCN) in Huh7 transduced with LV-ET-DBTw/co at different MOI. (e) VCN in Huh7 transduced with LV-hAAT-DBTw/co at different MOI. (f) DBT protein expression as determined by WB densitometry analysis and normalized on GAPDH house-keeping gene in Huh7 cells transduced with LV-ET-DBTw/co or LV-hAAT-DBTw/co.

[0050] FIG. 2—In vivo gene transfer with LV-ET-DBT

[0051] (a) Vector copy number (VCN) in liver of C571/6 WT mice administered with LV-ET-DBTw or LV-ET-DBTco LV. (b) DBT mRNA, normalized on HPRT expression, in mice treated with LV-ET-DBTw or LV-ET-DBTco. UT, untreated mice. (c) Western blot analysis of DBT in liver samples from C571/6 mice treated with LV-ET-DBTw or LV-ET-DBTco. (d) DBT protein expression as determined by WB densitometry analysis and normalized on  $\beta$ -actin house-keeping gene.

[0052] FIG. 3—Evaluation of LV gene therapy in iMSUD mice

[0053] (a) Survival of iMSUD mice treated with LV-ET-DBTco (KO+LV) compared to iMSUD untreated (KO UT) and WT control littermates (WT). \*\*\*p<0.0001, Long-rank test. (b) Branched chain amino-acids levels normalized on alanine (BCAA/Ala) in iMSUD mice treated with LV-ET-

DBTco (KO+LV) compared to iMSUD untreated (KO UT) and WT control littermates (WT). (c) Alloisoleucine ( $\mu\text{M}$ ) in iMSUD mice treated with LV-ET.DBTco (KO+LV) compared to iMSUD untreated (KO UT) and WT control littermates (WT).

#### DETAILED DESCRIPTION

[0054] It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural referents unless the context clearly dictates otherwise.

[0055] The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps. The terms "comprising", "comprises" and "comprised of" also include the term "consisting of".

[0056] Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, any nucleic acid sequences are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[0057] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that such publications constitute prior art to the claims appended hereto.

[0058] The skilled person will understand that they can combine all features of the invention disclosed herein without departing from the scope of the invention as disclosed.

#### Lentiviral Vector

[0059] In one aspect, the present invention provides a lentiviral vector comprising a protein-coding sequence, wherein the protein-coding sequence is a nucleotide sequence encoding a branched-chain alpha-ketoacid dehydrogenase complex (BCKDC) subunit.

[0060] A "lentiviral vector" may refer to an enveloped lentiviral genome (i.e. a lentiviral particle). For example, the pharmaceutical composition of the present invention preferably comprises the lentiviral vector in the form of a lentiviral particle and the lentiviral vector is preferably administered in the form of a lentiviral particle. In other embodiments, a "lentiviral vector" may comprise a lentiviral genome, optionally wherein the lentiviral genome is enveloped. As used herein, a "lentiviral genome" may refer to a genome that comprises at least one element derived or derivable from a lentivirus genome. Suitably, a lentiviral genome comprises at least one element that is involved in the mechanisms by which a lentivirus infects cells, expresses genes, and/or is replicated.

[0061] Lentivirus is a genus of retroviruses, which contain an RNA genome that is converted to DNA in the transduced cell by a virally encoded reverse transcriptase. Lentiviral vectors can transduce a wide range of cell types and integrate into the host genome in both dividing and post-mitotic cells, resulting in long-term expression of the protein-coding sequence both in vitro and in vivo (see e.g. Tiscornia, G., et al., 2006. *Nature protocols*, 1(1), pp. 241-245).

[0062] The basic genes required for lentivirus survival and function are the gag, pol, and env genes: gag encodes structural proteins; pol encodes enzymes required for

reverse transcription and integration into the host cell genome; and env encodes the viral envelope glycoprotein (see e.g. Milone, M. C. and O'Doherty, U., 2018. *Leukemia*, 32(7), pp. 1529-1541). Lentiviruses may also have additional cis-acting elements, such as a rev response element (RRE), which enables the efficient export of RNA transcripts of the integrated provirus from the nucleus to the cytoplasm of an infected target cell; a retroviral psi packaging element, which is involved in regulating the essential process of packaging the retroviral RNA genome into the viral capsid during replication; a primer binding site (PBS), where reverse transcription is initiated; the TAT activation region (TAR); splice donor and acceptor sites; and central and terminal polypurine tracts, which allow initiation of plus-strand synthesis.

[0063] In a lentivirus genome, these elements are typically flanked at both ends by regions called long terminal repeats (LTRs). The LTRs are responsible for integration and transcription. LTRs may also serve as enhancer-promoter sequences and can control the expression of the lentiviral genes. The LTRs themselves are identical or near-identical sequences that can typically be divided into three regions: U3, R and U5. LTRs may be naturally occurring or may be modified. For example, U3 and U5 modifications are described in Iwakuma, T., et al., 1999. *Virology*, 261(1), pp. 120-132.

[0064] The lentiviral vector of the present invention may comprise a minimal lentiviral genome. As used herein, a "minimal lentiviral genome" may mean that the lentiviral genome has been manipulated so as to remove the non-essential elements and to retain the essential elements in order to provide the required functionality to infect, transduce and deliver a nucleotide sequence of interest to a target host cell (see e.g. Kim, V. N., et al., 1998. *Journal of virology*, 72(1), pp. 811-816 and Sertkaya, H., et al., 2021. *Scientific reports*, 11(1), pp. 1-15).

[0065] A lentiviral genome may comprise from 5' to 3': a 5' LTR, one or more lentiviral-derived cis-acting elements, and a 3' LTR. Suitably, a lentiviral genome may comprise from 5' to 3': a 5' LTR, a RRE, and a 3' LTR. Suitably, a lentiviral genome may comprise from 5' to 3': a 5' LTR, a retroviral psi packaging element, a RRE, and a 3' LTR. Suitably, a lentiviral genome may comprise from 5' to 3': a 5' LTR, a retroviral psi packaging element, a RRE, a cPPT, and a 3' LTR. Suitably, a lentiviral genome may comprise from 5' to 3': a 5' LTR, a retroviral psi packaging element, a RRE, a cPPT, and a 3' LTR.

[0066] A lentiviral genome may further comprise a protein-coding sequence and, optionally, one or more regulatory elements (e.g. operably linked to the protein-coding sequence). Suitably, a lentiviral genome may comprise from 5' to 3': a 5' LTR, a RRE, a protein-coding sequence, and a 3' LTR. Suitably, a lentiviral genome may comprise from 5' to 3': a 5' LTR, a retroviral psi packaging element, a RRE, a protein-coding sequence, and a 3' LTR. Suitably, a lentiviral genome may comprise from 5' to 3': a 5' LTR, a retroviral psi packaging element, a RRE, a cPPT, a protein-coding sequence, and a 3' LTR. Suitably, lentiviral genome may comprise from 5' to 3': a 5' LTR, a PBS, a retroviral psi packaging element, a RRE, a cPPT, a protein-coding sequence, and a 3' LTR.

[0067] The lentiviral vector of the present invention may be replication-defective. Typically, at least part of one or more protein coding regions essential for replication may be

removed from the lentiviral genome. This makes the lentiviral vector “replication-defective” or “replication-incompetent”. Suitably, one or more of gag, pol, rev, and env genes are deleted (at least partially) in a replication-defective lentiviral vector. Suitably, each of the gag, pol, rev, and env genes are deleted (at least partially) in a replication-defective lentiviral vector. Optionally, the lentiviral vector lacks a functional gag-pol and/or env gene and/or other genes essential for replication.

[0068] The lentiviral vector of the present invention may be derived from any lentivirus. As used herein “lentivirus-derived” or “lentivirus-based” may mean that the lentiviral genome comprises one or more elements from said lentivirus. For example, the coding regions of viral proteins may be deleted, but one or more cis-acting element may be retained from said lentivirus.

[0069] The lentiviral vector may be derived from a primate lentivirus. Examples of “primate” lentiviruses include, but are not limited to, human immunodeficiency virus (HIV) and simian immunodeficiency virus (SIV). The lentiviral vector may be derived from a non-primate lentivirus (i.e. derived from a lentivirus which does not primarily infect primates, especially humans). Examples of “non-primate” lentiviruses include, but are not limited to, the prototype “slow virus” visna/maedi virus (VMV), caprine arthritis-encephalitis virus (CAEV), equine infectious anaemia virus (EIAV), feline immunodeficiency virus (FIV), and bovine immunodeficiency virus (BIV).

[0070] Suitably, the lentiviral vector of the present invention is a HIV-derived lentiviral vector. As used herein “HIV-derived” or “HIV-based” may mean that the lentiviral genome comprises one or more element from HIV. For example, the coding regions of HIV viral proteins may be deleted, and one or more HIV cis-acting element may be retained in the lentiviral genome (see e.g. Johnson, N. M., et al., 2021. Molecular Therapy—Methods & Clinical Development, 21, pp. 451-465). A HIV-derived lentiviral genome may comprise from 5' to 3': a 5' LTR, one or more HIV-derived cis-acting elements (e.g. RRE and/or cPPT), and a 3' LTR.

[0071] Suitably, the lentiviral vector of the present invention is a HIV-1-derived lentiviral vector. The prototype lentiviral vector system is based on HIV-1 (see e.g. Merten, O. W., et al., 2016. Molecular Therapy-Methods & Clinical Development, 3, p. 16017). It has been shown that sequences that extend into the gag open reading frame may be important for packaging of HIV-1. Therefore, HIV-1 vectors often contain the relevant portion of gag in which the translational initiation codon has been mutated. In addition, HIV-1 vectors often also contain a portion of the env gene that includes the RRE. Rev binds to RRE, which permits the transport of full-length or singly spliced mRNAs from the nucleus to the cytoplasm. In the absence of rev and/or a RRE, full-length HIV-1 RNAs may accumulate in the nucleus. Alternatively, a constitutive transport element from certain simple retroviruses such as Mason-Pfizer monkey virus can be used to relieve the requirement for rev and a RRE. A HIV-1-derived lentiviral genome may comprise from 5' to 3': a 5' LTR, one or more HIV-1-derived cis-acting elements (e.g. a PBS, a retroviral psi packaging element, a RRE and/or a cPPT), and a 3' LTR.

[0072] The lentiviral vector of the present invention may be a self-inactivating lentiviral vector. As used herein, “self-inactivating” or “SIN” lentiviral vectors may comprise

lentiviral genomes in which the lentiviral enhancer and promoter sequences have been deleted (see e.g. Zufferey, R., et al., 1998. Journal of virology, 72(12), pp. 9873-9880 and Miyoshi, H., et al., 1998. Journal of virology, 72(10), pp. 8150-8157). SIN lentiviral vectors can be generated and transduce non-dividing cells in vivo with an efficacy similar to that of wild-type vectors. The transcriptional inactivation of the long terminal repeat (LTR) in the SIN provirus can prevent mobilisation by replication-competent virus. This can also enable the regulated expression of genes from internal promoters by eliminating any cis-acting effects of the LTR.

[0073] The lentiviral vector of the present invention may be integration competent. As used herein, an “integration competent” lentiviral vector is capable of integrating into the genome of a host cell. In contrast to integration competent lentiviral vectors, integration defective lentiviral vectors can be produced, for example, either by packaging the lentiviral vector with catalytically inactive integrase (such as an HIV integrase bearing the D64V mutation in the catalytic site) or by modifying or deleting essential att sequences from the lentiviral genome LTR, or by a combination of the above (see e.g. Wanisch, K. and Yenez-Munoz, R. J., 2009. Molecular Therapy, 17(8), pp. 1316-1332).

[0074] The lentiviral vector of the present invention may be replication-defective and integrating. The lentiviral vector of the present invention may be replication-defective, integrating, and self-inactivating. The lentiviral vector of the present invention may be replication-defective, integrating, self-inactivating, and HIV-derived.

[0075] The lentiviral vector of the present invention may be a lentiviral particle. A “lentiviral particle” may refer to an enveloped lentiviral genome. Lentiviral particles may be generated by co-transfection of a plasmid containing a lentiviral genome (e.g. a “transfer vector”) with helper plasmids (e.g. “packaging vectors” encoding gag-pol and/or rev and “envelope vectors” encoding env) into host cells and harvesting of the lentivirus-containing supernatant afterwards.

[0076] The lentiviral vector of the present invention may be pseudotyped. Pseudotyping lentiviral vectors with naturally occurring or engineered lentiviral envelopes can allow targeted transduction of specific cell types (see e.g. Joglekar, A. V. and Sandoval, S., 2017. Human Gene Therapy Methods, 28(6), pp. 291-301). Suitably, the lentiviral vector is pseudotyped to allow transduction of liver cells (e.g. hepatocytes).

[0077] The lentiviral vector of the present invention may be VSV-G pseudotyped. Vesicular stomatitis virus G protein (VSV-G) is a commonly used envelope protein for pseudotyping. VSV-G is a trimeric protein that binds phosphatidylserine and low-density lipoprotein receptors on a cell surface to endocytose into the cell. VSV-G pseudotyped lentiviral vectors may efficiently transduce liver cells (e.g. hepatocytes).

[0078] The lentiviral vector of the present invention may be replication-defective, integrating, and VSV-G pseudotyped. The lentiviral vector of the present invention may be replication-defective, integrating, self-inactivating, and VSV-G pseudotyped. The lentiviral vector of the present invention may be replication-defective, integrating self-inactivating, HIV-derived, and VSV-G pseudotyped.

### Immune-Shielded Lentiviral Vector

[0079] The lentiviral vector of the present invention may be an immune-shielded lentiviral vector.

[0080] As used herein, an “immune shielded lentiviral vector” may refer to a lentiviral vector which is modified to reduce immune responses following administration. For example, the immune-shielding may reduce activation of acute inflammatory response after administration.

[0081] In some embodiments, the lentiviral vector of the present invention is: (i) a CD47<sup>high</sup> lentiviral vector; (ii) a MHC-I<sup>free</sup> lentiviral vector; and/or (iii) comprises one or more miRNA target sequences (e.g. which suppress transgene expression in antigen-presenting cells). In preferred embodiments, the lentiviral vector of the present invention is a CD47<sup>high</sup>/MHC-I<sup>free</sup> lentiviral vector and comprises one or more miRNA target sequences (e.g. which suppress transgene expression in antigen-presenting cells). Each of these modifications may act to reduce immune responses following administration.

[0082] An immune-shielded lentiviral vector may be “phagocytosis-shielded” to reduce uptake by professional phagocytes. In some embodiments, the lentiviral vector is a phagocytosis-shielded lentiviral vector. For example, the lentiviral vector may be a CD47<sup>high</sup> lentiviral vector.

### CD47<sup>high</sup> Lentiviral Vectors

[0083] The lentiviral vector of the present invention may be a CD47<sup>high</sup> lentiviral vector. As used herein, a “CD47<sup>high</sup> lentiviral vector” may refer to a lentiviral vector with increased levels of CD47 (or a fragment thereof) on its surface. A CD47<sup>high</sup> lentiviral vector may have reduced uptake by professional phagocytes.

[0084] CD47 (Cluster of Differentiation 47) also known as integrin associated protein (IAP) is a transmembrane protein that in humans is encoded by the CD47 gene. Phagocytosis is physiologically inhibited by CD47, which is a ubiquitously expressed ligand of signal regulatory protein α (SIRP-α) receptor, that is expressed by professional phagocytes. CD47 may be incorporated into lentiviral vectors when they bud from producer cells.

[0085] The lentiviral vector of the present invention may comprise one or more CD47 polypeptides (or a fragment thereof) on its surface. The amount of CD47 (or a fragment thereof) on the surface may be enough to reduce uptake by professional phagocytes. Any suitable assay to quantify the amount of CD47 polypeptides (or fragments thereof) present on the surface of the lentiviral vector may be used.

[0086] In some embodiments, the density of CD47 polypeptides (or fragments thereof) may be determined by immunostaining for CD47 and total internal reflection fluorescence microscopy, e.g. as described in US2010/0316570A1. The CD47 polypeptides (or fragments thereof) may be present in a density of at least about 20 molecules/ $\mu\text{m}^2$ , at least about 25 molecules/ $\mu\text{m}^2$ , at least about 30 molecules/ $\mu\text{m}^2$ , at least about 35 molecules/ $\mu\text{m}^2$ , at least about 40 molecules/ $\mu\text{m}^2$ , at least about 45 molecules/ $\mu\text{m}^2$ , at least about 50 molecules/ $\mu\text{m}^2$ , at least about 60 molecules/ $\mu\text{m}^2$ , at least about 70 molecules/ $\mu\text{m}^2$ , at least about 80 molecules/ $\mu\text{m}^2$ , at least about 90 molecules/ $\mu\text{m}^2$ , at least about 100 molecules/ $\mu\text{m}^2$ , at least about 150 molecules/ $\mu\text{m}^2$ , at least about 200 molecules/ $\mu\text{m}^2$ , at least about 250 molecules/ $\mu\text{m}^2$ , at least about 300 molecules/ $\mu\text{m}^2$ , at least about 350 molecules/ $\mu\text{m}^2$ , at least about 400 molecules/ $\mu\text{m}^2$ , at

least about 450 molecules/ $\mu\text{m}^2$ , at least about 500 molecules/ $\mu\text{m}^2$ , at least about 600 molecules/ $\mu\text{m}^2$ , at least about 700 molecules/ $\mu\text{m}^2$ , at least about 800 molecules/ $\mu\text{m}^2$ , at least about 900 molecules/ $\mu\text{m}^2$ , or at least about 1000 molecules/ $\mu\text{m}^2$ . The CD47 polypeptides (or fragments thereof) may be present in a density of about 1000 molecules/ $\mu\text{m}^2$  or less, about 500 molecules/ $\mu\text{m}^2$  or less, or about 250 molecules/ $\mu\text{m}^2$  or less. The CD47 polypeptides (or fragments thereof) may be present in a density of from about 20 molecules/ $\mu\text{m}^2$  to about 1000 molecules/ $\mu\text{m}^2$ , from about 20 molecules/ $\mu\text{m}^2$  to about 500 molecules/ $\mu\text{m}^2$ , or from about 20 molecules/ $\mu\text{m}^2$  to about 250 molecules/ $\mu\text{m}^2$ .

[0087] In some embodiments, the amount of CD47 polypeptides (or fragments thereof) may be determined by immunostaining for CD47 and electron microscopy, as described in Milani, M., et al., 2019. Science Translational Medicine, 11(493), p.eaav7325. The CD47 polypeptides (or fragments thereof) may be detected in an amount of at least about 10 gold particles/lentiviral particle, at least about 15 gold particles/lentiviral particle, or at least about 20 gold particles/lentiviral particle. The CD47 polypeptides (or fragments thereof) may be detected in an amount of about 100 gold particles/lentiviral particle or less, about 80 gold particles/lentiviral particle or less, or about 60 gold particles/lentiviral particle or less. The CD47 polypeptides (or fragments thereof) may be detected in an amount of from about 10 to about 100 gold particles/lentiviral particle, from about 15 to about 80 gold particles/lentiviral particle, or from about 20 to about 60 gold particles/lentiviral particle.

[0088] The lentiviral vector of the present invention may be obtained from a CD47<sup>high</sup> host cell. As used herein, a “CD47<sup>high</sup> host cell” may refer to a host cell with increased levels of CD47 (or a fragment thereof) on its surface.

[0089] A CD47<sup>high</sup> host cell may be genetically engineered to increase expression of CD47 (or a fragment thereof) on the cell surface. For example, the host cell may comprise a vector encoding CD47 (or a fragment thereof) or may be edited to introduce a nucleotide sequence encoding CD47 (or a fragment thereof) into its genome. Suitably, the host cell is transduced with a viral vector encoding a CD47 polypeptide (or a fragment thereof).

[0090] A CD47<sup>high</sup> host cell may have a higher concentration of CD47 (or a fragment thereof) on its surface than an unmodified host cell (e.g. an unmodified producer cell or packaging cell, as described herein). Suitably, the host cell has at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, or at least about 30-fold more CD47 on its cell surface than an unmodified host cell. Suitably, the host cell has from about 5-fold to about 30-fold more CD47 (or a fragment thereof) on its cell surface than an unmodified host cell.

[0091] Suitably, the lentiviral vector of the present has a higher concentration of CD47 (or a fragment thereof) on its surface than a lentiviral vector obtained from an unmodified host cell (e.g. an unmodified producer cell or packaging cell, as described herein). Suitably, the lentiviral vector has at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at least about 10-fold, at least about 20-fold, at least about 30-fold, at least about 40-fold, or at least about 50-fold more CD47 (or a fragment thereof) on its

surface than a lentiviral vector obtained from an unmodified host cell. Suitably, the lentiviral vector has from about 5-fold to about 30-fold more CD47 (or a fragment thereof) on its surface than a lentiviral vector obtained from an unmodified host cell.

[0092] CD47 is a member of the immunoglobulin (Ig) superfamily of membrane proteins, with a single IgV-like domain at its N-terminus, a highly hydrophobic stretch with five membrane-spanning segments and an alternatively spliced cytoplasmic C-terminus ranging in length from 3 to 36 amino acids. Mouse, rat, bovine and human CD47 molecules have been cloned and show about 70% overall amino acid identity (see e.g. Brown, E. J. and Frazier, W. A., 2001. Trends in cell biology, 11(3), pp. 130-135).

[0093] The CD47 polypeptide (or a fragment thereof) may be a human CD47 polypeptide (or a fragment thereof). A CD47 polypeptide may have an amino acid sequence of UniProtKB Q08722.

[0094] Exemplary CD47 polypeptides are provided by SEQ ID NOs: 1-4. Suitably, a CD47 polypeptide comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to any of SEQ ID NOs: 1-4. Suitably, a CD47 polypeptide comprises or consists of the amino acid sequence of any of SEQ ID NOs: 1-4.

#### Exemplary CD47 polypeptide

(SEQ ID NO: 1)

```
MWPLVAALLLGSACCGSAQLLFNKTKSVEFTFCNDTVVIPCFTVNMEAQNTTEVYVKWKFKG  
RDIYTFDGALNKSTVPTDFSSAKIEVSQLLKGDAKMDKSDAVSHTGNYTCEVTELTRGE  
TIIELKYRVVSWFSPNENILIVIFPIFAILLFWGQFGIKTLKYRSGGMDEKTIALLVAGLVI  
TVIVIVGAILFVPGEYSLKNATGLGLIVTSTGILILLHYYVFSTAIGLTSFVIAILVIQVIA  
YILAVVGLSLCIAACIPMHGPLLISGLSILALAQLLGLVYMKVASNQKTIQPPRNN
```

#### Exemplary CD47 polypeptide

(SEQ ID NO: 2)

```
MWPLVAALLLGSACCGSAQLLFNKTKSVEFTFCNDTVVIPCFTVNMEAQNTTEVYVKWKFKG  
RDIYTFDGALNKSTVPTDFSSAKIEVSQLLKGDAKMDKSDAVSHTGNYTCEVTELTRGE  
TIIELKYRVVSWFSPNENILIVIFPIFAILLFWGQFGIKTLKYRSGGMDEKTIALLVAGLVI  
TVIVIVGAILFVPGEYSLKNATGLGLIVTSTGILILLHYYVFSTAIGLTSFVIAILVIQVIA  
YILAVVGLSLCIAACIPMHGPLLISGLSILALAQLLGLVYMKVASNQKTIQPPRKAVEEPL  
NAFKESKGMMNDE
```

#### Exemplary CD47 polypeptide

(SEQ ID NO: 3)

```
MWPLVAALLLGSACCGSAQLLFNKTKSVEFTFCNDTVVIPCFTVNMEAQNTTEVYVKWKFKG  
RDIYTFDGALNKSTVPTDFSSAKIEVSQLLKGDAKMDKSDAVSHTGNYTCEVTELTRGE  
TIIELKYRVVSWFSPNENILIVIFPIFAILLFWGQFGIKTLKYRSGGMDEKTIALLVAGLVI  
TVIVIVGAILFVPGEYSLKNATGLGLIVTSTGILILLHYYVFSTAIGLTSFVIAILVIQVIA  
YILAVVGLSLCIAACIPMHGPLLISGLSILALAQLLGLVYMKV
```

#### Exemplary CD47 polypeptide

(SEQ ID NO: 4)

```
MWPLVAALLLGSACCGSAQLLFNKTKSVEFTFCNDTVVIPCFTVNMEAQNTTEVYVKWKFKG  
RDIYTFDGALNKSTVPTDFSSAKIEVSQLLKGDAKMDKSDAVSHTGNYTCEVTELTRGE  
TIIELKYRVVSWFSPNENILIVIFPIFAILLFWGQFGIKTLKYRSGGMDEKTIALLVAGLVI  
TVIVIVGAILFVPGEYSLKNATGLGLIVTSTGILILLHYYVFSTAIGLTSFVIAILVIQVIA  
YILAVVGLSLCIAACIPMHGPLLISGLSILALAQLLGLVYMKVASNQKTIQPPRKAVEEPL
```

**[0095]** Exemplary CD47 polypeptides excluding the signal peptide are provided by SEQ ID NOs: 5-8. Suitably, a CD47 polypeptide comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to any of SEQ ID NOs: 5-8. Suitably, a CD47 polypeptide comprises or consists of the amino acid sequence of any of SEQ ID NOs: 5-8.

and/or variant may have at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 100% of the activity of a full-length CD47 polypeptide.

**[0098]** Suitably, a variant of SEQ ID NO: 1 may comprise one or more variation selected from V51, C14W, C15R, F22L, S27F, F30L, F32Y, T36S, V38L, V38I, F42V, T44A, N50S, T51A, T52S, T52A, V56I, R63K, A71T, S75Y, T76A, P78L, P78S, P78A, S82R, S82N, S83T, K85N, K85E, V88A, V88L, V88I, Q90R, L91F, K93N, M100I, M100V,

```

Exemplary CD47 polypeptide excluding signal peptide
(SEQ ID NO: 5)
QLLFNKTGSVEFTCNDTVVIPCFTVNMEAQNTEVYVKWFKGRDIYTFDGALNKSTVPTD

FSSAKIEVSQLLKGDAALKMDKSDAVSHGNYTCEVTELTRGETIIELKYRVRVSWFSPNEN

ILIVIFPIFAILFWGQFGIKTLKYRSGGMDEKTIALLVAGLVITVIVVGAILFVPGEYSL

KNATGLGLIVTSTGILILLHYYVSTAIGLTSFVIAILVIQVIAYILAVVGLSLCIAACIPM

HGPLLISGLSILALAQLLGLVYMKEVASNQKTIQPPRNN

Exemplary CD47 polypeptide excluding signal peptide
(SEQ ID NO: 6)
QLLFNKTGSVEFTCNDTVVIPCFTVNMEAQNTEVYVKWFKGRDIYTFDGALNKSTVPTD

FSSAKIEVSQLLKGDAALKMDKSDAVSHGNYTCEVTELTRGETIIELKYRVRVSWFSPNEN

ILIVIFPIFAILFWGQFGIKTLKYRSGGMDEKTIALLVAGLVITVIVVGAILFVPGEYSL

KNATGLGLIVTSTGILILLHYYVSTAIGLTSFVIAILVIQVIAYILAVVGLSLCIAACIPM

HGPLLISGLSILALAQLLGLVYMKFVASNQKTIQPPRKAVEEPLNAFKESKGMMNDE

Exemplary CD47 polypeptide excluding signal peptide
(SEQ ID NO: 7)
QLLFNKTGSVEFTCNDTVVIPCFTVNMEAQNTEVYVKWFKGRDIYTFDGALNKSTVPTD

FSSAKIEVSQLLKGDAALKMDKSDAVSHGNYTCEVTELTRGETIIELKYRVRVSWFSPNEN

ILIVIFPIFAILFWGQFGIKTLKYRSGGMDEKTIALLVAGLVITVIVVGAILFVPGEYSL

KNATGLGLIVTSTGILILLHYYVSTAIGLTSFVIAILVIQVIAYILAVVGLSLCIAACIPM

HGPLLISGLSILALAQLLGLVYMKFV

Exemplary CD47 polypeptide excluding signal peptide
(SEQ ID NO: 8)
QLLFNKTGSVEFTCNDTVVIPCFTVNMEAQNTEVYVKWFKGRDIYTFDGALNKSTVPTD

FSSAKIEVSQLLKGDAALKMDKSDAVSHGNYTCEVTELTRGETIIELKYRVRVSWFSPNEN

ILIVIFPIFAILFWGQFGIKTLKYRSGGMDEKTIALLVAGLVITVIVVGAILFVPGEYSL

KNATGLGLIVTSTGILILLHYYVSTAIGLTSFVIAILVIQVIAYILAVVGLSLCIAACIPM

HGPLLISGLSILALAQLLGLVYMKFVASNQKTIQPPRKAVEEPLN

```

**[0096]** A person skilled in the art would be able to generate variants and/or fragments based on conservative substitutions and/or the known structural and functional features of CD47. These are described, for instance in Fenalti, G., et al., 2021. *Nature communications*, 12(1), pp. 1-14.

**[0097]** Suitably, a fragment of CD47 and/or CD47 variant retains the ability to inhibit phagocytosis. Suitably, a CD47 fragment and/or CD47 variant may comprise the extracellular domain of CD47. The extracellular domain of human CD47 may interact with SIRP- $\alpha$  and inhibit phagocytosis. Optionally, a CD47 fragment and/or CD47 variant comprises the transmembrane domain of CD47. The domains may be linked by inter-domain linker(s). The fragment

D101G, K102R, K102T, S107L, I126F, I127V, K130Q, R132H, S138F, V146I, I150V, I153V, S169A, G170R, G170S, G171S, D173Y, I177V, A178G, V181I, V185A, I186V, V188A, I189T, I191V, V198I, A207S, T215I, I219M, Y226C, A231S, T235A, S236F, A240V, A240T, I241V, V243I, I244T, V246L, Y249F, A252S, A252T, V254A, S257T, I264M, I264L, M266I, M266T, M266V, V287I, V292A, N295S, N295D, Q296L, P302S, N304S, and N304D. These are considered to be tolerated, benign, and/or likely benign variations as predicted by SIFT, PolyPhen, CADD, REVEL, and MetaLR.

**[0099]** Suitably, a variant of SEQ ID NO: 2 may comprise one or more variation selected from P3L, A6P, F22L, S27F,

F30L, F32Y, T36S, V38L, V38I, F42V, N50S, T51A, T52S, T52A, V56I, R63K, A71T, S75Y, T76A, P78L, P78S, P78A, S82R, S82N, S83T, K85N, K85E, V88A, V88L, V88I, Q90R, L91F, K93N, M100I, M100V, D101G, K102R, K102T, S107L, I126F, I127V, K130Q, R132H, S138F, V146I, I150V, I153V, S169A, G170R, G170S, G171S, I177V, A178G, V181I, I186V, V188A, I189T, I191V, V198I, A207S, T215I, I219M, Y226C, A231S, T235A, A240V, A240T, I241V, V243I, I244T, V246L, Y249F, A252S, A252T, V254A, I264M, I264L, M266I, M266T, M266V, V287I, V292A, N295S, N295D, and Q296L. These are considered to be tolerated, benign, and/or likely benign variations as predicted by SIFT, PolyPhen, CADD, REVEL, and MetaLR.

**[0100]** An exemplary CD47 fragment is provided by SEQ ID NO: 9. Suitably, a CD47 fragment comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to SEQ ID NO: 9. Suitably, a CD47 fragment comprises or consists of the amino acid sequence of SEQ ID NO: 9.

#### Exemplary CD47 fragment

(SEQ ID NO: 9)

```
MWPLVAALLGSACCGSAQLLFNKTKSVEFTFCNDTVVIPCFVTNMEAQ
NTTEVVVKWFKGRDIYTFDGALNKSTVPTDFSSAKIEVSQLLKGDASL
KMDKSDAVSHTGNYTCEVTEL TREGETIIELKYRVSWSESPN
```

**[0101]** An exemplary CD47 fragment excluding the signal peptide is provided by SEQ ID NO: 10. Suitably, a CD47 fragment comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to SEQ ID NO: 10. Suitably, a CD47 fragment comprises or consists of the amino acid sequence of SEQ ID NO: 10.

Exemplary CD47 fragment excluding signal peptide  
(SEQ ID NO: 10)

```
QLLFNKTKSVEFTFCNDTVVIPCFVTNMEAQNTTEVVVKWFKGRDIYT
EDGALNKSTVPTDESSAKIEVSQLLKGDASLKMDKSDAVSHTGNYTCEV
TEL TREGETIIELKYRVSWSESPN
```

#### MHC-I<sup>low</sup> or MHC-I<sup>free</sup> Lentiviral Vectors

**[0102]** The lentiviral vector of the present invention may be a MHC-I<sup>low</sup> lentiviral vector or a MHC-I<sup>free</sup> lentiviral vector. In preferred embodiments, the lentiviral vector of the present invention is a MHC-I<sup>free</sup> lentiviral vector.

**[0103]** As used herein, a “MHC-I<sup>low</sup> lentiviral vector” may refer to a lentiviral vector with reduced levels of one or more MHC-I molecules on its surface (i.e. reduced levels of surface-exposed MHC-I molecules). The number of surface-exposed MHC-I molecules may be reduced such that the immune response to the MHC-I is decreased to a therapeutically relevant degree.

**[0104]** As used herein, a “MHC-I<sup>free</sup> lentiviral vector” may refer to a lentiviral vector which is substantially devoid of

(or free of) one or more MHC-I molecules on its surface (i.e. substantially devoid of (or free of) surface-exposed MHC-I molecules).

**[0105]** The major histocompatibility complex class I (MHC-I) is a heterodimeric membrane protein that is displayed on the outer leaflet of the cell membrane (see e.g. Penn, D. J. and Ilmonen, P., 2005. Major histocompatibility complex (MHC). eLS). MHC-I functions to bind and display peptide fragments of proteins to the extracellular environment where they may be recognised by CD8+ cytotoxic T cells. Peptide fragments generated from normal cellular proteins will not activate cytotoxic T cells due to central and peripheral tolerance mechanisms. However, foreign peptides (e.g. those originating from viral proteins) will cause activation of an immune response to destroy the cell. An allogeneic MHC-I protein itself may be recognised by the immune system. For example, antibodies may bind MHC-I epitopes directly. As a result, lentiviral vectors that comprise MHC-I molecules originating from an allogeneic source may be targeted and neutralised by the immune system.

**[0106]** The term “MHC-I molecules” may refer to human MHC-I molecules. Human MHC-I is also referred to as human leukocyte antigen class I (HLA-I) and is expressed on almost all nucleated cells. HLA-I consists of two polypeptide chains, an HLA-I heavy chain ( $\alpha$  chain) and  $\beta 2$  microglobulin ( $\beta 2M$  or  $\beta$  chain). The HLA-I  $\alpha$  chain and  $\beta 2M$  are linked non-covalently. The HLA-I  $\alpha$  chain is polymorphic. Six HLA-I  $\alpha$  chains have been identified to date, including three classical, highly polymorphic  $\alpha$  chains (HLA-A, HLA-B and HLA-C) and three non-classical, less polymorphic (HLA-E, HLA-F and HLA-G)  $\alpha$  chains. The MHC-I molecules may comprise or consist of HLA-A, HLA-B, and HLA-C molecules, which comprise an invariant  $\beta 2M$  sequence.

**[0107]** The term “MHC-1 molecules” may also include variant MHC-I sequences, such as polymorphisms of HLA-I  $\alpha$  chain sequences and/or  $\beta 2M$  sequences. For example, variant MHC-I sequences may include HLA-I  $\alpha$  chain sequences and/or  $\beta 2M$  sequences with single nucleotide polymorphisms (SNPs) or multiple SNPs.

**[0108]** Any suitable assay to quantify the amount of MHC-I molecules present on the surface of the lentiviral vector may be used.

**[0109]** In some embodiments, the amount of MHC-I molecules may be determined by immunostaining for MHC-I and electron microscopy, as described in Milani, M., et al., 2017. EMBO molecular medicine, 9(11), pp. 1558-1573. The MHC-I molecules may be detected in an amount of less than about 10 gold particles/lentiviral particle, less than about 9 gold particles/lentiviral particle, less than about 8 gold particles/lentiviral particle, less than about 7 gold particles/lentiviral particle, less than about 6 gold particles/lentiviral particle, less than about 5 gold particles/lentiviral particle, less than about 4 gold particles/lentiviral particle, less than about 3 gold particles/lentiviral particle, less than about 2 gold particles/lentiviral particle, less than about 1 gold particle/lentiviral particle, or about 0 gold particles/lentiviral particle. The MHC-I molecules may be undetectable (e.g. the amount of gold particles detected may not be significantly higher than background levels).

**[0110]** The lentiviral vector of the present invention may be obtained from a MHC-I<sup>low</sup> host cell or a MHC-I<sup>free</sup> host cell. In preferred embodiments, the lentiviral vector of the present invention is obtained from a MHC-I<sup>free</sup> host cell. As

used herein, a “MHC-I<sup>low</sup> host cell” may refer to a host cell with reduced levels of one or more MHC-I molecule on its surface. As used herein, a “MHC-I<sup>free</sup> host cell” may refer to a host cell which is substantially devoid of or free of one or more MHC-I molecule on its surface.

[0111] A MHC-I<sup>low</sup> or MHC-I<sup>free</sup> host cell may be genetically engineered to decrease expression of MHC-I on the cell surface. For example, the cell may comprise a genetically engineered disruption of a gene encoding β2-microglobulin and/or a genetically engineered disruption of a gene encoding an MHC-I α chain.

[0112] Methods for genetic engineering to decrease protein expression are known in the art. For example, this may be achieved by targeted gene knockout. To decrease protein expression, the gene encoding the protein itself or its regulatory sequence (e.g. its promoter) may be knocked out. Knockout may be achieved by deletion of a section of the coding nucleic acid sequence, which may delete a section of the protein essential for expression or stability, or alter the reading frame of the coding sequence or by base-editing. Suitable methods for targeted gene knockout include use of zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR/Cas-based RNA-guided nucleases (see e.g. Gaj, T. et al., 2013. Trends Biotechnol, 31, pp. 397-405). For example, the CRISPR/Cas9 RNA-guided nuclease may be used to catalyse a double strand break at a specific locus in the genome if provided with appropriate RNA guides designed to bind that locus. Cas9 and the guide RNA may be delivered to a target cell by transfection of vectors encoding the protein and RNA. Cells attempt to repair any double strand breaks in their DNA using the non-homologous end joining (NHEJ) pathway. This is an error-prone mechanism which inserts random nucleotides and often disrupts the reading frame of the targeted gene. Alternatively, the genetic engineering to decrease protein expression may be accomplished using RNAi techniques, microRNA or antisense RNA to suppress expression of the target gene.

[0113] Once the targeted gene knockout or suppression of expression approach has been carried out, the resulting population of cells may be screened to select and enrich for those cells exhibiting the phenotype of interest, for example decreased expression of surface-exposed MHC-I. Suitable techniques for screening and enrichment are known in the art and include flow cytometry and fluorescence-activated cell sorting (FACS).

[0114] In some embodiments, the host cell comprises a genetically engineered disruption of a gene encoding β2-microglobulin. β2-microglobulin stabilises MHC-I, thus cells deficient in β2-microglobulin will exhibit decreased expression of MHC-I on the surface of the cell. The cell may comprise genetically engineered disruptions in all copies of the gene encoding β2-microglobulin.

[0115] In another embodiment, the cell comprises a genetically engineered disruption of one or more gene encoding an MHC-I α chain. The cell may comprise genetically engineered disruptions in all copies of the gene encoding an MHC-I α chain.

[0116] The cell may comprise both genetically engineered disruptions of genes encoding β2-microglobulin and genetically engineered disruptions of genes encoding an MHC-I α chain.

[0117] Decreased expression of MHC-I on the surface of the cell may refer to a decrease in the number of MHC-I molecules that are expressed on the surface of the cell that has been genetically engineered, in comparison to the number of MHC-I molecules that are expressed on the surface of a cell lacking the genetic engineering, but under otherwise substantially identical conditions. The expression of MHC-I on the surface of the cell may be decreased such that the number of surface-exposed MHC-I molecules is, for example, less than about 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 4%, 3%, 2% or 1% of the number of surface-exposed MHC-I molecules that are displayed in the absence of the genetic engineering. In some embodiments, the expression of MHC-I on the surface of the cell is decreased such that the number of surface-exposed MHC-I molecules is 0% of the number of surface-exposed MHC-I molecules that are displayed in the absence of the genetic engineering.

[0118] The expression of MHC-I on the surface of the cell is preferably decreased such that the cell is substantially devoid of surface-exposed MHC-I molecules. In this context, “substantially devoid” may mean that there is a substantial decrease in the number of MHC-I molecules that are expressed on the surface of the cell that has been genetically engineered, in comparison to the number of MHC-I molecules that are expressed on the surface of a cell lacking the genetic engineering, such that the immune response to MHC-I on lentiviral vectors produced by the cell is decreased to a therapeutically useful degree.

[0119] Suitably, the lentiviral vector of the present invention has a lower concentration of MHC-I molecules on its surface than a lentiviral vector obtained from an unmodified host cell (e.g. an unmodified producer cell or packaging cell, as described herein). Suitably, the lentiviral vector has less than about 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 4%, 3%, 2% or 1% of the number of surface-exposed MHC-I molecules that are displayed on a lentiviral vector obtained from an unmodified host cell. In some embodiments, the lentiviral vector has less than about 20% of the number of surface-exposed MHC-I molecules that are displayed on a lentiviral vector obtained from an unmodified host cell.

[0120] In some embodiments, the lentiviral vector of the present invention is substantially devoid of MHC-I molecules on its surface. In this context, “substantially devoid” may mean that there is no detectable immune response due to the molecules on the surface of the lentiviral vector.

[0121] In some embodiments, the lentiviral vector of the present invention is free of MHC-I molecules on its surface. In this context, “free” may mean that there are no detectable molecules (e.g. by immunostaining and electron microscopy) on the surface of the lentiviral vector. As used herein, “not detectable” may refer to levels which are not statistically significantly different compared to background levels.

[0122] In some embodiments, the lentiviral vector of the present invention has decreased HLA-A, HLA-B, and/or HLA-C molecules on its surface. Suitably, the lentiviral vector has less than about 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 4%, 3%, 2% or 1% of the number of surface-exposed HLA-A molecules that are displayed on a lentiviral vector obtained from an unmodified host cell. Suitably, the lentiviral vector has less than about 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 4%, 3%, 2% or 1% of the number of surface-exposed HLA-B molecules that are displayed on a lentiviral vector obtained from an unmodified host cell. Suitably, the lentiviral vector has less than about 50%, 40%,

30%, 25%, 20%, 15%, 10%, 5%, 4%, 3%, 2% or 1% of the number of surface-exposed HLA-C molecules that are displayed on a lentiviral vector obtained from an unmodified host cell.

[0123] In some embodiments, the lentiviral vector of the present invention is substantially devoid of HLA-A, HLA-B, and/or HLA-C molecules on its surface. In some embodiments, the lentiviral vector of the present invention is substantially devoid of HLA-A, HLA-B, and HLA-C molecules on its surface. In some embodiments, the lentiviral vector of the present invention is free of HLA-A, HLA-B, and/or HLA-C molecules on its surface. In some embodiments, the lentiviral vector of the present invention is free of HLA-A, HLA-B, and HLA-C molecules on its surface.

[0124] As described above, an HLA-I molecule consists of two polypeptide chains, an HLA-I heavy chain ( $\alpha$  chain) and  $\beta$ 2 microglobulin ( $\beta$ 2M or  $\beta$  chain). The HLA-I  $\alpha$  chain and  $\beta$ 2M are linked non-covalently.

[0125] The skilled person would readily be able to determine amino acid and nucleic acid sequences of HLA-I  $\alpha$  chains. For example, the HLA-I  $\alpha$  chains may be identified in a genome sequence using their location within the major histocompatibility complex region of the chromosome (see e.g. Penn, D. J. and Ilmonen, P., 2005. Major histocompatibility complex (MHC). eLS).

[0126] HLA-A alpha chains may have an amino acid sequence of UniProtKB P04439. Exemplary HLA-A alpha chains are provided by SEQ ID NOS: 11 and 12. Suitably, an HLA-A alpha chain comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to SEQ ID NO: 11 or 12. Suitably, a HLA-A alpha chain comprises or consists of the amino acid sequence of SEQ ID NO: 11 or 12.

#### Exemplary HLA-A alpha chain

(SEQ ID NO: 11)

```
MAVMAPRTLLLLLSSGALALTQTWAGSHSMRYFFTSVSRPGRGEPRFIAV
GYVDDTQFVRFDSAASQRMEPAPWIEQEGPEYWDQETRNVKAQSQTD
RVDLGTLRGYYNQSEAGSHTIQIMYGCDVGSDGRFLRGYRQDAYDGKDY
IALNEDLRSWTAADMAAQITKRKWEAAHEAEQLRAYLDGTCVEWLRRYL
ENGKETLQRTDPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDG
EDQTQDTELVELTRPAGDGTTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPL
TLRWELOSSOPTIPIVGIAGLVLLGAVITGAVVAAMWRRKSSDRKGGS
YTQAASSDSAQGSDVSLTACKV
```

#### Exemplary HLA-A alpha chain

(SEQ ID NO: 12)

```
MAVMAPRTLLLLLSSGALALTQTWAGSHSMRYFFTSVSRPGRGEPRFIAV
GYVDDTQFVRFDSAASQRMEPAPWIEQEGPEYWDQETRNVKAQSQTD
RVDLGTLRGYYNQSEAGSHTIQIMYGCDVGSDGRFLRGYRQDAYDGKDY
IALNEDLRSWTAADMAAQITKRKWEAAHAAEQLRAYLEGRVCVEWLRRYL
ENGKETLQRTDPKTHMTHHPISDHEATLRCWALGFYPAEITLTWQRDG
EDQTQDTELVELTRPAGDGTTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPL
```

-continued

```
TLRWELOSSOPTIPIVGIAGLVLLGAVITGAVVAAMWRRKSSGGEGVK
DRKGGSYTOQAASSDSAQGSDVSLTACKV
```

[0127] HLA-B alpha chains may have an amino acid sequence of UniProtKB P01889. An exemplary HLA-B alpha chain is provided by SEQ ID NO: 13. Suitably, an HLA-B alpha chain comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to SEQ ID NO: 13. Suitably, a HLA-B alpha chain comprises or consists of the amino acid sequence of SEQ ID NO: 13.

#### Exemplary HLA-B alpha chain

(SEQ ID NO: 13)

```
MLVMAPRTVLLLLSAALALTETWAGSHSMRYFYTSVSRPGRGEPRFISV
GYVDDTQFVRFDSAASPREEPRAPWIEQEGPEYWDRNTQIYKAQAQTD
RESLRNLRGYYNQSEAGSHTLQSMYGCVDGPDGRLLRGHDQYAYDGKDY
IALNEDLRSWTAADTAQITQRKWEAAREAEQRRAYLEGECEVWLRRYL
ENGKDKLERADPPKTHVTHHPISDHEATLRCWALGFYPAEITLTWQRDG
EDQTQDTELVELTRPAGDRTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPL
TLRWEPSQSTPIVGIAGLVLLGAVITGAVVAAMWRRKSSGGKGS
YSQAACSDSAQGSDVSLTA
```

[0128] HLA-C alpha chains may have an amino acid sequence of UniProtKB P10321. Exemplary HLA-C alpha chains are provided by SEQ ID NOS: 14 and 15. Suitably, an HLA-C alpha chain comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to SEQ ID NO: 14 or 15. Suitably, a HLA-C alpha chain comprises or consists of the amino acid sequence of SEQ ID NO: 14 or 15.

#### Exemplary HLA-C alpha chain

(SEQ ID NO: 14)

```
MRVMAPRALLLLSGGLALTETWACSHSMRYFDATVSRPGRGEPRFISV
GYVDDTQFVRFDSAASPRGEPRAPWIEQEGPEYWDRETOKYKRQAQAD
RVSLRNLRGYYNQSEDGSHTLQRMSGCSDLGPDRLLRGYDQSAYDGKDY
IALNEDLRSWTAADTAQITQRKWEAAREAEQLRAYLEGTCEVWLRRYL
ENGKETLQRAEPPKTHVTHPLSDHEATLRCWALGFYPAEITLTWQRDG
EDQTQDTELVELTRPAGDGTTFQKWAAVVVPSGEEQRYTCHMQHEGLQPL
TLSWEPSSQPTIPIMGIVAGLAVVLAFLGAVVTAMMCRRKSSGGKGG
SCSQAACNSAQGSDESPLITCKA
```

#### Exemplary HLA-C alpha chain

(SEQ ID NO: 15)

```
MRVMAPRALLLLSGGLALTETWACSHSMRYEDTAVSRPGRGEPRFISV
GYVDDTQFVRFDSAASPRGEPRAPWIEQEGPEYWDRETQKYKRQAQAD
RVSLRNLRGYYNQSEDGSHTLQRMSGCSDLGPDRLLRGYDQSAYDGKDY
```

- continued

```
IALNEHLSCTAADTAAQITQRKLEAARAAEQLRAYLEGTCVEWLRRYL  
ENGKETLQRAEPPKTHVTHHPLSDHEATLRCWALGFYPAEITLTWQRDG  
EDQTQDTTELVETRPAGDGTFKWAADVVPSCQEQRYTCHMQHEGLQEP  
TLRWGGKGGSQAACNSAQQGSDSLITCKA
```

**[0129]** Amino acid and nucleic acid sequences encoding β2M are also known in the art. For example, a nucleic acid sequence of a human β2M is deposited as GenBank Accession No. NM\_004048.

**[0130]** An HLA β chain may be that of UniProtKB P61769. An exemplary HLA β chain is provided by SEQ ID NO: 16. Suitably, a HLA β chain comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identity to SEQ ID NO: 16. Suitably, a HLA β chain comprises or consists of the amino acid sequence of SEQ ID NO: 16.

#### Exemplary HLA beta chain

```
(SEQ ID NO: 16)  
MSRSVALVALLSLSGLEAIQRTPKIQVYSRHPAENGKSNFLNCYVSG  
FHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEPTPTKEDEY  
ACRVNHVTLSQPKIVKWDRDM
```

**[0131]** The lentiviral vector of the present invention may be a CD47<sup>high</sup>/MHC-I<sup>free</sup> lentiviral vector or a CD47<sup>high</sup>/MHC-I<sup>low</sup> lentiviral vector. In preferred embodiments, the lentiviral vector of the present invention is a CD47<sup>high</sup>/MHC-I<sup>free</sup> lentiviral vector.

**[0132]** The lentiviral vector of the present invention may be obtained from a CD47<sup>high</sup>/MHC-I<sup>free</sup> host cell or a CD47<sup>high</sup>/MHC-I<sup>low</sup> host cell. In preferred embodiments, the lentiviral vector of the present invention is obtained from a CD47<sup>high</sup>/MHC-I<sup>free</sup> host cell.

#### miRNA Target Sequence

**[0133]** The lentiviral vector of the present invention may comprise one or more miRNA target sequences. The one or more miRNA target sequences may be operably linked to the protein-coding sequence. The term “operably linked” may mean that the components described are in a relationship permitting them to function in their intended manner.

**[0134]** MicroRNA (miRNA) genes are scattered across all human chromosomes, except for the Y chromosome. Similar to protein-coding genes, miRNAs are usually transcribed from polymerase-II promoters, generating a so-called primary miRNA transcript (pri-miRNA). From the pri-miRNA, a stem loop of about 60 nucleotides in length, called miRNA precursor (pre-miRNA), is excised leaving a 5' phosphate and a 2 bp long, 3' overhang. The pre-miRNA is then actively transported from the nucleus to the cytoplasm. Then, Dicer performs a double strand cut at the other end of the stem loop, generating a 19-24 bp duplex, which is composed of the mature miRNA and the opposite strand of the duplex, called miRNA\*. One strand of the duplex is selectively loaded into the RNA-induced silencing complex (RISC), and accumulates as the mature microRNA. This strand is usually the one whose 5' end is less tightly paired

to its complement. However, there are some miRNAs that support accumulation of both duplex strands to similar extent.

**[0135]** Once loaded into RISC, the guide strand of the mature microRNA interacts with mRNA target sequences preferentially found in the 3' untranslated region (3'UTR) of protein-coding genes. If the whole guide strand sequence is perfectly complementary to the mRNA target, the mRNA is endonucleolytically cleaved. If only the seed sequence (i.e. nucleotides 2-8 counted from the 5' end of the miRNA) is perfectly complementary to the target mRNA, RNAi may act through alternative mechanisms leading to translational repression.

**[0136]** Expression of the protein from the protein-coding sequence (i.e. “transgene expression”) may be regulated by one or more endogenous miRNAs using one or more corresponding miRNA target sequences. Using this method, one or more miRNAs endogenously expressed in a cell prevent or reduce transgene expression in that cell by interacting with its corresponding miRNA target sequence positioned in the lentiviral genome (see e.g. Brown, B. D. et al. (2007) Nat Biotechnol 25: 1457-1467).

**[0137]** Suitable miRNA target sequences which suppress transgene expression in specific cells will be known to the skilled person. Determining a miRNA with the desired expression profile may be achieved using techniques known to those skilled in the art. For example, a mammalian microRNA expression atlas is described in Landgraf, P., et al., 2007. Cell, 129(7), pp. 1401-1414 and the distribution of miRNA expression across human tissues is described in Ludwig, N., et al., 2016. Nucleic acids research, 44(8), pp. 3865-3877. Once a miRNA has been identified, the corresponding target sequence can readily be determined using, for example, a microRNA database, such as miRBase (Griffiths-Jones, S., et al., 2007. Nucleic acids research, 36(suppl\_1), pp. D154-D158).

**[0138]** A miRNA target sequence may be fully or partially complementary to the corresponding miRNA. The term “fully complementary”, as used herein, may mean that the target sequence has a nucleic acid sequence which is 100% complementary to the sequence of the miRNA which recognises it. The term “partially complementary”, as used herein, may mean that the target sequence is only in part complementary to the sequence of the miRNA which recognises it, whereby the partially complementary sequence is still recognised by the miRNA. In other words, a partially complementary target sequence in the context of the present invention is effective in recognising the corresponding miRNA and effecting prevention or reduction of transgene expression in cells expressing that miRNA. Suitably, a partially complementary miRNA target sequence may be fully complementary to the miRNA seed sequence.

**[0139]** Including more than one copy of a miRNA target sequence in a lentiviral vector may increase the effectiveness of the system. Also, different miRNA target sequences can be included. For example, the protein-coding sequence may be operably linked to more than one miRNA target sequence, which may or may not be different. The miRNA target sequences may be in tandem, but other arrangements are envisaged. The lentiviral vector may, for example, comprise 1, 2, 3, 4, 5, 6, 7 or 8 copies of the same or different miRNA target sequences. Suitably, the lentiviral vector comprises 4 miRNA target sequences of each miRNA target sequence.

[0140] Copies of miRNA target sequences may be separated by a spacer sequence. A spacer sequence may comprise, for example, at least one, at least two, at least three, at least four or at least five nucleotide bases.

[0141] Suitably, the lentiviral vector comprises one or more miRNA target sequence, two or more miRNA target sequences, three or more miRNA target sequences, or four or more miRNA target sequences. Suitably, the protein-coding sequence is operably linked to one or more miRNA target sequence, two or more miRNA target sequences, three or more miRNA target sequences, or four or more miRNA target sequences. In some embodiments, the protein-coding sequence is operably linked to four miRNA target sequences.

[0142] The miRNA target sequence may be a human miRNA target sequence. Suitably, the miRNA target sequence is a -5p or -3p miRNA target sequence.

[0143] The one or more miRNA target sequence may suppress transgene expression in one or more cells other than liver cells (e.g. hepatocytes).

[0144] The one or more miRNA target sequence may suppress transgene expression in hematopoietic-lineage cells. Hematopoietic stem cells give rise to different types of blood cells, in lines called myeloid and lymphoid. As used herein, "hematopoietic-lineage cells" may include myeloid cells and lymphoid cells. Myeloid cells may include monocytes, macrophages, neutrophils, basophils, and eosinophils. Lymphoid cells may include T cells, B cells, natural killer cells, and innate lymphoid cells.

[0145] The one or more miRNA target sequences may suppress transgene expression in antigen-presenting cells. As used herein, an "antigen presenting cell" (APC) may refer to a cell that displays antigen bound by major histocompatibility complex (MHC) proteins on its surface. APCs may be hematopoietic-lineage cells. The antigen-presenting cells may be professional antigen-presenting cells. Professional APCs specialise in presenting antigens to T cells and may include macrophages, B cells and dendritic cells. Suitably, the APCs are splenic and/or hepatic APCs.

[0146] The one or more miRNA target sequences may suppress transgene expression in hematopoietic-lineage antigen-presenting cells.

[0147] By preventing transgene expression in antigen-presenting cells, while permitting high levels of expression in other cells, miRNA regulation may enable strong and stable gene transfer in the absence of an immune response.

[0148] As used herein, the term "suppress expression" may refer to a reduction of expression in the relevant cell type(s) of a transgene to which the one or more miRNA target sequence is operably linked as compared to transgene expression in the absence of the one or more miRNA target sequence, but under otherwise substantially identical conditions. In some embodiments, transgene expression is suppressed by at least 50%. In some embodiments, transgene expression is suppressed by at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%. In some embodiments, transgene expression is substantially prevented, e.g. not detectable.

[0149] The miRNA-mediated approach for restricting gene expression has several advantages over other strategies of regulating transgenes. Although using tissue-specific promoters can successfully limit expression to target cells, leaky expression in a fraction of non-target cells is observed. This occurs because the reconstituted promoter, modified for

inclusion into a vector system, often loses some of its cell specificity and also because vector integration near active promoters and enhancers can activate the tissue-specific promoter and drive transgene expression. In contrast, because miRNA-mediated silencing occurs at the post-transcriptional level, promoter and enhancer trapping is irrelevant. As such, miRNA-regulation can be used to effectively de-target transgene expression from a particular cell type, while still allowing for broad tissue expression. miRNA regulation may also be used as in combination with tissue-specific promoter/enhancers. By including the miRNA target sequence in expression cassettes already under the control of a tissue-specific promoter, an additional layer of regulation is added which may eliminate off-target expression.

[0150] Exemplary miRNA target sequences which suppress transgene expression in hematopoietic-lineage cells and/or antigen-presenting cells, include, but are not limited to, miR-181, miR-142, miR-223, and miR-155 target sequences. Other miRNA target sequences which suppress transgene expression in hematopoietic-lineage cells and/or antigen-presenting cells are known in the art (see e.g. Ghafouri-Fard, S., et al., 2021. Non-coding RNA research, 6(1), pp. 8-14). miRNAs which are expressed in hematopoietic-lineage cells and/or antigen-presenting cells interact with the corresponding miRNA target sequence and reduce the expression of the target gene (see e.g. Brown, B. D., et al., 2006. Nature medicine, 12(5), pp. 585-591 and Brown, B. D., et al., 2007. Nature biotechnology, 25(12), pp. 1457-1467).

[0151] Further miRNA target sequences that suppress transgene expression in hematopoietic-lineage cells and/or antigen-presenting cells can be identified by any suitable method, for example miRNA expression analysis as described in Monticelli, S., et al., 2005. Genome biology, 6(8), pp. 1-15.

[0152] Suitably, the one or more miRNA target sequence comprise or consist of: (i) one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8) miR-142 target sequence; (ii) one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8) miR-181 target sequence; (iii) one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8) miR-223 target sequence; and/or (iv) one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8) miR-155 target sequence.

[0153] In some embodiments, the one or more miRNA target sequence comprise or consist of: (i) two or more miR-142 target sequences; (ii) two or more miR-181 target sequences; (iii) two or more miR-223 target sequences; and/or (iv) two or more miR-155 target sequences. In some embodiments, the one or more miRNA target sequence comprise or consist of: (i) at least four miR-142 target sequences; (ii) at least four miR-181 target sequences; (iii) at least four miR-223 target sequences; and/or (iv) at least four miR-155 target sequences. In some embodiments, the one or more miRNA target sequence comprise or consist of: (i) four miR-142 target sequences; (ii) four miR-181 target sequences; (iii) four miR-223 target sequences; and/or (iv) four miR-155 target sequences. Suitably, the target sequences are separated by spacer sequences.

[0154] In some embodiments, the one or more miRNA target sequence comprise or consist of one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8) miR-142 target sequence. In some embodiments, the one or more miRNA target sequence comprise or consist of two or more miR-142 target sequences. In some embodiments, the one or more miRNA target sequence comprise or consist of three or more miR-

142 target sequences. In some embodiments, the one or more miRNA target sequence comprise or consist of four or more miR-142 target sequences. In some embodiments, the one or more miRNA target sequence comprise or consist of four miR-142 target sequences. Suitably, the target sequences are separated by spacer sequences.

[0155] The miR-142 target sequence may be a human miRNA target sequence. Suitably, the miR-142 target sequence is a miR-142-5p or miR-142-3p miRNA target sequence. In some embodiments, the miR-142 target sequence is a miR-142-3p miRNA target sequence.

[0156] In some embodiments, the miR-142 target sequence comprises or consists of a nucleotide sequence which is at least 80% identical to SEQ ID NO: 17 or a fragment thereof. Suitably, the miR-142 target sequence comprises or consists of a nucleotide sequence which is at least 85%, at least 90%, or at least 95% identical to SEQ ID NO: 17 or a fragment thereof.

[0157] In some embodiments, the miR-142 target sequence comprises or consists of the nucleotide sequence SEQ ID NO: 17 or a fragment thereof.

Exemplary miR-142 target sequence  
(SEQ ID NO: 17)  
TCCATAAAGTAGGAAACACTACA

[0158] In some embodiments, the one or more miRNA target sequence comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 18 or a fragment thereof. Suitably, the one or more miRNA target sequence comprises or consists of a nucleotide sequence which is at least 75%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 18 or a fragment thereof.

[0159] In some embodiments, the one or more miRNA target sequence comprises or consists of the nucleotide sequence of SEQ ID NO: 18 or a fragment thereof.

Exemplary 4x miR-142 target sequence  
(SEQ ID NO: 18)  
TCCATAAAGTAGGAAACACTACACGATTCCATAAAGTAGGAAACACTAC  
AACCGGTTCCATAAAGTAGGAAACACTACATCACTCCATAAAGTAGGAA  
ACACTACA

#### BCKDC Subunit

[0160] The protein-coding sequence delivered by the lentiviral vector of the present invention is a nucleotide sequence encoding a branched-chain alpha-ketoacid dehydrogenase complex (BCKDC) subunit. Once delivered to a cell, the protein encoded by the protein-coding sequence may be expressed in the cell (i.e. "transgene expression" may occur).

[0161] As used herein, a "BCKDC subunit" may be a polypeptide corresponding to a subunit of a branched-chain alpha-ketoacid dehydrogenase complex (BCKDC), or a variant and/or fragment thereof. The branched-chain alpha-ketoacid dehydrogenase complex (BCKDC or BCKDH complex) is a multi-subunit complex of enzymes that is found on the mitochondrial inner membrane that catalyses the oxidative decarboxylation of branched, short-chain alpha-ketoacids.

[0162] The BCKD complex is composed of three catalytic components: alpha-ketoacid dehydrogenase (E1 component), dihydrolipoyl transacylase (E2 component), and dihydrolipoamide dehydrogenase (E3 component). E2 forms the core of the complex with either 24 copies in octahedral symmetry or 60 copies in icosahedral symmetry depending on the type and source of the complex. Multiple copies of E1, E3 are attached to the E2 core through non-covalent bonds (see e.g. Ævarsson, A., et al., 2000. Structure, 8(3), pp. 277-291). Human BCKDC also contains two regulatory enzymes: BCKD kinase (BDK) and BCKD phosphatase (BDP), these enzymes regulate activity of BCKDC through the phosphorylation/dephosphorylation of the E1a subunits (see e.g. Wynn, R. M., et al., 2004. Structure, 12(12), pp. 2185-2196; and Tso, S. C., et al., 2013. PNAS, 110(24), pp. 9728-9733).

[0163] Suitably, the BCKDC subunit is selected from BCKD E1 alpha subunit (BCKDE1A), or a fragment and/or variant thereof; BCKD E1 beta subunit (BCKDE1B), or a fragment and/or variant thereof; BCKD E2 subunit (DBT), or a fragment and/or variant thereof; and BCKD E3 subunit (DLD), or a fragment and/or variant thereof. In preferred embodiments, the BCKDC subunit is BCKD E2 subunit (DBT), or a fragment and/or variant thereof.

[0164] The protein-coding sequence may be codon-optimised. For example, the protein-coding sequence may be codon-optimised for expression in a mammalian (e.g. human) cell.

[0165] Different cells differ in their usage of particular codons. This codon bias corresponds to a bias in the relative abundance of particular tRNAs in the cell type. By altering the codons in the sequence so that they are tailored to match with the relative abundance of corresponding tRNAs, it is possible to increase expression. By the same token, it is possible to decrease expression by deliberately choosing codons for which the corresponding tRNAs are known to be rare in the particular cell type. Thus, an additional degree of translational control is available. Codon usage tables are known in the art for mammalian cells (e.g. humans), as well as for a variety of other organisms.

#### BCKDE1

[0166] In some embodiments, the BCKDC subunit is BCKDE1 or a fragment and/or variant thereof.

[0167] BCKD E1 (EC 1.2.4.4) is a branched-chain alpha-ketoacid decarboxylase that catalyses both the decarboxylation of the  $\alpha$ -ketoacid and the subsequent reductive acylation of the lipoyl moiety (another catalytic cofactor) that is covalently bound to E2. E1 is a heterotetramer consisting of two alpha subunits and two beta subunits. As used herein, "BCKDE1" may refer to either an E1 alpha subunit or an E1 beta subunit.

#### BCKDE1A

[0168] In some embodiments, the BCKDC subunit is BCKDE1A or a fragment and/or variant thereof.

[0169] "BCKDE1A" is the abbreviated name of the polypeptide encoded by the BCKDHA gene and is also known as branched-chain alpha-ketoacid dehydrogenase E1 component alpha chain and 2-oxoisovalerate dehydrogenase subunit alpha, mitochondrial. BCKDE1A may have 3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-

transferring) activity. BCKDE1A may bind thiamine diphosphate and/or stabilise the BCKDC complex.

[0170] A fragment and/or variant of BCKDE1A may retain BCKDE1A activity. For example, a fragment and/or variant of BCKDE1A may have methyl-2-oxobutanoate dehydrogenase activity, bind thiamine diphosphate and/or stabilise the BCKDC complex. Suitably, a fragment and/or variant of BCKDE1A may have the same or similar activity to BCKDE1A, e.g. may have at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 100% of the activity of BCKDE1A.

[0171] A person skilled in the art would be able to generate fragments and/or variants using conservative substitutions, based on the known structural and functional features of BCKDE1A (see e.g. Åvarsson, A., et al., 2000. Structure, 8(3), pp. 277-291), and/or based on known variants (see e.g. NCBI Gene ID: 593 and NCBI HomoloGene: 569). Suitably, a fragment of BCKDE1A and/or a BCKDE1A variant comprises a thiamine pyrophosphate binding domain.

[0172] The BCKDHA gene is conserved in human, chimpanzee, Rhesus monkey, dog, cow, mouse, and rat. The BCKDE1A may be a human BCKDE1A. Suitably, the BCKDE1A may comprise or consist of a polypeptide sequence of UniProtKB accession P12694, or a fragment and/or variant thereof.

[0173] In some embodiments, the BCKDE1A comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 37 or a fragment thereof. Suitably, the BCKDE1A comprises or consists of an amino acid sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 37 or a fragment thereof.

[0174] In some embodiments, the BCKDE1A comprises or consists of SEQ ID NO: 37 or a fragment thereof.

#### Exemplary BCKDE1A

(SEQ ID NO: 38)

```
MAVAIAAAARVWRLNRGLSQAAALLLRQPGARGLARSHPPRQQQFSSLDDKPFPGASAEFIDKLEFIQPN
VISGIPPIYRVMDRQGQIINPSEDPHLPKEKVLKLYKSMTLLNTMDRILY
ESQRQGRISFYMTNYGEEGTHVGSAALNDNTDLVFGQYREAGVLMYRDY
PLELFMAQCYGNISDLGKGRQMPVHYGCKERHFVTISSPLATOQIPQAVG
AYAAAKRANANRVVICYFGEAASEGDAHAGFNFAATLECPIIFFCRNN
GYAISTPTSEQYRGDGIAARGPGYGMISRVDGNDVFAVYNATKEARRR
AVAENQPFLIEAMTYSSSPILPPDPHSREPTLTWGPLPLCRIGHHSTSD
DSSAYRSVDEVNYWDKQDHPISLRHYLLSQGWWDEEQEKAWRKQSRK
VMEAFEQAERKPKNPNLLFSDVYQEMPAQLRKQQESLARHLQTYGEHYPLD
PLDHFDK
```

[0178] Suitably, the nucleotide sequence encoding BCKDE1A, or a fragment and/or variant thereof, may comprise or consist of a nucleotide sequence of NCBI reference sequence NM\_000709 or NM\_001164783, or a fragment and/or variant thereof.

[0179] In some embodiments, the nucleotide sequence encoding BCKDE1A, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 39 or a fragment thereof. Suitably, the nucleotide sequence encoding BCKDE1A, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 39 or a fragment thereof.

[0180] In some embodiments, the nucleotide sequence encoding BCKDE1A, or a fragment and/or variant thereof, comprises or consists of the nucleotide sequence SEQ ID NO: 39 or a fragment thereof.

#### Exemplary nucleotide encoding BCKDE1A

(SEQ ID NO: 39)

```
ATGGCGGTAGCGATCGCTGCAGCGAGGGCTGGCGCTAACCGTGGTTGAGCCAGGCTGCCCTCCTG
CTGCTGCGGCAGCCTGGGCTCGGGACTGGCTAGATCTCACCCCCCAGGCAGCAGCAGCTTTCAG
TCTCTGGATGACAAGCCCCAGTCCCAAGGGCTCGGCGAGTTATAGATAAGTTGAAATCATCCAG
CCCAACGTCATCTCTGGAATCCCCATCTACCGCGTATGGACCGGAAGGCCAGATCATCAACCCCAGC
```

- continued

```

GAGGACCCCCACCTGCCGAAGGAGAAGGTGCTGAAGCTCTACAAGAGCATGACACTGCTTAACACCATG
GACCGCATTCTATGAGTCTAGCGGCAGGGCCGGATCTCCTTCTACATGACCAACTATGGTGAGGAG
GGCACGCACGTGGGGAGTGCCGCCGCTGGACAACACGGACCTGGTGTGCCCCAGTACCGGGAGGCA
GGTGTGCTGATGTATCGGACTACCCCTGAACTATTCTGGCCAGTGTCTATGGCAACATCAGTGAC
TTGGCAAGGGCGCCAGATGCCCTGTCACACTACGGCTGCAAGGAACGCCACTCGTCACTATCTCCT
CACTGGCCACGCCAGATCCCTCAGGGGTGGGGCGCGTACGCAGCCAAGCGGGCAATGCCAACAGG
GTCGTCATCTGTTACTTCGGCGAGGGGGCAGCCAGTGAGGGGACGCCATGCCGCTTCAACTCGCT
GCCACACTTGAGTGCCCCATCATCTTCTGCCGAACAATGGCTACGCCATCTCACGCCACCTCT
GAGCAGTATCGGGCGATGGCATTCGAGCAGGCCCCGGGTATGGCATATGTCATCCGCGTGGAT
GGTAATGATGTGTTGCGTATAAACGCCACAAAGGAGGCCAGGGCTGTGGCAGAGAACCCAG
CCCTCCTCATCGAGGCATGACCTACAGGATCGGGCACCAAGCACCAGTGACGACAGTCAGCGTAC
CGCTCGGTGGATGAGGTCATTACTGGGATAAACAGGACCACCCATCTCCGGCTGCGGACTATCTG
CTGAGCCAAGGCTGGGGATGAGGAGCAGGAGAAGGCCCTGGAGGAAGCAGTCCGCAGGAAGGTGATG
GAGGCCTTGAAGCGAGGCCAGGCCAAGCCCACCCAAACCTACTCTTCAGACGTGTATCAG
GAGATGCCGCCAGCTCCGAAGCAGCAGGAGTCTGGCCCGACCTGCAGACCTACGGGGAGCAC
TACCCACTGGATCACTCGATAAGTGA

```

#### BCKDE1B

**[0181]** In some embodiments, the BCKDC subunit is BCKDE1B or a fragment and/or variant thereof.

**[0182]** “BCKDE1B” is the abbreviated name of the polypeptide encoded by the BCKDHB gene and is also known as branched chain ketoacid dehydrogenase E1 subunit beta and 2-oxoisovalerate dehydrogenase subunit beta, mitochondrial. BCKDE1B may have 3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring) activity.

**[0183]** A fragment and/or variant of BCKDE1B may retain BCKDE1B activity. For example, a fragment and/or variant of BCKDE1B may have 3-methyl-2-oxobutanoate dehydrogenase activity. Suitably, a fragment and/or variant of BCKDE1B may have the same or similar activity to BCKDE1B, e.g. may have at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 100% of the activity of BCKDE1B.

**[0184]** A person skilled in the art would be able to generate fragments and/or variants using conservative substitutions, based on the known structural and functional features of BCKDE1B (see e.g. Åvarsson, A., et al., 2000. Structure, 8(3), pp. 277-291), and/or based on known variants (see e.g. NCBI Gene ID: 594 and NCBI HomoloGene: 39). Suitably, a fragment of BCKDE1B and/or a BCKDE1B variant comprises a transketolase, C-terminal domain and a pyrimidine-binding domain (of TPP-dependent enzymes). The domains may be linked by inter-domain linker(s).

**[0185]** The BCKDHB gene is conserved in human, chimpanzee, dog, cow, mouse, and rat. The BCKDE1B may be a human BCKDE1B. Suitably, the BCKDE1B may comprise or consist of a polypeptide sequence of UniProtKB accession P21953, or a fragment and/or variant thereof.

**[0186]** In some embodiments, the BCKDE1B comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 40 or a fragment thereof. Suitably, the BCKDE1B comprises or consists of an amino acid sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 40 or a fragment thereof.

**[0187]** In some embodiments, the BCKDE1B comprises or consists of SEQ ID NO: 40 or a fragment thereof.

#### Exemplary BCKDE1B

(SEQ ID NO: 40)

```

MAVVAAGWLLRLRAAGAEGHWRRLPGAGLARGFLHPAATVEDAAQRR
QVAHFTFQDPPEPREYQTKMNLFQSNTSALDNLSAKDPTAVIFGEDDV
AFGGVFRCVGLRDKYKDRVFNTPLCEQGIVFGIGIAVTGATAIAEI
QFADYIFFPAFDQIVNEAAKYRYRSQDLFNCGSLTIRSPWGVGHGALYH
SQSPEAFFAHCPGIKVVIPRSPFQAKGLLSCIEDKNPCIFFEPKILYR
AAAEVEPIEPYNIPLSQAEVIQEGSDVTLVAWGTQVHVIREVASMAKEK
LGVSCEDVIDLRTIIPWDVDTICKSVIKTGRLLISHEAPLTGGFASEISS
TVQEECFLNLEAPISRVCYDTPFPFHIFEPFYIPDKWKCYDALRKMINY

```

**[0188]** Suitably, a BCKDE1B variant may comprise one or more variation selected from: A6V, G9S, A29V, T411, K211N. These are considered to be benign (or likely benign) variations based on clinical data.

**[0189]** In some embodiments, the BCKDE1B comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 41 or a fragment thereof. Suitably,

the BCKDE1B comprises or consists of an amino acid sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 41 or a fragment thereof.

**[0190]** In some embodiments, the BCKDE1B comprises or consists of SEQ ID NO: 41 or a fragment thereof.

least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 42 or a fragment thereof.

**[0193]** In some embodiments, the nucleotide sequence encoding BCKDE1B, or a fragment and/or variant thereof, comprises or consists of the nucleotide sequence SEQ ID NO: 42 or a fragment thereof.

#### Exemplary nucleotide encoding BCKDE1B

(SEQ ID NO: 42)

```

ATGGCGGTGTTAGCGCGCTGCCGCTGGTACTCAGGCTCAGGGCGGCAGGGCTGAGGGGACTGG
CGTCGGCTTCCTGGCGGGGCTGGCGGGGCTTTGCACCCCGCCGACTGTCGAGGATGCGGCC
CAGAGGCGCAGGTGGCTATTTACTTCCAGCCAGATCCGGAGCCCCGGAGTACGGCAAACCTAG
AAAATGAATCTTCCAGTCTGTAACAAGTGCCTGGATAACTCATGGCAAAGATCCTACTGCAGTA
ATATTGGTGAAGATGTTGCCTTGGTAGTCTAGTCAGTGGCTTGCGAGACAAATATGGA
AAAGATAGAGTTTAATACCCATTGCTGAACAAGGAATTGGATTGGAATCGAATTGCGTC
ACTGGAGCTACTGCCATTGGAAATTCACTGTTGAGATTATTTCCCTGCATTGATCAGATTGTT
AATGAAGCTGCCAAGTATCGCTATCGCTCTGGGATCTTTAACGTGGAAGCCTACTATCCGGCC
CCTGGGGCTGTTGGCATGGGCTCTATCATTCTCAGAGTCTGAAGCATTGGCCATTG
CCAGGAATCAAGGTGGTTACCCAGAACGCCCTTCCAGGCCAAGGACTCTTTGTCATGCATAGAG
GATAAAAATCCTGTATATTTGAACTAAATACCTTACAGGGCAGCAGCGGAAGAAGTCCCTATA
GAACCATAAACATCCCAGTCCCAGGCCAGTCATACAGGAAGGGAGTGTACTCTAGTTGCC
TGGGCACTCAGGTTCATGTGATCCAGAGGTAGCTTCCATGGCAAAGAAAAGCTTGGAGTGTCTTGT
GAAGTCATTGATCAGGACTATAACCTTGGATGTGGACACAATTGTAAGTCTGTGATCAAACA
GGGCACTGCTAATCAGTCAGCAGGGCTCCCTGACAGCGGGCTTGCATCGAAATCAGCTCTACAGTT
CAGGAGGAATGTTCTGAA CCTAGAGGCTCTATCAAGAGTATGTGGTATGACACACCATTCC
CACATTGGAAACCATTCTACATCCCAGACAAATGGAAGTGTATGATGCCCTCGAAAAATGATCAAC
TATTGA

```

#### Exemplary BCKDE1B

(SEQ ID NO: 41)

```

MAVVAAGWLLRLRAAGAEGHWRRLPAGLARGFLHPAATVEDAQRR
QVAHFTFQPDPEPREYQQTQMNLFQSNTSALDNLSAKDPTAVIFGEDV
AFGGVFRCVGLRDKYGDRLVFNPLCEQGIVFGIGIAVTGATAIEI
QFADYIPFAFDQIVNEAKYRYRSQDLFNCGSLTIRSPWGCVGHGALYH
SQSPEAFFAHCPGIKIKVISLS

```

**[0191]** Suitably, the nucleotide sequence encoding BCKDE1B, or a fragment and/or variant thereof, may comprise or consist of a nucleotide sequence of NCBI reference sequence NM\_000056, NM\_001318975, or NM\_183050.4, or a fragment and/or variant thereof.

**[0192]** In some embodiments, the nucleotide sequence encoding BCKDE1B, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 42 or a fragment thereof. Suitably, the nucleotide sequence encoding BCKDE1B, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at

#### DBT

**[0194]** In some embodiments, the BCKDC subunit is DBT or a fragment and/or variant thereof.

**[0195]** “DBT” is the abbreviated name of the polypeptide encoded by DBT and is also known as Dihydrolipoamide Branched Chain Transacylase E2 or lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial. Within the BCKD complex, the catalytic function of DBT is to accept, and to transfer to coenzyme A, acyl groups that are generated by the branched-chain alpha-keto acid decarboxylase component. DBT also forms the core of the BCKD complex.

**[0196]** A fragment and/or variant of DBT may retain DBT activity (see EC 2.3.1.168). For example, a fragment and/or variant of DBT may be able to accept and/or transfer acyl groups to coenzyme A, and/or form the core of a BCKD complex. Suitably, a fragment and/or variant of DBT may have the same or similar activity to DBT, e.g. may have at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 100% of the activity of DBT. Assays to determine BCKDC activity are described, for instance in Wynn, R. M., et al., 2004. Structure, 12(12), pp. 2185-2196. For example, BCKDC activity

may be assayed spectrophotchemically, as described in Chuang, J. L., et al., 2000. Methods in enzymology, 324, pp. 192-200.

**[0197]** A “fragment of DBT” may refer to a portion or region of a full-length DBT that has the same or similar activity as a full-length DBT (e.g. acyl-group transferring and/or core-forming activity), i.e. the fragment may be a functional fragment. The fragment may have at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 100% of the activity of a full-length DBT.

**[0198]** A “DBT variant” may include an amino acid sequence or a nucleotide sequence which may be at least 50%, at least 55%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85% or at least 90% identical, optionally at least 95% or at least 97% or at least 99% identical to a wild-type DBT. DBT variants may have the same or similar activity to a wild-type DBT (e.g. acyl-group transferring and/or core-forming activity). A DBT variant may have at least at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 100% of the activity of a wild-type DBT.

**[0199]** A person skilled in the art would be able to generate a fragment of DBT and/or DBT variants using conservative substitutions, based on the known structural and functional features of DBT (see e.g. Chang, C. F., et al., 2002. Journal of Biological Chemistry, 277(18), pp. 15865-15873; Kato, M., et al., 2006. The EMBO journal, 25(24), pp. 5983-5994; and Brautigam, C. A., et al., 2011. Journal of Biological Chemistry, 286(26), pp. 23476-23488) and/or based on known variants (see e.g. Tsuruta, M., et al., 1998. Journal of human genetics, 43(2), pp. 91-100; NCBI Gene ID: 1629 and NCBI HomoloGene: 1444).

**[0200]** Suitably, a fragment of DBT and/or a DBT variant comprises an E3 binding domain, a 2-oxoacid dehydrogenase acyltransferase domain, and a lipoyl binding domain. The domains may be linked by inter-domain linker(s).

**[0201]** The DBT gene is conserved in chimpanzee, Rhesus monkey, dog, cow, mouse, and rat. The DBT may be a human DBT. Suitably, the DBT may comprise or consist of a polypeptide sequence of UniProtKB accession P11182, or a fragment and/or variant thereof.

**[0202]** In some embodiments, the DBT comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 43 or a fragment thereof. Suitably, the DBT comprises or consists of an amino acid sequence which is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 43 or a fragment thereof.

**[0203]** In some embodiments, the DBT comprises or consists of SEQ ID NO: 43 or a fragment thereof.

#### Exemplary DBT

(SEQ ID NO: 43)

```
MAAVRMLRTWSRNAGKLICVRYFQTCGNVHVLKPNEYCFFGYPSFKYSH
PHHFLKTTAALRGQQVQFKLSDIGEGIREVTVKEWYVKEGDTVSQFDSTI
CEVQSDKASVTITSRYDGVIKLYNLDIAYVGKPLVDIETEALKDSE
EDVVETPAVSHDEHTHQEIKGRTLATPAVRRLAMENNIKLSEVVGSKG
DGRILKEDILNYLEKQTGAILPPSPKVEIMPPPCKDMTPVILVSKPP
VFTGKDKEPIKGQFQKAMVKTMSAALKIPHFGYCDEIDLTELVKLREEL
```

-continued

```
KPIAFARGIKLSFMPFFLKAASLGLLQFPILNASVDENCQNITYKASHN
IGIAMDTEQGLIVPNVKNVQICSIFDIATELNRLQKLGSVGQLSTTDLT
GGTFTLSNIGSIGGTFAKPVIMPPEVAIGALGSIKAIPRFNQKGEVYKA
QIMNVWSADHRVIDGATMSRFNSNLWKSYLENPAPMFLDLK
```

**[0204]** Suitably, a DBT variant may comprise one or more variation selected from A3T, R5P, M6I, M6V, M6L, T9S, T9N, W10C, W10L, W10G, R12S, R12G, N13K, N13D, G15R, R21H, R21C, Y22C, C26R, G27V, G27D, V29A, V31A, P34S, Y36F, Y36D, V37A, V37M, C38Y, F39L, G41V, G41R, F45L, K46R, H49N, P50R, H51R, H51Y, F53L, L60F, L60I, R61H, R61C, R61S, R77S, R77K, T80A, V81L, Y122H, N124K, N124D, Y130C, A142T, E148D, V151I, T153N, P154S, A155V, A155S, V156L, S157P, H158R, H158P, D159N, E160G, E160A, H161R, T162A, H163Y, R169Q, T171I, M182V, N185H, K211R, K211E, K211Q, T213I, G214A, I216M, V223A, V223D, V223I, I225S, I225L, P227S, D234E, D234G, T236A, V237L, V237I, P238T, I239V, I239L, L240R, L240Q, V241A, S242P, K243N, K243E, P245L, P245R, P245T, F247L, K252Q, Q260L, A269V, I273V, E286K, K295Q, I297M, F299S, A300V, A315V, Q321H, Q321R, N335K, I346V, E351Q, I364L, C365S, I367V, S382F, V383M, G384N, G384S, S387R, T389S, V412L, P415L, P415A, I426V, I429L, K435R, E437D, E437Q, E437K, V438I, Y439H, K440E, I443V, M444I, N445S, M460V, L466V, A475D, A475S, F476C, D480E, and D480N. These are considered to be tolerated, benign, and/or likely benign variations as predicted by SIFT, PolyPhen, CADD, REVEL, and MetaLR.

**[0205]** Suitably, a DBT variant may comprise one or more variation selected from: N13D, R21H, C26R, R169Q, G214del, I239V, S242P, P245T, and G384S. In some embodiments, DBT variant may comprise G384S. These are considered to be benign (or likely benign) variations based on clinical data.

**[0206]** Suitably, the nucleotide sequence encoding DBT, or a fragment and/or variant thereof, may comprise or consist of a nucleotide sequence of NCBI reference sequence NM\_001918, NM\_001399969, or NM\_001399972, or a fragment and/or variant thereof.

**[0207]** In some embodiments of the invention, the nucleotide sequence encoding DBT, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 44 or a fragment thereof. Suitably, the nucleotide sequence encoding DBT, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97% at least 98% or at least 99% identical to SEQ ID NO: 44 or a fragment thereof.

**[0208]** In some embodiments of the invention, the nucleotide sequence encoding DBT, or a fragment and/or variant thereof, comprises or consists of the nucleotide sequence SEQ ID NO: 44 or a fragment thereof.

## Exemplary nucleotide encoding DBT

(SEQ ID NO: 44)

```

ATGGCTGCAGTCGTGAGAACCTGGAGCAGGAATGCGGGAAAGCTGATTGTTGCTATTCTT
CAAACATGTGGTAATGTTCATGTTGAAAGCCAAATTATGTTGTTCTTGTTATCCTCATTCAAG
TATAGTCATCCACATCACTTCCTGAAAACAACGTGCTCTCCGTGGACAGGTTGTTCAAGCTC
TCAGACATTGGAGAAGGGATTAGAGAAGTAACGTGTTAAAGAATGGTATGTAAGAAGGGAGATA
CTCAGTTGATAGCATCTGTGAAGTTCAAAGTGATAAAAGCTCTGTTACCATCACTAGTCGTTATG
GGAGTCATAAAAAAACTCTATTATAATCTAGACGATATTGCCATGTGGGAAGCCATTAGTAGACATA
GAAACGGAAGCTTAAAGATTCAAGAAGAAGATGTTGAAACTCCTGCAGTGTCTCATGATGAACAT
ACACACCAAGAGAGATAAAGGGCCGAAAAACACTGGCAACTCCTGCAGTCGCCGTCTGGCAATGGAA
AATATTAAGCTGAGTGAAGTTGGCTCAGGAAAGATGGCAGAATACTTAAAGAAGATATCCTCAAC
TATTGGAAAAGCAGACAGGGACTATTCGCTTCCCTCACCCAAAGTTGAAATTATGCCACCTCCACCA
AAGCCAAAAGACATGACTGTTCTATACTAGTATCAAACCTCCGTATTCAAGGCAAAGACAAAACA
GAACCCATAAAGGCTTCAAAAGCAATGGTCAAGACTATGCTGCAGCCCTGAAGATACTTCATT
GGTTATTGTGATGAGATTGACCTTACTGAACGGTTAAGCTCCGAGAAGAATTAAACCCATTGCAATT
GCTCGTGGAAATTAAACTCTCCTTATGCCCTTCTTAAAGGCTGTTCTGGGATTACTACAGTTT
CCTATCCTTAACGCTCTGTGGATGAAAAGCCAGAATAACATATAAGGCTCTCATAACATTGGG
ATAGCAATGGGAACTTGAGCAGGGTTTGATTGTCCTAATGTGAAAATGTCAGATCTGCTCTATATT
GACATGCCACTGAACCGCCTCCAGAAATTGGCTCTGTGGTCAGCTCAGCACCACGTGATCTT
ACAGGAGGAACATTACTCTTCAACATTGGATCAATTGGGTACCTTGCACAAACCAGTGATAATG
CCACCTGAAGTAGCCATTGGAGCCCTGGATCAATTAGGCCATTCCCCGATTTAACCAAGGAGAA
GTATATAAGGCACAGATAATGAATGTGAGCTGGTCAGCTGATCACAGAGTTGATGGTGTACAATG
TCACGCTCTCCAATTGTGAAATCCTATTAGAAAACCCAGCTTTATGCTACTAGATCTGAAATGA

```

**[0209]** In some embodiments, the nucleotide sequence encoding DBT, or a fragment and/or variant thereof, is codon optimised. An exemplary codon-optimised sequence is provided in SEQ ID NO: 45.

**[0210]** In some embodiments, the nucleotide sequence encoding DBT, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 45 or a fragment thereof. Suitably, the nucleotide sequence encoding DBT, or a frag-

ment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 75%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 45 or a fragment thereof.

**[0211]** In some embodiments, the nucleotide sequence encoding DBT, or a fragment and/or variant thereof, comprises or consists of the nucleotide sequence SEQ ID NO: 45 or a fragment thereof.

## Exemplary codon-optimised nucleotide encoding DBT

(SEQ ID NO: 45)

```

ATGGCAGCCGTGAGGATGCTGAGGACCTGGAGCCGAACGCAGGAAGCTGATCTGCGTGAGATACTTT
CAGACATGTGGCAACGTGACGTGCTGAAGCCAAATTACGTGTCGCTCTCGGCTACCCCTCCCTCAAG
TATTCTCACCTCACCACTTCTGAAGACCAACAGCCGCTGAGGGACAGGTGGTCAGTCAAGCTG
AGCGACATCGCGAGGGCATCCGCGAGGTGACCGTGAAGGAGTGTTACGTGAAGGAGGGCGACACAGTG
AGCCAGTTGATTCCATCTGTGAGGTGACGTCTGACAAGGCCAGCGTACCATCACATCCGGTACGAT
GGCGTGTCAAGAAGCTGTACTATAACCTGGACGACATGCCATGTGGCAAGCCACTGGTGACATC
GAGACAGAGGCCCTGAAGGACAGCGAGGAGATGTTGGAGACACCCGCGTGTCCCACGATGAGCAC
ACACACCAGGAGATCAAGGAAAGGAAGCACCCCTGGCACACCAGCCGTGCGGAGACTGGCATGGAGAAC
AATATCAAGCTGAGCGAGGTGGTGGGATCCGGCAAGGACGGCAGAATCCTGAAGGAGGACATCCTGAAC

```

- continued

```

TATCTGGAGAACGAGACCGGCGCAATCCTGCCACCTCCCCTAAGGTGGAGATCATGCCACCCCTCCA
AAGCCTAAGGACATGACAGTGCCAATCCTGGTGTCAAGCCCCCGTGGTACCGGAAGGATAAGACA
GAGCCTATCAAGGGCTTCAGAAGGCCATGGTAAGACCATGAGCGCCGCCCTGAAGATCCCACACTTC
GGCTACTCGGAGAGATCGATCTGACAGAGCTGGTAAGCTGGGGAGGAGCTGAAGCCAATCGCCTC
GCCAGAGGCATCAAGCTGTCCTTATGCCCTTCTCTGAAGGCCCTCTCTGGGCTGCTGCAGTTT
CCTATCCTGAACGCCCTGTGGACGAGAACTGCCAGAATATCACCTATAAGGCCAGCCACAATATCGGC
ATCGCCATGGATACAGAGCAGGGCTGATCGTCCAAACGTGAAGAAATGTGCAGATCTGTTCCATCTTC
GACATGCCACCGAGCTGAACAGGCTGCAGAAGCTGGGCTCTGTGGGCCAGCTGAGCACACAGATCTG
ACCGGCCACCTTCACACTGTCCAATATCGGCTCTATCGGGCACATTGCCAAGGCCGTGATCATG
CCACCAGAGGTGGCAATCGGCCCTGGGCTCTATCAAGGCCATCCCTCGCTCAACCAGAAGGGCGAG
GTGTACAAGGCCAGATCATGAATGTGAGCTGGTCCGCCGACCACAGAGTATCGATGGGCCACCATG
TCTCGCTTCAGCAACCTGTGGAAGTCCTATCTGGAGAATCCGCCCTTATGCTGCTGGATCTGAAGTG

```

#### DLD

**[0212]** In some embodiments, the BCKDC subunit is DLD or a fragment and/or variant thereof.

**[0213]** “DLD” is the abbreviated name of the polypeptide encoded by the DLD gene and is also known as dihydro-lipoamide dehydrogenase or the E3 Component of Pyruvate Dehydrogenase Complex. DLD is a component of the glycine cleavage system as well as an E3 component of three alpha-ketoacid dehydrogenase complexes.

**[0214]** A fragment and/or variant of DLD may retain DLD activity (see e.g. EC 1.8.1.4). For example, a fragment and/or variant of DLD may have dihydrolipoyl dehydrogenase activity. Suitably, DLD may have the same or similar activity to DLD, e.g. may have at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 100% of the activity of DLD.

**[0215]** A person skilled in the art would be able to generate fragments and/or variants using conservative substitutions, based on the known structural and functional features of DLD (see e.g. Brautigam, C. A., et al., 2005. Journal of molecular biology, 350(3), pp. 543-552), and/or based on known variants (see e.g. NCBI Gene ID: 1738 and NCBI HomoloGene: 84). Suitably, a fragment of DLD and/or a DLD variant comprises a NAD(P)-binding Rossmann-like domain, a pyridine nucleotide-disulphide oxidoreductase domain, and a pyridine nucleotide-disulphide oxidoreductase dimerisation domain. The domains may be linked by inter-domain linker(s).

**[0216]** The DLD gene is conserved in chimpanzee, Rhesus monkey, dog, cow, mouse, and rat. The DLD may be a human DLD. Suitably, the DLD may comprise or consist of a polypeptide sequence of UniProtKB accession P09622, or a fragment and/or variant thereof.

**[0217]** In some embodiments, the DLD comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 46 or a fragment thereof. Suitably, the DLD comprises or consists of an amino acid sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 46 or a fragment thereof.

**[0218]** In some embodiments, the DLD comprises or consists of SEQ ID NO: 46 or a fragment thereof.

#### Exemplary DLD

(SEQ ID NO: 46)

```

MQSWSRVYCSLAKRGHFNRISSHGLQQLSAVPLRTYADQPIDADVTIVGS
GPGGYVAAIKAAQLGFKTVCIKENETLGGTCLNVGCIPS KALLNNNSHY
HMAHGKDFASRGIEJMSEVRNLNDKMMEQKSTAVKALTGGIAHLFKQNKV
VHVNNGYKITGKNQVTATKADGGTQVIDTKNIIATGSEVTPFPGITID
EDTIVSSTGALSLKKVPEKMVVIGAGVIGVELGSVWQRLGADVTAVEFL
GHVGGVGIDMEISKNFQRLQKQGFKPFLNTKVTGATKSDGKIDVSIE
AASGGKAEVITCDVLLVCIGRRPFTKNLGLEELGIELDPRGRIPVNTRF
QTKIPNIYAIGDVVAGPMLAHKAEGDEGIICVEGMAGGAHVIDYNCPSV
IYTHPEVAVGKSSEQLKEEGIEYKVGKFPAANSRAKTNADTDGMVKI
LGQKSTDRLVLAHILGPAGEMVNNEAALALEYGASCEDIARVCHAHPTL
SEAFREANLAASFHKINF

```

**[0219]** Suitably, a DLD variant may comprise one or more variation selected from: A12T, G170S, T276S, and D290G. These are considered to be benign (or likely benign) variations based on clinical data.

**[0220]** In some embodiments, the DLD comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 47 or a fragment thereof. Suitably, the DLD comprises or consists of an amino acid sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 47 or a fragment thereof.

**[0221]** In some embodiments, the DLD comprises or consists of SEQ ID NO: 47 or a fragment thereof.

#### Exemplary DLD

(SEQ ID NO: 47)

```

MAHGKDFASRGIEJMSEVRNLNDKMMEQKSTAVKALTGGIAHLFKQNKV
VHVNNGYKITGKNQVTATKADGGTQVIDTKNIIATGSEVTPFPGITID
DTIVSSTGALSLKKVPEKMVVIGAGVIGVELGSVWQRLGADVTAVEFLG

```

- continued

```
HVGGVGIDMEISKNFQRILQKQGFKFKLNTKVGTGATKKSDGKIDVSIEA
ASGGKAEVITCDVLLVCIGRRPFTKNLGLEELGIELDPRGRIPVNTRFQ
TKIPNIYAI GDVVAGPMLAHKAEDEGIICVEGMAGGAVIDYNCPVSI
YTHPEVAWVGKSEEQLKEEGIEYKVGKFPFAANSRAKTNADTDGMVKIL
GQKSTDRVLAHILGPAGEMVNEAALEYGASCEDIARVCHAHPTLS
EAFREANLAASFPGKSINF
```

**[0222]** In some embodiments, the DLD comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 48 or a fragment thereof. Suitably, the DLD comprises or consists of an amino acid sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 48 or a fragment thereof.

**[0223]** In some embodiments, the DLD comprises or consists of SEQ ID NO: 48 or a fragment thereof.

#### Exemplary DLD

(SEQ ID NO: 48)

```
MQSWSRVYCSLAKRGHFNRI SHGLQGLSAVPLRTYADQPIDADVTIVGS
GPGGYVAAIKAAQLGFKTVCI EKNETLGGTCLNVGCIPS KALLNNSHYY
HMAHGKDFASRGIEMSEVRNLNDKMMEQKSTAVKALTGGIAHLFKQNKI
```

- continued

```
DEDTIVSSTGALSLKKVPEKMVVIGAGVIGVELGSVWQRLGADVTAVEF
LGHVGGVGIDMEISKNFQRILQKQGFKFKLNTKVGTGATKKSDGKIDVSIE
EAASGGKAEVITCDVLLVCIGRRPFTKNLGLEELGIELDPRGRIPVNTRFQ
FQTKIPNIYAI GDVVAGPMLAHKAEDEGIICVEGMAGGAVIDYNCPVSI
VIYTHPEVAWVGKSEEQLKEEGIEYKVGKFPFAANSRAKTNADTDGMVKIL
ILGQKSTDRVLAHILGPAGEMVNEAALEYGASCEDIARVCHAHPTLS
LSEAFREANLAASFPGKSINF
```

**[0224]** Suitably, the nucleotide sequence encoding DLD, or a fragment and/or variant thereof, may comprise or consist of a nucleotide sequence of NCBI reference sequence NM\_000108, NM\_001289750, NM\_001289751 or NM\_001289752, or a fragment and/or variant thereof.

**[0225]** In some embodiments, the nucleotide sequence encoding DLD, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 49 or a fragment thereof. Suitably, the nucleotide sequence encoding DLD, or a fragment and/or variant thereof, comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to SEQ ID NO: 49 or a fragment thereof.

**[0226]** In some embodiments, the nucleotide sequence encoding DLD, or a fragment and/or variant thereof, comprises or consists of the nucleotide sequence SEQ ID NO: 49 or a fragment thereof.

#### Exemplary nucleotide encoding DLD

(SEQ ID NO: 49)

```
ATGCAGAGCTGGAGTCGTGTACTGCTCCTGGCCAAGAGAGGCCATTCAATCGAACATCTCATGGC
CTACAGGGACTTTCTGCAGTGCCTCTGAGAACTTACGCAGATCAGCCGATTGATGCTGATGTAACAGTT
ATAGGTTCTGGCTCTGGAGGATATGGCTGCTATTAAAGCTGCCAGTTAGGCTCAAGACAGTCTGC
ATTGAGAAAAATGAAACACTTGGGAAACATGCTTGATGTTGGTTGATTCTCTAAGGTTTATTG
AACAACTCTCATTATTACCATATGCCCATGGAAAAGATTGCTAGAGGAATTGAAATGTCGAA
GTTCTGAATTAGACAAGATGATGGAGCAGAAGAGTACTGCAGTAAAGCTTAAACAGGTGGAATT
GCCCACTTATTCAAACAGAATAAGGTTGTTCATGTCATGGATATGGAAAGATAACTGGCAAAATCAA
GTCACTGCTACGAAGACTGATGGCGCACTCAGGTATTGATACAAAGAACATTCTATAGCCACGGGT
TCAGAAGTTACTCCTTTCTGGAATCACGATAGATGAAGATAACATAGTGTCTACAGGTGCTTA
TCTTAAAAAAAGTCCAGAAAAGATGGTTTATTGGTCAGGGATAATGGTGTAGATTGGTTCA
GTTTGGCAAAGACTTGGCAGATGTGACAGCAGTTGAATTAGTCATGTTAGGTGGAGTTGGAATT
GATATGGAGATATCTAAACCTCAACGCATCCTCAAAACAGGGTTAAATTAAATTGAATACA
AAGGTTACTGGCTACCAAGAAGTCAGATGGAAAATTGATGTTCTATTGAAGCTGCTCTGGTGGT
AAAGCTGAAGTTACTGTGATGTTACTCTGGCTTGCATTGGCGACGACCCTTACTAAGAATTG
GGACTAGAAGAGCTGGAAATTGAACTAGATCCCAGAGGTAGAATTCCAGTCAATACCAAGATTC
AAAATCCAATATCTATGCCATTGGTGTAGTTGCTGGCCAATGCTGGCTCACAAAGCAGAGGAT
GAAGGCAATTCTGTGTTGAAGGAATGGCTGGTGTGTCACATTGACTACAATTGTGCTCATCA
GTGATTACACACACCCCTGAAGTTGCTGGTTGGCAAATCAGAAGAGCAGTTGAAAGAAGAGGGTATT
GAGTACAAAGTTGGAAATTCCATTGCTGCTAACAGCAGAGCTAACAGACAATGCTGACACAGATGGC
```

- continued

```
ATGGTGAAGATCTTGGCAGAAATCGACAGACAGACTGGGAGCACATATTCTGGACCAGGTGCT
GGAGAAAATGGTAAATGAAGCTCTTGCTTTGAAATATGGAGCATCCTGTGAAGATATAGCTAGAGTC
TGTCAATGCACATCCGACCTTATCAGAAGCTTTAGAGAAGCAAATCTGCTGCGTCATTGGCAAATCA
ATCAACTTTGA
```

#### Regulatory Elements

**[0227]** The lentiviral vector of the present invention may further comprise one or more regulatory elements which may act pre- or post-transcriptionally. Suitably, the protein-coding sequence is operably linked to one or more regulatory elements which may act pre- or post-transcriptionally. The one or more regulatory elements may facilitate expression of the transgene in liver cells (e.g. hepatocytes).

**[0228]** As used herein, a “regulatory element” may refer any nucleotide sequence that facilitates expression of a polypeptide, e.g. acts to increase expression of a transcript or to enhance mRNA stability. Suitable regulatory elements include for example promoters, enhancer elements, post-transcriptional regulatory elements, polyadenylation sites, and Kozak sequences.

#### Promoter

**[0229]** The lentiviral vector of the present invention may comprise a promoter, preferably a liver-specific (e.g. hepatocyte-specific) promoter. Suitably, the protein-coding sequence is operably linked to a promoter, preferably a liver-specific (e.g. hepatocyte-specific) promoter.

**[0230]** A “promoter” may refer to a region of DNA that leads to initiation of transcription of a gene. Promoters are located near the transcription start sites of genes, upstream on the DNA (towards the 5' region of the sense strand).

**[0231]** As used herein, a “tissue-specific promoter” may refer to a promoter which preferentially facilitates expression of a transgene in a specific type of cells or tissue. Suitably, a tissue-specific promoter may facilitate higher expression of a transgene in one cell-type as compared to other cell-types. Higher expression may be measured for example by measuring the expression of a transgene, e.g. green fluorescence protein (GFP), operably linked to the promoter, wherein expression of the transgene correlates with the ability of the promoter to facilitate expression of a gene. For example, a tissue-specific promoter may be a promoter which facilitates transgene expression levels at least 10% higher, at least 20% higher, at least 30% higher, at least 40% higher, at least 50% higher, at least 100% higher, at least 200% higher, at least 300% higher, at least 400% higher, at least 500% higher, or at least 1000% higher in one cell-type as compared to expression levels in other cell-types.

**[0232]** In some embodiments, the promoter is a liver-specific promoter. In some embodiments, the promoter is a hepatocyte-specific promoter.

**[0233]** Suitably, the promoter may be (or may be derived from) a promoter associated with a gene with selective expression in human liver cells (e.g. hepatocytes). Suitably, the promoter may be (or may be derived from) a promoter associated with a gene with selective expression in human hepatocyte cells. Methods to identify promoters associated with genes will be well known to those of skill in the art.

**[0234]** Exemplary liver-specific and/or hepatocyte-specific promoters are described in Kattenhorn, L. M., et al., 2016. Human gene therapy, 27(12), pp. 947-961 and include transthyretin (TTR) promoters, alpha-1-antitrypsin (AAT) promoters, thyroxine-binding globulin (TBG) promoters, APoE/hAAT promoters, HCR-hAAT promoters, LP1 promoters, and HLP promoters.

**[0235]** An engineered promoter variant derived from any of these promoters may be used, provided that the variant retains the capacity to drive liver-specific and/or hepatocyte-specific expression of a transgene which is operably coupled to the promoter. A skilled person will be arrive at such variants using methods known in the art. The variant may have at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identity to any of the promoters.

**[0236]** A fragment of any of these promoters (or variants thereof) may be used, provided that the fragment retains the capacity to drive liver-specific and/or hepatocyte-specific expression of a transgene which is operably coupled to the promoter. A skilled person will be able to arrive at such fragments using methods known in the art. The fragment may be, for example, at least 200 nucleotides, at least 300 nucleotides, at least 400 nucleotides, at least 500 nucleotides, or at least 1000 nucleotides in length.

**[0237]** In some embodiments, the promoter is selected from the group consisting of: a transthyretin (TTR) promoter, an alpha-1-antitrypsin (AAT) promoter, a thyroxine-binding globulin (TBG) promoter, an APoE/hAAT promoter, a HCR-hAAT promoter, a LP1 promoter, and a HLP promoter.

**[0238]** In some embodiments, the promoter is a TTR promoter, or a variant and/or fragment thereof. In some embodiments, the promoter is an Enh1mTTR (ET) promoter, or a variant and/or fragment thereof.

**[0239]** An exemplary ET promoter is provided in GenBank accession number AY661265. In some embodiments, the ET promoter comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 19 or a fragment thereof. Suitably, the ET promoter comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 19 or a fragment thereof.

**[0240]** In some embodiments, the ET promoter comprises or consists of the nucleotide sequence SEQ ID NO: 19 or a fragment thereof.

#### Exemplary ET promoter

(SEQ ID NO: 19)

```
CGCGAGTTAATAATTACCAAGCGCGGGCAAATAATAATCCGCAGGGG
CAGGTGACGTTGCCAGCGCGCTGTAATTATTACCTCGCGAATA
TTGATTCGAGGCCGCGATTGCCGCAATCGCGAGGGGCAGGTGACCTTG
```

- continued

```
CCCAAGCGCGCTTCGCCCGCCCCGGACGGTATCGATAAGCTTAGGAGC
TTGGGCTGCAGGTCGAGGGCACTGGGAGGATGTTGAGTAAGATGGAAAA
CTACTGATGACCCCTTGAGAGACAGAGATTAGGACATGTTGAACAGG
GGCCGGCGATCAGCAGGTAGCTAGAGGATCCCGTCTGCTGCACA
TTTCGTAGAGCGAGTGTCCGATACTCTAATCTCCCTAGGCAAGGTTCA
TATTGTGTAGGTTACTTATTCTCCTTTGTTGACTAAGTCATAATCA
GAATCAGCAGGTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTG
GAAGGGAGGGGTATAAAAGCCCCTCACCAAGGAGAACCGTCACACAGA
TCCACAAGCTCCGT
```

**[0241]** In some embodiments, the promoter is an AAT promoter, or a variant and/or fragment thereof. In some embodiments, the promoter is a human AAT (hAAT) promoter, or a variant and/or fragment thereof.

**[0242]** In some embodiments, the hAAT promoter comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 20 or a fragment thereof. Suitably, the hAAT promoter comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 20 or a fragment thereof.

**[0243]** In some embodiments, the hAAT promoter comprises or consists of the nucleotide sequence SEQ ID NO: 20 or a fragment thereof.

#### Exemplary hAAT promoter

(SEQ ID NO: 20)

```
GATCTTGCTACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAG
AGGGCCAGCTAAGTGGTACTCTCCAGAGACTGTCTGACTCACGCCACC
CCCTCCACCTTGGACACAGGACGCTGTGGTTCTGAGCCAGGTACAATG
ACTCCTTCTGGTAAGTGCAGTGGAGCTGTACACTGCCAGGCAAAGCG
TCCGGGCAGCGTAGGGCGACTCAGATCCAGCCAGTGGACTTAGCC
CCTGTTGCTCCCGATACTGGGTGACCTGGTTAATATTCCACAG
CAGCCTCCCCGTTGCCCTCTGGATCCACTGCTAAATACGGACGAGG
ACAG
```

**[0244]** The promoter may be a constitutive promoter. As used herein, a “constitutive promoter” is a promoter which is always active.

**[0245]** Alternatively, the promoter may be an inducible promoter. As used herein, an “inducible promoter” is a promoter which is only active under specific conditions. For example, expression of the transgene may be induced by a small molecule or drug (e.g. which binds to a promoter, regulatory sequence or to a transcriptional repressor or activator molecule) or by using an environmental trigger. Types of inducible promoter include chemically-inducible promoters (e.g. a Tet-on system); temperature-inducible promoters (e.g. Hsp70 or Hsp90-derived promoters); and light-inducible promoters. Suitably, the promoter is chemically-inducible. Any suitable method for engineering an inducible promoter may be used.

#### Enhancer Elements

**[0246]** The lentiviral vector of the present invention may comprise an enhancer, preferably a liver-specific (e.g. hepatocyte-specific) enhancer. Suitably, the protein-coding sequence is operably linked to an enhancer, preferably a liver-specific (e.g. hepatocyte-specific) enhancer.

**[0247]** An “enhancer” or “enhancer element” may refer a region of DNA that can be bound by proteins (activators) to increase the likelihood that transcription of a particular gene will occur. Enhancers are cis-acting. They can be located up to 1 Mbp (1,000,000 bp) away from the gene, upstream or downstream from the start site.

**[0248]** As used herein, a “tissue-specific enhancer” is an enhancer which preferentially facilitates expression of a gene in specific cells or tissues. Suitably, a tissue-specific enhancer may facilitate higher expression of a gene in specific cells-types as compared to other cell-types. Higher expression may be measured for example by measuring the expression of a transgene, e.g. green fluorescence protein (GFP), operably linked to the enhancer, wherein expression of the transgene correlates with the ability of the enhancer to facilitate expression of a gene. For example, a tissue-specific enhancer may be an enhancer which facilitates gene expression levels at least 10% higher, at least 20% higher, at least 30% higher, at least 40% higher, at least 50% higher, at least 100% higher, at least 200% higher, at least 300% higher, at least 400% higher, at least 500% higher, or at least 1000% higher in a specific cell-type compared to expression levels in other cell types.

**[0249]** Suitable tissue-specific enhancers will be well known to those of skill in the art. The enhancer may be a liver-specific enhancer, preferably a hepatocyte-specific enhancer.

**[0250]** Suitably, the enhancer may be (or may be derived from) an enhancer associated with a gene with selective expression in human liver cells (e.g. hepatocytes). Suitably, the enhancer may be (or may be derived from) an enhancer associated with a gene with selective expression in human hepatocyte cells. Methods to identify the enhancer regions associated with genes will be well known to those of skill in the art.

**[0251]** Exemplary liver-specific and/or hepatocyte-specific enhancers are described in Kramer, M. G., et al., 2003. Molecular therapy, 7(3), pp. 375-385, and include enhancer regions of the albumin,  $\alpha$ 1-antitrypsin, hepatitis B virus core protein, and hemopexin genes. Other liver-specific and/or hepatocyte-specific enhancers include apolipoprotein E (APoE) enhancers, hepatic control region (HCR) enhancers, and alpha-1-antitrypsin (AAT) enhancers.

**[0252]** An engineered enhancer variant derived from any of these enhancers may be used, provided that the variant retains the capacity to drive liver-specific and/or hepatocyte-specific expression of a transgene which is operably coupled to the enhancer. A skilled person will be arrive at such variants using methods known in the art. The variant may have at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identity to any of the enhancers.

**[0253]** A fragment of any of these enhancers (or variants thereof) may be used, provided that the fragment retains the capacity to drive liver-specific and/or hepatocyte-specific expression of a transgene which is operably coupled to the enhancer. A skilled person will be able to arrive at such fragments using methods known in the art. The fragment

may be at least 200 nucleotides, at least 300 nucleotides, at least 400 nucleotides, at least 500 nucleotides, or at least 1000 nucleotides in length.

**[0254]** The vector of the present invention may comprise a liver-specific promoter and/or a liver-specific enhancer, i.e. a liver-specific promoter and/or enhancer. Suitably, the protein-coding sequence is operably linked to a liver-specific promoter and/or enhancer. Suitably, the protein-coding sequence is operably linked to a hepatocyte-specific promoter and/or enhancer. The promoter and enhancer may be a combination of any of the above, for example a hAAAT promoter and an ApoE or HCR enhancer.

#### Post-Transcriptional Regulatory Elements

**[0255]** The lentiviral vector of the present invention may comprise one or more further post-transcriptional regulatory elements (e.g. in addition to one or more miRNA target sequence). Suitably, the protein-coding sequence is operably linked to one or more further post-transcriptional regulatory elements. The further post-transcriptional regulatory element may improve gene expression.

**[0256]** The lentiviral vector of the present invention may comprise a Woodchuck Hepatitis Virus Post-transcriptional Regulatory Element (WPRE). Suitably, the protein-coding sequence is operably linked to a WPRE.

**[0257]** Suitable WPRE sequences will be well known to those of skill in the art (see e.g. Zufferey, R., et al., 1999. Journal of virology, 73(4), pp. 2886-2892; and Zanta-Bousif, M. A. et al., 2009. Gene therapy, 16(5), pp. 605-619). Suitably, the WPRE is a wild-type WPRE or is a mutant WPRE. For example, the WPRE may be mutated to abrogate translation of the woodchuck hepatitis virus X protein (WHX) e.g. by mutating the WHX ORF translation start codon.

**[0258]** In some embodiments, the WPRE comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 21 or a fragment thereof. Suitably, the WPRE comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 21 or a fragment thereof.

**[0259]** In some embodiments, the WPRE comprises or consists of the nucleotide sequence SEQ ID NO: 21 or a fragment thereof.

#### Exemplary WPRE

(SEQ ID NO: 21)

```
AATCACCTCTGGATTACAAAATTGTGAAAGATTGACTGGTATTCTTA
ACTATGTTGCTCTTTACGCTATGTTGAGTACGCTGCTTTAATGCCTT
GTATCATGCTATTGCTTCCGTATGGCTTCATTTCTCCTCTTGAT
AAATCCTGGTTGCTGTCTTTATGAGGAGTTGTGGCCCGTGTCAAGC
AACGTGGCGTGGTGTGCACTGTGTTGCTGACGCAACCCCCACTGGTT
GGCATTGCCACCACCTGTCAGCTCCTTCCGGACTTCGCTTCCC
CTCCCTATTGCCACGGCGGAACCTCATGCCGCTGCTTGCCCGCTGCT
GGACAGGGGCTGGCTGTTGGGACTGACAATTCCGTGGTGTGCGG
GAAATCATCGTCCTTCCCTGGCTGCTCGCCTGTGTTGCCACCTGGATT
```

- continued

```
CTGCCGGGACGTCTCTGCTACGTCCCTCGGCCCTCAATCCAGCGG
ACCTTCCTCCCGC
```

#### Polyadenylation Sequence

**[0260]** The lentiviral vector of the present invention may comprise a polyadenylation sequence. Suitably, the protein-coding sequence is operably linked to a polyadenylation sequence. A polyadenylation sequence may be inserted after the protein-coding sequence to improve transgene expression.

**[0261]** A polyadenylation sequence typically comprises a polyadenylation signal, a polyadenylation site and a downstream element: the polyadenylation signal comprises the sequence motif recognised by the RNA cleavage complex; the polyadenylation site is the site of cleavage at which a poly-A tails is added to the mRNA; the downstream element is a GT-rich region which usually lies just downstream of the polyadenylation site, which is important for efficient processing.

**[0262]** Suitable polyadenylation sequences will be well known to those of skill in the art (see e.g. Schambach, A., et al., 2007. Molecular Therapy, 15(6), pp. 1167-1173; and Choi, J. H. et al., 2014. Molecular brain, 7(1), pp. 1-10). Exemplary polyadenylation sequences include the bGH poly(A) signal sequence and SV40 pA signal sequence.

**[0263]** Suitably, the polyadenylation sequence may be present in the 3' LTR (i.e. the lentiviral vector does not comprise an additional polyadenylation sequence).

#### Kozak Sequence

**[0264]** The lentiviral vector of the present invention may comprise a Kozak sequence. Suitably, the protein-coding sequence is operably linked to a Kozak sequence. A Kozak sequence may be inserted before the start codon to improve the initiation of translation.

**[0265]** Suitable Kozak sequences will be well known to those of skill in the art (see e.g. Kozak, M., 1987. Nucleic acids research, 15(20), pp. 8125-8148).

**[0266]** In some embodiments, the Kozak sequence comprises or consists of a nucleotide sequence which is at least 80% identical to SEQ ID NO: 22 or a fragment thereof.

**[0267]** In some embodiments, the Kozak sequence comprises or consists of the nucleotide sequence SEQ ID NO: 22 or a fragment thereof.

#### Exemplary Kozak sequence

(SEQ ID NO: 22)

```
GCCACC
```

#### Other Cis-Acting Elements

**[0268]** The lentiviral vector of the present invention may comprise any other suitable cis-acting elements, such as one or more of a rev response element (RRE); a retroviral psi packaging element; a primer binding site (PBS); a TAT activation region (TAR); splice donor and acceptor sites; and central and terminal polypyrimidine tracts.

## Long Terminal Repeats (LTRs)

**[0269]** The lentiviral vector of the present invention may comprise one or more long terminal repeat (LTR). As described above, LTRs are responsible for proviral integration and transcription. Typically, a naturally occurring LTR comprises U3, R, and U5 regions.

**[0270]** The lentiviral vector may comprise a 5' LTR and/or a 3' LTR. The lentiviral vector may comprise a 5' LTR and a 3' LTR. Suitably, a 5' LTR comprises R and U5 regions, and optionally comprise a U3 region. Suitably, a 3' LTR comprises U3, R, and U5 regions.

**[0271]** Suitable LTR sequences will be well known to those of skill in the art (see e.g. Frech, K., et al., 1996. *Virology*, 224(1), pp. 256-267).

**[0272]** In some embodiments, a LTR comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 23 or a fragment thereof. Suitably, a LTR comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 23 or a fragment thereof.

**[0273]** In some embodiments, a LTR comprises or consists of the nucleotide sequence SEQ ID NO: 23 or a fragment thereof.

## Exemplary LTR

(SEQ ID NO: 23)

```
TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTGCTTG  
TACTGGGTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTGGC  
TAACTAGGGAACCCACTGCTTAAGCCTCAATAAGCTTCCTGAGTG  
TTCAAGTAGTGTGTGCCGTCTGGTGTGACTCTGGTAACTAGAGATC  
CCTCAGACCCTTTAGTCAGTGTGGAAAATCTCTAGCAG
```

**[0274]** The lentiviral vector of the present invention may comprise one or more self-inactivating long terminal repeat (SIN-LTR). A "SIN-LTR" may comprise a deletion that abolishes transcription of the full-length virus after it has incorporated into a host cell. For example, a 3' SIN-LTR may comprise a deletion in the U3 region removing the promoter/enhancer elements (see e.g. Zufferey, R., et al., 1998. *Journal of virology*, 72(12), pp. 9873-9880). This deletion is copied into the 5' LTR after reverse transcription, thereby making the gene expression in target cells dependent on an internal promoter of choice.

**[0275]** Suitable SIN-LTR sequences will be well known to those of skill in the art (see e.g. Zufferey, R., et al., 1998. *Journal of virology*, 72(12), pp. 9873-9880 and Miyoshi, H., et al., 1998. *Journal of virology*, 72(10), pp. 8150-8157).

**[0276]** In some embodiments, the 5' LTR comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 24 or a fragment thereof. Suitably, the 5' LTR comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 24 or a fragment thereof.

**[0277]** In some embodiments, the 5' LTR comprises or consists of the nucleotide sequence SEQ ID NO: 24 or a fragment thereof.

## Exemplary 5' LTR

(SEQ ID NO: 24)

```
GGGTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTGGCTAAC  
TAGGGAACCCACTGCTTAAGCCTCAATAAGCTTCCTGAGTGCTTC  
AGTAGTGTGTGCCGTCTGGTGTGACTCTGGTAACTAGAGATCCCTC  
AGACCCCTTTAGTCAGTGTGGAAAATCTCTAGCAG
```

**[0278]** In some embodiments, the 5' LTR and/or the 3' LTR comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 25 or a fragment thereof. Suitably, the 5' LTR and/or the 3' LTR comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 25 or a fragment thereof.

**[0279]** In some embodiments, the 5' LTR and/or the 3' LTR comprises or consists of the nucleotide sequence SEQ ID NO: 25 or a fragment thereof.

**[0280]** In some embodiments, the 5' LTR and the 3' LTR comprise or consist of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 25 or a fragment thereof. Suitably, the 5' LTR and the 3' LTR comprise or consist of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 25 or a fragment thereof.

**[0281]** In some embodiments, the 5' LTR and the 3' LTR comprise or consist of the nucleotide sequence SEQ ID NO: 25 or a fragment thereof.

## Exemplary 3' LTR

(SEQ ID NO: 25)

```
TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTGCTTG  
TACTGGGTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTGGC  
TAACTAGGGAACCCACTGCTTAAGCCTCAATAAGCTTCCTGAGTG  
TTCAAGTAGTGTGTGCCGTCTGGTGTGACTCTGGTAACTAGAGATC  
CCTCAGACCCTTTAGTCAGTGTGGAAAATCTCTAGCAG
```

## Primer Binding Site (PBS)

**[0282]** The lentiviral vector of the present invention may comprise a primer binding site (PBS). A PBS is a cis-acting element where a primer may bind to initiate reverse transcription of the RNA genome (see e.g. Lanchy, J. M., et al., 1998. *Journal of Biological Chemistry*, 273(38), pp. 24425-24432).

**[0283]** Suitable retroviral PBSs will be well known to those of skill in the art.

**[0284]** In some embodiments, a PBS comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 26 or a fragment thereof. Suitably, a PBS comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 26 or a fragment thereof.

**[0285]** In some embodiments, a PBS comprises or consists of the nucleotide sequence SEQ ID NO: 26 or a fragment thereof.

Exemplary primer binding site  
 (SEQ ID NO: 26)  
 TGGCGCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGGAGCTCT  
 CTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGGCAGGG  
 GCGCGACTGGTGAGTACGCCAAAAATTGACTAGCGGAGGCTAGAAG  
 GAGAGAG

**[0286]** In some embodiments, a PBS comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 27 or a fragment thereof. Suitably, a PBS comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 27 or a fragment thereof.

**[0287]** In some embodiments, a PBS comprises or consists of the nucleotide sequence SEQ ID NO: 27 or a fragment thereof.

Exemplary primer binding site  
 (SEQ ID NO: 27)  
 TGGCGCCCGAACAGGGACCTGAAAGCGAAAGGGAAACCAGAGCTCTC  
 GACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGGCAGGGCG  
 GCGACTGGTGAGTACGCCAAAAATTGACTAGCGGAGGCTAGAAGGG  
 AGAG

#### Retroviral Psi Packaging Element

**[0288]** The lentiviral vector of the present invention may comprise a retroviral psi packaging element. As described above, a retroviral psi packaging element is a cis-acting element which is involved in regulating the essential process of packaging the retroviral RNA genome into the viral capsid during replication (see e.g. McBride, M. S., et al., 1997. Journal of virology, 71(6), pp. 4544-4554). A retroviral psi packaging element may form part of the 5' region of the gag gene.

**[0289]** Suitable retroviral psi packaging elements will be well known to those of skill in the art.

**[0290]** In some embodiments, a retroviral psi packaging element comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 28 or a fragment thereof. Suitably, a retroviral psi packaging element comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 28 or a fragment thereof.

**[0291]** In some embodiments, a retroviral psi packaging element comprises or consists of the nucleotide sequence SEQ ID NO: 28 or a fragment thereof.

Exemplary retroviral psi packaging element  
 (SEQ ID NO: 28)  
 ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGAGAATTAGATCGCGATG  
 GGAAAAAAATTGGTTAAGGCCAGGGGGAAAGAAAAAAATAAATTAAAA  
 CATATAGTATGGGCAAGCAGGGAGCTAGAACGATTGCGAGTAACTCTG  
 GCCTGTTAGAACATCAGAAGGCTGTAGACAAATCTGGGACAGCTACA

-continued

ACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACA  
 GTAGCAACCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCA  
 AGGAAGCTTAGACAAGATAGAGGAAGAGGAAACAAAAGTAAGACACCAC  
 CGCACAGCAAGCGCCCGCTGAT

#### Rev Response Element (RRE)

**[0292]** The lentiviral vector of the present invention may comprise a rev response element (RRE). As described above, a RRE is a cis-acting element that enables the efficient export of RNA transcripts of the integrated provirus from the nucleus to the cytoplasm of an infected target cell (see e.g. Pollard, V. W. and Malim, M. H., 1998. Annual review of microbiology, 52(1), pp. 491-532).

**[0293]** Suitable RRE sequences will be well known to those of skill in the art.

**[0294]** In some embodiments, a RRE comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 29 or a fragment thereof. Suitably, a RRE comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 29 or a fragment thereof.

**[0295]** In some embodiments, a RRE comprises or consists of the nucleotide sequence SEQ ID NO: 29 or a fragment thereof.

#### Exemplary rev response element

(SEQ ID NO: 29)  
 GGAGCTTGTCTGGTTCTGGGAGCAGCAGGAAGCACTATGGCG  
 CAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTAT  
 AGTGCAGCAGCAGAACAAATTGCTGAGGGCTATTGAGGCACACACAT  
 CTGTTGCAACTCACAGTCTGGGCATCAAGCAGCTCCAGGCAAGAACATC  
 TGGCTGTGAAAGATAACCTAAAGGATCAACAGCTCTGGGATT

**[0296]** In some embodiments, a RRE comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 30 or a fragment thereof. Suitably, a RRE comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 30 or a fragment thereof.

**[0297]** In some embodiments, a RRE comprises or consists of the nucleotide sequence SEQ ID NO: 30 or a fragment thereof.

#### Exemplary rev response element

(SEQ ID NO: 30)  
 GGAGCTTGTCTGGTTCTGGGAGCAGCAGGAAGCACTATGGCG  
 CAGCCTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTAT  
 AGTGCAGCAGCAGAACAAATTGCTGAGGGCTATTGAGGCACACACAT  
 CTGTTGCAACTCACAGTCTGGGCATCAAGCAGCTCCAGGCAAGAACATC  
 TGGCTGTGAAAGATAACCTAAAGGATCAACAGCTCTGGGATT

#### Central Polypurine Tract (cPPT)

**[0298]** The lentiviral vector of the present invention may comprise a central polypurine tract (cPPT). As described

above, a cPPT may allow initiation of plus-strand synthesis (see e.g. Follenzi, A. et al., 2000. *Nature genetics*, 25(2), pp. 217-222).

**[0299]** Suitable cPPT sequences will be well known to those of skill in the art.

**[0300]** In some embodiments, a cPPT comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 31 or a fragment thereof. Suitably, a cPPT comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 31 or a fragment thereof.

**[0301]** In some embodiments, a cPPT comprises or consists of the nucleotide sequence SEQ ID NO: 31 or a fragment thereof.

Exemplary central polypurine tract  
(SEQ ID NO: 31)  
AACTTTAAAAAGAAAAGGGGGATTGGGGGTACAGTCAGGGAAAGA  
ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAAC  
AAATTACAAAAATTCAAATTTTATC

#### Other Elements

**[0302]** The lentiviral vector of the present invention may comprise any other suitable elements.

**[0303]** In some embodiments, the lentiviral vector of the present invention comprises an element comprising or consisting of a nucleotide sequence which is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 32 or a fragment thereof. In some embodiments, the lentiviral vector of the present invention comprises an element comprising or consisting of the nucleotide sequence of SEQ ID NO: 32 or a fragment thereof.

Exemplary delta ENV1  
(SEQ ID NO: 32)  
TCTTCAGACCTGGAGGAGATATGAGGGACAATTGGAGAACGTGAATT  
ATATAAATATAAGTAGTAAAATTGAACCATTAGGAGTAGCACCCACC  
AAGGCAAAGAGAAGAGTGGTGAGAGAGAAAAAGAGCAGTGGAAATA

**[0304]** In some embodiments, the lentiviral vector of the present invention comprises an element comprising or consisting of a nucleotide sequence which is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 33 or a fragment thereof. In some embodiments, the lentiviral vector of the present invention comprises an element comprising or consisting of the nucleotide sequence of SEQ ID NO: 33 or a fragment thereof.

Exemplary delta ENV2  
(SEQ ID NO: 33)  
GGGTTGCTCTGGAAACTCATTGCACCACTGCTGTGCCCTGGAATGCT  
AGTTGGAGTAATAATCTCTGGAACAGATTGGAATCACACGACCTGGA  
TGGAGTGGACAGAGAAATTACAATTACACAAGCTTAATACACTCCTT

-continued

AATTGAAGAATCGAAAACCAGCAAGAAAAGAATGAACAAGAATTATTG  
GAATTAGATAAAATGGGCAAGTTGTGGAATTGGTTAACATAACAAATT  
GGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGG  
TTAAGAATAGTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAG  
GGATATTACCAATTATCGTTTCAGACCCACCTCCAACCCCGAGGGAC  
CCGACAGGCCGCAAGGAATAGAAGAAGAGGGAGAGAGACAGAGA  
CAGATCCATTGATTAGTGAACGGATC

#### Exemplary Cis-Acting Elements

**[0305]** The lentiviral vector of the present invention may comprise a cis-acting element comprising a PBS, a retroviral psi packaging element, and a rev response element (RRE).

**[0306]** In some embodiments, a cis-acting element comprising a PBS, a retroviral psi packaging element, and a RRE comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 34 or a fragment thereof. Suitably, a cis-acting element comprising a PBS, a retroviral psi packaging element, and a RRE comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 34 or a fragment thereof.

**[0307]** In some embodiments, a cis-acting element comprising a PBS, a retroviral psi packaging element, and a RRE comprises or consists of the nucleotide sequence SEQ ID NO: 34 or a fragment thereof.

Exemplary cis-acting element comprising a PBS,  
a retroviral psi packaging element, and a RRE  
(SEQ ID NO: 34)  
TGGGCCCGAACAGGGACCTGAAAGCGAAAGGGAAACCCAGAGCTCTC  
GACGCAGGACTCGGCTGCTGAAGCGCGCACGGCAAGAGGCAGGGCG  
GCGACTGGTAGACTGCCAAAATTGACTAGCGGAGGCTAGAAGGGAG  
AGAGATGGGTGCGAGCGTCAGTATTAAGCGGGGAGAATTAGATCGC  
GATGGGAAAAAATTGCGTTAAGGCCAGGGGGAAAGAAAAAAATATAAATT  
AAAACATATAGTATGGCAAGCAGGGAGCTAGAACGATTGCGAGTTAAT  
CCTGGCCTGTTAGAACATCAGAAGGCTGTAGACAAATACTGGGACAGC  
TACAACCATCCCTCAAGACAGGATCAGAAGAACTTAGATCATTATAAA  
TACAGTAGCAACCCCTCTATTGTCGTCATCAAAGGATAGAGATAAAAGAC  
ACCAAGGAAGCTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGA  
CCACCGCACAGCAAGCGCCGCGTGTCTCAGACCTGGAGGAGGAGATA  
TGAGGGACAATTGGAGAAGTGAATTATAAAATATAAGTAGTAAAAT  
TGAACCATAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAG  
AGAGAAAAAGAGCAGTGGGAATAGGAGCTTGTCTGGTTCTTGG  
GAGCAGCAGGAAGCACTATGGCGCAGCCTCAATGACGCTGACGGTACA  
GGCCAGACAATTATTGTCGCTGTAGTCAGCAGCAGAACAATTGCTG  
AGGGCTATTGAGGCGAACAGCATCTGTTGCAACTCACAGTCTGGGCA

- continued

```
TCAAGCAGCTCCAGGAAGAACTCTGGCTGTGGAAAGATAACCTAAAGGA  
TCAACAGCTCCTGGGATT
```

**[0308]** The lentiviral vector of the present invention may comprise a cis-acting element comprising a PBS, a retroviral psi packaging element, a rev response element (RRE), and a central polypurine tract (cPPT).

**[0309]** In some embodiments, a cis-acting element comprising a PBS, a retroviral psi packaging element, a RRE, and a cPPT comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 35 or a fragment thereof. Suitably, a cis-acting element comprising a PBS, a retroviral psi packaging element, a RRE, and a cPPT comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 35 or a fragment thereof.

**[0310]** In some embodiments, a cis-acting element comprising a PBS, a retroviral psi packaging element, a RRE, and a cPPT comprises or consists of the nucleotide sequence SEQ ID NO: 35 or a fragment thereof.

elements, a protein-coding sequence, optionally one or more post-transcriptional regulatory sequences, and a 3' LTR.

**[0312]** For example, the lentiviral genome of the present invention may comprise from 5' to 3': a 5' LTR, a PBS, a retroviral psi packaging element, a RRE, a cPPT, a liver-specific promoter, a protein-coding sequence, a WPRE, one or more miRNA target sequence, and a 3' LTR.

**[0313]** The lentiviral genome of the present invention may further comprise any other suitable elements, such as any other elements described herein or one or more spacer sequence. The spacer sequence(s) may comprise, for example, at least one (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10), at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least ten nucleotide bases.

**[0314]** In some embodiments, the lentiviral genome comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 36 or a fragment thereof. Suitably, the lentiviral genome comprises or consists of a nucleotide sequence which is at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 36 or a fragment thereof.

Exemplary cis-acting element comprising a PBS, a retroviral psi packaging element, a RRE, and a cPPT

(SEQ ID NO: 35)

```
GGCGCCCGAACAGGGACTGAAAGCGAAAGGGAAACCCAGAGCTCTCGACGCAGGACTCGGCTTGCTGAA  
GCGCGCACGGCAAGAGGCAGGGCGCGACTGGTAGTACGCCAAAATTTGACTAGCGGAGGCTAGAAG  
GAGAGAGATGGGTGCGAGAGCCTCAGTATTAGCGGGGAGAATTAGATCGCATGGAAAAATTGGTTA  
AGGCCAGGGGAAAGAAAAATATAAATAAACATATAGTATGGCAAGCAGGGAGCTAGAACGATTGCA  
GTTAATCCTGGCTGTTAGAACATCAGAAGGCTGTAGACAAATACTGGACAGCTACAACCATCCCTCAG  
ACAGGATCAGAAGAACTTAGTCATTATAACAGTAGCAACCCCTATTGTGTGCATCAAAGGATAGAG  
ATAAAAGACACCAAGGAAGCTTAGACAAGATAGAGGAAGAGCAAAAGTAAGACCACCGCACAGCAA  
GCGGCCGCTGATCTTCAGACCTGGAGGAGGATATGGGGACAATTGGAGAAGTGAATTATAAATATAA  
AGTAGTAAAATTGAACCATAGGAGTAGCACCCCACCAAGGCCAAAGAGAAGAGTGGTGCAGAGGAGAAAAA  
AGCAGTGGGAATAGGAGCTTGTCTGGGAGCAGCAGGAAGCACTATGGCGCAGCCTCAAT  
GACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTCAGCAGCAGACAATTGCTGAGGGCTAT  
TGAGGCGCACAGCATCTGTCACAGCTGGGACATCAAGCAGCTCAGGCAAGAACATTGGCTGT  
GGAAAGATACTAAAGGATCACAGCTCCTGGGATTGGGTTGCTCTGGAAACTCATTGACCAACTG  
TGTGCTGGAAATGCTAGTTGGAGTAATAATCTCTGGAACAGATTGGAAATCACAGCACCTGGATGGAGT  
GGACAGAGAAATTAAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAACGCAAAACAGCAAGAAAA  
GAATGAACAAGAATTATGGAAATTAGATAAATGGCAAGTTGTGGAATTGGTTAACATAACAAATTGGCT  
GTGGTATATAAATTATTCAATGATAGTAGGAGGCTGGTAGGTTAAGAATAGTTTGCTGTACTTTC  
TATAGTGAATAGAGTTAGGCAGGGATTACCACTTACGTTCAAGCCCACCTCCAAACCCGAGGGGACC  
CGACAGGCCGAAGGAATAGAAGAAGAGGTGGAGAGAGACAGAGACAGATCCATTGATTAGTGAACGG  
ATCTCGACGGTACGGTTAACTTTAAAGAAAAGGGGGATTGGGGTACAGTCAGGGAAAGAATAGT  
AGACATAATAGCAACAGACATAAAACTAAAGAATTACAAAACAAATTACAAAATTCAAAATTATC
```

Exemplary Lentiviral Genomes

**[0311]** The lentiviral genome of the present invention may comprise from 5' to 3': a 5' LTR, one or more cis-acting

**[0315]** In some embodiments, the lentiviral genome comprises or consists of the nucleotide sequence SEQ ID NO: 36 or a fragment thereof.

Exemplary lentiviral genome  
(SEQ\_ID NO: 36)  
GGTCTCTGGTTAGACCAGATCTGAGCCTGGAGCTCTGGTAACTAGGAAACCCACTGCTTAAGCCT  
CAATAAACGCTTGCTTGAAGTAGTGTGCCCCCTGTTGTGACTCTGGTAACAGAGATCC  
CTCAGACCCCTTGTAGTCAGTGAAAATCTAGCAGTGGGCCCGAACAGGGACTGAAAGCGAAAGGGA  
AACCAGAGCTCTCGACGCAGGACTCGGCTGCTGAAGCGCGCACCGCAAGAGGGGAGGGCGACTGG  
TGAGTACGCCAAAATTGACTAGCGGAGGCTAGAAGGAGAGATGGTGCAGAGCGTCAGTATTAAGC  
GGGGGAGAATTAGATCGCGATGGGAAAAATTGGTTAAGGCCAGGGGGAAAGAAAAATATAAATTAAAC  
ATATAGTATGGCAAGCAGGGAGCTAGAACGATTGCACTTAATCTGGCTGTTAGAACATCAGAAGGCT  
GTAGACAATACTGGGACAGCTACAACCACCCCTCAGACAGGATCAGAAGAACTTAGATCATTATATAATA  
CAGTAGCAACCCCTATTGTGTCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTAGACAAGATAG  
AGGAAGAGCAAAACAAAAGTAAGACCAACCGCACAGCAAGCGGCCGCTGATCTCAGACCTGGAGGAGGAGAT  
ATGAGGGACAATTGGAGAAGTGAATTATAAATATAAAGTAGTAAAAATTGAAACATTAGGAGTAGCACCC  
ACCAAGGCAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTGTCCCTGGTTC  
TTGGGAGCAGCAGGAAGCACTATGGCGCAGCCTCAATGACGCTGACGGTACAGGCAGACAATTATTGTCT  
GGTATAGTGCAGCAGCAGAACATTGCTGAGGGTATTGAGGCGCAACAGCATCTGGCAACTCACAGTC  
TGGGGCATCAAGCAGCTCCAGGCAAGAACCTGGCTGTGAAAGATACCTAAAGGATCAACAGCTCCTGGGG  
ATTTGGGGTGTGCTGGAAAACCTATTGCAACACTGGATGGAGTGGACAGAGAAATTACAATTACACAAGCTTAATA  
CTGGAACAGATTGGAATCACACGACCTGGATGGAGTGGACAGAGAAATTACAATTACACAAGCTTAATA  
CACTCCTTAATTGAAGAATCGAAAACAGCAAGAAAAGAATGAACAAGAATTATTGAAATTAGATAATGG  
GCAAGTTGTGAAATTGGTTAACATAACAAATTGGCTGTGGTATATAAAATTATTCAAATGATAGTAGGA  
GGCTTGGTAGGTTAAGAATAGTTTGTACTTCTATAGTGAATAGAGTTAGGCAGGGATATTCA  
TTATCGTTCAGACCCACCTCCAACCCGGGGGACCCGACAGGCCGAAGGAATAGAAGAAGGGTGA  
GAGAGAGACAGAGACAGATCCATTGATTAAGTGAACGGATCTCGACGGTATCGTTAACTTTAAAAGAAAA  
GGGGGGATTGGGGGTACAGTGCAGGGAAAGAATAGTAGACATAATAGAACAGACATAAAACTAAAGAA  
TTACAAAACAAATTACAAAATTCAAAATTATTATCGATCACGAGACTAGCCTCGAGCACGCGAGTTAA  
TTACCGCGGGCAAATAATAATCCGCGAGGGGCAGGTGACGTTGCCAGCGCGCTGTTAATTATT  
AACCTCGCGAATTGATTGAGGCCGATTGCCAATCGCGAGGGCAGGTGACCTTGCCCAGCGCGCTGTTAATT  
GTTCGCCCCGCCGGTACGATAAGCTAGGAGCTGGGCTGCAAGTGGGACTGGGAGGATGT  
TGAGTAAGATGGAAAATACTGATGACCTGGCAGAGACAGAGTATTAGGACATGTTGAAACAGGGGGGG  
CGATCAGCAGGTAGCTAGAGGATCCCGTCTGTCGACATTGCTAGAGCGAGTGTCCGATACTCTAA  
TCTCCCTAGGCAAGGTCATATTGTGAGGTTACTTATTCTCCTTGTGACTAAGTCATAATCAGAAT  
CAGCAGGTTGGAGTCAGCTGGCAGGGATCAGCAGCCTGGGTTGAAAGGAGGGGTATAAGCCCTCA  
CCAGGAGAACCGTCACACAGTCCACAAGCTCTGGTAGCGTACGCCACCATGGCAGCGTGAGGATGCT  
GAGGACCTGGAGCGGAACGCAGGCAAGCTGATCTGCGTAGAGATACTTCAGACATGTGGCAACGTGACGT  
GCTGAAGCCAATTACGTGTGCTTCTCGCTACCCCTCTTCAAGTATTCTCACCCCTCACCACCTTGAA  
GACCACAGCCGCGCTGAGGGGACAGGTGGTGCAGTCAAGCTGAGCGACACAGTGAGCCAGTTGATCCATCTGAGG  
GACCGTGAAGGAGTGGTACGTGAAGGAGGGCAGACACAGTGAGCCAGTTGATCCATCTGAGGTGAGCAGT  
TGACAAGGCCAGCGTGCACATCACATCCGGTAGCGTACGATGGCGTACAGAGCTGACTATAACCTGGACGA  
CATCGCCTATGTGGCAAGCCACTGGGGACATCGAGACAGAGGCCCTGAAGGACAGCGAGGAGGATGTGGT

- continued

```

GGAGACACCCGCGTGTCCCACGATGAGCACACACACCAGGAGATCAAGGGAGGAAGACCCCTGGCACACC
AGCCGTGCGGAGACTGGCATGGAGAACAAATATCAAGCTGAGCGAGGTGGTGGATCCGGCAAGGACGGCAG
AATCCTGAAGGAGGACATCCTGAACATCTGGAGAACGAGACCCGGCGAACCTGCCACCTTCCCTAAGGTT
GGAGATCATGCCACCCCCCTCAAAGCCTAAGGACATGACAGTGCAATCCTGGTGTCAAGCCCCCGTGT
CACCGGCAAGGATAAGACAGAGCCTATCAAGGGCTTCAGAGGCATGGTGAAGACCATGAGCGCCGCC
GAAGATCCCACACTCGGCTACTGCGACGAGATCGATCTGACAGAGCTGGTGAAGCTGCGGGAGGAGCTGAA
GCCAATCGCCTCGCCAGGGCATCAAGCTGTCCTTATGCCCTTCTGAAAGGCCGCTCTGGGCC
GCTGCAGTTCCATCCTGAACGCCCTGTGGACGAGAACGCCAGAATATCACCTATAAGGCCAGGCC
TATCGGCATCGCCATGGATAACAGAGCAGGGCTGATCGTGCACAGGGCTGATCGACAGATCTGTTCCAT
CTTCGACATGCCACCGAGCTGAACAGGCTGCAGAAGCTGGCTCTGTGGGCCAGCTGAGCACACAGATCT
GACCGGGCGGACCTTCACACTGTCCAATATCGGCCATCGGCCGACATTGCAAGGCCGTGATCATGCC
ACACAGAGGTGGCAATCGGCCCTGGCTCATCAAGGCCATCCCTCGCTTCAACCAAGGGGAGGTGTA
CAAGGCCAGATCATGAATGTGAGCTGGTCCGCCAACAGAGTGTGATGCCGACCCATGTCTCGCTT
CAGCAACCTGTGGAAGTCTATCTGGAGAATCCGCCCTTATGCTGCTGGATCTGAAGTGATGAGTCGACTC
GACAATCAACCTCTGGATTACAAAATTGTGAAAGATTGACTGGTATTCTTAACATGTTGCTCCTTTACG
CTATGTGGATACGCTGCTTAAATGCCCTTGTATCATGCTATTGCTCCCGTATGGCTTCACTTCTCC
TTGTATAATCTGGTTGCTGTCCTTTATGAGGAGTTGTGGGCCGTGTCAGGCAACGTGGCGTGGTGC
ACTGTGTTGCTGACGCAACCCCCACTGGTGGGCATTGCCACACCTGTCAGCTCCTTCCGGACTT
GCTTCCCCCTCCATTGCCACGGGAACCTCATGCCCGCTGCCCTGCCGCTGTCAGGACAGGGCTCG
CTGTTGGCAGTACAATTCCGGTGTGTCCTTTATGAGGAGTTGTGGGCCGTGTCAGGCAACGTGGCGTGGTGC
GCCACCTGGATTCTGCGCGGGACGTCCCTGCTACGTCCCTCGGCCCTCAATCCAGCGGACCTTCC
CGCGGCCCTGCTGCCGGCTAGATAATCCATAAAGTAGGAAACACTACACGATTCCATAAAGTAGGAAACAC
TACAACCGGTTCCATAAAGTAGGAAACACTACATCACTCCATAAAGTAGGAAACACTACACCCGGTCGAC
TCGGTACCTTAAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTAAAGAAAAGGGGG
CTGGAAAGGCTAATTCACTCCAAACGAAGACAAGATCTGCTTGTACTGGGTCTCTGGTTAGAC
CAGATCTGAGCCTGGAGCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAGCTGCCTGA
GTGCTTCAAGTAGTGTGCCCCGTGTTGTGACTCTGGTAAGAGATCCCTCAGACCCTTTAGTCA
GTGTGGAAAATCTCTAGCA

```

#### Variants, Derivatives, Analogues, and Fragments

**[0316]** In addition to the specific polypeptides and polynucleotides mentioned herein, the invention also encompasses variants, derivatives, and fragments thereof.

**[0317]** In the context of the invention, a “variant” of any given sequence is a sequence in which the specific sequence of residues (whether amino acid or nucleic acid residues) has been modified in such a manner that the polypeptide or polynucleotide in question retains at least one or all of its endogenous functions. A variant sequence can be obtained by addition, deletion, substitution, modification, replacement and/or variation of at least one residue present in the naturally occurring polypeptide or polynucleotide.

**[0318]** The term “derivative” as used herein in relation to proteins or polypeptides of the invention includes any substitution of, variation of, modification of, replacement of, deletion of and/or addition of one (or more) amino acid

residues from or to the sequence, providing that the resultant protein or polypeptide retains at least one or all of its endogenous functions.

**[0319]** Typically, amino acid substitutions may be made, for example from 1, 2 or 3, to 10 or 20 substitutions, provided that the modified sequence retains the required activity or ability. Amino acid substitutions may include the use of non-naturally occurring analogues.

**[0320]** Polypeptides used in the invention may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent polypeptide. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues as long as the endogenous function is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and

amino acids with uncharged polar head groups having similar hydrophilicity values include asparagine, glutamine, serine, threonine and tyrosine.

[0321] Conservative substitutions may be made, for example according to the table below. Amino acids in the same block in the second column and in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P I L V
	Polar - uncharged	C S T M N Q
	Polar - charged	D E K R H
AROMATIC		F W Y

[0322] The effect of additions, deletions, substitutions, modifications, replacements and/or variations may be predicted using any suitable prediction tool e.g. SIFT (Vaser, R., et al., 2016. Nature protocols, 11(1), pp. 1-9), PolyPhen-2 (Adzhubei, I., et al., 2013. Current protocols in human genetics, 76(1), pp. 7-20), CADD (Rentzsch, P., et al., 2021. Genome medicine, 13(1), pp. 1-12), REVEL (Ioannidis, N. M., et al., 2016. The American Journal of Human Genetics, 99(4), pp. 877-885), MetaLR (Dong, C., et al., 2015. Human molecular genetics, 24(8), pp. 2125-2137), and/or MutationAssessor (Reva, B., et al., 2011. Nucleic acids research, 39(17), pp. e118-e118) or based on clinical data e.g. ClinVar (Landrum, M. J., et al., 2016. Nucleic acids research, 44(D1), pp. D862-D868). Suitable additions, deletions, substitutions, modifications, replacements and/or variations may be considered tolerated, benign, and/or likely benign.

[0323] Typically, a variant may have a certain identity with the wild type amino acid sequence or the wild type nucleotide sequence.

[0324] In the present context, a variant sequence is taken to include an amino acid sequence which may be at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85% or at least 90% identical, suitably at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to the subject sequence. Although a variant can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express in terms of sequence identity.

[0325] In the present context, a variant sequence is taken to include a nucleotide sequence which may be at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85% or at least 90% identical, suitably at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to the subject sequence. Although a variant can also be considered in terms of similarity, in the context of the present invention it is preferred to express it in terms of sequence identity.

[0326] Suitably, reference to a sequence which has a percent identity to any one of the SEQ ID NOs detailed herein refers to a sequence which has the stated percent identity over the entire length of the SEQ ID NO referred to.

[0327] Sequence identity comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate percent identity between two or more sequences.

[0328] Percent identity may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid or nucleotide in one sequence is directly compared with the corresponding amino acid or nucleotide in the other sequence, one residue at a time. This is called an “ungapped” alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

[0329] Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion in the amino acid or nucleotide sequence may cause the following residues or codons to be put out of alignment, thus potentially resulting in a large reduction in percent identity when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall identity score. This is achieved by inserting “gaps” in the sequence alignment to try to maximise local identity.

[0330] However, these more complex methods assign “gap penalties” to each gap that occurs in the alignment so that, for the same number of identical amino acids or nucleotides, a sequence alignment with as few gaps as possible, reflecting higher relatedness between the two compared sequences, will achieve a higher score than one with many gaps. “Affine gap costs” are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

[0331] Calculation of maximum percent identity therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (see e.g. Devereux, J., et al., 1984. Nucleic acids research, 12(1), pp. 387-395). Examples of other software that can perform sequence comparisons include, but are not limited to, the BLAST package (see e.g. Altschul, S. F., et al., 1990. Journal of molecular biology, 215(3), pp. 403-410), BLAST 2 (see e.g. Tatusova, T. A. and Madden, T. L., 1999. FEMS microbiology letters, 174(2), pp. 247-250), FASTA (see e.g. Pearson, W. R. and Lipman, D. J., 1988. PNAS, 85(8), pp. 2444-2448.), EMBOSS Needle (Madeira, F., et al., 2019. Nucleic acids research, 47(W1), pp. W636-W641) and the GENWORKS suite of comparison tools. For some applications, it is preferred to use EMBOSS Needle.

[0332] Although the final percent identity can be measured, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix.

[0333] Once the software has produced an optimal alignment, it is possible to calculate percent sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result. The percent sequence identity may be calculated as the number of identical residues as a percentage of the total residues in the SEQ ID NO referred to.

[0334] "Fragments" are also variants and the term typically refers to a selected region of the polypeptide or polynucleotide that is of interest either functionally or, for example, in an assay. "Fragment" thus refers to an amino acid or nucleic acid sequence that is a portion of a full-length polypeptide or polynucleotide.

[0335] Such variants, derivatives, and fragments may be prepared using standard recombinant DNA techniques such as site-directed mutagenesis. Where insertions are to be made, synthetic DNA encoding the insertion together with 5' and 3' flanking regions corresponding to the naturally-occurring sequence either side of the insertion site may be made. The flanking regions will contain convenient restriction sites corresponding to sites in the naturally-occurring sequence so that the sequence may be cut with the appropriate enzyme(s) and the synthetic DNA ligated into the cut. The DNA is then expressed in accordance with the invention to make the encoded polypeptide. These methods are only illustrative of the numerous standard techniques known in the art for manipulation of DNA sequences and other known techniques may also be used.

#### Method of Production

[0336] In one aspect, the present invention provides a method of producing a lentiviral vector according to the present invention.

[0337] Suitable methods to produce lentiviral vectors will be well known to those of skill in the art (see e.g. Merten, O. W., et al., 2016. Molecular Therapy-Methods & Clinical Development, 3, p. 16017).

[0338] The method of production may comprise: (a) introducing a transfer vector and one or more helper vectors into a host cell; (b) culturing the host cell under conditions suitable to produce lentiviral vectors according to the present invention; and (c) obtaining the lentiviral vectors from the host cell.

[0339] As used herein, a "transfer vector" may encode the lentiviral genome of the present invention. Suitably, the transfer vector used to produce the lentiviral viral genome within a host cell/packaging cell will have sufficient lentiviral genetic information to allow packaging of an RNA genome, in the presence of packaging components (e.g. gag-pol, rev, env), into a viral particle which is capable of infecting a target cell, but is incapable of independent replication to produce infectious viral particles within the final target cell.

[0340] The transfer vector used to produce the viral genome within a host cell/packaging cell may include transcriptional regulatory control sequences operably linked to the lentiviral genome to direct transcription of the genome in a host cell/packaging cell. These regulatory sequences may be the natural sequences associated with the transcribed viral sequence (i.e. the 5' U3 region), or they may be a heterologous promoter, such as another viral promoter (e.g. the CMV promoter). The transfer vector may be a plasmid.

[0341] As used herein, a "helper vector" may encode one or more packaging components (e.g. gag-pol, rev, env). The nucleotide sequence encoding the packaging component(s) may be operably linked to a promoter (e.g. a CMV promoter or a RSV promoter) and/or a polyadenylation signal. The term "helper vector" may include "packaging vectors" (e.g. encoding gag-pol or rev) and "envelope vectors" (e.g. encoding an env gene, such as VSV-g). The helper vectors, packaging vectors, and/or envelope vectors may be plasmids.

[0342] The transfer vector and one or more helper vectors may be introduced into the host cell by any suitable technique known in the art, such as transfection, transduction and/or transformation. Suitably, the helper vectors may be transiently transfected or transduced into the host cell or may be stably maintained (e.g. stably integrated into the cell genome or episomally maintained) within the host cell. Alternatively, a combination of transient transfection or transduction and stable maintenance may be used to introduce the helper vectors into the host cell.

[0343] Suitably, the transfer vector and/or the helper vectors may be plasmids and introduced by transfection. Suitably, a four plasmid system may be used consisting of a transfer plasmid and three helper plasmids. The three helper plasmids may consist of: a first helper plasmid encoding a gag-pol gene; a second helper plasmid encoding a rev gene; and a third helper plasmid encoding an env gene. Alternatively, a three plasmid system may be used which consists of a transfer plasmid, one helper plasmid encoding a gag-pol gene and a rev gene; and one helper plasmid encoding an env gene. Alternatively, a two plasmid system may be used in which all helper functions (e.g. gag-pol, rev, and env) are encoded by one helper plasmid.

[0344] Any suitable host cell may be used to produce the lentiviral vector. Suitable host cells include producer cells and packaging cells, such as those described below (e.g. HEK 293, or derivatives thereof). Suitable conditions for culturing the host cell will be well known to the skilled person. For example, the host cells may be incubated in chemically defined medium for from about 1 day to about 5 days (e.g. about 48 hours, about 72 hours, or about 96 hours).

[0345] The lentiviral vector may be obtained using in any suitable methods known in the art. For example the culture supernatant may be harvested and lentiviral vector subsequently purified from the culture supernatant (e.g. by centrifugation, membrane filtration and/or chromatography). The method of production may further comprise any other suitable process steps e.g. DNA reduction, concentration, formulation and/or sterilization.

#### Vectors, Kits and Systems

[0346] In one aspect, the present invention provides a vector encoding the lentiviral genome of the present invention. The vector may be a transfer vector, as described herein. For example, the vector may be a plasmid and/or the lentiviral genome may be operably linked to a promoter (e.g. a viral promoter, such as a CMV promoter).

[0347] In one aspect, the present invention provides a kit or system for producing the lentiviral vector of the present invention.

[0348] The kit or system may be a lentivirus packaging kit or system or a lentivirus production kit or system. As used herein, a "lentivirus packaging kit or system" may comprise

one or more components, and optionally instructions, for packaging the lentiviral vector of the present invention. As used herein, a “lentivirus production kit or system” may comprise one or more components, and optionally instructions, for producing the lentiviral vector of the present invention.

[0349] The kit or system may comprise a transfer vector encoding the lentivirus genome of the present invention and optionally one or more helper vectors. The kit or system may further comprise host cells (e.g. packaging cells or producer cells) and/or other reagents (e.g. transfection reagent, culture medium, etc.). The kit or system may further comprise any other suitable components, and optionally instructions for packaging and/or producing the lentiviral vector of the present invention.

#### Cells

[0350] In one aspect, the present invention provides a cell comprising the lentiviral vector of the present invention. The cell may be an isolated cell. Suitably, the cell is a mammalian cell, for example a human cell. The cell may be an isolated human cell.

[0351] Suitably, the cell may be a producer cell. The term “producer cell” includes a cell that produces viral particles, after transient transfection, stable transfection or vector transduction of all the elements necessary to produce the viral particles or any cell engineered to stably comprise the elements necessary to produce the viral particles. Suitable producer cells will be well known to those of skill in the art and may include HEK293, COS-1, COS-7, CV-1, HeLa, CHO, and A549 cell lines. In some embodiments, the producer cell is a HEK293 cell, or a derivative thereof (e.g. a HEK293T cell, a HEK293T Lenti-X, a HEK293T-Rex cell, a HEK293FT cell, a HEK293SF-3F6 cell, a HEK293SF-3F9 cell, a HEK293-EBNA1 cell, or a SJ293TS cell).

[0352] Suitably, the cell may be a packaging cell. The term “packaging cell” includes a cell which contains some or all of the elements necessary for packaging a recombinant virus genome. Typically, such packaging cells contain one or more vectors which are capable of expressing viral structural proteins (e.g. gag-pol, rev, env) and/or one or more genes encoding the viral structural proteins have been integrated into the genome of the packaging cell. Cells comprising only some of the elements required for the production of enveloped viral particles are useful as intermediate reagents in the generation of viral particle producer cell lines, through subsequent steps of transient transfection, transduction or stable integration of each additional required element. These intermediate reagents are encompassed by the term “packaging cell”. Suitable packaging cells will be well known to those of skill in the art (see e.g. Merten, O. W., et al., 2016. Molecular Therapy-Methods & Clinical Development, 3, p. 16017).

[0353] Suitably, the cell may be a liver cell, for example a hepatocyte. Suitably, the cell may be an immortalized liver cell, for example an immortalized hepatocyte. Suitable cell lines will be well known to those of skill in the art, for example HepG2, Hep3B, HBG, and HepaRG cell lines. Methods to generate immortalized liver cells (e.g. immortalized hepatocytes) will be well known to those of skill in the art (see e.g. Ramboer, E., et al., 2015. Methods Mol Biol, 1250, pp. 53-76). Suitably, the cell may be a stem cell.

#### Pharmaceutical Compositions

[0354] In one aspect, the present invention provides pharmaceutical composition comprising the lentiviral vector or cell of the present invention. In preferred embodiments, the pharmaceutical composition comprises the lentiviral vector of the present invention in the form of a lentiviral particle.

[0355] A pharmaceutical composition is a composition that comprises or consists of a therapeutically effective amount of a pharmaceutically active agent e.g. the lentiviral vector. A pharmaceutical composition preferably includes a pharmaceutically acceptable carrier, diluent or excipient (including combinations thereof).

[0356] By “pharmaceutically acceptable” is included that the formulation is sterile and pyrogen free. The carrier, diluent, and/or excipient must be “acceptable” in the sense of being compatible with the lentiviral vector and not deleterious to the recipients thereof. Typically, the carriers, diluents, and excipients will be saline or infusion media which will be sterile and pyrogen free, however, other acceptable carriers, diluents, and excipients may be used.

[0357] Acceptable carriers, diluents, and excipients for therapeutic use are well known in the pharmaceutical art. The choice of pharmaceutical carrier, excipient or diluent can be selected with regard to the intended route of administration and standard pharmaceutical practice. The pharmaceutical compositions may comprise as (or in addition to) the carrier, excipient or diluent any suitable binder(s), lubricant(s), suspending agent(s), coating agent(s) or solubilising agent(s).

[0358] Examples of pharmaceutically acceptable carriers include, for example, water, salt solutions, alcohol, silicone, waxes, petroleum jelly, vegetable oils, polyethylene glycols, propylene glycol, liposomes, sugars, gelatin, lactose, amylose, magnesium stearate, talc, surfactants, silicic acid, viscous paraffin, perfume oil, fatty acid monoglycerides and diglycerides, petroethral fatty acid esters, hydroxymethylcellulose, polyvinylpyrrolidone, and the like.

[0359] The lentiviral vector, cell, or pharmaceutical composition according to the present invention may be administered in a manner appropriate for treating and/or preventing the diseases described herein. Suitable administration routes will be known to the skilled person (see e.g. Fumoto, S., et al., 2013. Novel Gene Therapy Approaches, pp. 3-31). The quantity and frequency of administration may be determined by the skilled person, for example depending by such factors as the condition of the subject, and the type and severity of the subject’s disease. The pharmaceutical composition may be formulated accordingly.

[0360] The lentiviral vector, cell or pharmaceutical composition according to the present invention may be administered parenterally, (e.g. intravenous, intra-arterial, intramuscular, intrathecal, subcutaneous), or by infusion techniques. The lentiviral vector, cell or pharmaceutical composition may be administered in the form of a sterile aqueous solution which may contain other substances, for example, enough salts or glucose to make the solution isotonic with blood. The aqueous solution may be suitably buffered (preferably to a pH of from 3 to 9). The pharmaceutical composition may be formulated accordingly. The preparation of suitable parenteral formulations under sterile conditions is readily accomplished by standard pharmaceutical techniques well-known to those skilled in the art.

[0361] The lentiviral vector, cell or pharmaceutical composition according to the present invention may be administered systemically, for example by intravenous injection or intraperitoneal injection. In some embodiments, the lentiviral vector, cell or pharmaceutical composition according to the present invention is administered by intravenous injection. The pharmaceutical composition may be formulated accordingly.

[0362] The lentiviral vector, cell or pharmaceutical composition according to the present invention may be administered locally, for example by direct injection, intra-arterial injection, or intraportal injection. In some embodiments, the lentiviral vector, cell or pharmaceutical composition according to the present invention is administered locally to the liver. In some embodiments, the lentiviral vector, cell or pharmaceutical composition according to the present invention is administered by intrahepatic injection, intrahepatic arterial injection, or intraportal injection. The pharmaceutical composition may be formulated accordingly.

[0363] The pharmaceutical compositions may comprise lentiviral vectors or cells of the invention in infusion media, for example sterile isotonic solution. The pharmaceutical composition may be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0364] The lentiviral vector, cell or pharmaceutical composition may be administered in a single or in multiple doses. Suitably, the lentiviral vector, cell or pharmaceutical composition may be administered in a single, one off dose. The pharmaceutical composition may be formulated accordingly.

[0365] The lentiviral vector, cell or pharmaceutical composition may be administered at varying doses (e.g. measured in Transducing Units (TU) per kg). The physician in any event may determine the actual dosage which will be most suitable for any individual subject and the dosage may, for example, vary with the age, weight and response of the particular subject. Suitably, the lentiviral vector of the present invention is administered at a dose of at least about  $10^8$  TU/kg, at least about  $10^9$  TU/kg, or at least about  $10^{10}$  TU/kg. Suitably, the lentiviral vector of the present invention is administered at a dose of about  $10^{13}$  TU/kg or less, about  $10^{12}$  TU/kg or less, or about  $10^{11}$  TU/kg or less. Suitably, the lentiviral vector of the present invention is administered in a dose of from about  $10^8$  to about  $10^{13}$  TU/kg, from about  $10^9$  to about  $10^{13}$  TU/kg, or from about  $10^{10}$  to about  $10^{13}$  TU/kg. Suitably, the lentiviral vector of the present invention is administered in a dose of from about  $10^8$  to about  $10^{12}$  TU/kg, from about  $10^9$  to about  $10^{12}$  TU/kg, or from about  $10^{10}$  to about  $10^{12}$  TU/kg. Suitably, the lentiviral vector of the present invention is administered in a dose of from about  $10^8$  to about  $10^{11}$  TU/kg, from about  $10^9$  to about  $10^{11}$  TU/kg, or from about  $10^{10}$  to about  $10^{11}$  TU/kg. In some embodiments, the lentiviral vector of the present invention is administered in a dose of from about  $10^8$  to about  $10^{11}$  TU/kg, from about  $10^9$  to about  $10^{10}$  TU/kg, or from about  $10^{10}$  to about  $10^{11}$  TU/kg. In some embodiments, the lentiviral vector of the present invention is administered in a dose of from about  $10^9$  to about  $10^{10}$  TU/kg. The pharmaceutical composition may be formulated accordingly.

[0366] The pharmaceutical composition may further comprise one or more other therapeutic agents.

[0367] The invention further includes kits comprising the lentiviral vector, cell and/or pharmaceutical composition of the present invention. Preferably said kits are for use in the methods and used as described herein, e.g., the therapeutic methods as described herein. Preferably said kits comprise instructions for use of the kit components.

#### Methods for Treating and/or Preventing Disease

[0368] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use as a medicament.

[0369] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament.

[0370] In one aspect, the present invention provides a method of administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

[0371] The lentiviral vector mediated gene therapy described herein may allow for a stable gene transfer even in paediatric patients at the first disease stages by virtue of lentiviral vector genomic integration.

[0372] Following administration of the lentiviral vector of the present invention, the lentiviral vector may integrate into the genome of liver cells (e.g. hepatocyte cells). Subsequently, the lentiviral vector may be maintained in the genome of liver cells (e.g. hepatocyte cells) as they duplicate. The integration of the lentiviral vector in the genome of liver cells may be determined by integration site (IS) analysis (e.g. quantitative high-throughput vector IS analysis). Suitable methods are known in the art (see e.g. Cantore, A., et al., 2015. Science translational medicine, 7(277), p. 277ra28).

[0373] The lentiviral vector, cell or pharmaceutical composition may be administered to any subject in need thereof. The subject may be a mammal (e.g. a human). In some embodiments, the subject is male. In some embodiments, the subject is female. In preferred embodiments, the lentiviral vector is administered in the form of a lentiviral particle.

[0374] In some embodiments, the subject is a juvenile, an adolescent, or a child. The term "juvenile" may refer to an individual that has not yet reached adulthood. The term "adolescent" may refer to an individual during the period from the onset of puberty to adulthood. The term "child" may refer an individual between the stages of birth and puberty.

[0375] In some embodiments, the subject is a young child, a toddler, or an infant. The term "young child" may refer to a human subject aged from 3 years to 5 years. The term "toddler" may refer to a human subject aged from 1 year to 3 years. The term "infant" may mean refer to a human subject under the age of 12 months.

[0376] In some embodiments, the subject is a paediatric patient. The term "paediatric patient" may refer to a human subject until about 18-21 years of age (see e.g. Sawyer, S. M., et al., 2019. The Lancet Child & Adolescent Health, 3(11), pp. 822-830).

[0377] In some embodiments, the subject is a neonatal patient or an infantile patient. The term "neonatal patient" may refer to a human subject who is aged about 4 weeks old or younger. The term "infantile patient" may refer to a human subject who is aged from about 4 weeks to about 1 year.

[0378] In other embodiments, the subject is an adult. Human liver is expected to completely renew every 5 years in humans, so integrating vectors are expected to be more persisting compared to mostly episomal vectors (e.g. AAV).

#### Maple Syrup Urine Disease (MSUD)

[0379] The vector, cell or pharmaceutical composition according to the present invention may be used to prevent and/or treat maple syrup urine disease (MSUD).

[0380] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating maple syrup urine disease (MSUD).

[0381] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating maple syrup urine disease (MSUD).

[0382] In one aspect, the present invention provides a method of preventing or treating maple syrup urine disease (MSUD), the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

[0383] As described above, MSUD (MCID: MPL001; OMIM: 248600) is an autosomal recessive metabolic disease caused by decreased activity of the branched-chain alpha-ketoacid dehydrogenase complex (BCKDC). There are five distinct clinical phenotypes of MSUD: classic, intermediate, intermittent, thiamine-responsive, and E3-deficient. The MSUD forms can be categorized based on age at onset, severity of symptoms, response to thiamine supplementation, and biochemical findings (see e.g. Blackburn, P. R., et al., 2017. The application of clinical genetics, 10, pp. 57-66).

[0384] In some embodiments, the MSUD is selected from classic MSUD, intermediate MSUD, intermittent MSUD, thiamine-responsive MSUD, and E3-deficient MSUD. In some embodiments, the MSUD is classic MSUD, intermediate MSUD, or thiamine-responsive MSUD. In some embodiments, the MSUD is classic MSUD.

[0385] Biochemical derangement caused by biallelic pathogenic variants in BCKDHA, encoding BCKA decarboxylase (E1) alpha subunit, may be referred to as MSUD type 1A; biochemical derangement caused by biallelic pathogenic variants in BCKDHB encoding BCKA decarboxylase (E1) beta subunit may be referred to as MSUD type 1B; biallelic pathogenic variants in DBT encoding dihydrolipoil transacylase (E2) subunit may be referred to as MSUD type 2; and biallelic pathogenic variants in DLD encoding the E3 subunit (lipoamide dehydrogenase) of BCKD, may be referred to as MSUD type 3 (see e.g. Strauss, K. A. et al., 2006. In Gene Reviews).

[0386] In some embodiments, the MSUD is selected from MSUD type 1A, MSUD type 1B, MSUD type 2, and MSUD type 3. In preferred embodiments, the MSUD is MSUD type 2. In some embodiments, the MSUD is classic MSUD, type 2.

[0387] Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) levels of branched-chain amino acids (BCAA) may be prevented from increasing or reduced. Suitably, the BCAs may be selected from one or more of leucine, isoleucine and valine. Suitably, serum (or plasma) BCAA

levels may be reduced by at least 10%, at least 20%, at least 30%, at least 40%, or at least 50%. Suitably, following administration, the subject's serum (or plasma) BCAA levels may be about 20 BCAA/Ala or less, about 15 BCAA/Ala or less, or about 10 BCAA/Ala or less. Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum BCAA levels may be normalised. A normal serum BCAA level may be about 2 BCAA/Ala or less or about 1 BCAA/Ala or less. Serum (or plasma) BCAA levels may be determined by any method known in the art, for example as described in Terrlink, T., et al., 1994. Clinical chemistry, 40(2), pp. 245-249 and Oglesbee, D., et al., 2008. Clinical chemistry, 54(3), pp. 542-54.

[0388] Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) levels of leucine may be prevented from increasing or reduced. Suitably, serum (or plasma) leucine levels may be reduced by at least 10%, at least 20%, at least 30%, at least 40%, or at least 50%. Suitably, following administration, the subject's serum (or plasma) leucine level may be from about 150 to about 300  $\mu\text{mol/L}$ . Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) leucine levels may be normalised. A normal serum (or plasma) leucine level may be from about 62 to about 200  $\mu\text{mol/L}$  or about  $119 \pm 38 \mu\text{mol/L}$  (see e.g. Strauss, K. A., et al., 2020. Molecular genetics and metabolism, 129(3), pp. 193-206).

[0389] Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) levels of isoleucine may be prevented from increasing or reduced. Suitably, serum (or plasma) isoleucine levels may be reduced by at least 10%, at least 20%, at least 30%, at least 40%, or at least 50%. Suitably, following administration, the subject's serum (or plasma) isoleucine level may be from about 150 to about 300  $\mu\text{mol/L}$ . Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) isoleucine levels may be normalised. A normal serum (or plasma) isoleucine level may be from about 26 to about 121  $\mu\text{mol/L}$  or about  $65 \pm 25 \mu\text{mol/L}$  (see e.g. Strauss, K. A., et al., 2020. Molecular genetics and metabolism, 129(3), pp. 193-206).

[0390] Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) levels of valine may be prevented from increasing or reduced. Suitably, serum (or plasma) valine levels may be reduced by at least 10%, at least 20%, at least 30%, at least 40%, or at least 50%. Suitably, following administration, the subject's serum (or plasma) valine level may be from about 300 to about 600  $\mu\text{mol/L}$ . Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) valine levels may be normalised. A normal serum (or plasma) valine level may be from about 118 to about 335  $\mu\text{mol/L}$  or about  $208 \pm 61 \mu\text{mol/L}$  (see e.g. Strauss, K. A., et al., 2020. Molecular genetics and metabolism, 129(3), pp. 193-206).

[0391] Suitably, following administration, the subject's serum (or plasma) leucine-to-valine concentration may be about 0.5 or less and/or the subject's serum (or plasma) leucine-to-isoleucine concentration may be about 2.0.

[0392] Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) levels of alloisoleucine may be prevented from increasing or reduced. Suitably, serum (or plasma) alloisoleucine levels may be reduced by at least 10%, at least 20%,

at least 30%, at least 40%, or at least 50%. Suitably, following administration, the subject's serum (or plasma) alloisoleucine level may be about 10 µmol/L or less. Following administration of the lentiviral vector of the present invention to a subject in need thereof, serum (or plasma) alloisoleucine levels may be normalised. A normal serum (or plasma) alloisoleucine level may be about <5 µmol/L. Serum (or plasma) alloisoleucine levels may be determined by any method known in the art, for example as described in Schadewaldt, P., et al., 1999. Clinical chemistry, 45(10), pp. 1734-1740 and Oglesbee, D., et al., 2008. Clinical chemistry, 54(3), pp. 542-54.

[0393] Following administration of the lentiviral vector of the present invention to a subject in need thereof the branched-chain alpha-ketoacid dehydrogenase (BCKD) activity in the liver of the subject may be improved and/or normalised. BCKD activity may be determined by any suitable method known in the art, for example BCKDC activity may be assayed spectrophotocochromically, as described in Chuang, J. L., et al., 2000. Methods in enzymology, 324, pp. 192-200.

#### Classic MSUD

[0394] Individuals with the classic form of MSUD may have <2% of BCKDC activity. Classic MSUD is typically presents in the neonatal period. Clinical features include maple syrup odour in cerumen and urine, irritability, poor feeding, lethargy, intermittent apnea, opisthotonus, and "bicycling" movements. Biochemical features include elevated BCAAs and alloisoleucine in plasma; and elevated branched-chain ketoacids in urine. Classic MSUD may be caused by defects in BCKDHA, BCKDHB, or DBT genes (see e.g. Blackburn, P. R., et al., 2017. The application of clinical genetics, 10, pp. 57-66).

[0395] In some embodiments, the MSUD is classic MSUD and the BCKDC subunit is BCKDE1A, or a fragment and/or variant thereof; BCKDE1B, or a fragment and/or variant thereof; or DBT, or a fragment and/or variant thereof. In some embodiments, the MSUD is classic MSUD and the BCKDC subunit is DBT, or a fragment and/or variant thereof.

[0396] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating classic MSUD.

[0397] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating classic MSUD.

[0398] In one aspect, the present invention provides a method of preventing or treating classic MSUD, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

#### Intermediate MSUD

[0399] Individuals with the intermediate form of MSUD may have up to about 30% BCKDC residual activity. Intermediate MSUD may present at any time of life, with decompensations occurring during periods of illness or

stress. Clinical features include maple syrup odour in cerumen and urine, feeding problems, poor growth, and developmental delay. Biochemical features are similar to classic MSUD but less severe. Intermediate MSUD may be caused by defects in BCKDHA, BCKDHB, or DBT genes (see e.g. Blackburn, P. R., et al., 2017. The application of clinical genetics, 10, pp. 57-66).

[0400] In some embodiments, the MSUD is intermediate MSUD and the BCKDC subunit is BCKDE1A, or a fragment and/or variant thereof; BCKDE1B, or a fragment and/or variant thereof; or DBT, or a fragment and/or variant thereof. In some embodiments, the MSUD is intermediate MSUD and the BCKDC subunit is DBT, or a fragment and/or variant thereof.

[0401] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating intermediate MSUD.

[0402] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating intermediate MSUD.

[0403] In one aspect, the present invention provides a method of preventing or treating intermediate MSUD, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

#### Intermittent MSUD

[0404] Intermittent MSUD may present at any time of life, with decompensations occurring during periods of illness or stress. Clinical features include normal growth and neurological development, however in stress situations, patients may present with encephalopathy. Biochemical features are normal BCAAs when well, but similar to classic MSUD during illness. Intermittent MSUD may be caused by defects in BCKDHA, BCKDHB, or DBT genes (see e.g. Blackburn, P. R., et al., 2017. The application of clinical genetics, 10, pp. 57-66).

[0405] In some embodiments, the MSUD is intermittent MSUD and the BCKDC subunit is BCKDE1A, or a fragment and/or variant thereof; BCKDE1B, or a fragment and/or variant thereof; or DBT, or a fragment and/or variant thereof. In some embodiments, the MSUD is intermittent MSUD and the BCKDC subunit is DBT, or a fragment and/or variant thereof.

[0406] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating intermittent MSUD.

[0407] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating intermittent MSUD.

[0408] In one aspect, the present invention provides a method of preventing or treating intermittent MSUD, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

### Thiamine-Responsive MSUD

[0409] Thiamine-responsive MSUD may present at any time of life, with decompensations occurring during periods of illness or stress. Clinical features are similar to intermediate MSUD. Biochemical features include improvement of leucine tolerance and levels of BCAAs when on thiamine supplementation. Thiamine-responsive MSUD may be caused by defects in the DBT gene (see e.g. Blackburn, P. R., et al., 2017. The application of clinical genetics, 10, pp. 57-66).

[0410] In some embodiments, the MSUD is thiamine-responsive MSUD and the BCKDC subunit is DBT, or a fragment and/or variant thereof.

[0411] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating thiamine-responsive MSUD.

[0412] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating thiamine-responsive MSUD.

[0413] In one aspect, the present invention provides a method of preventing or treating thiamine-responsive MSUD, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

### E3-Deficient MSUD

[0414] E3-deficient MSUD typically presents in the neonatal period. Clinical features include hypotonia, developmental delay, emesis, hepatomegaly, lethargy, seizures, spasticity, Leigh syndrome, and failure to thrive. Biochemical features include elevated BCAAs, alloisoleucine, lactate, pyruvate, and alanine in plasma; and elevated branched-chain ketoacids and  $\alpha$ -ketoglutarate in urine. E3-deficient MSUD may be caused by defects in the DLD gene.

[0415] In some embodiments, the MSUD is E3-deficient MSUD and the BCKDC subunit is DLD, or a fragment and/or variant thereof.

[0416] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating E3-deficient MSUD.

[0417] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating E3-deficient MSUD.

[0418] In one aspect, the present invention provides a method of preventing or treating E3-deficient MSUD, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

### MSUD Type 1A

[0419] MSUD type 1A is caused by a variety of mutations in BCKDHA, the gene coding for the alpha chain of the BCKDC E1 subunit (BCKDE1A).

[0420] In some embodiments, the MSUD is MSUD type 1A and the BCKDC subunit is BCKDE1A, or a fragment and/or variant thereof.

[0421] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating MSUD type 1A.

[0422] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating MSUD type 1A.

[0423] In one aspect, the present invention provides a method of preventing or treating MSUD type 1A, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

### MSUD Type 1B

[0424] MSUD type 1B is caused by a variety of mutations in BCKDHB, the gene coding for the beta chain of the BCKDC E1 subunit (BCKDE1B).

[0425] In some embodiments, the MSUD is MSUD type 1B and the BCKDC subunit is BCKDE1B, or a fragment and/or variant thereof.

[0426] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating MSUD type 1B.

[0427] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating MSUD type 1B.

[0428] In one aspect, the present invention provides a method of preventing or treating MSUD type 1B, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

### MSUD Type 2

[0429] MSUD type 2 is caused by a variety of mutations in DBT, the gene coding for the BCKDC E2 subunit (DBT).

[0430] In some embodiments, the MSUD is MSUD type 2 and the BCKDC subunit is DBT, or a fragment and/or variant thereof.

[0431] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating MSUD type 2.

[0432] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating MSUD type 2.

[0433] In one aspect, the present invention provides a method of preventing or treating MSUD type 2, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

### MSUD Type 3

[0434] MSUD type 3 is caused by a variety of mutations in DLD, the gene coding for the BCKDC E3 subunit (DLD).

[0435] In some embodiments, the MSUD is MSUD type 3 and the BCKDC subunit is DLD, or a fragment and/or variant thereof.

[0436] In one aspect, the present invention provides the lentiviral vector, cell or pharmaceutical composition according to the present invention for use in preventing or treating MSUD type 3.

[0437] In one aspect, the present invention provides use of the lentiviral vector, cell or pharmaceutical composition according to the present invention in the manufacture of a medicament for preventing or treating MSUD type 3.

[0438] In one aspect, the present invention provides a method of preventing or treating MSUD type 3, the method comprising administering a therapeutically effective amount of the lentiviral vector, cell or pharmaceutical composition according to the present invention to a subject in need thereof.

## EXAMPLES

[0439] Preferred features and embodiments of the invention will now be described by way of non-limiting examples.

### Example 1—Lentiviral Vector (LV) Encoding a DBT Transgene

[0440] We generated and produced a lentiviral vector (LV) encoding a containing DBT transgene encoding the E2 subunit of BCKDC and evaluated in vitro in hepatocyte cell lines and in vivo both in wild-type (WT) mice and in an intermediate mouse model of MSUD due to E2 deficiency, the iMSUD mice.

## Results

[0441] We generated and produced LV carrying human wild type or codon optimized DBT transgenes under the control of a hepatocyte-specific cassette based on the enhanced transthyretin promoter (ET.DBT) or the human alpha-1 antitrypsin promoter (hAAT.DBT) (FIG. 1a).

[0442] To compare between the different DBT transgenes, we transduced Huh7 cells with LV-ET.DBTwt and LV-ET.DBTco\_1 at increasing multiplicity of infection (MOI). We detected higher DBT expression in cells transduced with LV-ET.DBTco\_1 compared to LV-ET.DBTwt, as detected by western blot analysis (FIG. 1b). The same result was obtained by transducing Huh7 cells with LV-hAAT.DBTwt or LV-hAAT.DBTco\_1 at different MOI (FIG. 1c).

[0443] By DBT protein quantification and normalization on the GAPDH housekeeper, we confirmed a higher transgene expression (~7-8 fold) obtained from the expression cassette driven by the ET promoter compared to the hAAT and by the codon optimized transgene compared to the wt form at similar vector DNA copies per cell (vector copy number, VCN) (FIG. 1d-f).

[0444] To directly compare the in vivo liver transduction efficiency of the LV-ET.DBTwt and LV-ET.DBTco, we administered them to juvenile C57Bl/6 WT mice at a dose of  $2.5 \times 10^{10}$  transducing units (TU)/kg by i.v. injection. At the end of the experiment, 4 weeks post LV administration, the vector copy number (VCN) in the liver, was ~0.4 in both treated groups (FIG. 2a). The DBT mRNA was slightly higher in the group injected with LV-ET.DBTwt while the DBT protein output in the liver was slightly higher for the LV-ET.DBTco group (FIG. 2b-d).

[0445] We administered LV-ET.DBTco\_1 to juvenile iMSUD mice at a dose of  $5 \times 10^{10}$  transducing units (TU)/kg by i.v. injection. The treatment resulted in an improved survival in treated mice compared to knock out littermates (50% survival probability at day 111 and 35, respectively) (FIG. 3a). The measurement of circulating BCAAs and alloisoleucine showed that these biomarkers remained stable in LV-treated animals up week 12 post LV (FIG. 3b,c). At the last time points, these parameters could not be evaluated in iMSUD untreated animals, which had previously died.

## Materials and Methods

### Plasmid Construction

[0446] The DBT coding sequences (WT or codon optimized) used in the study were synthesized by GeneScript and cloned into a third-generation self-inactivating (SIN) LV transfer plasmid (Milani, M., et al., 2019. Science Translational Medicine, 11(493), p.eav7325) under the control of the enhanced transthyretin promoter (ET) or the human alpha-1 antitrypsin promoter (hAAT).

### Vector Production

[0447] Lab-grade VSV.G-pseudotyped third-generation SIN LV were produced by calcium phosphate transient transfection into 293T cells. 293T cells were transfected with a solution containing a mix of the selected LV genome transfer plasmid, the packaging plasmids pMDLg/pRRE and pCMV.REV, pMD2.G and pAdvantage, as previously described (Milani, M., et al., 2017. EMBO molecular medicine, 9(11), pp. 1558-1573; and Milani, M., et al., 2019. Science Translational Medicine, 11(493), p.eav7325).

[0448] Medium was replaced 14-16 hours post transfection and supernatant was collected around 30 hours after medium change. LV-containing supernatants were sterilized through a 0.22  $\mu$ m filter (Millipore) and transferred into sterile poliallomer tubes (Beckman) and centrifuged at 20,000 g for 120 min at 20° C. (Beckman Optima XL-100K Ultracentrifuge). LV pellet was resuspended in the appropriate volume of PBS to allow 500-1000x concentration.

### LV Titration

[0449] For LV titration,  $1 \times 10^5$  293T cells were transduced with serial LV dilutions in the presence of polybrene (8  $\mu$ g/ml). Genomic DNA (gDNA) was extracted 10 days after transduction, using Maxwell 16 Cell DNA Purification Kit (Promega), following manufacturer's instructions. VCN was determined by digital droplet PCR (ddPCR) starting from 5-20 ng of template gDNA using primers (HIV fw: 5'-T ACTGACGCTCTGCACC-3'; HIV rv: 5'-TCTCGACGCAAGGACTCG-3') and a probe (FAM 5'-ATCTCTCTCCTTAGCCTC-3') designed on the primer binding site region of LV. The amount of endogenous DNA was quantified by a primers/probe set designed on the human GAPDH gene (Applied Biosystems HS00483111\_cm). The PCR reaction was performed with each primer (900 nM) and the probe (250 nM) following manufacturer's instructions (Biorad), read with QX200 reader and analyzed with QuantaSoft software (Biorad). Infectious titer, expressed as TU/mL, was calculated using the formula  $TU/mL = (VCN \times 100000 \times (1/dilution\ factor))$ . LV physical

particles were measured by HIV-1 Gag p24 antigen immunocapture assay (Perkin Elmer) following manufacturer's instructions. LV specific infectivity was calculated as the ratio between infectious titer and physical particles.

#### Cell Culture and In Vitro Transduction Experiments

**[0450]** HuH7 cells were maintained under 37°C., 5% CO<sub>2</sub> condition in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% FBS (Thermo Fisher Scientific, Waltham, MA). Cells were seeded into 12-well plates (2×10<sup>5</sup> cells/well) and transduced with LV variants at different MOI in the presence of polybrene (8 µg/ml). Ten days post-LV transduction, cells were harvested for molecular analyses.

#### Mice Experiments

**[0451]** All animal experiments were performed in strict accordance with good animal practices following Italian and European legislation on animal care and experimentation (2010/63/EU). Wild-type C57Bl/6 or iMSUD juvenile mice were used in these studies. iMSUD is an intermediate MSUD mouse model, which is DBT KO and show low activity of BCKAD in the liver (Homanics, G. E., et al., 2006. BMC medical genetics, 7(1), pp. 1-13). Animals were administered with LV at a dose of 2.5-5×10<sup>10</sup> TU/kg via the retro-orbital plexus. Blood samples were collected monthly throughout the duration of the study for analysis of the metabolites. For the experiment conducted in WT C57Bl/6 mice, animals were killed 4 weeks post LV administration and liver samples were collected and snap-frozen for vector copy number, mRNA and protein analyses.

#### RNA Analysis

**[0452]** RNA samples were extracted from mouse livers using Maxwell 16 simplyRNA Tissue Kit (Promega). cDNA was synthesized starting from 1 µg of total RNA using the SuperScript IV Vilo Master Mix (Invitrogen) according to manufacturer's instructions. LV gene expression was assessed by ddPCR starting from 25-50 ng of template cDNA using a primers/probe set designed on the WPRE region of LV (WPRE: primer fw 5'-GGCTGTTGGGCACTGACAAT-3'; primer rv 5'-ACGTCCCCGCGCAGAACATC-3'; probe FAM 5'-TTTCCTTGGCTGCTGCCCTGTGT-3' NGB). Murine HPRT was used as reference gene (Bio Rad, Mmu 10031256 dMmu CPE5095493). The PCR reaction was performed with each primer (900 nM) and the probe (250 nM) following manufacturer's instructions (Biorad), read with QX200 reader and analyzed with QuantaSoft software (Biorad).

#### VCN Determination

**[0453]** DNA was extracted from cells or liver samples using Maxwell 16 Cell or Tissue DNA Purification Kits (Promega). VCN was determined in Huh7 samples as described above (see "LV titration"). VCN in murine DNA was determined by ddPCR, starting from 5-20 ng of template gDNA using a primers/probe set designed on the primer binding site region of LV (see "LV titration" above).

The amount of endogenous murine DNA was quantified by a primers/probe set designed on the murine Sema3a gene (Sema3A fw: 5'-ACCGAITCCAGATGATTGGC-3'; Sema3A rv: 5'-TCCATATTAATGCAGTGCTTGC-3'; Sema3A probe: HEX 5'-AGAGGCCTGTCCTGCAGCTC

**[0454]** ATGG-3' BHQ1). The PCR reaction was performed with each primer (900 nM) and the probe (250 nM) following manufacturer's instructions (Biorad), read with QX200 reader and analyzed with QuantaSoft software (Biorad).

#### Western Blot Analysis

**[0455]** Protein extracts from HuH7 cell or from mouse liver samples were prepared using RIPA buffer (EMD Millipore) and protease inhibitors (Roche). Protein concentration was determined using the DC Protein Assay (Bio Rad Laboratories). SDS-page electrophoresis was performed in a Bis-Tris 4-12% gradient polyacrylamide gel. After transfer, the membrane was blocked with 1×TBS+5% milk and incubated with an anti-DBT antibody (Sigma-Aldrich HPA026481) or anti-β actin (Sigma Aldrich A2228). The membrane was incubated with an anti-Mouse and an anti-Rabbit IgG secondary antibodies (Jackson Immunoresearch 115-035-003 and 111-035-144), then with clarity Western ECL substrate (Bio Rad) and visualized by Uvitec Imaging System (Cleaver Scientific). For Western blot quantification, we used the Image J software.

#### Blood BCAA and Alloisoleucine Determination

**[0456]** Blood samples were collected via the retro-orbital plexus in 0.5M EDTA-filled tubes. Dried blood spots (DBS) were obtained from spotted blood on filter paper (903; Whatman GmbH, Dassel, Germany) after withdrawal of the samples. BCAA and alloisoleucine measurement in DBS was performed by liquid chromatography-tandem mass spectrometry, as previously described.

#### Statistical Analysis

**[0457]** Statistical analyses were performed by using Prism 9 software. Comparison of survival curves was performed by applying a Long-rank (Mantel-Cox) test.

**[0458]** All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the disclosed vectors, cells, compositions, kits and uses of the invention will be apparent to the skilled person without departing from the scope and spirit of the invention. Although the invention has been disclosed in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the disclosed modes for carrying out the invention, which are obvious to the skilled person are intended to be within the scope of the following claims.

## SEQUENCE LISTING

Sequence total quantity: 58

SEQ ID NO: 1 moltype = AA length = 305  
 FEATURE Location/Qualifiers  
 source 1..305  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 1  
 MWPLVAALLL GSACCGSAQL LFNKTKSVEF TFCNDTVVIP CFVTNMEAQN TTEVYVKWF 60  
 KGRDIYTFDG ALNKSTVPTD FSSAKIEVSQ LLKGDAASLMK DKSDAVSHG NYTCEVTELT 120  
 REGETIIELK YRVVSWFSPN ENILIVIFPI FAILLFWGQF GIKTLKYRSG GMDEKTIALL 180  
 VAGLVITVIV IVGAILFVPG EYSLNATGL GLIVTSTGIL ILLHYYVFST AIGLTSFVIA 240  
 ILVIQVIAYI LAVVGLSLCI AACIPMHGPL LISGLSILAL AQLLGLVYMK FVASNQKTIQ 300  
 PPRNN 305

SEQ ID NO: 2 moltype = AA length = 323  
 FEATURE Location/Qualifiers  
 source 1..323  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 2  
 MWPLVAALLL GSACCGSAQL LFNKTKSVEF TFCNDTVVIP CFVTNMEAQN TTEVYVKWF 60  
 KGRDIYTFDG ALNKSTVPTD FSSAKIEVSQ LLKGDAASLMK DKSDAVSHG NYTCEVTELT 120  
 REGETIIELK YRVVSWFSPN ENILIVIFPI FAILLFWGQF GIKTLKYRSG GMDEKTIALL 180  
 VAGLVITVIV IVGAILFVPG EYSLNATGL GLIVTSTGIL ILLHYYVFST AIGLTSFVIA 240  
 ILVIQVIAYI LAVVGLSLCI AACIPMHGPL LISGLSILAL AQLLGLVYMK FVASNQKTIQ 300  
 PPRKAVEEPL NAFKESKGMM NDE 323

SEQ ID NO: 3 moltype = AA length = 292  
 FEATURE Location/Qualifiers  
 source 1..292  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 3  
 MWPLVAALLL GSACCGSAQL LFNKTKSVEF TFCNDTVVIP CFVTNMEAQN TTEVYVKWF 60  
 KGRDIYTFDG ALNKSTVPTD FSSAKIEVSQ LLKGDAASLMK DKSDAVSHG NYTCEVTELT 120  
 REGETIIELK YRVVSWFSPN ENILIVIFPI FAILLFWGQF GIKTLKYRSG GMDEKTIALL 180  
 VAGLVITVIV IVGAILFVPG EYSLNATGL GLIVTSTGIL ILLHYYVFST AIGLTSFVIA 240  
 ILVIQVIAYI LAVVGLSLCI AACIPMHGPL LISGLSILAL AQLLGLVYMK FVASNQKTIQ 292

SEQ ID NO: 4 moltype = AA length = 311  
 FEATURE Location/Qualifiers  
 source 1..311  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 4  
 MWPLVAALLL GSACCGSAQL LFNKTKSVEF TFCNDTVVIP CFVTNMEAQN TTEVYVKWF 60  
 KGRDIYTFDG ALNKSTVPTD FSSAKIEVSQ LLKGDAASLMK DKSDAVSHG NYTCEVTELT 120  
 REGETIIELK YRVVSWFSPN ENILIVIFPI FAILLFWGQF GIKTLKYRSG GMDEKTIALL 180  
 VAGLVITVIV IVGAILFVPG EYSLNATGL GLIVTSTGIL ILLHYYVFST AIGLTSFVIA 240  
 ILVIQVIAYI LAVVGLSLCI AACIPMHGPL LISGLSILAL AQLLGLVYMK FVASNQKTIQ 300  
 PPRKAVEEPL N 311

SEQ ID NO: 5 moltype = AA length = 287  
 FEATURE Location/Qualifiers  
 source 1..287  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 5  
 QLLFNKTKSV EFTFCNDTVV IPCFVTNMEA QNTTEVYVKW KFKGRDIYTF DGALNKSTVP 60  
 TDFSSAKIEV SQQLGDAASL KMDKSDAVSH TGNYTCEVTE LTREGETIIE LKYRVVSWFS 120  
 PNENILIVIF PIFAILFWG QFGIKTLKYR SGGMDEKTIA LLVAGLVITV IVIVGAILFV 180  
 PGHEYSLKNAT GLGLIVTSTG ILLILLHYYVF STAIGLTSFV IAILVIQVIA YILAVVGLSL 240  
 CIAACIPMHG PLLISGLSIL ALAQLLGLVY MKPVASNQKT IQPPRNN 287

SEQ ID NO: 6 moltype = AA length = 305  
 FEATURE Location/Qualifiers  
 source 1..305  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 6  
 QLLFNKTKSV EFTFCNDTVV IPCFVTNMEA QNTTEVYVKW KFKGRDIYTF DGALNKSTVP 60  
 TDFSSAKIEV SQQLGDAASL KMDKSDAVSH TGNYTCEVTE LTREGETIIE LKYRVVSWFS 120  
 PNENILIVIF PIFAILFWG QFGIKTLKYR SGGMDEKTIA LLVAGLVITV IVIVGAILFV 180  
 PGHEYSLKNAT GLGLIVTSTG ILLILLHYYVF STAIGLTSFV IAILVIQVIA YILAVVGLSL 240  
 CIAACIPMHG PLLISGLSIL ALAQLLGLVY MKPVASNQKT IQPPRKAEE PLNAFKESKG 300

-continued

---

MMNDE	305
SEQ ID NO: 7	moltype = AA length = 274
FEATURE	Location/Qualifiers
source	1..274
	mol_type = protein
	organism = unidentified
SEQUENCE: 7	
QLLFNKTCSV EFTFCNDTVV IPCFVNMEA QNTTEVYVKW KFKGRDIYTF DGALNKSTVP	60
TDFSSAKIEV SQLLGKDASL KMDKSDAVSH TGNYTCEVTE LTREGETIIE LKYRVVSWFS	120
PENENILIVIF PIFAILLFWG QFGIKTLKVR SGGMDEKTIA LLVAGLVITV IVIVGAILFV	180
PGEYSLKNAT GLGLIVTSTG ILILLHHYVF STAIGLTSFV IAILVIQVIA YILAVVGLSL	240
CIAACIPMHG PLLISGLSIL ALAQLLGLVY MKFV	274
SEQ ID NO: 8	moltype = AA length = 293
FEATURE	Location/Qualifiers
source	1..293
	mol_type = protein
	organism = unidentified
SEQUENCE: 8	
QLLFNKTCSV EFTFCNDTVV IPCFVNMEA QNTTEVYVKW KFKGRDIYTF DGALNKSTVP	60
TDFSSAKIEV SQLLGKDASL KMDKSDAVSH TGNYTCEVTE LTREGETIIE LKYRVVSWFS	120
PENENILIVIF PIFAILLFWG QFGIKTLKVR SGGMDEKTIA LLVAGLVITV IVIVGAILFV	180
PGEYSLKNAT GLGLIVTSTG ILILLHHYVF STAIGLTSFV IAILVIQVIA YILAVVGLSL	240
CIAACIPMHG PLLISGLSIL ALAQLLGLVY MKFVASNQKT IQPPRKAVEE PLN	293
SEQ ID NO: 9	moltype = AA length = 140
FEATURE	Location/Qualifiers
source	1..140
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 9	
MWPLVAALLL GSACCGSAQL LFNKTKSVEF TFCNDTVVIP CFVTNMEAQN TTEVYVKWF	60
KGRDIYTFDG ALNKSTVPTD FSSAKIEVSQ LLKGDASLKM DKSDAVSHTG NYTCEVTELT	120
REGETIIELK YRVVSWFSPN	140
SEQ ID NO: 10	moltype = AA length = 122
FEATURE	Location/Qualifiers
source	1..122
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 10	
QLLFNKTCSV EFTFCNDTVV IPCFVNMEA QNTTEVYVKW KFKGRDIYTF DGALNKSTVP	60
TDFSSAKIEV SQLLGKDASL KMDKSDAVSH TGNYTCEVTE LTREGETIIE LKYRVVSWFS	120
PN	122
SEQ ID NO: 11	moltype = AA length = 365
FEATURE	Location/Qualifiers
source	1..365
	mol_type = protein
	organism = Homo sapiens
SEQUENCE: 11	
MAMVAPRTLL LLLSGALALT QTWAGSHSMR YFFTSVSRPG RGEPRFIAVG YVDDTQFVRF	60
DSDAASQRME PRAPWIEQEG PEYWDQETRN VKAQSQTRDV DLGTLRGYYN QSEAGSHTIQ	120
IMYGCDVGSQ CRFLRGYRQD AYDGKDYLAL NEDLRSWTAA DMAAQITKRK WEAHAAEQL	180
RAYLEGRCVE WLRRYLENGK ETLQRTDPPK THMTHHPISD HEATLRCWAL GFYPAEITLT	240
WQRDGEDQTO DTELVETRPA GDGTQKWA VVPSGEBQR YTCHVQHEGL PKPLTLRWEI	300
SSQPTIPIVG IIAGLVLLGA VITGAVVAAV MWRRKSSDRK GGSYTQAASS DSAQGSDVSL	360
TACKV	365
SEQ ID NO: 12	moltype = AA length = 371
FEATURE	Location/Qualifiers
source	1..371
	mol_type = protein
	organism = Homo sapiens
SEQUENCE: 12	
MAMVAPRTLL LLLSGALALT QTWAGSHSMR YFFTSVSRPG RGEPRFIAVG YVDDTQFVRF	60
DSDAASQRME PRAPWIEQEG PEYWDQETRN VKAQSQTRDV DLGTLRGYYN QSEAGSHTIQ	120
IMYGCDVGSQ CRFLRGYRQD AYDGKDYLAL NEDLRSWTAA DMAAQITKRK WEAHAAEQQ	180
RAYLEGRCVE WLRRYLENGK ETLQRTDPPK THMTHHPISD HEATLRCWAL GFYPAEITLT	240
WQRDGEDQTO DTELVETRPA GDGTQKWA VVPSGEBQR YTCHVQHEGL PKPLTLRWEI	300
SSQPTIPIVG IIAGLVLLGA VITGAVVAAV MWRRKSSGGE GVKDRKGGSY TQAASSDSAQ	360
GSDVSLTACK V	371
SEQ ID NO: 13	moltype = AA length = 362
FEATURE	Location/Qualifiers

- continued

---

-continued

---

```

gcatcagca ggttagctcta gaggatcccc gtctgtctgc acatttcgta gagcgagtgt 360
tccgatactc taatccccc aggaagggtt catattttgt taggttactt atttccttt 420
tgttgactaa gtcaataatc agaatcagca gggttggagt cagctggca gggatcagca 480
gcctgggtg gaaggagggg gtataaaagc cccttcacca ggagaagccg tcacacagat 540
ccacaagctc ctg 553

SEQ ID NO: 20 moltype = DNA length = 347
FEATURE Location/Qualifiers
source 1..347
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 20
gatcttgcata ccagtggAAC agccactaAG gattctgcAG tgagAGcAGA gggccAGCTA 60
agtggtaACTC tcccAGAGAC tggctgtACTC acggccACCC CTCACCTG gacacAGGAC 120
gctgtggTTT ctgAGCAGGAG tacaATGACT ctttCGGTAG aGTCAGTGG aAGCTGTACA 180
ctgcccAGGC aaAGCCTCGG ggcAGCGTAG qGGGGCAGACT cAGATCCAG ccAGTGGACT 240
tagccccCTGT ttgctctcc gataACTGGG gtgacCTGG ttaatATTCA ccAGCAGCCT 300
ccccCGTGC ccctctGGAT ccACTGCTTA aatacGGACG aggACAG 347

SEQ ID NO: 21 moltype = DNA length = 504
FEATURE Location/Qualifiers
source 1..504
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 21
aatcaacCTC tggattacAA aattttgtgAA agatttGACT gtattttAA ctatgttgCT 60
cctttacgc tATGTTGATA cgctgttCA atgcctttG ATCATGOTAT tgctccCGT 120
atggcTTTCA ttttCTCCTC ctgttataAA tccTGGTGC tGTCCTTCA tgaggAGTTG 180
tggccCGTTG tcAGGCAACG tggcgtggTG tGACTGTGTT tgctgacGC aACCCCACT 240
gggtggggCA ttggcaccAC ctgtcAGCTC ctttccGGGA ctttcgCTT cccctCCCT 300
attggcACCTG cggAAACTCAT cgccgCCTG cttggccGGT GTCGGACAGG ggcteggtG 360
ttggcactG acaattCCGT ggttgtGCG gggAAACTCAT cgtcTTTCC ttggctGCTC 420
gcctgttg CCACCTGGAT tctgacGGG acgtcCTTCT gctacgtCCC ttggccCTC 480
aatccAGCGG accttcCTTC ccgc 504

SEQ ID NO: 22 moltype = length =
SEQUENCE: 22
000

SEQ ID NO: 23 moltype = DNA length = 235
FEATURE Location/Qualifiers
source 1..235
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 23
tggaaaggGCT aatttcactCC caacGAAGAC aagatctGCT ttttGCTTGT actgggtCTC 60
tctggtaga ccagatCTGA gcctGGGAGC tctctGGCTA actAGGGAAc ccactGCTTA 120
agcctcaata aagcttgcCT tgagtGCTtA aagttagtGtG tgccCGtCTG ttgtgtGACT 180
ctggtaacta gagatCCCTT AGCAGTGTG gaaaatcttC agcAG 235

SEQ ID NO: 24 moltype = DNA length = 182
FEATURE Location/Qualifiers
source 1..182
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 24
gggtctctCT ggttagACCA gatctgAGCC tgggAGGCTC ctggctAACT agggAACCCA 60
ctgcttaAGC ctcaataAAAG cttgccttGA gtgttcaAG tagtGtGtGc ccgtctGTTG 120
tGtGACTCTG gtaactAGAG atccctcAGA ccctttAGT cAGtGtGAA aatctctAGC 180
ag 182

SEQ ID NO: 25 moltype = DNA length = 235
FEATURE Location/Qualifiers
source 1..235
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 25
tggaaaggGCT aatttcactCC caacGAAGAC aagatctGCT ttttGCTTGT actgggtCTC 60
tctggtaga ccagatCTGA gcctGGGAGC tctctGGCTA actAGGGAAc ccactGCTTA 120
agcctcaata aagcttgcCT tgagtGCTtA aagttagtGtG tgccCGtCTG ttgtgtGACT 180
ctggtaacta gagatCCCTT AGCAGTGTG gaaaatcttC agcAG 235

SEQ ID NO: 26 moltype = DNA length = 154
FEATURE Location/Qualifiers
source 1..154
mol_type = other DNA

```

-continued

organism = synthetic construct

SEQUENCE: 26  
tggcgcggca acagggactt gaaagcgaaa gggaaaccag aggagcttc tcgacgcagg 60  
actcggttt ctgaagcgcg cacggcaaga ggcgaggggc ggcgacttgt gatgtacgcca 120  
aaaaattttta cttagcgagg cttagaggag agag 154

SEQ ID NO: 27 moltype = DNA length = 151  
FEATURE Location/Qualifiers  
source 1..151  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 27  
tggcgcggca acagggactt gaaagcgaaa gggaaaccag agctctctcg acgcaggactt 60  
cgccctgtctt aagcgccgac ggcgaaggc gagggggggc gactgttgat taccgcggaa 120  
attttgacta gcggaggcta gaaggagaga g 151

SEQ ID NO: 28 moltype = DNA length = 365  
FEATURE Location/Qualifiers  
source 1..365  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 28  
atgggtgcga gagcgtcagt attaagcggtt ggagaatttt atcgcgatgg gaaaaatttt 60  
gggttaaggccg agggggaaag aaaaatata aattttaaaata tatagttatgg gcaagcagg 120  
agctagaacg attcgcgtt aatcttgcgc ttgttagaaac atcagaaggc tgtagacaaa 180  
tactggacca gctacaacca tcccttcaga caggatcaga agaactttaga tcattttata 240  
atacgtatgc aaccctctat tttgtgtcatc aaaggataga gataaaaagac accaaggaaag 300  
ctttagacaa gatagaggaa gagcaaaaca aaagtaagac caccgcacag caaggcccg 360  
ctgtat 365

SEQ ID NO: 29 moltype = DNA length = 241  
FEATURE Location/Qualifiers  
source 1..241  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 29  
ggagctttgt tccttgggtt cttgggagca gcaggaaagca ctatgggcgc agcgtaatgt 60  
acgctgcggc tacaggccag acaatttttgc ttgttatag tgcaaggcaca gaacaatttt 120  
ctggggctta ttgaggccca acagcatgtt tcgtcaactca cagtttttttgc catcaagg 180  
ctccaggccaa gaatcttgcgc ttgtggaaaga tacctaaagg atcaacagct cttggggattt 240  
t 241

SEQ ID NO: 30 moltype = DNA length = 241  
FEATURE Location/Qualifiers  
source 1..241  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 30  
ggagctttgt tccttgggtt cttgggagca gcaggaaagca ctatgggcgc agcctcaatgt 60  
acgctgcggc tacaggccag acaatttttgc ttgttatag tgcaaggcaca gaacaatttt 120  
ctggggctta ttgaggccca acagcatgtt tcgtcaactca cagtttttttgc catcaagg 180  
ctccaggccaa gaatcttgcgc ttgtggaaaga tacctaaagg atcaacagct cttggggattt 240  
t 241

SEQ ID NO: 31 moltype = DNA length = 124  
FEATURE Location/Qualifiers  
source 1..124  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 31  
aactttaaa agaaaagggg ggattttttt gtacagtgc gggggaaagaa tagtagacat 60  
aatagcaaca gacatacaaaa cttaaaattt acaaaaacaa attacaaaaaa ttcaaaattt 120  
tatac 124

SEQ ID NO: 32 moltype = DNA length = 146  
FEATURE Location/Qualifiers  
source 1..146  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 32  
tcttcagacc tggaggaggg gatatgggg acaattgggg aagtgaatta tataaatata 60  
aagttagaaa aattgaacca tttaggttag cacccaccaa ggcaaaagaga agagttgtgc 120  
agagagaaaa aagagcgtt ggaata 146

SEQ ID NO: 33 moltype = DNA length = 468  
FEATURE Location/Qualifiers

- continued

source 1..468  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 33  
gggttgcctt ggaaaaactca tttgcaccac tgctgtgcct tggaaatgcta gttggagtaa 60  
taaatctctg gaacagattt ggaatcacac gacctggatg gagttggaca gagaattaa 120  
caattacaca agcttaatc actccctaat tgaagaatcg caaaaccgc aaaaaaggaa 180  
tgaacaagaa ttatggat tagataatcg ggcaagtgg tggaaatgg ttaacataac 240  
aaatggctg tggtatataa aattatccat aatgatgatgaa ggaggatgg taggtttaa 300  
aatagttttt gctgtactt ctatagtgaa tagagtttag cagggatatt caccatttc 360  
gttcagacc caccccccac ccccgagggg acccgacagg cccgaaggaa tagaagaaga 420  
aggtggagag agagacagag acagatccat tcgattatgt aacggatc 468

SEQ ID NO: 34 moltype = DNA length = 902  
FEATURE Location/Qualifiers  
source 1..902  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 34  
tggcgccccgaa acaggggacact gaaagcgaaa gggaaaccag agctctctcg accgaggact 60  
cggttgcgtg aagcgcgcac ggcaagggc gagggggggc gacttggtag tacgcacaaa 120  
attttgcata gggggggcgtt gaaggagaga gatgggtgcg agaggctcg tattaagcg 180  
gggagaattt gatcgcgcgtt ggaaaaattt cggttaaggc cagggggaaa gaaaaaatat 240  
aaattaaac atatgtatg ggcgaaggcgg gagctagaac gattcgact taatctggc 300  
ctgttagaaaa catcagaagg ctgttagacaa atactggac agctacaacc atcccttcag 360  
acaggatcag aagaacttag atcattatata aatagatcg caaccctcta ttgttgcata 420  
caaaggatag agataaaaaa caccacggaa gctttagaca agatagagga agagcaaaac 480  
aaaatgttggaa ccacccgcaca gcaagcgccc gctgttgc ttttccatggag gaggat 540  
ggggacataat ttggggatgtt aattatataa atataaagta gtaaaaaattt aaccattagg 600  
agtagcacc accaaggccaa agagaagat ggttgcacgaaa gaaaaaaaggggatcgttggat 660  
aggagctttt ttccttgggt tcttggggc agcaggaaac actatggcg cagccctcaat 720  
gacgctgacg gtacaggccaa gacaattttt gtctgttata gtgcacgcgg agaaacattt 780  
gctgggggtt attggggcgc aacagcatct gttcaactc acagtctgg gcatcaagca 840  
gttccaggca agaatccctgg ctgtggaaag atacctaag gatcaacagc tcctggggat 900  
ttt 902

SEQ ID NO: 35 moltype = DNA length = 1510  
FEATURE Location/Qualifiers  
source 1..1510  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 35  
tggcgccccgaa acaggggacact gaaagcgaaa gggaaaccag agctctctcg accgaggact 60  
cggttgcgtg aagcgcgcac ggcaagggc gagggggggc gacttggtag tacgcacaaa 120  
attttgcata gggggggcgtt gaaggagaga gatgggtgcg agaggctcg tattaagcg 180  
gggagaattt gatcgcgcgtt ggaaaaattt cggttaaggc cagggggaaa gaaaaaatat 240  
aaattaaac atatgtatg ggcgaaggcgg gagctagaac gattcgact taatctggc 300  
ctgttagaaaa catcagaagg ctgttagacaa atactggac agctacaacc atcccttcag 360  
acaggatcag aagaacttag atcattatata aatagatcg caaccctcta ttgttgcata 420  
caaaggatag agataaaaaa caccacggaa gctttagaca agatagagga agagcaaaac 480  
aaaatgttggaa ccacccgcaca gcaagcgccc gctgttgc ttttccatggag gaggat 540  
ggggacataat ttggggatgtt aattatataa atataaagta gtaaaaaattt aaccattagg 600  
agtagcacc accaaggccaa agagaagat ggttgcacgaaa gaaaaaaaggggatcgttggat 660  
aggagctttt ttccttgggt tcttggggc agcaggaaac actatggcg cagccctcaat 720  
gacgctgacg gtacaggccaa gacaattttt gtctgttata gtgcacgcgg agaaacattt 780  
gctgggggtt attggggcgc aacagcatct gttcaactc acagtctgg gcatcaagca 840  
gttccaggca agaatccctgg ctgtggaaag atacctaag gatcaacagc tcctggggat 900  
tttgggttgtt tcttggaaac tcatttgcac cactgtgtg ccttggatgt ctgttggag 960  
taataatctt ctggaaacaa ttttggatca caccatctgg atgggtggg acagagaaat 1020  
taacaattac acaagcttta tacatctttt aatttggaaat tcgcaaaacc agcaaaaaaa 1080  
gaatgaacaa gaattttttt aatttggaaat atggggaaat ttgttggaaat ggtttaacat 1140  
aacaattttt ctgtggata taaaattttt cataatgtat gtagggggat ttgttggggat 1200  
agaatgttggat ttgttggatcc ttcttataatgtt gaatgttggat aggccggat attcattt 1260  
atcggttgcg accccatctttt ccaccccccgg gggacccggc agggccggc gatcaagca 1320  
agaaggttggaa gagagagaca gagacagatc catcgatca gtagacggat ctgcacggta 1380  
tcgggttact tttaaaagaa aaggggggat tggggggatc agtgcagggg aaagaatgtt 1440  
agacataata gcaacagaca tacaaactaa agaattacaa aaacaattaa caaaaattca 1500  
aaattttatac 1510

SEQ ID NO: 36 moltype = DNA length = 4699  
FEATURE Location/Qualifiers  
source 1..4699  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 36  
gggttctctt ggttagacca gatctgaccc tggggatctt ctggctaaat agggaaaccca 60

-continued

---

ctgtttaagc	ctcaataaaag	cttgcccttga	gtgtttcaag	tagtgtgtgc	ccgtctgttg	120
tgtgacttgc	gtaaatagag	atccctcaga	cccttttagt	cagtgtggaa	aatctcttagc	180
agtggcgccc	gaacaggggac	ctgaaagcga	aaggaaaacc	agactctct	cgacgcagga	240
ctcggcttgc	tgaagcgcgc	acggcaagag	gcgagggggc	gcgactgtgt	agtacgccaa	300
aaattttgac	tagcggaggc	tagaaggaga	gagatgggtg	cgagagcgtc	agtattaagc	360
gggggagaat	tagatcgca	tggaaaaaaa	tccggtaaag	gccaggggga	aagaaaaaat	420
ataaaattaaa	acatatagta	tggcaagca	gggagctaga	acgatcgca	gttaatcctg	480
gcctgttaga	aacatcagaa	ggctgttaga	aaatactggg	acagctaca	ccatcccttc	540
agacaggatc	agaagaactt	agatcattat	ataatacagt	agcaaccctc	tattgtgtgc	600
atcaaaggat	agagataaaa	gacaccaagg	aqgttttaga	caagatagag	gaagagcaaa	660
acaaaaaggtaa	gaccacgcga	cageaaggcg	ccgctgatct	tcaagcttgg	aggaggagat	720
atgagggaca	atggagaag	tgaattatata	aaatataaaag	tagtaaaaat	tgaaccattta	780
ggagtagcac	ccaccaaggc	aaagagaaga	gtgtgtcaga	gagaaaaaaag	agcagtggga	840
ataggaggott	tgttcttggc	gttcttggga	gcagcaggaa	gcactatggg	cgccgcctca	900
atgacgtcga	cggtacaggc	cagacaatta	ttgtcttggta	tagtgcaga	gcagaacaat	960
ttctgtgggg	ctatttggc	gcaacagcat	cttgcacac	tcacagtctg	gggcacatca	1020
cagctccagg	caagaatcct	ggctgtggaa	agataacctaa	aggatcaaca	gctcttgggg	1080
atttgggggtt	gtctcgaaat	actcttttgc	accactgtct	tgcccttggaa	tgcttagttgg	1140
agtaataat	acatacgaa	gatttggaa	cacacgcac	ggatgggtg	ggacagagaa	1200
attaacaattt	acacaaggctt	ataatactcc	ttaatttgaag	aatcgcaaaa	ccagcaagaa	1260
aagaatgaac	aagaattattt	ggaatttagat	aatgggca	gtttgtggaa	ttgggttaac	1320
ataacaattt	ggctgtggta	tataaaattt	ttcataatga	tagttagggg	tttggtaggt	1380
ttaagaatag	ttttttgtgt	actttctata	tgtaatagag	tttaggcaggg	atattcacca	1440
ttatcgttt	agacccaccc	cccaaaaa	ccggggcccc	acaggccccg	aggaatagaa	1500
gaagaaggtg	gagagagaga	cagagacaga	tccatcgtat	tagtgaacgg	atctcgacgg	1560
tatcggttaa	ctttttaaaag	aaaagggggg	attgggggtt	acagtgcagg	ggaaagaata	1620
gtagacataa	tagcaacaga	catacaaaat	aaagaatttac	aaaacaataat	tacaaaaatt	1680
caaaaattttt	tgcatacgc	gactggctc	gagcacgcga	gttaattata	accagcgcgg	1740
gccaaataaa	taatccgcga	ggggcagggt	acgtttgcc	agcgcgcgct	ggttaattat	1800
aacctcgcga	atattgttac	gaggccgcga	ttggccgaat	cgcgaggggc	aggtgacctt	1860
tgcccgacgc	cggttgcggcc	cgccccggd	ggtagtgcata	agcttagggag	tttgggtctc	1920
aggtcgaggg	cactggggg	atgttgat	agatggaaaa	ctactgtata	cccttgcaga	1980
gacagagttat	taggacatgt	ttgaacaggg	gccccggat	cagcaggtag	ctctagagga	2040
tccccgtctg	tctgcacatt	tcgttagagcg	agtgttccga	tactctaata	tcccttaggca	2100
aggttcatat	tttgtttaggt	tacttattatc	cctttttgttgc	actaagtccaa	taatcagaat	2160
cagcagggtt	ggagtgcgt	ttggcagggt	cagcgcgttgc	ggtttggaa	agggggtata	2220
aaagccccctt	caccaggaga	agccgtcaca	cajatccaca	agctcttgc	tagcgtacgc	2280
caccatggca	gccgtgagga	tgctgaggac	ctggagccgg	aaacgcaggca	agctgtatcg	2340
cgttgagatac	tttcaagacat	gttggacatgt	gcacgtgtct	aaccccaatt	acgtgtgtct	2400
cttcggctac	cccttcata	agtttctca	ccctcaccac	ttttctgaaga	ccacagccgc	2460
cctcgagggg	cagggtgtgc	agttaaagct	gagcgcacatc	ggcgaggggc	tccgcgagggt	2520
gaccgtgaag	gagtgttacg	tgaaggaggg	cgacacatgt	agccagtttg	attccatctg	2580
tgaggctgcac	tctgcacagg	ccaggctgtac	catcacatcc	cggtacgtat	gctgtatcaa	2640
gaagctgtac	tataaacctgg	acgacatcg	ctatgtggc	aagccacttgg	ttggacatcg	2700
gacagaggccc	ctgaaggaca	gcccgggg	ttgtgtggaa	acacccccc	ttgtccacacg	2760
tgagcacaca	caccaggaga	tcaagggaa	gaagaccctg	gccacaccag	ccgtgcggag	2820
actggccatgt	gagaacataa	tcaactgtac	cgaggtgtgg	ggatccggca	aggacggcag	2880
aatccctgaag	gaggacatcc	tgaactatct	ggagaagcag	acccggccaa	tcctgcacc	2940
ttccccctaa	gttggagatc	tgcacccccc	tccaaagccct	aaggacatga	cagtgcacat	3000
cctgggtct	aagcccccc	tgttcacccg	caaggataa	acagagcc	tcaagggttt	3060
tcagaagggcc	atgggtgaaa	ccatggacgc	cgccctgaat	atcccacat	tccgtactgt	3120
cgacgagatc	gatctgacac	agctgtgtaa	gttgggtggaa	gtctgcgggg	gagctgaacgc	3180
cggccaggcc	atcaagctgt	cttttatcgc	tttttttctgt	aaggccgcct	ctctggccct	3240
gctgcagttt	cttatctga	acgcctctgt	ggacgagaac	tgccagaata	tcacatataa	3300
ggccaggccac	aatatccgcga	tcgcctat	tgacagac	ggctctgtatcg	tgccaaacgt	3360
gaagaatgtt	cgatgtttgc	ccatcttgc	catgcgcac	gagctgaaca	ggctgcagaa	3420
gtctggctct	gttggggccgc	tgagcaccac	agatctgacc	ggccgcacat	tcacatgttc	3480
caataatcgcc	tctatcgcc	gcacatttgc	caagccgttgc	atcatgcac	cagaggtggc	3540
aatccggccc	ctgggctcta	tcaaggccat	ccctcgcttc	aaccagaagg	gcgagggtgt	3600
caaggcccg	atcatgtat	tgatgtatgt	cgccgcacac	agatgtatcg	atgggcac	3660
catgtctgc	ttcagcaacc	ttgtgtggatc	ctatctggat	aatccgcct	ttatgtgtct	3720
ggatctgtac	tgatgtatgt	actcgacaa	caacctctgg	attacaaaat	ttgtgaaaga	3780
tttactgttgc	ttcttaacta	ttgtgttgcct	tttacgtat	gtggatcgc	tgctttaatg	3840
cccttgcata	atgttatttc	ttccctgtat	ttcccttcctt	gtataaaatcc	3900	
ttgttgtgtt	ctctttatgt	ggagttgtgt	ccctgtgtca	ggcaacgtgg	ctgtgtgtgc	3960
actgtgttttgc	tcgacgcac	ccccactgtt	ttggggcatttgc	ccaccacatgc	tcagctctt	4020
tcggggactt	tcgttttccc	cctcccttatt	gccacggccgg	aactcatcg	cgccctgcct	4080
ggccgcgtct	ggacaggggc	tcgggtgttg	ggcactgaca	attccgttgt	gttgtcgccc	4140
aaatctatcg	ctttcccttg	gtctgtcgcc	cttgcgttgc	ccgcggacgc	4200	
tccttctgt	acgtcccttc	ggcccttcata	ccagcggaccc	ttcccttcct	cgccctgtct	4260
ccggctctat	ataatccata	aaatggggaa	caactacac	ttccataaa	taggaaacac	4320
tacaacccgtt	tccataaaat	agggaaacact	acatcactcc	ataaaatgtt	aaacactata	4380
cccgccgtcg	gtctcggtacc	ttttagagcca	atgacttaca	aggcagctgt	agatcttagc	4440
cactttttaa	aagaaaagg	gggactggaa	gggctaatttcc	actcccaac	aagacaagat	4500
ctgttttttgc	tttgtactgg	gtctctgtgt	ttagaccaga	tctgagcc	ggagctct	4560
ggctcaactat	ggaaaccact	gtcttaagct	caataaaatgc	tgccttgcgt	gtctcaagta	4620

-continued

---

gtgtgtgcc	gtctgttg	tgactctgg	aactagagat	ccctcagacc	ctttagtca	4680
gtgtggaaa	tctctagca					4699
<b>SEQ ID NO: 37</b>			<b>moltype = AA length = 445</b>			
<b>FEATURE</b>			<b>Location/Qualifiers</b>			
<b>source</b>			1..445			
			<b>mol_type = protein</b>			
			<b>organism = unidentified</b>			
<b>SEQUENCE: 37</b>						
MAVAIAAARV	WRLNRGLSQA	ALLLRQPGA	RGLARSHPPR	QQQQFSSLDD	KPQFPGASAE	60
FIDKLEFIQP	NVISGIPYR	VMDRCGQIIN	PSEDPHLPKE	KVLKLYKSMT	LLNTMDRILY	120
ESQRQRISF	YMTNYGEEGT	HVGSAALDN	TDLVFGQYRE	AGVLMYRDYP	LELFMAQCYG	180
NISDLGKGRQ	MPVHYGCKER	HFTVISSPLA	TQIPQAVGAA	YAAKRANANR	VVICYFGEGA	240
ASEGDAHAGF	NFAATLECP	IIFCCRNNGYA	ISTPTSEQYR	GDGIAARGPG	YGIMSIRVDG	300
NDVFAYVNAT	KEARRRAVAE	NQPFYLIEAMT	YRIGHSTSTD	DSSAYRSVDE	VNYWDKQDHP	360
ISRLRHYLLS	QGWWDEEQEE	AWRKQSRRKV	MEAFEQAEERK	PKPNPNLILFS	DVYQEMPAQL	
RKQQESLARH	LQTYGEHYPL	DHFDK				445
<b>SEQ ID NO: 38</b>			<b>moltype = AA length = 448</b>			
<b>FEATURE</b>			<b>Location/Qualifiers</b>			
<b>source</b>			1..448			
			<b>mol_type = protein</b>			
			<b>organism = unidentified</b>			
<b>SEQUENCE: 38</b>						
MAVAIAAARV	WRLNRGLSQQ	QQFSSLDDKP	QFPGASAEFI	DKLEFIQPNV	ISGIPYRVM	60
DRQGQIINPS	EDPHLPKEKV	LKLYKSMTLL	NTMDRILYES	QRQGRISFYM	TNYGEEGTHV	120
GSAALDNTD	LVFGQYREAG	VLMYRDYPLE	LFLMAQCYGNI	SDLGKGRQMP	VHYGCKERHF	180
VTISSLATQ	IPQAVGAAAYA	AKRANANRVV	ICYFGEGAAS	EGDAHAGFNF	AATLECPPIIF	240
FCRNNNGYAI	TPTSEQYRGD	GIAARGPGYG	IMSIRVDGND	VFAVYNATKE	ARRRAVAENQ	300
PFLIEAMTYS	SSPILPPDPH	SREPTLTWGP	LPLCRIGHHS	TSDDSSAYRS	VDEVNYWDKQ	360
DHPISRRLHY	LLSQGWWDEE	QEKAWRKQSR	RKVMEEAFEQA	ERKPKNPNL	LFSDVYQEMP	420
AQLRKQQESL	ARHLQTYGEH	YPLDHFDK				448
<b>SEQ ID NO: 39</b>			<b>moltype = DNA length = 1338</b>			
<b>FEATURE</b>			<b>Location/Qualifiers</b>			
<b>source</b>			1..1338			
			<b>mol_type = other DNA</b>			
			<b>organism = unidentified</b>			
<b>SEQUENCE: 39</b>						
atggcggtag	cgtatcgctgc	agcggagggtc	tggcggtcaa	accgtgggtt	gagccaggct	60
gcgcctctgc	tgctcgccga	gcctggggct	cggggactgg	ctagatctca	cccccccaagg	120
cagcagcagc	agttttcatc	tctggatgac	aagccccagt	tcccaggggc	ctcgccggag	180
tttataataga	atttggattt	catccagggc	aacgtcatct	ctggaatccc	catctaccgc	240
gtcatggacc	ggcaaggcca	gatcatcaac	cccaagcgagg	accccccacact	ggcgaaggag	300
aaaggctgtga	agctctacaa	gagcatgaca	ctgtttaaca	ccatggaccc	catctctat	360
gagtcctcagc	ggcaggggccg	gatctccctc	atacatgacca	actatggta	ggaggggcacg	420
cacggtgggg	gtggccggcc	cctggacaaa	acggacccat	tgtttggcca	gtacccggag	480
gcagggtgtc	tgtatgtatc	ggactatccc	ttatggccca	gtatggccca	gtgtatggc	540
aacatcatgt	acttggccaa	ggggcccaag	atgcctgtcc	actacggctg	caaggAACG	600
cacttcgtca	ctatcttc	tcactggcc	acgcagatcc	ctcaggcggt	ggggccggcg	660
taacgcgocca	acggggccaa	tgccaaacagg	gtcgtcatct	gttacttcgg	cgagggggca	720
gcccaggatgg	ggggccggcca	tgccggatcc	aacttcgtcc	ccacacttga	gtgccccatc	780
atcttccttc	gcccggaaacaa	tggctacccc	atctccaccc	ccaccttga	gcagtatcgc	840
ggcgatggca	ttgcagcacg	aggccccggg	tatggcatca	tgtcaatccg	cgtggatgt	900
aatgtatgtt	ttgcgtata	caaccccaaa	aaggaggccc	gacggccggc	tgtggcagag	960
aaccaggccct	tcctcatcg	ggccatgacc	taacggatcc	ggcaccacag	caccatgtac	1020
gacagtttcag	cgtacccgtc	gggtggatag	gtcaattact	gggataaaaca	ggaccacccc	1080
atctccggc	tgccggacta	tctgtgtacc	caaggctgtt	gggatgagga	gcaggagaag	1140
gcctggagga	agcagtcccg	caggaagggt	atggaggcc	ttgagcaggc	cgagcggaaag	1200
cccaaaccacca	accccaaccat	actttctca	gacgtgtatc	aggagatgcc	cgcccacgtc	1260
cgcaggcagc	aggagtctct	ggccggccac	ctgcagacct	acggggagca	ctacccactg	1320
gatcacttcg	ataagtga					1338
<b>SEQ ID NO: 40</b>			<b>moltype = AA length = 392</b>			
<b>FEATURE</b>			<b>Location/Qualifiers</b>			
<b>source</b>			1..392			
			<b>mol_type = protein</b>			
			<b>organism = unidentified</b>			
<b>SEQUENCE: 40</b>						
MAVAAAAGW	LLRLRAAGAE	GHWRRLPGAG	LARGFLHPAA	TVEDAAQRRQ	VAHFTFQPD	60
EPREYGGTQK	MNLFQSVTSA	LDNSLAKDPT	AVIFGEDVAF	GGVFRCTVGL	RDKYGDVRF	120
NTPLCEQGIV	GFGIGIAVTG	ATAIAEIQFA	DYIFFPAFDQI	VNEAAKYRYR	SGDLFNCGSL	180
TIRSPWGCVG	HGALYHSQSP	EAFFAHCPGI	KVVIIPRSPFQ	AKGLLLSCIE	DKNPCIFFEP	240
KILYRAAAEE	VPIEPYNIPL	SQAEVIQEGS	DVTLVANGTQ	VHVIREVASM	AKEKLGVSC	300
VIDLRTIIPW	DVDTICKSVI	KTGRLLISHE	APLTGGFASE	ISSTVQEECF	LNLEAPISRV	360
CGYDTPFPPhi	PEPFYIPDKW	KCYDALRKMI	NY			392

---

-continued

---

```

SEQ ID NO: 41      moltype = AA  length = 218
FEATURE          Location/Qualifiers
source           1..218
                 mol_type = protein
                 organism = unidentified

SEQUENCE: 41
MAVVAAAAGW LLRLRAAGAE GHWRRLPGAG LARGFLHPAA TVEDAAQRRO VAHFTFQPD 60
EPREYQQTOK MNLFQSNTSA LDNSLAKDPT AVIFGEDVAF GGVFRCTVGL RDKYGD 120
NPLCEGIV GFGIGIAVTG ATATAEIQFA DYIFPAFDQI VNEAAKYRYR SGDFNCGSL 180
TIRSPWGCVG HGALYHSQSP EAFFAHCPGI KIKVISLS                         218

SEQ ID NO: 42      moltype = DNA  length = 1179
FEATURE          Location/Qualifiers
source           1..1179
                 mol_type = other DNA
                 organism = unidentified

SEQUENCE: 42
atggcggttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 60
gggcactggc gtcggcttcc tggcgccggg ctggcgccgg ctttttgcaccccgcccg 120
actgtcgagg atgcggccca gaggcgccgg gtggctatt ttactttcca gccagatccg 180
gaggccccggg agtacggggca aactcagaaa atgaatctt ccagtctgt aacaagtgc 240
tttgataact cattggccaa agatccatgc gcagataat ttgttgaaga ttgtgccttt 300
ggtgaggctt ttagatgcac ttgtgccttg cagacaaaat atggaaaaaa tagagtttt 360
aatacccccatt ttgtgtaaaca aggaattttt ggatttggaa tcggatttc ggtacttgg 420
gtcaactgcac ttgcggaaat tcagtttgcattt tccctgcatt tgatcagatt 480
gttaatcgaaat cgccaaatgc tgcgtatcgatcgtttttactgc tggaaagctc 540
actatccggc ccccttgggg ctgtgttgc tcatgggttc tctatcatc tcagagtcc 600
gaagcattttt ttgcccattt cccaggaaatc aagggtggta tacccagaag cccttcac 660
gccaaaggac ttcttttgc atgcatacggagataaaaaatc ttgtatattttttaac 720
aaaatactttt acaggggcage agcggaaagaa gttccctatataaaccatcaa catcccactg 780
tcccgaggcc aagtcataca ggaaggaggat gatgttacttc tagttgcctg gggactc 840
gttcatgtga tccgagaggat agtccatcgatc gcaaaagaaa agcttggagt gtcttgtaa 900
gtcattgtatc tgaggactat aataccttgg gatgtggaca caatttgcatttgc 960
aaaacaggcc gactgtatc cagtcacggat gttcccttgc caggcgccgt tgcatcggaa 1020
atcagctatac cagtttcaggaa ggaatgttgcatttttgcacccatgttgcatttgc 1080
tgcgttgcatttgcacccat tccctacatttttgcacccatgttgcatttgc 1140
aagtgttgcatttgcacccat tccctacatttttgcacccatgttgcatttgc 1179

SEQ ID NO: 43      moltype = AA  length = 482
FEATURE          Location/Qualifiers
source           1..482
                 mol_type = protein
                 organism = unidentified

SEQUENCE: 43
MAAVRMLRTW SRNAGKLICV RYFQTGCVNH VLKPNVYVCFF GYPSFKYSHP HHFLKTTAAL 60
RGQVVFQFKLS DIGEGIREVT VKEWVYVKEGD TVSQFDSICE VQSDKASVTI TSRYDGVVIKK 120
LYYNLDDIAY VGKPLVDIET EALKDSEEDV VETPAVSHDE HTHQEIKGRK TLATPAVRRL 180
AMENNKLSE VVGSKGDRG LKEDILNYLE KQTGAILPPS PKVEIMPPP KPKDMTPVIL 240
VSKPPVFTGK DKTEPIKGQ KAMVKTMSAA LKIPHFGYCD EIDLTELVLK REELKPIAFA 300
RGIKLFSMPF FLKAASLGLL QFPILNASVD ENCQNTYKA SHNIGIAMD TEQGLIVPNVK 360
NVQICSIKFDI ATELNRLQKL GSFGVQLSTTD LTGGTFTLSN IGSIGGTFAK PVIMPPEVAI 420
GALGSKIAIP RFNQKGEVY AQIMMVSWSA DHRVIDGATM SRFSNLWKST LENPAFMILL 480
LK                                         482

SEQ ID NO: 44      moltype = DNA  length = 1449
FEATURE          Location/Qualifiers
source           1..1449
                 mol_type = other DNA
                 organism = unidentified

SEQUENCE: 44
atggctcgac tccgtatgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 60
cgctattttc aaacatgtgg taatgttcat gttttgaagc caaattatgt gtgtttttttt 120
ggttatcttccatcaacttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 180
cgtggacagg ttgttgcatttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 240
gtttaaaggat ggttatgttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 300
gttcaaaatgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 360
ctctattata atcttagacgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 420
gaagctttaa aagatgttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 480
catacacacc aagatgttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 540
gcaatggaaa acaatattaa ggttatgttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 600
cttaaaggat ggttatgttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 660
cccaaaatgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 720
gtatcaaaatgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 780
aaagcaatgttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 840
gaggatgttgc tggcgccggc tgccggctgg ctactcaggc tcagggcgcc aggggcttag 900

```

-continued

---

```
cgttggattt aacttcctt tatgccttc ttcttaaagg ctgcttcctt gggattacta 960
cagtccca tccttaaccc ttctgtggat gaaaactgcc agaatataac atataaggct 1020
tctcataaca ttgggatagc aatggatact gagcagggtt tgattgtccc taatgtgaaa 1080
aatgttcaga tctgtctat attgacatc gccactgaac tgaaccgcct ccagaattg 1140
ggctctgtgg gtcagtcag caccactgtat cttacaggag gaacattac tctttccaac 1200
atggatcaa ttgggtgtac ctggccaaa ccagtgtata tgccacctga agtagccatt 1260
ggagccctt gatcaattaa ggccatccc cgatthaacc agaaaggaga agtatataag 1320
gcacagataa tgaatgttag ctggtcagct gtcacacagag ttattgttag tgctacaatg 1380
tcacgcttcc ccaatttgtg gaaatctat ttagaaaacc cagctttat gctactagat 1440
ctgaaatga 1449
```

```
SEQ ID NO: 45 moltype = DNA length = 1449
FEATURE Location/Qualifiers
source 1..1449
mol_type = other DNA
organism = synthetic construct
```

```
SEQUENCE: 45
atggcagccg tgaggatgtc gaggacctgg agccggaaacg caggcaagct gatctgcgtg 60
agatacttc agacatgtgg caacgtgcac gtgtcaaggc ccaattacgt gtgcttc 120
ggctaccctt cttcaagttt ctctcaccc caccacttc tgaagaccac agccgcctg 180
aggggacagg tgggtcagtt caagtcgago gacatcgccg agggcatccg cgagggtgacc 240
gtgaaggagt ggtacgtgaa ggaggccgc acagtgcggc agtttgcatt catctgtgag 300
gtgcagtcg acaaggccgc cgtgaccatc acatccccgtt acgtggcggt gatcaagaag 360
ctgtactata acctggacga catcgccat gtggcaacgc cactgggtga catcgagaca 420
gaggccctga aggacagcga ggaggatgtg gtggagacac cccgcgtgtc ccacgatgag 480
cacacacacc aggagatcaa gggaaaggaa accctggcca caccagccgt gcggagactg 540
cccatggaga acaatcatcaa gtcgagcgg gttgtggat cccgcggaa cggcagaatc 600
cttgaaggagg acatcttgc aaatgtggatc aacgcggccgc ggcgcattcc 660
cctaaggcttgg agatcatgcc acccccttca aacgcctaaacg acatgcacgt gccaatctcg 720
gtgtctaaacg ccccccgttgc caccggcaag gataagacag accctatcaa gggcttc 780
aaggccatgg tgaagaccatc gageggccgc ctgaaatgcg caccatccg ctactgcgc 840
gagatgcattc tgacagatgtc ggttgcggc cccggggggc tgaagccat cgccttc 900
agaggcatca agctgtcctt tatgccttc ttctgtggat cccgccttctt gggctgtg 960
cagtttcata tcctgaacgc ctctgtggac gagaactgcg agaatatcac ctataaggcc 1020
agccatggggccatccatc tccggatcgc catggatataa ggcggggccgc tgatctggcc 1080
aatatgtcgcata tctgttccat ctgcgcgcgc ggcaccgcgc tgaacaggct gcacgcgc 1140
ggctctgtgg gocagctgatc caccacatgc ctgaccggccgc gcacccatc actgtccat 1200
atcggtctata tcggccgcac attttgcggatc tgccaccaga ggtggcaatc 1260
ggccgccttgg gctctatcaa ggccatccctt cgcctcaacc agaaggccgc ggttacaatg 1320
gcccagatca tgaatgttag ctggccgcgc gaccacagag tgatctggccgc cccaccatg 1380
tctcgcttca gcaaccctatc tggagaatc cccgcctttat gctgtgtggat 1440
ctgaaatga 1449
```

```
SEQ ID NO: 46 moltype = AA length = 509
FEATURE Location/Qualifiers
source 1..509
mol_type = protein
organism = unidentified
```

```
SEQUENCE: 46
MQSWSRVYCS LAKRGHFNRI SHGLQGLSAV PLRTYADQPI DADTVIGSG PGGYVAIAKA 60
AQOLGFKTCI EKNETLGGT LNVGCIPSKA LLNNSHYHHM AHGKDFASRG IEMSEVRNLN 120
DKMMEQKSTA VKALTGGIAH LFKQNKGKHN NGYKKITGKVN QVTATKADGG TQVIDTKNIL 180
IATGSEVTPF PGITIDEDTI VSSTGALSLK KVPEKVMVIG AVGIVGELGS VWWQLGADVT 240
AVEFLGHVG VGINDEISKN FQRILQKQGF KFKLNKTVTG ATKKSDGKID VSIEAASGGK 300
AEVITCDVLL VCIQGRPFTH NLGELLELIE LDPRGRIPVN TRFQTKIPNI YAIGDVVAGP 360
MLAHKAEDEG IICVEGMAGG AVHIDYNCVP SVIYTHPEVA WVGKSEEQLK EEEIEYKVKG 420
FPFAANSRAK TNADTDGMVI ILGQKSTDRL LGAHILGPAG EMVNEAALA LEYGASCEI 480
ARVCAHAPL SEAFREANLA ASFGKSINF 509
```

```
SEQ ID NO: 47 moltype = AA length = 410
FEATURE Location/Qualifiers
source 1..410
mol_type = protein
organism = unidentified
```

```
SEQUENCE: 47
MAHGKDFASR GIEMSEVRNL LDKMMEQKST AVKALTGGIA HLFKQNKVH VNGYKKITGK 60
NQVTATKADG GTQVIDTKNI LIATGSEVTP PGITIDEDTI VSSTGALSL KKVPEKVMVI 120
GAGVIVGELG SWWQRLGADV TAVERLGHVG GVGIDMEISK NFQRILQKQG FKFKLNTKVT 180
GATKKSDGKI DVSIIEASGG KAEVITCDVL LVCIGRRPFT KNLGLLELGI ELDPRGRIPV 240
NTRFQTKIPN IYAIQDVVAG PMLAHKAEDE GIICVEGMAG GAVHIDYNCVP SVIYTHPEV 300
AWVGKSEEQL KEEGIEYKVKG KFPFAANSRA KTNADTDGMV KILGQKSTDRL VLGAHILGP 360
AGEMVNEAAL ALEYGASCED IARVCAHAPL LSEAFREANLA AASFGKSINF 410
```

```
SEQ ID NO: 48 moltype = AA length = 461
FEATURE Location/Qualifiers
source 1..461
```

-continued

---

```

mol_type = protein
organism = synthetic construct

SEQUENCE: 48
MQSWSRVYCS LAKRGHFNRI SHGLQGLSAV PLRTYADQPI DADTVIGSG PGGYVAIAKA 60
AQLGFKTCI EKNETLGGTC LNVCICPSKA LLNNSHYYHM AHGKDFASRG IEMSEVRLNL 120
DKMMEQKSTA VKALTGIAH LFKQNPKIDED TIVSSTGALS LKKVPEKMMV IGAGVIGVEL 180
GSVWQRLGAD VTAVEFLGHV GGVGIDMEIS KNFQRILQKQ GFKEKLNTKV TGATKKSDGK 240
IDVSIEAASG GKAEVITCDV LLVICGRPPF TKNLGLEELG IEELDPRGRIP VNTRPQTKIP 300
NIYAIGDVVA GPMLAHKAED EGIICVEGMA GGAVHIDYNC VPSVIYTHEP VAWVGKSEQ 360
LKEEGIYKV GKFPFAANSE AKTNAADTDM VKİLGQKSTD RVLGAHILGP GAGEMVNEAA 420
LALEYGASCE DIARVCHAHP TLSEAFREAN LAASFGKSIN F 461

SEQ ID NO: 49      moltype = DNA length = 1530
FEATURE           Location/Qualifiers
source            1..1530
mol_type = other DNA
organism = unidentified

SEQUENCE: 49
atgcagagct ggagtctgtgt gtactgctcc ttggccaaga gaggccattt caatcgaata 60
ttccatggcc tacaggact ttctgcgtg cctctgaaata ctacgcaga tcagccgatt 120
gatgctgatg taacagttat aggttctgtt cctggaggat atgttgc tattaaagct 180
gcccaggtag gttcaagac agtctgcgtt gaaaaaaaatgg aaacacttgg tggAACATGC 240
ttgaatgttg ttgttattcc ttcttaaggct ttatgaaaca actctcatta ttccatatg 300
gccccatggaa aagatttgc attctggaggat attgaaatgtt ccggaaatgttcc ttgaaatcta 360
gacaagatgaa tggagcagaa ggtactgca gtaaaagctt taacagggtt aattggccac 420
ttattcaaac agaaataaggat ttgttcatgtc aatggatgat gaaagataac tggcaaaaat 480
caagtcaactt ctacgaaaggc tgatggcgcc acttcagggtt ttgatcacaaa gaacatttt 540
atagccacgg gttcagaatgt tactctttt ccttggaaatcc cgtatgatgaa agatacaata 600
gtgtcatctt cagggtctttt atctttaaaaaaa aaaaatcccg aaaaatgttgg tttttttttt 660
gcaggagttaa taggtttaga attgggttca gtttggcaaa gacttgggtc agatgtgaca 720
gcagttgaat ttttaggttca ttgttggatgatgat gtttggatgat atctaaaaaac 780
tttcaacggca ttcttcaaaa acagggggtt aatattttat tgaatcacaaa ggttacttgg 840
gctaccaaga agtcaagatgg aaaaattgtt gtttctattt aagctgtttc tgggtgtt 900
gctgaagttt tcaacttgtt gtttacttgc gtttgcattt gccgacgacc cttaactaag 960
aattttggggat tagaaagatgg gggaaatggaa ctatggatccca gagatggat tccatgtt 1020
accaggatcc aactaaatcc tccaaatatac tccaaatccatgg tttttttttt tttttttttt 1080
atgtctggc tccaaatccatgg tttttttttt tttttttttt tttttttttt tttttttttt 1140
gctgtgcaca ttgactacaa ttgtgtgcac tcaatgtt acacacaccc tgaagttgt 1200
ttgggttgcac aatcaagaaga gcaatgttggaa gaagagggtt ttgatgttggaa 1260
ttcccatgtt ctgtcaacagc cagatgttgg tttttttttt tttttttttt tttttttttt 1320
atcttggcc aaaaatccatgg tttttttttt tttttttttt tttttttttt tttttttttt 1380
ggagaaaatgg taaaatgttgg tttttttttt tttttttttt tttttttttt tttttttttt 1440
gttagatgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1500
gcgtcatttgc gcaaatcaat tttttttttt tttttttttt tttttttttt tttttttttt 1530

SEQ ID NO: 50      moltype = DNA length = 17
FEATURE           Location/Qualifiers
source            1..17
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 50
actgacgctc tcgcacc 17

SEQ ID NO: 51      moltype = DNA length = 17
FEATURE           Location/Qualifiers
source            1..17
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 51
tctcgacgca ggactcg 17

SEQ ID NO: 52      moltype = DNA length = 20
FEATURE           Location/Qualifiers
source            1..20
mol_type = other DNA
organism = synthetic construct
misc_feature      1
note = FAM linked at 5' terminus

SEQUENCE: 52
atctctctcc ttcttagcctc 20

SEQ ID NO: 53      moltype = DNA length = 20
FEATURE           Location/Qualifiers
source            1..20
mol_type = other DNA
organism = synthetic construct

```

---

-continued

---

```

SEQUENCE: 53
ggctgttggg cactgacaat                                20

SEQ ID NO: 54      moltype = DNA  length = 17
FEATURE
source           Location/Qualifiers
1..17
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 54
acgtcccgcg cagaatc                                17

SEQ ID NO: 55      moltype = DNA  length = 23
FEATURE
source           Location/Qualifiers
1..23
mol_type = other DNA
organism = synthetic construct
misc_feature     1
note = FAM linked at 5' terminus
23
note = NGB linked at 3' terminus

SEQUENCE: 55
ttcccttggc tgctcgctg tgt                                23

SEQ ID NO: 56      moltype = DNA  length = 20
FEATURE
source           Location/Qualifiers
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 56
accgattcca gatgattggc                                20

SEQ ID NO: 57      moltype = DNA  length = 22
FEATURE
source           Location/Qualifiers
1..22
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 57
tccatattaa tgcagtgcctt gc                                22

SEQ ID NO: 58      moltype = DNA  length = 24
FEATURE
source           Location/Qualifiers
1..24
mol_type = other DNA
organism = synthetic construct
misc_feature     1
note = HEX linked at 5' terminus
24
note = BHQ1 linked at 3' terminus

SEQUENCE: 58
agaggcctgt cctgcagctc atgg                                24

```

---

**1.** A lentiviral vector comprising a nucleotide sequence encoding a branched-chain alpha-ketoacid dehydrogenase complex (BCKDC) subunit.

**2.** The lentiviral vector according to claim 1, wherein the lentiviral vector is an immune-shielded lentiviral vector.

**3.** The lentiviral vector according to claim 1 or 2, wherein the BCKDC subunit is selected from BCKD E1 alpha subunit (BCKDE1A), or a fragment thereof; BCKD E1 beta subunit (BCKDE1B), or a fragment thereof; BCKD E2 subunit (DBT), or a fragment thereof; and BCKD E3 subunit (DLD), or a fragment thereof; preferably wherein the BCKDC subunit is DBT or a fragment thereof.

**4.** The lentiviral vector according to any preceding claim, wherein the BCKDC subunit comprises or consists of an amino acid sequence which is at least 70% identical to one of SEQ ID NOs: 37, 38, 40, 41, 43, 46, 47 or 48, or a fragment thereof, preferably wherein the BCKDC subunit comprises or consists of an amino acid sequence which is at least 70% identical to SEQ ID NO: 43, or a fragment thereof.

**5.** The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC sub-unit comprises or consists of a nucleotide sequence which is at least 70% identical to one of SEQ ID NOs: 39, 42, 44, 45 or 49, or a fragment thereof, preferably wherein the nucleotide sequence encoding a BCKDC subunit comprises or consists of a nucleotide sequence which is at least 70% identical to one of SEQ ID NOs: 44 or 45, or a fragment thereof.

**6.** The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC sub-unit is codon-optimised, preferably wherein the nucleotide sequence encoding a BCKDC subunit comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 45, or a fragment thereof.

**7.** The lentiviral vector according to any preceding claim, wherein the lentiviral vector is a CD47<sup>high</sup> lentiviral vector.

**8.** The lentiviral vector according to any preceding claim, wherein the lentiviral vector is obtained from a CD47<sup>high</sup>

host cell, optionally wherein the host cell is genetically engineered to increase expression of CD47 on the cell surface.

9. The lentiviral vector according to any preceding claim, wherein the lentiviral vector has at least about 2-fold more CD47 on its surface than a lentiviral vector obtained from an unmodified host cell.

10. The lentiviral vector according to any preceding claim, wherein the lentiviral vector is a MHC-I<sup>free</sup> lentiviral vector.

11. The lentiviral vector according to any preceding claim, wherein the lentiviral vector is obtained from a MHC-I<sup>free</sup> host cell, optionally wherein the host cell is genetically engineered to disrupt expression of MHC-I on the cell surface.

12. The lentiviral vector according to any preceding claim, wherein MHC-I is not detectable on the surface of the lentiviral vector.

13. The lentiviral vector according to any preceding claim, wherein the lentiviral vector is a CD47<sup>high</sup>/MHC-I<sup>free</sup> lentiviral vector.

14. The lentiviral vector according to any preceding claim, wherein the lentiviral vector is obtained from a CD47<sup>high</sup>/MHC-I<sup>free</sup> host cell.

15. The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to one or more miRNA target sequences.

16. The lentiviral vector according to claim 15, wherein the one or more miRNA target sequences suppress transgene expression in one or more cells other than hepatocytes, preferably wherein the one or more miRNA target sequence suppress transgene expression in hematopoietic-lineage cells and/or antigen-presenting cells.

17. The lentiviral vector according to claim 15 or 16, wherein the one or more miRNA target sequences are selected from miR-181, miR-142, miR-223, and miR-155 target sequences.

18. The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to one or more mir-142 target sequence, two or more mir-142 target sequences, three or more mir-142 target sequences, or four or more mir-142 target sequences.

19. The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to four mir-142 target sequences.

20. The lentiviral vector according to any of claims 15-19, wherein the one or more miRNA target sequences comprise or consist of a nucleotide sequence which is at least 90% identical to SEQ ID NO: 17, preferably wherein the one or more miRNA target sequences comprise or consist of a nucleotide sequence which is at least 90% identical to SEQ ID NO: 18.

21. The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to a liver-specific promoter, preferably wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to a hepatocyte-specific promoter.

22. The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to a transthyretin (TTR) promoter, an alpha-1-antitrypsin (AAT) promoter, a thyroxine-binding

globulin (TBG) promoter, a APoE/hAAT promoter, a HCR-hAAT promoter, a LP1 promoter, or a HLP promoter.

23. The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to a transthyretin (TTR) promoter, preferably wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to a Enh1mTTR (ET) promoter.

24. The lentiviral vector according to any preceding claim, wherein the nucleotide sequence encoding a BCKDC subunit is operably linked to a promoter which comprises or consists of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 19.

25. The lentiviral vector according to any preceding claim, wherein the lentiviral vector is pseudotyped, preferably wherein the lentiviral vector is VSV.G-pseudotyped.

26. The lentiviral vector according to any preceding claim, wherein the lentiviral vector is a self-inactivating (SIN) lentiviral vector, preferably wherein the lentiviral vector comprises self-inactivating (SIN) LTRs which comprise or consist of a nucleotide sequence which is at least 70% identical to SEQ ID NO: 23, or a fragment thereof.

27. The lentiviral vector according to any preceding claim, wherein the lentiviral vector is an integrating lentiviral vector and/or a replication-defective lentiviral vector.

28. The lentiviral vector according to any preceding claim, wherein the lentiviral vector comprises a nucleotide sequence which is at least 70% identical to SEQ ID NO: 36.

29. An isolated cell comprising a lentiviral vector according to any of claims 1-28.

30. A pharmaceutical composition comprising a lentiviral vector according to any of claims 1-28, in combination with a pharmaceutically acceptable carrier, diluent or excipient.

31. The lentiviral vector according to any of claims 1-28 or pharmaceutical composition according to claim 30, for use as a medicament.

32. Use of a lentiviral vector according to any of claims 1-28 or a pharmaceutical composition according to claim 30, for the manufacture of a medicament.

33. A method comprising administering a therapeutically effective amount of a lentiviral vector according to any of claims 1-28 or a pharmaceutical composition according to claim 30, to a subject in need thereof.

34. The lentiviral vector according to any of claims 1-28 or pharmaceutical composition according to claim 30, for use in preventing or treating maple syrup urine disease (MSUD).

35. Use of a lentiviral vector according to any of claims 1-28 or a pharmaceutical composition according to claim 30, for the manufacture of a medicament for preventing or treating maple syrup urine disease (MSUD).

36. A method of preventing or treating maple syrup urine disease (MSUD), comprising administering a therapeutically effective amount of a lentiviral vector according to any of claims 1-28 or a pharmaceutical composition according to claim 30, to a subject in need thereof.

37. The lentiviral vector or pharmaceutical composition for use according to claim 34, the use according to claim 35, or the method according to claim 36, wherein the BCKDC subunit is BCKDE1A, or a fragment thereof, and the MSUD is MSUD type 1A; the BCKDC subunit is BCKDE1B, or a fragment thereof, and the MSUD is MSUD type 1B; the BCKDC subunit is DBT, or a fragment thereof, and the

MSUD is MSUD type 2; or the BCKDC subunit is DLD, or a fragment thereof, and the MSUD is MSUD type 3.

**38.** The lentiviral vector or pharmaceutical composition for use according to claim **34** or **37**, the use according to claim **35** or **37**, or the method according to claim **36** or **37**, wherein the BCKDC subunit is DBT, or a fragment thereof, and the MSUD is MSUD type 2.

**39.** The lentiviral vector or pharmaceutical composition for use according to any of claims claim **31**, **34**, **37-38**, the use according to any of claims **32**, **35**, **37-38**, or the method according to any of claims **33**, **36-39**, wherein the subject is a human subject.

**40.** The lentiviral vector or pharmaceutical composition for use according to any of claims claim **31**, **34**, **37-39**, the use according to any of claims **32**, **35**, **37-39**, or the method according to any of claims **33**, **36-39**, wherein the subject is a juvenile.

**41.** The lentiviral vector or pharmaceutical composition for use according to any of claims claim **31**, **34**, **37-39**, the use according to any of claims **32**, **35**, **37-39**, or the method according to any of claims **33**, **36-39**, wherein the subject is a paediatric patient, preferably wherein the subject is a neonatal patient or an infantile patient.

**42.** The lentiviral vector or pharmaceutical composition for use according to any of claims **31**, **34**, **37-41**, the use according to any of claims **32**, **35**, **37-41**, or the method according to any of claims **33**, **36-41**, wherein said lentiviral vector or said pharmaceutical composition is administered systemically, preferably wherein said lentiviral vector or said pharmaceutical composition is administered by intravenous injection or intraperitoneal injection.

**43.** The lentiviral vector or pharmaceutical composition for use according to any of claims **31**, **34**, **37-41**, the use according to any of claims **32**, **35**, **37-41**, or the method according to any of claims **33**, **36-41**, wherein said lentiviral vector or said pharmaceutical composition is administered locally, preferably wherein said lentiviral vector or said pharmaceutical composition is administered by direct injection, intra-arterial injection, or intraportal injection.

**44.** The lentiviral vector or pharmaceutical composition for use according to claim **43**, the use according to claim **43**, or the method according to claim **43**, wherein said lentiviral vector or said pharmaceutical composition is administered locally to the liver, preferably wherein said lentiviral vector or said pharmaceutical composition is administered by intrahepatic injection, intrahepatic arterial injection, or intraportal injection.

**45.** The lentiviral vector or pharmaceutical composition for use according to any of claims **31**, **34**, **37-44**, the use according to any of claims **32**, **35**, **37-43**, or the method according to any of claims **33**, **36-44**, wherein the lentiviral vector is administered at a dose of at least about  $10^8$  TU/kg, at least about  $10^9$  TU/kg, or at least about  $10^{10}$  TU/kg.

**46.** The lentiviral vector or pharmaceutical composition for use according to any of claims **31**, **34**, **37-45**, the use according to any of claims **32**, **35**, **37-45**, or the method according to any of claims **33**, **36-45**, wherein the lentiviral vector is administered in a dose of from about  $10^8$  to about  $10^{11}$  TU/kg, from about  $10^8$  to about  $10^{10}$  TU/kg, or from about  $10^9$  to about  $10^{10}$  TU/kg.

**47.** The lentiviral vector or pharmaceutical composition for use according to any of claims **31**, **34**, **37-46**, the use according to any of claims **32**, **35**, **37-46**, or the method according to any of claims **33**, **36-46**, wherein the lentiviral vector integrates into the genome of liver cells and is maintained as the liver cells duplicate, preferably wherein the lentiviral vector integrates into the genome of hepatocytes and is maintained as the hepatocytes duplicate.

**48.** The lentiviral vector or pharmaceutical composition for use according to any of claims **34**, **37-47**, the use according to any of claims **35**, **37-47**, or the method according to any of claims **36-47**, wherein serum levels of branched-chain amino acids (BCAA) and/or alloisoleucine are reduced and/or normalised.

**49.** The lentiviral vector or pharmaceutical composition for use according to any of claims **34**, **37-48**, the use according to any of claims **35**, **37-48**, or the method according to any of claims **36-48**, wherein branched-chain alpha-ketoacid dehydrogenase (BCKD) activity in the liver is improved and/or normalised.

**50.** An immune-shielded lentiviral vector for use in a method of therapy, wherein the method comprises administration of the immune-shielded lentiviral vector to a juvenile or paediatric subject.

**51.** A cell for use in a method of therapy, wherein the cell comprises an immune-shielded lentiviral vector, and wherein the method comprises administration of the cell to a juvenile or paediatric subject.

**52.** The immune-shielded lentiviral vector or cell for use according to claim **50** or **51**, wherein the subject is a neonatal subject or an infantile subject.

\* \* \* \* \*