LENTIVIRAL VECTORS FOR DELIVERY OF PKLR TO TREAT PYRUVATE KINASE DEFICIENCY

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 62/573,037, filed on October 16, 2017, which is incorporated by reference herein in its entirety.

DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY

[0002] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is ROPA_004_01WO_SeqList_ST25.txt. The text file is 29 KB, was created on October 15, 2018, and is being submitted electronically via EFS-Web.

FIELD OF THE INVENTION

[0003] This invention pertains to gene therapy of Pyruvate Kinase Deficiency.

BACKGROUND OF THE INVENTION

[0004] Pyruvate Kinase Deficiency (PKD) is a monogenic metabolic disease caused by mutations in the *PKLR* gene that leads to hemolytic anemia of variable symptomatology and that can be fatal during the neonatal period. PKD recessive inheritance trait and its curative treatment by allogeneic bone marrow transplantation provide an ideal scenario for developing gene therapy approaches.

[0005] Among many other hereditary enzymatic defects affecting the erythrocytes, Pyruvate Kinase deficiency (PKD) is the most frequent one causing chronic nonspherocytic hemolytic anemia (CNSHA) (Zanella et al. 2007). Onset and severity of PKD are very variable and range from mild to severe neonatal anemia, becoming fatal during the childhood in the most severe cases (Pissard et al. 2006). Growth retardation, *hydrops fetalis* and death during the neonatal period have also been reported with low frequency (Gilsanz et al. 1993). PKD prevalence has been estimated at 1:20,000 in the general Caucasian population (Beutler et al. 2000) and, so far, more than 195 different mutations in the *PKLR* gene have been identified (http://www.lovd.nl/pklr).

[0006] Allogeneic bone marrow transplantation (BMT) has been successfully used to cure severe PKD patients (Tanphaichitr et al. 2000), but the low availability of histocompatible donors and the serious complications associated with the BMT of these patients (*i.e.* graft

versus host disease, opportunistic infections, etc.) make periodic blood transfusions and splenectomy the main therapeutic options for most of the severe forms of PKD (Zanella et al. 2005), dramatically increasing patient morbidity and mortality (Hilgard et al. 2005). The limited efficacy and side effects of the therapeutic options for severe PKD patients and its recessive inheritance trait make PKD a suitable disease to be treated by gene therapy.

[0007] PKD is caused by defects in the Pyruvate Kinase (PK) enzyme (Zanella 2005) that catalyses the last ATP-generating reaction of the glycolysis pathway in all cells. In mature erythrocytes PK becomes essential as RBCs only express the R-type specific isoform (RPK) (Kanno et al. 1992) due to the regulation of the erythroid specific alternative promoter of the *PKLR* locus (Noguchi et al. 1987). Thus, any loss of RPK activity impairs RBC metabolism and lifespan (Zanella 2005), leading to CNSHA.

[0008] A promising approach to treating and preventing genetic and other diseases and disorders is delivery of therapeutic agents with a gene therapy vector. Currently, viral vectors show the greatest efficiency in gene transfer, and for correction of genetic diseases such that persistent gene expression is required, herpesvirus, retrovirus, lentivirus, adenovirus, or AAV based vectors are desirable due to the integrating nature of the viral life cycle.

[0009] Gene therapy for monogenic diseases, particularly those affecting the hematopoietic system, has provided convincing evidence that genetic correction of autologous hematopoietic stem cells (HSCs) is an alternative therapeutic option to allogeneic HSCT, avoiding its major complications (Cartier et al 2009; Cavazzana-Calvo et al 2010; Cartier et al 2012; Aiuti et al 2013; Biffi et al 2013). Genetic correction for diseases affecting the erythrocyte such as β-thalassemia and sickle cell disease, have been addressed in animal models (Pestina et al 2009l Breda et al 2012) and also in humans (Cavazzana-Calvo et al 2010). However, gene therapy approaches for inherited erythroid metabolic deficiencies such as PKD are still limited.

[0010] The feasibility of HSC gene therapy for PKD has been demonstrated both in mouse (Tani et al 1994; Meza et al 2009) and in dog RPK deficient experimental models (Trobridge et al 2012) showing that donor chimerism or level of gene-modified cells are key points to reach an efficient correction of the hemolytic phenotype (Richard et al 2004) given the lack of selective advantage of gene-corrected HSCs. Previous work with a PKD mouse model demonstrated that retrovirally-derived human RPK expression was capable of fully correcting PKD phenotype when over 25% genetically corrected cells were transplanted (Meza et al 2009). A similar therapeutic threshold of corrected cells was recently reported in one PKD Basenji dog infused with *in vivo* expanded and foamy vector-corrected HSCs (Trobridge et al 2012).

[0011] A number of challenges remain with regard to designing polynucleotide cassettes and expression vectors for use in gene therapy. One significant challenge is obtaining sufficient expression of the transgene in target cells. A longstanding unmet need in the art has been sufficiently robust expression of transgenes following gene transfer. In some cases, more efficient expression is required for the efficacy of certain vectors, for example plasmid DNA vectors. In other cases, more efficient gene expression cassettes are desirable to allow for a lower therapeutic dose that has a more favorable safety profile or a less invasive route of administration.

[0012] High levels of transgene expression can be achieved when gamma retroviral (gamma-RV) vectors are used due to the fact that the therapeutic transgene expression is regulated by their long terminal repeats (LTR). However, the first clinical trials based on this type of vector raised safety concerns, as several patients developed unexpected leukemias (Hacein-Bey-Abina et al 2008). The strong promoter activity of LTRs could affect the regulation of surrounding genes, either by activation of proto-oncogene promoters or by inhibition of tumour suppressor genes, leading to insertional mutagenesis (Ott et al 2006; Howe et al 2008; Stein et al 2010; Braun et al 2014). These findings highlighted the need to use safer and more efficient vectors than gamma-RV vectors for the PKD gene therapy.

SUMMARY OF THE INVENTION

[0013] In one embodiment, the present invention provides an expression cassette comprising a polynucleotide sequence comprising: a) a promoter sequence; b) a sequence encoding a gene product; and c) a ribonucleic acid (RNA) export signal, wherein the promoter sequence is operably linked to the sequence encoding the pyruvate kinase polypeptide, and optionally where a)-c) are present in the expression cassette in 5' to 3' order. In certain embodiments, the promoter is a phosphoglycerate kinase (PGK) promoter. In some embodiments, the gene product is a therapeutic gene product. In some embodiments, the therapeutic gene product is a pyruvate kinase (PK) polypeptide, optionally a pyruvate kinase, liver and red blood cell (PKLR) polypeptide. In certain embodiments, the sequence encoding the gene product is codon-optimized. In some embodiments, the sequence encoding the gene product is a codon-optimized version of the human pyruvate kinase cDNA having at least 85% identity to SEQ ID NO:8. In particular embodiments, the RNA export signal is a mutated post-transcriptional regulatory element of the woodchuck hepatitis virus (wPRE).

[0013a] Disclosed herein is an expression cassette for pyruvate kinase deficiency (PKD), comprising a polynucleotide sequence comprising in the following 5' to 3' order: a) a promoter

sequence, b) a codon-optimized sequence encoding a human PKLR gene product and c) a mutated post-transcriptional regulatory element of woodchuck hepatitis virus (WPRE); wherein the promoter sequence is operably linked to the codon-optimized sequence encoding a human PKLR gene product, and wherein the codon-optimized sequence encoding a human PKLR gene product shares at least 95% identity to SEQ ID NO: 8 and wherein the mutated WPRE is a chimeric WPRE comprising a sequence having at least 90% identity to SEQ ID NO: 24.

[0014] In certain embodiments, the mutated wPRE is a chimeric wPRE comprising a sequence having at least 80% identity to SEQ ID NO:24. In some embodiments, the expression cassette further comprising one or more enhancer sequences. In some embodiments, the expression cassette further comprises a polypurine tract (PPT) or polyadenylation (polyA) signal sequence. In some embodiments, the expression cassette further comprises one or more of the following sequences: i) a packing signal sequence; ii) a truncated Gag sequence; iii) a Rev responsive element (RRE); iv) a central polypurine tract (cPPT); v) a central terminal sequence (CTS); and vi) an upstream sequence element (USE), optionally from simian virus 40 (SV40-USE). In some embodiments, the expression cassette further comprises 5' and 3' long terminal repeat (LTR) sequences.

[0015] In a related embodiment, the present invention provides a recombinant gene delivery vector comprising an expression cassette disclosed herein. In certain embodiments, the recombinant gene delivery vector is a virus or viral vector. In certain embodiments, the virus or viral vector is a lentivirus (LV).

[0015a] Disclosed herein is a recombinant gene delivery vector comprising the expression cassette as defined herein, wherein the recombinant gene delivery vector is a lentiviral vector.

[0016] Disclosed herein is a cell comprising an expression cassette or transduced by a gene delivery vector disclosed herein. In some embodiments, the cell is a blood cell. In some embodiments, the cell is an erythroid cell. In some embodiments, the cell is a bone marrow cell, *e.g.*, a lineage depleted bone marrow cell. In some embodiments, the cell is a hematopoietic stem cell. In some embodiments, the cell is a CD34+ hematopoietic stem cell. In some embodiments, the cell is a committed hematopoietic erythroid progenitor cell.

[0017] Disclosed herein is a pharmaceutical composition comprising a pharmaceutically acceptable excipient and recombinant gene delivery vector or cell disclosed herein.

[0018] In another embodiment, the present invention provides a method of treating or preventing a disease or disorder in a subject in need thereof, comprising providing to the subject an expression cassette, gene delivery vector, or pharmaceutical composition disclosed herein.

In one embodiment, the disease or disorder is a PKD and the gene product is a PK polypeptide, optionally a PKLR polypeptide. In certain embodiments, the pharmaceutical composition comprises the recombinant gene delivery vector. In other embodiments, the pharmaceutical composition comprises the cell. In one embodiment, the cell is autologous to the subject.

[0018a] Disclosed herein is the use of the pharmaceutical composition as defined herein in the manufacture of a medicament for treating or preventing Pyruvate Kinase Deficiency (PKD) in a subject in need thereof.

[0019] In a related embodiment, the present invention provides a method for expressing a transgene in erythroid cells, comprising contacting one or more erythroid cells with an effective amount of a recombinant viral vector, wherein the vector comprises a human PGK promoter, a codon optimized version of a human PKLR cDNA transgene, and a mutated wPRE, wherein following said contacting, PKLR is expressed at detectable levels in the one or more erythroid cells.

[0019a] Disclosed herein is an *in vitro* or *ex vivo* method for expressing a human PKLR gene in erythroid cells or erythroid progenitor cells, comprising contacting one or more erythroid cells or erythroid progenitor cells with an effective amount of a recombinant viral vector, wherein the vector comprises a human phosphoglycerate kinase promoter, a codon-optimized PKLR transgene, and a mutated post-transcriptional regulatory element of the woodchuck hepatitis virus (WPRE), wherein the codon-optimized PKLR transgene shares at least 95% identity with SEQ ID NO: 8, wherein the mutated WPRE is a chimeric WPRE comprising a sequence having at least 90% identity to SEQ ID NO: 24, and wherein following said contacting, PKLR is expressed at detectable levels in the one or more erythroid cells or erythroid progenitor cells.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0021] FIG. 1 depicts a scheme reflecting the position of different elements present in the backbone of an illustrative LV vector.

[0022] FIG. 2A is a schematic representation of illustrative self-inactivating (SIN) LV vectors used throughout gene therapy experiments harboring the human PGK promoter regulating the expression of the EGFP transgene in the control vector (upper diagram) or the expression of a codon-optimized sequence of the *PKLR* gene cDNA (coRPK) in the therapeutic vector (lower diagram).

[0023] FIG. 2B is a schematic of a gene therapy protocol performed to address the functionality of the developed PGK-coRPK LV vector.

[0024] FIGS. 3A-3D depict data showing correction of PKD phenotype in peripheral blood of primary recipients after genetic correction. **FIGS. 3A and 3B** show RBC and reticulocyte levels, respectively, in healthy (black bar, n=5) and PKD anemic mice (gray bar, n=6), and PKD anemic mice that were transplanted with the EGFP (white bar, n=9) or coRPK transduced cells (scratched bar, n=17). Data are represented as the average ± SEM and were analyzed by non-parametric Kruskal-Wallis test. **FIG. 3C** shows the flow cytometry strategy used to detect the biotin labelled RBCs throughout the time and **FIG. 3D** shows RBC survival kinetics in healthy (black line, n=2), anemic (gray line, n=2) and genetically corrected mice (discontinuous line, n=4). Data are represented as the average ± SEM and were analyzed by two-way ANOVA test. Healthy, non-transplanted control mice; PKD, non-transplanted PKD mice; coRPK, PKD mice expressing the therapeutic transgene.

[0025] FIGS. 4A-4C show multi-lineage hematopoietic reconstitution in secondary transplanted mice. **FIG. 4A** is a diagram of the flow cytometry strategy used to identify the different hematopoietic lineages by labeling with CD3-PE, B220-PE, B220-PECy5, Gr1-Biotin and Mac1-Biotin antibodies plus SAV-PE-Cy5. **FIG. 4B** depicts representative dot-plots and **FIG. 4C** depicts percentages of each lineage in PB at 140 days after transplant. Bars represent the average percentage ± SEM of healthy (n=2, black bar) and PKD mouse (n=2, grey bar) controls and secondary transplanted mice expressing the coRPK therapeutic transgene (n=4, scratched bar).

[0026] FIGS. 5A-5D depict PKD phenotype correction in secondary transplanted mice. **FIG. 5A** shows Brilliant Cresyl blue staining of blood smears from non-transplanted mice and secondary recipients to identify reticulocyte population (in blue). **FIG. 5B** is flow cytometry analysis of reticulocyte levels in peripheral blood. **FIG. 5C** represents RBC percentage and **FIG. 5D** represents reticulocyte percentage in secondary transplanted mice expressing the coRPK transgene (scratched bar, n=4), in healthy mice (black bar, n=3) and in anemic control mice (grey bar, n=3). Data are represented the average ± SEM and were analyzed by non-parametric two-tailed Mann-Whitney test.

[0027] FIGS. 6A-6C show quantification of proviral integrations. **FIG. 6A** shows vector copy number per cell in BM CFUs from individual transplanted mice at 120 to 170 days after transplant. Transduction and chimerism percentages are also shown. **FIG. 6B** shows provirus copy number in cells from different hematopoietic compartments. Columns represent the average ± SEM of the different groups of transplanted mice. **FIG. 6C** shows the kinetics of proviral integrations in BM cells from individual transplanted EGFP-expressing mice (grey lines) and mice carrying the coRPK transgene (black lines).

[0028] FIGS. 7A-7C depict the normalization of the erythroid differentiation pattern in genetically corrected mice. **FIG. 7A** shows percentages of the different erythroid subpopulation in bone marrow and spleen at 140 days after transplant. **FIG. 7B** depicts representative dot plots of the flow cytometry strategy used. The expression intensity of the CD71 and Ter119 markers allows for identifying four erythroid subpopulations. population I: early proerythroblasts ($Ter119^{med}$ $CD71^{high}$), population II: basophilic erythroblasts ($Ter119^{high}$ $CD71^{high}$), population III: late basophilic and polychromatophilic erythroblasts ($Ter119^{high}$ $CD71^{med}$) and population IV: orthochromatophilic erythroblasts, reticulocytes and mature erythroid cells ($Ter119^{high}$ $CD71^{low}$). **FIG. 7C** shows plasma Epo levels measured by ELISA in non-transplanted and transplanted mice. Dots represent values of individual mice. Lines represent average ± SEM and were analyzed by non-parametric Kruskal-Wallis test. Healthy, non-transplanted control mice; PKD, non-transplanted PKD mice; EGFP, PKD mice expressing the EGFP transgene; coRPK, PKD mice expressing the therapeutic transgene.

[0029] FIGS. 8A-8B show hematopoietic progenitor assays in control mice and transplanted mice with transduced cells. The data demonstrate total CFUs from spleen (**FIG. 8A**) and bone marrow (**FIG. 8B**) at 140 days after transplant. Dots represent number of colonies per mouse analyzed and lines represent average ± SEM in each group. Data were statistically analyzed by non-parametric Kruskal-Wallis test.

[0030] FIGS. 9A-9C demonstrate reversion of splenomegaly and organ pathology in genetically corrected mice at 140 days post-transplantation. **FIG. 9A** shows pictures of representative spleens, and **FIG. 9B** shows ratio of spleen weight to total body weight from primary and secondary transplanted PKD mice. Dots represent values of individual mice. Lines represent average ± SEM per group. Data were analyzed by non-parametric Kruskal-Wallis test. **FIG. 9C** shows the histological study of spleen and liver from primary transplanted PKD mice. First and second column show the representative histology sections of spleen and liver stained with hematoxylin-eosin and photographed using a 4x and 10x objective, respectively,

in a light microscope. Arrows point to erythroid cell clusters indicative of extramedullary erythropoiesis. Third column shows Prussian blue staining (Fe) of liver sections to detect iron deposits indicated by arrowheads. Photographs were taken using a 20x objective. Group legends as in **Fig. 7**. 2nd coRPK, secondary recipients.

[0031] FIGS. 10A-10G depict metabolic profiling in RBC samples from mice transplanted with genetically modified cells. Analysis of significant metabolic profile changes in healthy and transplanted mice by comparison to PKD animals in two independent experiments. **FIG. 10A** shows the complete RBC heat map obtained by untargeted profiling, where higher and lower metabolite levels are represented in red and blue respectively. Metabolites listed have at least one comparison that is significant using the following criteria: absolute fold change>1.5; minimal signal>2000; Adjusted p-value<0.01. Black boxes highlight cluster of metabolite changes with distinct profile among the groups. **FIGS. 10B, 10C, and 10D** depict ATP, ADP and pyruvate levels in RBCs, respectively, measured by untargeted profiling by comparison to PKD mice at 140 days after transplant. Assay 1: Healthy mice (black bars) n=1, PKD (grey bars) n=1, hPGK-EGFP (white bars) n=2, hPGK-coRPK (scratched bars) n=3. Assay 2: Healthy mice n=2, PKD n=2, hPGK-EGFP n=6, hPGK-coRPK n=10. **FIGS. 10Ee, 10F, and 10G** depict RBC targeted metabolic profiling of a selected number of metabolites involved in the glycolytic pathway (PEP, 3-phosphoglyceric acid and D-lactic acid, respectively) at 280 days post-transplantation. Dots represent values of individual mice. Lines represent average ± SEM and were analyzed by non-parametric Kruskal-Wallis test. Assay 2: Healthy mice n=7, PKD n=5, hPGK-EGFP n=3, hPGK-coRPK n=5.

[0032] FIGS. 11A-11C depict pyruvate kinase activity, hexokinase activity and ratio of pyruvate kinase and hexokinase enzymatic activities, respectively, in RBCs from control mice and mice transplanted with transduced cells. RBCs were purified from blood samples through a cellulose column to avoid leukocyte PK activity contamination and subjected to enzyme activity evaluation. Black bars, healthy mice (n=2); white bars, mice transplanted with cells transduced with the EGFP expressing vector (n=3); scratched bars, mice transplanted with cells transduced with the coRPK expressing vector (n=3). Checkered bars represent values from a healthy volunteer (n=1). Data represent the average ± SEM of each group.

[0033] FIGS. 12A-12D show untargeted metabolic profiling in WBC samples from mice transplanted with genetically modified cells. **FIG. 12A** represents the principal component analysis of untargeted metabolite profile in RBCs (red dots; left and center) and WBCs (blue dots; cluster on right) in control and transplanted mice. **FIGS. 12B, 12C, and 12D** depict ATP,

ADP and pyruvate levels, respectively, in WBCs by comparison to PKD mice. Assay 1: healthy mice (black bars) n=1, PKD (grey bars) n=1, hPGK-EGFP (white bars) n=2, hPGK-coRPK (scratched bars) n=3. Assay 2: healthy mice n=2, PKD n=2, hPGK-EGFP n=6, hPGK-coRPK n=10. Data represent the average ± SEM per group and were analyzed by non-parametric Kruskal-Wallis test.

[0034] FIG. 13 shows a gel image of LAM-PCR products generated with Tsp509I enzyme for samples harvested from all mice at different time points and tissues. Vector integration sites were identified by LAM-PCR amplification of 3'vector LTR-genome junctions. A MultiNA automated system was used, generating a pattern characterized by several bands. Vector backbone derived Tsp509I internal control band (IC) is indicated by an arrow.

[0035] FIG. 14 shows a gel image of LAM-PCR products generated with HpyCH4IV5 enzyme for samples harvested from all mice at different time points and tissues. Vector integration sites were identified by LAM-PCR amplification of 3' vector LTR-genome junctions. A MultiNA automated system was used, generating a pattern characterized by several bands. Vector backbone derived HpyCH4IV5 IC is indicated by an arrow.

[0036] FIG. 15 depicts a general scheme of the analysis of integration site mapping performed in mice transplanted with genetically modified hematopoietic progenitors. Bone marrow and white blood cell samples from transplanted mice belonging to two independent experiments (**Table 3**) and harvested at different time-points after transplant were analyzed as described in supplementary methods following the showed pipeline.

[0037] FIGS. 16A-16B show the distribution of LV integrations along the genome of transplanted mice. **FIG. 16A** depicts integration site (IS) frequency distribution around Transcription Start Site (TSS) of the nearest RefSeq gene, spanning 500 Kb upstream and downstream the TSS. Numbers on the top are the number of IS detected for all samples and time-points. **FIG. 16B** depicts chromosomal distribution of LV integration sites in transplanted mice expressing the EGFP transgene (black bars) or the coRPK therapeutic transgene (grey bars), showing no skewing towards any particular chromosome.

[0038] FIGS. 17A-17C demonstrates clonal abundance analysis of coRPK-LV transduced cells. Dots plot representation of clonal abundance of pooled integrations in each mouse from assays 1 **(FIG. 17A)** and 2 **(FIGS. 17B and 17C)**. The relative percentage (y-axis) for each IS is relative to the total number of sequences reads obtained in each dataset. BM, bone marrow; PB, peripheral blood; coRPK1-14, mice transplanted with hematopoietic cells transduced with the therapeutic vector.

[0039] FIG. 18 presents the tracked shared integrations between primary and secondary recipient mice carrying the therapeutic PGK-coRPK LV vector. Integrations detected in either mouse in any organ and at any time are pooled. Secondary recipients received the pooled BM from transplanted mice coRPK11 to 14. The rest of the IS detected were detected or in the primary or in the secondary recipients. Numbers in the boxes show the representativeness in percentage of the corresponding integration in the referred mouse. In addition to ≥ 5% filter applied on integration analysis, all integration with a sequence count < 3 were eliminated.

[0040] FIG. 19 demonstrates clonal abundance analysis of EGFP- LV transduced cells. Dots plot representation of clonal abundance of pooled integrations in each mouse in bone marrow**.** The relative percentage (y-axis) for each IS is relative to the total number of sequences reads obtained in each dataset. Similarly to co-RPK transduced cells (**Fig. 17**), the graph indicates that the vast majority of transplanted mice show a polyclonal pattern of hematopoietic repopulation.

[0041] FIG. 20 depicts the LV genomic integration profile. Gene Ontology (GO) analysis was performed using the GREAT software on samples from transplanted mouse. All integrations retrieved from this study (N=2220) showed overrepresentations of the gene functions indicated on the left part of the figure. To address if the most abundant integrations were enriched on specific gene classes, all ISs with a relative sequence count >5% of the entire dataset (shown in **Fig. 17**) were selected, showing no GO gene classes overrepresented.

[0042] FIG. 21 depicts a schematic diagram of the PGK-coRPK LV lentiviral vector.

[0043] FIG. 22 depicts a diagram of a plasmid (CPcoRPKW-17) comprising the PGK-coRPK LV lentiviral vector.

[0044] FIGS. 23A-23H represent the nucleotide sequence of the PGK-coRPK LV lentiviral vector, annotated to show the location of the various functional elements.

[0045] FIGS. 24A-24B depicts the mechanism of action of the PGK-coRPK LV lentiviral vector. **FIG. 24A** depicts the ectopic expression of the PGK-coRPK LV vector will rescue the wild-type phenotype of PKD erythrocytes, otherwise unable to generate a functional RPK protein to produce sufficient energy to carry out their functions. **FIG. 24B** shows the gene therapy strategy for PKD patients based on the *ex vivo* transduction of RPK deficient $CD34^+$ hematopoietic progenitors with the PGK-coRPK LV lentiviral vector and subsequent transplantation into the patient. The developed PGK-coRPK LV lentiviral vector carrying the therapeutic human *PKLR* gene cDNA will be integrated in the patient $CD34^+$ cell genome upon *ex vivo* transduction. These genetically corrected cells will be then infused back into the patient, where they will produce RBCs expressing the therapeutic transgene, and therefore, producing

functional RPK proteins that will correct the PKD pathological phenotype. Figure modified from the Boston Children's Hospital blog.

DETAILED DESCRIPTION OF THE INVENTION

DEFINITIONS

[0046] A "vector" as used herein refers to a macromolecule or association of macromolecules that comprises or associates with a polynucleotide and which can mediate delivery of the polynucleotide to a cell. Illustrative vectors include, for example, plasmids, viral vectors, liposomes, and other gene delivery vehicles.

[0047] The term "LV" is an abbreviation for lentivirus, and may refer to the virus itself or derivatives thereof. The term covers all subtypes and both naturally occurring and recombinant forms, except where required otherwise.

[0048] As used herein, the term "gene" or "coding sequence" refers to a nucleotide sequence in vitro or in vivo that encodes a gene product. In some instances, the gene consists or consists essentially of coding sequence, that is, sequence that encodes the gene product. In other instances, the gene comprises additional, non-coding, sequence. For example, the gene may or may not include regions preceding and following the coding region, *e.g.* 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).

[0049] As used herein, a "therapeutic gene" refers to a gene that, when expressed, confers a beneficial effect on the cell or tissue in which it is present, or on a mammal in which the gene is expressed. Examples of beneficial effects include amelioration of a sign or symptom of a condition or disease, prevention or inhibition of a condition or disease, or conferral of a desired characteristic. Therapeutic genes include genes that correct a genetic deficiency in a cell or mammal.

[0050] As used herein, a transgene is a gene delivered to a cell by a vector.

[0051] As used herein, the term "gene product" refers to the desired expression product of a polynucleotide sequence such as a polypeptide, peptide, protein or interfering RNA including short interfering RNA (siRNA), miRNA or small hairpin RNA (shRNA).

[0052] As used herein, the terms "polypeptide," "peptide," and "protein" refer to polymers of amino acids of any length. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, phosphorylation, or conjugation with a labeling component.

[0053] By "comprising" it is meant that the recited elements are required in, for example, the composition, method, kit, etc., but other elements may be included to form the, for example, composition, method, kit etc. within the scope of the claim. For example, an expression cassette "comprising" a gene encoding a therapeutic polypeptide operably linked to a promoter is an expression cassette that may include other elements in addition to the gene and promoter, *e.g.* poly-adenylation sequence, enhancer elements, other genes, linker domains, etc.

[0054] By "consisting essentially of", it is meant a limitation of the scope of the, for example, composition, method, kit, etc., described to the specified materials or steps that do not materially affect the basic and novel characteristic(s) of the, for example, composition, method, kit, etc. For example, an expression cassette "consisting essentially of" a gene encoding a therapeutic polypeptide operably linked to a promoter and a polyadenylation sequence may include additional sequences, *e.g.* linker sequences, so long as they do not materially affect the transcription or translation of the gene. As another example, a variant, or mutant, polypeptide fragment "consisting essentially of" a recited sequence has the amino acid sequence of the recited sequence plus or minus about 10 amino acid residues at the boundaries of the sequence based upon the full length naïve polypeptide from which it was derived, *e.g.* 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 residue less than the recited bounding amino acid residue, or 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 residues more than the recited bounding amino acid residue.

[0055] By "consisting of", it is meant the exclusion from the composition, method, or kit of any element, step, or ingredient not specified in the claim. For example, an expression cassette "consisting of" a gene encoding a therapeutic polypeptide operably linked to a promoter, and a post-transcriptional regulatory element consists only of the promoter, polynucleotide sequence encoding the therapeutic polypeptide, and post-transcriptional regulatory element. As another example, a polypeptide "consisting of" a recited sequence contains only the recited sequence.

[0056] An "expression vector" as used herein encompasses a vector, *e.g.* plasmid, minicircle, viral vector, liposome, and the like as discussed above or as known in the art, comprising a polynucleotide encoding a gene product of interest, and effecting the expression of a gene product in an intended target cell. An expression vector also comprises control elements operatively linked to the encoding region to facilitate expression of the gene product in the target. The combination of control elements, *e.g.* promoters, enhancers, UTRs, miRNA targeting sequences, etc., and a gene or genes to which they are operably linked for expression is sometimes referred to as an "expression cassette." Many such control elements are known

and available in the art or can be readily constructed from components that are available in the art.

[0057] A "promoter" as used herein encompasses a DNA sequence that directs the binding of RNA polymerase and thereby promotes RNA synthesis, *i.e.*, a minimal sequence sufficient to direct transcription. Promoters and corresponding protein or polypeptide expression may be ubiquitous, meaning strongly active in a wide range of cells, tissues and species or cell-type specific, tissue-specific, or species specific. Promoters may be "constitutive," meaning continually active, or "inducible," meaning the promoter can be activated or deactivated by the presence or absence of biotic or abiotic factors. Also included in the nucleic acid constructs or vectors of the invention are enhancer sequences that may or may not be contiguous with the promoter sequence. Enhancer sequences influence promoter-dependent gene expression and may be located in the 5' or 3' regions of the native gene.

[0058] An "enhancer" as used herein encompasses a cis-acting element that stimulates or inhibits transcription of adjacent genes. An enhancer that inhibits transcription also is termed a "silencer". Enhancers can function (*i.e.*, can be associated with a coding sequence) in either orientation, over distances of up to several kilobase pairs (kb) from the coding sequence and from a position downstream of a transcribed region.

[0059] A "termination signal sequence" as used herein encompasses any genetic element that causes RNA polymerase to terminate transcription, such as for example a polyadenylation signal sequence.

[0060] As used herein, the terms "operatively linked" or "operably linked" refers to a juxtaposition of genetic elements, *e.g.* promoter, enhancer, termination signal sequence, polyadenylation sequence, etc., wherein the elements are in a relationship permitting them to operate in the expected manner. For instance, a promoter is operatively linked to a coding region if the promoter helps initiate transcription of the coding sequence. There may be intervening residues between the promoter and coding region so long as this functional relationship is maintained.

[0061] As used herein, the term "heterologous" means derived from a genotypically distinct entity from that of the rest of the entity to which it is being compared. For example, a polynucleotide introduced by genetic engineering techniques into a plasmid or vector derived from a different species is a heterologous polynucleotide. As another example, a promoter removed from its native coding sequence and operatively linked to a coding sequence with which it is not naturally linked is a heterologous promoter. Thus, for example, an LV vector that includes a heterologous nucleic acid encoding a heterologous gene product is an LV vector

that includes a nucleic acid not normally included in a naturally-occurring, wild-type LV, and the encoded heterologous gene product is a gene product not normally encoded by a naturally-occurring, wild-type LV.

[0062] The term "endogenous" as used herein with reference to a nucleotide molecule or gene product refers to a nucleic acid sequence, *e.g.* gene or genetic element, or gene product, *e.g.* RNA, protein, that is naturally occurring in or associated with a host virus or cell.

[0063] The term "native" as used herein refers to a nucleotide sequence, *e.g.* gene, or gene product, *e.g.* RNA, protein, that is present in a wild-type virus or cell.

[0064] The term "variant" as used herein refers to a mutant of a reference polynucleotide or polypeptide sequence, for example a native polynucleotide or polypeptide sequence, *i.e.* having less than 100% sequence identity with the reference polynucleotide or polypeptide sequence. Put another way, a variant of a polynucleotide sequence comprises at least one nucleotide difference (*e.g.*, nucleotide substitution, nucleotide insertion, or nucleotide deletion) relative to a reference polynucleotide sequence, *e.g.*, a native polynucleotide sequence. A variant of a polypeptide sequence comprises at least one amino acid difference (*e.g.*, amino acid substitution, amino acid insertion, amino acid deletion) relative to a reference polypeptide sequence, *e.g.*, a native polypeptide sequence. For example, a variant may be a polynucleotide having a sequence identity of 70% or more with a full-length native polynucleotide sequence, *e.g.* an identity of 75% or 80% or more, such as 85%, 90%, or 95% or more, for example, 98% or 99% identity with the full-length native polynucleotide sequence. As another example, a variant may be a polypeptide having a sequence identity of 70% or more with a full-length native polypeptide sequence, *e.g.* an identity of 75% or 80% or more, such as 85%, 90%, or 95% or more, for example, 98% or 99% identity with the full-length native polypeptide sequence. Variants may also include variant fragments of a reference, *e.g.* native, sequence sharing a sequence identity of 70% or more with a fragment of the reference, *e.g.* native, sequence, *e.g.* an identity of 75% or 80% or more, such as 85%, 90%, or 95% or more, for example, 98% or 99% identity with the native sequence.

[0065] As used herein, the terms "biological activity" and "biologically active" refer to the activity attributed to a particular biological element in a cell. For example, the "biological activity" of an "immunoglobulin", "antibody" or fragment or variant thereof refers to the ability to bind an antigenic determinant and thereby facilitate immunological function. As another example, the biological activity of a polypeptide or functional fragment or variant thereof refers to the ability of the polypeptide or functional fragment or variant thereof to carry out its native functions of, *e.g.*, binding, enzymatic activity, etc. As a third example, the biological activity

of a gene regulatory element, *e.g.* promoter, enhancer, Kozak sequence, and the like, refers to the ability of the regulatory element or functional fragment or variant thereof to regulate, *i.e.* promote, enhance, or activate the translation of, respectively, the expression of the gene to which it is operably linked.

[0066] The terms "administering" or "introducing", as used herein, refer to delivery of a vector for recombinant protein expression to a cell, to cells and/or organs of a subject, or to a subject. Such administering or introducing may take place in vivo, in vitro or ex vivo. A vector for expression of a gene product may be introduced into a cell by transfection, which typically means insertion of heterologous DNA into a cell by physical means (*e.g.*, calcium phosphate transfection, electroporation, microinjection or lipofection); infection, which typically refers to introduction by way of an infectious agent, *i.e.*, a virus; or transduction, which typically means stable infection of a cell with a virus or the transfer of genetic material from one microorganism to another by way of a viral agent (*e.g.*, a bacteriophage).

[0067] "Transformation" is typically used to refer to bacteria comprising heterologous DNA, or, cells that express an oncogene and have been converted into a continuous growth mode such as tumor cells. A vector used to "transform" a cell may be a plasmid, virus or other vehicle.

[0068] Typically, a cell is referred to as "transduced", "infected"; "transfected" or "transformed" dependent on the means used for administration, introduction or insertion of heterologous DNA (*i.e.*, the vector) into the cell. The terms "transduced", "transfected" and "transformed" may be used interchangeably herein regardless of the method of introduction of heterologous DNA.

[0069] The term "host cell", as used herein refers to a cell, which has been transduced, infected, transfected or transformed with a vector. The vector may be a plasmid, a viral particle, a phage, etc. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art. It will be appreciated that the term "host cell" refers to the original transduced, infected, transfected or transformed cell and progeny thereof.

[0070] The terms "treatment", "treating" and the like are used herein to generally mean obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof, *e.g.* reducing the likelihood that the disease or symptom thereof occurs in the subject, and/or may be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease in a mammal, and includes: (a) preventing the disease from occurring in a subject which may be predisposed to

the disease but has not yet been diagnosed as having it; (b) inhibiting the disease, *i.e.*, arresting its development; or (c) relieving the disease, *i.e.*, causing regression of the disease. The therapeutic agent may be administered before, during or after the onset of disease or injury. The treatment of ongoing disease, where the treatment stabilizes or reduces the undesirable clinical symptoms of the patient, is of particular interest. Such treatment is desirably performed prior to complete loss of function in the affected tissues. The subject therapy will desirably be administered during the symptomatic stage of the disease, and in some cases after the symptomatic stage of the disease.

[0071] The terms "individual," "host," "subject," and "patient" are used interchangeably herein, and refer to a mammal, including, but not limited to, human and non-human primates, including simians and humans; mammalian sport animals (*e.g.*, horses); mammalian farm animals (*e.g.*, sheep, goats, etc.); mammalian pets (dogs, cats, etc.); and rodents (*e.g.*, mice, rats, etc.).

[0072] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. Furthermore, to the extent that the terms "including", "includes", "having", "has", "with", or variants thereof are used in either the detailed description and/or the claims, such terms are intended to be inclusive in a manner similar to the term "comprising".

[0073] The term "about" or "approximately" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, *i.e.*, the limitations of the measurement system. For example, "about" can mean within 1 or more than 1 standard deviation, per the practice in the art. Alternatively, "about" can mean a range of up to 20%, preferably up to 10%, more preferably up to 5%, and more preferably still up to 1% of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, preferably within 5-fold, and more preferably within 2-fold, of a value. Where particular values are described in the application and claims, unless otherwise stated the term "about" meaning within an acceptable error range for the particular value should be assumed.

[0074] Unless otherwise indicated, all terms used herein have the same meaning as they would to one skilled in the art and the practice of the present invention will employ conventional techniques of microbiology and recombinant DNA technology, which are within the knowledge of those of skill of the art.

[0075] The practice of the present invention employs, unless otherwise indicated, conventional techniques of cell biology, molecular biology (including recombinant techniques), microbiology, biochemistry and immunology, which are within the scope of those of skill in the art. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al., 1989); "Oligonucleotide Synthesis" (M. J. Gait, ed., 1984); "Animal Cell Culture" (R. I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Handbook of Experimental Immunology" (D. M. Weir & C. C. Blackwell, eds.); "Gene Transfer Vectors for Mammalian Cells" (J. M. Miller & M. P. Calos, eds., 1987); "Current Protocols in Molecular Biology" (F. M. Ausubel et al., eds., 1987); "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994); and "Current Protocols in Immunology" (J. E. Coligan et al., eds., 1991), each of which is expressly incorporated by reference herein.

[0076] In certain embodiments, the present disclosure provides polynucleotides, polynucleotide cassettes and expression vectors for the expression of a gene in cells. Also provided are pharmaceutical compositions and methods for the use of any of the compositions in promoting the expression of a gene in cells, for example, in an individual, *e.g.* for the treatment or prophylaxis of a disorder. These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the compositions and methods as more fully described below.

[0077] The present invention relates generally to the fields of molecular biology and virology, and in particular, to genetic expression cassettes, and vectors comprising them useful for the delivery of nucleic acid segments encoding selected therapeutic constructs (including for example, peptides, polypeptides, ribozymes, and catalytic RNA molecules), to selected cells and tissues of vertebrate animals. In particular, these genetic constructs are useful in the development of gene therapy vectors, including for example, LV vectors, for the treatment of mammalian, and in particular, human, diseases, disorders, and dysfunctions.

[0078] The disclosed compositions may be utilized in a variety of investigative, diagnostic and therapeutic regimens, including the prevention and treatment of a variety of human diseases. The various compositions and methods of the invention are described below.

[0079] Although particular compositions and methods are exemplified herein, it is understood that any of a number of alternative compositions and methods are applicable and suitable for use in practicing the invention. It will also be understood that an evaluation of the expression constructs and methods of the invention may be carried out using procedures standard in the art.

[0080] In certain embodiments, methods and compositions are provided for preparation of gene therapy vector compositions, *e.g.*, viral vectors, comprising these genetic expression cassettes for use in the preparation of medicaments useful in central and targeted gene therapy of diseases, disorders, and dysfunctions in an animal, and in humans in particular.

[0081] In some embodiments, the present invention provides for gene therapy for PKD based on a LV vector harbouring the PGK eukaryotic promoter that drives the expression of the *PKLR* cDNA. This therapeutic vector may be used to transduce hematopoietic stem cells (HSCs), which may be subsequently transplanted into a subject. Ectopic RPK expression normalizes the erythroid compartment correcting the hematological phenotype and reverting organ pathology. Metabolomic studies demonstrate functional correction of the glycolytic pathway in RBCs derived from genetically corrected PKD HSCs, with no metabolic disturbances in leukocytes. The analysis of the LV ISs in the genome of transplanted hematopoietic cells demonstrates no evidence of genotoxicity in any of the transplanted animals. Overall, the results underscore the therapeutic potential of the PGK-coRPK LV vector and provide high expectations towards the gene therapy of PKD and other erythroid metabolic genetic disorders.

[0082] In certain embodiments, the present invention provides an RPK LV vector for the genetic correction of PKD. Genetic modification of murine PKD-HSCs with this vector can efficiently correct the hemolytic phenotype and the RBC metabolite profile in transplanted PKD mice. No evidence of metabolic disturbances in leukocytes and genotoxicity derived from the vector integration are observed, supporting the therapeutic potential of the PGK-coRPK LV vector. Overall, results provide encouraging evidence of the feasibility of gene therapy for PKD with a LV designed for clinical application.

[0083] Certain embodiments of the present invention comprise a self-inactivating (SIN) LV vector expressing a codon-optimized version of human *PLKR* gene. The expression vector comprises a promoter region, a coding sequence, and a post-transcriptional regulatory element.

[0084] Certain embodiments of polynucleotide cassettes of the present invention comprise a promoter region comprising a promoter sequence, or a functional fragment thereof. In one embodiment, the promoter is a human PGK promoter.

[0085] Some embodiments of the present invention comprise polynucleotide cassettes for the enhanced expression of pyruvate kinase (*e.g.*, human *PLKR* gene). In some embodiments, the polynucleotide cassette comprises a codon-optimized version of the human *PKLR* cDNA (coRPK) to increase mRNA stability upon transcription. For the optimization, GeneArt® software may be used, increasing the GC content and removing cryptic splice sites in order to avoid transcriptional silencing and therefore increase transgene expression. The coRPK

optimized sequence showed 80.4% homology with the human *PKLR* gene, with no changes in the amino acids sequence of the protein. Alternatively, any optimization method known in the art may be used.

[0086] In some embodiments, the polynucleotide cassette comprises an RNA export signal downstream of the second enhancer. The RNA export signal may comprise a wPRE sequence. In some embodiments, a mutated wPRE, lacking any residual open reading frame (Schambach, Bohne et al. 2006) is also included to improve the level of expression and stability of the therapeutic gene. The term "RNA export signal" may refer to any of the "post-transcriptional regulatory elements" known in the art. The terms "RNA export signal" and "post-transcriptional regulatory element" and the like are used interchangeably herein, as they are generally in the field of vector technology and development. Exemplary RNA export signals or post-transcriptional regulatory elements used in some embodiments of the present disclosure include without limitation, hepatitis B virus posttranscriptional regulatory element (HBVPRE) or a constitutive transport element (CTE) originating from different simian retroviruses, including simian retrovirus type 1 (SRV-1) and type-2 (SRV-2) and Mason-Pfizer (MPV). Post-transcriptional regulatory elements are provided by U.S. Patent No. 6,136,597; Donello et al. J. Virol. 72:5085-92 (1998); Hlavaty et al. *Virology* 341:1-11 (2005); and Oh et al. Retrovirology 4:38 (2007). Further alternative wPRE sequences are provided by Zanta-Boussif et al. *Gene Therapy* 16:605-19 (2009). The disclosures of the foregoing reference are incorporated herein in their entirety.

[0087] In some aspects of the invention, gene delivery vectors are provided comprising a polynucleotide cassette of the present invention. In some embodiments, the gene delivery vector is a LV.

[0088] In some aspects of the invention, pharmaceutical compositions are provided comprising a polynucleotide cassette of the invention and a pharmaceutical excipient. In some embodiments, the pharmaceutical composition comprises a gene delivery vector of the invention and a pharmaceutical excipient.

[0089] In some aspects of the invention, methods are provided for expressing a transgene in mammalian cells. In some embodiments, the method comprises contacting one or more mammalian cells with an effective amount of a polynucleotide cassette of the invention or a gene delivery vector of the invention, wherein the transgene is expressed at detectable levels in the one or more mammalian cells. In some embodiments, the method comprises contacting one or more mammalian cells with an effective amount of a polynucleotide cassette of the invention or a gene delivery vector of the invention, wherein the transgene is expressed at

therapeutic levels in the one or more mammalian cells. In some embodiments, the method is in vitro. In other embodiments, the method is in vivo.

[0090] In some aspects of the invention, methods are provided for the treatment or prophylaxis of a disease or disorder in a mammal in need of treatment or prophylaxis for a disease or disorder. In some embodiments, the method comprises administering to the mammal an effective amount of a pharmaceutical composition of the invention, wherein the coding sequence encodes a therapeutic gene product.

COMPOSITIONS

[0091] In some aspects of the disclosure, compositions are provided for the expression of a transgene in a eukaryotic cell(s). In some aspects, the eukaryotic cell is a mammalian cell. In some aspects, the mammalian cell is a hematopoietic stem cell. In some embodiments, the cell is a bone marrow cell, *e.g.*, a lineage depleted bone marrow cell. In some aspects, the mammalian cell is a committed hematopoietic erythroid progenitor.

[0092] In some embodiments of the disclosure, the composition is a polynucleotide cassette. By a "polynucleotide cassette" is meant a polynucleotide sequence comprising two or more functional polynucleotide sequences, *e.g.* regulatory elements, translation initiation sequences, coding sequences, and/or termination sequences, etc., typically in operable linkage to one another. Likewise, by a "polynucleotide cassette for the expression of a transgene in a mammalian cell," it is meant a combination of two or more functional polynucleotide sequences, *e.g.* promoter, enhancer, 5'UTR, translation initiation sequence, coding sequence, and/or termination sequences, etc. that promotes the expression of the transgene in a cell.

[0093] In some embodiments, the polynucleotide cassettes of the present disclosure provide for enhanced expression of a transgene in mammalian cells. As used herein "expression of a transgene" or "expressing a transgene" refers to both the transcription of the transgene into a messenger RNA (mRNA) by the host cell and translation of that mRNA into a polypeptide (*i.e.*, the transgene product). Thus, expression of the transgene can be measured either at the mRNA level (*i.e.*, by sequence-specific quantification of the level of mRNA in a cell) or at the polypeptide level (*i.e.*, by measuring the level of polypeptide gene product using a Western Blot, Enzyme-Linked Immunoabsorbent Assay (ELISA), Immunofluorescence Microscopy, or other means of quantifying a specific polypeptide). As demonstrated by the working examples of the present disclosure, the present inventors have discovered a number of polynucleotide elements, *i.e.*, improved elements as compared to those known in the art, which individually and synergistically provide for the enhanced expression of transgenes in mammalian cells. In

certain embodiments, the arrangement of the two or more functional polynucleotide sequences within the polynucleotide cassettes of the present disclosure provide for enhanced expression of a transgene in mammalian cells. By "enhanced" it is meant that expression of the transgene is increased, augmented, or stronger, in cells carrying the polynucleotide cassettes of the present disclosure relative to in cells carrying the transgene operably linked to comparable regulatory elements, *e.g.* as known in the art. Put another way, expression of the transgene is increased, augmented, or stronger, from the polynucleotide cassettes of the present disclosure relative to expression from a polynucleotide cassette not comprising the one or more optimized elements of the present disclosure, *i.e.* a reference control. In certain embodiments, the enhanced expression is specific for or limited to one or more desired cell types.

[0094] For example, expression of the transgene may be enhanced, or augmented, or increased, in cells comprising a polynucleotide cassette comprising a promoter disclosed herein than in cells that carry the transgene operably linked to a different promoter, *e.g.* as known in the art. As another example, expression of the transgene may be enhanced, or increased, augmented, or stronger, in cells comprising a polynucleotide cassette comprising an enhancer sequence disclosed herein than in cells that carry the transgene operably linked to a different enhancer sequence.

[0095] Without wishing to be bound by theory, enhanced expression of a transgene in cells is believed to be due to a more rapid build-up of gene product in the cells or a more stable gene product in the cells. Thus, enhanced expression of a transgene by the polynucleotide cassettes of the subject disclosure may be observed in a number of ways. For example, enhanced expression may be observed by detecting the expression of the transgene following contact of the polynucleotide cassette to the cells sooner, *e.g.* 2 days sooner, 7 days sooner, 2 weeks sooner, 3 weeks sooner, 4 weeks sooner, 8 weeks sooner, 12 weeks sooner or more, than expression would be detected if the transgene were operably linked to comparable regulatory elements, *e.g.*, as known in the art. Enhanced expression may also be observed as an increase in the amount of gene product per cell. For example, there may be a 2-fold increase or more, *e.g.* a 3-fold increase or more, a 4-fold increase or more, a 5-fold increase or more, or a 10-fold increase or more in the amount of gene product per mammalian cell. Enhanced expression may also be observed as an increase in the number of mammalian cells that express detectable levels of the transgene carried by the polynucleotide cassette. For example, there may be a 2-fold increase or more, *e.g.* a 3-fold increase or more, a 4-fold increase or more, a 5-fold increase or more, or a 10-fold increase or more in the number of mammalian cells that express detectable levels of the transgene.

[0096] As another example, the polynucleotide of the present invention may promote detectable levels of the transgene in a greater percentage of cells as compared to a conventional polynucleotide cassette; for example, where a conventional cassette may promote detectable levels of transgene expression in, for example, less than 5% of the cells in a certain region, the polynucleotide of the present invention promotes detectable levels of expression in 5% or more of the cells in that region; *e.g.* 10% or more, 15% or more, 20% or more, 25% or more, 30% or more, 35% or more, 40% or more, or 45% or more, in some instances 50% or more, 55% or more; 60% or more, 65% or more, 70% or more, or 75% or more, for example 80% or more, 85% or more, 90% or more, or 95% or more of the cells that are contacted, will express detectable levels of gene product. Enhanced expression may also be observed as an alteration in the viability and/or function of the cells.

[0097] The polynucleotide cassettes of the present disclosure typically comprise a promoter region. Any suitable promoter region or promoter sequence therein can be used in the subject polynucleotide cassettes, so long as the promoter region promotes expression of a coding sequence in eukaryotic cells. In certain embodiments, the promoter region promotes expression of a coding sequence in mammalian cells. In some instances, the promoter is a ubiquitous promoter, *i.e.*, it is a promoter that is active in a wide range of cells, tissues and species. In other instances, the promoter is a human PGK promoter.

[0098] Promoter and enhancer elements can be tissue specific or stage-specific. For example, a tissue–specific promoter or enhancer preferentially drives expression (or a higher level of expression) in one or more particular cell types. Examples of cell types include but are not limited to: hematopoietic stem cells, long term hematopoietic stem cells, short term hematopoietic stem cells, multipotent progenitors, hematopoietic CD34+ cells and any cluster of differentiation subpopulation within the CD34+ population. A stage-specific promoter or enhancer preferentially drives expression (or higher level of expression) during one or more specific stages of the cell cycle or development. These include but are not limited to beta-globin locus control region, spectrin promoter, and an erythroid specific promoter.

[0099] In some embodiments, the polynucleotide comprises one or more enhancers. Enhancers are nucleic acid elements known in the art to enhance transcription, and can be located anywhere in association with the gene they regulate, *e.g.* upstream, downstream, within an intron, etc. Any enhancer element can be used in the polynucleotide cassettes and gene therapy vectors of the present disclosure, so long as it enhances expression of the gene when used in combination with the promoter.

[0100] The coding sequence to be expressed in the cells can be any polynucleotide sequence, *e.g.* gene or cDNA that encodes a gene product, *e.g.* a polypeptide or RNA-based therapeutic (siRNA, antisense, ribozyme, shRNA, etc.). The coding sequence may be heterologous to the promoter sequence to which it is operably linked, *i.e.* not naturally operably associated with it. Alternatively, the coding sequence may be endogenous to the promoter sequence to which it is operably linked, *i.e.* is associated in nature with that promoter. The gene product may act intrinsically in the mammalian cell, or it may act extrinsically, *e.g.*, it may be secreted. For example, when the transgene is a therapeutic gene, the coding sequence may be any gene that encodes a desired gene product or functional fragment or variant thereof that can be used as a therapeutic for treating a disease or disorder. In various preferred embodiments, the transgene encodes human PKLR.

[0101] In one embodiment of the invention, the transgene coding sequence is modified, or "codon optimized" to enhance expression by replacing infrequently represented codons with more frequently represented codons. The coding sequence is the portion of the mRNA sequence that encodes the amino acids for translation. During translation, each of 61 trinucleotide codons are translated to one of 20 amino acids, leading to a degeneracy, or redundancy, in the genetic code. However, different cell types, and different animal species, utilize tRNAs (each bearing an anticodon) coding for the same amino acids at different frequencies. When a gene sequence contains codons that are infrequently represented by the corresponding tRNA, the ribosome translation machinery may slow, impeding efficient translation. Expression can be improved via "codon optimization" for a particular species, where the coding sequence is altered to encode the same protein sequence, but utilizing codons that are highly represented, and/or utilized by highly expressed human proteins (Cid-Arregui et al., 2003; J. Virol. 77: 4928). In one aspect of the present invention, the coding sequence of the transgene is modified to replace codons infrequently expressed in mammal or in primates with codons frequently expressed in primates. For example, in some embodiments, the coding sequence encoded by the transgene encodes a polypeptide having at least 85% sequence identity to a polypeptide encoded by a sequence disclosed above or herein, for example at least 90% sequence identity, *e.g.* at least 95% sequence identity, at least 98% identity, at least 99% identity, wherein at least one codon of the coding sequence has a higher tRNA frequency in humans than the corresponding codon in the sequence disclosed above or herein.

[0102] In an additional embodiment of the invention, the transgene coding sequence is modified to enhance expression by termination or removal of open reading frames (ORFs) that do not encode the desired transgene. An open reading frame (ORF) is the nucleic acid sequence

that follows a start codon and does not contain a stop codon. ORFs may be in the forward or reverse orientation, and may be "in frame" or "out of frame" compared with the gene of interest. Such open reading frames have the potential to be expressed in an expression cassette alongside the gene of interest, and could lead to undesired adverse effects. In one aspect of the present invention, the coding sequence of the transgene has been modified to remove open reading frames by further altering codon usage. This was done by eliminating start codons (ATG) and introducing stop codons (TAG, TAA, or TGA) in reverse orientation or out-of-frame ORFs, while preserving the amino acid sequence and maintaining highly utilized codons in the gene of interest (*i.e.*, avoiding codons with frequency < 20%). In the present invention, the transgene coding sequence may be optimized by either of codon optimization and removal of non-transgene ORFs or using both techniques. As will be apparent to one of ordinary skill in the art, it is preferable to remove or minimize non-transgene ORFs after codon optimization in order to remove ORFs introduced during codon optimization.

[0103] The RRE sequence improves the efficiency of gene transfer. In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the RRE sequence comprises or consists of the following sequence, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:

AGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGC
GTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCA
GCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCAC
AGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCT
AAAGGATCAACAGCTCCT (SEQ ID NO:1).

[0104] The retroviral leader region contains the packaging signal (Ψ), which is involved in packaging the retroviral genome into the viral capsid. LV vectors were thought to require approximately 300 bp of the *Gag* gene in this region. Currently, this *Gag* sequence has been reduced to just 40 bp (Figure 1). In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the ψ sequence is an HIV-1 ψ sequence or the ψ sequence comprises or consists of the following sequence, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:

CTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGG
CGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGA
GATGGGTGCGAGAGCGTC (SEQ ID NO:2).

[0105] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the truncated HIV-1 5' LTR comprises or consists of the following sequence, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:
GGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGA
ACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGC
CCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGT
GGAAAATCTCTAGCA (SEQ ID NO:3).

[0106] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the HIV-1 self-inactivating 3' LTR comprises or consists of the following sequence, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:
TGGAAGGGCTAATTCACTCCCAACGAAGACAAGATCTGCTTTTTGCTTGTACTGG
GTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAAC
CCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCC
GTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTG
GAAAATCTCTAGCA (SEQ ID NO:4).

[0107] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the human cytomegalovirus (CMV) immediate early promoter comprises or consists of the following sequence, a functional fragment thereof, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:
GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGG
GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAA
TCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGC
GGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT (SEQ ID NO: 5).

[0108] The cPPT, which facilitates nuclear translocation of the pre-integration complexes, together with the CTS involved in the separation of reverse transcriptase, has been seen to improve viral titer (Zennou, et al. 2000; Follenzi et al. 2000). In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the central polypurine tract and central termination sequence of HIV-1 (cPPT/CTS) comprises or consists of the following sequence, a functional fragment thereof, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:

TTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAG
ACATAATAGCAACAGACATACAAACTAAAGAATTACAAAAACAAATTACAAAA
ATTCAAAATTTT (SEQ ID NO: 6).

[0109] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the human phosphoglycerate kinase 1 (hPGK) promoter comprises or consists of the following sequence, a functional fragment thereof, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:

GGGGTTGGGGTTGCGCCTTTTCCAAGGCAGCCCTGGGTTTGCGCAGGGACGCGG
CTGCTCTGGGCGTGGTTCCGGGAAACGCAGCGGCGCCGACCCTGGGTCTCGCAC
ATTCTTCACGTCCGTTCGCAGCGTCACCCGGATCTTCGCCGCTACCCTTGTGGGCC
CCCCGGCGACGCTTCCTGCTCCGCCCCTAAGTCGGGAAGGTTCCTTGCGGTTCGC
GGCGTGCCGGACGTGACAAACGGAAGCCGCACGTCTCACTAGTACCCTCGCAGA
CGGACAGCGCCAGGGAGCAATGGCAGCGCGCCGACCGCGATGGGCTGTGGCCA
ATAGCGGCTGCTCAGCAGGGCGCGCCGAGAGCAGCGGCCGGGAAGGGGCGGTG
CGGGAGGCGGGGTGTGGGGCGGTAGTGTGGGCCCTGTTCCTGCCCGCGCGGTGT
TCCGCATTCTGCAAGCCTCCGGAGCGCACGTCGGCAGTCGGCTCCCTCGTTGACC
GAATCACCGACCTCTCTCCCCAG (SEQ ID NO:7).

[0110] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the codon-optimized version of the human *PKLR* cDNA (coRPK) comprises or consists of the following sequence, a functional fragment thereof, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:

ATGAGCATCCAGGAAAATATCAGCTCTCTGCAGCTGCGGTCCTGGGTGTCCAAG
AGCCAGAGAGACCTGGCCAAGAGCATCCTGATCGGAGCCCCTGGCGGACCAGCC
GGATACCTGAGAAGGGCTAGCGTGGCCCAGCTGACCCAGGAACTGGGCACCGCC
TTTTTCCAGCAGCAGCAGCTGCCAGCCGCCATGGCCGACACCTTTCTGGAACACC
TGTGCCTGCTGGACATCGACTCTGAGCCCGTGGCCGCCAGAAGCACCAGCATCAT
TGCCACCATCGGCCCTGCCAGCAGAAGCGTGGAGCGGCTGAAAGAGATGATCAA
GGCCGGCATGAATATCGCCCGGCTGAACTTCTCCCACGGCAGCCACGAGTACCA
CGCAGAGAGCATTGCCAACGTCCGGGAGGCCGTGGAGAGCTTTGCCGGCAGCCC
CCTGAGCTACAGACCCGTGGCCATTGCCCTGGACACCAAGGGCCCCGAGATCAG
AACAGGAATTCTGCAGGGAGGGCCTGAGAGCGAGGTGGAGCTGGTGAAGGGCA
GCCAAGTGCTGGTGACCGTGGACCCCGCCTTCAGAACCAGAGGCAACGCCAACA

CAGTGTGGGTGGACTACCCCAACATCGTGCGGGTGGTGCCTGTGGGCGGCAGAA
TCTACATCGACGACGGCCTGATCAGCCTGGTGGTGCAGAAGATCGGACCTGAGG
GCCTGGTGACCCAGGTCGAGAATGGCGGCGTGCTGGGCAGCAGAAAGGGCGTGA
ATCTGCCAGGCGCCCAGGTGGACCTGCCTGGCCTGTCTGAGCAGGACGTGAGAG
ACCTGAGATTTGGCGTGGAGCACGGCGTGGACATCGTGTTCGCCAGCTTCGTGCG
GAAGGCCTCTGATGTGGCCGCCGTGAGAGCCGCTCTGGGCCCTGAAGGCCACGG
CATCAAGATCATCAGCAAGATCGAGAACCACGAGGGCGTGAAGCGGTTCGACGA
GATCCTGGAAGTGTCCGACGGCATCATGGTGGCCAGAGGCGACCTGGGCATCGA
GATCCCCGCCGAGAAGGTGTTCCTGGCCCAGAAAATGATGATCGGACGGTGCAA
CCTGGCCGGCAAACCTGTGGTGTGCGCCACCCAGATGCTGGAAAGCATGATCAC
CAAGCCCAGACCCACCAGAGCCGAGACAAGCGACGTGGCCAACGCCGTGCTGGA
TGGCGCTGACTGCATCATGCTGTCCGGCGAGACAGCCAAGGGCAACTTCCCCGT
GGAGGCCGTGAAGATGCAGCACGCCATTGCCAGAGAAGCCGAGGCCGCCGTGTA
CCACCGGCAGCTGTTCGAGGAACTGCGGAGAGCCGCCCCTCTGAGCAGAGATCC
CACCGAAGTGACCGCCATCGGAGCCGTGGAAGCCGCCTTCAAGTGCTGCGCCGC
TGCAATCATCGTGCTGACCACCACAGGCAGAAGCGCCCAGCTGCTGTCCAGATA
CAGACCCAGAGCCGCCGTGATCGCCGTGACAAGATCCGCCCAGGCCGCTAGACA
GGTCCACCTGTGCAGAGGCGTGTTCCCCCTGCTGTACCGGGAGCCTCCCGAGGCC
ATCTGGGCCGACGACGTGGACAGACGGGTGCAGTTCGGCATCGAGAGCGGCAAG
CTGCGGGGCTTCCTGAGAGTGGGCGACCTGGTGATCGTGGTGACAGGCTGGCGG
CCTGGCAGCGGCTACACCAACATCATGAGGGTGCTGTCCATCAGCTGA (SEQ ID
NO:8).

[0111] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the human CMV enhancer comprises or consists of the following sequence, a functional fragment thereof, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:
GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC
TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGAC
GTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGG
ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG (SEQ ID
NO:9).

[0112] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the simian virus 40 (SV40) poly(A) signal comprises or consists of the following sequence, a functional fragment thereof, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence: AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATT TCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATC AATGTATCTTA (SEQ ID NO:10).

[0113] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the SV40 origin of replication comprises or consists of the following sequence, a functional fragment thereof, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence: ATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAA TTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAA GTAGTGAGGAGGCTTTTTTGGAGGCC (SEQ ID NO:11).

[0114] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the cPPT present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence: TTTAAAAGAAAAGGGGGGATTGGGGGGT (SEQ ID NO:12), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO:12.

[0115] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the dNEF signal present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence: GAATTCGAGCTCGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTT AGCCACTTTTTAAAAGAAAAGGGGGGAC (SEQ ID NO:13), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO:13.

[0116] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the NeoR/KanR sequence present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence: ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCGGCTTGGGTGGAGAGG CTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGT TCCGGCTGTCAGCGCAGGGGCGTCCGGTTCTTTTTGTCAAGACCGACCTGTCCGG TGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCGACGAC

GGGCGTTCCTTGCGCGGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTG
GCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCT
GCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGAT
CCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGT
ACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAG
GGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGTCTATGCCCGACGGC
GAGGATCTCGTCGTGACCCACGGCGATGCCTGCTTGCCGAATATCATGGTGGAA
AATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGTCTGGGTGTGGCGGACCGCT
ATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAAT
GGGCTGACCGCTTCCTTGTGCTTTACGGTATCGCCGCGCCCGATTCGCAGCGCAT
CGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGA (SEQ ID NO:14), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO:14.

[0117] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the rrnG terminator (transcription terminator from the *E.coli* ribosomal RNA rrnG operon (Albrechtsen et al., 1991)) present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence:
GCATTGGCGCAGAAAAAAATGCCTGATGCGACGCTGCGCGTCTTATACTCCCAC
ATATGCCAGATTCAGCAACGGATACGGCTTCCCCAACTTGCCCACTTCCATACGT
GTCCTCCTTACCAGAAATTTATCCTTAA (SEQ ID NO:15), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO:15.

[0118] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the ori (high-copy-number ColE1/pMB1/pBR322/pUC origin of replication) present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence:
TTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCG
CTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGG
TAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTA
GTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTA
ATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG
ACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTT
CGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTAC
AGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGG

TATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG
GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAG
CGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAA (SEQ ID NO:
16), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO:16.

[0119] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the CAP binding site present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence: TAATGTGAGTTAGCTCACTCAT (SEQ ID NO:17), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO:17.

[0120] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the *E.coli lac* promoter present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence: TTTACACTTTATGCTTCCGGCTCGTATGTTG (SEQ ID NO:18), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO:18.

[0121] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the lac operator present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence: TTGTGAGCGGATAACAA (SEQ ID NO: 19), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO: 19.

[0122] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the T3 promoter (promoter for bacteriophage T3 RNA polymerase) present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence: AATTAACCCTCACTAAAGG (SEQ ID NO: 20), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO: 20.

[0123] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the CMV enhancer present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence:
GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAG
TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC

TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGAC
GTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGG
ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG (SEQ ID NO: 21), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO: 21.

[0124] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the T7 promoter (promoter for bacteriophage T7 RNA polymerase) present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence: CCTATAGTGAGTCGTATTA (SEQ ID NO: 22), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO: 22.

[0125] In some embodiments of any of the expression cassettes and gene delivery vectors described herein, the f1 ori (f1 bacteriophage origin of replication) present in any of the expression cassettes or gene delivery vectors described herein comprises or consists of the following sequence:

ACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCC
TTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTT
AGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGT
GATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT
TGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAA
CCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATT
GGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATT (SEQ ID NO: 23), or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to SEQ ID NO: 23.

[0126] In some embodiments, the polynucleotide cassette of the present invention further comprises an RNA export signal. Exemplary RNA export sequences include but are not limited to wPRE. The wPRE significantly increases transgene expression in target cells, by increasing RNA stability in a transgene, promoter and vector-independent manner (Zuffrey et al, 1999). However, it can express a truncated 60-amino acid protein derived from the WHV X gene involved in liver cancer (Kingsman et al, 2005). Therefore, most pre-clinical protocols and clinical trials include a mutated version of the wPRE element (Zanta-Boussif et al, 2009). On the other hand, the use of two SV40-USE elements in SIN-LV vectors has been seen to be more efficient than the wPRE sequence in supressing transcriptional read through (Schambach et al,

2007). More precisely, the wPRE disclosed herein is a chimeric wPRE that carries 589 nucleotides from the modified WPRE performed by Axel Schambach (nucleotides 1-589) (WO 2008136670 A2; [5]) and 88 from a former wPRE (nucleotide 590-677) (Zuffrey et al, 1999). Data disclosed herein shows this chimeric wPRE works better than the former wPRE. The chimeric wPRE sequence comprises the following sequence:

CGAGCATCTTACCGCCATTTATTCCCATATTTGTTCTGTTTTTCTTGATTTGGGTAT
ACATTTAAATGTTAATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAGGGA
TATGTAATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAACTTT
CCCGTTATTTACGCTCTGTTCCTGTTAATCAACCTCTGGATTACAAAATTTGTGAA
AGATTGACTGATATTCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATATGCTGC
TTTAATGCCTCTGTATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTCCTT
GTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTCAAC
GTGGCGTGGTGTGCTCTGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCATTGC
CACCACCTGTCAACTCCTTTCTGGGACTTTCGCTTTCCCCCTCCCGATCGCCACGG
CAGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTAGGTTGCTGGG
CACTGATAATTCCGTGGTGTTGTCGGGGAAGGGCCTGCTGCCGGCTCTGCGGCCT
CTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTC
CCCGCCTG (SEQ ID NO:24).

[0127] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the wPRE sequence comprises or consists of the sequence of SEQ ID NO:24, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the sequence of SEQ ID NO:24.

CGAGCATCTTACCGCCATTTATTCCCATATTTGTTCTGTTTTTCTTGATTTGGGTAT
ACATTTAAATGTTAATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAGGGA
TATGTAATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAACTTT
CCCGTTATTTACGCTCTGTTCCTGTTAATCAACCTCTGGATTACAAAATTTGTGAA
AGATTGACTGATATTCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATATGCTGC
TTTAATGCCTCTGTATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTCCTT
GTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTCAAC
GTGGCGTGGTGTGCTCTGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCATTGC
CACCACCTGTCAACTCCTTTCTGGGACTTTCGCTTTCCCCCTCCCGATCGCCACGG
CAGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTAGGTTGCTGGG
CACTGATAATTCCGTGGTGTTGTCGGGGAAGGGCCTGCTGCCGGCTCTGCGGCCT

CTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTC CCCGCCTG (SEQ ID NO:24)

[0128] Other combinations of elements both as disclosed herein or as known in the art will be readily appreciated by the ordinarily skilled artisan.

[0129] Additionally, as will be recognized by one of ordinary skill in the art, the polynucleotide cassettes may optionally contain other elements including, but not limited to restriction sites to facilitate cloning and regulatory elements for a particular gene expression vector.

[0130] In some aspects of the present invention, the subject polynucleotide cassettes are used to deliver a gene to cells of an animal, *e.g.* to determine the effect that the gene has on cell viability and/or function, to treat a cell disorder, etc. Accordingly, in some aspects of the invention, the composition that provides for the expression of a transgene in mammalian cells is a gene delivery vector, wherein the gene delivery vector comprises the polynucleotide cassettes of the present disclosure.

[0131] Any convenient gene therapy vector that finds use delivering polynucleotide sequences to mammalian cells is encompassed by the gene delivery vectors of the present disclosure. For example, the vector may comprise single or double stranded nucleic acid, *e.g.* single stranded or double stranded DNA. For example, the gene delivery vector may be DNA, *e.g.*, a naked DNA, *e.g.*, a plasmid, a minicircle, etc. The vector may comprise single-stranded or double-stranded RNA, including modified forms of RNA. In another example, the gene delivery vector may be an RNA, *e.g.*, an mRNA or modified mRNA.

[0132] As another example, the gene delivery vector may be a viral vector derived from a virus, *e.g.*, an adenovirus, an adeno-associated virus, a LV, a herpes virus, an alphavirus or a retrovirus, *e.g.*, Moloney murine leukemia virus (M-MuLV), Moloney murine sarcoma virus (MoMSV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), gibbon ape leukemia virus (GaLV), feline leukemia virus (FLV), spumavirus, Friend murine leukemia virus, Murine Stem Cell Virus (MSCV) or Rous Sarcoma Virus (RSV). While embodiments encompassing the use of LV are described in greater detail below, it is expected that the ordinarily skilled artisan will appreciate that similar knowledge and skill in the art can be brought to bear on non-LV gene therapy vectors as well.

[0133] In some embodiments, the gene delivery vector is a self-limiting LV. In some embodiments, the gene delivery vector is a self-inactivating LV. In some embodiments, the gene delivery vector is a non-integrating LV. In a specific embodiment of any of the expression cassettes and gene delivery vectors described herein, the self-limiting LV (Figures 22 and 23)

of the disclosure comprises or consists of the following sequence, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:

GTGTTTAAACCTAGATATTGATAGTCTGATCGGTCAACGTATAATCGAGTCCTAG
CTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGGCTTTCGCATGATTGA
ACAAGATGGATTGCACGCAGGTTCTCCGGCGGCTTGGGTGGAGAGGCTATTCGG
CTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTG
TCAGCGCAGGGGCGTCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGA
ATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCGACGACGGGCGTTC
CTTGCGCGGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATT
GGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAA
GTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCT
GCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGG
AAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGC
CAGCCGAACTGTTCGCCAGGCTCAAGGCGTCTATGCCCGACGGCGAGGATCTCG
TCGTGACCCACGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTT
TTCTGGATTCATCGACTGTGGCCGTCTGGGTGTGGCGGACCGCTATCAGGACATA
GCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCT
TCCTTGTGCTTTACGGTATCGCCGCGCCCGATTCGCAGCGCATCGCCTTCTATCGC
CTTCTTGACGAGTTCTTCTGACCGATTCTAGGTGCATTGGCGCAGAAAAAAATGC
CTGATGCGACGCTGCGCGTCTTATACTCCCACATATGCCAGATTCAGCAACGGAT
ACGGCTTCCCCAACTTGCCCACTTCCATACGTGTCCTCCTTACCAGAAATTTATCC
TTAACGATCGGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCG
CTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTC
ATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGA
AGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCAC
TGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTC
TGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTG
TTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGA
GCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCA
AGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGC
TGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTA
CCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGC
TTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAA

AGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAG
GGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCT
TTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCT
CGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGT
TCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT
CTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCG
AACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATAC
GCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAG
GTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCT
CACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTG
GAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACG
CCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTGCAAGCTT
GGCCATTGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGT
CCAACATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAA
TTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTAC
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAAT
AATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGG
GTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC
CAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC
CCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGC
GGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTT
GTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCA
TTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT
CGTTTAGTGAACCGGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTC
TGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTT
CAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGAC
CCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAA
AGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGC
GCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGAC
TAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGG
GGAGAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAA
AAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCA
GTTAATCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAG

CTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAG
TAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAG
CTTTAGACAAGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAA
GCGGCCGCTGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAG
TGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCAC
CAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTGGGAATAGGA
GCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAA
TGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGA
ACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTG
GGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGA
TCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCT
GTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAACAGATTTGGAATCACA
CGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACT
CCTTAATTGAAGAATCGCAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGG
AATTAGATAAATGGGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTG
GTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTT
TTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTT
TCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGA
AGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCTCG
ACGGTATCGGTTAACTTTTAAAAGAAAAGGGGGGATTGGGGGGTACAGTGCAGG
GGAAAGAATAGTAGACATAATAGCAACAGACATACAAACTAAAGAATTACAAA
AACAAATTACAAAAATTCAAAATTTTATCGATCACGAGACTAGCCTCGAGAAGC
TTGATATCGAATTCCACGGGGTTGGGGTTGCGCCTTTTCCAAGGCAGCCCTGGGT
TTGCGCAGGGACGCGGCTGCTCTGGGCGTGGTTCCGGGAAACGCAGCGGCGCCG
ACCCTGGGTCTCGCACATTCTTCACGTCCGTTCGCAGCGTCACCCGGATCTTCGC
CGCTACCCTTGTGGGCCCCCCGGCGACGCTTCCTGCTCCGCCCCTAAGTCGGGAA
GGTTCCTTGCGGTTCGCGGCGTGCCGGACGTGACAAACGGAAGCCGCACGTCTC
ACTAGTACCCTCGCAGACGGACAGCGCCAGGGAGCAATGGCAGCGCGCCGACCG
CGATGGGCTGTGGCCAATAGCGGCTGCTCAGCAGGGCGCGCCGAGAGCAGCGGC
CGGGAAGGGGCGGTGCGGGAGGCGGGGTGTGGGGCGGTAGTGTGGGCCCTGTTC
CTGCCCGCGCGGTGTTCCGCATTCTGCAAGCCTCCGGAGCGCACGTCGGCAGTCG
GCTCCCTCGTTGACCGAATCACCGACCTCTCTCCCCAGGGGGATCCGTCGACACC
GGTGCCACCATGAGCATCCAGGAAAATATCAGCTCTCTGCAGCTGCGGTCCTGG
GTGTCCAAGAGCCAGAGAGACCTGGCCAAGAGCATCCTGATCGGAGCCCCTGGC

GGACCAGCCGGATACCTGAGAAGGGCTAGCGTGGCCCAGCTGACCCAGGAACTG
GGCACCGCCTTTTTCCAGCAGCAGCAGCTGCCAGCCGCCATGGCCGACACCTTTC
TGGAACACCTGTGCCTGCTGGACATCGACTCTGAGCCCGTGGCCGCCAGAAGCA
CCAGCATCATTGCCACCATCGGCCCTGCCAGCAGAAGCGTGGAGCGGCTGAAAG
AGATGATCAAGGCCGGCATGAATATCGCCCGGCTGAACTTCTCCCACGGCAGCC
ACGAGTACCACGCAGAGAGCATTGCCAACGTCCGGGAGGCCGTGGAGAGCTTTG
CCGGCAGCCCCCTGAGCTACAGACCCGTGGCCATTGCCCTGGACACCAAGGGCC
CCGAGATCAGAACAGGAATTCTGCAGGGAGGGCCTGAGAGCGAGGTGGAGCTG
GTGAAGGGCAGCCAAGTGCTGGTGACCGTGGACCCCGCCTTCAGAACCAGAGGC
AACGCCAACACAGTGTGGGTGGACTACCCCAACATCGTGCGGGTGGTGCCTGTG
GGCGGCAGAATCTACATCGACGACGGCCTGATCAGCCTGGTGGTGCAGAAGATC
GGACCTGAGGGCCTGGTGACCCAGGTCGAGAATGGCGGCGTGCTGGGCAGCAGA
AAGGGCGTGAATCTGCCAGGCGCCCAGGTGGACCTGCCTGGCCTGTCTGAGCAG
GACGTGAGAGACCTGAGATTTGGCGTGGAGCACGGCGTGGACATCGTGTTCGCC
AGCTTCGTGCGGAAGGCCTCTGATGTGGCCGCCGTGAGAGCCGCTCTGGGCCCTG
AAGGCCACGGCATCAAGATCATCAGCAAGATCGAGAACCACGAGGGCGTGAAG
CGGTTCGACGAGATCCTGGAAGTGTCCGACGGCATCATGGTGGCCAGAGGCGAC
CTGGGCATCGAGATCCCCGCCGAGAAGGTGTTCCTGGCCCAGAAAATGATGATC
GGACGGTGCAACCTGGCCGGCAAACCTGTGGTGTGCGCCACCCAGATGCTGGAA
AGCATGATCACCAAGCCCAGACCCACCAGAGCCGAGACAAGCGACGTGGCCAAC
GCCGTGCTGGATGGCGCTGACTGCATCATGCTGTCCGGCGAGACAGCCAAGGGC
AACTTCCCCGTGGAGGCCGTGAAGATGCAGCACGCCATTGCCAGAGAAGCCGAG
GCCGCCGTGTACCACCGGCAGCTGTTCGAGGAACTGCGGAGAGCCGCCCCTCTG
AGCAGAGATCCCACCGAAGTGACCGCCATCGGAGCCGTGGAAGCCGCCTTCAAG
TGCTGCGCCGCTGCAATCATCGTGCTGACCACCACAGGCAGAAGCGCCCAGCTG
CTGTCCAGATACAGACCCAGAGCCGCCGTGATCGCCGTGACAAGATCCGCCCAG
GCCGCTAGACAGGTCCACCTGTGCAGAGGCGTGTTCCCCCTGCTGTACCGGGAGC
CTCCCGAGGCCATCTGGGCCGACGACGTGGACAGACGGGTGCAGTTCGGCATCG
AGAGCGGCAAGCTGCGGGGCTTCCTGAGAGTGGGCGACCTGGTGATCGTGGTGA
CAGGCTGGCGGCCTGGCAGCGGCTACACCAACATCATGAGGGTGCTGTCCATCA
GCTGACCGCGGTCTAGAGGATCCCCCGGGCTGCAGGAATTCGAGCATCTTACCG
CCATTTATTCCCATATTTGTTCTGTTTTTCTTGATTTGGGTATACATTTAAATGTTA
ATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAGGGATATGTAATTACTAG
TTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAACTTTCCCGTTATTTACGC

TCTGTTCCTGTTAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGATAT
TCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATATGCTGCTTTAATGCCTCTGT
ATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTCCTTGTATAAATCCTGG
TTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTCAACGTGGCGTGGTGTG
CTCTGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCATTGCCACCACCTGTCAA
CTCCTTTCTGGGACTTTCGCTTTCCCCCTCCCGATCGCCACGGCAGAACTCATCGC
CGCCTGCCTTGCCCGCTGCTGGACAGGGGCTAGGTTGCTGGGCACTGATAATTCC
GTGGTGTTGTCGGGGAAGGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTC
GCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCCTGGAAT
TCGAGCTCGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCC
ACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGAC
AAGATCTGCTTTTTGCTTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGG
GAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT
GAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATC
CCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTAGTAGTTCATGTCAT
CTTATTATTCAGTATTTATAACTTGCAAAGAAATGAATATCAGAGAGTGAGAGGA
ACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTT
CACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA
ATGTATCTTATCATGTCTGGCTCTAGCTATCCCGCCCCTAACTCCGCCCATCCCGC
CCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTT
ATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGA
GGAGGCTTTTTTGGAGGCCTAGGGACGTACCCAATTCGCCCTATAGTGAGTCGTA
TTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGC
GTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATA
GCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCG
AATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGC
GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTC
CCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCT
CCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGAT
TAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTT
TGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAAC
ACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGG
CCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAA
AATCGTTCCCTCAGGACGTC (SEQ ID NO:25).

[0134] In a specific embodiment of the expression cassettes described herein, the expression comprises or consists of the following sequence, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to the following sequence:
GGGGTTGGGGTTGCGCCTTTTCCAAGGCAGCCCTGGGTTTGCGCAGGGACGCGG
CTGCTCTGGGCGTGGTTCCGGGAAACGCAGCGGCGCCGACCCTGGGTCTCGCAC
ATTCTTCACGTCCGTTCGCAGCGTCACCCGGATCTTCGCCGCTACCCTTGTGGGCC
CCCCGGCGACGCTTCCTGCTCCGCCCCTAAGTCGGGAAGGTTCCTTGCGGTTCGC
GGCGTGCCGGACGTGACAAACGGAAGCCGCACGTCTCACTAGTACCCTCGCAGA
CGGACAGCGCCAGGGAGCAATGGCAGCGCGCCGACCGCGATGGGCTGTGGCCA
ATAGCGGCTGCTCAGCAGGGCGCGCCGAGAGCAGCGGCCGGGAAGGGGCGGTG
CGGGAGGCGGGGTGTGGGGCGGTAGTGTGGGCCCTGTTCCTGCCCGCGCGGTGT
TCCGCATTCTGCAAGCCTCCGGAGCGCACGTCGGCAGTCGGCTCCCTCGTTGACC
GAATCACCGACCTCTCTCCCCAGGGGGATCCGTCGACACCGGTGCCACCATGAG
CATCCAGGAAAATATCAGCTCTCTGCAGCTGCGGTCCTGGGTGTCCAAGAGCCA
GAGAGACCTGGCCAAGAGCATCCTGATCGGAGCCCCTGGCGGACCAGCCGGATA
CCTGAGAAGGGCTAGCGTGGCCCAGCTGACCCAGGAACTGGGCACCGCCTTTTT
CCAGCAGCAGCAGCTGCCAGCCGCCATGGCCGACACCTTTCTGGAACACCTGTG
CCTGCTGGACATCGACTCTGAGCCCGTGGCCGCCAGAAGCACCAGCATCATTGCC
ACCATCGGCCCTGCCAGCAGAAGCGTGGAGCGGCTGAAAGAGATGATCAAGGCC
GGCATGAATATCGCCCGGCTGAACTTCTCCCACGGCAGCCACGAGTACCACGCA
GAGAGCATTGCCAACGTCCGGGAGGCCGTGGAGAGCTTTGCCGGCAGCCCCCTG
AGCTACAGACCCGTGGCCATTGCCCTGGACACCAAGGGCCCCGAGATCAGAACA
GGAATTCTGCAGGGAGGGCCTGAGAGCGAGGTGGAGCTGGTGAAGGGCAGCCA
AGTGCTGGTGACCGTGGACCCCGCCTTCAGAACCAGAGGCAACGCCAACACAGT
GTGGGTGGACTACCCCAACATCGTGCGGGTGGTGCCTGTGGGCGGCAGAATCTA
CATCGACGACGGCCTGATCAGCCTGGTGGTGCAGAAGATCGGACCTGAGGGCCT
GGTGACCCAGGTCGAGAATGGCGGCGTGCTGGGCAGCAGAAAGGGCGTGAATCT
GCCAGGCGCCCAGGTGGACCTGCCTGGCCTGTCTGAGCAGGACGTGAGAGACCT
GAGATTTGGCGTGGAGCACGGCGTGGACATCGTGTTCGCCAGCTTCGTGCGGAA
GGCCTCTGATGTGGCCGCCGTGAGAGCCGCTCTGGGCCCTGAAGGCCACGGCAT
CAAGATCATCAGCAAGATCGAGAACCACGAGGGCGTGAAGCGGTTCGACGAGAT
CCTGGAAGTGTCCGACGGCATCATGGTGGCCAGAGGCGACCTGGGCATCGAGAT
CCCCGCCGAGAAGGTGTTCCTGGCCCAGAAAATGATGATCGGACGGTGCAACCT
GGCCGGCAAACCTGTGGTGTGCGCCACCCAGATGCTGGAAAGCATGATCACCAA

GCCCAGACCCACCAGAGCCGAGACAAGCGACGTGGCCAACGCCGTGCTGGATGG
CGCTGACTGCATCATGCTGTCCGGCGAGACAGCCAAGGGCAACTTCCCCGTGGA
GGCCGTGAAGATGCAGCACGCCATTGCCAGAGAAGCCGAGGCCGCCGTGTACCA
CCGGCAGCTGTTCGAGGAACTGCGGAGAGCCGCCCCTCTGAGCAGAGATCCCAC
CGAAGTGACCGCCATCGGAGCCGTGGAAGCCGCCTTCAAGTGCTGCGCCGCTGC
AATCATCGTGCTGACCACCACAGGCAGAAGCGCCCAGCTGCTGTCCAGATACAG
ACCCAGAGCCGCCGTGATCGCCGTGACAAGATCCGCCCAGGCCGCTAGACAGGT
CCACCTGTGCAGAGGCGTGTTCCCCCTGCTGTACCGGGAGCCTCCCGAGGCCATC
TGGGCCGACGACGTGGACAGACGGGTGCAGTTCGGCATCGAGAGCGGCAAGCTG
CGGGGCTTCCTGAGAGTGGGCGACCTGGTGATCGTGGTGACAGGCTGGCGGCCT
GGCAGCGGCTACACCAACATCATGAGGGTGCTGTCCATCAGCTGACCGCGGTCT
AGAGGATCCCCCGGGCTGCAGGAATTCGAGCATCTTACCGCCATTTATTCCCATA
TTTGTTCTGTTTTTCTTGATTTGGGTATACATTTAAATGTTAATAAAACAAAATGG
TGGGGCAATCATTTACATTTTTAGGGATATGTAATTACTAGTTCAGGTGTATTGCC
ACAAGACAAACATGTTAAGAAACTTTCCCGTTATTTACGCTCTGTTCCTGTTAAT
CAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGATATTCTTAACTATGTTG
CTCCTTTTACGCTGTGTGGATATGCTGCTTTAATGCCTCTGTATCATGCTATTGCT
TCCCGTACGGCTTTCGTTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTAT
GAGGAGTTGTGGCCCGTTGTCCGTCAACGTGGCGTGGTGTGCTCTGTGTTTGCTG
ACGCAACCCCCACTGGCTGGGGCATTGCCACCACCTGTCAACTCCTTTCTGGGAC
TTTCGCTTTCCCCCTCCCGATCGCCACGGCAGAACTCATCGCCGCCTGCCTTGCCC
GCTGCTGGACAGGGGCTAGGTTGCTGGGCACTGATAATTCCGTGGTGTTGTCGGG
GAAGGGCC (SEQ ID NO: 26).

[0135] In such embodiments, the subject polynucleotide cassette is flanked on the 5' and 3' ends by functional LTR sequences. In one embodiment, the position of different elements present in the backbone of the LV vector is depicted in **Figure 1**. Both LTR sequences have been modified to generate SIN LV vectors. SIN vectors have a 400 bp deletion in the 3'-LTR, covering the promoter/enhancer elements from the U3 region. Expression of the transgene is thereby dependent on internal promoters, reducing the risk of RCLs and decreasing promoter interference (Ginn et al, 2003). This 3'-LTR deletion removes the TATA box, preventing transcription initiation (Miyoshi et al. 1998; Zuffrey et al 1998); and therefore inactivating the vector. The U3 region of the 5'-LTR has been replaced by other heterologous promoting sequences (*i.e.* CMV or RSV) to achieve a Tat-independent transcription and to increase genomic RNA synthesis, resulting in the increase of the viral titer. Because 5'-U3 region drives

the expression of primary transcripts, its modifications will not be present in transduced cells (Schambach et al. 2009).

[0136] Exogenous elements, such as β-globin or SV40 polyadenylation signals (Iwakuma et al, 1999) or the USE from SV40-USE (Schambach et al. 2007), have also been included in the R region of the viral 3'LTR in order to decrease the transcriptional readthrough from the internal promoters (Zaiss et al, 2002) or from remnants of the deleted U3 region of SIN-LV vectors (Almarza et al. 2011) preventing the potential transcriptional activation of the downstream genes.

[0137] In certain embodiments, an expression cassette or gene delivery vector, *e.g.*, a LV, comprises a polynucleotide sequence comprising the following sequences in 5' to 3' order:

a) a PGK promoter sequence, optionally a human PGK promoter sequence;

b) a sequence encoding a pyruvate kinase polypeptide, optionally a coRPK coding or cDNA sequence; and

c) a mutant wPRE sequence, optionally comprising or consisting of the sequence of SEQ ID NO:24.

[0138] In certain embodiments, an expression cassette or gene delivery vector, *e.g.*, a LV, comprises a polynucleotide sequence comprising the following sequences in 5' to 3' order:

a) a cPPT sequence;

b) PGK promoter sequence, optionally a human PGK promoter sequence;

c) a sequence encoding a pyruvate kinase polypeptide, optionally a coRPK coding or cDNA sequence; and

d) a mutant wPRE sequence, optionally comprising or consisting of the sequence of SEQ ID NO:24.

[0139] In certain embodiments, an expression cassette or gene delivery vector, *e.g.*, a LV, comprises a polynucleotide sequence comprising the following sequences in 5' to 3' order:

a) a 5' LTR, optionally a modified 5' LTR;

b) a cPPT sequence;

c) PGK promoter sequence, optionally a human PGK promoter sequence;

d) a sequence encoding a pyruvate kinase polypeptide, optionally a coRPK coding or cDNA sequence;

e) a mutant wPRE sequence, optionally comprising or consisting of the sequence of SEQ ID NO:24; and

f) a 3' LTR, optionally a modified 3' LTR.

[0140] In certain embodiments, an expression cassette or gene delivery vector, *e.g.*, a LV, comprises a polynucleotide sequence comprising the following sequences in 5' to 3' order:

a) a 5' LTR, optionally a modified 5' LTR;

b) a cPPT sequence;

c) PGK promoter sequence, optionally a human PGK promoter sequence;

d) a sequence encoding a pyruvate kinase polypeptide, optionally a coRPK coding or cDNA sequence;

e) a mutant wPRE sequence, optionally comprising or consisting of the sequence of SEQ ID NO:24;

f) a 3' LTR, optionally a modified 3' LTR;

g) an SV40 poly(A) signal; and

h) an SV40 origin of replication (ori).

[0141] In certain embodiments, an expression cassette or gene delivery vector, *e.g.*, a LV, comprises a polynucleotide sequence comprising the following sequences in 5' to 3' order:

a) a 5' LTR, optionally a modified 5' LTR;

b) a cPPT sequence;

c) PGK promoter sequence, optionally a human PGK promoter sequence;

d) a sequence encoding a pyruvate kinase polypeptide, optionally a coRPK coding or cDNA sequence;

e) a mutant wPRE sequence, optionally comprising or consisting of the sequence of SEQ ID NO:24;

f) a 3' LTR, optionally a modified 3' LTR;

g) an SV40 poly(A) signal;

h) an SV40 origin of replication (ori);

i) a T7 promoter;

j) an f1 ori; and

k) a NeoR/KanR sequence.

[0142] In certain embodiments, an expression cassette or gene delivery vector, *e.g.*, a LV, comprises a polynucleotide sequence comprising the following sequences in 5' to 3' order:

a) a 5' LTR, optionally a modified 5' LTR;

b) an HIV-1 ψ sequence;

c) an RRE;

d) a cPPT/CTS sequence;

e) a PGK promoter sequence, optionally a human PGK promoter sequence;

f) a sequence encoding a pyruvate kinase polypeptide, optionally a coRPK coding or cDNA sequence;
g) a mutant wPRE sequence, optionally comprising or consisting of the sequence of SEQ ID NO:24;
h) a dNEF signal;
i) a 3' LTR, optionally a modified 3' LTR;
j) an SV40 poly(A) signal;
k) an SV40 origin of replication (ori);
l) a T7 promoter;
m) an f1 ori;
n) a NeoR/KanR sequence;
o) a rrnG terminator;
p) an ori sequence;
q) a CAP binding site;
r) a lac promoter sequence;
s) a lac operator sequence;
t) a T3 promoter sequence;
u) a CMV enhancer sequence; and
v) a CMV promoter sequence.

[0143] In certain embodiments, the gene delivery vector is PGK-coRPK LV, or comprises the elements depicted in FIG. 21.

[0144] In certain embodiments, an expression cassette or gene delivery vector, *e.g.*, a LV, comprises a polynucleotide sequence comprising the following sequences in 5' to 3' order:
a) a 5' LTR, optionally a modified 5' LTR;
b) an HIV-1 ψ sequence;
c) an RRE;
d) a cPPT/CTS sequence;
e) a PGK promoter sequence, optionally a human PGK promoter sequence;
f) a sequence encoding a pyruvate kinase polypeptide, optionally a coRPK coding or cDNA sequence;
g) a mutant wPRE sequence, optionally comprising or consisting of the sequence of SEQ ID NO: 24;
h) optionally, a dNEF signal;
i) a 3' LTR, optionally a modified 3' LTR;

j) an SV40 poly(A) signal;

k) an SV40 origin of replication (ori);

l) a T7 promoter;

m) pUC ori;

u) a CMV enhancer sequence; and

v) a CMV promoter sequence.

[0145] In particular embodiments of any of the expression cassettes and gene delivery vectors described herein, the coRPK cDNA or coding sequence encodes a PKLR polypeptide that comprises or consists of the sequence disclosed in any of GenBank accession Nos. XP_016856982.1, XP_011507942.1, XP_006711449.1, NP_870986.1, or NP_000289.1, or a functional fragment of any of these sequences, or a sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to any of these sequences.

[0146] Gene therapy vectors encapsulating the polynucleotide cassettes of the present disclosure may be produced using standard methodology. For example, in the case of LV virions, an LV expression vector according to the invention may be introduced into a producer cell, followed by introduction of an LV helper construct, where the helper construct includes LV coding regions capable of being expressed in the producer cell and which complement LV helper functions absent in the LV vector. This is followed by introduction of helper virus and/or additional vectors into the producer cell, wherein the helper virus and/or additional vectors provide accessory functions capable of supporting efficient LV virus production. The producer cells are then cultured to produce LV. These steps are carried out using standard methodology.

[0147] Any suitable method for producing viral particles for delivery of the subject polynucleotide cassettes can be used, including but not limited to those described in the examples that follow. Any concentration of viral particles suitable to effectively transducer mammalian cells can be prepared for contacting mammalian cells in vitro or in vivo. For example, the viral particles may be formulated at a concentration of 10^8 vector genomes per ml or more, for example, $5x10^8$ vector genomes per mL; 10^9 vector genomes per mL; 5 x 10^9 vector genomes per mL, 10^{10} vector genomes per mL, $5x10^{10}$ vector genomes per mL; 10^{11} vector genomes per mL; 5 $x10^{11}$ vector genomes per mL; 10^{12} vector genomes per mL; $5x10^{12}$ vector genomes per mL; 10^{13} vector genomes per mL; 1.5 $x10^{13}$ vector genomes per mL; $3x10^{13}$ vector genomes per mL; $5x10^{13}$ vector genomes per mL; $7.5x10^{13}$ vector genomes per mL; $9x10^{13}$ vector genomes per mL; 1 x 10^{14} vector genomes per mL, 5 x 10^{14} vector genomes per mL or more, but typically not more than 1 x 10^{15} vector genomes per mL.

[0148] In preparing the subject LV compositions, any host cells for producing LV virions may be employed, including, for example, mammalian cells (*e.g.* 293 cells), insect cells (*e.g.* SF9 cells), microorganisms and yeast. Host cells can also be packaging cells in which the LV rep and cap genes are stably maintained in the host cell or producer cells in which the LV vector genome is stably maintained and packaged. Exemplary packaging and producer cells are derived from SF-9, 293, A549 or HeLa cells. LV vectors are purified and formulated using standard techniques known in the art.

[0149] In certain embodiments, the present invention includes a cell comprising an expression cassette or gene delivery vector disclosed herein. In related embodiments, the cell is transduced with a viral vector comprising an expression cassette disclosed herein or has an expression cassette disclosed herein integrated into the cell's genome. In certain embodiments, the cell is a cell used to produce a viral gene delivery vector. In other embodiments, the cell is a cell to be delivered to a subject in order to provide to the subject the gene product encoded by the expression cassette. Thus, in certain embodiments, the cell is autologous to the subject to be treated or was obtained from the subject to be treated. In other embodiments, the cell is allogeneic to the subject to be treated or was obtained from a donor other than the subject to be treated. In particular embodiments, the cell is a mammalian cell, *e.g.*, a human cell. In certain embodiments, the cell is a blood cell, an erythrocyte, a hematopoietic progenitor cell, a bone marrow cell, *e.g.*, a lineage depleted bone marrow cell, a hematopoietic stem cell (*e.g.*, CD34+) or a committed hematopoietic erythroid progenitor cell.

[0150] The present invention includes pharmaceutical compositions comprising a polynucleotide cassette, gene delivery vector, or cell described herein and a pharmaceutically-acceptable carrier, diluent or excipient. The subject polynucleotide cassette, gene delivery vector, or cell can be combined with pharmaceutically-acceptable carriers, diluents and reagents useful in preparing a formulation that is generally safe, non-toxic, and desirable, and includes excipients that are acceptable for primate use. Such excipients can be solid, liquid, semisolid, or, in the case of an aerosol composition, gaseous. Examples of such excipients, carriers or diluents include, but are not limited to, water, saline, Ringer's solutions, dextrose solution, and 5% human serum albumin. Supplementary active compounds can also be incorporated into the formulations. Solutions or suspensions used for the formulations can include a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial compounds such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating compounds such as ethylenediaminetetraacetic acid (EDTA); buffers such as

acetates, citrates or phosphates; detergents such as Tween 20 to prevent aggregation; and compounds for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. In particular embodiments, the pharmaceutical compositions are sterile.

[0151] Pharmaceutical compositions suitable for use in the present invention further include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion.

[0152] Sterile solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0153] In one embodiment, the compositions are prepared with carriers that will protect the gene cassette or expression vector against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially.

[0154] It is especially advantageous to formulate oral, ocular or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

[0155] The pharmaceutical compositions can be included in a container, pack, or dispenser, *e.g.* syringe, *e.g.* a prefilled syringe, together with instructions for administration.

[0156] The pharmaceutical compositions of the invention encompass any pharmaceutically acceptable salts, esters, or salts of such esters, or any other compound which, upon administration to an animal comprising a human, is capable of providing (directly or indirectly) the biologically active metabolite or residue thereof.

[0157] The term "pharmaceutically acceptable salt" refers to physiologically and pharmaceutically acceptable salts of the compounds of the invention: *i.e.*, salts that retain the desired biological activity of the parent compound and do not impart undesired toxicological effects thereto. A variety of pharmaceutically acceptable salts are known in the art and described, *e.g.*, in in "Remington's Pharmaceutical Sciences", 17th edition, Alfonso R. Gennaro (Ed.), Mark Publishing Company, Easton, PA, USA, 1985 (and more recent editions thereof), in the "Encyclopaedia of Pharmaceutical Technology", 3rd edition, James Swarbrick (Ed.), Informa Healthcare USA (Inc.), NY, USA, 2007, and in J. Pharm. Sci. 66: 2 (1977). Also, for a review on suitable salts, see Handbook of Pharmaceutical Salts: Properties, Selection, and Use by Stahl and Wermuth (Wiley-VCH, 2002).

[0158] Pharmaceutically acceptable base addition salts are formed with metals or amines, such as alkali and alkaline earth metals or organic amines. Metals used as cations comprise sodium, potassium, magnesium, calcium, and the like. Amines comprise N-N'-dibenzylethylenediamine, chloroprocaine, choline, diethanolamine, dicyclohexylamine, ethylenediamine, N-methylglucamine, and procaine (see, for example, Berge et al., "Pharmaceutical Salts," J. Pharma Sci., 1977, 66, 119). The base addition salts of said acidic compounds are prepared by contacting the free acid form with a sufficient amount of the desired base to produce the salt in the conventional manner. The free acid form may be regenerated by contacting the salt form with an acid and isolating the free acid in the conventional manner. The free acid forms differ from their respective salt forms somewhat in certain physical properties such as solubility in polar solvents, but otherwise the salts are equivalent to their respective free acid for purposes of the present invention.

[0159] The subject polynucleotide cassette, gene delivery vector, *e.g.*, recombinant virus (virions), or cell (*e.g.*, transduced with a gene delivery vector disclosed herein) can be incorporated into pharmaceutical compositions for administration to mammalian patients, particularly primates and more particularly humans. The subject polynucleotide cassette, gene delivery vector, *e.g.* virions, or cell can be formulated in nontoxic, inert, pharmaceutically acceptable aqueous carriers, preferably at a pH ranging from 3 to 8, more preferably ranging from 6 to 8. Such sterile compositions will comprise the vector or virion containing the nucleic

acid encoding the therapeutic molecule dissolved in an aqueous buffer having an acceptable pH upon reconstitution.

[0160] In some embodiments, the pharmaceutical composition provided herein comprise a therapeutically effective amount of a cell, vector or virion disclosed herein in admixture with a pharmaceutically acceptable carrier and/or excipient, for example saline, phosphate buffered saline, phosphate and amino acids, polymers, polyols, sugar, buffers, preservatives and other proteins. Exemplary amino acids, polymers and sugars and the like are octylphenoxy polyethoxy ethanol compounds, polyethylene glycol monostearate compounds, polyoxyethylene sorbitan fatty acid esters, sucrose, fructose, dextrose, maltose, glucose, mannitol, dextran, sorbitol, inositol, galactitol, xylitol, lactose, trehalose, bovine or human serum albumin, citrate, acetate, Ringer's and Hank's solutions, cysteine, arginine, carnitine, alanine, glycine, lysine, valine, leucine, polyvinylpyrrolidone, polyethylene and glycol. Preferably, this formulation is stable for at least six months at 4º C.

[0161] In some embodiments, the pharmaceutical composition provided herein comprises a buffer, such as phosphate buffered saline (PBS) or sodium phosphate/sodium sulfate, tris buffer, glycine buffer, sterile water and other buffers known to the ordinarily skilled artisan such as those described by Good et al. (1966) Biochemistry 5:467. The pH of the buffer in which the pharmaceutical composition comprising the tumor suppressor gene contained in the adenoviral vector delivery system, may be in the range of 6.5 to 7.75, preferably 7 to 7.5, and most preferably 7.2 to 7.4.

[0162] In certain embodiments, viral vectors may be formulated into any suitable unit dosage, including, without limitation, $1x10^8$ vector genomes or more, for example, $1x10^9$, $1x10^{10}$, $1x10^{11}$, $1x10^{12}$, or $1x10^{13}$ vector genomes or more, in certain instances, $1x10^{14}$ vector genomes, but usually no more than $4x10^{15}$ vector genomes. In some cases, the unit dosage is at most about $5x10^{15}$ vector genomes, *e.g.* $1x10^{14}$ vector genomes or less, for example $1x10^{13}$, $1x10^{12}$, $1x10^{11}$, $1x10^{10}$, or $1x10^9$ vector genomes or less, in certain instances $1x10^8$vector genomes or less, and typically no less than $1x10^8$ vector genomes. In some cases, the unit dosage is $1x10^{10}$ to $1x10^{11}$ vector genomes. In some cases, the unit dosage is $1x10^{10}$ to $3x10^{12}$ vector genomes. In some cases, the unit dosage is $1x10^9$ to $3x10^{13}$ vector genomes. In some cases, the unit dosage is $1x10^8$ to $3x10^{14}$ vector genomes. In one embodiment, the range is from about $5x10^{10}$ to about $1x10^{11}$ vector genomes. In some embodiments, the range is from about $1x10^9$ to about $1x10^{10}$ vector genomes.

[0163] In some cases, the unit dosage of a pharmaceutical composition may be measured using multiplicity of infection (MOI). By MOI it is meant the ratio, or multiple, of vector or viral

genomes to the cells to which the nucleic acid may be delivered. In some cases, the MOI may be $1x10^6$. In some cases, the MOI may be $1x10^5$ -$1x10^7$. In some cases, the MOI may be $1x10^4$ -$1x10^8$. In some cases, recombinant viruses of the disclosure are at least about $1x10^1$, $1x10^2$, $1x10^3$, $1x10^4$, $1x10^5$, $1x10^6$, $1x10^7$, $1x10^8$, $1x10^9$, $1x10^{10}$, $1x10^{11}$, $1x10^{12}$, $1x10^{13}$, $1x10^{14}$, $1x10^{15}$, $1x10^{16}$, $1x10^{17}$, and $1x10^{18}$ MOI. In some cases, recombinant viruses of this disclosure are $1x10^8$ to $3x10^{14}$ MOI. In some cases, recombinant viruses of the disclosure are at most about $1x10^1$, $1x10^2$, $1x10^3$, $1x10^4$, $1x10^5$, $1x10^6$, $1x10^7$, $1x10^8$, $1x10^9$, $1x10^{10}$, $1x10^{11}$, $1x10^{12}$, $1x10^{13}$, $1x10^{14}$, $1x10^{15}$, $1x10^{16}$, $1x10^{17}$, and $1x10^{18}$ MOI. In some, embodiments the range is from about 20 to about 400 MOI.

[0164] In some aspects, the amount of pharmaceutical composition comprises about 1 x 10^8 to about 1 x 10^{15} recombinant viruses, about 1 x 10^9 to about 1 x 10^{14} recombinant viruses, about 1 x 10^{10} to about 1 x 10^{13} recombinant viruses, or about 1 x 10^{11} to about 3 x 10^{12} recombinant viruses.

METHODS

[0165] As disclosed herein, the subject polynucleotide cassettes and gene delivery vectors, referred to collectively herein as the "subject compositions", find use in expressing a transgene in cells of an animal. For example, the subject compositions may be used in research, *e.g.* to determine the effect that the gene has on cell viability and/or function. As another example, the subject compositions may be used in medicine, *e.g.* to treat or prevent a disease or disorder. Thus, in some aspects of the invention, methods are provided for the expression of a gene in cells, the method comprising contacting cells with a composition of the present disclosure. In some embodiments, contacting occurs in vitro or ex vivo. In some embodiments, contacting occurs in vivo, *i.e.*, the subject composition is administered to a subject.

[0166] For instances in which mammalian cells are to be contacted in vitro or ex vivo with a subject polynucleotide cassette or gene delivery vector comprising a subject polynucleotide cassette, the cells may be from any mammalian species, *e.g.* rodent (*e.g.* mice, rats, gerbils, squirrels), rabbit, feline, canine, goat, ovine, pig, equine, bovine, primate, human. Cells may be from established cell lines or they may be primary cells, where "primary cells", "primary cell lines", and "primary cultures" are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow *in vitro* for a limited number of passages, *i.e.* splittings, of the culture. For example, primary cultures are cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not

enough times go through the crisis stage. Typically, the primary cell lines of the present invention are maintained for fewer than 10 passages in vitro.

[0167] In particular embodiments, the cell to be contacted with a disclosed polynucleotide cassette or gene delivery vector are hematopoietic stem cells (HSCs) from the subject or from a donor, *e.g.*, an allogenic donor. In some embodiments, a biological sample, *e.g.*, peripheral blood, is obtained from the subject following mobilization of hematopoietic stem cells (HSCs). In one embodiment, HSCs and/or progenitor cells are mobilized by treating the subject with G-CSF or an analog thereof. HSCs and progenitor cells (HSPC) in peripheral blood may be mobilized prior to collection of the biological sample. Peripheral blood HSCs and HSPC can be mobilized by any method known in the art. Peripheral blood HSCs and HSPC can be mobilized by treating the subject with any agent(s), described herein or known in the art, that increase the number of HSPC circulating in the peripheral blood of the subject. For example, in particular embodiments, peripheral blood is mobilized by treating the subject with one or more cytokines or growth factors (*e.g.*, G-CSF, kit ligand (KL), IL-I, IL-7, IL-8, IL-11, Flt3 ligand, SCF, thrombopoietin, or GM-CSF (such as sargramostim)). Different types of G-CSF that can be used in the methods for mobilization of peripheral blood include filgrastim and longer acting G-CSF: pegfilgrastim. In particular embodiments, peripheral blood is mobilized by treating the subject with one or more chemokines (e.g., macrophage inflammatory protein-1a (MIP1a/CCL3)), chemokine receptor ligands (e.g., chemokine receptor 2 ligands GROl3 and GR013M), chemokine receptor analogs (e.g., stromal cell derived factor-1a (SDF- 1a) protein analogs such as CTCE-0021, CTCE-0214, or SDF-1a such as Met-SDF-113), or chemokine receptor antagonists (e.g., chemokine (C-X-C motif) receptor 4 (CXCR4) antagonists such as AMD3100). In particular embodiments, peripheral blood is mobilized by treating the subject with one or more anti-integrin signaling agents (e.g., function blocking anti-very late antigen 4 (VLA-4) antibody, or anti-vascular cell adhesion molecule 1 (VCAM-1)). In particular embodiments, peripheral blood is mobilized by treating the subject with one or more cytotoxic drugs such as cyclophosphamide, etoposide or paclitaxel. In particular embodiments, peripheral blood can be mobilized by administering to a subject one or more of the agents listed above for a certain period of time. For example, the subject can be treated with one or more agents (e.g., G-CSF) via injection (e.g., subcutaneous, intravenous or intraperitoneal), once daily or twice daily, for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 days prior to collection of HSPC. In specific embodiments, HSPC are collected within 1, 2, 3, 4, 5, 6,7, 8, 12, 14, 16, 18, 20 or 24 hours after the last dose of an agent used for mobilization of HSPC into peripheral blood. In particular embodiments, HSCs and HSPC are mobilized by

treating the subject with two or more different types of agents described above or known in the art, such as a growth factor (e.g., G-CSF) and a chemokine receptor antagonist (e.g., CXCR4 receptor antagonist such as AMD3100), or a growth factor (e.g., G-CSF or KL) and an anti-integrin agent (e.g., function blocking VLA-4 antibody). In one embodiment, HSCs and/or progenitor cells are mobilized by treating the subject with G-CSF or an analog thereof. In one embodiment, the G-CSF is filgrastim. In one embodiment, HSCs and/or progenitor cells are mobilized by treating the subject with plerixafor. In a certain embodiment, HSCs and/or progenitor cells are mobilized using a combination of filgrastim and plerixafor, by filgrastim alone, or by plerixafor alone. In particular embodiments, different types of mobilizing agents are administered concurrently or sequentially. For additional information regarding methods of mobilization of peripheral blood see, e.g., Craddock et al., 1997, Blood 90(12):4779-4788; Jin et al., 2008, Journal of Translational Medicine 6:39; Pelus, 2008, Curr. Opin. Hematol. 15(4):285-292; Papayannopoulou et al., 1998, Blood 91(7):2231-2239; Tricot et al., 2008, Haematologica 93(11):1739-1742; and Weaver et al., 2001, Bone Marrow Transplantation 27(2):S23-S29).

[0001] In some embodiments, the cells are enriched prior to contacting the cells with the disclosed polynucleotide cassette or gene delivery vector. As used herein, “enriched” or “high enriched” refers to a method of enriching for a cell population intended to result in substantial enrichment of cells for cells expressing a particular biological marker, *e.g.* CD34. For example, CD34 enrichment used clinically results in mean: 61.6% and median: 65.7% yield of $CD34^+$ cells and mean: 88.5% and median: 95.9% relative purity (N=166) (Clin Lab. 2016 Jul 1;62(7):1243-1248 (PMID: 28164638)). “Enrichment” refers to a process with the goal of substantial enrichment of a relatively rare cell type, CD34+, which usually comprises between 0.2-2% of the cell product in a mobilized leukopheresis or bone marrow collection. Enrichment of CD34+ cells from a mobilized leukopheresis or bone marrow collection targets final CD34+ percentages that have increased from 0.2-2% to >80%. To accomplish this, following initial application of a biological sample to a capture matrix, repeated buffer exchanges, termed herein “washes,” are carried out with the goal of removing cells weakly or non-specifically bound to the capture matrix. Generally, cells are removed from the capture matrix and reapplied for every wash cycle. Removal and reapplication can be accomplished manually by pipetting from tubes or automated using a pump and tubing system. For example, using Quad Technologies MagCloudz® coupled with Dynabeads® magnetic cell separation system, cell-magnetic particle complexes are separated in tubes on a magnetic stand and washes are done manually.

Using the Miltenyi Biotec CliniMACS® System, a pre-set automated program applies the cell-magnetic particle complexes to a magnetic column in a tubing set and washes/reapplications are done using a valve pump system. In certain embodiments, selection may be performed on various instruments, including without limitation the Miltenyi Biotec MACSQuant Tyto®, Quad Technologies MagCloudz®, GE Sepax® Cell Separation System, Terumo Elutra® Cell Separation System, COBE Spectra® Cell Separator, SynGen LAB® or WASH® Systems, Fresenius-Kabi Lovo®, Miltenyi Biotec CliniMACS® System or CliniMACS Prodigy® System. Selection may be performed in a laboratory or at point-of-care. Detailed methods for preparation and enrichment of cells and cell populations, including exemplary methods for selection of CD34+ cells are described, *e.g.*, in Int'l Patent Pub. No. WO 2016/118790. Illustrative selection method useful for high-stringency selection are provided by U.S. Patent No. 8,727,132..

[0002] In certain embodiments, peripheral blood is obtained through a syringe or catheter inserted into a subject's or donor's vein. For example, the peripheral blood can be collected using an apheresis machine. Blood flows from the vein through the catheter into an apheresis machine, which separates the white blood cells, including HSPC from the rest of the blood and then returns the remainder of the blood to the subject's body. Apheresis can be performed for several hours over successive days (*e.g.*, 1 to 5 days) until enough HSPC have been collected.

[0003] In certain embodiments, bone marrow is obtained from the posterior iliac crest of the subject by needle aspiration (see, e.g., Koda et al., 1984, J. Clin Invest. 73:1377-1384).

[0004] In certain embodiments, a hematocrit level of the biological sample may be determined. The hematocrit level may be determined by centrifuging the sample within a treatment chamber to separate RBCs of a sample into a layer such that the packed cell volume may be determined. It should be appreciated that the sample may be combined with an anticoagulant in order to assist with determining the hematocrit level and that such an anticoagulant may be added to the treatment chamber prior to or during centrifugation. Alternatively, the hematocrit level may be determined by measuring optical properties of the sample. For example, a spectrometer may be used to analyze the sample. It should be appreciated that any type of known spectroscopic methods of determining hematocrit level may be used such as, for example, Raman spectroscopy and/or light scattering techniques.

[0005] In certain embodiments, the biological sample is depleted of erythrocytes, *e.g.*, before preparing the one or more cell populations enriched for CD34+ cells from the biological sample. In some embodiments, the cells remaining after depletion techniques are washed. In another embodiment, non-specific IgG is added to the washed cells. In some embodiments, the non-specific IgG is flebogamma.

[0168] Embodiments of the present invention comprise mammalian cells (*e.g.*, CD34+ cells) transduced with a viral delivery vector, *e.g.*, a LV vector containing the human PKLR gene. Accordingly, the present invention includes a method of transducing a mammalian cell, *e.g.* a human hematopoietic stem cell or other cell described herein, comprising contacting the cell with a viral delivery vector, *e.g.*, a LV vector, comprising an expression cassette described herein. In certain embodiments, the cell was previously obtained from a subject to be treated, or from another donor. In particular embodiments, the subject was diagnosed with PKD, and the cell is transduced with a LV comprising an expression cassette encoding pyruvate kinase, *e.g.*, a coRPK coding region or cDNA. It is understood that the disclosed methods, *e.g.*, those used to deliver a pyruvate kinase gene product, *e.g.*, using a coPRK cDNA sequence, to a subject may also be used to treat hemolytic anemia, and/or normalize erythroid differentiation, increase the number of functional mature erythrocytes, reduce extramedullar erythropoiesis, reduce splenomegaly and other secondary effects of hemolytic anemia or PKD.

[0169] To promote expression of the transgene, the subject polynucleotide cassette or gene delivery vector comprising a subject polynucleotide cassette will be contacted with the cells for about 30 minutes to 24 hours or more, *e.g.*, 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 12 hours, 16 hours, 18 hours, 20 hours, 24 hours, etc.

[0170] The subject polynucleotide cassette or gene delivery vector comprising a subject polynucleotide cassette may be provided to the subject cells one or more times, *e.g.* one time, twice, three times, or more than three times, and the cells allowed to incubate with the agent(s) for some amount of time following each contacting event *e.g.* 16-24 hours, after which time the media is replaced with fresh media and the cells are cultured further. Contacting the cells may occur in any culture media and under any culture conditions that promote the survival of the cells. The culture may contain growth factors to which the cells are responsive. Growth factors, as defined herein, are molecules capable of promoting survival, growth and/or differentiation of cells, either in culture or in the intact tissue, through specific effects on a transmembrane receptor. Growth factors include polypeptides and non-polypeptide factors.

[0171] Typically, an effective amount of subject polynucleotide cassette or gene delivery vector comprising a subject polynucleotide cassette is provided to produce the expression of the transgene in cells. As discussed elsewhere herein, the effective amount may be readily determined empirically, *e.g.* by detecting the presence or levels of transgene gene product, by detecting an effect on the viability or function of the cells, etc. Typically, an effect amount of subject polynucleotide cassette or gene delivery vector comprising a subject polynucleotide cassette will promote greater expression of the transgene in cells than the same amount of a polynucleotide cassette as known in the art. Typically, expression will be enhanced 2-fold or more relative to the expression from a reference, or control, polynucleotide cassette *e.g.* as known in the art, for example 3-fold, 4-fold, or 5-fold or more, in some instances 10-fold, 20-fold or 50-fold or more, *e.g.* 100-fold.

[0172] For instances in which cells are to be contacted in vivo with a subject polynucleotide cassette or gene delivery vector comprising a subject polynucleotide cassette, the subject may be any mammal, *e.g.* rodent (*e.g.* mice, rats, gerbils), rabbit, feline, canine, goat, ovine, pig, equine, bovine, or primate. In preferred embodiments, the primate is a human. In some embodiment, the cells are CD34+ cells.

[0173] The methods and compositions of the present disclosure find use, *e.g.*, in the treatment of pyruvate kinase deficiency.

[0174] In another embodiment, the present invention includes a method of treating a disease in a subject in need thereof comprising providing to the subject an effective amount of cells transduced with a gene delivery vector, *e.g.*, a viral vector, that expresses a therapeutic gene product in the cells. In particular embodiments, the cells are autologous to the subject. In certain embodiments, the cells are erythroid cells, *e.g.*, hematopoietic stem cells or committed hematopoietic erythroid progenitor cells. In some embodiments, the cell is a bone marrow cell, *e.g.*, a lineage depleted bone marrow cell. In particular embodiments, the method is used to treat PKD, and the viral vector is a LV comprising an expression construct disclosed herein comprising a human PGK promoter operably linked to a coRPK gene cDNA or coding sequence, and a mutated wPRE disclosed herein. In particular embodiments, the cells are provided to the subject parenterally, *e.g.*, via intravenous injection. In some embodiments, the cells are provided to the subject by transfusion.

[0175] In another embodiment, the present invention includes a method of treating PKD in a subject in need thereof, comprising providing to the subject an effective amount of autologous CD34+ stem cells transduced with a LV vector that expresses a coRPK cDNA in the cells, wherein the LV vector comprises a human PGK promoter operably linked to the coRPK cDNA

or coding sequence, and a mutated wPRE sequence disclosed herein. In particular embodiments, the cells are hematopoietic stem cells or committed hematopoietic erythroid progenitor cells, *e.g.*, bone marrow cells. In particular embodiments, the cells are provided to the subject parenterally, *e.g.*, via intravenous injection.

[0176] In another embodiment, the present invention provides a of treating a disease in a subject in need thereof comprising providing to the subject an effective amount of a gene delivery vector, *e.g.*, a viral vector, that expresses a therapeutic gene product in the subject. In particular embodiments, the method is used to treat PKD, and the viral vector is a LV comprising an expression construct disclosed herein comprising a human PGK promoter operably linked to a coRPK gene cDNA or coding sequence, and a mutated wPRE disclosed herein. In particular embodiments, the gene delivery vector are provided to the subject parenterally, *e.g.*, via intravenous injection.

[0177] In particular embodiments, the cells or gene delivery vectors are provided to the subject in pharmaceutical compositions.

[0178] In some embodiments, the subject methods result in a therapeutic benefit, *e.g.* preventing the development of a disorder, halting the progression of a disorder, reversing the progression of a disorder, etc. In some embodiments, the subject method comprises the step of detecting that a therapeutic benefit has been achieved. The ordinarily skilled artisan will appreciate that such measures of therapeutic efficacy will be applicable to the particular disease being modified, and will recognize the appropriate detection methods to use to measure therapeutic efficacy.

[0179] Expression of the transgene using the subject transgene is expected to be robust. Accordingly, in some instances, the expression of the transgene, *e.g.* as detected by measuring levels of gene product, by measuring therapeutic efficacy, etc. may be observed two months or less after administration, *e.g.* 4, 3 or 2 weeks or less after administration, for example, 1 week after administration of the subject composition. Expression of the transgene is also expected to persist over time. Accordingly, in some instances, the expression of the transgene, *e.g.* as detected by measuring levels of gene product, by measuring therapeutic efficacy, etc., may be observed 2 months or more after administration of the subject composition, *e.g.*, 4, 6, 8, or 10 months or more, in some instances 1 year or more, for example 2, 3, 4, or 5 years, in certain instances, more than 5 years.

[0180] In certain embodiments, the method comprises the step of detecting expression of the transgene in the cells or in the subject, wherein expression is enhanced relative to expression from a polynucleotide cassette not comprising the one or more improved elements of the

present disclosure. Typically, expression will be enhanced 2-fold or more relative to the expression from a reference, *i.e.* a control polynucleotide cassette, *e.g.* as known in the art, for example 3-fold, 4-fold, or 5-fold or more, in some instances 10-fold, 20-fold or 50-fold or more, *e.g.* 100-fold, as evidenced by, *e.g.* earlier detection, higher levels of gene product, a stronger functional impact on the cells, etc.

[0181] Typically, if the subject composition is an LV comprising the subject a polynucleotide cassette of the present disclosure, an effective amount to achieve a change in will be about $1x10^8$ vector genomes or more, in some cases $1x10^9$, $1x10^{10}$, $1x10^{11}$, $1x10^{12}$, or $1x10^{13}$ vector genomes or more, in certain instances, $1x10^{14}$ vector genomes or more, and usually no more than $1x10^{15}$ vector genomes. In some cases, the amount of vector genomes that is delivered is at most about $1x10^{15}$ vector genomes, *e.g.* $1x10^{14}$ vector genomes or less, for example $1x10^{13}$, $1x10^{12}$, $1x10^{11}$, $1x10^{10}$, or $1x10^9$ vector genomes or less, in certain instances $1x10^8$ vector genomes, and typically no less than $1x10^8$ vector genomes. In some cases, the amount of vector genomes that is delivered is $1x10^{10}$ to $1x10^{11}$ vector genomes. In some cases, the amount of vector genomes that is delivered is $1x10^{10}$ to $3x10^{12}$ vector genomes. In some cases, the amount of vector genomes that is delivered is $1x10^9$ to $3x10^{13}$ vector genomes. In some cases, the amount of vector genomes that is delivered is $1x10^8$ to $3x10^{14}$ vector genomes.

[0182] In some cases, the amount of pharmaceutical composition to be administered may be measured using multiplicity of infection (MOI). In some cases, MOI may refer to the ratio, or multiple of vector or viral genomes to the cells to which the nucleic may be delivered. In some cases, the MOI may be $1x10^6$. In some cases, the MOI may be $1x10^5$ -$1x10^7$. In some cases, the MOI may be $1x10^4$ -$1x10^8$. In some cases, recombinant viruses of the disclosure are at least about $1x10^1$, $1x10^2$, $1x10^3$, $1x10^4$, $1x10^5$, $1x10^6$, $1x10^7$, $1x10^8$, $1x10^9$, $1x10^{10}$, $1x10^{11}$, $1x10^{12}$, $1x10^{13}$, $1x10^{14}$, $1x10^{15}$, $1x10^{16}$, $1x10^{17}$, and $1x10^{18}$ MOI. In some cases, recombinant viruses of this disclosure are $1x10^8$ to $3x10^{14}$ MOI. In some cases, recombinant viruses of the disclosure are at most about $1x10^1$, $1x10^2$, $1x10^3$, $1x10^4$, $1x10^5$, $1x10^6$, $1x10^7$, $1x10^8$, $1x10^9$, $1x10^{10}$, $1x10^{11}$, $1x10^{12}$, $1x10^{13}$, $1x10^{14}$, $1x10^{15}$, $1x10^{16}$, $1x10^{17}$, and $1x10^{18}$ MOI.

[0183] In some aspects, the amount of pharmaceutical composition comprises about 1 x 10^8 to about 1 x 10^{15} particles of recombinant viruses, about 1 x 10^9 to about 1 x 10^{14} particles of recombinant viruses, about 1 x 10^{10} to about 1 x 10^{13} particles of recombinant viruses, or about 1 x 10^{11} to about 3 x 10^{12} particles of recombinant viruses.

[0184] Any total number of viral particles suitable to provide appropriate transduction of cells to confer the desired effect or treat the disease can be administered to the mammal. In various preferred embodiments, at least 10^8; $5x10^8$; 10^9; 5 x 10^9, 10^{10}, $5x10^{10}$; 10^{11}; 5 $x10^{11}$; 10^{12};

$5x10^{12}$; 10^{13}; 1.5 $x10^{13}$; $3x10^{13}$; $5x10^{13}$; $7.5x10^{13}$; $9x10^{13}$, 1 x 10^{14} viral particles, or 5 x 10^{14} viral particles or more, but typically not more than 1 x 10^{15} viral particles are injected. Any suitable number of administrations of the vector to the mammal or the primate eye can be made. In one embodiment, the methods comprise a single administration; in other embodiments, multiple administrations are made over time as deemed appropriate by an attending clinician. In some embodiments at least 2 x 10^8 VG/ml of 5 x 10^5 cells/ml is required in a single administration (24 hours transduction) to result in high transduction efficiencies. Individual doses are typically not less than an amount required to produce a measurable effect on the subject, and may be determined based on the pharmacokinetics and pharmacology for absorption, distribution, metabolism, and excretion ("ADME") of the subject composition or its by-products, and thus based on the disposition of the composition within the subject. This includes consideration of the route of administration as well as dosage amount. Effective amounts of dose and/or dose regimen can readily be determined empirically from preclinical assays, from safety and escalation and dose range trials, individual clinician-patient relationships, as well as in vitro and in vivo assays such as those described herein and illustrated in the Examples.

[0185] Several aspects of the invention are described herein with reference to example applications for illustration. It should be understood that numerous specific details, relationships, and methods are set forth to provide a full understanding of the invention. One having ordinary skill in the relevant art, however, will readily recognize that the invention can be practiced without one or more of the specific details or with other methods. The present invention is not limited by the illustrated ordering of acts or events, as some acts may occur in different orders and/or concurrently with other acts or events. Furthermore, not all illustrated acts or events are required to implement a methodology in accordance with the present invention.

[0186] It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely", "only" and the like in connection with the recitation of claim elements, or the use of a "negative" limitation.

[0187] 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 the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

[0188] All of the above U.S. patents, U.S. patent application publications, U.S. patent applications, foreign patents, foreign patent applications and non-patent publications referred to in this specification and/or listed in the Application Data Sheet, are incorporated herein by reference, in their entirety, *e.g.*, to disclose and describe the methods and/or materials in connection with which the publications are cited. It is understood that the present disclosure supersedes any disclosure of an incorporated publication to the extent there is a contradiction.

[0189] From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

EXAMPLES

[0190] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.* amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Experimental Methods

[0191] Vectors and LV supernatant production. LVs were generated as described herein. CoRPK sequence was designed using the GeneArt® software to increase the GC content of the sequence and to prevent cryptic splice sites. Vectors were developed using the pCCL.sin.ppt.hPGK-EGFP-wPRE* construct as a backbone, generously provided by Dr. Naldini (HSR-TIGET, San Raffaele Telethon Institute, Milano, Italy). The original pCCL lentivirus transfer vector and its elements are described in Dull et al., J Virol, 1998; 72(11): 8463-8471, which is incorporated herein in its entirety. Vector stocks of VSV-G pseudotyped LVs were prepared by 3-plasmid calcium phosphate-mediated transfection in 293T cells (ATCC: CRL-1573, Rockeville, MD, USA), as previously described [Follenzi A, et al. (2000). Nat Genet 25: 217-222]. Titers of infective LVs were determined in HT1080 cells (ATCC: CCL-121) by qPCR as described elsewhere [Charrier S, et al. (2005). Gene Ther 12: 597-606]. LV stocks of 10^7-10^8 viral particles (vp)/mL titers were routinely obtained.

[0192] Purification and transduction of murine HSCs. BM from 8-14 week-old male PKD mice was harvested from the leg bones and lineage negative cells (Lin^-) were purified using the Lin^- Cell Depletion kit (Miltenyi Biotec, Gladbach, Germany), obtaining 70-90% purity. Lin^- cells were pre-stimulated with 100 ng/mL of recombinant human IL-11 (Peprotech EC Ltd., London, UK) and 100 ng/mL of recombinant murine SCF (R&D Systems Inc., Minneapolis, MN) in IMDM-Glutamax medium supplemented with 20% FBS and 0.5% antibiotics (50 U/mL penicillin and 50 μg/mL streptomycin, (Thermo Fisher Scientific, Waltham, MA) for 24h, and then transduced with EGFP or coRPK carrying LVs in two cycles of transduction at MOIs of 1-10 vp/cell. Each transduction was carried out for 24h in the presence of the aforementioned cytokines on plates previously coated with CH-296 fibronectin fragment (2μg/cm^2; Retronectin, TakaraShuzo, Otsu, Japan) overnight at 4ºC.

[0193] ***In vivo*** **RBC survival.** Transplanted mice carrying the coRPK transgene were injected with three consecutive intravenous injections (12h apart) of Biotin 3-sulfo-N-hydroxysuccinimide ester sodium salt (50 mg/kg) (Sigma Aldrich, Saint Louis, MO). Twelve hours after the last injection, tail vein blood was harvested and labelled with 2 μg/mL of anti-mouse Ter119-PE (BD Bioscience, San Jose, CA) and streptavidin-FITC (50 μg/mL, BD Biosciences, San Jose, CA) for 30 min at 4ºC. Samples were analyzed in an EPICS XL flow cytometer (Beckman Coulter, Brea, CA) every 2-4 days for 40 days after the injection. RBC survival kinetics was measured by the percentage of biotinylated cells within the total RBC population.

[0194] CFC Assay. CFC assay was performed in BM and spleen from control and transplanted mice according to manufacturer's procedure from Methocult medium GF M3434 (Stem Cell Technologies, Vancouver, Canada). BM cells were harvested at different time-points after transplant from all groups of mice, and CFUs (clusters of 30 or more cells) were scored 7 days after seeding in a Nikon Diaphot-TMD microscope.

[0195] Identification of hematopoietic lineages. PBMCs were obtained from the tail vein of transplanted animals and labelled with a panel of antibodies to detect different hematopoietic cells. Myeloid cells were detected with anti-GR-1 and anti-Mac-1 biotinylated antibodies (BD Bioscience, San Jose, CA, 5 μg/mL), while lymphoid cells were detected using anti-CD3-PE antibody for T-cells, and anti-B220-PE and anti-B220-PECy5 antibodies for B-cells (BD Bioscience, San Jose, CA, 10 μg/mL), together with SAV-TRC secondary antibody (Invitrogen, Thermo Fisher Scientific, Waltham, MA). Samples were analysed in a BD LSR Fortessa Cytometer (BD Bioscience, San Jose, CA,, USA) adding DAPI (Boehringer, Ingelheim, Germany, 2 μg/mL) to exclude death cells.

[0196] Structural and histological studies. Spleens were collected, photographed and weighed on precision scales to determine the presence of splenomegaly. Histological studies were performed on spleen and liver sections obtained following conventional histological methods, and stained with hematoxylin (Gill-2 Haematoxylin, Thermo, Pittsburgh, USA) and eosin (Eosin Alcoholic, Thermo Fisher Scientific, Waltham, MA). Iron deposits were also studied in the spleen by Prussian Blue or Perls' staining (Sigma Aldrich, Saint Louis, MO) following manufacturer's instructions. All sections were examined using an Olympus BX40 light microscope and photographed with an Olympus DP21 camera, with a final magnification of 100x or 200x.

[0197] Erythroid differentiation. Flow cytometry analysis of Ter119 and CD71 marker intensities in BM and spleen were used to identify the different erythroid subpopulations as described elsewhere [Socolovsky M, et al. (2001). Blood 98: 3261-3273] using 4 μg/mL of anti-mouse Ter119-PE antibody (BD Bioscience, San Jose, CA,), 10 μg/mL of biotinylated anti-CD71 antibody (BD Bioscience, San Jose, CA,) and streptavidin-tricolor (Invitrogen, Thermo Fisher Scientific, Waltham, MA). Cells were then analyzed in an EPICS XL flow cytometer (Beckman Coulter, Brea, CA) using propidium iodide (IP, 2 μg/mL) to detect live cells.

[0198] Provirus quantification. Detection and quantification of integrated provirus per cell was accomplished using complementary primers to the packaging proviral sequence (Ψ) and the mouse Titin housekeeping gene. Total BM and peripheral blood samples were collected periodically, and genomic DNA from nucleated cells was isolated using the DNeasy Blood & Tissue kit (Qiagen, Venlo, Limburg, The Netherlands). Twenty to 50 ng of genomic DNA (gDNA) were amplified by multiplex qPCR using the 7500 Fast Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA) and primers and probes previously described [Charrier S, et al. (2011). Gene Ther 18: 479-487].

[0199] Chimerism. Presence of donor cells was quantified by qPCR detecting the Y chromosome SRY gene and the mouse β-Actin housekeeping gene. Primers and probes previously described [Navarro S et al (2006). Mol Ther 14: 525-535] were used and genomic DNA from PB of transplanted mice was amplified using the 7500 Fast Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA). Standard curves were prepared using gDNA extracts from samples containing 0% to 100% of BM cells from male/female mouse mixtures and chimerism was calculated as: % of donor engraftment = 100 x $2^{(Ct\beta Act - CtSRY).}$

[0200] LAM-PCR procedure. In order to identify vector integration sites, 3' vector LTR-genome junctions were amplified by LAM-PCR following the method published by Schmidt et al. 2007 [Nat Methods 4: 1051-1057]. The starting linear amplification (100 cycles) was performed using biotinylated LTR specific primers and up to 100 ng of gDNA as template. Linear amplification products were purified using streptavidin magnetic beads and followed by complementary strand synthesis, parallel digestion with 2 different restriction enzymes (Tsp509I and HpyCH4IV) and two ligation reactions using linker cassettes complementary to the ends left by the enzyme's cut. The fragments generated were amplified by two additional exponential PCR steps. LAM-PCR products were separated and quantified by gel electrophoresis on a MultiNA automated system (Shimadzu).

[0201] Setup of LAM-PCR products for Illumina MiSeq sequencing. Following the method published by Parazynski et al [Paruzynski A, et al. (2010). Nat Protoc 5: 1379-1395], 40 ng of the second exponential PCR products generated by Tsp509I and HpyCH4IV enzymes were re-amplified using fusion primers containing specific sequences that allow paired end sequencing on an Illumina MiSeq sequencer. LAM-PCR samples were adapted for 454-pyrusequencing by fusion PCR to add the Roche 454 GS-FLX adaptors: adaptor A plus an 8-nucleotide barcode was added to the LTR end of the LAM-PCR amplicon; adaptor B was added to the linker cassette side. In 5'-3' orientation the final amplicon was composed as follow: adaptor A, barcode, LTR sequence, unknown genome sequence, linker cassette sequence and primer B. Purified fusion primer PCR products were run and quantified on a MultiNA automated electrophoresis system, and pooled together in order to obtain a final equimolar library of 10 nM. The final library was then re-quantified using a KAPA Library Quantification Kit for Illumina Sequencing Platform (Kapa Biosystems, Wilmington, MA) on a Viia7 real-time PCR system (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA), obtaining an estimated concentration of 16.35 nM. Finally, libraries were sequenced using the Illumina MiSeq Reagent Kit.

[0202] Bioinformatics analysis. To extract vector IS from a high-throughput sequencing platform, both Roche 454 and Illumina MiSeq/HiSeq, a pipeline taking in input the row data (typically in FastQ file format) was designed, providing the list of reliable IS and the nearest gene. Superior level analyses for clonal abundance quantification and gene ontology enrichment were performed using Excel, GraphPad Prism (TM) and available online tools.

[0203] NGS data processing and pipeline usage. The step of NGS data processing deals with the management of high-throughput data from Illumina MiSeq sequencing platforms and aims at identify IS in which all valid sequence reads are aligned to the reference genome. Data

processing comprises two main activities: 1. **Data quality inspection and analysis**, in which LV vector sequences and other contaminants are trimmed. 2. I**ntegration site identification**, in which all valid sequence reads are aligned to the genome of reference and valid ISs are retrieved.

[0204] Data quality analysis. In order to identify IS from Illumina MiSeq raw data a bioinformatics pipeline was developed. Standard LAM-PCR products contain a LTR sequence, a flanking human genomic sequence and a linker cassette (LC) sequence. The 459 technology allowed retrieval of LAM-PCR sequences with length ranging from 10bp to 900 bp. Similar results were retrieved from Illumina MiSeq paired-ends reads. These length boundaries are important parameters to consider in the quality analysis process since they affect both, the subsequent alignment procedure and the algorithm of the vector components identification. Sequences too short to be correctly aligned to the reference gene were discarded, as well as those exceeding the maximum size reachable with NSG technology to avoid missing part or all of the LC sequence. Once the pipeline ends for each pool, all integration sites were collected both in files (archived in the TIGET network attached file storage -NAS-) and in the internal database, and maintained in a storage server that keeps track of the modified copies.

[0205] Integration site identification. To identify unique integration sites and extract the excel file with all IS in rows and each sample in columns (IS matrix) with the closest gene annotations, we run the following steps: 1. Creation of the IS matrix using the program called *create_matrix*, enabling the collision detection inter projects. This program will produce a tab-separated file (TSV); 2.Annotation of the IS matrix file using the program *annotate_bed*, that will be called as follows for each pool using the input TSV file: *awk '{print "chr"$1"\t"$2"\t"$2}' TSV_FILE | tail -n +2 > TSV_FILE.bed; annotate_bed –a /opt/genome/mouse/mm9/annotation/mm9.refGene.TIGET.gtf -b TSV_FILE.bed -o TSV_FILE.annotated.bed; 3.*Import both annotation and matrix file into a new Excel worksheet, here on called XLS.

[0206] Collision detection. In order to obtain a reliable dataset of ISs from each transplanted mouse, we filtered data from potential contaminations/collisions and from false positives based on sequence counts. An additional step of data normalization was required to combine integration sites resulting from different experiments.

[0207] The term "collision" is used to identify the presence of identical IS in independent samples. In our experimental setting, the integration of vector in the very same genomic position in different cells is a very low probability event. Thus, the detection of identical ISs in independent samples likely derives from contamination, which may occur at different stages

of wet laboratory procedures (sample purification, DNA extraction, LAM-PCRs and sequencing). Although our working pipeline is designed to minimize the occurrence of inter-samples contacts, the high-throughput analysis of ISs intrinsically carries a certain degree of background contamination. Identification of the extent of contamination between samples is crucial also because the retrieval of the same IS in different samples obtained from the same mouse is used in subsequent steps to make inference on biological properties of the vector-marked hematopoietic cells (*i.e.* multi-lineage potential and sustained clonogenic activity). Thus, we must be able to distinguish the actual occurrence of the same IS in different samples (from the same mouse) from a contamination/collision.

[0208] To address these issues, we assessed the extent of shared IS among samples derived from different test items and mice as a way to measure the extent of collision in our analyses and then design rules to discard from each mouse's data set those IS that can be ascribed to collision and minimize the likelihood of scoring false positive when looking for shared IS between samples from the same mouse. We designed a collision detection process allowing the validation of each integration locus. The overall result should be that, given the set I of integration loci, in case of classification of an integration locus i in I as collision, i is discarded from I. We applied collision detection process between 3 independent transplantation groups: 1. **coPKR170s**: mice from assay 1 euthanized at 170 days after transplant with Lin^- cells transduced with the coRPK expressing LV vector (coRPK 1-3). 2. **EGFP**: mice from assay 2 transplanted with Lin^- cells carrying the EGFP expressing LV vector (EGFP 1-6). 3.**coPKR-TC**: mice from assay 2 transplanted with Lin^- cells transduced with the coPKR expressing LV vector (coRPK 1-14), whose blood and BM was analysed at different time-points, including secondary recipients transplanted with a pooled BM from a sub-group of primary transplanted mice (coRPK 11-14).

[0209] Each identical IS has different sequence reads (sequence count) among the different mice. Sequence counts can be used to determine whether samples from one mouse contaminated the other mice' samples based on the abundance criterion. In our rationale, an integration found in two mice will be assigned to the mouse that shows the highest abundance, while in the other mouse it will be considered as a contaminant. Therefore, we could identify a threshold of differential sequence count that allows assigning a given collision to a mouse and removing from the others. We retrieved the threshold value from our data obtaining a value of 10, meaning that for each IS, among all TI, if an IS has got an abundance value (percentage sequence count ratio) 10 times lower than the highest abundance value (percentage sequence count) of the other TIs, then it is discarded from the current TI. We applied these rules both

among the TI and the selected groups, and Excel files were used to compute the collisions detection by applying the following rules (here detailed for TI filtering but that are extended to groups filtering as well): 1. Isolating each TI, group all samples of the same TI together by summing the sequence counts. 2. For the three TIs obtained, for each IS, compute the percentage ratio of the IS sequence count versus the overall sum of reads for the TI. 3. Then, applying the following rule to compute the decision step with the threshold of 10 that allowed to assigning each IS to a reliable TI.

[0210] Once an IS was detected to remove, reads of that IS were removed from the group so that it will no longer assigned also to that group. The filter described above was applied between the mice transplanted with different *ex-vivo* transduced cell populations (one cohort of EGFP expressing mice from assay 2, and two cohorts of coPKR mice belonging to two independent transplantation experiments). Moreover for the coPKR-TC group (assay 2), the filter method mentioned above was modified in two ways: a) To better highlight the sharing of integrations between time-points in the context of the **clonal abundance analysis** we added the following rule: if one integration is shared between one or more mice then the integration will be kept for all time points even if their sequence count is less than the 10% of the maximum sequence count among mice; b) For **lineage tracking** relationships we applied a more stringent filter by eliminating the IS with a sequence count lower than 3 and the 10% sequence count filter for sharing between time-points. Meaning that an integration shared between two time points will be kept or discarded only if is more or less than the other respectively.

[0211] Gene ontology analysis. All gene ontology analyses were made using the GREAT online software (http://bejerano.stanford.edu/great/public/html/). The web page allow to upload the genomic coordinates of the integrations of each dataset and calculates the enrichment levels in the tested dataset by correlating positional information (based on the binomial distribution analysis for p-value calculations) and annotated function of the genes nearest to the integration sites (based on the hypergeometric distribution analysis for p-value calculations) [Groeschel S, et al. (2011). J Inherit Metab Dis 34: 1095-1102]. Biological processes and molecular functions of the GO database were chosen for enrichment analysis. Only the gene classes with a false discovery rate <0.05 for both statistical analyses were considered (**Fig. 20**).

[0212] Data storage. All data, both row data and results, are stored in TIGET network attached file storage (NAS) in the root folder, in which all alignments from the pipeline are available, as well as the abundance matrix and plots. NAS storage is secured by authentication and

authorization policies, and was built on a reliable and scalable infrastructure using redundancy array of disks RAID 5, and it is under backup on our CrashPlan software registered in TIGET.

Example 1

PGK-coRPK therapeutic LV vector leads to a stable and long-term correction of the anemic phenotype in genetically corrected PKD mice.

[0213] *In vivo* efficacy of the PGK-coRPK LV (**Fig. 2a**) was assayed by transduction and transplantation of lineage depleted BM cells (Lin^- cells) from PKD mice (**Fig. 2b**). **Fig. 2a** is a schematic representation of the SIN LV vectors used throughout gene therapy experiments harboring the human PGK promoter regulating the expression of the EGFP transgene in the control vector (upper diagram) or the expression of coRPK cDNA in the therapeutic vector (lower diagram). The coRPK sequence showed 80.4% homology with the human *PKLR* cDNA and 76.5% homology with the mouse *Pklr* cDNA, with no changes in the amino acid sequence. **Figure 2b** is a schematic of the gene therapy protocol performed to address the functionality of the developed PGK-coRPK LV vector. Correction of the PKD phenotype was studied for 4 to 9 months after transplant in PB and BM through hematological analysis and metabolic profiling. Integration analysis was performed in different tissues and time-points from all mice to address LV vector safety. At 280 days post-transplantation, total BM from primary transplanted mice carrying the coRPK transgene was transplanted again into lethally-irradiated female PKD mice (secondary recipients) to test the stability and safety of the engraftment. Lethally irradiated PKD mice transplanted with deficient cells transduced with the coRPK LV showed a significant improvement in all tested blood erythroid parameters when compared to non-transplanted PKD littermates or to mice transplanted with cells transduced with an EGFP LV (**Fig. 3 and Table 2**).

Table 2. Hematological variables recorded 140 days post-transplantation in peripheral blood.

Group	HGB (g/dL)	HTC (%)	MCV	MCH (pg)
Healthy (n=5)	14.64 ± 0.99	36.15 ± 2.64	38.20 ± 0.86	15.44 ± 0.21
PKD (n=9)	9.70 ± 0.55	28.91 ± 1.53	51.56 ± 0.50	17.23 ± 0.31
EGFP (n=8)	6.81 ± 0.68	21.32 ± 1.51	49.13 ± 0.72	15.46 ± 0.62
coRPK (n=17)	10.67 ± 0.53*	31.09 ± 1.45*	45.65 ± 0.84	15.35 ± 0.53
2nd coRPK	12.40 ± 0.67**	34.11 ±	46.25 ± 0.85	16.78 ± 0.23

Group	HGB (g/dL)	HTC (%)	MCV	MCH (pg)
(n=4)		1.49*		

[0214] Data represent the mean ± SEM and were statistically analysed by comparison to EGFP-expressing mice using the Kruskal-Wallis non-parametric test. *p<0.05; **p<0.01. RBC counts increased as soon as 40 days post-transplantation (**Fig. 3a)** and constitutive reticulocytosis, which is one of the most common signs of PKD, was significantly reverted in mice carrying the PGK-coRPK transgene, reaching levels close to those observed in healthy controls for at least 9 months after transplant (**Fig. 3b**). On the contrary, PKD animals transplanted with EGFP LV-transduced cells showed anemia and remarkable reticulocytosis in parallel with PKD mice at all time-points analyzed. Hemoglobin levels (HGB), hematocrit index (HTC), mean corpuscular volume (MCV) and mean corpuscular hemoglobin (MCH) values were also corrected in mice transplanted with LV-corrected cells when compared to non-transplanted PKD littermates (**Table 2**). This hematological correction was achieved with 63.66 ± 4.45 % of donor chimerism and transduction efficacies ranging from 60% to 90% (**Table 3**).

Table 3 Relevant molecular parameters in mice transplanted with genetically modified cells.

Assay	Groups	Vector copy number (VCN/cell)			Transduction % Provirus+ CFUs	Donor chimerism % SRY+ PB cells
		WBC	Total BM	CFU		
1 MOI	EGFP (n=2)	0.83 ± 0.05	0.42 ± 0.03	0.42 ± 0.00	57.73 ± 12.28	n.d
1-4	coRPK (n=3)	4.76 ± 0.28	3.58 ± 0.34	3.07 ± 0.76	91.08 ± 3.67	n.d
2 MOI	EGFP (n=6)	4.56 ± 0.50	4.19 ± 1.29	3.93 ± 0.98	92.32 ± 7.68	61.82 ± 3.61
10	coRPK	1.65 ±	0.99 ±	1.89 ±	62.06 ± 11.73	63.66 ±

(n=14)	0.08	0.13	0.42		4.45
2nd coRPK (n=4)	1.44 ± 0.08	n.d	n.d	63.15 ± 0.31[a]	62.89 ± 5.61

[0215] Data represent the mean ± SEM, n.d, not determined. [a] estimated transduction percentage obtained by interpolation in the linear regression built from experiment 1 (X axis: VCN/WBC, Y axis: % provirus$^+$ CFUs).

[0216] Transduced cells showed on average 1.65 ± 0.08 integrated vector copies per cell, indicating that PGK-coRPK LV-vector provided enough human RPK transgenic expression to revert the hemolytic anemia. Remarkably, the expression of the coRPK transgene led to an extension of erythroid cell half-life when compared to non-transplanted PKD mice (**Figs. 3c,d**). On average, PKD mice showed a RBC half-life of 19 days, while in genetically corrected mice this was extended to 25 days (6 days' extension) reaching values close to wild–type RBC half-life (**Fig. 3d**). Thus, RBCs of coRPK-expressing mice displayed intermediate survival kinetics between healthy and deficient control mice (**Fig.3c**), most likely because full chimerism was not achieved in these animals (**Table 3**).

[0217] Nine months after transplant hematopoietic progenitors from primary recipients were transplanted into secondary recipients that maintained engraftment levels (62.89 ± 5.61 %) and VCN (1.44 ± 0.08 copies) (**Table 3**). Secondary transplanted recipients showed a multi-lineage hematopoietic reconstitution up to 5 months post-transplantation (**Fig. 4**) and a significant improvement in all PB erythroid parameters (**Fig. 5 and Table 2**). **Figure 4a** is a diagram of the flow cytometry strategy used to identify the different hematopoietic lineages by labeling with CD3-PE, B220-PE, B220-PECy5, Gr1-Biotin and Mac1-Biotin antibodies plus SAV-PE-Cy5. **Figure 4b** depicts representative dot-plots and percentages (**Figure 4c**) of each lineage in PB at 140 days after transplant. Bars represent the average percentage ± SEM of healthy (n=2, black bar) and PKD mouse (n=2, grey bar) controls and secondary transplanted mice expressing the coRPK therapeutic transgene (n=4, scratched bar). In addition, proviral integrations were detected in differently committed hematopoietic progenitors (**Figs. 6a,b**) and its number remained constant over time (**Fig.6c**) demonstrating the stability of the genetic correction and highlighting the safety of the PGK-coRPK LV. **Figure 6a** shows the vector copy number per cell in BM CFUs from individual transplanted mice at 120 to 170 days after transplant. Transduction and chimerism percentages are also shown. **Figure 6b** shows provirus copy number in cells from different hematopoietic compartments. Columns represent the

average ± SEM of the different groups of transplanted mice. **Figure 6c** shows the kinetics of proviral integrations in BM cells from individual transplanted EGFP-expressing mice (grey lines) and mice carrying the coRPK transgene (black lines).

Example 2

LV-derived RPK expression normalizes erythroid differentiation and allows the production of functional mature erythrocytes.

[0218] PKD mice show a characteristic expansion of the erythroid compartment caused by the compensatory erythropoiesis mechanism (Min-oo et al 2004). The study of the erythroid differentiation pattern in transplanted mice indicated that the ectopic RPK expression reverted this mechanism (**Figs. 7a,b**). PKD and EGFP-expressing mice showed a predominance of immature erythroid precursors (subpopulation I: proerythroblasts and subpopulation II: basophilic erythroblasts) in BM and spleen, and a remarkable fall in late erythroid cells (population IV: reticulocytes and mature erythrocytes), while mice transplanted with cells transduced with the coRPK LV showed a significant reduction of immature erythroid precursors (subpopulations I and II) in BM and spleen, and a significant increase of the latest erythroid compartment (subpopulation IV) equivalent to healthy mice (**Figs. 7a,b**). In addition, unlike PKD and EGFP-expressing mice, those carrying the coRPK transgene showed a significant reduction of erythropoietin (Epo) levels in plasma (**Fig. 7c**). **Figure 8a** shows the total CFUs from spleen and **Figure 8b** shows the bone marrow at 140 days after transplant. Dots represent number of colonies per mouse analyzed and lines represent average ± SEM in each group. Data were statistically analyzed by non-parametric Kruskal-Wallis test. The normalization of the erythropoiesis in PKD mice treated with the therapeutic vector was accompanied by a reduction of the splenic number of progenitors content to normal levels (**Fig. 8a**), although no changes in the BM CFU content were noted (**Fig. 8b**).

Example 3

Transplantation of cells transduced with the coRPK LV vector reverts extramedullar erythropoiesis and organ pathology.

[0219] Due to the active destruction of RPK deficient erythrocytes, PKD and EGFP-expressing mice displayed acute splenomegaly with an increase of spleen weight and size of over 200% when compared to healthy controls (**Figs. 9a,b**). A disorganized structure of the splenic tissue and the expansion of the spleen red pulp was also observed in these animals, indicating an intense extramedullar erythropoiesis also supported by the presence of erythroid cell clusters

in PKD and EGFP-expressing liver sections (**Fig. 9c**). Remarkably, the ectopic expression of coRPK transgene completely reverted spleen and liver pathology in genetically corrected mice, reducing the RBC accumulations and normalizing spleen histological structure and size (**Fig. 9**). Additionally, histological studies revealed the total absence of iron deposits in the liver of genetically corrected mice, whereas PKD mice either from the non-transplanted group or the group transplanted with HSCs transduced with the EGFP-carrying vector displayed an intense iron overload due to the continuous hemolytic process (**Fig. 9c**). Overall, the transplant of genetically corrected HSCs in PKD mice restored the normal status of the erythropoiesis and all the secondary effects caused by the hemolytic anemia.

Example 4

PGK-coRPK LV-derived expression restores the glycolysis pathway in RBCs without modifying the WBC metabolic balance.

[0220] Next, we performed an extensive metabolomic analysis of all transplanted and control mice to study the functional correction of RPK enzymatic activity. Following an untargeted profiling strategy, we observed significant changes of glycolytic intermediates in RBCs among the different groups, identifying three broad clusters of metabolite patterns with distinct trends (**Fig.10a**). RBCs from coRPK-expressing mice showed an increase of metabolites from cluster 1 similar to healthy controls but different from transplanted mice carrying the EGFP transgene. Likewise, cluster 3 reflected a reduced metabolite trend in genetically corrected mice similar to wild type mice and different from EGFP-expressing mice. Nevertheless, cluster 2 from assay 1 showed no differences of metabolite profile between transplanted mouse groups (EGFP and coRPK expressing mice) (**Fig. 10a**). The untargeted metabolic profiling also showed that the genetic modification was capable of modifying some important glycolytic intermediates, achieving an increase in ATP (**Fig. 10b**), ADP (**Fig. 10c**) and pyruvate (**Fig. 10d**) levels in erythrocytes isolated from mice transplanted with PGK-coRPK LV-transduced HSCs. Considering these metabolic trends, we then analysed other metabolites located closer to the PK-catalysed reaction using a targeted profiling approach. Levels of the direct PK substrate phosphoenolpyruvate (PEP) (**Fig. 10e**) and the 3-phosphoglycerate (3-PG) (**Fig. 10f**), located upstream of the PK-catalysed reaction, approached to that of healthy control mice. Deficient erythrocytes expressing the coRPK transgene also produced an increase of D-lactate (**Fig. 10g**), the final product of the anaerobic glycolysis, when compared to PKD and EGFP expressing mice. To test whether the compensation in the glycolytic metabolites was a result of the normalization in the PK activity in mature erythrocytes, we measured the activity of this

enzyme and normalized it in relation to the Hexokinase activity in order to avoid the influence of high amounts of reticulocytes in the deficient animals. RBCs were purified through a cellulose column to prevent leukocyte PK activity contamination. A complete compensation of PK activity was observed in the animals expressing the coRPK that reach ratios similar to those obtained from wild type healthy animals and from a normal healthy blood donor volunteer (**Fig. 11**). **Fig. 11a** shows pyruvate kinase activity, **Fig. 11b** shows hexokinase activity and **Fig. 11c** shows ratio of pyruvate kinase and hexokinase enzymatic activities in RBCs from control mice and mice transplanted with transduced cells. RBCs were purified from blood samples through a cellulose column to avoid leukocyte PK activity contamination and subjected to enzyme activity evaluation. Black bars, healthy mice (n=2); white bars, mice transplanted with cells transduced with the EGFP expressing vector (n=3); scratched bars, mice transplanted with cells transduced with the coRPK expressing vector (n=3). Checkered bars represent values from a healthy volunteer (n=1). Data represent the average ± SEM of each group.

[0221] Principal component analysis showed that metabolite pattern of RBCs was different depending on the group and markedly different to WBC profile (**Fig. 12a**). On the contrary, WBC sub-groups clustered together with very little difference among groups, indicating no changes in the metabolic balance of leukocytes when expressing the ectopic coRPK (**Fig. 12a**). Additionally, specific metabolite changes observed in the RBC untargeted profiling were not present in WBCs (**Figs. 12b-d**).

Example 5

PGK-coRPK LV transduced cells render polyclonal hematopoietic reconstitution without evidence of vector genotoxicity.

[0222] Integration profile of LVs carrying either the coRPK or the EGFP transgene was analysed in transplanted mice. Resulting from the genome-wide integration profile of LVs, each insertion created a unique genetic mark that can be used to track the clonal behaviour in individual transduced cells. Genomic DNA (gDNA) was obtained from WBCs and from BM cells and from primary and secondary transplanted mice, as well as from transduced cell pools before transplant (Lin^- cells). Linear Amplification Mediated PCR (LAM-PCR) (**Fig. 13 and 14**) was used to amplify vector/genome junctions, and to identify vector ISs. **Figure 13** demonstrates vector ISs were identified by LAM-PCR amplification of 3'vector LTR-genome junctions. A MultiNA automated system was used, generating a pattern characterized by several bands. Vector backbone derived Tsp509I internal control band (IC) is indicated by an arrow. **Figure 14** demonstrates vector ISs were identified by LAM-PCR amplification of 3'

vector LTR-genome junctions. A MultiNA automated system was used, generating a pattern characterized by several bands. Vector backbone derived HpyCH4IV5 IC is indicated by an arrow.

[0223] PCR products were sequenced by MiSeq Illumina platform and the obtained sequences were mapped onto the mouse genome by bioinformatics pipeline and filtered for collisions as described in the Methods section above (**Fig. 15**). **Figure 15** is the general scheme of the analysis of IS mapping performed in mice transplanted with genetically modified hematopoietic progenitors. Bone marrow and white blood cell samples from transplanted mice belonging to two independent experiments (**Table 3**) and harvested at different time-points after transplant were analyzed as described in supplementary methods following the showed pipeline.

[0224] Overall, we mapped 5,173,892 sequencing reads on the transplanted mouse genome, resulting in 2,220 unique vector integration sites. The genomic distribution of ISs from two independent experiments matched the previously reported LV preference for integration within transcriptional units (particularly within the first 50 Kb downstream of the transcription start site -TSS-) (**Fig. 16a**), showing no skewing towards any particular chromosome in the mouse genome (**Fig. 16b**). **Figure 16a** shows IS frequency distribution around TSS of the nearest RefSeq gene, spanning 500 Kb upstream and downstream the TSS. Numbers on the top are the number of IS detected for all samples and time-points. **Figure 16b** shows chromosomal distribution of LV ISs in transplanted mice expressing the EGFP transgene (black bars) or the coRPK therapeutic transgene (grey bars), showing no skewing towards any particular chromosome.

[0225] Safety of the PGK-coRPK LV-based gene therapy was studied by clonal abundance estimations, calculating the percentage of sequence count for each IS (a clonal mark) with respect to the total number of sequences of the dataset. Dot plots and heat map representations of the relative abundance of each IS retrieved for each mouse (**Figs.17, 18 and 19**) showed strong fluctuations in clonal composition of the different mice and of *in vitro* cultured Lin^- cells. Figure 17 is a chart of tracked shared integrations between primary and secondary recipient mice carrying the therapeutic PGK-coRPK LV vector. Integrations detected in either mouse in any organ and at any time are pooled. Secondary recipients received the pooled BM from transplanted mice coRPK11 to 14. The rest of the IS detected were detected in the primary or in the secondary recipients. Numbers in the boxes show the representativeness in percentage of the corresponding integration in the referred mouse. In addition to $\geq 5\%$ filter applied on integration analysis, all integrations with a sequence count < 3 were eliminated. **Figure 19**

presents dot plot representation of clonal abundance of pooled integrations in each mouse in bone marrow**.** The relative percentage (y-axis) for each IS is relative to the total number of sequences reads obtained in each dataset. Similarly to co-RPK transduced cells (**Fig. 17**), the graph indicates that the vast majority of transplanted mice show a polyclonal pattern of hematopoietic repopulation.

[0226] Also, it was possible to appreciate that for several samples, a small number of integrations contributed to a large amount of sequence reads (**Figs. 17 and 18**), revealing a polyclonal pattern of repopulation of transduced HSCs. In addition, tracked shared integration between primary mice carrying the therapeutic PGK-coRPK LV and subsequently transplanted secondary mice showed no strong sharing of integrations between the groups, confirming the absence of clonal dominance (**Fig. 18**).

[0227] To determine whether hallmarks of insertional mutagenesis were present in transplanted mice, we assessed the occurrence of Common Insertion Sites (CIS) similar to currently ongoing LV-mediated clinical trials. CIS are insertional hotspots that may result from integration bias at the time of transduction or *in vivo* selection of clones harbouring vector integrations that confer growth advantage. CIS were identified using an algorithm based on Abel and cols and the Grubbs test for outliers finding no CIS and thus no alarming signs of genotoxicity by this readout. Moreover, GO analysis revealed no skewing towards gene classes involved in cancer, cell proliferation or regulation of apoptosis in any of the integration datasets sorted by tissue-distribution, time point or abundance of repopulating hematopoietic cell clones (**Fig. 20**). **Figure 20** represents LV genomic integration profile. GO analysis was performed using the GREAT software on samples from transplanted mouse. All integrations retrieved from this study (N=2220) showed overrepresentations of the gene functions indicated on the left part of the figure. To address if the most abundant integrations were enriched on specific gene classes, all ISs with a relative sequence count >5% of the entire dataset (shown in **Fig. 17**) were selected, showing no GO gene classes overrepresented.

[0228] These results suggest neutrality of vector integration and demonstrate the safety of the PGK-coRPK LV in a preclinical setting.

Example 6

Human Clinical Trial

[0229] A clinical trial is conducted to evaluate safety and preliminary efficacy of autologous hematopoietic stem cell transplantation (HSCT) using the EU/3/14/1130 medicinal product (autologous CD34+ hematopoietic stem cells transduced with the LV vector containing the

RPK gene) in patients with PKD with a history of severe and transfusion-dependent anemia refractory to splenectomy.

[0230] The ODD EU/3/14/1130 comprises a SIN LV vector expressing the coRPK gene (**Figure 21**).

[0231] SIN LV vectors provide a more robust expression (Ellis 2005) and are less susceptible to transcriptional silencing than gamma-RV vectors (Pfeifer, Ikawa et al. 2002). They also show a much safer integration profile (Schroder, Shinn et al. 2002) (Mitchell, Beitzel et al. 2004) (Wu, Li et al. 2003), and because of the 400 bp deletion that they carry in the 3' LTR sequence (Miyoshi, Blomer et al. 1998) (Zufferey, Dull et al. 1998), transgene expression is regulated by internal promoters, increasing the safety of the LV-based genetic modification.

[0232] The sequence of the accepted LV vector also includes several modifications to improve transgene expression and safety in target cells.

[0233] One modification is the use of the human PGK promoter, already characterized by its stable *in vivo* activity and improved safety properties compared to other promoters used in gene therapy (Montini, Cesana et al. 2006, Modlich, Navarro et al. 2009, Montini, Cesana et al. 2009, Biffi, Montini et al. 2013). Incorporation of the PGK promoter leads to a physiological expression of the transgene and a lower susceptibility to transcriptional silencing (Gerolami, Uch et al. 2000, Zychlinski, Schambach et al. 2008).

[0234] Another modification is the coRPK gene to increase mRNA stability upon transcription. For the optimization the GeneArt® software has been used, increasing the GC content and removing cryptic splice sites in order to avoid transcriptional silencing and therefore increase transgene expression. The coRPK optimized sequence showed 80.4% homology with the human *PKLR* gene, with no changes in the amino acids sequence of the protein.

[0235] Another modification is a mutated wPRE, lacking any residual open reading frame (Schambach, Bohne et al. 2006) is also included to improve the level of expression and stability of the therapeutic gene. The backbone, promoter and wPRE* sequences of this LV vector (PGK-coRPK LV) are the same as the one corresponding to the medicinal product "Lentiviral vector containing the Fanconi anemia A (FANCA) gene for the therapy of Fanconi anemia Type A patients" (Ref 141/2000), as well as vector backbone used in the currently ongoing clinical trial for the metachromatic leukodystrophy (MLD) (Biffi, Montini et al. 2013).

[0236] Mode of action

[0237] After harvesting the $CD34^+$ progenitor cells from PKD patients, either from bone marrow (BM) or mobilized peripheral blood cells, they are transduced *ex vivo* with the medicinal product and the therapeutic vector will be integrated in the genome of the cells. Once

integrated, the therapeutic human gene (coRPK) are transcribed and translated within deficient cells to produce the therapeutic RPK protein that is missing or reduced in PKD mature erythrocytes. Transduced PKD hematopoietic progenitors are then genetically corrected, and thus able to produce RBCs with sufficient amounts of ATP to accomplish their functions (**Figure 22**). These genetically corrected hematopoietic progenitors (which will constitute the medicinal product) are then transplanted back into the patient, where once engrafted will generate normal erythrocytes, providing a life-long cure of the disease.

[0238] The active principle consists in a cellular suspension of corrected hematopoietic stem cells ($CD34^+$) cells with the therapeutic LV vector PGK.coRPK.wpre, designated as orphan drug by the European Commission (ODD EU/3/14/1130) for the treatment of PKD. Thus, this new medicament should be included in the groups of advanced therapy developments within the gene therapy sub-class.

[0239] The active principle is composed of a genetically modified cellular suspension of at least $2x10^6$ $CD34^+$ cells/kg of body weight with at least 0.1 copies of the therapeutic vector per cell. The cells are suspended in a saline buffer with 2% HSA.

[0240] The final therapeutic product is produced according to GMP rules, so the product requirements for its liberation and its infusion in the patient will be related with the quality of the product. In relation to this, these specifications will include cellular viability $\geq 30\%$, sterility (Gram test and sterility by Pharmacopeia), absence of mycoplasma, absence of replication-competent LV particles, and demonstration of the potency of the therapeutic potential by the detection of the presence of at least 0.1 vector copies per cell by quantitative PCR. Additionally, investigation and research studies of the content of hematopoietic progenitors and the vector copy number in these are performed. Three independent validations are performed with healthy control cells to assure the procedure reaches the above-described requirements.

[0241] The final product is packaged in a transfer bag sealed by heat and especially for freezing and storage until its infusion to the patient; previously, samples will be collected for the precise corresponding quality controls.

Mobilization

[0242] Patients are mobilized in their respective hospitals, although the two first patients are mobilized in the Hospital del Niño Jesús, Madrid (Spain). Mobilization process includes in the administration of recombinant stimulating factor granulocyte colony (G-CSF, Neupogen, Amgen, Thousand Oaks, CA, USA) at doses of 12 mg/kg twice a day for up to eight days from birth, Plerixafor 4th day 240 mg/kg/d (Mozobil®, Genzyme Europe BV, Naarden,

Netherlands) to four subcutaneous doses for 4 consecutive days. The hematopoietic progenitor cells from peripheral blood are collected by leukapheresis, and large volumes from day 5 of mobilization are sent through a cell separator according to standard protocols of the Hospital Niño Jesús in Madrid. All the instruments and solutions are CE marked and meet the specifications of the legislation for medical devices.

CD34⁺ cell purification

[0243] Consistent with the mobilization process, the apheresis takes place in the hospital where the patient is mobilized. Apheresis product is processed immediately to select hematopoietic progenitors ($CD34^+$ cells) through MACS "magnetic cell sorting" technology (Miltenyi Biotec, Germany) which permits separation of cells by a high magnetic field gradient, through a powerful permanent magnet and a separation column with a ferromagnetic matrix. The CliniMACS (Miltenyi Biotec, Bergisch Gladbach, Germany) system consists of a computer (CliniMACS®plus Instrument), a specific $CD34^+$ selection software, a set of sterile tubes (CliniMACS Tubing Sets), magnetic-controlled reactive sterile instrument (CliniMACS CD34 Reagent) and a sterile buffer (CliniMACS PBS/EDTA Buffer). The instrument and reagents used are CE marked and will meet the specifications of the legislation for medical devices. During this phase and in later washes nonspecific immunoglobulin (intravenous Flebogamma 5% 0.5 gr, Grifols) and human albumin (human albumin Grifols® 20%, Grifols) are employed, which are later removed by washing after centrifugation. The CD34+ cells are then quantified. Microbiological controls on the products obtained will be performed by taking standard fungal, aerobic bacteria and anaerobes samples for cultures following specific protocols.

CD34⁺ transduction

[0244] Transduction of the purified $CD34^+$ cells with the ODD EU/3/14/1130 is performed under GMP conditions in a time window within 48h since the extraction of the cells from the patient (apheresis). *Ex-vivo* culture of the cells lasts less than 48 hours, and will be cultured following stablished standards including the use of properly formulated media X-vivo-20 (Lonza), addition of hematopoietic growth factors (100ng/ml hrSCF, 100ng/ml hrFlt-3, 100ng/ml TPO and 20 ng/mL IL-3 (all from Prepotech), 1µg/mL Pulmozyme and controlled 5% O_2 concentration. Transduction will be performed with a GMP LV batch of the ODD EU/3/14/1130 produced by VIVEbiotech (San Sebastian, Spain). After transduction, cells are washed in X-vivo-20 (Lonza) to be finally packaged in a transfer bag suitable for cryopreservation. Specific samples are collected to determine if the final product meets all the

specifications already mentioned for its final liberation. Three independent validations are performed to evaluate the stability of the product. All products and solutions, including the vector, will meet the specifications of the legislation for medical devices and clinical use. Before production, all raw materials (including consumables, biological reagents and chemical powders) will be inspected by the Quality Control (QC) Unit of CliniStem according to standard operating procedures (SOPs).

Conditioning

[0245] Patients are conditioned according to standardized and specific protocols considered for the trial. To consider a patient for conditioning as an alternative, a back-up of $2x10^6$ unmanipulated $CD34^+$ cells/kg is kept frozen to be used in case the prepared product does not completely reconstitute the hematopoiesis of the treated patient.

Infusion

[0246] Before infusion, patient eligibility is checked to ensure they meet the study requirements. The day of infusion, premedication and prophylactic medication used are recorded.

Example 7

Non-clinical development

[0247] Previous work demonstrated the feasibility of HSC gene therapy for PKD in mice when above 25% genetically corrected cells were transplanted. These results suggest that a significant number of donor gene-corrected HSCs (Zaucha, Yu et al. 2001) and high levels of transgene expression are needed to achieve a therapeutic effect in PKD. We have developed a new therapeutic LV vector proposed for this clinical trial, harbouring the hPGK eukaryotic promoter driving the expression of the *PKLR* cDNA that was designated as an Orphan Drug on August 2014 (EU/3/14/1130). With this vector, we conducted a preclinical gene therapy protocol for PKD in a mouse model of the disease. With LV dosages based on clinical standards, ectopic RPK expression was able to normalize the erythroid compartment, correcting the haematological phenotype and reverting organ pathology. Metabolomic studies demonstrated the functional correction of the glycolytic pathway in genetically corrected RBCs, with no metabolic disturbances observed in leukocytes. Remarkably, WBCs analyzed in parallel showed no alterations of the metabolic balance in leukocytes when RPK is ectopically expressed under the activity of an ubiquitous promoter such as PGK, ruling out a

leukocyte metabolic advantage as possible safety concern and reinforces the therapeutic potential of the EU/3/14/1130 vector.

[0248] The multi-lineage reconstitution and the absence of any leukemic event or clonal expansion in secondary recipients after the proliferative stress induced by BM re-transplant demonstrate the long-term stability and safety of the PGK-coRPK LV vector-based protocol. The use of the human PGK eukaryotic promoter that i) likely led to a more physiological expression of the RPK transgene, ii) has been proven to be a weak transactivator and iii) is being currently used in the clinical trial for metachromatic leukodystrophy (MLD), could also account for the safety of the whole procedure.

[0249] To assess the long-term safety of HSC gene therapy through the analysis of vector ISs, next generation sequencing was used to predict the risk of insertional oncogenesis in HSC. More than 5,173,892 sequences reads were mapped on the mouse genome to a total of 2220 unique vector IS, finding no evidence of *in vivo* expansion or selection of clones harboring IS. Rather, our data shows the clonal composition and dynamics of hematopoiesis after transplantation of transduced HSCs in mice, suggesting a genuine and stable genetic *in vivo* modification of HSC over time. Overall, the analysis of the vector integration pattern emphasizes the safety properties of the PGK-coRPK LV vector that provides PKD genetic correction with no evidence of genotoxicity.

Example 8

Clinical development

[0250] Protocol assistance for clinical studies is requested to a Regulatory Agency. Our aim is to perform a clinical trial sponsored by the European Commission. The ForGeTPKD Consortium, composed by different clinicians and basic researchers in Europe, has been established to focus in PKD research and development of new therapeutic strategies. The ForGeTPKD clinical trial will be the first administration of this medicinal product in humans. It is designed as an International, Multicenter, Phase I/II Open Label Study to evaluate the Safety and Efficacy of Transplantation of Autologous CD34+ Cells Transduced Ex Vivo with a Lentiviral vector containing the red-cell type Pyruvate Kinase (RPK) gene (EU/3/14/1130) in patients with Severe Pyruvate Kinase Deficiency.

Regulatory status

[0251] The medicinal product has no marketing authorization at present time. The objective of the PKD Consortium is to move forward to the clinical development of the medicinal product in order to eventually receive a marketing authorization.

[0252] The mentioned final product will be produced with a LV vector that has received the Orphan Drug Designation related to:

[0253] Indication: Treatment of Pyruvate Kinase Deficiency

[0254] Criteria: The only curative treatment for PKD is allogeneic BMT, which has been used in patients with transfusion-dependent severe anemia refractory to other measures. However, allogeneic BMT is not a widely accepted treatment for PKD (only one patient reported in the literature (Tanphaichitr, Suvatte et al. 2000)) as it is associated with severe complications related to intensive pre-allo-BMT conditioning by chemotherapy or chemo-radiotherapy, as well as acute and chronic graft-versus-host disease (GVHD). Our hypothesis is that gene therapy using autologous hematopoietic stem cells transduced with viral vectors containing the wild-type version of the gene, provided by the ODD EU/3/14/1130, may represent a potential curative opportunity for these patients, avoiding the risks of GVHD, the main cause of failure of a hematopoietic progenitors transplant.

[0255] Active substance: Autologous $CD34^+$ hematopoietic stem cells, transduced with Lentiviral vector containing the red-cell type Pyruvate Kinase (RPK) gene (ODD EU/3/14/1130), expressing the wild type version of the protein

[0256] Finished product: Frozen bag containing at least $2x10^6$ active substances/kg body weight of the patient, suspended in a saline buffer with 2% HAS

Example 9

Pharmacology

[0257] Completed studies: The medicinal product developed includes several modifications in its sequence that provides some advantages for gene therapy of PKD: 1) The use of a SIN-LV vector design allowed a relatively easy and safe production of viral stocks, able to efficiently transduce HSCs; 2) the use of a weak eukaryotic promoter such as hPGK, which is less susceptible to silencing by methylation (Gerolami, Uch et al. 2000), leads to physiological expression of the transgene, achieving therapeutic levels with a viral dosage (1.65 VCN) within the clinical standards (Matrai, Chuah et al. 2010); and 3) the codon optimized transgene sequence and the presence of the mutated wPRE sequence increases transgene mRNA stability. No reporter gene was included in the therapeutic vector sequence, avoiding possible

immunogenic problems (Morris, Conerly et al. 2004); (Stripecke, Carmen Villacres et al. 1999).

[0258] The developed hPGK-coRPK LV medicinal product efficiently reverted PKD pathology in both primary and secondary deficient mice transplanted with progenitors transduced and corrected with the ODD EU/3/14/1130. The correction was achieved with cells carrying on average 1.65 copies per cell of the therapeutic transgene.

[0259] Human PGK promoter was potent enough to express clinically relevant levels of coRPK protein, restoring the hemolytic phenotype in transplanted mice.

[0260] The genetic correction was able to: extend RBC half-life; normalize the hematological variables and reticulocyte levels; revert the compensatory erythropoiesis constitutively activated in PKD mice; and rescue the pathology in spleen and liver, remarkably reducing the iron overload, which is one of the life-threatening complications of PKD. In addition, the ectopic expression of human RPK corrected the energetic defect in RBCs without altering the metabolic balance in WBCs, emphasizing the efficacy and safety of the medicinal product.

Example 10

Ongoing studies

[0261] Transduction of human hematopoietic progenitors from healthy donors and PKD patients for the study of: 1) determining the efficiency of transduction of the ODD EU/3/14/1130 in human cells; 2) defining the optimal vector copy number/cell to get efficient and therapeutic expression of the RPK therapeutic protein; and 3) defining the optimal conditions to get therapeutic transduction levels without losing hematopoietic stem cell ability.

[0262] Planned studies include set up of conditions for large-scale transduction at the GMP facility, and pre-validation and 3 validation studies to set up the optimal conditions to reach the required specifications defined for the final therapeutic product.

[0263] Toxicology

[0264] Completed study results include: 1) The ectopic expression of human RPK corrected the energetic defect in RBCs without altering the metabolic balance in WBCs; 2) Genome integration analysis of the vector has demonstrated that (i) The analysis of the relative abundance of specific cell clones revealed an oligoclonal hematopoietic reconstitution for some mice, with no clonal dominance for any primary and secondary transplanted mice; (ii) CIS analysis, considered a hallmark of insertional mutagenesis, did not show any sign of genotoxicity, nor an abnormal enrichment of CIS over time, as the detected CIS from the two independent gene therapy experiments performed in mice were not represented by high

sequence counts and did not preferentially target oncogenes; (iii) GO analysis of the genes targeted by the LV integration and analysis of the position of the vector integrations in specific regions of the genome demonstrated no skewing towards gene classes involved in cancer, cell proliferation or regulation of apoptosis; and (iv) Overall, medicinal product integration analysis did not show any evidence of genotoxicity.

[0265] Planned studies include 1) Analysis of Recombinant Competent Lentiviruses (RCL) production: Human T lymphocytes from healthy donors and from PKD patients will be transduced with the ODD EU/3/14/1130 and cultured in vitro for extended periods of time. Presence of the viral p24 protein will be analyzed in supernatants by ELISA to evaluate the potential generation of RCLs and 2) Biodistribution of the medicinal product: Mouse hematopoietic progenitors will be transduced with the ODD EU/3/14/1130 and transplanted into lethally irradiated recipients. One month post-transplant, animals will be sacrificed, and different organs (gonads, liver, kidney, brain, bone marrow, spleen and peripheral blood) will be analyzed for the presence of vector DNA; and 3) Vector integrome in human cells: Hematopoietic progenitors from healthy donors and from PKD patients will be transduced with the ODD EU/3/14/1130 and transplanted into severe immunodeficient mice to allow the engraftment and proliferation of human hematopoietic cells. At different time points (1, 2 and 3 months post-transplant) blood and BM transplants will be taken, sorted for human cells and subjected to vector integrome analysis as already performed with mouse cells.

Example 11

Human Clinical Trial

[0266] To test clinical efficacy, the ForGeTPKD trial will be conducted. The proposed clinical trial aims to evaluate safety and preliminary efficacy of autologous HSCT using the EU/3/14/1130 medicinal product (autologous $CD34^+$ hematopoietic stem cells transduced with Lentiviral vector containing the red-cell type Pyruvate Kinase (RPK) gene) in patients with PKD with a history of severe and transfusion-dependent anemia refractory to splenectomy.

[0267] The primary objective is to evaluate treatment, safety, and tolerability/feasibility. The following endpoints will be measured in accordance: 1) incidence and characterization of Adverse Events (AE), including: AE related to the infusion of the transduced cells, AE derived from conditioning prior to cell infusion, and AE derived from clonal evolution related with the transduced cells; and 2) number of patients with stem cell engraftment at 30 days post-transplant.

[0268] Secondary objectives are to evaluate preliminary treatment efficacy. The following endpoints will be measured in accordance: 1) Number of patients who become "transfusion independent" at the end of the study, and the number of patients who still need transfusions after treatment; 2) Ratio between the mean numbers of transfusions needed within the study period (1 year) with respect to the mean number of transfusions in the last 1.5 years before baseline evaluation; 3) Clinically significant reduction of anemia, defined as the number of patients with a rise in hemoglobin levels in 2 gr/dL from baseline at the end of the study; 4) Clinically significant reduction of reticulocytosis, defined as the number of patients with a reduction of 50% from baseline evaluation at the end of the study; and 5) Number of patients with stem cell engraftment where 1% transduced cells can be detected at 6 and 12 month post-cell infusion at the end of the study.

[0269] An exploratory objective is to evaluate treatment impact on patient´s quality of life. The following end-point will be measured in accordance: Improvement in quality of life from baseline at the end of the study, using a quality of life questionnaire (SF-36 for adults or PEDSQL for children) and its translated validated versions in the language of the participant countries (Italian, Dutch and Spanish).

[0270] The ForGetPKD trial is a multicenter, international trial, which will be carried out in 3 EU member states: Spain, Italy and The Netherlands. Participating centers include Reference National Investigators and Institutions for PKD diagnosis and treatment.

[0271] The trial will represent the first administration to humans of the described product. It is designed as a non-comparative, open-label, single-dose, Phase I/II study.

[0272] Global Study duration will be 2 years from the first visit of the first patient to the last visit of the last patient. This includes 1 year of recruitment period and 1 year of treatment period and early (immediate) follow-up. After the end of the trial, included subjects will be asked to participate in a subsequent follow-up study that will monitor safety and efficacy for up to 5 years after the transplant.

[0273] Study procedures include a Screening Period, a Treatment Period and a Follow-up Period. Details of visits on each phase and associated study procedures are detailed below and summarized in Table 4.

Table 4. Study Procedures

Period	**Screening Period**		**Treatment Period**	**Follow up period**			
Visit	**-1 Pre-screening**	**0 Baseline**	**1 Treatment period**	**2 + 1 m**	**3 +3m**	**4 +6m**	**5 +12m**
Study procedures							
Informed consent	**X**						
Inclusion/exclusion Criteria	**X**	**X**					
Medical History	**X**						
Physical Examination	**X**	**X**	**X**				
CBCs	**X**		**X**	**X**	**X**	**X**	**X**
Biochemistry	**X**		**X**	**X**	**X**	**X**	**X**
Coagulation	**X**		**X**	**X**	**X**	**X**	**X**
Conditioning regimen			**X**				
Transduced cells infusion			**X**				
Discharge			**X**				
Bone Marrow sampling							**X**
Cell engraftment in peripheral blood				**X**	**X**	**X**	**X**
Average vector copy number in peripheral blood				**X**	**X**	**X**	**X**
Vector integration pattern in peripheral blood				**X**	**X**	**X**	**X**
Cell engraftment in					**X**		**X**

Period	Screening Period		Treatment Period	Follow up period			
Visit	**-1 Pre-screening**	**0 Baseline**	**1 Treatment period**	**2 + 1 m**	**3 +3m**	**4 +6m**	**5 +12m**
bone marrow							
Average vector copy number in bone marrow					**X**		**X**
Vector integration pattern in bone marrow					**X**		**X**
Cell extraction	**X**						
Availability of viable cells	**X**						
Quality of life questionnaire (SF-36 or PEDSQL)	**X**						**X**
Concomitant Medication	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Adverse Events	**X**	**X**	**X**	**X**	**X**	**X**	**X**

Screening Period

Visit -1: Pre-screening visit

[0274] Potential candidates will be informed of the aims and characteristics of the trial, and written informed consent will be taken, fulfilled and signed by the patient (or legal representative if underage), by duplicate. To be eligible for the study, patients must fulfill all inclusion criteria and none of the exclusion criteria, which will be checked. This includes a pre-treatment procedure to mobilize and obtain viable $CD34^+$ cells, to A. Store $2x10^6$ CD34+ cells/kg body weight to serve as a backup in case of non-engraftment, B. transduce at least $6x10^6$ CD34+ cells/kg body weight with the EU/3/14/1130 vector to generate the medicinal product and to perform all the quality control needed for the liberation of the medicinal product.

Only patients with enough transduced cells available after liberation of the medicinal product ($2x10^6$ transduced $CD34^+$ cells/kg body weight) will be included in the study.

[0275] The following procedures will be also performed in this visit:

- Register of relevant medical o surgery history.
- Register of demographic data and clinically relevant physical examination findings
- Register of relevant concomitant medications.
- Peripheral blood testing for routine cell blood counts (CBC), Biochemistry, Coagulation determination and serology.
- Echocardiogram, Lung function test and thorax X-ray.
- Quality of life questionnaire (SF-36 or PEDSQL).
- Genetic diagnosis of PKD.

[0276] To be eligible for the study, patients have to fulfill all of the following inclusion criteria and none of the exclusion criteria.

[0277] Inclusion criteria are male or female patients, Age >2 year old at the time of recruitment, willing to give signed informed consent (which will be signed by their parents or legal representative in case of children under 18 years old), previous diagnosis for PKD confirmed by genetic testing, history of severe transfusion-dependent anemia, not responsive to splenectomy, Candidate to Autologous Hematopoietic Stem Cell Transplant, $\geq 2x10^6$ transduced $CD34^+$ cells/kg body weight available, and treated and followed for at least the past 2 years in a specialized center that maintained detailed medical records, including transfusion history.

[0278] Exclusion criteria are: positive for presence of human immunodeficiency virus type 1 or 2 (HIV 1 and HIV 2), uncorrected bleeding disorder, presence of other causes of hemolysis, any prior or current malignancy or myeloproliferative or immunodeficiency disorder, immediate family member with a known or suspected Familial Cancer Syndrome (including but not limited to hereditary breast and ovarian cancer syndrome, hereditary non-polyposis colorectal cancer syndrome and familial adenomatous polyposis), in patients with previous allogeneic transplant, presence of residual cells of donor origin, patients with severe complications that after medical evaluation are considered to suffer grade III/IV cardiac, pulmonary, hepatic or renal function abnormalities, uncontrolled seizure disorder, diffusion capacity of carbon monoxide (DLco) <50% of predicted (corrected for hemoglobin), any other evidence of severe iron overload that, in the Investigator's opinion, warrants exclusion, participation in another clinical study with an investigational drug within 30 days of Screening, availability of a HLA-identical family donor for allogeneic bone marrow transplant, pregnant or breast-feeding women, patients that, according to investigator criteria, will not be able to

understand study purposes, benefits and risks and/or to comply with study procedures, an poor functional status, evidenced by a Karnofsky Index ≤80 in adults or Lansky ≤80 in children.

[0279] The statistical analysis of the study will be descriptive. Qualitative endpoints will be described by frequencies and percentages. Qualitative endpoints include the adverse events, the number of patients with stem cell engraftment, the number of patients who become "transfusion independent", the number of patients who have a clinically significant reduction of anemia, the number of patients who have a clinically significant reduction of reticulocytosis, the number of patients with stem cell engraftment where presence of transduced cells can be detected, and the improvement in quality of life from baseline measured by SF-36 or PEDSQL questionnaire. Quantitative endpoints will be described by mean and standard deviation or by median and quartiles. Quantitative endpoints include the reductions of anemia and reticulocytosis from baseline, the number of transfusions needed within the study period with respect to the number of transfusion in the last 1 year before baseline evaluation, and the vector copy number in peripheral blood and bone marrow. All endpoints will be described at the end of the study.

[0280] Previous work demonstrated the feasibility of HSC gene therapy for PKD in mice when above 25% genetically corrected cell were transplanted. These results suggest that a significant number of donor gene-corrected HSCs (Zaucha, Yu et al. 2001) and high levels of transgene expression are needed to achieve a therapeutic effect in PKD. We have developed a new therapeutic LV vector proposed for this clinical trial, harboring the hPGK eukaryotic promoter driving the expression of the *PKLR* cDNA that was designated as Orphan Drug on August 2014 (EU/3/14/1130). With this vector we conducted a preclinical gene therapy protocol for PKD in a mouse model of the disease. With LV dosages based on clinical standards, ectopic RPK expression was able to normalize the erythroid compartment, correcting the hematological phenotype and reverting organ pathology. Metabolomic studies demonstrated the functional correction of the glycolytic pathway in genetically corrected RBCs, with no metabolic disturbances observed in leukocytes. Remarkably, WBCs analyzed in parallel showed no alterations of the metabolic balance in leukocytes when RPK is ectopically expressed under the activity of an ubiquitous promoter such as PGK, ruling out a leukocyte metabolic advantage as possible safety concern and reinforces the therapeutic potential of the EU/3/14/1130 vector.

[0281] The multi-lineage reconstitution and the absence of any leukemic event or clonal expansion in secondary recipients after the proliferative stress induced by BM re-transplant demonstrate the long-term stability and safety of the PGK-coRPK LV vector-based protocol. The use of the human PGK eukaryotic promoter that likely led to a more physiological

expression of the RPK transgene, that has been proven to be a weak transactivator and is being currently used in the clinical trial for metachromatic leukodystrophy (MLD) could also account for the safety of the whole procedure.

[0282] To assess the long-term safety of HSC gene therapy through the analysis of vector ISs, next generation sequencing was used to predict the risk of insertional oncogenesis in HSC. More than 5,173,892 sequences reads were mapped on the mouse genome to a total of 2220 unique vector ISs, finding no evidence of *in vivo* expansion or selection of clones harboring IS. Rather, our data show the clonal composition and dynamics of hematopoiesis after transplantation of transduced HSCs in mice, suggesting a genuine and stable genetic *in vivo* modification of HSC over time. Overall, the analysis of the vector integration pattern emphasizes the safety properties of the PGK-coRPK LV vector that provides PKD genetic correction with no evidence of genotoxicity.

[0283] The invention may be further defined by reference to the following illustrative clauses:

1. An expression cassette comprising a polynucleotide sequence comprising in the following 5' to 3' order:

a) a promoter sequence;

b) a sequence encoding a gene product; and

c) an ribonucleic acid (RNA) export signal,

wherein the promoter sequence is operably linked to the sequence encoding the gene product.

2. The expression cassette of clause 1, wherein the promoter is a phosphoglycerate kinase (PGK) promoter.

3. The expression cassette of clause 1 or clause 2, wherein the gene product is a therapeutic gene product.

4. The expression cassette of clause 3, wherein the therapeutic gene product is a pyruvate kinase (PK) polypeptide, optionally a pyruvate kinase, liver and red blood cell (PKLR) polypeptide.

5. The expression cassette of any of clauses 1-4, wherein the sequence encoding the gene product is codon-optimized.

6. The expression cassette of clause 5, wherein the codon-optimized sequence comprises a sequence having at least 85% identity to SEQ ID NO: 8.

7. The expression cassette of any of clauses 1-6, wherein the RNA export signal is a mutated post-transcriptional regulatory element of the woodchuck hepatitis virus (wPRE).

8. The expression cassette of clause 7, wherein the mutated wPRE is a chimeric wPRE comprising a sequence having at least 80% identity to SEQ ID NO:24.

9. The expression cassette of any of clauses 1-8, further comprising one or more enhancer sequences.

10. The expression cassette of any of clauses 1-9, further comprising a polypurine tract (PPT) or polyadenylation (polyA) signal sequence.

11. The expression cassette of any of clauses 1-10, further comprising one or more of the following sequences:

i) a packing signal sequence;

ii) a truncated Gag sequence;

iii) a Rev responsive element (RRE);

iv) a central polypurine tract (cPPT);

v) a central terminal sequence (CTS); and

vi) an upstream sequence element (USE), optionally from simian virus 40 (SV40-USE).

12. The expression cassette of any one of clauses 1-11, further comprising 5' and 3' long terminal repeat sequences.

13. A recombinant gene delivery vector comprising the expression cassette of any one of clauses 1-12.

14. The recombinant gene delivery vector of clause 13, wherein the recombinant

gene delivery vector is a virus or viral vector.

15. The recombinant gene delivery vector of clause 14, wherein the virus or viral vector is a lentivirus (LV).

16. A cell comprising the expression cassette of any of clauses 1-12 or the recombinant gene delivery vector of any of clauses 13-15.

17. The cell of clause 16, wherein the cell is a hematopoietic stem cell.

18. The cell of clause 16, wherein the cell is a committed hematopoietic erythroid progenitor cell.

19. A pharmaceutical composition comprising a pharmaceutically acceptable excipient and the recombinant gene delivery vector of any of clauses 13-15 or the cell of any of clauses 16-18.

20. A method of treating or preventing a disease or disorder in a subject in need thereof, comprising providing to the subject the pharmaceutical composition of clause 19.

21. The method of clause 20, wherein the disease or disorder is a Pyruvate Kinase Deficiency (PKD) and the gene product is a pyruvate kinase (PK) polypeptide, optionally a pyruvate kinase, liver and red blood cell (PKLR) polypeptide.

22. The method of clause 20 or clause 21, wherein the pharmaceutical composition comprises the recombinant gene delivery vector.

23. The method of clause 20 or clause 21, wherein the pharmaceutical composition comprises the cell.

24. The method of clause 23, wherein the cell is autologous to the subject.

25. A method for expressing a transgene in erythroid cells, comprising contacting one or more erythroid cells with an effective amount of a recombinant viral vector, wherein

the vector comprises a human phosphoglycerate kinase promoter, a codon optimized version of a human pyruvate kinase, liver and red blood cell (PKLR) cDNA transgene, and a mutated post-transcriptional regulatory element of the woodchuck hepatitis virus, wherein following said contacting, PKLR is expressed at detectable levels in the one or more erythroid cells.

Claims

1. An expression cassette for pyruvate kinase deficiency (PKD), comprising a polynucleotide sequence comprising in the following 5' to 3' order:

a) a promoter sequence,

b) a codon-optimized sequence encoding a human *PKLR* gene product; and

c) a mutated post-transcriptional regulatory element of woodchuck hepatitis virus (WPRE),

wherein the promoter sequence is operably linked to the codon-optimized sequence encoding a human *PKLR* gene product, and

wherein the codon-optimized sequence encoding a human *PKLR* gene product shares at least 95% identity to SEQ ID NO: 8, and

wherein the mutated WPRE is a chimeric WPRE comprising a sequence having at least 90% identity to SEQ ID NO: 24.

2. The expression cassette of claim 1, wherein the promoter is a phosphoglycerate kinase (PGK) promoter.

3. The expression cassette of claim 1 or claim 2, wherein the codon-optimized sequence encoding a human *PKLR* gene product comprises a sequence having at least 99% identity to SEQ ID NO: 8.

4. The expression cassette of claim 1, wherein the mutated WPRE comprises a sequence having at least 95% identity to SEQ ID NO:24.

5. The expression cassette of claim 4, wherein the mutated WPRE comprises a sequence having at least 99% identity to SEQ ID NO:24.

6. The expression cassette of any of claims 1-5, further comprising a polyadenylation (polyA) signal sequence.

7. The expression cassette of any of claims 1-6, wherein the expression cassette shares at least 95% identity to SEQ ID NO: 26.

8. The expression cassette of claim 7, wherein the expression cassette shares at least 99% identity to SEQ ID NO: 26.

9. A recombinant gene delivery vector comprising the expression cassette of any of claims 1-8, wherein the recombinant gene delivery vector is a lentiviral vector.

10. The recombinant gene delivery vector of claim 9, comprising the following sequences in 5' to 3' order:

a) a CMV enhancer sequence;

b) a CMV promoter sequence;

c) a 5' LTR, optionally a modified 5' LTR;

d) an HIV-1 packaging signal sequence (ψ or Psi) sequence;

e) a truncated Gag sequence;

f) an RRE;

g) a cPPT/CTS sequence;

h) a PGK promoter sequence, optionally a human PGK promoter sequence;

i) a codon-optimized sequence encoding a human *PKLR* gene product having at least 95% identity to SEQ ID NO: 8;

j) a mutant WPRE sequence having at least 90% identity to SEQ ID NO:24;

k) a modified 3' LTR;

l) an SV40 poly(A) signal;

m) an SV40 origin of replication (ori).

11. The recombinant gene delivery vector of claim 9 or 10, wherein the codon-optimized sequence encoding a human *PKLR* gene product comprises a sequence having at

least 99% identity to SEQ ID NO: 8.

12. The recombinant gene delivery vector of any one of claims 9-11, wherein the mutated WPRE comprises a sequence having at least 95% identity to SEQ ID NO:24.

13. The recombinant gene delivery vector of claim 12, wherein the mutated WPRE comprises a sequence having at least 99% identity to SEQ ID NO:24.

14. The recombinant gene delivery vector of any one of claims 9-13, wherein the recombinant gene delivery vector is a therapeutic gene delivery vector.

15. The recombinant gene delivery vector of claim 14, wherein the recombinant gene delivery vector is capable of long-term stability and safety in clinical use.

16. A cell comprising the expression cassette of any of claims 1-8 or transduced by the recombinant gene delivery vector of any of claims 9-15.

17. The cell of claim 16, wherein the cell is a hematopoietic stem cell.

18. The cell of claim 17, wherein the cell is a committed hematopoietic erythroid progenitor cell.

19. A pharmaceutical composition comprising a pharmaceutically acceptable excipient and the recombinant gene delivery vector of any of claims 9-15 or the cell of any of claims 16-18.

20. Use of the pharmaceutical composition of claim 19 in the manufacture of a medicament for treating or preventing Pyruvate Kinase Deficiency (PKD) in a subject in need thereof.

21. The use of claim 20, wherein the pharmaceutical composition comprises the recombinant gene delivery vector.

22. The use of claim 20, wherein the pharmaceutical composition comprises the cell.

23. The use of claim 22, wherein the cell is autologous to the subject.

24. The use of claim 22, wherein the cell is allogeneic to the subject.

25. A *in vitro* or *ex vivo* method for expressing a human *PKLR* gene in erythroid cells or erythroid progenitor cells, comprising contacting one or more erythroid cells or erythroid progenitor cells with an effective amount of a recombinant viral vector, wherein the vector comprises a human phosphoglycerate kinase promoter, a codon-optimized *PKLR* transgene, and a mutated post-transcriptional regulatory element of the woodchuck hepatitis virus (WPRE), wherein the codon-optimized *PKLR* transgene shares at least 95% identity with SEQ ID NO: 8, wherein the mutated WPRE is a chimeric WPRE comprising a sequence having at least 90% identity to SEQ ID NO: 24, and wherein following said contacting, *PKLR* is expressed at detectable levels in the one or more erythroid cells or erythroid progenitor cells.

26. The method of claim 25, wherein the codon-optimized sequence encoding a human *PKLR* gene product comprises a sequence having at least 99% identity to SEQ ID NO: 8.

27. The method of claim 25 or 26, wherein the mutated WPRE comprises a sequence having at least 95% identity to SEQ ID NO:24.

28. The method of claim 27, wherein the mutated WPRE comprises a sequence having at least 99% identity to SEQ ID NO:24.