



US 20190098879A1

(19) **United States**

(12) **Patent Application Publication**
Drummond-Samuels et al.

(10) **Pub. No.: US 2019/0098879 A1**
(43) **Pub. Date: Apr. 4, 2019**

(54) **NON-HUMAN ANIMALS COMPRISING A HUMANIZED TTR LOCUS AND METHODS OF USE**

(71) Applicant: **Regeneron Pharmaceuticals, Inc.**, Tarrytown, NY (US)

(72) Inventors: **Meghan Drummond-Samuels**, Katonah, NY (US); **Jeffery Haines**, New York, NY (US); **Suzanne Hartford**, Putnam Valley, NY (US); **David Frenthewey**, New York, NY (US); **Brian Zambrowicz**, Sleepy Hollow, NY (US); **Andrew J. Murphy**, Croton-on-Hudson, NY (US)

(21) Appl. No.: **16/145,859**

(22) Filed: **Sep. 28, 2018**

Related U.S. Application Data

(60) Provisional application No. 62/720,292, filed on Aug. 21, 2018, provisional application No. 62/679,142, filed on Jun. 1, 2018, provisional application No. 62/565,980, filed on Sep. 29, 2017.

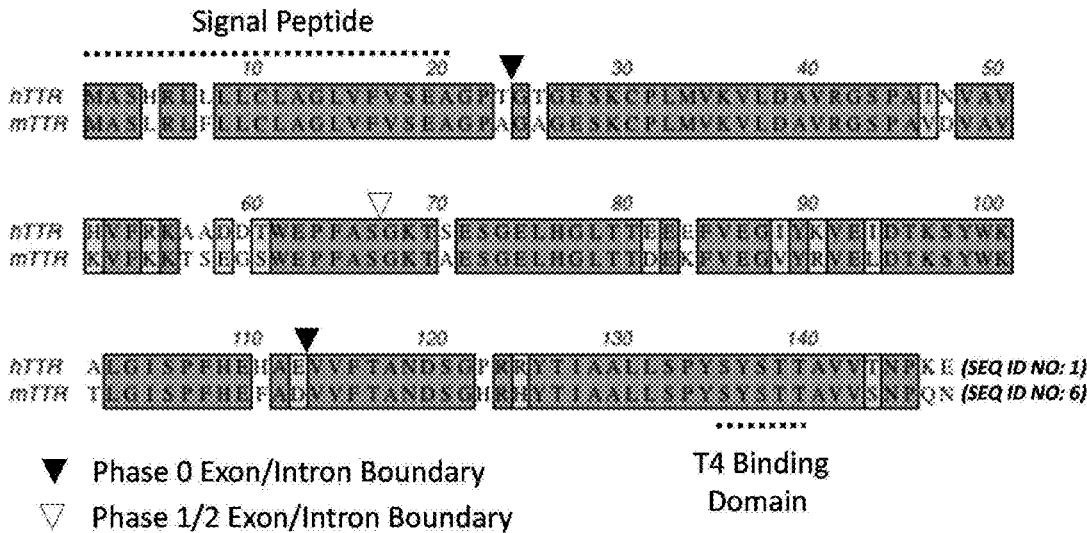
Publication Classification

(51) **Int. Cl.**
A01K 67/027 (2006.01)
C12N 15/86 (2006.01)
(52) **U.S. Cl.**
CPC *A01K 67/0275* (2013.01); *C12N 15/86* (2013.01); *A01K 2217/072* (2013.01); *A01K 2227/105* (2013.01); *A01K 2217/058* (2013.01); *C12N 2710/10041* (2013.01)

(57) **ABSTRACT**

Non-human animal genomes, non-human animal cells, and non-human animals comprising a humanized TTR locus and methods of using such non-human animal genomes, non-human animal cells, and non-human animals are provided. Non-human animal cells or non-human animals comprising a humanized TTR locus express a human transthyretin protein or a chimeric transthyretin protein, fragments of which are from human transthyretin. Methods are provided for using such non-human animals comprising a humanized TTR locus to assess in vivo efficacy of human-TTR-targeting reagents such as nuclease agents designed to target human TTR. Methods are also provided for making such non-human animals comprising a humanized TTR locus.

Specification includes a Sequence Listing.



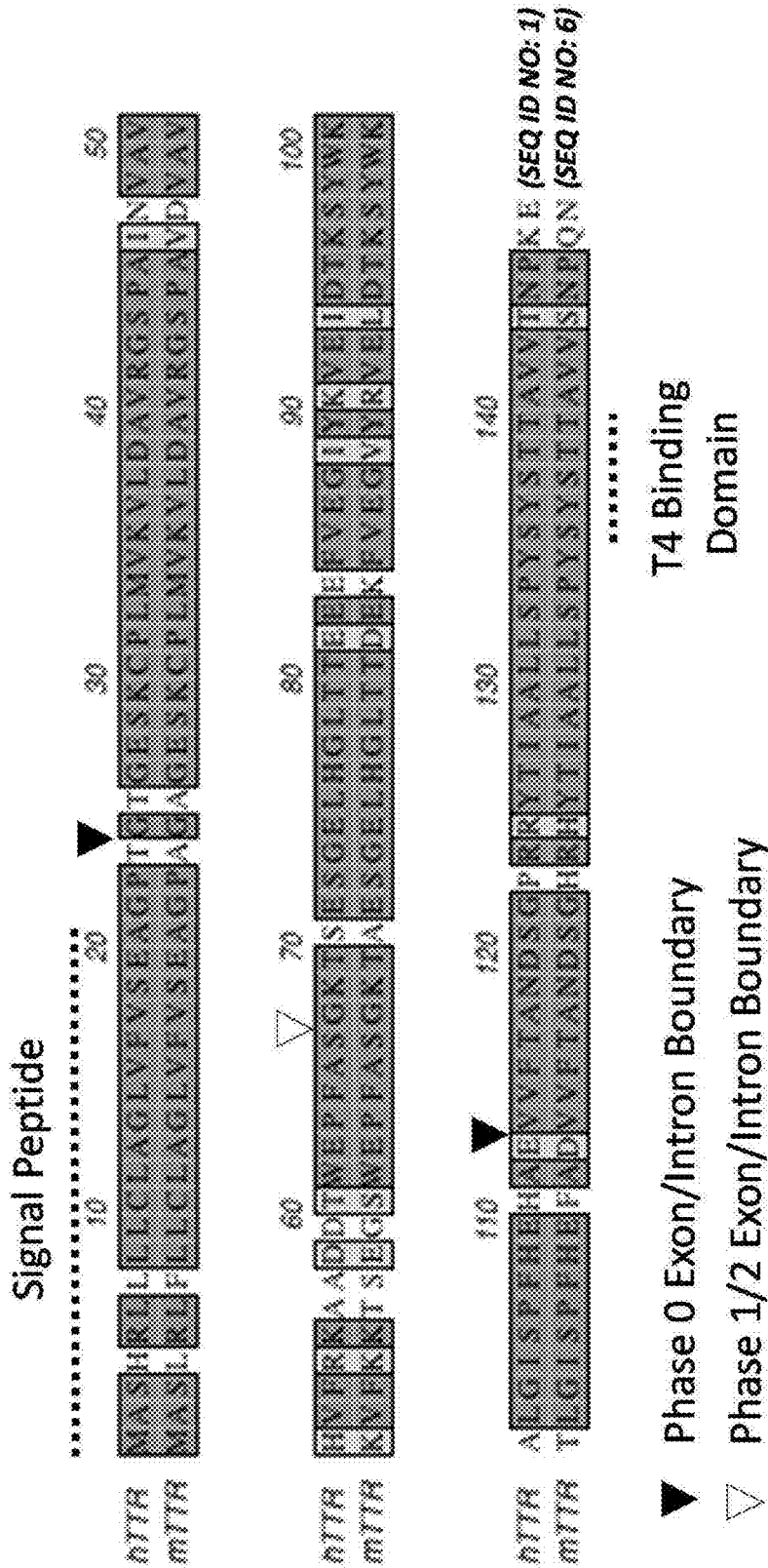


FIG. 1A

m = mTTR CDS
h = hTTR CDS

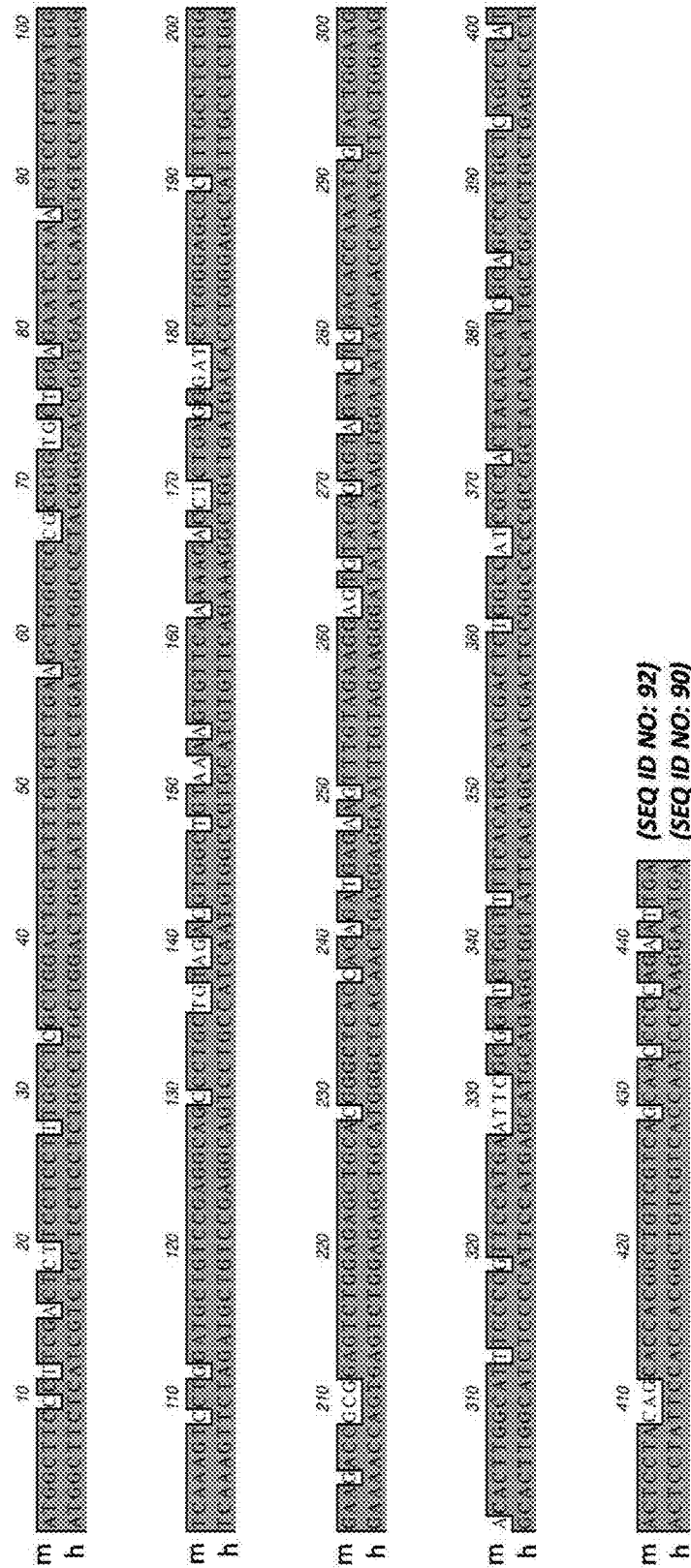


FIG. 1B

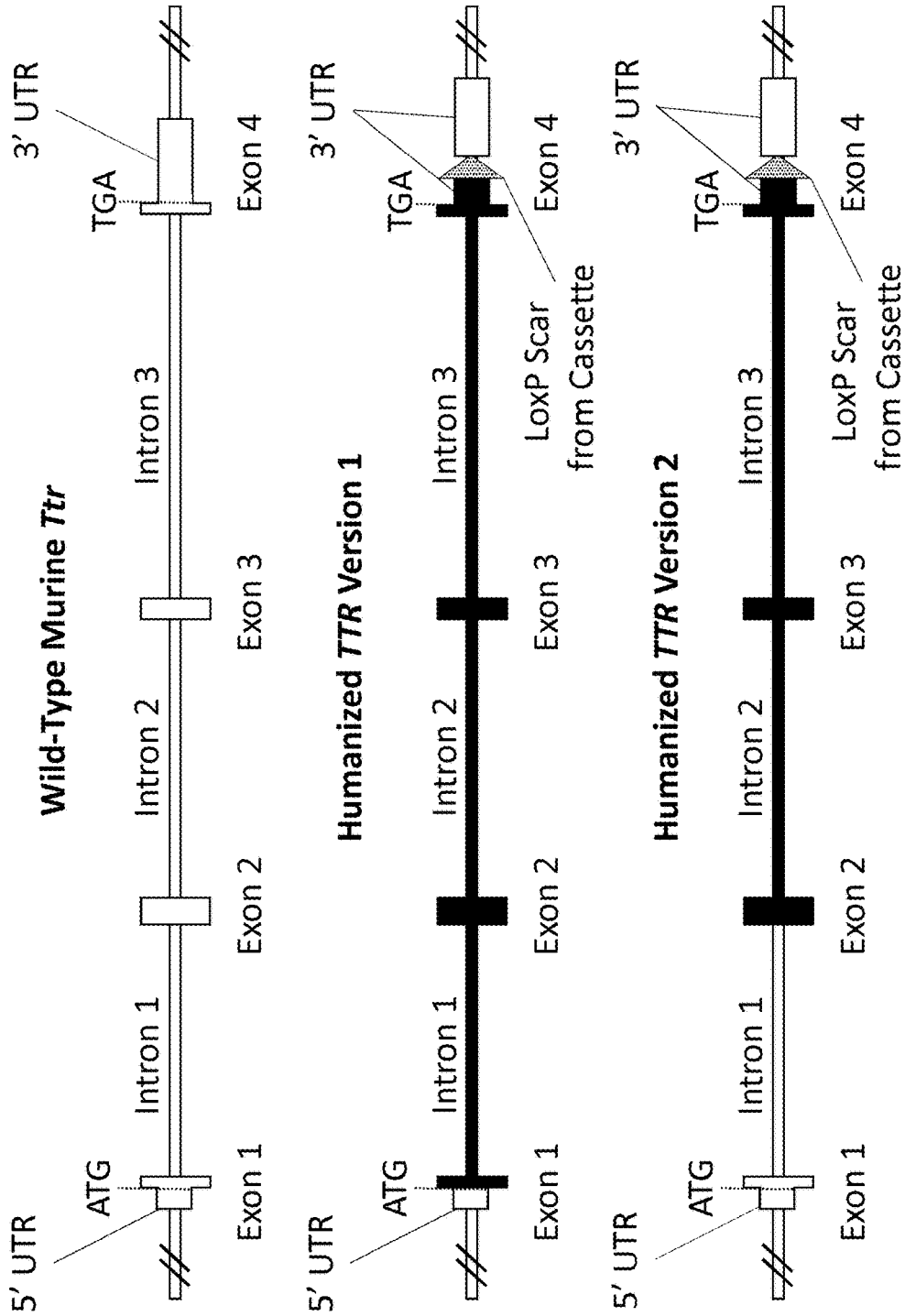


FIG. 2

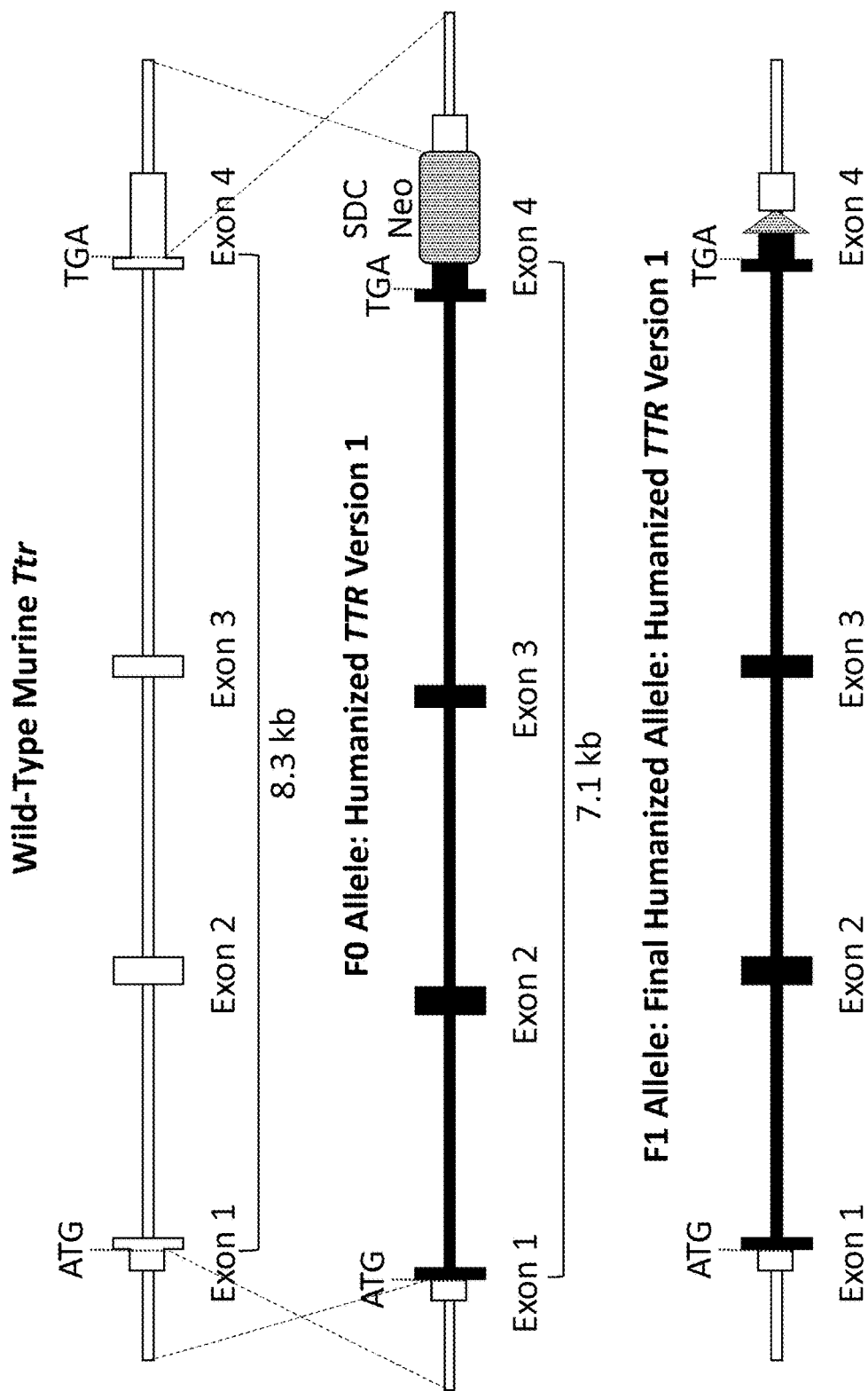


FIG. 3

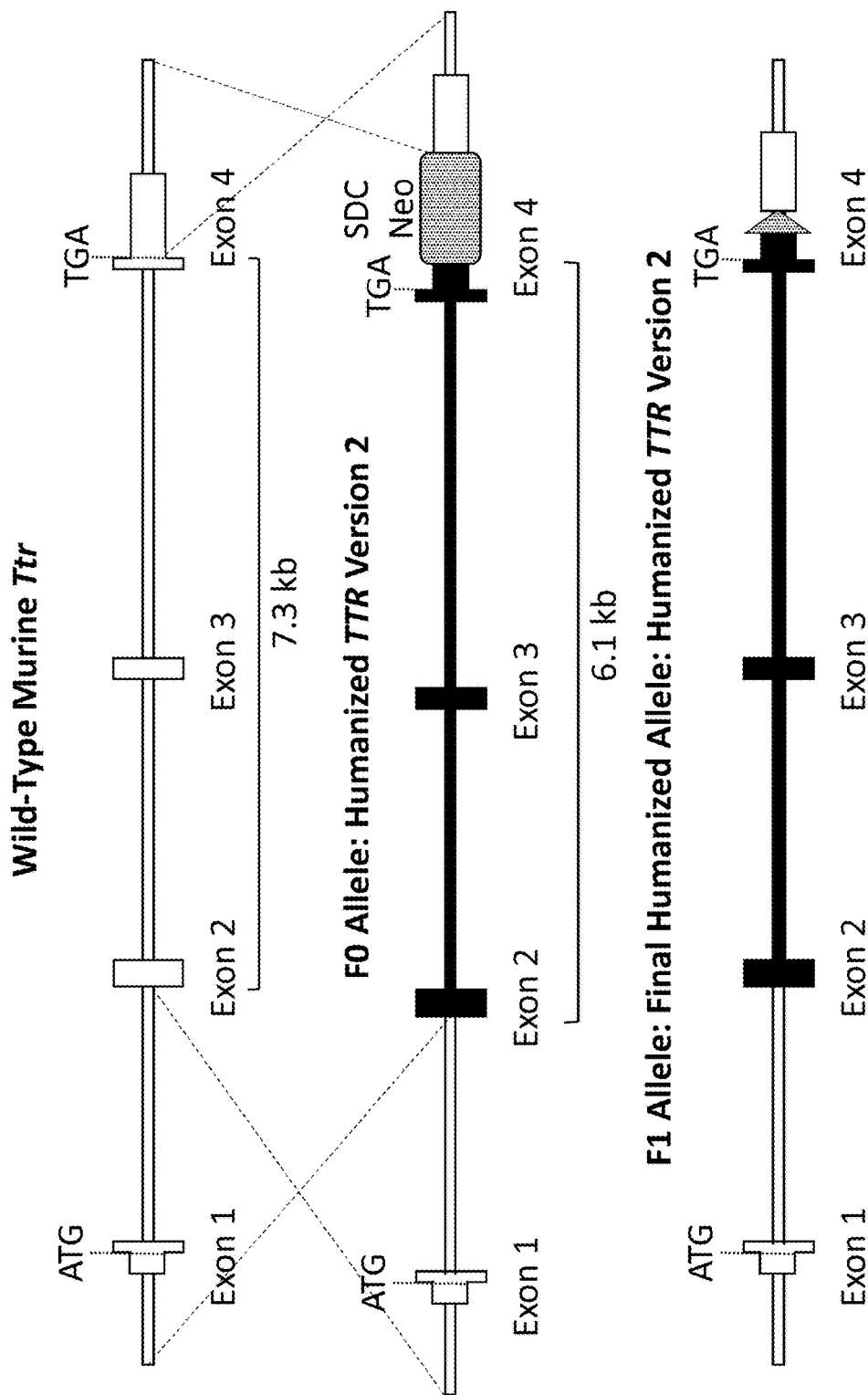


FIG. 4

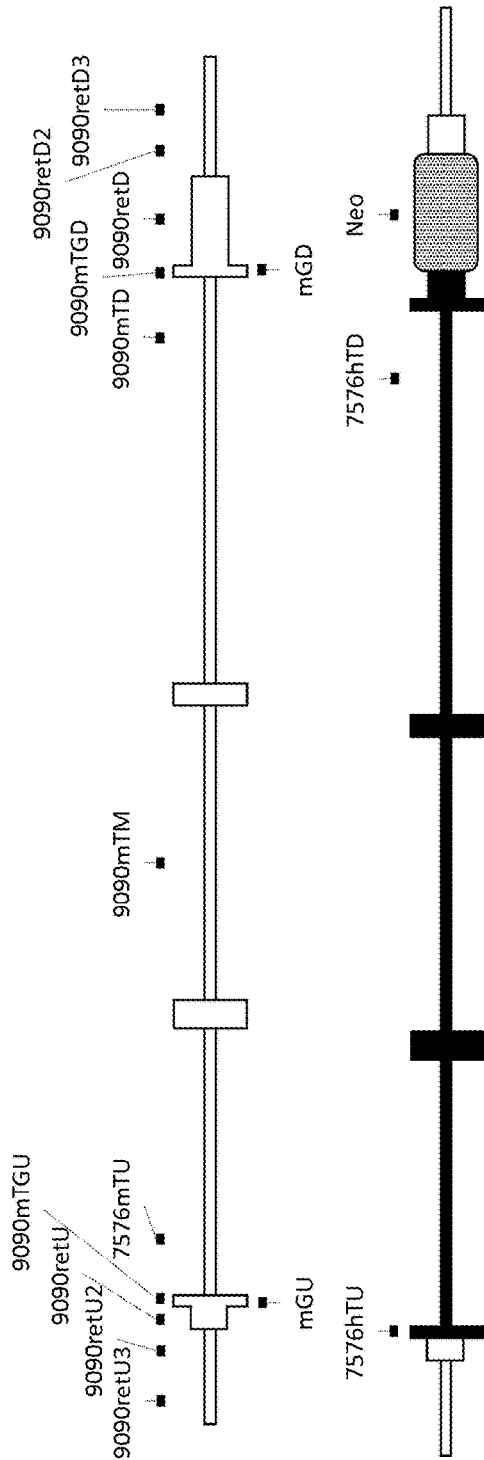


FIG. 5A

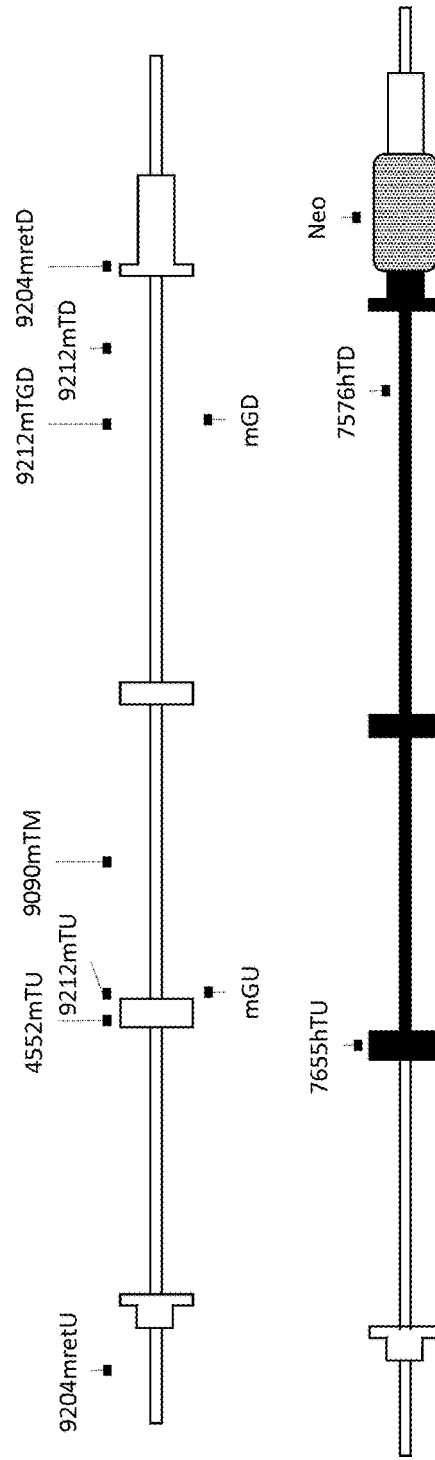


FIG. 5B

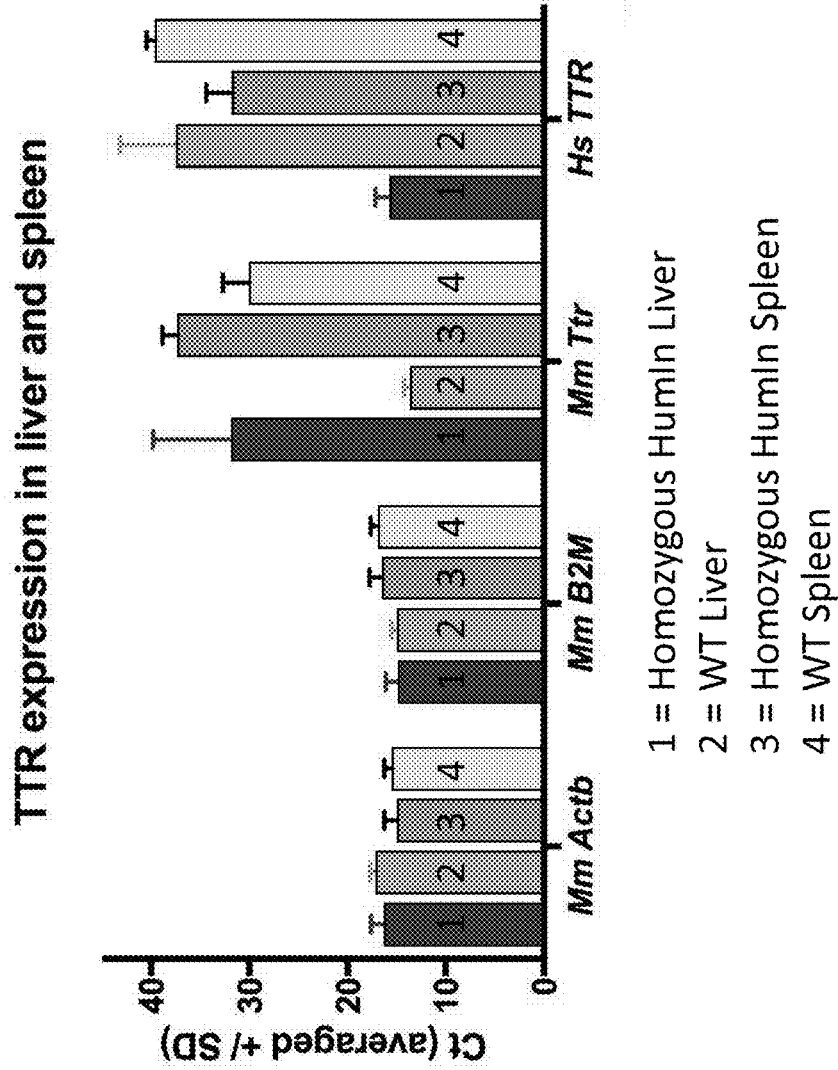


FIG. 6

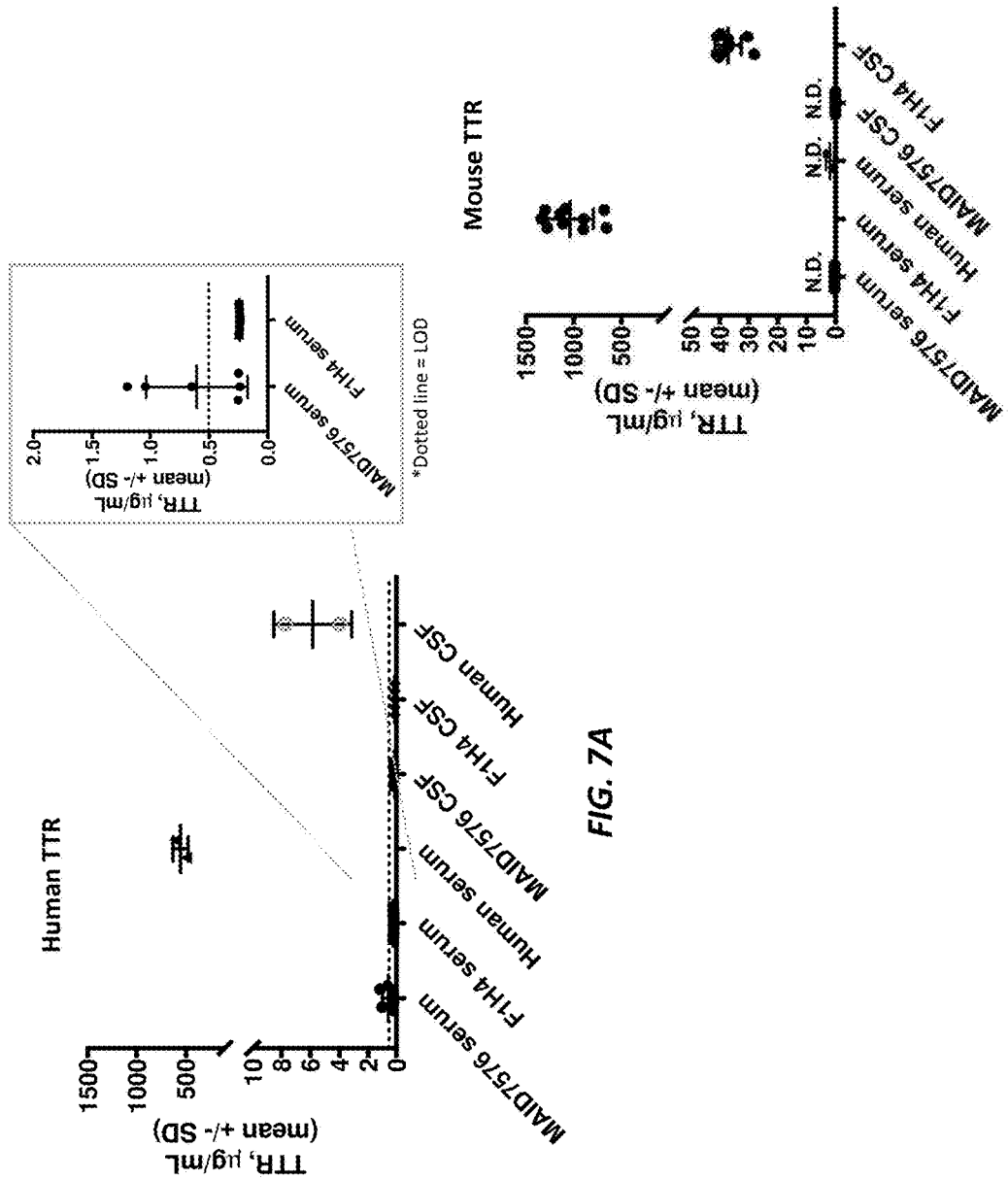


FIG. 7B

FIG. 7A

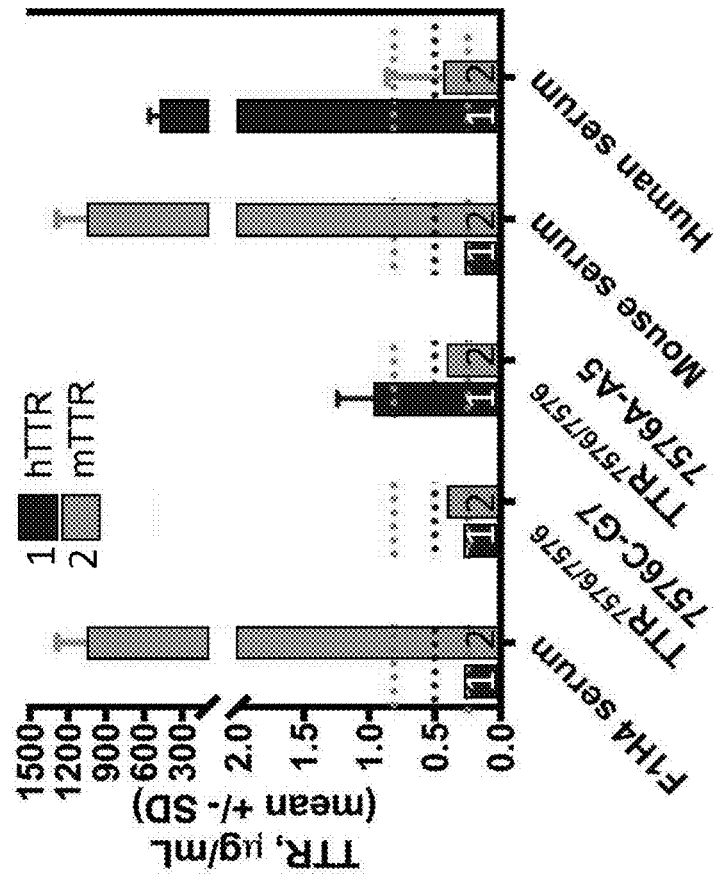


FIG.7C

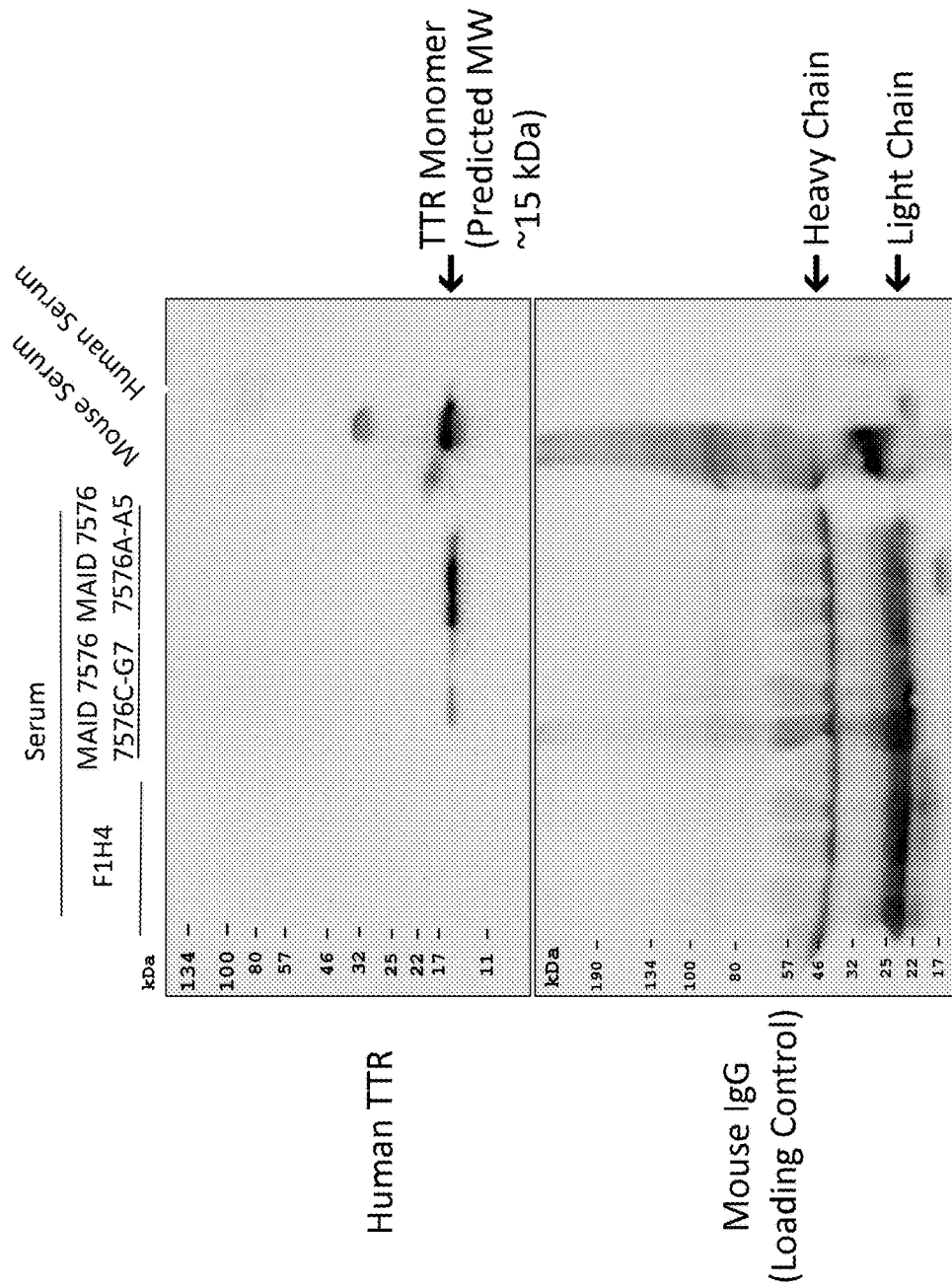


FIG. 8

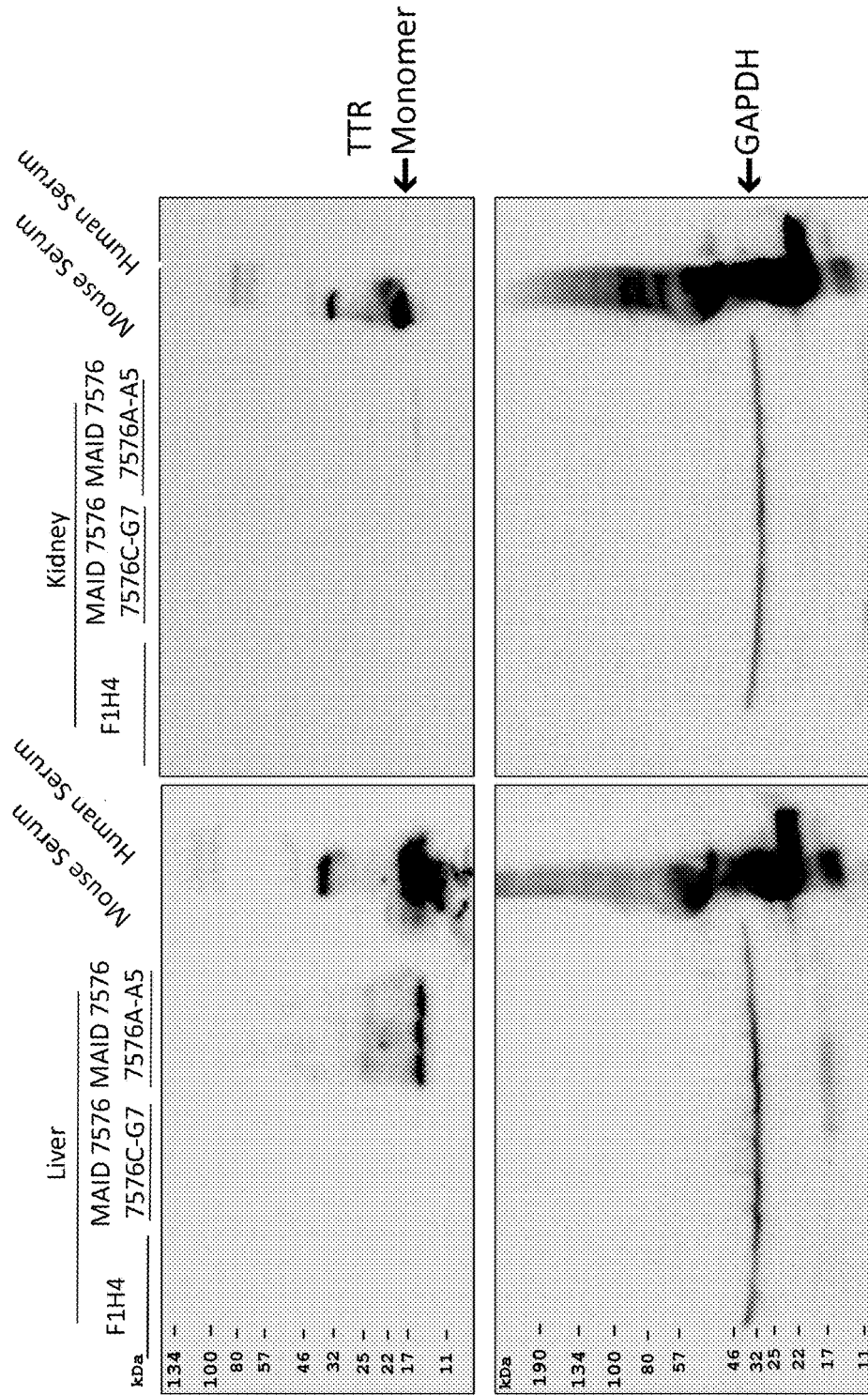


FIG. 9

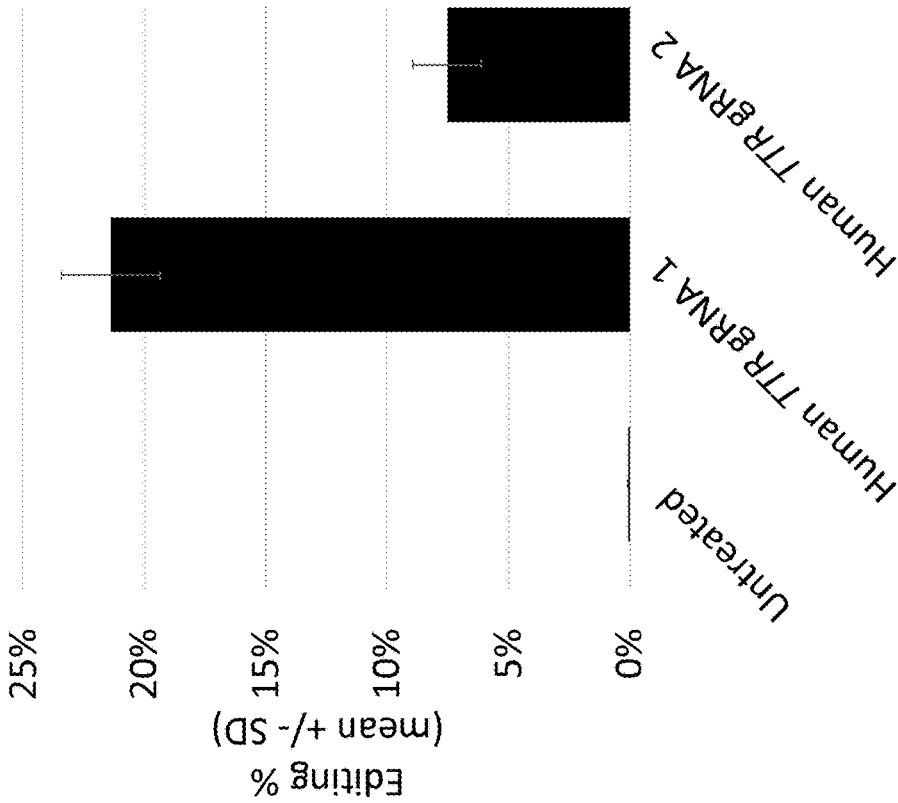


FIG. 10

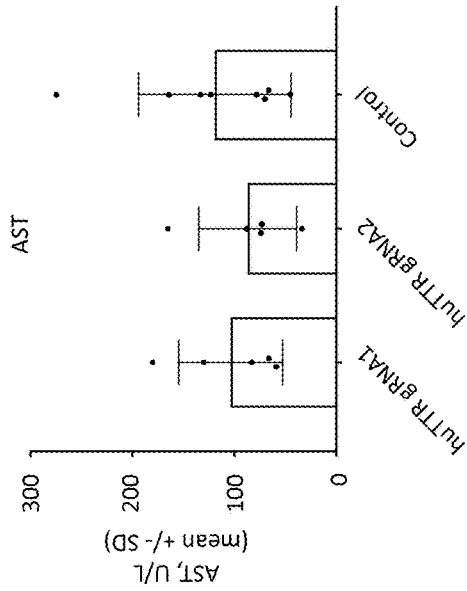


FIG. 11B

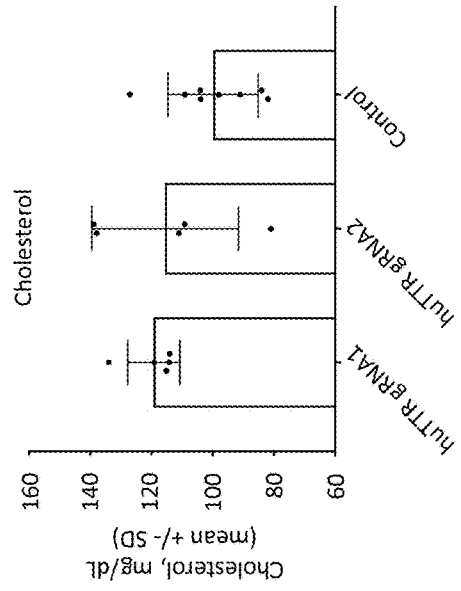


FIG. 11D

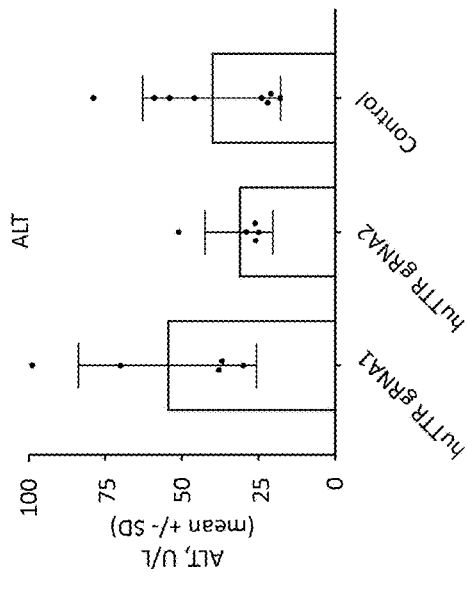


FIG. 11A

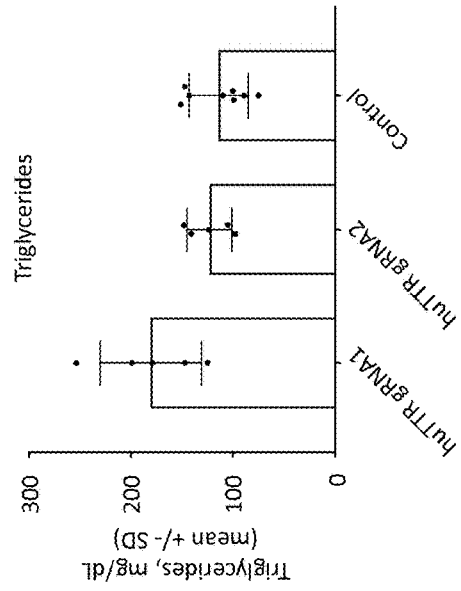


FIG. 11C

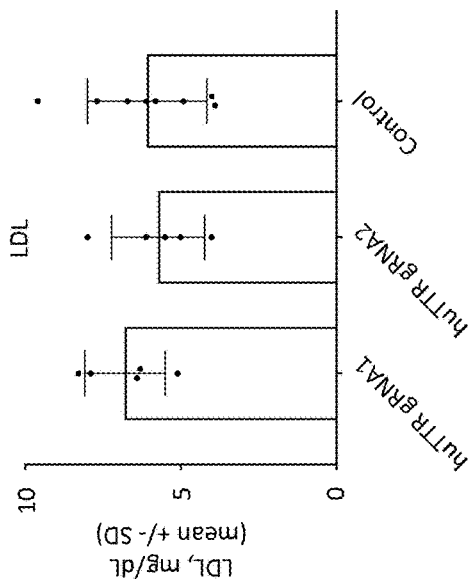


FIG. 11F

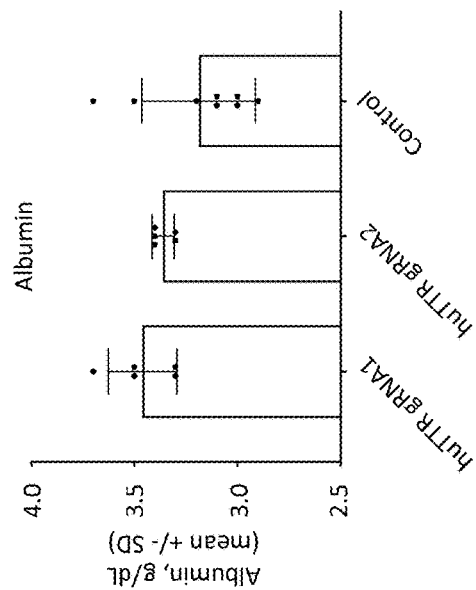


FIG. 11H

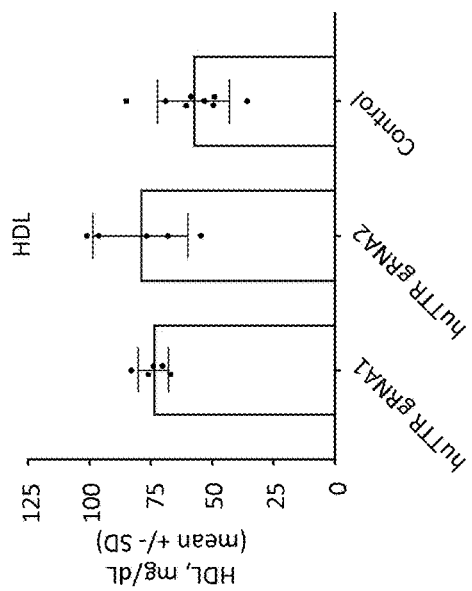


FIG. 11E

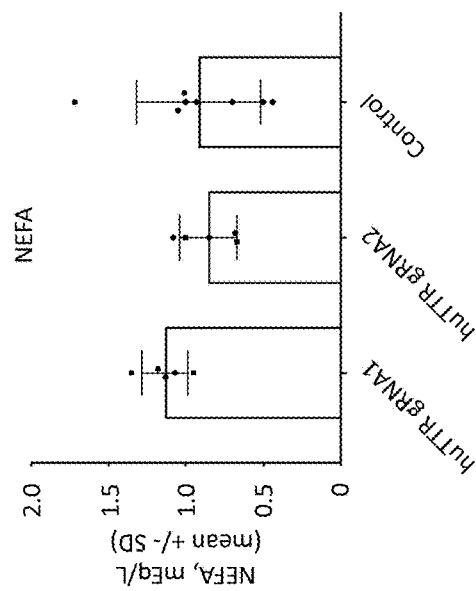


FIG. 11G

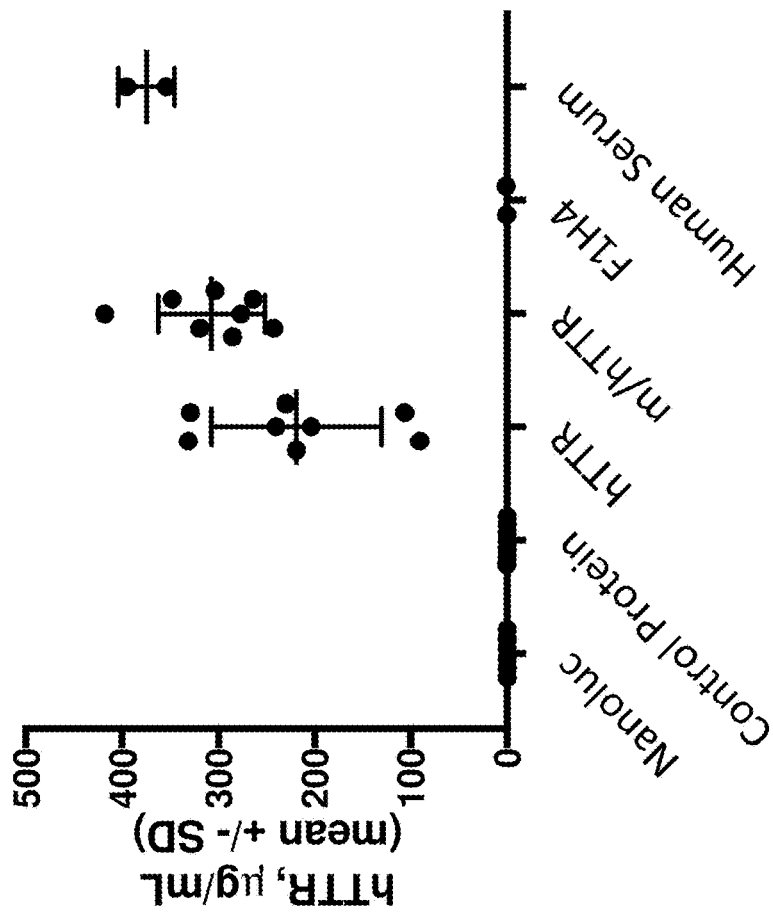


FIG. 13

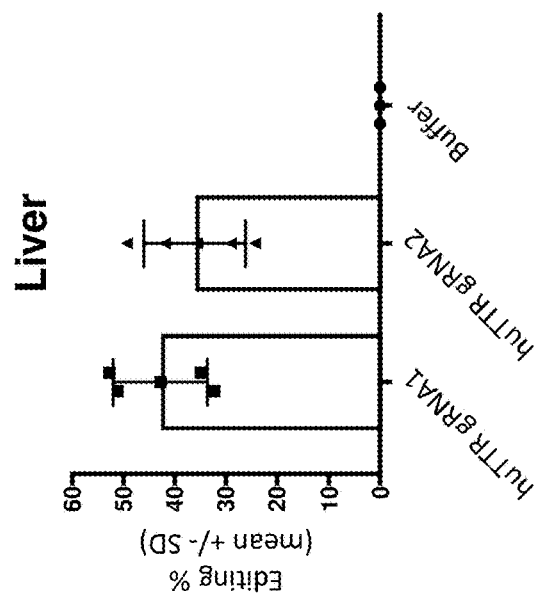


FIG. 12

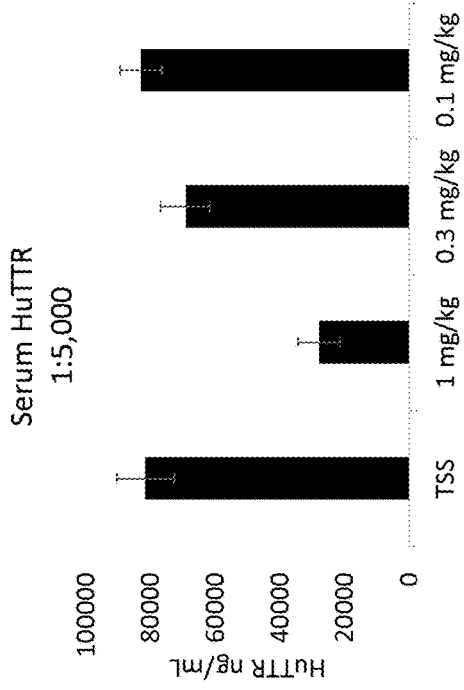


FIG. 15A

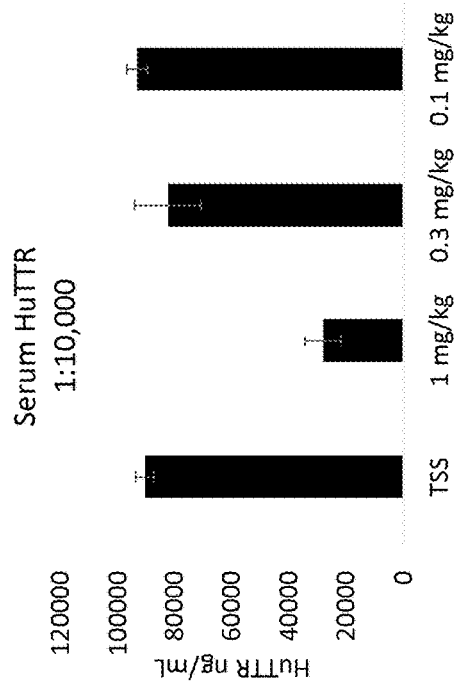


FIG. 15B

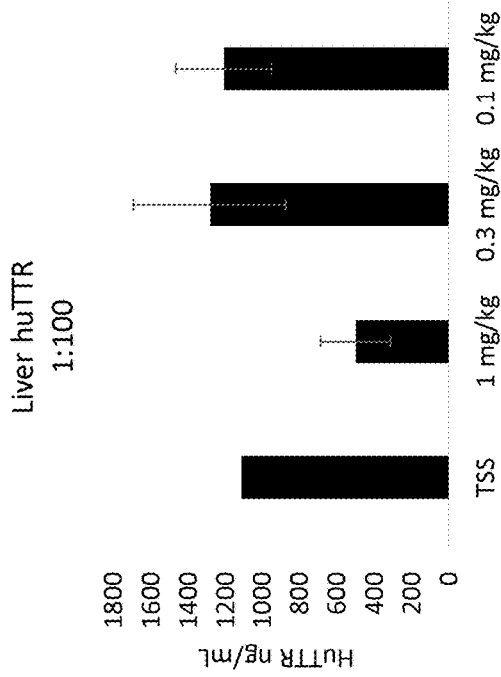


FIG. 14

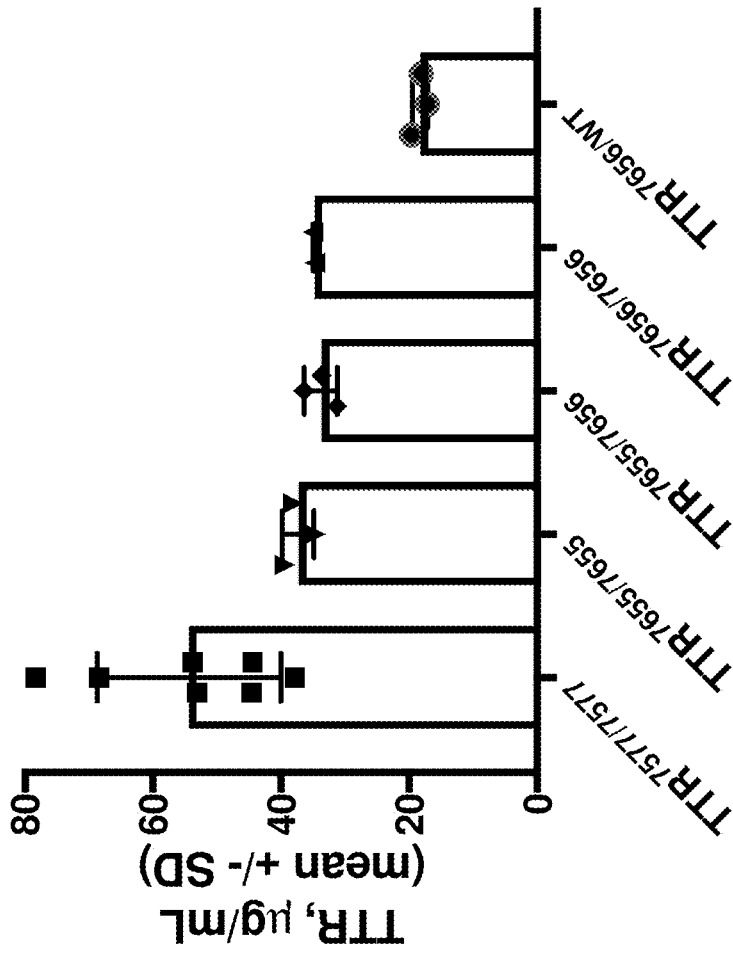


FIG. 16

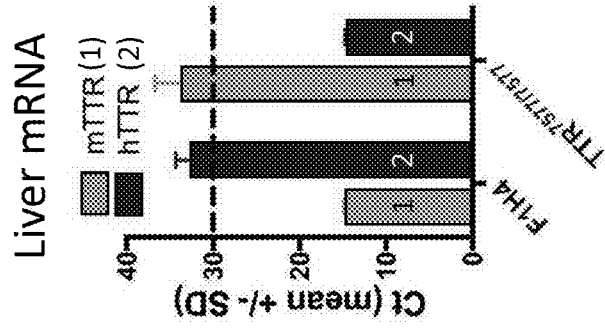


FIG. 17C

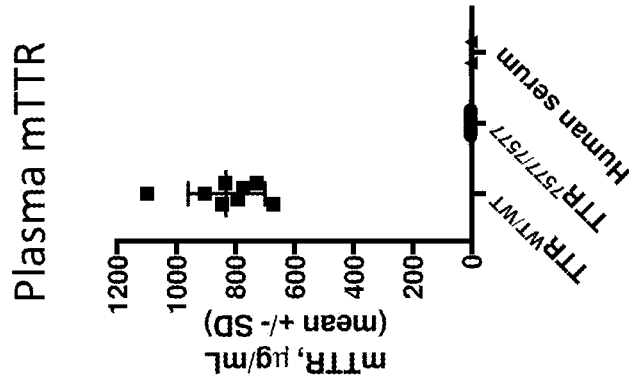


FIG. 17B

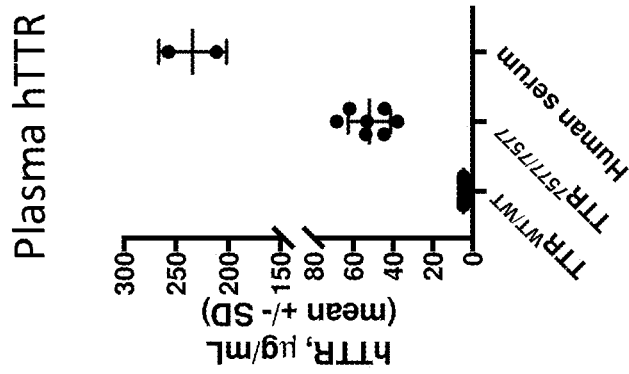


FIG. 17A

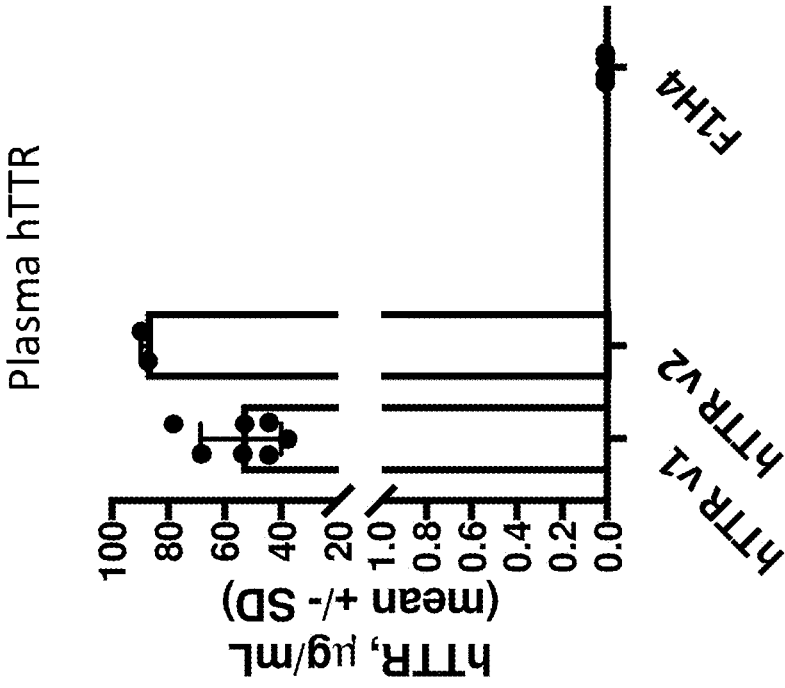


FIG. 18

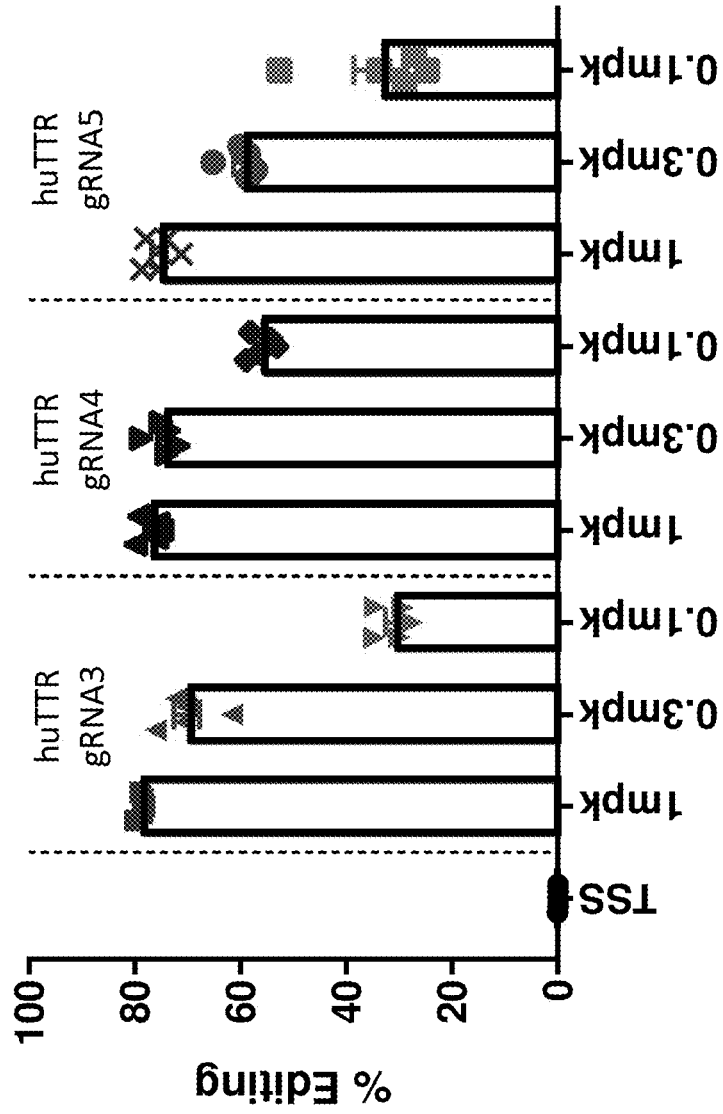


FIG. 19

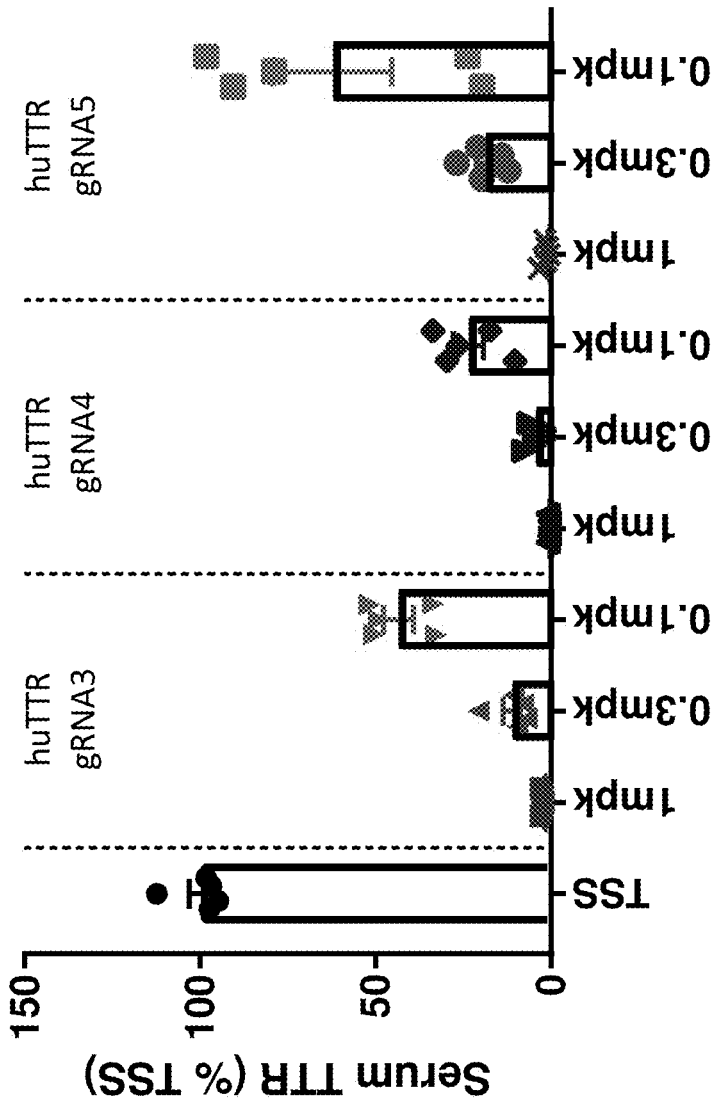


FIG. 20

**NON-HUMAN ANIMALS COMPRISING A
HUMANIZED TTR LOCUS AND METHODS
OF USE**

CROSS-REFERENCE TO RELATED
APPLICATIONS

[0001] This application claims the benefit of U.S. Application No. 62/565,980 filed Sep. 29, 2017, U.S. Application No. 62/679,142 filed Jun. 1, 2018, and U.S. Application No. 62/720,292 filed Aug. 21, 2018, each of which is herein incorporated by reference in its entirety for all purposes.

REFERENCE TO A SEQUENCE LISTING
SUBMITTED AS A TEXT FILE VIA EFS WEB

[0002] The Sequence Listing written in file 519829SEQLIST.txt is 139 kilobytes, was created on Sep. 28, 2018, and is hereby incorporated by reference.

BACKGROUND

[0003] Transthyretin (TTR) is a protein found in the serum and cerebrospinal fluid that carries thyroid hormone and retinol-binding protein to retinol. The liver secretes TTR into the blood, while the choroid plexus secretes it into the cerebrospinal fluid. TTR is also produced in the retinal pigmented epithelium and secreted into the vitreous. Misfolded and aggregated TTR accumulates in multiple tissues and organs in the amyloid diseases senile systemic amyloidosis (SSA), familial amyloid polyneuropathy (FAP), and familial amyloid cardiomyopathy (FAC).

[0004] One promising therapeutic approach for the TTR amyloidosis diseases is to reduce the TTR load in the patient. However, there remains a need for suitable non-human animals providing the true human target or a close approximation of the true human target of human-TTR-targeting reagents at the endogenous Ttr locus, thereby enabling testing of the efficacy and mode of action of such agents in live animals as well as pharmacokinetic and pharmacodynamics studies in a setting where the humanized protein and humanized gene are the only version of TTR present.

SUMMARY

[0005] Non-human animals comprising a humanized TTR locus are provided, as well as methods of using such non-human animals. Non-human animal genomes or cells comprising a humanized TTR locus are also provided.

[0006] In one aspect, provided are non-human animal genomes, non-human animal cells, or non-human animals comprising a humanized TTR locus. Such a non-human animal genome, non-human animal cell, or non-human animal can comprise in its genome a genetically modified endogenous Ttr locus comprising a human TTR sequence comprising both TTR coding sequence and non-coding sequence. Some such non-human animal genomes, non-human animal cells, or non-human animals can comprise a genetically modified endogenous Ttr locus, wherein a region of the endogenous Ttr locus comprising both Ttr coding sequence and non-coding sequence has been deleted and replaced with a corresponding human TTR sequence comprising both TTR coding sequence and non-coding sequence. Optionally, the genetically modified endogenous Ttr locus comprises the endogenous Ttr promoter. Optionally, the human TTR sequence is operably linked to the endogenous Ttr promoter. Optionally, at least one intron and

at least one exon of the endogenous Ttr locus have been deleted and replaced with the corresponding human TTR sequence.

[0007] In some such non-human animal genomes, non-human animal cells, or non-human animals, the entire Ttr coding sequence of the endogenous Ttr locus has been deleted and replaced with the corresponding human TTR sequence. Optionally, the region of the endogenous Ttr locus from the Ttr start codon to the Ttr stop codon has been deleted and replaced with the corresponding human TTR sequence.

[0008] In some such non-human animal genomes, non-human animal cells, or non-human animals, the genetically modified endogenous Ttr locus comprises a human TTR 3' untranslated region. In some such non-human animals, the endogenous Ttr 5' untranslated region has not been deleted and replaced with the corresponding human TTR sequence.

[0009] In some such non-human animal genomes, non-human animal cells, or non-human animals, the region of the endogenous Ttr locus from the Ttr start codon to the Ttr stop codon has been deleted and replaced with a human TTR sequence comprising the corresponding human TTR sequence and a human TTR 3' untranslated region, and the endogenous Ttr 5' untranslated region has not been deleted and replaced with the corresponding human TTR sequence, and the endogenous Ttr promoter has not been deleted and replaced with the corresponding human TTR sequence. Optionally, the human TTR sequence at the genetically modified endogenous Ttr locus comprises, consists essentially of, or consists of a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 18. Optionally, the genetically modified endogenous Ttr locus encodes a protein comprising, consisting essentially of, or consisting of a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 1. Optionally, the genetically modified endogenous Ttr locus comprises a coding sequence comprising, consisting essentially of, or consisting of a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 90. Optionally, the genetically modified endogenous Ttr locus comprises, consists essentially of, or consists of a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 14 or 15.

[0010] In some such non-human animal genomes, non-human animal cells, or non-human animals, the genetically modified endogenous Ttr locus encodes a transthyretin precursor protein comprising a signal peptide, and the region of the endogenous Ttr locus encoding the signal peptide has not been deleted and replaced with the corresponding human TTR sequence. Optionally, the first exon of the endogenous Ttr locus has not been deleted and replaced with the corresponding human TTR sequence. Optionally, the first exon and first intron of the endogenous Ttr locus have not been deleted and replaced with the corresponding human TTR sequence. Optionally, the region of the endogenous Ttr locus from the start of the second Ttr exon to the Ttr stop codon has been deleted and replaced with the corresponding human TTR sequence. Optionally, the genetically modified endogenous Ttr locus comprises a human TTR 3' untranslated region.

[0011] In some such non-human animal genomes, non-human animal cells, or non-human animals, the region of the endogenous Ttr locus from the second Ttr exon to the Ttr

stop codon has been deleted and replaced with a human TTR sequence comprising the corresponding human TTR sequence and a human TTR 3' untranslated region, and the endogenous Ttr 5' untranslated region has not been deleted and replaced with the corresponding human TTR sequence, and the endogenous Ttr promoter has not been deleted and replaced with the corresponding human TTR sequence. Optionally, the human TTR sequence at the genetically modified endogenous Ttr locus comprises, consists essentially of, or consists of a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 19. Optionally, the genetically modified endogenous Ttr locus encodes a protein comprising, consisting essentially of, or consisting of a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 2. Optionally, the genetically modified endogenous Ttr locus comprises a coding sequence comprising, consisting essentially of, or consisting of a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 91. Optionally, the genetically modified endogenous Ttr locus comprises, consists essentially of, or consists of a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 16 or 17.

[0012] In some such non-human animal genomes, non-human animal cells, or non-human animals, the genetically modified endogenous Ttr locus does not comprise a selection cassette or a reporter gene. In some such non-human animal genomes, non-human animal cells, or non-human animals, the genetically modified endogenous Ttr locus does comprise a selection cassette or a reporter gene. In some such non-human animal genomes, non-human animal cells, or non-human animals, the non-human animal genome, non-human animal cell, or non-human animal is homozygous for the genetically modified endogenous Ttr locus. In some such non-human animal genomes, non-human animal cells, or non-human animals, the non-human animal genome, non-human animal cell, or non-human animal is heterozygous for the genetically modified endogenous Ttr locus.

[0013] In some such non-human animal genomes, non-human animal cells, or non-human animals, the non-human animal is a mammal. Optionally, the mammal is a rodent. Optionally, the rodent is a rat or mouse. Optionally, the non-human animal is a mouse.

[0014] In another aspect, provided are methods of using the non-human animals comprising a humanized TTR locus to assess the activity of human-TTR-targeting reagents in vivo. Such methods can comprise: (a) administering the human-TTR-targeting reagent to any of the above non-human animals; and (b) assessing the activity of the human-TTR-targeting reagent in the non-human animal.

[0015] In some such methods, the administering comprises adeno-associated virus (AAV)-mediated delivery, lipid nanoparticle (LNP)-mediated delivery, or hydrodynamic delivery (HDD). Optionally, the administering comprises LNP-mediated delivery, and optionally the LNP dose is between about 0.1 mg/kg and about 2 mg/kg. Optionally, the administering comprises AAV8-mediated delivery.

[0016] In some such methods, step (b) comprises isolating a liver from the non-human animal and assessing activity of the human-TTR-targeting reagent in the liver. Optionally,

step (b) further comprises assessing activity of the human-TTR-targeting reagent in an organ or tissue other than the liver.

[0017] In some such methods, the human-TTR-targeting reagent is a genome-editing agent, and the assessing comprises assessing modification of the genetically modified Ttr locus. Optionally, the assessing comprises measuring the frequency of insertions or deletions within the genetically modified Ttr locus. In some such methods, the assessing comprises measuring expression of a Ttr messenger RNA encoded by the genetically modified Ttr locus. In some such methods, the assessing comprises measuring expression of a TTR protein encoded by the genetically modified Ttr locus. Optionally, measuring expression of the TTR protein comprises measuring serum levels of the TTR protein in the non-human animal. Optionally, the activity is assessed in the liver of the non-human animal.

[0018] In some such methods, the human-TTR-targeting reagent comprises a nuclease agent designed to target a region of a human TTR gene. Optionally, the nuclease agent comprises a Cas protein and a guide RNA designed to target a guide RNA target sequence in the human TTR gene. Optionally, the Cas protein is a Cas9 protein. Optionally, the human-TTR-targeting reagent further comprises an exogenous donor nucleic acid, wherein the exogenous donor nucleic acid is designed to recombine with the human TTR gene. Optionally, the exogenous donor nucleic acid is a single-stranded oligodeoxynucleotide (ssODN).

[0019] In another aspect, provided are methods of optimizing the activity of a human-TTR-targeting reagent in vivo. Such methods can comprise: (I) performing any of the above methods of assessing the activity of human-TTR-targeting reagents in vivo a first time in a first non-human animal comprising in its genome a genetically modified endogenous Ttr locus comprising a human TTR sequence comprising both TTR coding sequence and non-coding sequence; (II) changing a variable and performing the method of step (I) a second time with the changed variable in a second non-human animal comprising in its genome the genetically modified endogenous Ttr locus comprising the human TTR sequence comprising both TTR coding sequence and non-coding sequence; and (III) comparing the activity of the human-TTR-targeting reagent in step (I) with the activity of the human-TTR-targeting reagent in step (II), and selecting the method resulting in the higher activity. Optionally, step (III) can comprise selecting the method resulting in the higher efficacy, higher precision, higher consistency, or higher specificity.

[0020] Optionally, the changed variable in step (II) is the delivery method of introducing the human-TTR-targeting reagent into the non-human animal. Optionally, the administering comprises LNP-mediated delivery, and the changed variable in step (II) is the LNP formulation. Optionally, the changed variable in step (II) is the route of administration of introducing the human-TTR-targeting reagent into the non-human animal. Optionally, the changed variable in step (II) is the concentration or amount of the human-TTR-targeting reagent introduced into the non-human animal. Optionally, the changed variable in step (II) is the form of the human-TTR-targeting reagent introduced into the non-human animal. Optionally, the changed variable in step (II) is the human-TTR-targeting reagent introduced into the non-human animal.

[0021] In some such methods, the human-TTR-targeting reagent comprises a Cas protein (e.g., a Cas9 protein) and a guide RNA designed to target a guide RNA target sequence in the human TTR gene. Optionally, the changed variable in step (II) is the guide RNA sequence or the guide RNA target sequence. Optionally, the Cas protein and the guide RNA are each administered in the form of RNA, and the changed variable in step (II) is the ratio of Cas mRNA to guide RNA. Optionally, the changed variable in step (II) is guide RNA modifications.

[0022] In another aspect, provided are methods of making the non-human animals comprising a humanized TTR locus. Some such methods comprise: (a) introducing into a non-human animal embryonic stem (ES) cell: (i) a nuclease agent that targets a target sequence in the endogenous Ttr locus; and (ii) a targeting vector comprising a nucleic acid insert comprising the human TTR sequence flanked by a 5' homology arm corresponding to a 5' target sequence in the endogenous Ttr locus and a 3' homology arm corresponding to a 3' target sequence in the endogenous Ttr locus, wherein the targeting vector recombines with the endogenous Ttr locus to produce a genetically modified non-human ES cell comprising in its genome the genetically modified endogenous Ttr locus comprising the human TTR sequence; (b) introducing the genetically modified non-human ES cell into a non-human animal host embryo; and (c) gestating the non-human animal host embryo in a surrogate mother, wherein the surrogate mother produces an F0 progeny genetically modified non-human animal comprising in its genome the genetically modified endogenous Ttr locus comprising the human TTR sequence.

[0023] In some such methods, the nuclease agent comprises a Cas protein (e.g., a Cas9 protein) and a guide RNA. In some such methods, the targeting vector is a large targeting vector at least 10 kb in length or in which the sum total of the 5' and 3' homology arms is at least 10 kb in length. In some such methods, the non-human animal is a mouse or a rat. In some such methods, the non-human animal is a mouse.

BRIEF DESCRIPTION OF THE FIGURES

[0024] FIG. 1A shows an alignment of human and mouse transthyretin (TTR) precursor proteins (SEQ ID NOS: 1 and 6, respectively). The signal peptide, T4 binding domain, phase 0 exon/intron boundaries, and phase 1/2 exon/intron boundaries are denoted.

[0025] FIG. 1B shows an alignment of human and mouse transthyretin (TTR) coding sequences (SEQ ID NOS: 90 and 92, respectively).

[0026] FIG. 2 shows schematics (not drawn to scale) of the wild-type murine Ttr locus, a first version of a humanized mouse Ttr locus, and a second version of a humanized mouse Ttr locus. Exons, introns, 5' untranslated regions (UTRs), 3' UTRs, start codons (ATG), stop codons (TGA), and loxP scars from selection cassettes are denoted. White boxes indicate murine sequence; black boxes indicate human sequence.

[0027] FIG. 3 shows a schematic (not drawn to scale) of the targeting to create the first version of the humanized mouse Ttr locus. The wild type mouse Ttr locus, the F0 allele of the humanized mouse Ttr locus with the self-deleting neomycin (SDC-Neo) selection cassette (MAID 7576), and the F1 allele of the humanized mouse Ttr locus with the loxP scar from removal of the SDC-Neo selection

cassette (MAID 7577) are shown. White boxes indicate murine sequence; black boxes indicate human sequence.

[0028] FIG. 4 shows a schematic (not drawn to scale) of the targeting to create the second version of the humanized mouse Ttr locus. The wild type mouse Ttr locus, the F0 allele of the humanized mouse Ttr locus with the SDC-Neo selection cassette, and the F1 allele of the humanized mouse Ttr locus with the loxP scar from removal of the SDC-Neo selection cassette are shown. White boxes indicate murine sequence; black boxes indicate human sequence.

[0029] FIG. 5A shows a schematic (not drawn to scale) of the strategy for screening of the first targeted mouse Ttr locus, including loss-of-allele assays (7576mTU, 9090mTM, and 9090mTD), gain of allele assays (7576hTU, 7576hTD, Neo), retention assays (9090retU, 9090retU2, 9090retU3, 9090retD, 9090retD2, 9090retD3), and CRISPR assays designed to cover the region that is disrupted by the CRISPR guides (9090mTGU, mGU, 9090mTGD, and mGD). White boxes indicate murine sequence; black boxes indicate human sequence.

[0030] FIG. 5B shows a schematic (not drawn to scale) of the strategy for screening of the second targeted mouse Ttr locus, including loss-of-allele assays (4552mTU, 9212mTU, 9090mTM, 9212mTD), gain of allele assays (7655hTU, 7576hTD, Neo), retention assays (9204mretU, 9204mretD), and CRISPR assays designed to cover the region that is disrupted by the CRISPR guides (mGU, mGD, and 9212mTGD). White boxes indicate murine sequence; black boxes indicate human sequence.

[0031] FIG. 6 shows beta-actin (Actb), beta-2-microglobulin (B2M), *Mus musculus* transthyretin (Mm Ttr), and *Homo sapiens* transthyretin (Hs TTR) mRNA expression in liver samples from (1) F0 generation mice homozygous for the first version of the humanized mouse Ttr locus (MAID 7576; F0 allele from FIG. 3), (2) liver samples from wild type mice, (3) spleen samples from F0 generation mice homozygous for the first version of the humanized mouse Ttr locus, and (4) spleen samples from wild type mice. Lower Ct values indicate higher expression.

[0032] FIGS. 7A and 7B show results of ELISA assays for human TTR protein levels (FIG. 7A) and mouse TTR protein levels (FIG. 7B) in serum and cerebrospinal fluid (CSF). The samples tested include serum and CSF from F0 generation mice homozygous for the first version of the humanized mouse Ttr locus (MAID 7576; F0 allele from FIG. 3), human serum and CSF controls, and mouse (F1H4) serum and CSF controls.

[0033] FIG. 7C shows results of ELISA assays for (1) human TTR and (2) mouse TTR protein levels in serum. The samples tested include serum samples from F0 generation mice homozygous for the first version of the humanized mouse Ttr locus (MAID 7576; F0 allele from FIG. 3) generated from a first clone (clone 7576C-G7), F0 generation mice homozygous for the first version of the humanized mouse Ttr locus (MAID 7576; F0 allele from FIG. 3) generated from a second clone (clone 7576A-A5), and wild type mice (F1H4). Mouse serum and human serum were used as controls.

[0034] FIG. 8 shows human TTR protein expression as determined by western blot in serum samples from wild type mice (F1H4), F0 generation mice homozygous for the first version of the humanized mouse Ttr locus (MAID 7576; F0 allele from FIG. 3) generated from a first clone (clone 7576C-G7), and F0 generation mice homozygous for the

first version of the humanized mouse *Ttr* locus generated from a second clone (7576A-A5). Mouse serum was used as a negative control, and human serum was used as a positive control. Mouse IgG was used as a loading control.

[0035] FIG. 9 shows human TTR protein expression as determined by western blot in liver and kidney samples from wild type mice (F1H4), F0 generation mice homozygous for the first version of the humanized mouse *Ttr* locus (MAID 7576; F0 allele from FIG. 3) generated from a first clone (clone 7576C-G7), and F0 generation mice homozygous for the first version of the humanized mouse *Ttr* locus generated from a second clone (7576A-A5). Mouse serum was used as a negative control, and human serum was used as a positive control. GAPDH was used as a loading control.

[0036] FIG. 10 shows percent genome editing (total number of insertions or deletions observed over the total number of sequences read in the PCR reaction from a pool of lysed cells) at the humanized mouse *Ttr* locus as determined by next-generation sequencing (NGS) in primary hepatocytes isolated from F0 generation mice homozygous for the first version of the humanized mouse *Ttr* locus (MAID 7576; F0 allele from FIG. 3). The samples tested included untreated hepatocytes and hepatocytes treated with lipid nanoparticles containing Cas9 mRNA and guide RNAs designed to target human TTR.

[0037] FIGS. 11A-11H show serum chemistry analysis of alanine aminotransferase (ALT) (FIG. 11A), aspartate aminotransferase (AST) (FIG. 11B), triglycerides (FIG. 11C), cholesterol (FIG. 11D), high-density lipoprotein (HDL) (FIG. 11E), low-density lipoprotein (LDL) (FIG. 11F), non-esterified fatty acids (NEFA) (FIG. 11G), and albumin (FIG. 11H) 14 days post-injection of lipid nanoparticles containing Cas9 mRNA and guide RNAs designed to target human TTR into F0 generation mice homozygous for the first version of the humanized mouse *Ttr* locus (MAID 7576; F0 allele from FIG. 3). U/L refers to units per liter, mg/dL refers to milligrams per deciliter, mEq/L refers to milliequivalents per liter, and g/dL refers to grams per deciliter.

[0038] FIG. 12 shows percent genome editing (total number of insertions or deletions observed over the total number of sequences read in the PCR reaction from a pool of lysed cells) at the humanized mouse *Ttr* locus as determined by next-generation sequencing (NGS) in samples from liver 14 days post-injection of buffer control or lipid nanoparticles containing Cas9 mRNA and guide RNAs designed to target human TTR into F0 generation mice homozygous for the first version of the humanized mouse *Ttr* locus (MAID 7576; F0 allele from FIG. 3).

[0039] FIG. 13 shows results of an ELISA assaying serum levels of human TTR in wild type mice (F1H4), mice in which human TTR plasmids were introduced by hydrodynamic delivery (HDD), and mice in which a chimeric mouse/human TTR plasmid (region encoded by exon 1 is mouse, region encoded by exons 2-4 is human) was introduced by HDD. Two negative controls are shown, and human serum was used as a positive control.

[0040] FIG. 14 shows results of an ELISA assaying human TTR levels in liver lysates 8 days post-injection of buffer control or lipid nanoparticles containing Cas9 mRNA and human TTR guide RNA 1 designed to target human TTR into F2 generation mice homozygous for the first version of the humanized mouse *Ttr* locus (MAID 7576; F1 allele from FIG. 3; derived from clone 7576B-F10).

[0041] FIGS. 15A and 15B show results of an ELISA assaying human TTR levels in serum samples (1:5000 dilution in FIG. 15A, and 1:10000 dilution in FIG. 15B) 8 days post-injection of buffer control or lipid nanoparticles containing Cas9 mRNA and human TTR guide RNA 1 designed to target human TTR into F2 generation mice homozygous for the first version of the humanized mouse *Ttr* locus (MAID 7576; F1 allele from FIG. 3; derived from clone 7576B-F10).

[0042] FIG. 16 shows results of an ELISA assaying human TTR levels in blood plasma samples of hTTR^{7577/7577}, hTTR^{7655/7655}, hTTR^{7655/7656}, and hTTR^{7656/7656}, and hTTR^{7656/WT} mice.

[0043] FIGS. 17A and 17B show results of an ELISA assaying human TTR and mouse TTR levels in blood plasma samples of hTTR^{WT/WT} and hTTR^{7577/7577} mice (3 months of age). Human serum was used as a control.

[0044] FIG. 17C shows mTTR (1) and hTTR (2) mRNA expression in liver samples from 3-month old hTTR^{WT/WT} and hTTR^{7577/7577} mice. Lower Ct values indicate higher expression.

[0045] FIG. 18 shows results of an ELISA assaying human TTR levels in blood plasma samples of wild type (F1H4), hTTR^{7577/7577} (hTTR v1), and hTTR^{7656/7656} (hTTRv2) mice (ages 2-3 months).

[0046] FIG. 19 shows percent genome editing at the humanized mouse *Ttr* locus as determined by next-generation sequencing (NGS) in samples from liver post-injection of buffer control or lipid nanoparticles containing Cas9 mRNA and guide RNAs designed to target human TTR into mice homozygous for the first version of the humanized mouse *Ttr* locus.

[0047] FIG. 20 shows results of an ELISA assaying human TTR levels in serum samples post-injection of buffer control or lipid nanoparticles containing Cas9 mRNA and guide RNAs designed to target human TTR into mice homozygous for the first version of the humanized mouse *Ttr* locus.

DEFINITIONS

[0048] The terms “protein,” “polypeptide,” and “peptide,” used interchangeably herein, include polymeric forms of amino acids of any length, including coded and non-coded amino acids and chemically or biochemically modified or derivatized amino acids. The terms also include polymers that have been modified, such as polypeptides having modified peptide backbones. The term “domain” refers to any part of a protein or polypeptide having a particular function or structure.

[0049] Proteins are said to have an “N-terminus” and a “C-terminus.” The term “N-terminus” relates to the start of a protein or polypeptide, terminated by an amino acid with a free amine group (—NH₂). The term “C-terminus” relates to the end of an amino acid chain (protein or polypeptide), terminated by a free carboxyl group (—COOH).

[0050] The terms “nucleic acid” and “polynucleotide,” used interchangeably herein, include polymeric forms of nucleotides of any length, including ribonucleotides, deoxyribonucleotides, or analogs or modified versions thereof. They include single-, double-, and multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, and polymers comprising purine bases, pyrimidine bases, or other natural, chemically modified, biochemically modified, non-natural, or derivatized nucleotide bases.

[0051] Nucleic acids are said to have “5' ends” and “3' ends” because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. An end of an oligonucleotide is referred to as the “5' end” if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring. An end of an oligonucleotide is referred to as the “3' end” if its 3' oxygen is not linked to a 5' phosphate of another mononucleotide pentose ring. A nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being “upstream” or 5' of the “downstream” or 3' elements.

[0052] The term “genomically integrated” refers to a nucleic acid that has been introduced into a cell such that the nucleotide sequence integrates into the genome of the cell. Any protocol may be used for the stable incorporation of a nucleic acid into the genome of a cell.

[0053] The term “expression vector” or “expression construct” or “expression cassette” refers to a recombinant nucleic acid containing a desired coding sequence operably linked to appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host cell or organism. Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome binding site, as well as other sequences. Eukaryotic cells are generally known to utilize promoters, enhancers, and termination and polyadenylation signals, although some elements may be deleted and other elements added without sacrificing the necessary expression.

[0054] The term “targeting vector” refers to a recombinant nucleic acid that can be introduced by homologous recombination, non-homologous-end-joining-mediated ligation, or any other means of recombination to a target position in the genome of a cell.

[0055] The term “viral vector” refers to a recombinant nucleic acid that includes at least one element of viral origin and includes elements sufficient for or permissive of packaging into a viral vector particle. The vector and/or particle can be utilized for the purpose of transferring DNA, RNA, or other nucleic acids into cells either *ex vivo* or *in vivo*. Numerous forms of viral vectors are known.

[0056] The term “isolated” with respect to proteins, nucleic acids, and cells includes proteins, nucleic acids, and cells that are relatively purified with respect to other cellular or organism components that may normally be present *in situ*, up to and including a substantially pure preparation of the protein, nucleic acid, or cell. The term “isolated” also includes proteins and nucleic acids that have no naturally occurring counterpart or proteins or nucleic acids that have been chemically synthesized and are thus substantially uncontaminated by other proteins or nucleic acids. The term “isolated” also includes proteins, nucleic acids, or cells that have been separated or purified from most other cellular components or organism components with which they are naturally accompanied (e.g., other cellular proteins, nucleic acids, or cellular or extracellular components).

[0057] The term “wild type” includes entities having a structure and/or activity as found in a normal (as contrasted

with mutant, diseased, altered, or so forth) state or context. Wild type genes and polypeptides often exist in multiple different forms (e.g., alleles).

[0058] The term “endogenous sequence” refers to a nucleic acid sequence that occurs naturally within a cell or non-human animal. For example, an endogenous Ttr sequence of a non-human animal refers to a native Ttr sequence that naturally occurs at the Ttr locus in the non-human animal.

[0059] “Exogenous” molecules or sequences include molecules or sequences that are not normally present in a cell in that form. Normal presence includes presence with respect to the particular developmental stage and environmental conditions of the cell. An exogenous molecule or sequence, for example, can include a mutated version of a corresponding endogenous sequence within the cell, such as a humanized version of the endogenous sequence, or can include a sequence corresponding to an endogenous sequence within the cell but in a different form (i.e., not within a chromosome). In contrast, endogenous molecules or sequences include molecules or sequences that are normally present in that form in a particular cell at a particular developmental stage under particular environmental conditions.

[0060] The term “heterologous” when used in the context of a nucleic acid or a protein indicates that the nucleic acid or protein comprises at least two segments that do not naturally occur together in the same molecule. For example, the term “heterologous,” when used with reference to segments of a nucleic acid or segments of a protein, indicates that the nucleic acid or protein comprises two or more sub-sequences that are not found in the same relationship to each other (e.g., joined together) in nature. As one example, a “heterologous” region of a nucleic acid vector is a segment of nucleic acid within or attached to another nucleic acid molecule that is not found in association with the other molecule in nature. For example, a heterologous region of a nucleic acid vector could include a coding sequence flanked by sequences not found in association with the coding sequence in nature. Likewise, a “heterologous” region of a protein is a segment of amino acids within or attached to another peptide molecule that is not found in association with the other peptide molecule in nature (e.g., a fusion protein, or a protein with a tag). Similarly, a nucleic acid or protein can comprise a heterologous label or a heterologous secretion or localization sequence.

[0061] “Codon optimization” takes advantage of the degeneracy of codons, as exhibited by the multiplicity of three-base pair codon combinations that specify an amino acid, and generally includes a process of modifying a nucleic acid sequence for enhanced expression in particular host cells by replacing at least one codon of the native sequence with a codon that is more frequently or most frequently used in the genes of the host cell while maintaining the native amino acid sequence. For example, a nucleic acid encoding a Cas9 protein can be modified to substitute codons having a higher frequency of usage in a given prokaryotic or eukaryotic cell, including a bacterial cell, a yeast cell, a human cell, a non-human cell, a mammalian cell, a rodent cell, a mouse cell, a rat cell, a hamster cell, or any other host cell, as compared to the naturally occurring nucleic acid sequence. Codon usage tables are readily available, for example, at the “Codon Usage Database.” These tables can be adapted in a number of ways. See Nakamura et al. (2000) *Nucleic Acids Research* 28:292, herein incorporated by

reference in its entirety for all purposes. Computer algorithms for codon optimization of a particular sequence for expression in a particular host are also available (see, e.g., Gene Forge).

[0062] The term “locus” refers to a specific location of a gene (or significant sequence), DNA sequence, polypeptide-encoding sequence, or position on a chromosome of the genome of an organism. For example, a “Ttr locus” may refer to the specific location of a Ttr gene, Ttr DNA sequence, transthyretin-encoding sequence, or Ttr position on a chromosome of the genome of an organism that has been identified as to where such a sequence resides. A “Ttr locus” may comprise a regulatory element of a Ttr gene, including, for example, an enhancer, a promoter, 5' and/or 3' untranslated region (UTR), or a combination thereof.

[0063] The term “gene” refers to a DNA sequence in a chromosome that codes for a product (e.g., an RNA product and/or a polypeptide product) and includes the coding region interrupted with non-coding introns and sequence located adjacent to the coding region on both the 5' and 3' ends such that the gene corresponds to the full-length mRNA (including the 5' and 3' untranslated sequences). The term “gene” also includes other non-coding sequences including regulatory sequences (e.g., promoters, enhancers, and transcription factor binding sites), polyadenylation signals, internal ribosome entry sites, silencers, insulating sequence, and matrix attachment regions. These sequences may be close to the coding region of the gene (e.g., within 10 kb) or at distant sites, and they influence the level or rate of transcription and translation of the gene.

[0064] The term “allele” refers to a variant form of a gene. Some genes have a variety of different forms, which are located at the same position, or genetic locus, on a chromosome. A diploid organism has two alleles at each genetic locus. Each pair of alleles represents the genotype of a specific genetic locus. Genotypes are described as homozygous if there are two identical alleles at a particular locus and as heterozygous if the two alleles differ.

[0065] The “coding region” or “coding sequence” of a gene consists of the portion of a gene’s DNA or RNA, composed of exons, that codes for a protein. The region begins at the start codon on the 5' end and ends at the stop codon on the 3' end.

[0066] A “promoter” is a regulatory region of DNA usually comprising a TATA box capable of directing RNA polymerase II to initiate RNA synthesis at the appropriate transcription initiation site for a particular polynucleotide sequence. A promoter may additionally comprise other regions which influence the transcription initiation rate. The promoter sequences disclosed herein modulate transcription of an operably linked polynucleotide. A promoter can be active in one or more of the cell types disclosed herein (e.g., a eukaryotic cell, a non-human mammalian cell, a human cell, a rodent cell, a pluripotent cell, a one-cell stage embryo, a differentiated cell, or a combination thereof). A promoter can be, for example, a constitutively active promoter, a conditional promoter, an inducible promoter, a temporally restricted promoter (e.g., a developmentally regulated promoter), or a spatially restricted promoter (e.g., a cell-specific or tissue-specific promoter). Examples of promoters can be found, for example, in WO 2013/176772, herein incorporated by reference in its entirety for all purposes.

[0067] “Operable linkage” or being “operably linked” includes juxtaposition of two or more components (e.g., a

promoter and another sequence element) such that both components function normally and allow the possibility that at least one of the components can mediate a function that is exerted upon at least one of the other components. For example, a promoter can be operably linked to a coding sequence if the promoter controls the level of transcription of the coding sequence in response to the presence or absence of one or more transcriptional regulatory factors. Operable linkage can include such sequences being contiguous with each other or acting in trans (e.g., a regulatory sequence can act at a distance to control transcription of the coding sequence).

[0068] “Complementarity” of nucleic acids means that a nucleotide sequence in one strand of nucleic acid, due to orientation of its nucleobase groups, forms hydrogen bonds with another sequence on an opposing nucleic acid strand. The complementary bases in DNA are typically A with T and C with G. In RNA, they are typically C with G and U with A. Complementarity can be perfect or substantial/sufficient. Perfect complementarity between two nucleic acids means that the two nucleic acids can form a duplex in which every base in the duplex is bonded to a complementary base by Watson-Crick pairing. “Substantial” or “sufficient” complementary means that a sequence in one strand is not completely and/or perfectly complementary to a sequence in an opposing strand, but that sufficient bonding occurs between bases on the two strands to form a stable hybrid complex in set of hybridization conditions (e.g., salt concentration and temperature). Such conditions can be predicted by using the sequences and standard mathematical calculations to predict the T_m (melting temperature) of hybridized strands, or by empirical determination of T_m by using routine methods. T_m includes the temperature at which a population of hybridization complexes formed between two nucleic acid strands are 50% denatured (i.e., a population of double-stranded nucleic acid molecules becomes half dissociated into single strands). At a temperature below the T_m , formation of a hybridization complex is favored, whereas at a temperature above the T_m , melting or separation of the strands in the hybridization complex is favored. T_m may be estimated for a nucleic acid having a known G+C content in an aqueous 1 M NaCl solution by using, e.g., $T_m=81.5+0.41(\% \text{ G+C})$, although other known T_m computations take into account nucleic acid structural characteristics.

[0069] “Hybridization condition” includes the cumulative environment in which one nucleic acid strand bonds to a second nucleic acid strand by complementary strand interactions and hydrogen bonding to produce a hybridization complex. Such conditions include the chemical components and their concentrations (e.g., salts, chelating agents, formamide) of an aqueous or organic solution containing the nucleic acids, and the temperature of the mixture. Other factors, such as the length of incubation time or reaction chamber dimensions may contribute to the environment. See, e.g., Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2^{sup}.nd ed., pp. 1.90-1.91, 9.47-9.51, 11.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), herein incorporated by reference in its entirety for all purposes.

[0070] Hybridization requires that the two nucleic acids contain complementary sequences, although mismatches between bases are possible. The conditions appropriate for hybridization between two nucleic acids depend on the length of the nucleic acids and the degree of complemen-

tation, variables which are well known. The greater the degree of complementation between two nucleotide sequences, the greater the value of the melting temperature (T_m) for hybrids of nucleic acids having those sequences. For hybridizations between nucleic acids with short stretches of complementarity (e.g. complementarity over 35 or fewer, 30 or fewer, 25 or fewer, 22 or fewer, 20 or fewer, or 18 or fewer nucleotides) the position of mismatches becomes important (see Sambrook et al., supra, 11.7-11.8). Typically, the length for a hybridizable nucleic acid is at least about 10 nucleotides. Illustrative minimum lengths for a hybridizable nucleic acid include at least about 15 nucleotides, at least about 20 nucleotides, at least about 22 nucleotides, at least about 25 nucleotides, and at least about 30 nucleotides. Furthermore, the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the region of complementation and the degree of complementation.

[0071] The sequence of polynucleotide need not be 100% complementary to that of its target nucleic acid to be specifically hybridizable. Moreover, a polynucleotide may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (e.g., a loop structure or hairpin structure). A polynucleotide (e.g., gRNA) can comprise at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence complementarity to a target region within the target nucleic acid sequence to which they are targeted. For example, a gRNA in which 18 of 20 nucleotides are complementary to a target region, and would therefore specifically hybridize, would represent 90% complementarity. In this example, the remaining noncomplementary nucleotides may be clustered or interspersed with complementary nucleotides and need not be contiguous to each other or to complementary nucleotides.

[0072] Percent complementarity between particular stretches of nucleic acid sequences within nucleic acids can be determined routinely using BLAST programs (basic local alignment search tools) and PowerBLAST programs (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; Zhang and Madden (1997) *Genome Res.* 7:649-656, each of which is herein incorporated by reference in its entirety for all purposes) or by using the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.), using default settings, which uses the algorithm of Smith and Waterman (1981) *Adv. Appl. Math.* 2:482-489, herein incorporated by reference in its entirety for all purposes.

[0073] The methods and compositions provided herein employ a variety of different components. Some components throughout the description can have active variants and fragments. Such components include, for example, Cas proteins, CRISPR RNAs, tracrRNAs, and guide RNAs. Biological activity for each of these components is described elsewhere herein. The term “functional” refers to the innate ability of a protein or nucleic acid (or a fragment or variant thereof) to exhibit a biological activity or function. Such biological activities or functions can include, for example, the ability of a Cas protein to bind to a guide RNA and to a target DNA sequence. The biological functions of functional fragments or variants may be the same or may in fact be changed (e.g., with respect to their specificity or selectivity or efficacy) in comparison to the original, but with retention of the basic biological function.

[0074] The term “variant” refers to a nucleotide sequence differing from the sequence most prevalent in a population (e.g., by one nucleotide) or a protein sequence different from the sequence most prevalent in a population (e.g., by one amino acid).

[0075] The term “fragment” when referring to a protein means a protein that is shorter or has fewer amino acids than the full-length protein. The term “fragment” when referring to a nucleic acid means a nucleic acid that is shorter or has fewer nucleotides than the full-length nucleic acid. A fragment can be, for example, an N-terminal fragment (i.e., removal of a portion of the C-terminal end of the protein), a C-terminal fragment (i.e., removal of a portion of the N-terminal end of the protein), or an internal fragment.

[0076] “Sequence identity” or “identity” in the context of two polynucleotides or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins, residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have “sequence similarity” or “similarity.” Means for making this adjustment are well known. Typically, this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif.).

[0077] “Percentage of sequence identity” includes the value determined by comparing two optimally aligned sequences (greatest number of perfectly matched residues) over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity. Unless otherwise specified (e.g., the shorter sequence includes a linked heterologous sequence), the comparison window is the full length of the shorter of the two sequences being compared.

[0078] Unless otherwise stated, sequence identity/similarity values include the value obtained using GAP Version 10 using the following parameters: % identity and % similarity for a nucleotide sequence using GAP Weight of 50 and Length Weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using GAP Weight of 8 and Length Weight of 2, and the

BLOSUM62 scoring matrix; or any equivalent program thereof “Equivalent program” includes any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10.

[0079] The term “conservative amino acid substitution” refers to the substitution of an amino acid that is normally present in the sequence with a different amino acid of similar size, charge, or polarity. Examples of conservative substitutions include the substitution of a non-polar (hydrophobic) residue such as isoleucine, valine, or leucine for another non-polar residue. Likewise, examples of conservative substitutions include the substitution of one polar (hydrophilic) residue for another such as between arginine and lysine, between glutamine and asparagine, or between glycine and serine. Additionally, the substitution of a basic residue such as lysine, arginine, or histidine for another, or the substitution of one acidic residue such as aspartic acid or glutamic acid for another acidic residue are additional examples of conservative substitutions. Examples of non-conservative substitutions include the substitution of a non-polar (hydrophobic) amino acid residue such as isoleucine, valine, leucine, alanine, or methionine for a polar (hydrophilic) residue such as cysteine, glutamine, glutamic acid or lysine and/or a polar residue for a non-polar residue. Typical amino acid categorizations are summarized in Table 1 below.

TABLE 1

Amino Acid Categorizations.					
Alanine	Ala	A	Nonpolar	Neutral	1.8
Arginine	Arg	R	Polar	Positive	-4.5
Asparagine	Asn	N	Polar	Neutral	-3.5
Aspartic acid	Asp	D	Polar	Negative	-3.5
Cysteine	Cys	C	Nonpolar	Neutral	2.5
Glutamic acid	Glu	E	Polar	Negative	-3.5
Glutamine	Gln	Q	Polar	Neutral	-3.5
Glycine	Gly	G	Nonpolar	Neutral	-0.4
Histidine	His	H	Polar	Positive	-3.2
Isoleucine	Ile	I	Nonpolar	Neutral	4.5
Leucine	Leu	L	Nonpolar	Neutral	3.8
Lysine	Lys	K	Polar	Positive	-3.9
Methionine	Met	M	Nonpolar	Neutral	1.9
Phenylalanine	Phe	F	Nonpolar	Neutral	2.8
Proline	Pro	P	Nonpolar	Neutral	-1.6
Serine	Ser	S	Polar	Neutral	-0.8
Threonine	Thr	T	Polar	Neutral	-0.7
Tryptophan	Trp	W	Nonpolar	Neutral	-0.9
Tyrosine	Tyr	Y	Polar	Neutral	-1.3
Valine	Val	V	Nonpolar	Neutral	4.2

[0080] A “homologous” sequence (e.g., nucleic acid sequence) includes a sequence that is either identical or substantially similar to a known reference sequence, such that it is, for example, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to the known reference sequence. Homologous sequences can include, for example, orthologous sequence and paralogous sequences. Homologous genes, for example, typically descend from a common ancestral DNA sequence, either through a speciation event (orthologous genes) or a genetic duplication event (paralogous genes). “Orthologous” genes include genes in different species that evolved from a

common ancestral gene by speciation. Orthologs typically retain the same function in the course of evolution. “Paralogous” genes include genes related by duplication within a genome. Paralogs can evolve new functions in the course of evolution.

[0081] The term “in vitro” includes artificial environments and to processes or reactions that occur within an artificial environment (e.g., a test tube). The term “in vivo” includes natural environments (e.g., a cell or organism or body) and to processes or reactions that occur within a natural environment. The term “ex vivo” includes cells that have been removed from the body of an individual and to processes or reactions that occur within such cells.

[0082] The term “reporter gene” refers to a nucleic acid having a sequence encoding a gene product (typically an enzyme) that is easily and quantifiably assayed when a construct comprising the reporter gene sequence operably linked to an endogenous or heterologous promoter and/or enhancer element is introduced into cells containing (or which can be made to contain) the factors necessary for the activation of the promoter and/or enhancer elements. Examples of reporter genes include, but are not limited, to genes encoding beta-galactosidase (lacZ), the bacterial chloramphenicol acetyltransferase (cat) genes, firefly luciferase genes, genes encoding beta-glucuronidase (GUS), and genes encoding fluorescent proteins. A “reporter protein” refers to a protein encoded by a reporter gene.

[0083] The term “fluorescent reporter protein” as used herein means a reporter protein that is detectable based on fluorescence wherein the fluorescence may be either from the reporter protein directly, activity of the reporter protein on a fluorogenic substrate, or a protein with affinity for binding to a fluorescent tagged compound. Examples of fluorescent proteins include green fluorescent proteins (e.g., GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, and ZsGreen1), yellow fluorescent proteins (e.g., YFP, eYFP, Citrine, Venus, YPet, PhiYFP, and ZsYellow1), blue fluorescent proteins (e.g., BFP, eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, and T-sapphire), cyan fluorescent proteins (e.g., CFP, eCFP, Cerulean, CyPet, AmCyan1, and Midoriishi-Cyan), red fluorescent proteins (e.g., RFP, mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, and Jred), orange fluorescent proteins (e.g., mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, and tdTomato), and any other suitable fluorescent protein whose presence in cells can be detected by flow cytometry methods.

[0084] Repair in response to double-strand breaks (DSBs) occurs principally through two conserved DNA repair pathways: homologous recombination (HR) and non-homologous end joining (NHEJ). See Kasparek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897, herein incorporated by reference in its entirety for all purposes. Likewise, repair of a target nucleic acid mediated by an exogenous donor nucleic acid can include any process of exchange of genetic information between the two polynucleotides.

[0085] The term “recombination” includes any process of exchange of genetic information between two polynucleotides and can occur by any mechanism. Recombination can occur via homology directed repair (HDR) or homologous

recombination (HR). HDR or HR includes a form of nucleic acid repair that can require nucleotide sequence homology, uses a “donor” molecule as a template for repair of a “target” molecule (i.e., the one that experienced the double-strand break), and leads to transfer of genetic information from the donor to target. Without wishing to be bound by any particular theory, such transfer can involve mismatch correction of heteroduplex DNA that forms between the broken target and the donor, and/or synthesis-dependent strand annealing, in which the donor is used to resynthesize genetic information that will become part of the target, and/or related processes. In some cases, the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA. See Wang et al. (2013) *Cell* 153:910-918; Mandalos et al. (2012) *PLOS ONE* 7:e45768:1-9; and Wang et al. (2013) *Nat Biotechnol.* 31:530-532, each of which is herein incorporated by reference in its entirety for all purposes.

[0086] NHEJ includes the repair of double-strand breaks in a nucleic acid by direct ligation of the break ends to one another or to an exogenous sequence without the need for a homologous template. Ligation of non-contiguous sequences by NHEJ can often result in deletions, insertions, or translocations near the site of the double-strand break. For example, NHEJ can also result in the targeted integration of an exogenous donor nucleic acid through direct ligation of the break ends with the ends of the exogenous donor nucleic acid (i.e., NHEJ-based capture). Such NHEJ-mediated targeted integration can be preferred for insertion of an exogenous donor nucleic acid when homology directed repair (HDR) pathways are not readily usable (e.g., in non-dividing cells, primary cells, and cells which perform homology-based DNA repair poorly). In addition, in contrast to homology-directed repair, knowledge concerning large regions of sequence identity flanking the cleavage site is not needed, which can be beneficial when attempting targeted insertion into organisms that have genomes for which there is limited knowledge of the genomic sequence. The integration can proceed via ligation of blunt ends between the exogenous donor nucleic acid and the cleaved genomic sequence, or via ligation of sticky ends (i.e., having 5' or 3' overhangs) using an exogenous donor nucleic acid that is flanked by overhangs that are compatible with those generated by a nuclease agent in the cleaved genomic sequence. See, e.g., US 2011/020722, WO 2014/033644, WO 2014/089290, and Maresca et al. (2013) *Genome Res.* 23(3):539-546, each of which is herein incorporated by reference in its entirety for all purposes. If blunt ends are ligated, target and/or donor resection may be needed to generation regions of microhomology needed for fragment joining, which may create unwanted alterations in the target sequence.

[0087] The term “antigen-binding protein” includes any protein that binds to an antigen. Examples of antigen-binding proteins include an antibody, an antigen-binding fragment of an antibody, a multispecific antibody (e.g., a bi-specific antibody), an scFV, a bis-scFV, a diabody, a triabody, a tetrabody, a V-NAR, a VHH, a VL, a F(ab), a F(ab)₂, a DVD (dual variable domain antigen-binding protein), an SVD (single variable domain antigen-binding protein), a bispecific T-cell engager (BiTE), or a Davisbody (U.S. Pat. No. 8,586,713, herein incorporated by reference herein in its entirety for all purposes).

[0088] The term “antigen” refers to a substance, whether an entire molecule or a domain within a molecule, which is capable of eliciting production of antibodies with binding specificity to that substance. The term antigen also includes substances, which in wild type host organisms would not elicit antibody production by virtue of self-recognition, but can elicit such a response in a host animal with appropriate genetic engineering to break immunological tolerance.

[0089] The term “epitope” refers to a site on an antigen to which an antigen-binding protein (e.g., antibody) binds. An epitope can be formed from contiguous amino acids or noncontiguous amino acids juxtaposed by tertiary folding of one or more proteins. Epitopes formed from contiguous amino acids (also known as linear epitopes) are typically retained on exposure to denaturing solvents whereas epitopes formed by tertiary folding (also known as conformational epitopes) are typically lost on treatment with denaturing solvents. An epitope typically includes at least 3, and more usually, at least 5 or 8-10 amino acids in a unique spatial conformation. Methods of determining spatial conformation of epitopes include, for example, x-ray crystallography and 2-dimensional nuclear magnetic resonance. See, e.g., Epitope Mapping Protocols, in *Methods in Molecular Biology*, Vol. 66, Glenn E. Morris, Ed. (1996), herein incorporated by reference in its entirety for all purposes.

[0090] An antibody paratope as described herein generally comprises at a minimum a complementarity determining region (CDR) that specifically recognizes the heterologous epitope (e.g., a CDR3 region of a heavy and/or light chain variable domain).

[0091] The term “antibody” includes immunoglobulin molecules comprising four polypeptide chains, two heavy (H) chains and two light (L) chains inter-connected by disulfide bonds. Each heavy chain comprises a heavy chain variable domain and a heavy chain constant region (C_H). The heavy chain constant region comprises three domains: C_H1, C_H2 and C_H3. Each light chain comprises a light chain variable domain and a light chain constant region (C_L). The heavy chain and light chain variable domains can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDR), interspersed with regions that are more conserved, termed framework regions (FR). Each heavy and light chain variable domain comprises three CDRs and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4 (heavy chain CDRs may be abbreviated as HCDR1, HCDR2 and HCDR3; light chain CDRs may be abbreviated as LCDR1, LCDR2 and LCDR3). The term “high affinity” antibody refers to an antibody that has a K_D with respect to its target epitope about of 10⁻⁹ M or lower (e.g., about 1×10⁻⁹ M, 1×10⁻¹⁰ M, 1×10⁻¹¹ M, or about 1×10⁻¹² M). In one embodiment, K_D is measured by surface plasmon resonance, e.g., BIACORE™; in another embodiment, K_D is measured by ELISA.

[0092] Specific binding of an antigen-binding protein to its target antigen includes binding with an affinity of at least 10⁶, 10⁷, 10⁸, 10⁹, or 10¹⁰ M⁻¹. Specific binding is detectably higher in magnitude and distinguishable from non-specific binding occurring to at least one unrelated target. Specific binding can be the result of formation of bonds between particular functional groups or particular spatial fit (e.g., lock and key type) whereas non-specific binding is usually the result of van der Waals forces. Specific binding

does not however necessarily imply that an antigen-binding protein binds one and only one target.

[0093] The term “antisense RNA” refers to a single-stranded RNA that is complementary to a messenger RNA strand transcribed in a cell.

[0094] The term “small interfering RNA (siRNA)” refers to a typically double-stranded RNA molecule that induces the RNA interference (RNAi) pathway. These molecules can vary in length (generally between 18-30 base pairs) and contain varying degrees of complementarity to their target mRNA in the antisense strand. Some, but not all, siRNAs have unpaired overhanging bases on the 5' or 3' end of the sense strand and/or the antisense strand. The term “siRNA” includes duplexes of two separate strands, as well as single strands that can form hairpin structures comprising a duplex region. The double-stranded structure can be, for example, less than 20, 25, 30, 35, 40, 45, or 50 nucleotides in length. For example, the double-stranded structure can be from about 21-23 nucleotides in length, from about 19-25 nucleotides in length, or from about 19-23 nucleotides in length.

[0095] The term “short hairpin RNA (shRNA)” refers to a single strand of RNA bases that self-hybridizes in a hairpin structure and can induce the RNA interference (RNAi) pathway upon processing. These molecules can vary in length (generally about 50-90 nucleotides in length, or in some cases up to greater than 250 nucleotides in length, e.g., for microRNA-adapted shRNA). shRNA molecules are processed within the cell to form siRNAs, which in turn can knock down gene expression. shRNAs can be incorporated into vectors. The term “shRNA” also refers to a DNA molecule from which a short, hairpin RNA molecule may be transcribed.

[0096] Compositions or methods “comprising” or “including” one or more recited elements may include other elements not specifically recited. For example, a composition that “comprises” or “includes” a protein may contain the protein alone or in combination with other ingredients. The transitional phrase “consisting essentially of” means that the scope of a claim is to be interpreted to encompass the specified elements recited in the claim and those that do not materially affect the basic and novel characteristic(s) of the claimed invention. Thus, the term “consisting essentially of” when used in a claim of this invention is not intended to be interpreted to be equivalent to “comprising.”

[0097] “Optional” or “optionally” means that the subsequently described event or circumstance may or may not occur and that the description includes instances in which the event or circumstance occurs and instances in which it does not.

[0098] Designation of a range of values includes all integers within or defining the range, and all subranges defined by integers within the range.

[0099] Unless otherwise apparent from the context, the term “about” encompasses values within a standard margin of error of measurement (e.g., SEM) of a stated value.

[0100] The term “and/or” refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative (“or”).

[0101] The term “or” refers to any one member of a particular list and also includes any combination of members of that list.

[0102] The singular forms of the articles “a,” “an,” and “the” include plural references unless the context clearly

dictates otherwise. For example, the term “a protein” or “at least one protein” can include a plurality of proteins, including mixtures thereof.

[0103] Statistically significant means $p \leq 0.05$.

DETAILED DESCRIPTION

I. Overview

[0104] Disclosed herein are non-human animal genomes, non-human animal cells, and non-human animals comprising in their genome a humanized TTR locus and methods of using such non-human animal cells and non-human animals. Non-human animal cells or non-human animals comprising a humanized TTR locus express a human transthyretin protein or a chimeric transthyretin protein comprising one or more fragments of a human transthyretin protein. Such non-human animal cells and non-human animals can be used to assess delivery or efficacy of human-TTR-targeting agents (e.g., CRISPR/Cas9 genome editing agents) *ex vivo* or *in vivo* and can be used in methods of optimizing the delivery of efficacy of such agents *ex vivo* or *in vivo*.

[0105] In some of the non-human animal cells and non-human animals disclosed herein, most or all of the human TTR genomic DNA is inserted into the corresponding orthologous non-human animal Ttr locus. In some of the non-human animal cells and non-human animals disclosed herein, most or all of the non-human animal Ttr genomic DNA is replaced one-for-one with corresponding orthologous human TTR genomic DNA. Compared to non-human animals with cDNA insertions, expression levels should be higher when the intron-exon structure and splicing machinery are maintained because conserved regulator elements are more likely to be left intact, and spliced transcripts that undergo RNA processing are more stable than cDNAs. In contrast, insertion of human TTR cDNA (e.g., along with insertion of an artificial beta-globin intron in the 5' UTR) into a non-human animal Ttr locus would abolish conserved regulatory elements such as those contained within the first exon and intron of the non-human animal Ttr. Replacing the non-human animal genomic sequence with the corresponding orthologous human genomic sequence or inserting human TTR genomic sequence in the corresponding orthologous non-human Ttr locus is more likely to result in faithful expression of the transgene from the endogenous Ttr locus. Similarly, transgenic non-human animals with transgenic insertion of human-TTR-coding sequences at a random genomic locus rather than the endogenous non-human animal Ttr locus will not as accurately reflect the endogenous regulation of Ttr expression. A humanized TTR allele resulting from replacing most or all of the non-human animal genomic DNA one-for-one with corresponding orthologous human genomic DNA or inserting human TTR genomic sequence in the corresponding orthologous non-human Ttr locus will provide the true human target or a close approximation of the true human target of human-TTR-targeting reagents (e.g., CRISPR/Cas9 reagents designed to target human TTR), thereby enabling testing of the efficacy and mode of action of such agents in live animals as well as pharmacokinetic and pharmacodynamics studies in a setting where the humanized protein and humanized gene are the only version of TTR present.

II. Non-Human Animals Comprising a Humanized TTR Locus

[0106] The cells and non-human animals disclosed herein comprise in their genome a humanized TTR locus. Cells or non-human animals comprising a humanized TTR locus express a human transthyretin protein or a partially humanized, chimeric transthyretin protein in which one or more fragments of the native transthyretin protein have been replaced with corresponding fragments from human transthyretin.

[0107] A. Transthyretin (TTR)

[0108] The cells and non-human animals described herein comprise a humanized transthyretin (Tr) locus. Transthyretin (TTR) is a 127-amino acid, 55 kDa serum and cerebrospinal fluid transport protein primarily synthesized by the liver but also produced by the choroid plexus. It has also been referred to as prealbumin, thyroxine binding prealbumin, ATTR, TBPA, CTS, CTS1, HEL111, HsT2651, and PALB. In its native state, TTR exists as a tetramer. In homozygotes, homo-tetramers comprise identical 127-amino-acid beta-sheet-rich subunits. In heterozygotes, TTR tetramers can be made up of variant and/or wild-type subunits, typically combined in a statistical fashion. TTR is responsible for carrying thyroxine (T4) and retinol-bound RBP (retinol-binding protein) in both the serum and the cerebrospinal fluid.

[0109] Unless otherwise apparent from context, reference to human transthyretin (TTR) or its fragments or domains includes the natural, wild type human amino acid sequences including isoforms and allelic variants thereof. Transthyretin precursor protein includes a signal sequence (typically 20 amino acids), whereas the mature transthyretin protein does not. Exemplary TTR polypeptide sequences are designated by Accession Numbers NP_000362.1 (NCBI) and P02766.1 (UniProt) (identical, each set forth SEQ ID NO: 1). Residues may be numbered according to UniProt Accession Number P02766.1, with the first amino acid of the mature protein (i.e., not including the 20 amino acid signal sequence) designated residue 1. In any other TTR protein, residues are numbered according to the corresponding residues in UniProt Accession Number P02766.1 on maximum alignment.

[0110] The human TTR gene is located on chromosome 18 and includes four exons and three introns. An exemplary human TTR gene is from residues 5001-12258 in the sequence designated by GenBank Accession Number NG_009490.1 (SEQ ID NO: 3). The four exons in SEQ ID NO: 3 include residues 1-205, 1130-1260, 3354-3489, and 6802-7258, respectively. The TTR coding sequence in SEQ ID NO: 3 includes residues 137-205, 1130-1260, 3354-3489, and 6802-6909. An exemplary human TTR mRNA is designated by NCBI Accession Number NM_000371.3 (SEQ ID NO: 4).

[0111] The mouse Ttr gene is located and chromosome 18 and also includes four exons and three introns. An exemplary mouse Ttr gene is from residues 20665250 to 20674326 the sequence designated by GenBank Accession Number NC_000084.6 (SEQ ID NO: 5). The four exons in SEQ ID NO: 5 include residues 1-258, 1207-1337, 4730-4865, and 8382-9077, respectively. The Ttr coding sequence in SEQ ID NO: 5 includes residues 190-258, 1207-1337, 4730-4865, and 8382-8489. An exemplary mouse TTR protein is designated by UniProt Accession Number P07309.1 or NCBI Accession Number NP_038725.1 (identical, each set forth SEQ ID NO: 6). An exemplary mouse Ttr

mRNA is designated by NCBI Accession Number NM_013697.5 (SEQ ID NO: 7).

[0112] An exemplary rat TTR protein is designated by UniProt Accession Number P02767. An exemplary pig TTR protein is designated by UniProt Accession Number P50390. An exemplary chicken TTR protein is designated by UniProt Accession Number P27731. An exemplary cow TTR protein is designated by UniProt Accession Number O46375. An exemplary sheep TTR protein is designated by UniProt Accession Number P12303. An exemplary chimpanzee TTR protein designated by UniProt Accession Number Q5U7I5. An exemplary orangutan TTR protein is designated by UniProt Accession Number Q5NVS2. An exemplary rabbit TTR protein is designated by UniProt Accession Number P07489. An exemplary cynomolgus monkey (macaque) TTR protein is designated by UniProt Accession Number Q8HXW1.

[0113] Transthyretin (TTR) amyloidosis is a systemic disorder characterized by pathogenic, misfolded TTR and the extracellular deposition of amyloid fibrils composed of TTR. TTR amyloidosis is generally caused by destabilization of the native TTR tetramer form (due to environmental or genetic conditions), leading to dissociation, misfolding, and aggregation of TTR into amyloid fibrils that accumulate in various organs and tissues, causing progressive dysfunction. The dissociated monomers have a propensity to form misfolded protein aggregates and amyloid fibrils.

[0114] In humans, both wild-type TTR tetramers and mixed tetramers made up of mutant and wild-type subunits can dissociate, misfold, and aggregate, with the process of amyloidogenesis leading to the degeneration of post-mitotic tissue. Thus, TTR amyloidoses encompass diseases caused by pathogenic misfolded TTR resulting from mutations in TTR or resulting from non-mutated, misfolded TTR.

[0115] Senile systemic amyloidosis (SSA) and senile cardiac amyloidosis (SCA) are age-related types of amyloidosis that result from the deposition of wild-type TTR amyloid outside and within the cardiomyocytes of the heart. TTR amyloidosis is also the most common form of hereditary (familial) amyloidosis, which is caused by mutations that destabilize the TTR protein. TTR amyloidoses associated with point mutations in the TTR gene include familial amyloid polyneuropathy (FAP), familial amyloid cardiomyopathy (FAC), and central nervous system selective amyloidosis (CNSA).

[0116] B. Humanized TTR Loci

[0117] A humanized TTR locus disclosed herein can be a Ttr locus in which the entire Ttr gene is replaced with the corresponding orthologous human TTR sequence, or it can be a Ttr locus in which only a portion of the Ttr gene is replaced with the corresponding orthologous human TTR sequence (i.e., humanized). A humanized TTR locus can also comprise human TTR sequence inserted into an endogenous Ttr locus without replacing the corresponding orthologous endogenous sequence. A human TTR sequence corresponding to a particular segment of endogenous Ttr sequence refers to the region of human TTR that aligns with the particular segment of endogenous Ttr sequence when human TTR and the endogenous Ttr are optimally aligned. Optionally, the human TTR sequence is modified to be codon-optimized based on codon usage in the non-human animal. Replaced or inserted (i.e., humanized) regions can include coding regions such as an exon, non-coding regions such as an intron, untranslated regions, or regulatory regions

(e.g., a promoter, an enhancer, or a transcriptional repressor-binding element), or any combination thereof.

[0118] A humanized TTR locus is one in which a region of the endogenous Ttr locus has been deleted and replaced with a corresponding orthologous human TTR sequence (e.g., orthologous wild type human TTR sequence). Alternatively, a humanized TTR locus is one in which a region of the human TTR locus has been inserted into a corresponding endogenous non-human-animal Ttr locus. As one example, the replaced or inserted region of the endogenous Ttr locus can comprise both a coding sequence (i.e., all or part of an exon) and a non-coding sequence (i.e., all or part of intron), such as at least one exon and at least one intron. For example, the replaced or inserted region can comprise at least one exon and at least one intron. The replaced or inserted region comprising both coding sequence and non-coding sequence can be a contiguous region of the endogenous Ttr locus, meaning there is no intervening sequence between the replaced or inserted coding sequence and the replaced or inserted non-coding sequence. For example, the replaced or inserted region can comprise at least one exon and at least one adjacent intron. The replaced or inserted region can comprise one exon, two exons, three exons, four exons, or all exons of the endogenous Ttr locus. The inserted human TTR sequence can comprise one exon, two exons, three exons, four exons, or all exons of a human TTR gene. Likewise, the replaced region can comprise one intron, two introns, three introns, or all introns of the endogenous Ttr locus. The inserted human TTR sequence can comprise one intron, two introns, three introns, or all introns of a human TTR gene. Optionally, one or more introns and/or one or more exons of the endogenous Ttr locus remain unmodified (i.e., not deleted and replaced). For example, the first exon of the endogenous Ttr locus can remain unmodified. Similarly, the first exon and the first intron of the endogenous Ttr locus can remain unmodified.

[0119] In one specific example, the entire coding sequence for the transthyretin precursor protein can be deleted and replaced with the corresponding orthologous human TTR sequence. For example, the region of the endogenous Ttr locus beginning at the start codon and ending at the stop codon can be deleted and replaced with the corresponding orthologous human TTR sequence. In another specific example, the entire coding sequence for the human transthyretin precursor protein can be inserted. For example, the region of the human TTR locus beginning at the start codon and ending at the stop codon can be inserted.

[0120] Flanking untranslated regions including regulatory sequences can also be humanized. The first exon of a Ttr locus typically include a 5' untranslated region upstream of the start codon. Likewise, the last exon of a Ttr locus typically includes a 3' untranslated region downstream of the stop codon. Regions upstream of the Ttr start codon and downstream of the Ttr stop codon can either be unmodified or can be deleted and replaced with the corresponding orthologous human TTR sequence. For example, the 5' untranslated region (UTR), the 3'UTR, or both the 5' UTR and the 3' UTR can be humanized, or the 5' UTR, the 3'UTR, or both the 5' UTR and the 3' UTR can remain endogenous. In one specific example, the 5' UTR remains endogenous. In another specific example, the 3' UTR is humanized, but the 5' UTR remains endogenous. In another specific example, the 5' UTR remains endogenous, and a human TTR 3' UTR is inserted into the endogenous Ttr locus. For example, the

human TTR 3' UTR can replace the endogenous 3' UTR or can be inserted without replacing the endogenous 3' UTR (e.g., it can be inserted upstream of the endogenous 3' UTR).

[0121] One or more regions of the endogenous Ttr locus encoding one or more domains of the transthyretin precursor protein can be humanized. Likewise, one or more regions of the endogenous Ttr locus encoding one or more domains of the transthyretin precursor protein can remain unmodified (i.e., not deleted and replaced). For example, transthyretin precursor proteins typically have a signal peptide at the N-terminus. The signal peptide can be, for example, about 20 amino acids in length. The region of the endogenous Ttr locus encoding the signal peptide can remain unmodified (i.e., not deleted and replaced), or can be deleted and replaced with the corresponding orthologous human TTR sequence. Similarly, a region of the endogenous Ttr locus encoding an epitope recognized by an anti-human-TTR antigen-binding protein can be humanized.

[0122] Depending on the extent of replacement by corresponding orthologous sequences or the extent of insertion of human TTR sequences, regulatory sequences, such as a promoter, can be endogenous or supplied by the replacing or inserted human orthologous sequence. For example, the humanized TTR locus can include the endogenous non-human animal Ttr promoter. The coding sequence for the transthyretin precursor protein at the genetically modified endogenous Ttr locus can be operably linked to the endogenous Ttr promoter. For example, the human TTR sequence can be operably linked to the endogenous Ttr promoter.

[0123] As a specific example, the humanized TTR locus can be one in which the region of the endogenous Ttr locus being deleted and replaced with the corresponding orthologous human TTR sequence or the region of the human TTR locus being inserted comprises, consists essentially of, or consists of the region from the Ttr start codon to the stop codon. The human TTR sequence being inserted can further comprise a human TTR 3' UTR. For example, the human TTR sequence at the humanized TTR locus can comprise, consist essentially of, or consist of the region from the TTR start codon to the end of the 3' UTR. Optionally, the Ttr coding sequence in the modified endogenous Ttr locus is operably linked to the endogenous Ttr promoter. The human TTR sequence at the humanized TTR locus can comprise, consist essentially of, or consist of a sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 18. The humanized TTR locus can comprise, consist essentially of, or consist of a sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 14 or 15. The coding sequence (CDS) at the humanized TTR locus can comprise, consist essentially of, or consist of a sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 90 (or degenerates thereof that encode the same protein). The resulting human transthyretin precursor protein encoded by the humanized TTR locus can comprise, consist essentially of, or consist of a sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 1.

[0124] As another specific example, the humanized TTR locus can be one in which the region of the endogenous Ttr locus being deleted and replaced with the corresponding orthologous human TTR sequence or the region of the human TTR locus being inserted comprises, consists essentially of, or consists of the region from the start of the second Ttr exon to the stop codon. The human TTR sequence being

inserted can further comprise a human TTR 3' UTR. For example, the human TTR sequence at the humanized TTR locus can comprise, consist essentially of, or consist of the region from the start of the second human TTR exon to the end of the 3' UTR. Optionally, the Ttr coding sequence in the modified endogenous Ttr locus is operably linked to the endogenous Ttr promoter. The human TTR sequence at the humanized TTR locus can comprise, consist essentially of, or consist of a sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 19. The humanized TTR locus can comprise, consist essentially of, or consist of a sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 16 or 17. The coding sequence (CDS) at the humanized TTR locus can comprise, consist essentially of, or consist of a sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 91 (or degenerates thereof that encode the same protein). The resulting chimeric transthyretin precursor protein encoded by the humanized TTR locus can comprise, consist essentially of, or consist of a sequence that is at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 2.

[0125] TTR protein expressed from a humanized TTR locus can be an entirely human TTR protein or a chimeric endogenous/human TTR protein (e.g., if the non-human animal is a mouse, a chimeric mouse/human TTR protein). For example, the signal peptide of the transthyretin precursor protein can be endogenous, and the remainder of the protein can be human. Alternatively, the N-terminus of the transthyretin precursor protein can be endogenous, and the remainder of the protein can be human. For example, the N-terminal 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 amino acids can be endogenous, and the remainder can be human. In a specific example, the 23 amino acids at the N-terminus are endogenous, and the remainder of the protein is human.

[0126] Optionally, a humanized TTR locus can comprise other elements. Examples of such elements can include selection cassettes, reporter genes, recombinase recognition sites, or other elements. As one example, a humanized TTR locus can comprise a removable selection cassette (e.g., a self-deleting selection cassette) flanked by recombinase recognition sequences (e.g., loxP sites). Alternatively, the humanized TTR locus can lack other elements (e.g., can lack a selection cassette and/or can lack a reporter gene). Examples of suitable reporter genes and reporter proteins are disclosed elsewhere herein. Examples of suitable selection markers include neomycin phosphotransferase (*neo_r*), hygromycin B phosphotransferase (*hyg_r*), puromycin-N-acetyltransferase (*puro_r*), blasticidin S deaminase (*bsr_r*), xanthine/guanine phosphoribosyl transferase (*gpt*), and herpes simplex virus thymidine kinase (*HSV-k*). Examples of recombinases include Cre, Flp, and Dre recombinases. One example of a Cre recombinase gene is Crei, in which two exons encoding the Cre recombinase are separated by an intron to prevent its expression in a prokaryotic cell. Such recombinases can further comprise a nuclear localization signal to facilitate localization to the nucleus (e.g., NLS-Crei). Recombinase recognition sites include nucleotide sequences that are recognized by a site-specific recombinase and can serve as a substrate for a recombination event. Examples of recombinase recognition sites include FRT, FRT11, FRT171, attp, att, rox, and lox sites such as loxP, lox511, lox2272, lox66, lox71, loxM2, and lox5171.

[0127] Other elements such as reporter genes or selection cassettes can be self-deleting cassettes flanked by recombinase recognition sites. See, e.g., U.S. Pat. No. 8,697,851 and US 2013/0312129, each of which is herein incorporated by reference in its entirety for all purposes. As an example, the self-deleting cassette can comprise a Crei gene (comprises two exons encoding a Cre recombinase, which are separated by an intron) operably linked to a mouse Prm1 promoter and a neomycin resistance gene operably linked to a human ubiquitin promoter. By employing the Prm1 promoter, the self-deleting cassette can be deleted specifically in male germ cells of F0 animals. The polynucleotide encoding the selection marker can be operably linked to a promoter active in a cell being targeted. Examples of promoters are described elsewhere herein. As another specific example, a self-deleting selection cassette can comprise a hygromycin resistance gene coding sequence operably linked to one or more promoters (e.g., both human ubiquitin and EM7 promoters) followed by a polyadenylation signal, followed by a Crei coding sequence operably linked to one or more promoters (e.g., an mPrm1 promoter), followed by another polyadenylation signal, wherein the entire cassette is flanked by loxP sites.

[0128] The humanized TTR locus can also be a conditional allele. For example, the conditional allele can be a multifunctional allele, as described in US 2011/0104799, herein incorporated by reference in its entirety for all purposes. For example, the conditional allele can comprise: (a) an actuating sequence in sense orientation with respect to transcription of a target gene; (b) a drug selection cassette (DSC) in sense or antisense orientation; (c) a nucleotide sequence of interest (NSI) in antisense orientation; and (d) a conditional by inversion module (COIN, which utilizes an exon-splitting intron and an invertible gene-trap-like module) in reverse orientation. See, e.g., US 2011/0104799. The conditional allele can further comprise recombinable units that recombine upon exposure to a first recombinase to form a conditional allele that (i) lacks the actuating sequence and the DSC; and (ii) contains the NSI in sense orientation and the COIN in antisense orientation. See, e.g., US 2011/0104799.

[0129] C. Non-Human Animal Genomes, Non-Human Animal Cells, and Non-Human Animals Comprising a Humanized TTR Locus

[0130] Non-human animal genomes, non-human animal cells, and non-human animals comprising a humanized TTR locus as described elsewhere herein are provided. The genomes, cells, or non-human animals can be heterozygous or homozygous for the humanized TTR locus. A diploid organism has two alleles at each genetic locus. Each pair of alleles represents the genotype of a specific genetic locus. Genotypes are described as homozygous if there are two identical alleles at a particular locus and as heterozygous if the two alleles differ.

[0131] The non-human animal genomes or cells provided herein can be, for example, any non-human cell comprising a Ttr locus or a genomic locus homologous or orthologous to the human TTR locus. The genomes can be from or the cells can be eukaryotic cells, which include, for example, animal cells, mammalian cells, non-human mammalian cells, and human cells. The term "animal" includes mammals, fishes, and birds. A mammalian cell can be, for example, a non-human mammalian cell, a rodent cell, a rat cell, a mouse cell, or a hamster cell. Other non-human

mammals include, for example, non-human primates, monkeys, apes, orangutans, cats, dogs, rabbits, horses, livestock (e.g., bovine species such as cows, steer, and so forth; ovine species such as sheep, goats, and so forth; and porcine species such as pigs and boars). Domesticated animals and agricultural animals are also included. The term “non-human” excludes humans.

[0132] The cells can also be any type of undifferentiated or differentiated state. For example, a cell can be a totipotent cell, a pluripotent cell (e.g., a human pluripotent cell or a non-human pluripotent cell such as a mouse embryonic stem (ES) cell or a rat ES cell), or a non-pluripotent cell. Totipotent cells include undifferentiated cells that can give rise to any cell type, and pluripotent cells include undifferentiated cells that possess the ability to develop into more than one differentiated cell types. Such pluripotent and/or totipotent cells can be, for example, ES cells or ES-like cells, such as an induced pluripotent stem (iPS) cells. ES cells include embryo-derived totipotent or pluripotent cells that are capable of contributing to any tissue of the developing embryo upon introduction into an embryo. ES cells can be derived from the inner cell mass of a blastocyst and are capable of differentiating into cells of any of the three vertebrate germ layers (endoderm, ectoderm, and mesoderm).

[0133] The cells provided herein can also be germ cells (e.g., sperm or oocytes). The cells can be mitotically competent cells or mitotically-inactive cells, meiotically competent cells or meiotically-inactive cells. Similarly, the cells can also be primary somatic cells or cells that are not a primary somatic cell. Somatic cells include any cell that is not a gamete, germ cell, gametocyte, or undifferentiated stem cell. For example, the cells can be liver cells, such as hepatoblasts or hepatocytes.

[0134] Suitable cells provided herein also include primary cells. Primary cells include cells or cultures of cells that have been isolated directly from an organism, organ, or tissue. Primary cells include cells that are neither transformed nor immortal. They include any cell obtained from an organism, organ, or tissue which was not previously passed in tissue culture or has been previously passed in tissue culture but is incapable of being indefinitely passed in tissue culture. Such cells can be isolated by conventional techniques and include, for example, hepatocytes.

[0135] Other suitable cells provided herein include immortalized cells. Immortalized cells include cells from a multicellular organism that would normally not proliferate indefinitely but, due to mutation or alteration, have evaded normal cellular senescence and instead can keep undergoing division. Such mutations or alterations can occur naturally or be intentionally induced. A specific example of an immortalized cell line is the HepG2 human liver cancer cell line. Numerous types of immortalized cells are well known. Immortalized or primary cells include cells that are typically used for culturing or for expressing recombinant genes or proteins.

[0136] The cells provided herein also include one-cell stage embryos (i.e., fertilized oocytes or zygotes). Such one-cell stage embryos can be from any genetic background (e.g., BALB/c, C57BL/6, 129, or a combination thereof for mice), can be fresh or frozen, and can be derived from natural breeding or in vitro fertilization.

[0137] The cells provided herein can be normal, healthy cells, or can be diseased or mutant-bearing cells.

[0138] In a specific example, the non-human animal cells are embryonic stem (ES) cells or liver cells, such as mouse or rat ES cells or liver cells.

[0139] Non-human animals comprising a humanized TTR locus as described herein can be made by the methods described elsewhere herein. The term “animal” includes mammals, fishes, and birds. Non-human mammals include, for example, non-human primates, monkeys, apes, orangutans, cats, dogs, horses, rabbits, rodents (e.g., mice, rats, hamsters, and guinea pigs), and livestock (e.g., bovine species such as cows and steer; ovine species such as sheep and goats; and porcine species such as pigs and boars). Domesticated animals and agricultural animals are also included. The term “non-human animal” excludes humans. Preferred non-human animals include, for example, rodents, such as mice and rats.

[0140] The non-human animals can be from any genetic background. For example, suitable mice can be from a 129 strain, a C57BL/6 strain, a mix of 129 and C57BL/6, a BALB/c strain, or a Swiss Webster strain. Examples of 129 strains include 129P1, 129P2, 129P3, 129X1, 129S1 (e.g., 129S1/SV, 129S1/SvIm), 129S2, 129S4, 129S5, 129S9/SvEvH, 129S6 (129/SvEvTac), 129S7, 129S8, 129T1, and 129T2. See, e.g., Festing et al. (1999) *Mammalian Genome* 10:836, herein incorporated by reference in its entirety for all purposes. Examples of C57BL strains include C57BL/A, C57BL/An, C57BL/GrFa, C57BL/Kal_wN, C57BL/6, C57BL/6J, C57BL/6ByJ, C57BL/6NJ, C57BL/10, C57BL/10ScSn, C57BL/10Cr, and C57BL/O1a. Suitable mice can also be from a mix of an aforementioned 129 strain and an aforementioned C57BL/6 strain (e.g., 50% 129 and 50% C57BL/6). Likewise, suitable mice can be from a mix of aforementioned 129 strains or a mix of aforementioned BL/6 strains (e.g., the 129S6 (129/SvEvTac) strain).

[0141] Similarly, rats can be from any rat strain, including, for example, an ACI rat strain, a Dark Agouti (DA) rat strain, a Wistar rat strain, a LEA rat strain, a Sprague Dawley (SD) rat strain, or a Fischer rat strain such as Fisher F344 or Fisher F6. Rats can also be obtained from a strain derived from a mix of two or more strains recited above. For example, a suitable rat can be from a DA strain or an ACI strain. The ACI rat strain is characterized as having black agouti, with white belly and feet and an RT1^{av1} haplotype. Such strains are available from a variety of sources including Harlan Laboratories. The Dark Agouti (DA) rat strain is characterized as having an agouti coat and an RT1^{av1} haplotype. Such rats are available from a variety of sources including Charles River and Harlan Laboratories. Some suitable rats can be from an inbred rat strain. See, e.g., US 2014/0235933, herein incorporated by reference in its entirety for all purposes.

III. Methods of Using Non-Human Animals Comprising a Humanized TTR Locus for Assessing Efficacy of Human-TTR-Targeting Reagents In Vivo or Ex Vivo

[0142] Various methods are provided for using the non-human animals comprising a humanized TTR locus as described elsewhere herein for assessing or optimizing delivery or efficacy of human-TTR-targeting reagents (e.g., therapeutic molecules or complexes) in vivo or ex vivo. Because the non-human animals comprise a humanized TTR locus, the non-human animals will more accurately reflect the efficacy of a human TTR-targeting reagent. Such non-human animals are particularly useful for testing genome-editing reagents designed to target the human TTR gene

because the non-human animals disclosed herein comprise humanized endogenous Ttr loci rather than transgenic insertions of human TTR sequence at random genomic loci, and the humanized endogenous Ttr loci comprise orthologous human genomic TTR sequence from both coding and non-coding regions rather than an artificial cDNA sequence.

[0143] A. Methods of Testing Efficacy of Human-TTR-Targeting Reagents In Vivo or Ex Vivo

[0144] Various methods are provided for assessing delivery or efficacy of human-TTR-targeting reagents in vivo using non-human animals comprising a humanized TTR locus as described elsewhere herein. Such methods can comprise: (a) introducing into the non-human animal a human-TTR-targeting reagent (i.e., administering the human-TTR-targeting reagent to the non-human animal); and (b) assessing the activity of the human-TTR-targeting reagent.

[0145] The human-TTR-targeting reagent can be any biological or chemical agent that targets the human TTR locus (the human TTR gene), the human TTR mRNA, or the human transthyretin protein. Examples of human-TTR-targeting reagents are disclosed elsewhere herein and include, for example, genome-editing agents. For example, the human-TTR-targeting reagent can be a TTR-targeting nucleic acid (e.g., CRISPR/Cas guide RNAs, short hairpin RNAs (shRNAs), or small interfering RNAs (siRNAs)) or nucleic acid encoding a TTR-targeting protein (e.g., a Cas proteins such as Cas9, a ZFN, or a TALEN). Alternatively, the human-TTR-targeting reagent can be a TTR-targeting antibody or antigen-binding protein, or any other large molecule or small molecule that targets human TTR.

[0146] Such human-TTR-targeting reagents can be administered by any delivery method (e.g., AAV, LNP, or HDD) as disclosed in more detail elsewhere herein and by any route of administration. Means of delivering therapeutic complexes and molecules and routes of administration are disclosed in more detail elsewhere herein. In particular methods, the reagents delivered via AAV-mediated delivery. For example, AAV8 can be used to target the liver. In other particular methods, the reagents are delivered by LNP-mediated delivery. In other particular methods, the reagents are delivered by hydrodynamic delivery (HDD). The dose can be any suitable dose. For example, in some methods in which the reagents (e.g., Cas9 mRNA and gRNA) are delivered by LNP-mediated delivery, the dose can be between about 0.01 and about 10 mg/kg, about 0.01 and about 5 mg/kg, between about 0.01 and about 4 mg/kg, between about 0.01 and about 3 mg/kg, between about 0.01 and about 2 mg/kg, between about 0.01 and about 1 mg/kg, between about 0.1 and about 10 mg/kg, between about 0.1 and about 6 mg/kg; between about 0.1 and about 5 mg/kg, between about 0.1 and about 4 mg/kg, between about 0.1 and about 3 mg/kg, between about 0.1 and about 2 mg/kg, between about 0.1 and about 1 mg/kg, between about 0.3 and about 10 mg/kg, between about 0.3 and about 6 mg/kg; between about 0.3 and about 5 mg/kg, between about 0.3 and about 4 mg/kg, between about 0.3 and about 3 mg/kg, between about 0.3 and about 2 mg/kg, between about 0.3 and about 1 mg/kg, about 0.1 mg/kg, about 0.3 mg/kg, about 1 mg/kg, about 2 mg/kg, or about 3 mg/kg. In a specific example, the dose is between about 0.1 and about 6 mg/kg; between about 0.1 and about 3 mg/kg, or between about 0.1 and about 2 mg/kg. In a specific example, the human-TTR-targeting reagent is a genome editing reagent, the LNP dose

is about 1 mg/kg, and the percent genome editing at the humanized TTR locus is between about 70% and about 80%. In another specific example, the human-TTR-targeting reagent is a genome editing reagent, the LNP dose is about 0.3 mg/kg, and the percent editing is between about 50% and about 80%. In another specific example, the human-TTR-targeting reagent is a genome editing reagent, the LNP dose is about 0.1 mg/kg, and the percent editing is between about 20% and about 80%. In another specific example, the LNP dose is about 1 mg/kg, and the serum TTR levels are reduced to between about 0% and about 10% or between about 0% and about 35% of control levels. In another specific example, the LNP dose is about 0.3 mg/kg, and the serum TTR levels are reduced to between about 0% and about 20% or about 0% and about 95% of control levels. In another specific example, the LNP dose is about 0.1 mg/kg, and the serum TTR levels are reduced to between about 0% and about 60% or about 0% and about 99% of control levels.

[0147] Methods for assessing activity of the human-TTR-targeting reagent are well-known and are provided elsewhere herein. Assessment of activity can be in any cell type, any tissue type, or any organ type as disclosed elsewhere herein. In some methods, assessment of activity is in liver cells. If the TTR-targeting reagent is a genome editing reagent (e.g., a nuclease agent), such methods can comprise assessing modification of the humanized TTR locus. As one example, the assessing can comprise measuring non-homologous end joining (NHEJ) activity at the humanized TTR locus. This can comprise, for example, measuring the frequency of insertions or deletions within the humanized TTR locus. For example, the assessing can comprise sequencing the humanized TTR locus in one or more cells isolated from the non-human animal (e.g., next-generation sequencing). Assessment can comprise isolating a target organ (e.g., liver) or tissue from the non-human animal and assessing modification of humanized TTR locus in the target organ or tissue. Assessment can also comprise assessing modification of humanized TTR locus in two or more different cell types within the target organ or tissue. Similarly, assessment can comprise isolating a non-target organ or tissue (e.g., two or more non-target organs or tissues) from the non-human animal and assessing modification of humanized TTR locus in the non-target organ or tissue.

[0148] Such methods can also comprise measuring expression levels of the mRNA produced by the humanized TTR locus, or by measuring expression levels of the protein encoded by the humanized TTR locus. For example, protein levels can be measured in a particular cell, tissue, or organ type (e.g., liver), or secreted levels can be measured in the serum. Methods for assessing expression of Ttr mRNA or protein expressed from the humanized TTR locus are provided elsewhere herein and are well-known.

[0149] The various methods provided above for assessing activity in vivo can also be used to assess the activity of human-TTR-targeting reagents ex vivo as described elsewhere herein.

[0150] As one example, if the human-TTR-targeting reagent is a genome editing reagent (e.g., a nuclease agent), percent editing at the humanized TTR locus can be assessed (e.g., in liver cells). For example, the percent editing (e.g., total number of insertions or deletions observed over the total number of sequences read in the PCR reaction from a pool of lysed cells) can be at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about

50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 99%, or, for example, between about 1% and about 99%, between about 10% and about 99%, between about 20% and about 99%, between about 30% and about 99%, between about 40% and about 99%, between about 50% and about 99%, between about 60% and about 99%, between about 1% and about 90%, between about 10% and about 90%, between about 20% and about 90%, between about 30% and about 90%, between about 40% and about 90%, between about 50% and about 90%, between about 60% and about 90%, between about 1% and about 80%, between about 10% and about 80%, between about 20% and about 80%, between about 30% and about 80%, between about 40% and about 80%, between about 50% and about 80%, or between about 60% and about 80%.

[0151] As another example, serum TTR levels can be assessed. For example, serum TTR levels can be reduced by at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 65%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 99%, or, for example, between about 1% and about 99%, between about 10% and about 99%, between about 20% and about 99%, between about 30% and about 99%, between about 40% and about 99%, between about 50% and about 99%, between about 60% and about 99%, between about 70% and about 99%, between about 80% and about 99%, between about 1% and about 90%, between about 10% and about 90%, between about 20% and about 90%, between about 30% and about 90%, between about 40% and about 90%, between about 50% and about 90%, between about 60% and about 90%, between about 70% and about 90%, or between about 80% and about 90%.

[0152] In some methods, the human-TTR-targeting reagent is a nuclease agent, such as a CRISPR/Cas nuclease agent, that targets the human TTR gene. Such methods can comprise, for example: (a) introducing into the non-human animal a nuclease agent designed to cleave the human TTR gene (e.g., Cas protein such as Cas9 and a guide RNA designed to target a guide RNA target sequence in the human TTR gene); and (b) assessing modification of the humanized TTR locus.

[0153] In the case of a CRISPR/Cas nuclease, for example, modification of the humanized TTR locus will be induced when the guide RNA forms a complex with the Cas protein and directs the Cas protein to the humanized TTR locus, and the Cas/guide RNA complex cleaves the guide RNA target sequence, triggering repair by the cell (e.g., via non-homologous end joining (NHEJ) if no donor sequence is present).

[0154] Optionally, two or more guide RNAs can be introduced, each designed to target a different guide RNA target sequence within the human TTR gene. For example, two guide RNAs can be designed to excise a genomic sequence between the two guide RNA target sequences. Modification of the humanized TTR locus will be induced when the first guide RNA forms a complex with the Cas protein and directs the Cas protein to the humanized TTR locus, the second guide RNA forms a complex with the Cas protein and directs the Cas protein to the humanized TTR locus, the first Cas/guide RNA complex cleaves the first guide RNA target

sequence, and the second Cas/guide RNA complex cleaves the second guide RNA target sequence, resulting in excision of the intervening sequence.

[0155] Optionally, an exogenous donor nucleic acid capable of recombining with and modifying a human TTR gene is also introduced into the non-human animal. Optionally, the nuclease agent or Cas protein can be tethered to the exogenous donor nucleic acid as described elsewhere herein. Modification of the humanized TTR locus will be induced, for example, when the guide RNA forms a complex with the Cas protein and directs the Cas protein to the humanized TTR locus, the Cas/guide RNA complex cleaves the guide RNA target sequence, and the humanized TTR locus recombines with the exogenous donor nucleic acid to modify the humanized TTR locus. The exogenous donor nucleic acid can recombine with the humanized TTR locus, for example, via homology-directed repair (HDR) or via NHEJ-mediated insertion. Any type of exogenous donor nucleic acid can be used, examples of which are provided elsewhere herein.

[0156] B. Methods of Optimizing Delivery or Efficacy of Human-TTR-Targeting Reagent In Vivo or Ex Vivo

[0157] Various methods are provided for optimizing delivery of human-TTR-targeting reagents to a cell or non-human animal or optimizing the activity or efficacy of human-TTR-targeting reagents in vivo. Such methods can comprise, for example: (a) performing the method of testing the efficacy of a human-TTR-targeting reagent as described above a first time in a first non-human animal or first cell comprising a humanized TTR locus; (b) changing a variable and performing the method a second time in a second non-human animal (i.e., of the same species) or a second cell comprising a humanized TTR locus with the changed variable; and (c) comparing the activity of the human-TTR-targeting reagent in step (a) with the activity of the human-TTR-targeting reagent in step (b), and selecting the method resulting in the higher activity.

[0158] Methods of measuring delivery, efficacy, or activity of human-TTR-targeting reagents are disclosed elsewhere herein. For example, such methods can comprise measuring modification of the humanized TTR locus. More effective modification of the humanized TTR locus can mean different things depending on the desired effect within the non-human animal or cell. For example, more effective modification of the humanized TTR locus can mean one or more or all of higher levels of modification, higher precision, higher consistency, or higher specificity. Higher levels of modification (i.e., higher efficacy) of the humanized TTR locus refers to a higher percentage of cells is targeted within a particular target cell type, within a particular target tissue, or within a particular target organ (e.g., liver). Higher precision refers to more precise modification of the humanized TTR locus (e.g., a higher percentage of targeted cells having the same modification or having the desired modification without extra unintended insertions and deletions (e.g., NHEJ indels)). Higher consistency refers to more consistent modification of the humanized TTR locus among different types of targeted cells, tissues, or organs if more than one type of cell, tissue, or organ is being targeted (e.g., modification of a greater number of cell types within the liver). If a particular organ is being targeted, higher consistency can also refer to more consistent modification throughout all locations within the organ (e.g., the liver). Higher specificity can refer to higher specificity with respect to the genomic locus or loci targeted, higher specificity with

respect to the cell type targeted, higher specificity with respect to the tissue type targeted, or higher specificity with respect to the organ targeted. For example, increased genomic locus specificity refers to less modification of off-target genomic loci (e.g., a lower percentage of targeted cells having modifications at unintended, off-target genomic loci instead of or in addition to modification of the target genomic locus). Likewise, increased cell type, tissue, or organ type specificity refers to less modification of off-target cell types, tissue types, or organ types if a particular cell type, tissue type, or organ type is being targeted (e.g., when a particular organ is targeted (e.g., the liver), there is less modification of cells in organs or tissues that are not intended targets).

[0159] The variable that is changed can be any parameter. As one example, the changed variable can be the packaging or the delivery method by which the human-TTR-targeting reagent or reagents are introduced into the cell or non-human animal. Examples of delivery methods, such as LNP, HDD, and AAV, are disclosed elsewhere herein. For example, the changed variable can be the AAV serotype. Similarly, the administering can comprise LNP-mediated delivery, and the changed variable can be the LNP formulation. As another example, the changed variable can be the route of administration for introduction of the human-TTR-targeting reagent or reagents into the cell or non-human animal. Examples of routes of administration, such as intravenous, intravitreal, intraparenchymal, and nasal instillation, are disclosed elsewhere herein.

[0160] As another example, the changed variable can be the concentration or amount of the human-TTR-targeting reagent or reagents introduced. As another example, the changed variable can be the concentration or the amount of one human-TTR-targeting reagent introduced (e.g., guide RNA, Cas protein, or exogenous donor nucleic acid) relative to the concentration or the amount another human-TTR-targeting reagent introduced (e.g., guide RNA, Cas protein, or exogenous donor nucleic acid).

[0161] As another example, the changed variable can be the timing of introducing the human-TTR-targeting reagent or reagents relative to the timing of assessing the activity or efficacy of the reagents. As another example, the changed variable can be the number of times or frequency with which the human-TTR-targeting reagent or reagents are introduced. As another example, the changed variable can be the timing of introduction of one human-TTR-targeting reagent introduced (e.g., guide RNA, Cas protein, or exogenous donor nucleic acid) relative to the timing of introduction of another human-TTR-targeting reagent introduced (e.g., guide RNA, Cas protein, or exogenous donor nucleic acid).

[0162] As another example, the changed variable can be the form in which the human-TTR-targeting reagent or reagents are introduced. For example, a guide RNA can be introduced in the form of DNA or in the form of RNA. A Cas protein (e.g., Cas9) can be introduced in the form of DNA, in the form of RNA, or in the form of a protein (e.g., complexed with a guide RNA). An exogenous donor nucleic acid can be DNA, RNA, single-stranded, double-stranded, linear, circular, and so forth. Similarly, each of the components can comprise various combinations of modifications for stability, to reduce off-target effects, to facilitate delivery, and so forth. As another example, the changed variable can be the human-TTR-targeting reagent or reagents that are introduced (e.g., introducing a different guide RNA with a

different sequence, introducing a different Cas protein (e.g., introducing a different Cas protein with a different sequence, or a nucleic acid with a different sequence but encoding the same Cas protein amino acid sequence), or introducing a different exogenous donor nucleic acid with a different sequence).

[0163] In a specific example, the human-TTR-targeting reagent comprises a Cas protein and a guide RNA designed to target a guide RNA target sequence in a human TTR gene. In such methods, the changed variable can be the guide RNA sequence and/or the guide RNA target sequence. In some such methods, the Cas protein and the guide RNA can each be administered in the form of RNA, and the changed variable can be the ratio of Cas mRNA to guide RNA (e.g., in an LNP formulation). In some such methods, the changed variable can be guide RNA modifications (e.g., a guide RNA with a modification is compared to a guide RNA without the modification).

[0164] C. Human-TTR-Targeting Reagents

[0165] A human-TTR-targeting reagent can be any reagent that targets a human TTR gene, a human TTR mRNA, or a human TTR protein. For example, it can be a genome editing reagent such as a nuclease agent that cleaves a target sequence within the human TTR gene, it can be an antisense oligonucleotide targeting a human TTR mRNA, it can be an antigen-binding protein targeting an epitope of a human TTR protein, or it can be a small molecule targeting human TTR. Human-TTR-targeting reagents in the methods disclosed herein can be known human-TTR-targeting reagents, can be putative-TTR-targeting reagents (e.g., candidate reagents designed to target human TTR), or can be reagents being screened for human-TTR-targeting activity.

[0166] (1) Nuclease Agents Targeting Human TTR Gene

[0167] A human-TTR-targeting reagent can be a genome editing reagent such as a nuclease agent that cleaves a target sequence within the human TTR gene. A nuclease target sequence includes a DNA sequence at which a nick or double-strand break is induced by a nuclease agent. The target sequence for a nuclease agent can be endogenous (or native) to the cell or the target sequence can be exogenous to the cell. A target sequence that is exogenous to the cell is not naturally occurring in the genome of the cell. The target sequence can also be exogenous to the polynucleotides of interest that one desires to be positioned at the target locus. In some cases, the target sequence is present only once in the genome of the host cell.

[0168] The length of the target sequence can vary, and includes, for example, target sequences that are about 30-36 bp for a zinc finger nuclease (ZFN) pair (i.e., about 15-18 bp for each ZFN), about 36 bp for a Transcription Activator-Like Effector Nuclease (TALEN), or about 20 bp for a CRISPR/Cas9 guide RNA.

[0169] Any nuclease agent that induces a nick or double-strand break at a desired target sequence can be used in the methods and compositions disclosed herein. A naturally occurring or native nuclease agent can be employed so long as the nuclease agent induces a nick or double-strand break in a desired target sequence. Alternatively, a modified or engineered nuclease agent can be employed. An "engineered nuclease agent" includes a nuclease that is engineered (modified or derived) from its native form to specifically recognize and induce a nick or double-strand break in the desired target sequence. Thus, an engineered nuclease agent can be derived from a native, naturally occurring nuclease

agent or it can be artificially created or synthesized. The engineered nuclease can induce a nick or double-strand break in a target sequence, for example, wherein the target sequence is not a sequence that would have been recognized by a native (non-engineered or non-modified) nuclease agent. The modification of the nuclease agent can be as little as one amino acid in a protein cleavage agent or one nucleotide in a nucleic acid cleavage agent. Producing a nick or double-strand break in a target sequence or other DNA can be referred to herein as “cutting” or “cleaving” the target sequence or other DNA.

[0170] Active variants and fragments of the exemplified target sequences are also provided. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the given target sequence, wherein the active variants retain biological activity and hence are capable of being recognized and cleaved by a nuclease agent in a sequence-specific manner. Assays to measure the double-strand break of a target sequence by a nuclease agent are well-known. See, e.g., Frendewey et al. (2010) *Methods in Enzymology* 476:295-307, which is incorporated by reference herein in its entirety for all purposes.

[0171] The target sequence of the nuclease agent can be positioned anywhere in or near the Ttr locus. The target sequence can be located within a coding region of the Ttr gene, or within regulatory regions that influence the expression of the gene. A target sequence of the nuclease agent can be located in an intron, a promoter, an enhancer, a regulatory region, or any non-protein coding region.

[0172] One type of nuclease agent is a Transcription Activator-Like Effector Nuclease (TALEN). TAL effector nucleases are a class of sequence-specific nucleases that can be used to make double-strand breaks at specific target sequences in the genome of a prokaryotic or eukaryotic organism. TAL effector nucleases are created by fusing a native or engineered transcription activator-like (TAL) effector, or functional part thereof, to the catalytic domain of an endonuclease, such as, for example, FokI. The unique, modular TAL effector DNA binding domain allows for the design of proteins with potentially any given DNA recognition specificity. Thus, the DNA binding domains of the TAL effector nucleases can be engineered to recognize specific DNA target sites and thus, used to make double-strand breaks at desired target sequences. See WO 2010/079430; Morbitzer et al. (2010) *PNAS* 10.1073/pnas.1013133107; Scholze & Boch (2010) *Virulence* 1:428-432; Christian et al. *Genetics* (2010) 186:757-761; Li et al. (2010) *Nuc. Acids Res.* (2010) doi:10.1093/nar/gkq704; and Miller et al. (2011) *Nature Biotechnology* 29:143-148, each of which is herein incorporated by reference in its entirety.

[0173] Examples of suitable TAL nucleases, and methods for preparing suitable TAL nucleases, are disclosed, e.g., in US 2011/0239315 A1, US 2011/0269234 A1, US 2011/0145940 A1, US 2003/0232410 A1, US 2005/0208489 A1, US 2005/0026157 A1, US 2005/0064474 A1, US 2006/0188987 A1, and US 2006/0063231 A1, each of which is herein incorporated by reference in its entirety. In various embodiments, TAL effector nucleases are engineered that cut in or near a target nucleic acid sequence in, e.g., a locus of interest or a genomic locus of interest, wherein the target nucleic acid sequence is at or near a sequence to be modified by a targeting vector. The TAL nucleases suitable for use with the various methods and compositions provided herein

include those that are specifically designed to bind at or near target nucleic acid sequences to be modified by targeting vectors as described herein.

[0174] In some TALENs, each monomer of the TALEN comprises 33-35 TAL repeats that recognize a single base pair via two hypervariable residues. In some TALENs, the nuclease agent is a chimeric protein comprising a TAL-repeat-based DNA binding domain operably linked to an independent nuclease such as a FokI endonuclease. For example, the nuclease agent can comprise a first TAL-repeat-based DNA binding domain and a second TAL-repeat-based DNA binding domain, wherein each of the first and the second TAL-repeat-based DNA binding domains is operably linked to a FokI nuclease, wherein the first and the second TAL-repeat-based DNA binding domain recognize two contiguous target DNA sequences in each strand of the target DNA sequence separated by a spacer sequence of varying length (12-20 bp), and wherein the FokI nuclease subunits dimerize to create an active nuclease that makes a double strand break at a target sequence.

[0175] The nuclease agent employed in the various methods and compositions disclosed herein can further comprise a zinc-finger nuclease (ZFN). In some ZFNs, each monomer of the ZFN comprises 3 or more zinc finger-based DNA binding domains, wherein each zinc finger-based DNA binding domain binds to a 3 bp subsite. In other ZFNs, the ZFN is a chimeric protein comprising a zinc finger-based DNA binding domain operably linked to an independent nuclease such as a FokI endonuclease. For example, the nuclease agent can comprise a first ZFN and a second ZFN, wherein each of the first ZFN and the second ZFN is operably linked to a FokI nuclease subunit, wherein the first and the second ZFN recognize two contiguous target DNA sequences in each strand of the target DNA sequence separated by about 5-7 bp spacer, and wherein the FokI nuclease subunits dimerize to create an active nuclease that makes a double strand break. See, e.g., US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308.A2; US20130123484; US20100291048; WO/2011/017293.A2; and Gaj et al. (2013) *Trends in Biotechnology*, 31(7):397-405, each of which is herein incorporated by reference.

[0176] Another type of nuclease agent is a meganuclease. Meganucleases have been classified into four families based on conserved sequence motifs, the families are the LAGLI-DADG, GIY-YIG, H-N-H, and His-Cys box families. These motifs participate in the coordination of metal ions and hydrolysis of phosphodiester bonds. Meganucleases are notable for their long target sequences, and for tolerating some sequence polymorphisms in their DNA substrates. Meganuclease domains, structure and function are known, see for example, Guhan and Muniyappa (2003) *Crit Rev Biochem Mol Biol* 38:199-248; Lucas et al., (2001) *Nucleic Acids Res* 29:960-9; Jurica and Stoddard, (1999) *Cell Mol Life Sci* 55:1304-26; Stoddard, (2006) *Q Rev Biophys* 38:49-95; and Moure et al., (2002) *Nat Struct Biol* 9:764. In some examples, a naturally occurring variant and/or engineered derivative meganuclease is used. Methods for modifying the kinetics, cofactor interactions, expression, optimal conditions, and/or target sequence specificity, and screening for activity are known. See, e.g., Epinat et al., (2003) *Nucleic Acids Res* 31:2952-62; Chevalier et al., (2002) *Mol Cell* 10:895-905; Gimble et al., (2003) *Mol Biol* 334:993-1008; Seligman et al., (2002) *Nucleic Acids Res* 30:3870-9; Suss-

man et al., (2004) *J Mol Biol* 342:31-41; Rosen et al., (2006) *Nucleic Acids Res* 34:4791-800; Chames et al., (2005) *Nucleic Acids Res* 33:e178; Smith et al., (2006) *Nucleic Acids Res* 34:e149; Gruen et al., (2002) *Nucleic Acids Res* 30:e29; Chen and Zhao, (2005) *Nucleic Acids Res* 33:e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; and WO2004031346, each of which is herein incorporated by reference in its entirety.

[0177] Any meganuclease can be used, including, for example, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TII, I-PpoI, PI-PspI, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-Amal, I-AniI, I-ChuI, I-Cmoel, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaI, I-NanI, I-NciIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-Sp-BetaIP, I-Scal, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIIP, PI-MtuHIIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TII, PI-TIIII, or any active variants or fragments thereof.

[0178] Meganucleases can recognize, for example, double-stranded DNA sequences of 12 to 40 base pairs. In some cases, the meganuclease recognizes one perfectly matched target sequence in the genome.

[0179] Some meganucleases are homing nucleases. One type of homing nuclease is a LAGLIDADG family of homing nucleases including, for example, I-SceI, I-CreI, and I-DmoI.

[0180] Nuclease agents can further comprise CRISPR/Cas systems as described in more detail below.

[0181] Active variants and fragments of nuclease agents (i.e., an engineered nuclease agent) are also provided. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the native nuclease agent, wherein the active variants retain the ability to cut at a desired target sequence and hence retain nick or double-strand-break-inducing activity. For example, any of the nuclease agents described herein can be modified from a native endonuclease sequence and designed to recognize and induce a nick or double-strand break at a target sequence that was not recognized by the native nuclease agent. Thus, some engineered nucleases have a specificity to induce a nick or double-strand break at a target sequence that is different from the corresponding native nuclease agent target sequence. Assays for nick or double-strand-break-inducing activity are known and generally measure the overall activity and specificity of the endonuclease on DNA substrates containing the target sequence.

[0182] The nuclease agent may be introduced into a cell or non-human animal by any known means. A polypeptide encoding the nuclease agent may be directly introduced into the cell or non-human animal. Alternatively, a polynucleotide encoding the nuclease agent can be introduced into the cell or non-human animal. When a polynucleotide encoding the nuclease agent is introduced, the nuclease agent can be transiently, conditionally, or constitutively expressed within

the cell. The polynucleotide encoding the nuclease agent can be contained in an expression cassette and be operably linked to a conditional promoter, an inducible promoter, a constitutive promoter, or a tissue-specific promoter. Examples of promoters are discussed in further detail elsewhere herein. Alternatively, the nuclease agent can be introduced into the cell as an mRNA encoding the nuclease agent.

[0183] A polynucleotide encoding a nuclease agent can be stably integrated in the genome of a cell and operably linked to a promoter active in the cell. Alternatively, a polynucleotide encoding a nuclease agent can be in a targeting vector.

[0184] When the nuclease agent is provided to the cell through the introduction of a polynucleotide encoding the nuclease agent, such a polynucleotide encoding a nuclease agent can be modified to substitute codons having a higher frequency of usage in the cell of interest, as compared to the naturally occurring polynucleotide sequence encoding the nuclease agent. For example, the polynucleotide encoding the nuclease agent can be modified to substitute codons having a higher frequency of usage in a given eukaryotic cell of interest, including a human cell, a non-human cell, a mammalian cell, a rodent cell, a mouse cell, a rat cell or any other host cell of interest, as compared to the naturally occurring polynucleotide sequence.

[0185] (2) CRISPR/Cas Systems Targeting Human TTR Gene

[0186] A particular type of human-TTR-targeting reagent can be a CRISPR/Cas system that targets the human TTR gene. CRISPR/Cas systems include transcripts and other elements involved in the expression of, or directing the activity of, Cas genes. A CRISPR/Cas system can be, for example, a type I, a type II, or a type III system. Alternatively, a CRISPR/Cas system can be a type V system (e.g., subtype V-A or subtype V-B). CRISPR/Cas systems used in the compositions and methods disclosed herein can be non-naturally occurring. A "non-naturally occurring" system includes anything indicating the involvement of the hand of man, such as one or more components of the system being altered or mutated from their naturally occurring state, being at least substantially free from at least one other component with which they are naturally associated in nature, or being associated with at least one other component with which they are not naturally associated. For example, non-naturally occurring CRISPR/Cas systems can employ CRISPR complexes comprising a gRNA and a Cas protein that do not naturally occur together, a Cas protein that does not occur naturally, or a gRNA that does not occur naturally.

[0187] Cas Proteins and Polynucleotides Encoding Cas Proteins.

[0188] Cas proteins generally comprise at least one RNA recognition or binding domain that can interact with guide RNAs (gRNAs, described in more detail below). Cas proteins can also comprise nuclease domains (e.g., DNase or RNase domains), DNA-binding domains, helicase domains, protein-protein interaction domains, dimerization domains, and other domains. Some such domains (e.g., DNase domains) can be from a native Cas protein. Other such domains can be added to make a modified Cas protein. A nuclease domain possesses catalytic activity for nucleic acid cleavage, which includes the breakage of the covalent bonds of a nucleic acid molecule. Cleavage can produce blunt ends or staggered ends, and it can be single-stranded or double-stranded. For example, a wild type Cas9 protein will typically create a blunt cleavage product. Alternatively, a wild

type Cpf1 protein (e.g., FnCpf1) can result in a cleavage product with a 5-nucleotide 5' overhang, with the cleavage occurring after the 18th base pair from the PAM sequence on the non-targeted strand and after the 23rd base on the targeted strand. A Cas protein can have full cleavage activity to create a double-strand break at a target genomic locus (e.g., a double-strand break with blunt ends), or it can be a nickase that creates a single-strand break at a target genomic locus.

[0189] Examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 or Csx12), Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, and Cu1966, and homologs or modified versions thereof.

[0190] An exemplary Cas protein is a Cas9 protein or a protein derived from Cas9 protein. Cas9 proteins are from a type II CRISPR/Cas system and typically share four key motifs with a conserved architecture. Motifs 1, 2, and 4 are RuvC-like motifs, and motif 3 is an HNH motif. Exemplary Cas9 proteins are from *Streptococcus pyogenes*, *Streptococcus thermophilus*, *Streptococcus* sp., *Staphylococcus aureus*, *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces viridochromogenes*, *Streptosporangium roseum*, *Streptosporangium roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycooides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microcilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas* sp., *Crocospaera watsonii*, *Cyanothece* sp., *Microcystis aeruginosa*, *Synechococcus* sp., *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor beccsii*, *Candidatus Desulforudis*, *Clostridium botulinum*, *Clostridium difficile*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter* sp., *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Kiedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc* sp., *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira* sp., *Lynghya* sp *Microcoleus chthonoplastes*, *Oscillatoria* sp *Petrotoga mobilis*, *Thermosipho africanus*, *Acaryochloris marina*, *Neisseria meningitidis*, or *Campylobacter jejuni*. Additional examples of the Cas9 family members are described in WO 2014/131833, herein incorporated by reference in its entirety for all purposes. Cas9 from *S. pyogenes* (SpCas9) (assigned SwissProt accession number Q99ZW2) is an exemplary Cas9 protein. Cas9 from *S. aureus* (SaCas9) (assigned UniProt accession number J7RUA5) is another exemplary Cas9 protein. Cas9 from *Campylobacter jejuni* (CjCas9) (assigned UniProt accession number Q0P897) is another exemplary Cas9 protein. See, e.g., Kim et al. (2017) *Nat. Comm.* 8:14500, herein incorporated by reference in its entirety for all purposes. SaCas9 is smaller than SpCas9, and CjCas9 is smaller than both SaCas9 and SpCas9. An exemplary Cas9 protein sequence can comprise, consist essentially of, or consist of SEQ ID NO: 94. An exemplary DNA

encoding the Cas9 protein can comprise, consist essentially of, or consist of SEQ ID NO: 93.

[0191] Another example of a Cas protein is a Cpf1 (CRISPR from *Prevotella* and *Francisella* 1) protein. Cpf1 is a large protein (about 1300 amino acids) that contains a RuvC-like nuclease domain homologous to the corresponding domain of Cas9 along with a counterpart to the characteristic arginine-rich cluster of Cas9. However, Cpf1 lacks the HNH nuclease domain that is present in Cas9 proteins, and the RuvC-like domain is contiguous in the Cpf1 sequence, in contrast to Cas9 where it contains long inserts including the HNH domain. See, e.g., Zetsche et al. (2015) *Cell* 163(3):759-771, herein incorporated by reference in its entirety for all purposes. Exemplary Cpf1 proteins are from *Francisella tularensis* 1, *Francisella tularensis* subsp. *novicida*, *Prevotella albensis*, *Lachnospiraceae bacterium* MC2017_1, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium* GW2011_GWA2_33_10, *Parcubacteria bacterium* GW2011_GWC2_44_17, *Smithella* sp. SCADC, *Acidaminococcus* sp. BV3L6, *Lachnospiraceae bacterium* MA2020, *Candidatus Methanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi* 237, *Leptospira inadai*, *Lachnospiraceae bacterium* ND2006, *Porphyromonas crevioricanis* 3, *Prevotella disiens*, and *Porphyromonas macacae*. Cpf1 from *Francisella novicida* U112 (FnCpf1; assigned UniProt accession number A0Q7Q2) is an exemplary Cpf1 protein.

[0192] Cas proteins can be wild type proteins (i.e., those that occur in nature), modified Cas proteins (i.e., Cas protein variants), or fragments of wild type or modified Cas proteins. Cas proteins can also be active variants or fragments with respect to catalytic activity of wild type or modified Cas proteins. Active variants or fragments with respect to catalytic activity can comprise at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the wild type or modified Cas protein or a portion thereof, wherein the active variants retain the ability to cut at a desired cleavage site and hence retain nick-inducing or double-strand-break-inducing activity. Assays for nick-inducing or double-strand-break-inducing activity are known and generally measure the overall activity and specificity of the Cas protein on DNA substrates containing the cleavage site.

[0193] Cas proteins can be modified to increase or decrease one or more of nucleic acid binding affinity, nucleic acid binding specificity, and enzymatic activity. Cas proteins can also be modified to change any other activity or property of the protein, such as stability. For example, one or more nuclease domains of the Cas protein can be modified, deleted, or inactivated, or a Cas protein can be truncated to remove domains that are not essential for the function of the protein or to optimize (e.g., enhance or reduce) the activity or a property of the Cas protein.

[0194] One example of a modified Cas protein is the modified SpCas9-HF1 protein, which is a high-fidelity variant of *Streptococcus pyogenes* Cas9 harboring alterations (N497A/R661A/Q695A/Q926A) designed to reduce non-specific DNA contacts. See, e.g., Kleinstiver et al. (2016) *Nature* 529(7587):490-495, herein incorporated by reference in its entirety for all purposes. Another example of a modified Cas protein is the modified eSpCas9 variant (K848A/K1003A/R1060A) designed to reduce off-target effects. See, e.g., Slaymaker et al. (2016) *Science* 351(6268):

84-88, herein incorporated by reference in its entirety for all purposes. Other SpCas9 variants include K855A and K810A/K1003A/R1060A.

[0195] Cas proteins can comprise at least one nuclease domain, such as a DNase domain. For example, a wild type Cpf1 protein generally comprises a RuvC-like domain that cleaves both strands of target DNA, perhaps in a dimeric configuration. Cas proteins can also comprise at least two nuclease domains, such as DNase domains. For example, a wild type Cas9 protein generally comprises a RuvC-like nuclease domain and an HNH-like nuclease domain. The RuvC and HNH domains can each cut a different strand of double-stranded DNA to make a double-stranded break in the DNA. See, e.g., Jinek et al. (2012) *Science* 337:816-821, herein incorporated by reference in its entirety for all purposes.

[0196] One or more or all of the nuclease domains can be deleted or mutated so that they are no longer functional or have reduced nuclease activity. For example, if one of the nuclease domains is deleted or mutated in a Cas9 protein, the resulting Cas9 protein can be referred to as a nickase and can generate a single-strand break at a guide RNA target sequence within a double-stranded DNA but not a double-strand break (i.e., it can cleave the complementary strand or the non-complementary strand, but not both). If both of the nuclease domains are deleted or mutated, the resulting Cas protein (e.g., Cas9) will have a reduced ability to cleave both strands of a double-stranded DNA (e.g., a nuclease-null or nuclease-inactive Cas protein, or a catalytically dead Cas protein (dCas)). An example of a mutation that converts Cas9 into a nickase is a D10A (aspartate to alanine at position 10 of Cas9) mutation in the RuvC domain of Cas9 from *S. pyogenes*. Likewise, H939A (histidine to alanine at amino acid position 839), H840A (histidine to alanine at amino acid position 840), or N863A (asparagine to alanine at amino acid position N863) in the HNH domain of Cas9 from *S. pyogenes* can convert the Cas9 into a nickase. Other examples of mutations that convert Cas9 into a nickase include the corresponding mutations to Cas9 from *S. thermophilus*. See, e.g., Sapranaukas et al. (2011) *Nucleic Acids Research* 39:9275-9282 and WO 2013/141680, each of which is herein incorporated by reference in its entirety for all purposes. Such mutations can be generated using methods such as site-directed mutagenesis, PCR-mediated mutagenesis, or total gene synthesis. Examples of other mutations creating nickases can be found, for example, in WO 2013/176772 and WO 2013/142578, each of which is herein incorporated by reference in its entirety for all purposes. If all of the nuclease domains are deleted or mutated in a Cas protein (e.g., both of the nuclease domains are deleted or mutated in a Cas9 protein), the resulting Cas protein (e.g., Cas9) will have a reduced ability to cleave both strands of a double-stranded DNA (e.g., a nuclease-null or nuclease-inactive Cas protein). One specific example is a D10A/H840A *S. pyogenes* Cas9 double mutant or a corresponding double mutant in a Cas9 from another species when optimally aligned with *S. pyogenes* Cas9. Another specific example is a D10A/N863A *S. pyogenes* Cas9 double mutant or a corresponding double mutant in a Cas9 from another species when optimally aligned with *S. pyogenes* Cas9.

[0197] Examples of inactivating mutations in the catalytic domains of *Staphylococcus aureus* Cas9 proteins are also known. For example, the *Staphylococcus aureus* Cas9 enzyme (SaCas9) may comprise a substitution at position

N580 (e.g., N580A substitution) and a substitution at position D10 (e.g., D10A substitution) to generate a nuclease-inactive Cas protein. See, e.g., WO 2016/106236, herein incorporated by reference in its entirety for all purposes.

[0198] Examples of inactivating mutations in the catalytic domains of Cpf1 proteins are also known. With reference to Cpf1 proteins from *Francisella novicida* U112 (FnCpf1), *Acidaminococcus* sp. BV3L6 (AsCpf1), *Lachnospiraceae bacterium* ND2006 (LbCpf1), and *Moraxella bovoculi* 237 (MbCpf1 Cpf1), such mutations can include mutations at positions 908, 993, or 1263 of AsCpf1 or corresponding positions in Cpf1 orthologs, or positions 832, 925, 947, or 1180 of LbCpf1 or corresponding positions in Cpf1 orthologs. Such mutations can include, for example one or more of mutations D908A, E993A, and D1263A of AsCpf1 or corresponding mutations in Cpf1 orthologs, or D832A, E925A, D947A, and D1180A of LbCpf1 or corresponding mutations in Cpf1 orthologs. See, e.g., US 2016/0208243, herein incorporated by reference in its entirety for all purposes.

[0199] Cas proteins (e.g., nuclease-active Cas proteins or nuclease-inactive Cas proteins) can also be operably linked to heterologous polypeptides as fusion proteins. For example, a Cas protein can be fused to a cleavage domain or an epigenetic modification domain. See WO 2014/089290, herein incorporated by reference in its entirety for all purposes. Cas proteins can also be fused to a heterologous polypeptide providing increased or decreased stability. The fused domain or heterologous polypeptide can be located at the N-terminus, the C-terminus, or internally within the Cas protein.

[0200] As one example, a Cas protein can be fused to one or more heterologous polypeptides that provide for subcellular localization. Such heterologous polypeptides can include, for example, one or more nuclear localization signals (NLS) such as the monopartite SV40 NLS and/or a bipartite alpha-importin NLS for targeting to the nucleus, a mitochondrial localization signal for targeting to the mitochondria, an ER retention signal, and the like. See, e.g., Lange et al. (2007) *J. Biol. Chem.* 282:5101-5105, herein incorporated by reference in its entirety for all purposes. Such subcellular localization signals can be located at the N-terminus, the C-terminus, or anywhere within the Cas protein. An NLS can comprise a stretch of basic amino acids, and can be a monopartite sequence or a bipartite sequence. Optionally, a Cas protein can comprise two or more NLSs, including an NLS (e.g., an alpha-importin NLS or a monopartite NLS) at the N-terminus and an NLS (e.g., an SV40 NLS or a bipartite NLS) at the C-terminus. A Cas protein can also comprise two or more NLSs at the N-terminus and/or two or more NLSs at the C-terminus.

[0201] Cas proteins can also be operably linked to a cell-penetrating domain or protein transduction domain. For example, the cell-penetrating domain can be derived from the HIV-1 TAT protein, the TLM cell-penetrating motif from human hepatitis B virus, MPG, Pep-1, VP22, a cell penetrating peptide from Herpes simplex virus, or a polyarginine peptide sequence. See, e.g., WO 2014/089290 and WO 2013/176772, each of which is herein incorporated by reference in its entirety for all purposes. The cell-penetrating domain can be located at the N-terminus, the C-terminus, or anywhere within the Cas protein.

[0202] Cas proteins can also be operably linked to a heterologous polypeptide for ease of tracking or purification,

such as a fluorescent protein, a purification tag, or an epitope tag. Examples of fluorescent proteins include green fluorescent proteins (e.g., GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), yellow fluorescent proteins (e.g., YFP, eYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), blue fluorescent proteins (e.g., eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), cyan fluorescent proteins (e.g., eCFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), red fluorescent proteins (e.g., mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), orange fluorescent proteins (e.g., mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato), and any other suitable fluorescent protein. Examples of tags include glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly(NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, hemagglutinin (HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, histidine (His), biotin carboxyl carrier protein (BCCP), and calmodulin.

[0203] Cas proteins can also be tethered to exogenous donor nucleic acids or labeled nucleic acids. Such tethering (i.e., physical linking) can be achieved through covalent interactions or noncovalent interactions, and the tethering can be direct (e.g., through direct fusion or chemical conjugation, which can be achieved by modification of cysteine or lysine residues on the protein or intein modification), or can be achieved through one or more intervening linkers or adapter molecules such as streptavidin or aptamers. See, e.g., Pierce et al. (2005) *Mini Rev. Med. Chem.* 5(1):41-55; Duckworth et al. (2007) *Angew. Chem. Int. Ed. Engl.* 46(46):8819-8822; Schaeffer and Dixon (2009) *Australian J. Chem.* 62(10):1328-1332; Goodman et al. (2009) *Chembiochem.* 10(9):1551-1557; and Khatwani et al. (2012) *Bioorg. Med. Chem.* 20(14):4532-4539, each of which is herein incorporated by reference in its entirety for all purposes. Noncovalent strategies for synthesizing protein-nucleic acid conjugates include biotin-streptavidin and nickel-histidine methods. Covalent protein-nucleic acid conjugates can be synthesized by connecting appropriately functionalized nucleic acids and proteins using a wide variety of chemistries. Some of these chemistries involve direct attachment of the oligonucleotide to an amino acid residue on the protein surface (e.g., a lysine amine or a cysteine thiol), while other more complex schemes require post-translational modification of the protein or the involvement of a catalytic or reactive protein domain. Methods for covalent attachment of proteins to nucleic acids can include, for example, chemical cross-linking of oligonucleotides to protein lysine or cysteine residues, expressed protein-ligation, chemoenzymatic methods, and the use of photoaptamers. The exogenous donor nucleic acid or labeled nucleic acid can be tethered to the C-terminus, the N-terminus, or to an internal region within the Cas protein. In one example, the exogenous donor nucleic acid or labeled nucleic acid is tethered to the C-terminus or the N-terminus of the Cas protein. Likewise, the Cas protein can be tethered to the 5' end, the 3' end, or to an internal region within the exogenous donor nucleic acid or labeled nucleic acid. That is, the exogenous donor nucleic acid or labeled nucleic acid can be tethered in any orientation and polarity. For example, the Cas protein can be

tethered to the 5' end or the 3' end of the exogenous donor nucleic acid or labeled nucleic acid.

[0204] Cas proteins can be provided in any form. For example, a Cas protein can be provided in the form of a protein, such as a Cas protein complexed with a gRNA. Alternatively, a Cas protein can be provided in the form of a nucleic acid encoding the Cas protein, such as an RNA (e.g., messenger RNA (mRNA)) or DNA. Optionally, the nucleic acid encoding the Cas protein can be codon optimized for efficient translation into protein in a particular cell or organism. For example, the nucleic acid encoding the Cas protein can be modified to substitute codons having a higher frequency of usage in a bacterial cell, a yeast cell, a human cell, a non-human cell, a mammalian cell, a rodent cell, a mouse cell, a rat cell, or any other host cell of interest, as compared to the naturally occurring polynucleotide sequence. When a nucleic acid encoding the Cas protein is introduced into the cell, the Cas protein can be transiently, conditionally, or constitutively expressed in the cell.

[0205] Cas proteins provided as mRNAs can be modified for improved stability and/or immunogenicity properties. The modifications may be made to one or more nucleosides within the mRNA. Examples of chemical modifications to mRNA nucleobases include pseudouridine, 1-methylpseudouridine, and 5-methyl-cytidine. For example, capped and polyadenylated Cas mRNA containing N1-methyl pseudouridine can be used. Likewise, Cas mRNAs can be modified by depletion of uridine using synonymous codons.

[0206] Nucleic acids encoding Cas proteins can be stably integrated in the genome of the cell and operably linked to a promoter active in the cell. Alternatively, nucleic acids encoding Cas proteins can be operably linked to a promoter in an expression construct. Expression constructs include any nucleic acid constructs capable of directing expression of a gene or other nucleic acid sequence of interest (e.g., a Cas gene) and which can transfer such a nucleic acid sequence of interest to a target cell. For example, the nucleic acid encoding the Cas protein can be in a targeting vector comprising a nucleic acid insert and/or a vector comprising a DNA encoding a gRNA. Alternatively, it can be in a vector or plasmid that is separate from the targeting vector comprising the nucleic acid insert and/or separate from the vector comprising the DNA encoding the gRNA. Promoters that can be used in an expression construct include promoters active, for example, in one or more of a eukaryotic cell, a human cell, a non-human cell, a mammalian cell, a non-human mammalian cell, a rodent cell, a mouse cell, a rat cell, a hamster cell, a rabbit cell, a pluripotent cell, an embryonic stem (ES) cell, or a zygote. Such promoters can be, for example, conditional promoters, inducible promoters, constitutive promoters, or tissue-specific promoters. Optionally, the promoter can be a bidirectional promoter driving expression of both a Cas protein in one direction and a guide RNA in the other direction. Such bidirectional promoters can consist of (1) a complete, conventional, unidirectional Pol III promoter that contains 3 external control elements: a distal sequence element (DSE), a proximal sequence element (PSE), and a TATA box; and (2) a second basic Pol III promoter that includes a PSE and a TATA box fused to the 5' terminus of the DSE in reverse orientation. For example, in the H1 promoter, the DSE is adjacent to the PSE and the TATA box, and the promoter can be rendered bidirectional by creating a hybrid promoter in which transcription in the reverse direction is controlled by

appending a PSE and TATA box derived from the U6 promoter. See, e.g., US 2016/0074535, herein incorporated by references in its entirety for all purposes. Use of a bidirectional promoter to express genes encoding a Cas protein and a guide RNA simultaneously allow for the generation of compact expression cassettes to facilitate delivery.

[0207] Guide RNAs.

[0208] A “guide RNA” or “gRNA” is an RNA molecule that binds to a Cas protein (e.g., Cas9 protein) and targets the Cas protein to a specific location within a target DNA. Guide RNAs can comprise two segments: a “DNA-targeting segment” and a “protein-binding segment.” “Segment” includes a section or region of a molecule, such as a contiguous stretch of nucleotides in an RNA. Some gRNAs, such as those for Cas9, can comprise two separate RNA molecules: an “activator-RNA” (e.g., tracrRNA) and a “targeter-RNA” (e.g., CRISPR RNA or crRNA). Other gRNAs are a single RNA molecule (single RNA polynucleotide), which can also be called a “single-molecule gRNA,” a “single-guide RNA,” or an “sgRNA.” See, e.g., WO 2013/176772, WO 2014/065596, WO 2014/089290, WO 2014/093622, WO 2014/099750, WO 2013/142578, and WO 2014/131833, each of which is herein incorporated by reference in its entirety for all purposes. For Cas9, for example, a single-guide RNA can comprise a crRNA fused to a tracrRNA (e.g., via a linker). For Cpf1, for example, only a crRNA is needed to achieve binding to and/or cleavage of a target sequence. The terms “guide RNA” and “gRNA” include both double-molecule (i.e., modular) gRNAs and single-molecule gRNAs.

[0209] An exemplary two-molecule gRNA comprises a crRNA-like (“CRISPR RNA” or “targeter-RNA” or “crRNA” or “crRNA repeat”) molecule and a corresponding tracrRNA-like (“trans-acting CRISPR RNA” or “activator-RNA” or “tracrRNA”) molecule. A crRNA comprises both the DNA-targeting segment (single-stranded) of the gRNA and a stretch of nucleotides (i.e., the crRNA tail) that forms one half of the dsRNA duplex of the protein-binding segment of the gRNA. An example of a crRNA tail, located downstream (3') of the DNA-targeting segment, comprises, consists essentially of, or consists of GUUUUAGAGCUAUGCU (SEQ ID NO: 87). Any of the DNA-targeting segments disclosed herein can be joined to the 5' end of SEQ ID NO: 87 to form a crRNA.

[0210] A corresponding tracrRNA (activator-RNA) comprises a stretch of nucleotides that forms the other half of the dsRNA duplex of the protein-binding segment of the gRNA. A stretch of nucleotides of a crRNA are complementary to and hybridize with a stretch of nucleotides of a tracrRNA to form the dsRNA duplex of the protein-binding domain of the gRNA. As such, each crRNA can be said to have a corresponding tracrRNA. An example of a tracrRNA sequence comprises, consists essentially of, or consists of AGCAUAGCAAGUUAAAUAAGGCUAGUC-CGUUAUCAACUUGAAAAAGUGGCACC GAGUCG-GUGCUUU (SEQ ID NO: 88).

[0211] In systems in which both a crRNA and a tracrRNA are needed, the crRNA and the corresponding tracrRNA hybridize to form a gRNA. In systems in which only a crRNA is needed, the crRNA can be the gRNA. The crRNA additionally provides the single-stranded DNA-targeting segment that targets a guide RNA target sequence by hybridizing to the opposite strand (i.e., the complementary strand). If used for modification within a cell, the exact sequence of

a given crRNA or tracrRNA molecule can be designed to be specific to the species in which the RNA molecules will be used. See, e.g., Mali et al. (2013) *Science* 339:823-826; Jinek et al. (2012) *Science* 337:816-821; Hwang et al. (2013) *Nat. Biotechnol.* 31:227-229; Jiang et al. (2013) *Nat. Biotechnol.* 31:233-239; and Cong et al. (2013) *Science* 339: 819-823, each of which is herein incorporated by reference in its entirety for all purposes.

[0212] The DNA-targeting segment (crRNA) of a given gRNA comprises a nucleotide sequence that is complementary to a sequence (i.e., the complementary strand of the guide RNA recognition sequence on the strand opposite of the guide RNA target sequence) in a target DNA. The DNA-targeting segment of a gRNA interacts with a target DNA in a sequence-specific manner via hybridization (i.e., base pairing). As such, the nucleotide sequence of the DNA-targeting segment may vary and determines the location within the target DNA with which the gRNA and the target DNA will interact. The DNA-targeting segment of a subject gRNA can be modified to hybridize to any desired sequence within a target DNA. Naturally occurring crRNAs differ depending on the CRISPR/Cas system and organism but often contain a targeting segment of between 21 to 72 nucleotides length, flanked by two direct repeats (DR) of a length of between 21 to 46 nucleotides (see, e.g., WO 2014/131833, herein incorporated by reference in its entirety for all purposes). In the case of *S. pyogenes*, the DRs are 36 nucleotides long and the targeting segment is 30 nucleotides long. The 3' located DR is complementary to and hybridizes with the corresponding tracrRNA, which in turn binds to the Cas protein.

[0213] The DNA-targeting segment can have a length of at least about 12 nucleotides, at least about 15 nucleotides, at least about 17 nucleotides, at least about 18 nucleotides, at least about 19 nucleotides, at least about 20 nucleotides, at least about 25 nucleotides, at least about 30 nucleotides, at least about 35 nucleotides, or at least about 40 nucleotides. Such DNA-targeting segments can have a length from about 12 nucleotides to about 100 nucleotides, from about 12 nucleotides to about 80 nucleotides, from about 12 nucleotides to about 50 nucleotides, from about 12 nucleotides to about 40 nucleotides, from about 12 nucleotides to about 30 nucleotides, from about 12 nucleotides to about 25 nucleotides, or from about 12 nucleotides to about 20 nucleotides. For example, the DNA targeting segment can be from about 15 nucleotides to about 25 nucleotides (e.g., from about 17 nucleotides to about 20 nucleotides, or about 17 nucleotides, about 18 nucleotides, about 19 nucleotides, or about 20 nucleotides). See, e.g., US 2016/0024523, herein incorporated by reference in its entirety for all purposes. For Cas9 from *S. pyogenes*, a typical DNA-targeting segment is between 16 and 20 nucleotides in length or between 17 and 20 nucleotides in length. For Cas9 from *S. aureus*, a typical DNA-targeting segment is between 21 and 23 nucleotides in length. For Cpf1, a typical DNA-targeting segment is at least 16 nucleotides in length or at least 18 nucleotides in length.

[0214] TracrRNAs can be in any form (e.g., full-length tracrRNAs or active partial tracrRNAs) and of varying lengths. They can include primary transcripts or processed forms. For example, tracrRNAs (as part of a single-guide RNA or as a separate molecule as part of a two-molecule gRNA) may comprise, consist essentially of, or consist of all or a portion of a wild type tracrRNA sequence (e.g., about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more

nucleotides of a wild type tracrRNA sequence). Examples of wild type tracrRNA sequences from *S. pyogenes* include 171-nucleotide, 89-nucleotide, 75-nucleotide, and 65-nucleotide versions. See, e.g., Deltcheva et al. (2011) *Nature* 471:602-607; WO 2014/093661, each of which is herein incorporated by reference in its entirety for all purposes. Examples of tracrRNAs within single-guide RNAs (sgRNAs) include the tracrRNA segments found within +48, +54, +67, and +85 versions of sgRNAs, where “+n” indicates that up to the +n nucleotide of wild type tracrRNA is included in the sgRNA. See U.S. Pat. No. 8,697,359, herein incorporated by reference in its entirety for all purposes.

[0215] The percent complementarity between the DNA-targeting segment and the complementary strand of the guide RNA recognition sequence within the target DNA can be at least 60% (e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100%). The percent complementarity between the DNA-targeting segment and the complementary strand of the guide RNA recognition sequence within the target DNA can be at least 60% over about 20 contiguous nucleotides. As an example, the percent complementarity between the DNA-targeting segment and the complementary strand of the guide RNA recognition sequence within the target DNA is 100% over the 14 contiguous nucleotides at the 5' end of the complementary strand of the guide RNA recognition sequence within the target DNA and as low as 0% over the remainder. In such a case, the DNA-targeting segment can be considered to be 14 nucleotides in length. As another example, the percent complementarity between the DNA-targeting segment and the complementary strand of the guide RNA recognition sequence within the target DNA is 100% over the seven contiguous nucleotides at the 5' end of the complementary strand of the guide RNA recognition sequence within the complementary strand of the target DNA and as low as 0% over the remainder. In such a case, the DNA-targeting segment can be considered to be 7 nucleotides in length. In some guide RNAs, at least 17 nucleotides within the DNA-targeting segment are complementary to the target DNA. For example, the DNA-targeting segment can be 20 nucleotides in length and can comprise 1, 2, or 3 mismatches with the complementary strand of the guide RNA recognition sequence. Preferably, the mismatches are not adjacent to a protospacer adjacent motif (PAM) sequence (e.g., the mismatches are in the 5' end of the DNA-targeting segment, or the mismatches are at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 base pairs away from the PAM sequence).

[0216] The protein-binding segment of a gRNA can comprise two stretches of nucleotides that are complementary to one another. The complementary nucleotides of the protein-binding segment hybridize to form a double-stranded RNA duplex (dsRNA). The protein-binding segment of a subject gRNA interacts with a Cas protein, and the gRNA directs the bound Cas protein to a specific nucleotide sequence within target DNA via the DNA-targeting segment.

[0217] Single-guide RNAs have the DNA-targeting segment and a scaffold sequence (i.e., the protein-binding or Cas-binding sequence of the guide RNA). For example, such guide RNAs have a 5' DNA-targeting segment and a 3' scaffold sequence. Exemplary scaffold sequences comprise, consist essentially of, or consist of: GUUUUAGAGCUA-GAAAUAGCAAGUUAAAAUAAGGCUAGUC-

CGUUAUCAACUUGA AAAAGUGGCACCGAGUCG-GUGCU (version 1; SEQ ID NO: 89); GUUGGAACCAUUCAAAACAGCAUAG-CAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (version 2; SEQ ID NO: 8); GUUUUAGAGCUAGAAAUAG-CAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC-UUGA AAAAGUGGCACCGAGUCGGUGC (version 3; SEQ ID NO: 9); and GUUUUAGAGCUAUGCUG-GAAACAGCAUAGCAAGUUUUAAAAUAAGGCUAGUC-CGUU AUCAACUUGAAAAAGUGGCACCGAGUCG-GUGC (version 4; SEQ ID NO: 10). Guide RNAs targeting any guide RNA target sequence can include, for example, a DNA-targeting segment on the 5' end of the guide RNA fused to any of the exemplary guide RNA scaffold sequences on the 3' end of the guide RNA. That is, any of the DNA-targeting segments disclosed herein can be joined to the 5' end of any one of SEQ ID NOS: 89, 8, 9, or 10 to form a single guide RNA (chimeric guide RNA). Guide RNA versions 1, 2, 3, and 4 as disclosed elsewhere herein refer to DNA-targeting segments (i.e., guide sequences or guides) joined with scaffold versions 1, 2, 3, and 4, respectively.

[0218] Guide RNAs can include modifications or sequences that provide for additional desirable features (e.g., modified or regulated stability; subcellular targeting; tracking with a fluorescent label; a binding site for a protein or protein complex; and the like). Examples of such modifications include, for example, a 5' cap (e.g., a 7-methylguanylate cap (m7G)); a 3' polyadenylated tail (i.e., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and/or protein complexes); a stability control sequence; a sequence that forms a dsRNA duplex (i.e., a hairpin); a modification or sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, and so forth); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof. Other examples of modifications include engineered stem loop duplex structures, engineered bulge regions, engineered hairpins 3' of the stem loop duplex structure, or any combination thereof. See, e.g., US 2015/0376586, herein incorporated by reference in its entirety for all purposes. A bulge can be an unpaired region of nucleotides within the duplex made up of the crRNA-like region and the minimum tracrRNA-like region. A bulge can comprise, on one side of the duplex, an unpaired 5'-XXX-3' where X is any purine and Y can be a nucleotide that can form a wobble pair with a nucleotide on the opposite strand, and an unpaired nucleotide region on the other side of the duplex.

[0219] Unmodified nucleic acids can be prone to degradation. Exogenous nucleic acids can also induce an innate immune response. Modifications can help introduce stability and reduce immunogenicity. Guide RNAs can comprise modified nucleosides and modified nucleotides including, for example, one or more of the following: (1) alteration or replacement of one or both of the non-linking phosphate oxygens and/or of one or more of the linking phosphate oxygens in the phosphodiester backbone linkage; (2) altera-

tion or replacement of a constituent of the ribose sugar such as alteration or replacement of the 2' hydroxyl on the ribose sugar; (3) replacement of the phosphate moiety with dephospho linkers; (4) modification or replacement of a naturally occurring nucleobase; (5) replacement or modification of the ribose-phosphate backbone; (6) modification of the 3' end or 5' end of the oligonucleotide (e.g., removal, modification or replacement of a terminal phosphate group or conjugation of a moiety); and (7) modification of the sugar. Other possible guide RNA modifications include modifications of or replacement of uracils or poly-uracil tracts. See, e.g., WO 2015/048577 and US 2016/0237455, each of which is herein incorporated by reference in its entirety for all purposes. Similar modifications can be made to Cas-encoding nucleic acids, such as Cas mRNAs.

[0220] As one example, nucleotides at the 5' or 3' end of a guide RNA can include phosphorothioate linkages (e.g., the bases can have a modified phosphate group that is a phosphorothioate group). For example, a guide RNA can include phosphorothioate linkages between the 2, 3, or 4 terminal nucleotides at the 5' or 3' end of the guide RNA. As another example, nucleotides at the 5' and/or 3' end of a guide RNA can have 2'-O-methyl modifications. For example, a guide RNA can include 2'-O-methyl modifications at the 2, 3, or 4 terminal nucleotides at the 5' and/or 3' end of the guide RNA (e.g., the 5' end). See, e.g., WO 2017/173054 A1 and Finn et al. (2018) *Cell Reports* 22:1-9, each of which is herein incorporated by reference in its entirety for all purposes. In one specific example, the guide RNA comprises 2'-O-methyl analogs and 3' phosphorothioate internucleotide linkages at the first three 5' and 3' terminal RNA residues. In another specific example, the guide RNA is modified such that all 2'OH groups that do not interact with the Cas9 protein are replaced with 2'-O-methyl analogs, and the tail region of the guide RNA, which has minimal interaction with Cas9, is modified with 5' and 3' phosphorothioate internucleotide linkages. See, e.g., Yin et al. (2017) *Nat. Biotech.* 35(12):1179-1187, herein incorporated by reference in its entirety for all purposes. Other examples of modified guide RNAs are provided, e.g., in WO 2018/107028 A1, herein incorporated by reference in its entirety for all purposes.

[0221] Guide RNAs can be provided in any form. For example, the gRNA can be provided in the form of RNA, either as two molecules (separate crRNA and tracrRNA) or as one molecule (sgRNA), and optionally in the form of a complex with a Cas protein. The gRNA can also be provided in the form of DNA encoding the gRNA. The DNA encoding the gRNA can encode a single RNA molecule (sgRNA) or separate RNA molecules (e.g., separate crRNA and tracrRNA). In the latter case, the DNA encoding the gRNA can be provided as one DNA molecule or as separate DNA molecules encoding the crRNA and tracrRNA, respectively.

[0222] When a gRNA is provided in the form of DNA, the gRNA can be transiently, conditionally, or constitutively expressed in the cell. DNAs encoding gRNAs can be stably integrated into the genome of the cell and operably linked to a promoter active in the cell. Alternatively, DNAs encoding gRNAs can be operably linked to a promoter in an expression construct. For example, the DNA encoding the gRNA can be in a vector comprising a heterologous nucleic acid, such as a nucleic acid encoding a Cas protein. Alternatively, it can be in a vector or a plasmid that is separate from the vector comprising the nucleic acid encoding the Cas protein.

Promoters that can be used in such expression constructs include promoters active, for example, in one or more of a eukaryotic cell, a human cell, a non-human cell, a mammalian cell, a non-human mammalian cell, a rodent cell, a mouse cell, a rat cell, a hamster cell, a rabbit cell, a pluripotent cell, an embryonic stem (ES) cell, an adult stem cell, a developmentally restricted progenitor cell, an induced pluripotent stem (iPS) cell, or a one-cell stage embryo. Such promoters can be, for example, conditional promoters, inducible promoters, constitutive promoters, or tissue-specific promoters. Such promoters can also be, for example, bidirectional promoters. Specific examples of suitable promoters include an RNA polymerase III promoter, such as a human U6 promoter, a rat U6 polymerase III promoter, or a mouse U6 polymerase III promoter.

[0223] Alternatively, gRNAs can be prepared by various other methods. For example, gRNAs can be prepared by in vitro transcription using, for example, T7 RNA polymerase (see, e.g., WO 2014/089290 and WO 2014/065596, each of which is herein incorporated by reference in its entirety for all purposes). Guide RNAs can also be a synthetically produced molecule prepared by chemical synthesis.

[0224] Guide RNA Recognition Sequences and Guide RNA Target Sequences.

[0225] The term “guide RNA recognition sequence” includes nucleic acid sequences present in a target DNA to which a DNA-targeting segment of a gRNA will bind, provided sufficient conditions for binding exist. The term guide RNA recognition sequence as used herein encompasses both strands of the target double-stranded DNA (i.e., the sequence on the complementary strand to which the guide RNA hybridizes and the corresponding sequence on the non-complementary strand adjacent to the protospacer adjacent motif (PAM)). The term “guide RNA target sequence” as used herein refers specifically to the sequence on the non-complementary strand adjacent to the PAM (i.e., upstream or 5' of the PAM). That is, the guide RNA target sequence refers to the sequence on the non-complementary strand corresponding to the sequence to which the guide RNA hybridizes on the complementary strand. A guide RNA target sequence is equivalent to the DNA-targeting segment of a guide RNA, but with thymines instead of uracils. As one example, a guide RNA target sequence for a Cas9 enzyme would refer to the sequence on the non-complementary strand adjacent to the 5'-NGG-3' PAM. Guide RNA recognition sequences include sequences to which a guide RNA is designed to have complementarity, where hybridization between the complementary strand of a guide RNA recognition sequence and a DNA-targeting segment of a guide RNA promotes the formation of a CRISPR complex. Full complementarity is not necessarily required, provided that there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. Guide RNA recognition sequences or guide RNA target sequences also include cleavage sites for Cas proteins, described in more detail below. A guide RNA recognition sequence or a guide RNA target sequence can comprise any polynucleotide, which can be located, for example, in the nucleus or cytoplasm of a cell or within an organelle of a cell, such as a mitochondrion or chloroplast.

[0226] The guide RNA recognition sequence within a target DNA can be targeted by (i.e., be bound by, or hybridize with, or be complementary to) a Cas protein or a gRNA. Suitable DNA/RNA binding conditions include

physiological conditions normally present in a cell. Other suitable DNA/RNA binding conditions (e.g., conditions in a cell-free system) are known (see, e.g., *Molecular Cloning: A Laboratory Manual*, 3rd Ed. (Sambrook et al., Harbor Laboratory Press 2001), herein incorporated by reference in its entirety for all purposes). The strand of the target DNA that is complementary to and hybridizes with the Cas protein or gRNA can be called the “complementary strand,” and the strand of the target DNA that is complementary to the “complementary strand” (and is therefore not complementary to the Cas protein or gRNA) can be called “non-complementary strand” or “template strand.”

[0227] The Cas protein can cleave the nucleic acid at a site within or outside of the nucleic acid sequence present in the target DNA to which the DNA-targeting segment of a gRNA will bind. The “cleavage site” includes the position of a nucleic acid at which a Cas protein produces a single-strand break or a double-strand break. For example, formation of a CRISPR complex (comprising a gRNA hybridized to the complementary strand of a guide RNA recognition sequence and complexed with a Cas protein) can result in cleavage of one or both strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the nucleic acid sequence present in a target DNA to which a DNA-targeting segment of a gRNA will bind. If the cleavage site is outside of the nucleic acid sequence to which the DNA-targeting segment of the gRNA will bind, the cleavage site is still considered to be within the “guide RNA recognition sequence” or guide RNA target sequence. The cleavage site can be on only one strand or on both strands of a nucleic acid. Cleavage sites can be at the same position on both strands of the nucleic acid (producing blunt ends) or can be at different sites on each strand (producing staggered ends (i.e., overhangs)). Staggered ends can be produced, for example, by using two Cas proteins, each of which produces a single-strand break at a different cleavage site on a different strand, thereby producing a double-strand break. For example, a first nickase can create a single-strand break on the first strand of double-stranded DNA (dsDNA), and a second nickase can create a single-strand break on the second strand of dsDNA such that overhanging sequences are created. In some cases, the guide RNA recognition sequence or guide RNA target sequence of the nickase on the first strand is separated from the guide RNA recognition sequence or guide RNA target sequence of the nickase on the second strand by at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500, or 1,000 base pairs.

[0228] Site-specific binding and/or cleavage of target DNA by Cas proteins can occur at locations determined by both (i) base-pairing complementarity between the gRNA and the target DNA and (ii) a short motif, called the protospacer adjacent motif (PAM), in the target DNA. The PAM can flank the guide RNA target sequence on the non-complementary strand opposite of the strand to which the guide RNA hybridizes. Optionally, the guide RNA target sequence can be flanked on the 3' end by the PAM. Alternatively, the guide RNA target sequence can be flanked on the 5' end by the PAM. For example, the cleavage site of Cas proteins can be about 1 to about 10 or about 2 to about 5 base pairs (e.g., 3 base pairs) upstream or downstream of the PAM sequence. In some cases (e.g., when Cas9 from *S. pyogenes* or a closely related Cas9 is used), the PAM sequence of the non-complementary strand can be 5'-N₁GG-3', where N₁ is any DNA nucleotide and is immediately 3' of

the guide RNA recognition sequence of the non-complementary strand of the target DNA (i.e., immediately 3' of the guide RNA target sequence). As such, the PAM sequence of the complementary strand would be 5'-CCN₂-3', where N₂ is any DNA nucleotide and is immediately 5' of the guide RNA recognition sequence of the complementary strand of the target DNA. In some such cases, N₁ and N₂ can be complementary and the N₁-N₂ base pair can be any base pair (e.g., N₁=C and N₂=G; N₁=G and N₂=C; N₁=A and N₂=T; or N₁=T, and N₂=A). In the case of Cas9 from *S. aureus*, the PAM can be NNGRRT or NNGRR, where N can be A, G, C, or T, and R can be G or A. In the case of Cas9 from *C. jejuni*, the PAM can be, for example, NNNNACAC or NNNNRYAC, where N can be A, G, C, or T, and R can be G or A. In some cases (e.g., for Fncpf1), the PAM sequence can be upstream of the 5' end and have the sequence 5'-TTN-3'.

[0229] Examples of guide RNA target sequences or guide RNA target sequences in addition to a PAM sequence are provided below. For example, the guide RNA target sequence can be a 20-nucleotide DNA sequence immediately preceding an NGG motif recognized by a Cas9 protein. Examples of such guide RNA target sequences plus a PAM sequence are GN₁₉NGG (SEQ ID NO: 11) or N₂₀NGG (SEQ ID NO: 12). See, e.g., WO 2014/165825, herein incorporated by reference in its entirety for all purposes. The guanine at the 5' end can facilitate transcription by RNA polymerase in cells. Other examples of guide RNA target sequences plus a PAM sequence can include two guanine nucleotides at the 5' end (e.g., GGN₂₀NGG; SEQ ID NO: 13) to facilitate efficient transcription by T7 polymerase in vitro. See, e.g., WO 2014/065596, herein incorporated by reference in its entirety for all purposes. Other guide RNA target sequences plus a PAM sequence can have between 4-22 nucleotides in length of SEQ ID NOS: 11-13, including the 5' G or GG and the 3' GG or NGG. Yet other guide RNA target sequences can have between 14 and 20 nucleotides in length of SEQ ID NOS: 11-13.

[0230] The guide RNA recognition sequence or guide RNA target sequence can be any nucleic acid sequence endogenous or exogenous to a cell. The guide RNA recognition sequence or guide RNA target sequence can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory sequence) or can include both.

[0231] (3) Exogenous Donor Nucleic Acids Targeting Human TTR Gene

[0232] The methods and compositions disclosed herein can utilize exogenous donor nucleic acids to modify the humanized TTR locus following cleavage of the humanized TTR locus with a nuclease agent. In such methods, the nuclease agent protein cleaves the humanized TTR locus to create a single-strand break (nick) or double-strand break, and the exogenous donor nucleic acid recombines the humanized TTR locus via non-homologous end joining (NHEJ)-mediated ligation or through a homology-directed repair event. Optionally, repair with the exogenous donor nucleic acid removes or disrupts the nuclease target sequence so that alleles that have been targeted cannot be re-targeted by the nuclease agent.

[0233] Exogenous donor nucleic acids can comprise deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), they can be single-stranded or double-stranded, and they can be in linear or circular form. For example, an exogenous

donor nucleic acid can be a single-stranded oligodeoxy-nucleotide (ssODN). See, e.g., Yoshimi et al. (2016) *Nat. Commun.* 7:10431, herein incorporated by reference in its entirety for all purposes. An exemplary exogenous donor nucleic acid is between about 50 nucleotides to about 5 kb in length, is between about 50 nucleotides to about 3 kb in length, or is between about 50 to about 1,000 nucleotides in length. Other exemplary exogenous donor nucleic acids are between about 40 to about 200 nucleotides in length. For example, an exogenous donor nucleic acid can be between about 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140, 140-150, 150-160, 160-170, 170-180, 180-190, or 190-200 nucleotides in length. Alternatively, an exogenous donor nucleic acid can be between about 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 nucleotides in length. Alternatively, an exogenous donor nucleic acid can be between about 1-1.5, 1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, or 4.5-5 kb in length. Alternatively, an exogenous donor nucleic acid can be, for example, no more than 5 kb, 4.5 kb, 4 kb, 3.5 kb, 3 kb, 2.5 kb, 2 kb, 1.5 kb, 1 kb, 900 nucleotides, 800 nucleotides, 700 nucleotides, 600 nucleotides, 500 nucleotides, 400 nucleotides, 300 nucleotides, 200 nucleotides, 100 nucleotides, or 50 nucleotides in length. Exogenous donor nucleic acids (e.g., targeting vectors) can also be longer.

[0234] In one example, an exogenous donor nucleic acid is an ssODN that is between about 80 nucleotides and about 200 nucleotides in length. In another example, an exogenous donor nucleic acid is an ssODN that is between about 80 nucleotides and about 3 kb in length. Such an ssODN can have homology arms, for example, that are each between about 40 nucleotides and about 60 nucleotides in length. Such an ssODN can also have homology arms, for example, that are each between about 30 nucleotides and 100 nucleotides in length. The homology arms can be symmetrical (e.g., each 40 nucleotides or each 60 nucleotides in length), or they can be asymmetrical (e.g., one homology arm that is 36 nucleotides in length, and one homology arm that is 91 nucleotides in length).

[0235] Exogenous donor nucleic acids can include modifications or sequences that provide for additional desirable features (e.g., modified or regulated stability; tracking or detecting with a fluorescent label; a binding site for a protein or protein complex; and so forth). Exogenous donor nucleic acids can comprise one or more fluorescent labels, purification tags, epitope tags, or a combination thereof. For example, an exogenous donor nucleic acid can comprise one or more fluorescent labels (e.g., fluorescent proteins or other fluorophores or dyes), such as at least 1, at least 2, at least 3, at least 4, or at least 5 fluorescent labels. Exemplary fluorescent labels include fluorophores such as fluorescein (e.g., 6-carboxyfluorescein (6-FAM)), Texas Red, HEX, Cy3, Cy5, Cy5.5, Pacific Blue, 5-(and-6)-carboxytetramethylrhodamine (TAMRA), and Cy7. A wide range of fluorescent dyes are available commercially for labeling oligonucleotides (e.g., from Integrated DNA Technologies). Such fluorescent labels (e.g., internal fluorescent labels) can be used, for example, to detect an exogenous donor nucleic acid that has been directly integrated into a cleaved target nucleic acid having protruding ends compatible with the ends of the exogenous donor nucleic acid. The label or tag can be at the 5' end, the 3' end, or internally within the exogenous donor nucleic acid. For example, an exogenous donor nucleic acid

can be conjugated at 5' end with the IR700 fluorophore from Integrated DNA Technologies (5'IRDYE® 700).

[0236] Exogenous donor nucleic acids can also comprise nucleic acid inserts including segments of DNA to be integrated at the humanized TTR locus. Integration of a nucleic acid insert at a humanized TTR locus can result in addition of a nucleic acid sequence of interest to the humanized TTR locus, deletion of a nucleic acid sequence of interest at the humanized TTR locus, or replacement of a nucleic acid sequence of interest at the humanized TTR locus (i.e., deletion and insertion). Some exogenous donor nucleic acids are designed for insertion of a nucleic acid insert at the humanized TTR locus without any corresponding deletion at the humanized TTR locus. Other exogenous donor nucleic acids are designed to delete a nucleic acid sequence of interest at the humanized TTR locus without any corresponding insertion of a nucleic acid insert. Yet other exogenous donor nucleic acids are designed to delete a nucleic acid sequence of interest at the humanized TTR locus and replace it with a nucleic acid insert.

[0237] The nucleic acid insert or the corresponding nucleic acid at the humanized TTR locus being deleted and/or replaced can be various lengths. An exemplary nucleic acid insert or corresponding nucleic acid at the humanized TTR locus being deleted and/or replaced is between about 1 nucleotide to about 5 kb in length or is between about 1 nucleotide to about 1,000 nucleotides in length. For example, a nucleic acid insert or a corresponding nucleic acid at the humanized TTR locus being deleted and/or replaced can be between about 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140, 140-150, 150-160, 160-170, 170-180, 180-190, or 190-200 nucleotides in length. Likewise, a nucleic acid insert or a corresponding nucleic acid at the humanized TTR locus being deleted and/or replaced can be between 1-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 nucleotides in length. Likewise, a nucleic acid insert or a corresponding nucleic acid at the humanized TTR locus being deleted and/or replaced can be between about 1-1.5, 1.5-2, 2-2.5, 2.5-3, 3-3.5, 3.5-4, 4-4.5, or 4.5-5 kb in length or longer.

[0238] The nucleic acid insert can comprise a sequence that is homologous or orthologous to all or part of sequence targeted for replacement. For example, the nucleic acid insert can comprise a sequence that comprises one or more point mutations (e.g., 1, 2, 3, 4, 5, or more) compared with a sequence targeted for replacement at the humanized TTR locus. Optionally, such point mutations can result in a conservative amino acid substitution (e.g., substitution of aspartic acid [Asp, D] with glutamic acid [Glu, E]) in the encoded polypeptide.

[0239] Donor Nucleic Acids for Non-Homologous-End-Joining-Mediated Insertion.

[0240] Some exogenous donor nucleic acids have short single-stranded regions at the 5' end and/or the 3' end that are complementary to one or more overhangs created by nuclease-mediated cleavage at the humanized TTR locus. These overhangs can also be referred to as 5' and 3' homology arms. For example, some exogenous donor nucleic acids have short single-stranded regions at the 5' end and/or the 3' end that are complementary to one or more overhangs created by nuclease-mediated cleavage at 5' and/or 3' target sequences at the humanized TTR locus. Some such exog-

enous donor nucleic acids have a complementary region only at the 5' end or only at the 3' end. For example, some such exogenous donor nucleic acids have a complementary region only at the 5' end complementary to an overhang created at a 5' target sequence at the humanized TTR locus or only at the 3' end complementary to an overhang created at a 3' target sequence at the humanized TTR locus. Other such exogenous donor nucleic acids have complementary regions at both the 5' and 3' ends. For example, other such exogenous donor nucleic acids have complementary regions at both the 5' and 3' ends e.g., complementary to first and second overhangs, respectively, generated by nuclease-mediated cleavage at the humanized TTR locus. For example, if the exogenous donor nucleic acid is double-stranded, the single-stranded complementary regions can extend from the 5' end of the top strand of the donor nucleic acid and the 5' end of the bottom strand of the donor nucleic acid, creating 5' overhangs on each end. Alternatively, the single-stranded complementary region can extend from the 3' end of the top strand of the donor nucleic acid and from the 3' end of the bottom strand of the template, creating 3' overhangs.

[0241] The complementary regions can be of any length sufficient to promote ligation between the exogenous donor nucleic acid and the target nucleic acid. Exemplary complementary regions are between about 1 to about 5 nucleotides in length, between about 1 to about 25 nucleotides in length, or between about 5 to about 150 nucleotides in length. For example, a complementary region can be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. Alternatively, the complementary region can be about 5-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140, or 140-150 nucleotides in length, or longer.

[0242] Such complementary regions can be complementary to overhangs created by two pairs of nickases. Two double-strand breaks with staggered ends can be created by using first and second nickases that cleave opposite strands of DNA to create a first double-strand break, and third and fourth nickases that cleave opposite strands of DNA to create a second double-strand break. For example, a Cas protein can be used to nick first, second, third, and fourth guide RNA target sequences corresponding with first, second, third, and fourth guide RNAs. The first and second guide RNA target sequences can be positioned to create a first cleavage site such that the nicks created by the first and second nickases on the first and second strands of DNA create a double-strand break (i.e., the first cleavage site comprises the nicks within the first and second guide RNA target sequences). Likewise, the third and fourth guide RNA target sequences can be positioned to create a second cleavage site such that the nicks created by the third and fourth nickases on the first and second strands of DNA create a double-strand break (i.e., the second cleavage site comprises the nicks within the third and fourth guide RNA target sequences). Preferably, the nicks within the first and second guide RNA target sequences and/or the third and fourth guide RNA target sequences can be off-set nicks that create overhangs. The offset window can be, for example, at least about 5 bp, 10 bp, 20 bp, 30 bp, 40 bp, 50 bp, 60 bp, 70 bp, 80 bp, 90 bp, 100 bp or more. See Ran et al. (2013) *Cell* 154:1380-1389; Mali et al. (2013) *Nat. Biotech.* 31:833-838; and Shen et al. (2014) *Nat. Methods* 11:399-404, each of which is herein incorporated by reference in its entirety for

all purposes. In such cases, a double-stranded exogenous donor nucleic acid can be designed with single-stranded complementary regions that are complementary to the overhangs created by the nicks within the first and second guide RNA target sequences and by the nicks within the third and fourth guide RNA target sequences. Such an exogenous donor nucleic acid can then be inserted by non-homologous-end-joining-mediated ligation.

[0243] Donor Nucleic Acids for Insertion by Homology-Directed Repair.

[0244] Some exogenous donor nucleic acids comprise homology arms. If the exogenous donor nucleic acid also comprises a nucleic acid insert, the homology arms can flank the nucleic acid insert. For ease of reference, the homology arms are referred to herein as 5' and 3' (i.e., upstream and downstream) homology arms. This terminology relates to the relative position of the homology arms to the nucleic acid insert within the exogenous donor nucleic acid. The 5' and 3' homology arms correspond to regions within the humanized TTR locus, which are referred to herein as "5' target sequence" and "3' target sequence," respectively.

[0245] A homology arm and a target sequence "correspond" or are "corresponding" to one another when the two regions share a sufficient level of sequence identity to one another to act as substrates for a homologous recombination reaction. The term "homology" includes DNA sequences that are either identical or share sequence identity to a corresponding sequence. The sequence identity between a given target sequence and the corresponding homology arm found in the exogenous donor nucleic acid can be any degree of sequence identity that allows for homologous recombination to occur. For example, the amount of sequence identity shared by the homology arm of the exogenous donor nucleic acid (or a fragment thereof) and the target sequence (or a fragment thereof) can be at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, such that the sequences undergo homologous recombination. Moreover, a corresponding region of homology between the homology arm and the corresponding target sequence can be of any length that is sufficient to promote homologous recombination. Exemplary homology arms are between about 25 nucleotides to about 2.5 kb in length, are between about 25 nucleotides to about 1.5 kb in length, or are between about 25 to about 500 nucleotides in length. For example, a given homology arm (or each of the homology arms) and/or corresponding target sequence can comprise corresponding regions of homology that are between about 25-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, or 450-500 nucleotides in length, such that the homology arms have sufficient homology to undergo homologous recombination with the corresponding target sequences within the target nucleic acid. Alternatively, a given homology arm (or each homology arm) and/or corresponding target sequence can comprise corresponding regions of homology that are between about 0.5 kb to about 1 kb, about 1 kb to about 1.5 kb, about 1.5 kb to about 2 kb, or about 2 kb to about 2.5 kb in length. For example, the homology arms can each be about 750 nucleotides in length. The homology arms can be symmetrical (each about the same size in length), or they can be asymmetrical (one longer than the other).

[0246] When a nuclease agent is used in combination with an exogenous donor nucleic acid, the 5' and 3' target sequences are preferably located in sufficient proximity to the nuclease cleavage site (e.g., within sufficient proximity to a the nuclease target sequence) so as to promote the occurrence of a homologous recombination event between the target sequences and the homology arms upon a single-strand break (nick) or double-strand break at the nuclease cleavage site. The term "nuclease cleavage site" includes a DNA sequence at which a nick or double-strand break is created by a nuclease agent (e.g., a Cas9 protein complexed with a guide RNA). The target sequences within the targeted locus that correspond to the 5' and 3' homology arms of the exogenous donor nucleic acid are "located in sufficient proximity" to a nuclease cleavage site if the distance is such as to promote the occurrence of a homologous recombination event between the 5' and 3' target sequences and the homology arms upon a single-strand break or double-strand break at the nuclease cleavage site. Thus, the target sequences corresponding to the 5' and/or 3' homology arms of the exogenous donor nucleic acid can be, for example, within at least 1 nucleotide of a given nuclease cleavage site or within at least 10 nucleotides to about 1,000 nucleotides of a given nuclease cleavage site. As an example, the nuclease cleavage site can be immediately adjacent to at least one or both of the target sequences.

[0247] The spatial relationship of the target sequences that correspond to the homology arms of the exogenous donor nucleic acid and the nuclease cleavage site can vary. For example, target sequences can be located 5' to the nuclease cleavage site, target sequences can be located 3' to the nuclease cleavage site, or the target sequences can flank the nuclease cleavage site.

[0248] (4) Other Human-TTR-Targeting Reagents

[0249] The activity of any other known or putative human-TTR-targeting reagent can also be assessed using the non-human animals disclosed herein. Similarly, any other molecule can be screened for human-TTR-targeting activity using the non-human animals disclosed herein.

[0250] Examples of other human-TTR-targeting reagents include antisense oligonucleotides (e.g., siRNAs or shRNAs) that act through RNA interference (RNAi). Antisense oligonucleotides (ASOs) or antisense RNAs are short synthetic strings of nucleotides designed to prevent the expression of a targeted protein by selectively binding to the RNA that encodes the targeted protein and thereby preventing translation. These compounds bind to RNA with high affinity and selectivity through well characterized Watson-Crick base pairing (hybridization). RNA interference (RNAi) is an endogenous cellular mechanism for controlling gene expression in which small interfering RNAs (siRNAs) that are bound to the RNA-induced silencing complex (RISC) mediate the cleavage of target messenger RNA (mRNA). Examples of human-TTR-targeting antisense oligonucleotides are known. See, e.g., Ackermann et al. (2012) *Amyloid Suppl* 1:43-44 and Coelho et al. (2013) *N. Engl. J. Med.* 369(9):819-829, each of which is herein incorporated by reference in its entirety for all purposes.

[0251] Other human-TTR-targeting reagents include antibodies or antigen-binding proteins designed to specifically bind a human TTR epitope.

[0252] Other human-TTR-targeting reagents include small-molecule reagents. One example of such a small-molecule reagent is tafamidis, which functions by kinetic

stabilization of the correctly folded tetrameric form of the transthyretin (TTR) protein. See, e.g., Hammarstrom et al. (2003) *Science* 299:713-716, herein incorporated by reference in its entirety for all purposes.

[0253] D. Administering Human-TTR-Targeting Reagents to Non-Human Animals or Cells

[0254] The methods disclosed herein can comprise introducing into a non-human animal or cell various molecules (e.g., human-TTR-targeting reagents such as therapeutic molecules or complexes), including nucleic acids, proteins, nucleic-acid-protein complexes, or protein complexes. "Introducing" includes presenting to the cell or non-human animal the molecule (e.g., nucleic acid or protein) in such a manner that it gains access to the interior of the cell or to the interior of cells within the non-human animal. The introducing can be accomplished by any means, and two or more of the components (e.g., two of the components, or all of the components) can be introduced into the cell or non-human animal simultaneously or sequentially in any combination. For example, a Cas protein can be introduced into a cell or non-human animal before introduction of a guide RNA, or it can be introduced following introduction of the guide RNA. As another example, an exogenous donor nucleic acid can be introduced prior to the introduction of a Cas protein and a guide RNA, or it can be introduced following introduction of the Cas protein and the guide RNA (e.g., the exogenous donor nucleic acid can be administered about 1, 2, 3, 4, 8, 12, 24, 36, 48, or 72 hours before or after introduction of the Cas protein and the guide RNA). See, e.g., US 2015/0240263 and US 2015/0110762, each of which is herein incorporated by reference in its entirety for all purposes. In addition, two or more of the components can be introduced into the cell or non-human animal by the same delivery method or different delivery methods. Similarly, two or more of the components can be introduced into a non-human animal by the same route of administration or different routes of administration.

[0255] In some methods, components of a CRISPR/Cas system are introduced into a non-human animal or cell. A guide RNA can be introduced into a non-human animal or cell in the form of an RNA (e.g., in vitro transcribed RNA) or in the form of a DNA encoding the guide RNA. When introduced in the form of a DNA, the DNA encoding a guide RNA can be operably linked to a promoter active in a cell in the non-human animal. For example, a guide RNA may be delivered via AAV and expressed in vivo under a U6 promoter. Such DNAs can be in one or more expression constructs. For example, such expression constructs can be components of a single nucleic acid molecule. Alternatively, they can be separated in any combination among two or more nucleic acid molecules (i.e., DNAs encoding one or more CRISPR RNAs and DNAs encoding one or more tracrRNAs can be components of a separate nucleic acid molecules).

[0256] Likewise, Cas proteins can be provided in any form. For example, a Cas protein can be provided in the form of a protein, such as a Cas protein complexed with a gRNA. Alternatively, a Cas protein can be provided in the form of a nucleic acid encoding the Cas protein, such as an RNA (e.g., messenger RNA (mRNA)) or DNA. Optionally, the nucleic acid encoding the Cas protein can be codon optimized for efficient translation into protein in a particular cell or organism. For example, the nucleic acid encoding the Cas protein can be modified to substitute codons having a

higher frequency of usage in a mammalian cell, a rodent cell, a mouse cell, a rat cell, or any other host cell of interest, as compared to the naturally occurring polynucleotide sequence. When a nucleic acid encoding the Cas protein is introduced into a non-human animal, the Cas protein can be transiently, conditionally, or constitutively expressed in a cell in the non-human animal.

[0257] Nucleic acids encoding Cas proteins or guide RNAs can be operably linked to a promoter in an expression construct. Expression constructs include any nucleic acid constructs capable of directing expression of a gene or other nucleic acid sequence of interest (e.g., a Cas gene) and which can transfer such a nucleic acid sequence of interest to a target cell. For example, the nucleic acid encoding the Cas protein can be in a vector comprising a DNA encoding one or more gRNAs. Alternatively, it can be in a vector or plasmid that is separate from the vector comprising the DNA encoding one or more gRNAs. Suitable promoters that can be used in an expression construct include promoters active, for example, in one or more of a eukaryotic cell, a human cell, a non-human cell, a mammalian cell, a non-human mammalian cell, a rodent cell, a mouse cell, a rat cell, a hamster cell, a rabbit cell, a pluripotent cell, an embryonic stem (ES) cell, an adult stem cell, a developmentally restricted progenitor cell, an induced pluripotent stem (iPS) cell, or a one-cell stage embryo. Such promoters can be, for example, conditional promoters, inducible promoters, constitutive promoters, or tissue-specific promoters. Optionally, the promoter can be a bidirectional promoter driving expression of both a Cas protein in one direction and a guide RNA in the other direction. Such bidirectional promoters can consist of (1) a complete, conventional, unidirectional Pol III promoter that contains 3 external control elements: a distal sequence element (DSE), a proximal sequence element (PSE), and a TATA box; and (2) a second basic Pol III promoter that includes a PSE and a TATA box fused to the 5' terminus of the DSE in reverse orientation. For example, in the H1 promoter, the DSE is adjacent to the PSE and the TATA box, and the promoter can be rendered bidirectional by creating a hybrid promoter in which transcription in the reverse direction is controlled by appending a PSE and TATA box derived from the U6 promoter. See, e.g., US 2016/0074535, herein incorporated by references in its entirety for all purposes. Use of a bidirectional promoter to express genes encoding a Cas protein and a guide RNA simultaneously allows for the generation of compact expression cassettes to facilitate delivery.

[0258] Molecules (e.g., Cas proteins or guide RNAs) introduced into the non-human animal or cell can be provided in compositions comprising a carrier increasing the stability of the introduced molecules (e.g., prolonging the period under given conditions of storage (e.g., -20°C ., 4°C ., or ambient temperature) for which degradation products remain below a threshold, such below 0.5% by weight of the starting nucleic acid or protein; or increasing the stability in vivo). Non-limiting examples of such carriers include poly(lactic acid) (PLA) microspheres, poly(D,L-lactic-co-glycolic-acid) (PLGA) microspheres, liposomes, micelles, inverse micelles, lipid cochleates, and lipid microtubules.

[0259] Various methods and compositions are provided herein to allow for introduction of a nucleic acid or protein into a cell or non-human animal. Methods for introducing nucleic acids into various cell types are known and include,

for example, stable transfection methods, transient transfection methods, and virus-mediated methods.

[0260] Transfection protocols as well as protocols for introducing nucleic acid sequences into cells may vary. Non-limiting transfection methods include chemical-based transfection methods using liposomes; nanoparticles; calcium phosphate (Graham et al. (1973) *Virology* 52 (2): 456-67, Bacchetti et al. (1977) *Proc. Natl. Acad. Sci. USA* 74 (4): 1590-4, and Kriegler, M (1991). *Transfer and Expression: A Laboratory Manual*. New York: W. H. Freeman and Company, pp. 96-97); dendrimers; or cationic polymers such as DEAE-dextran or polyethylenimine. Non-chemical methods include electroporation, Sono-poration, and optical transfection. Particle-based transfection includes the use of a gene gun, or magnet-assisted transfection (Bertram (2006) *Current Pharmaceutical Biotechnology* 7, 277-28). Viral methods can also be used for transfection.

[0261] Introduction of nucleic acids or proteins into a cell can also be mediated by electroporation, by intracytoplasmic injection, by viral infection, by adenovirus, by adeno-associated virus, by lentivirus, by retrovirus, by transfection, by lipid-mediated transfection, or by nucleofection. Nucleofection is an improved electroporation technology that enables nucleic acid substrates to be delivered not only to the cytoplasm but also through the nuclear membrane and into the nucleus. In addition, use of nucleofection in the methods disclosed herein typically requires much fewer cells than regular electroporation (e.g., only about 2 million compared with 7 million by regular electroporation). In one example, nucleofection is performed using the LONZA® NUCLEOFECTOR™ system.

[0262] Introduction of nucleic acids or proteins into a cell (e.g., a zygote) can also be accomplished by microinjection. In zygotes (i.e., one-cell stage embryos), microinjection can be into the maternal and/or paternal pronucleus or into the cytoplasm. If the microinjection is into only one pronucleus, the paternal pronucleus is preferable due to its larger size. Microinjection of an mRNA is preferably into the cytoplasm (e.g., to deliver mRNA directly to the translation machinery), while microinjection of a Cas protein or a polynucleotide encoding a Cas protein or encoding an RNA is preferable into the nucleus/pronucleus. Alternatively, microinjection can be carried out by injection into both the nucleus/pronucleus and the cytoplasm: a needle can first be introduced into the nucleus/pronucleus and a first amount can be injected, and while removing the needle from the one-cell stage embryo a second amount can be injected into the cytoplasm. If a Cas protein is injected into the cytoplasm, the Cas protein preferably comprises a nuclear localization signal to ensure delivery to the nucleus/pronucleus. Methods for carrying out microinjection are well known. See, e.g., Nagy et al. (Nagy A, Gertsenstein M, Vintersten K, Behringer R., 2003, *Manipulating the Mouse Embryo*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); see also Meyer et al. (2010) *Proc. Natl. Acad. Sci. USA* 107:15022-15026 and Meyer et al. (2012) *Proc. Natl. Acad. Sci. USA* 109:9354-9359.

[0263] Other methods for introducing nucleic acid or proteins into a cell or non-human animal can include, for example, vector delivery, particle-mediated delivery, exosome-mediated delivery, lipid-nanoparticle-mediated delivery, cell-penetrating-peptide-mediated delivery, or implantable-device-mediated delivery. As specific examples, a nucleic acid or protein can be introduced into a cell or

non-human animal in a carrier such as a poly(lactic acid) (PLA) microsphere, a poly(D,L-lactic-co-glycolic-acid) (PLGA) microsphere, a liposome, a micelle, an inverse micelle, a lipid cochleate, or a lipid microtubule. Some specific examples of delivery to a non-human animal include hydrodynamic delivery, virus-mediated delivery (e.g., adeno-associated virus (AAV)-mediated delivery), and lipid-nanoparticle-mediated delivery.

[0264] Introduction of nucleic acids and proteins into cells or non-human animals can be accomplished by hydrodynamic delivery (HDD). Hydrodynamic delivery has emerged as a method for intracellular DNA delivery *in vivo*. For gene delivery to parenchymal cells, only essential DNA sequences need to be injected via a selected blood vessel, eliminating safety concerns associated with current viral and synthetic vectors. When injected into the bloodstream, DNA is capable of reaching cells in the different tissues accessible to the blood. Hydrodynamic delivery employs the force generated by the rapid injection of a large volume of solution into the incompressible blood in the circulation to overcome the physical barriers of endothelium and cell membranes that prevent large and membrane-impermeable compounds from entering parenchymal cells. In addition to the delivery of DNA, this method is useful for the efficient intracellular delivery of RNA, proteins, and other small compounds *in vivo*. See, e.g., Bonamassa et al. (2011) *Pharm. Res.* 28(4): 694-701, herein incorporated by reference in its entirety for all purposes.

[0265] Introduction of nucleic acids can also be accomplished by virus-mediated delivery, such as AAV-mediated delivery or lentivirus-mediated delivery. Other exemplary viruses/viral vectors include retroviruses, adenoviruses, vaccinia viruses, poxviruses, and herpes simplex viruses. The viruses can infect dividing cells, non-dividing cells, or both dividing and non-dividing cells. The viruses can integrate into the host genome or alternatively do not integrate into the host genome. Such viruses can also be engineered to have reduced immunity. The viruses can be replication-competent or can be replication-defective (e.g., defective in one or more genes necessary for additional rounds of virion replication and/or packaging). Viruses can cause transient expression, long-lasting expression (e.g., at least 1 week, 2 weeks, 1 month, 2 months, or 3 months), or permanent expression (e.g., of Cas9 and/or gRNA). Exemplary viral titers (e.g., AAV titers) include 10^{12} , 10^{13} , 10^{14} , 10^{15} and 10^{16} vector genomes/mL.

[0266] The ssDNA AAV genome consists of two open reading frames, Rep and Cap, flanked by two inverted terminal repeats that allow for synthesis of the complementary DNA strand. When constructing an AAV transfer plasmid, the transgene is placed between the two ITRs, and Rep and Cap can be supplied *in trans*. In addition to Rep and Cap, AAV can require a helper plasmid containing genes from adenovirus. These genes (E4, E2a, and VA) mediated AAV replication. For example, the transfer plasmid, Rep/Cap, and the helper plasmid can be transfected into HEK293 cells containing the adenovirus gene E1+ to produce infectious AAV particles. Alternatively, the Rep, Cap, and adenovirus helper genes may be combined into a single plasmid. Similar packaging cells and methods can be used for other viruses, such as retroviruses.

[0267] Multiple serotypes of AAV have been identified. These serotypes differ in the types of cells they infect (i.e., their tropism), allowing preferential transduction of specific

cell types. Serotypes for CNS tissue include AAV1, AAV2, AAV4, AAV5, AAV8, and AAV9. Serotypes for heart tissue include AAV1, AAV8, and AAV9. Serotypes for kidney tissue include AAV2. Serotypes for lung tissue include AAV4, AAV5, AAV6, and AAV9. Serotypes for pancreas tissue include AAV8. Serotypes for photoreceptor cells include AAV2, AAV5, and AAV8. Serotypes for retinal pigment epithelium tissue include AAV1, AAV2, AAV4, AAV5, and AAV8. Serotypes for skeletal muscle tissue include AAV1, AAV6, AAV7, AAV8, and AAV9. Serotypes for liver tissue include AAV7, AAV8, and AAV9, and particularly AAV8.

[0268] Tropism can be further refined through pseudotyping, which is the mixing of a capsid and a genome from different viral serotypes. For example AAV2/5 indicates a virus containing the genome of serotype 2 packaged in the capsid from serotype 5. Use of pseudotyped viruses can improve transduction efficiency, as well as alter tropism. Hybrid capsids derived from different serotypes can also be used to alter viral tropism. For example, AAV-DJ contains a hybrid capsid from eight serotypes and displays high infectivity across a broad range of cell types *in vivo*. AAV-DJ8 is another example that displays the properties of AAV-DJ but with enhanced brain uptake. AAV serotypes can also be modified through mutations. Examples of mutational modifications of AAV2 include Y444F, Y500F, Y730F, and S662V. Examples of mutational modifications of AAV3 include Y705F, Y731F, and T492V. Examples of mutational modifications of AAV6 include S663V and T492V. Other pseudotyped/modified AAV variants include AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5, AAV8.2, and AAV/SASTG.

[0269] To accelerate transgene expression, self-complementary AAV (scAAV) variants can be used. Because AAV depends on the cell's DNA replication machinery to synthesize the complementary strand of the AAV's single-stranded DNA genome, transgene expression may be delayed. To address this delay, scAAV containing complementary sequences that are capable of spontaneously annealing upon infection can be used, eliminating the requirement for host cell DNA synthesis. However, single-stranded AAV (ssAAV) vectors can also be used.

[0270] To increase packaging capacity, longer transgenes may be split between two AAV transfer plasmids, the first with a 3' splice donor and the second with a 5' splice acceptor. Upon co-infection of a cell, these viruses form concatemers, are spliced together, and the full-length transgene can be expressed. Although this allows for longer transgene expression, expression is less efficient. Similar methods for increasing capacity utilize homologous recombination. For example, a transgene can be divided between two transfer plasmids but with substantial sequence overlap such that co-expression induces homologous recombination and expression of the full-length transgene.

[0271] Introduction of nucleic acids and proteins can also be accomplished by lipid nanoparticle (LNP)-mediated delivery. For example, LNP-mediated delivery can be used to deliver a combination of Cas mRNA and guide RNA or a combination of Cas protein and guide RNA. Delivery through such methods results in transient Cas expression, and the biodegradable lipids improve clearance, improve tolerability, and decrease immunogenicity. Lipid formulations can protect biological molecules from degradation while improving their cellular uptake. Lipid nanoparticles

are particles comprising a plurality of lipid molecules physically associated with each other by intermolecular forces. These include microspheres (including unilamellar and multilamellar vesicles, e.g., liposomes), a dispersed phase in an emulsion, micelles, or an internal phase in a suspension. Such lipid nanoparticles can be used to encapsulate one or more nucleic acids or proteins for delivery. Formulations which contain cationic lipids are useful for delivering polyanions such as nucleic acids. Other lipids that can be included are neutral lipids (i.e., uncharged or zwitterionic lipids), anionic lipids, helper lipids that enhance transfection, and stealth lipids that increase the length of time for which nanoparticles can exist in vivo. Examples of suitable cationic lipids, neutral lipids, anionic lipids, helper lipids, and stealth lipids can be found in WO 2016/010840 A1, herein incorporated by reference in its entirety for all purposes. An exemplary lipid nanoparticle can comprise a cationic lipid and one or more other components. In one example, the other component can comprise a helper lipid such as cholesterol. In another example, the other components can comprise a helper lipid such as cholesterol and a neutral lipid such as DSPC. In another example, the other components can comprise a helper lipid such as cholesterol, an optional neutral lipid such as DSPC, and a stealth lipid such as S010, S024, S027, S031, or S033.

[0272] The LNP may contain one or more or all of the following: (i) a lipid for encapsulation and for endosomal escape; (ii) a neutral lipid for stabilization; (iii) a helper lipid for stabilization; and (iv) a stealth lipid. See, e.g., Finn et al. (2018) *Cell Reports* 22:1-9 and WO 2017/173054 A1, each of which is herein incorporated by reference in its entirety for all purposes. In certain LNPs, the cargo can include a guide RNA or a nucleic acid encoding a guide RNA. In certain LNPs, the cargo can include an mRNA encoding a Cas nuclease, such as Cas9, and a guide RNA or a nucleic acid encoding a guide RNA.

[0273] The lipid for encapsulation and endosomal escape can be a cationic lipid. The lipid can also be a biodegradable lipid, such as a biodegradable ionizable lipid. One example of a suitable lipid is Lipid A or LP01, which is (9Z,12Z)-3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl)octadeca-9,12-dienoate, also called 3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl) (9Z,12Z)-octadeca-9,12-dienoate. See, e.g., Finn et al. (2018) *Cell Reports* 22:1-9 and WO 2017/173054 A1, each of which is herein incorporated by reference in its entirety for all purposes. Another example of a suitable lipid is Lipid B, which is ((5-((dimethylamino)methyl)-1,3-phenylene)bis(oxy))bis(octane-8,1-diyl)bis(decanoate), also called ((5-((dimethylamino)methyl)-1,3-phenylene)bis(oxy))bis(octane-8,1-diyl)bis(decanoate). Another example of a suitable lipid is Lipid C, which is 2-((4-(((3-(dimethylamino)propoxy)carbonyloxy)hexadecanoyloxy)propane-1,3-diyl) (9Z,9'Z,12Z,12'Z)-bis(octadeca-9,12-dienoate). Another example of a suitable lipid is Lipid D, which is 3-(((3-(dimethylamino)propoxy)carbonyloxy)-13-(octanoyloxy)tridecyl 3-octylundecanoate. Other suitable lipids include heptatriaconta-6,9,28,31-tetraen-19-yl4-(dimethylamino)butanoate (also known as Dlin-MC3-DMA (MC3)).

[0274] Some such lipids suitable for use in the LNPs described herein are biodegradable in vivo. For example, LNPs comprising such a lipid include those where at least 75% of the lipid is cleared from the plasma within 8, 10, 12,

24, or 48 hours, or 3, 4, 5, 6, 7, or 10 days. As another example, at least 50% of the LNP is cleared from the plasma within 8, 10, 12, 24, or 48 hours, or 3, 4, 5, 6, 7, or 10 days. **[0275]** Such lipids may be ionizable depending upon the pH of the medium they are in. For example, in a slightly acidic medium, the lipids may be protonated and thus bear a positive charge. Conversely, in a slightly basic medium, such as, for example, blood where pH is approximately 7.35, the lipids may not be protonated and thus bear no charge. In some embodiments, the lipids may be protonated at a pH of at least about 9, 9.5, or 10. The ability of such a lipid to bear a charge is related to its intrinsic pKa. For example, the lipid may, independently, have a pKa in the range of from about 5.8 to about 6.2.

[0276] Neutral lipids function to stabilize and improve processing of the LNPs. Examples of suitable neutral lipids include a variety of neutral, uncharged or zwitterionic lipids. Examples of neutral phospholipids suitable for use in the present disclosure include, but are not limited to, 5-heptadecylbenzene-1,3-diol (resorcinol), dipalmitoylphosphatidylcholine (DPPC), distearoylphosphatidylcholine (DSPC), phosphocholine (DOPC), dimyristoylphosphatidylcholine (DMPC), phosphatidylcholine (PLPC), 1,2-distearoyl-sn-glycero-3-phosphocholine (DAPC), phosphatidylethanolamine (PE), egg phosphatidylcholine (EPC), dilauryloylphosphatidylcholine (DLPC), dimyristoylphosphatidylcholine (DMPC), 1-myristoyl-2-palmitoyl phosphatidylcholine (MPPC), 1-palmitoyl-2-myristoyl phosphatidylcholine (PMPC), 1-palmitoyl-2-stearoyl phosphatidylcholine (PSPC), 1,2-diarachidoyl-sn-glycero-3-phosphocholine (DBPC), 1-stearoyl-2-palmitoyl phosphatidylcholine (SPPC), 1,2-dieicosenoyl-sn-glycero-3-phosphocholine (DEPC), palmitoyloleoyl phosphatidylcholine (POPC), lysophosphatidyl choline, dioleoyl phosphatidylethanolamine (DOPE), dilinoleoylphosphatidylcholine distearoylphosphatidylethanolamine (DSPE), dimyristoyl phosphatidylethanolamine (DMPE), dipalmitoyl phosphatidylethanolamine (DPPE), palmitoyloleoyl phosphatidylethanolamine (POPE), lysophosphatidylethanolamine, and combinations thereof. For example, the neutral phospholipid may be selected from the group consisting of distearoylphosphatidylcholine (DSPC) and dimyristoyl phosphatidyl ethanolamine (DMPE).

[0277] Helper lipids include lipids that enhance transfection. The mechanism by which the helper lipid enhances transfection can include enhancing particle stability. In certain cases, the helper lipid can enhance membrane fusogenicity. Helper lipids include steroids, sterols, and alkyl resorcinols. Examples of suitable helper lipids suitable include cholesterol, 5-heptadecylresorcinol, and cholesterol hemisuccinate. In one example, the helper lipid may be cholesterol or cholesterol hemisuccinate.

[0278] Stealth lipids include lipids that alter the length of time the nanoparticles can exist in vivo. Stealth lipids may assist in the formulation process by, for example, reducing particle aggregation and controlling particle size. Stealth lipids may modulate pharmacokinetic properties of the LNP. Suitable stealth lipids include lipids having a hydrophilic head group linked to a lipid moiety.

[0279] The hydrophilic head group of stealth lipid can comprise, for example, a polymer moiety selected from polymers based on PEG (sometimes referred to as poly(ethylene oxide)), poly(oxazoline), poly(vinyl alcohol), poly(glycerol), poly(N-vinylpyrrolidone), polyaminoacids, and

poly N-(2-hydroxypropyl)methacrylamide. The term PEG means any polyethylene glycol or other polyalkylene ether polymer. In certain LNP formulations, the PEG, is a PEG-2K, also termed PEG 2000, which has an average molecular weight of about 2,000 daltons. See, e.g., WO 2017/173054 A1, herein incorporated by reference in its entirety for all purposes.

[0280] The lipid moiety of the stealth lipid may be derived, for example, from diacylglycerol or diacylglyceramide, including those comprising a dialkylglycerol or dialkylglycamide group having alkyl chain length independently comprising from about C4 to about C40 saturated or unsaturated carbon atoms, wherein the chain may comprise one or more functional groups such as, for example, an amide or ester. The dialkylglycerol or dialkylglycamide group can further comprise one or more substituted alkyl groups.

[0281] As one example, the stealth lipid may be selected from PEG-dilaurylglycerol, PEG-dimyristoylglycerol (PEG-DMG), PEG-dipalmitoylglycerol, PEG-distearoylglycerol (PEG-DSPE), PEG-dilaurylglycamide, PEG-dimyristylglycamide, PEG-dipalmitoylglycamide, and PEG-distearoylglycamide, PEG-cholesterol (1-[8'-(Cholest-5-en-3[beta]-oxy)carboxamido-3',6'-dioxaoctanyl]carbamoyl-[omega]-methyl-poly(ethylene glycol), PEG-DMB (3,4-ditetradecoxylbenzyl-[omega]-methyl-poly(ethylene glycol)ether), 1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000] (PEG2k-DMG), 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000] (PEG2k-DSPE), 1,2-distearoyl-sn-glycerol, methoxypoly ethylene glycol (PEG2k-DSG), poly(ethylene glycol)-2000-dimethacrylate (PEG2k-DMA), and 1,2-distearoxypropyl-3-amine-N-[methoxy(polyethylene glycol)-2000] (PEG2k-DSA). In one particular example, the stealth lipid may be PEG2k-DMG.

[0282] The LNPs can comprise different respective molar ratios of the component lipids in the formulation. The mol-% of the CCD lipid may be, for example, from about 30 mol-% to about 60 mol-%, from about 35 mol-% to about 55 mol-%, from about 40 mol-% to about 50 mol-%, from about 42 mol-% to about 47 mol-%, or about 45%. The mol-% of the helper lipid may be, for example, from about 30 mol-% to about 60 mol-%, from about 35 mol-% to about 55 mol-%, from about 40 mol-% to about 50 mol-%, from about 41 mol-% to about 46 mol-%, or about 44 mol-%. The mol-% of the neutral lipid may be, for example, from about 1 mol-% to about 20 mol-%, from about 5 mol-% to about 15 mol-%, from about 7 mol-% to about 12 mol-%, or about 9 mol-%. The mol-% of the stealth lipid may be, for example, from about 1 mol-% to about 10 mol-%, from about 1 mol-% to about 5 mol-%, from about 1 mol-% to about 3 mol-%, about 2 mol-%, or about 1 mol-%.

[0283] The LNPs can have different ratios between the positively charged amine groups of the biodegradable lipid (N) and the negatively charged phosphate groups (P) of the nucleic acid to be encapsulated. This may be mathematically represented by the equation N/P. For example, the N/P ratio may be from about 0.5 to about 100, from about 1 to about 50, from about 1 to about 25, from about 1 to about 10, from about 1 to about 7, from about 3 to about 5, from about 4 to about 5, about 4, about 4.5, or about 5. The N/P ratio can also be from about 4 to about 7 or from about 4.5 to about 6. In specific examples, the N/P ratio can be 4.5 or can be 6.

[0284] In some LNPs, the cargo can comprise Cas mRNA and gRNA. The Cas mRNA and gRNAs can be in different ratios. For example, the LNP formulation can include a ratio of Cas mRNA to gRNA nucleic acid ranging from about 25:1 to about 1:25, ranging from about 10:1 to about 1:10, ranging from about 5:1 to about 1:5, or about 1:1. Alternatively, the LNP formulation can include a ratio of Cas mRNA to gRNA nucleic acid from about 1:1 to about 1:5, or about 10:1. Alternatively, the LNP formulation can include a ratio of Cas mRNA to gRNA nucleic acid of about 1:10, 25:1, 10:1, 5:1, 3:1, 1:1, 1:3, 1:5, 1:10, or 1:25. Alternatively, the LNP formulation can include a ratio of Cas mRNA to gRNA nucleic acid of from about 1:1 to about 1:2. In specific examples, the ratio of Cas mRNA to gRNA can be about 1:1 or about 1:2.

[0285] In some LNPs, the cargo can comprise exogenous donor nucleic acid and gRNA. The exogenous donor nucleic acid and gRNAs can be in different ratios. For example, the LNP formulation can include a ratio of exogenous donor nucleic acid to gRNA nucleic acid ranging from about 25:1 to about 1:25, ranging from about 10:1 to about 1:10, ranging from about 5:1 to about 1:5, or about 1:1. Alternatively, the LNP formulation can include a ratio of exogenous donor nucleic acid to gRNA nucleic acid from about 1:1 to about 1:5, about 5:1 to about 1:1, about 10:1, or about 1:10. Alternatively, the LNP formulation can include a ratio of exogenous donor nucleic acid to gRNA nucleic acid of about 1:10, 25:1, 10:1, 5:1, 3:1, 1:1, 1:3, 1:5, 1:10, or 1:25.

[0286] A specific example of a suitable LNP has a nitrogen-to-phosphate (N/P) ratio of 4.5 and contains biodegradable cationic lipid, cholesterol, DSPC, and PEG2k-DMG in a 45:44:9:2 molar ratio. The biodegradable cationic lipid can be (9Z,12Z)-3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl octadeca-9,12-dienoate, also called 3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl (9Z,12Z)-octadeca-9,12-dienoate. See, e.g., Finn et al. (2018) *Cell Reports* 22:1-9, herein incorporated by reference in its entirety for all purposes. The Cas9 mRNA can be in a 1:1 ratio by weight to the guide RNA. Another specific example of a suitable LNP contains Dlin-MC3-DMA (MC3), cholesterol, DSPC, and PEG-DMG in a 50:38.5:10:1.5 molar ratio.

[0287] Another specific example of a suitable LNP has a nitrogen-to-phosphate (N/P) ratio of 6 and contains biodegradable cationic lipid, cholesterol, DSPC, and PEG2k-DMG in a 50:38:9:3 molar ratio. The biodegradable cationic lipid can be (9Z,12Z)-3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl octadeca-9,12-dienoate, also called 3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl (9Z,12Z)-octadeca-9,12-dienoate. The Cas9 mRNA can be in a 1:2 ratio by weight to the guide RNA.

[0288] The mode of delivery can be selected to decrease immunogenicity. For example, a Cas protein and a gRNA may be delivered by different modes (e.g., bi-modal delivery). These different modes may confer different pharmacodynamics or pharmacokinetic properties on the subject delivered molecule (e.g., Cas or nucleic acid encoding, gRNA or nucleic acid encoding, or exogenous donor nucleic acid/repair template). For example, the different modes can result in different tissue distribution, different half-life, or different temporal distribution. Some modes of delivery (e.g., delivery of a nucleic acid vector that persists in a cell

by autonomous replication or genomic integration) result in more persistent expression and presence of the molecule, whereas other modes of delivery are transient and less persistent (e.g., delivery of an RNA or a protein). Delivery of Cas proteins in a more transient manner, for example as mRNA or protein, can ensure that the Cas/gRNA complex is only present and active for a short period of time and can reduce immunogenicity caused by peptides from the bacterially-derived Cas enzyme being displayed on the surface of the cell by MHC molecules. Such transient delivery can also reduce the possibility of off-target modifications.

[0289] Administration *in vivo* can be by any suitable route including, for example, parenteral, intravenous, oral, subcutaneous, intra-arterial, intracranial, intrathecal, intraperitoneal, topical, intranasal, or intramuscular. Systemic modes of administration include, for example, oral and parenteral routes. Examples of parenteral routes include intravenous, intraarterial, intraosseous, intramuscular, intradermal, subcutaneous, intranasal, and intraperitoneal routes. A specific example is intravenous infusion. Nasal instillation and intravitreal injection are other specific examples. Local modes of administration include, for example, intrathecal, intracerebroventricular, intraparenchymal (e.g., localized intraparenchymal delivery to the striatum (e.g., into the caudate or into the putamen), cerebral cortex, precentral gyms, hippocampus (e.g., into the dentate gyrus or CA3 region), temporal cortex, amygdala, frontal cortex, thalamus, cerebellum, medulla, hypothalamus, tectum, tegmentum, or substantia nigra), intraocular, intraorbital, subconjunctival, intravitreal, subretinal, and transscleral routes. Significantly smaller amounts of the components (compared with systemic approaches) may exert an effect when administered locally (for example, intraparenchymal or intravitreal) compared to when administered systemically (for example, intravenously). Local modes of administration may also reduce or eliminate the incidence of potentially toxic side effects that may occur when therapeutically effective amounts of a component are administered systemically.

[0290] Administration *in vivo* can be by any suitable route including, for example, parenteral, intravenous, oral, subcutaneous, intra-arterial, intracranial, intrathecal, intraperitoneal, topical, intranasal, or intramuscular. A specific example is intravenous infusion. Compositions comprising the guide RNAs and/or Cas proteins (or nucleic acids encoding the guide RNAs and/or Cas proteins) can be formulated using one or more physiologically and pharmaceutically acceptable carriers, diluents, excipients or auxiliaries. The formulation can depend on the route of administration chosen. The term “pharmaceutically acceptable” means that the carrier, diluent, excipient, or auxiliary is compatible with the other ingredients of the formulation and not substantially deleterious to the recipient thereof.

[0291] The frequency of administration and the number of dosages can be depend on the half-life of the exogenous donor nucleic acids, guide RNAs, or Cas proteins (or nucleic acids encoding the guide RNAs or Cas proteins) and the route of administration among other factors. The introduction of nucleic acids or proteins into the cell or non-human animal can be performed one time or multiple times over a period of time. For example, the introduction can be performed at least two times over a period of time, at least three times over a period of time, at least four times over a period of time, at least five times over a period of time, at least six times over a period of time, at least seven times over a period

of time, at least eight times over a period of time, at least nine times over a period of times, at least ten times over a period of time, at least eleven times, at least twelve times over a period of time, at least thirteen times over a period of time, at least fourteen times over a period of time, at least fifteen times over a period of time, at least sixteen times over a period of time, at least seventeen times over a period of time, at least eighteen times over a period of time, at least nineteen times over a period of time, or at least twenty times over a period of time.

[0292] E. Measuring Delivery, Activity, or Efficacy of Human-TTR-Targeting Reagents *In Vivo* or *Ex Vivo*

[0293] The methods disclosed herein can further comprise detecting or measuring activity of human-TTR-targeting reagents. For example, if the human-TTR-targeting reagent is a genome editing reagent (e.g., CRISPR/Cas designed to target the human TTR locus), the measuring can comprise assessing the humanized TTR locus for modifications.

[0294] Various methods can be used to identify cells having a targeted genetic modification. The screening can comprise a quantitative assay for assessing modification of allele (MOA) of a parental chromosome. For example, the quantitative assay can be carried out via a quantitative PCR, such as a real-time PCR (qPCR). The real-time PCR can utilize a first primer set that recognizes the target locus and a second primer set that recognizes a non-targeted reference locus. The primer set can comprise a fluorescent probe that recognizes the amplified sequence. Other examples of suitable quantitative assays include fluorescence-mediated *in situ* hybridization (FISH), comparative genomic hybridization, isothermic DNA amplification, quantitative hybridization to an immobilized probe(s), INVADER® Probes, TAQMAN® Molecular Beacon probes, or ECLIPSE™ probe technology (see, e.g., US 2005/0144655, herein incorporated by reference in its entirety for all purposes).

[0295] Next-generation sequencing (NGS) can also be used for screening. Next-generation sequencing can also be referred to as “NGS” or “massively parallel sequencing” or “high throughput sequencing.” NGS can be used as a screening tool in addition to the MOA assays to define the exact nature of the targeted genetic modification and whether it is consistent across cell types or tissue types or organ types.

[0296] Assessing modification of the humanized TTR locus in a non-human animal can be in any cell type from any tissue or organ. For example, the assessment can be in multiple cell types from the same tissue or organ or in cells from multiple locations within the tissue or organ. This can provide information about which cell types within a target tissue or organ are being targeted or which sections of a tissue or organ are being reached by the human-TTR-targeting reagent. As another example, the assessment can be in multiple types of tissue or in multiple organs. In methods in which a particular tissue, organ, or cell type is being targeted, this can provide information about how effectively that tissue or organ is being targeted and whether there are off-target effects in other tissues or organs.

[0297] If the reagent is designed to inactivate the humanized TTR locus, affect expression of the humanized TTR locus, prevent translation of the humanized TTR mRNA, or clear the humanized TTR protein, the measuring can comprise assessing humanized TTR mRNA or protein expression. This measuring can be within the liver or particular cell

types or regions within the liver, or it can involve measuring serum levels of secreted humanized TTR protein.

[0298] Production and secretion of the humanized TTR protein can be assessed by any known means. For example, expression can be assessed by measuring levels of the encoded mRNA in the liver of the non-human animal or levels of the encoded protein in the liver of the non-human animal using known assays. Secretion of the humanized TTR protein can be assessed by measuring or plasma levels or serum levels of the encoded humanized TTR protein in the non-human animal using known assays.

IV. Methods of Making Non-Human Animals Comprising a Humanized TTR Locus

[0299] Various methods are provided for making a non-human animal comprising a humanized TTR locus as disclosed elsewhere herein. Any convenient method or protocol for producing a genetically modified organism is suitable for producing such a genetically modified non-human animal. See, e.g., Cho et al. (2009) *Current Protocols in Cell Biology* 42:19.11:19.11.1-19.11.22 and Gama Sosa et al. (2010) *Brain Struct. Funct.* 214(2-3):91-109, each of which is herein incorporated by reference in its entirety for all purposes. Such genetically modified non-human animals can be generated, for example, through gene knock-in at a targeted Ttr locus.

[0300] For example, the method of producing a non-human animal comprising a humanized TTR locus can comprise: (1) modifying the genome of a pluripotent cell to comprise the humanized TTR locus; (2) identifying or selecting the genetically modified pluripotent cell comprising the humanized TTR locus; (3) introducing the genetically modified pluripotent cell into a non-human animal host embryo; and (4) implanting and gestating the host embryo in a surrogate mother. Optionally, the host embryo comprising modified pluripotent cell (e.g., a non-human ES cell) can be incubated until the blastocyst stage before being implanted into and gestated in the surrogate mother to produce an F0 non-human animal. The surrogate mother can then produce an F0 generation non-human animal comprising the humanized TTR locus.

[0301] The methods can further comprise identifying a cell or animal having a modified target genomic locus (i.e., a humanized TTR locus). Various methods can be used to identify cells and animals having a targeted genetic modification.

[0302] The screening step can comprise, for example, a quantitative assay for assessing modification of allele (MOA) of a parental chromosome. For example, the quantitative assay can be carried out via a quantitative PCR, such as a real-time PCR (qPCR). The real-time PCR can utilize a first primer set that recognizes the target locus and a second primer set that recognizes a non-targeted reference locus. The primer set can comprise a fluorescent probe that recognizes the amplified sequence.

[0303] Other examples of suitable quantitative assays include fluorescence-mediated in situ hybridization (FISH), comparative genomic hybridization, isothermic DNA amplification, quantitative hybridization to an immobilized probe (s), INVADER® Probes, TAQMAN® Molecular Beacon probes, or ECLIPSE™ probe technology (see, e.g., US 2005/0144655, incorporated herein by reference in its entirety for all purposes).

[0304] An example of a suitable pluripotent cell is an embryonic stem (ES) cell (e.g., a mouse ES cell or a rat ES cell). The modified pluripotent cell can be generated, for example, through recombination by (a) introducing into the cell one or more targeting vectors comprising an insert nucleic acid flanked by 5' and 3' homology arms corresponding to 5' and 3' target sites, wherein the insert nucleic acid comprises a humanized TTR locus; and (b) identifying at least one cell comprising in its genome the insert nucleic acid integrated at the endogenous Ttr locus. Alternatively, the modified pluripotent cell can be generated by (a) introducing into the cell: (i) a nuclease agent, wherein the nuclease agent induces a nick or double-strand break at a target sequence within the endogenous Ttr locus; and (ii) one or more targeting vectors comprising an insert nucleic acid flanked by 5' and 3' homology arms corresponding to 5' and 3' target sites located in sufficient proximity to the target sequence, wherein the insert nucleic acid comprises the humanized TTR locus; and (c) identifying at least one cell comprising a modification (e.g., integration of the insert nucleic acid) at the endogenous Ttr locus. Any nuclease agent that induces a nick or double-strand break into a desired target sequence can be used. Examples of suitable nucleases include a Transcription Activator-Like Effector Nuclease (TALEN), a zinc-finger nuclease (ZFN), a meganuclease, and Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas) systems or components of such systems (e.g., CRISPR/Cas9). See, e.g., US 2013/0309670 and US 2015/0159175, each of which is herein incorporated by reference in its entirety for all purposes.

[0305] In a specific example, a method of making a non-human animal comprising a humanized TTR locus can comprise: (a) introducing into a non-human animal embryonic stem (ES) cell: (i) a nuclease agent that targets a target sequence in the endogenous Ttr locus; and (ii) a targeting vector comprising a nucleic acid insert comprising the human TTR sequence flanked by a 5' homology arm corresponding to a 5' target sequence in the endogenous Ttr locus and a 3' homology arm corresponding to a 3' target sequence in the endogenous Ttr locus, wherein the targeting vector recombines with the endogenous Ttr locus to produce a genetically modified non-human ES cell comprising in its genome the genetically modified endogenous Ttr locus comprising the human TTR sequence; (b) introducing the genetically modified non-human ES cell into a non-human animal host embryo; and (c) gestating the non-human animal host embryo in a surrogate mother, wherein the surrogate mother produces an F0 progeny genetically modified non-human animal comprising in its genome the genetically modified endogenous Ttr locus comprising the human TTR sequence.

[0306] In some such methods, the nuclease agent can comprise a Cas protein (e.g., a Cas9 protein) and a guide RNA that targets a target sequence in the endogenous Ttr locus, but other suitable nuclease agents can also be used. CRISPR/Cas and CRISPR/Cas9 systems are described in more detail elsewhere herein. Optionally, two or more (e.g., three or four) nuclease agents (e.g., guide RNAs) can be used. The target sequence(s) can be any suitable target sequence within the endogenous Ttr locus. For example, the target sequence(s) can be within about 10, about 20, about 30, about 40, about 50, about 100, about 200, about 300, about 400, about 500, about 1000 nucleotides about 2000,

about 3000, about 4000, or about 5000 nucleotides of the start codon and/or the stop codon (e.g., one target sequence in proximity to the start codon and one target sequence in proximity to the stop codon).

[0307] In some such methods, the targeting vector is a large targeting vector at least 10 kb in length or in which the sum total of the 5' and 3' homology arms is at least 10 kb in length, but other types of exogenous donor nucleic acids can also be used and are well-known. The 5' and 3' homology arms can correspond with 5' and 3' target sequences, respectively, that flank the region being replaced by the human TTR insert nucleic acid or that flank the region into which the human TTR insert nucleic acid is to be inserted. The exogenous donor nucleic acid or targeting vector can recombine with the target locus via homology directed repair or can be inserted via NHEJ-mediated insertion to generate the humanized TTR locus.

[0308] The donor cell can be introduced into a host embryo at any stage, such as the blastocyst stage or the pre-morula stage (i.e., the 4 cell stage or the 8 cell stage). Progeny that are capable of transmitting the genetic modification through the germline are generated. See, e.g., U.S. Pat. No. 7,294,754, herein incorporated by reference in its entirety for all purposes.

[0309] Alternatively, the method of producing the non-human animals described elsewhere herein can comprise: (1) modifying the genome of a one-cell stage embryo to comprise the humanized TTR locus using the methods described above for modifying pluripotent cells; (2) selecting the genetically modified embryo; and (3) implanting and gestating the genetically modified embryo into a surrogate mother. Progeny that are capable of transmitting the genetic modification through the germline are generated.

[0310] Nuclear transfer techniques can also be used to generate the non-human mammalian animals. Briefly, methods for nuclear transfer can include the steps of: (1) enucleating an oocyte or providing an enucleated oocyte; (2) isolating or providing a donor cell or nucleus to be combined with the enucleated oocyte; (3) inserting the cell or nucleus into the enucleated oocyte to form a reconstituted cell; (4) implanting the reconstituted cell into the womb of an animal to form an embryo; and (5) allowing the embryo to develop. In such methods, oocytes are generally retrieved from deceased animals, although they may be isolated also from either oviducts and/or ovaries of live animals. Oocytes can be matured in a variety of well-known media prior to enucleation. Enucleation of the oocyte can be performed in a number of well-known manners. Insertion of the donor cell or nucleus into the enucleated oocyte to form a reconstituted cell can be by microinjection of a donor cell under the zona pellucida prior to fusion. Fusion may be induced by application of a DC electrical pulse across the contact/fusion plane (electrofusion), by exposure of the cells to fusion-promoting chemicals, such as polyethylene glycol, or by way of an inactivated virus, such as the Sendai virus. A reconstituted cell can be activated by electrical and/or non-electrical means before, during, and/or after fusion of the nuclear donor and recipient oocyte. Activation methods include electric pulses, chemically induced shock, penetration by sperm, increasing levels of divalent cations in the oocyte, and reducing phosphorylation of cellular proteins (as by way of kinase inhibitors) in the oocyte. The activated reconstituted cells, or embryos, can be cultured in well-known media and then transferred to the womb of an animal.

See, e.g., US 2008/0092249, WO 1999/005266, US 2004/0177390, WO 2008/017234, and U.S. Pat. No. 7,612,250, each of which is herein incorporated by reference in its entirety for all purposes.

[0311] The various methods provided herein allow for the generation of a genetically modified non-human F0 animal wherein the cells of the genetically modified F0 animal comprise the humanized TTR locus. Depending on the method used to generate the F0 animal, the number of cells within the F0 animal that have the humanized TTR locus will vary. The introduction of the donor ES cells into a pre-morula stage embryo from a corresponding organism (e.g., an 8-cell stage mouse embryo) via for example, the VELOCIMOUSE® method allows for a greater percentage of the cell population of the F0 animal to comprise cells having the nucleotide sequence of interest comprising the targeted genetic modification. For example, at least 50%, 60%, 65%, 70%, 75%, 85%, 86%, 87%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% of the cellular contribution of the non-human F0 animal can comprise a cell population having the targeted modification.

[0312] The cells of the genetically modified F0 animal can be heterozygous for the humanized TTR locus or can be homozygous for the humanized TTR locus.

[0313] All patent filings, websites, other publications, accession numbers and the like cited above or below are incorporated by reference in their entirety for all purposes to the same extent as if each individual item were specifically and individually indicated to be so incorporated by reference. If different versions of a sequence are associated with an accession number at different times, the version associated with the accession number at the effective filing date of this application is meant. The effective filing date means the earlier of the actual filing date or filing date of a priority application referring to the accession number if applicable. Likewise, if different versions of a publication, website or the like are published at different times, the version most recently published at the effective filing date of the application is meant unless otherwise indicated. Any feature, step, element, embodiment, or aspect of the invention can be used in combination with any other unless specifically indicated otherwise. Although the present invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

BRIEF DESCRIPTION OF THE SEQUENCES

[0314] The nucleotide and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three-letter code for amino acids. The nucleotide sequences follow the standard convention of beginning at the 5' end of the sequence and proceeding forward (i.e., from left to right in each line) to the 3' end. Only one strand of each nucleotide sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand. When a nucleotide sequence encoding an amino acid sequence is provided, it is understood that codon degenerate variants thereof that encode the same amino acid sequence are also provided. The amino acid sequences follow the standard convention of beginning at the amino terminus of

the sequence and proceeding forward (i.e., from left to right in each line) to the carboxy terminus.

TABLE 2

Description of Sequences.		
SEQ ID NO	Type	Description
1	Protein	Human TTR Protein NP_000362.1 and P02766.1
2	Protein	Expected Chimeric Mouse/Human TTR Protein - Humanization V2
3	DNA	Human TTR Gene NG_009490.1
4	DNA	Human TTR mRNA NM_000371.3
5	DNA	Mouse Ttr gene NC_000084.6
6	Protein	Mouse TTR protein P07309.1 and NP_038725.1
7	DNA	Mouse Ttr mRNA NM_013697.5
8	RNA	Generic Guide RNA Scaffold v.2
9	RNA	Generic Guide RNA Scaffold v.3
10	RNA	Generic Guide RNA Scaffold v.4
11	DNA	Generic Guide RNA Target Sequence plus PAM v.1
12	DNA	Generic Guide RNA Target Sequence plus PAM v.2
13	DNA	Generic Guide RNA Target Sequence plus PAM v.3
14	DNA	Expected Humanization V1 - F0, with SDC Pmc12 UbiNeo cassette
15	DNA	Expected Humanization V1 - F1, Cassette-Deleted
16	DNA	Expected Humanization V2 - F0, with SDC Pmc12 UbiNeo cassette
17	DNA	Expected Humanization V2 - F1, Cassette-Deleted
18	DNA	Human TTR Sequence Inserted in Humanized TTR V1
19	DNA	Human TTR Sequence Inserted in Humanized TTR V2
20	DNA	Mouse Ttr Locus - Start Codon to Stop Codon
21	DNA	9090retU3 - F Primer
22	DNA	9090retU2 - F Primer
23	DNA	9090retU - F Primer
24	DNA	9090mTGU - F Primer
25	DNA	7576mTU - F Primer
26	DNA	9090mTM - F Primer
27	DNA	7576mTD - F Primer
28	DNA	9090mTGD - F Primer
29	DNA	9090retD - F Primer
30	DNA	9090retD2 - F Primer
31	DNA	9090retD3 - F Primer
32	DNA	7576hTU - F Primer
33	DNA	7576hTD - F Primer
34	DNA	Neo - F Primer
35	DNA	7655hTU - F Primer
36	DNA	9212mTU - F Primer
37	DNA	9212mTGD - F Primer
38	DNA	7655mTU - F Primer
39	DNA	7655mTD - F Primer
40	DNA	9204mretD - F Primer
41	DNA	9204mretU - F Primer
42	DNA	4552mTU - F Primer
43	DNA	9090retU3 - R Primer
44	DNA	9090retU2 - R Primer
45	DNA	9090retU - R Primer
46	DNA	9090mTGU - R Primer
47	DNA	7576mTU - R Primer
48	DNA	9090mTM - R Primer
49	DNA	7576mTD - R Primer
50	DNA	9090mTGD - R Primer
51	DNA	9090retD - R Primer
52	DNA	9090retD2 - R Primer
53	DNA	9090retD3 - R Primer
54	DNA	7576hTU - R Primer
55	DNA	7576hTD - R Primer
56	DNA	Neo - R Primer
57	DNA	7655hTU - R Primer
58	DNA	9212mTU - R Primer
59	DNA	9212mTGD - R Primer
60	DNA	7655mTU - R Primer
61	DNA	7655mTD - R Primer
62	DNA	9204mretD - R Primer
63	DNA	9204mretU - R Primer
64	DNA	4552mTU - R Primer
65	DNA	9090retU3 - Probe

TABLE 2-continued

Description of Sequences.		
SEQ ID NO	Type	Description
66	DNA	9090retU2 - Probe
67	DNA	9090retU - Probe
68	DNA	9090mTGU - Probe
69	DNA	7576mTU - Probe
70	DNA	9090mTM - Probe
71	DNA	7576mTD - Probe
72	DNA	9090mTGD - Probe
73	DNA	9090retD - Probe
74	DNA	9090retD2 - Probe
75	DNA	9090retD3 - Probe
76	DNA	7576hTU - Probe
77	DNA	7576hTD - Probe
78	DNA	Neo - Probe
79	DNA	7655hTU - Probe
80	DNA	9212mTU - Probe
81	DNA	9212mTGD - Probe
82	DNA	7655mTU - Probe
83	DNA	7655mTD - Probe
84	DNA	9204mretD - Probe
85	DNA	9204mretU - Probe
86	DNA	4552mTU - Probe
87	RNA	crRNA tail
88	RNA	tracrRNA
89	RNA	Generic Guide RNA Scaffold v.1
90	DNA	Humanized TTR CDS v1.0
91	DNA	Humanized TTR CDS v2.0
92	DNA	Mouse TTR CDS
93	DNA	Cas9 DNA Sequence
94	Protein	Cas9 Protein Sequence

EXAMPLES

Example 1. Generation of Mice Comprising a Humanized TTR Locus

[0315] One promising therapeutic approach for the TTR amyloidosis diseases is to reduce the TTR load in the patient by inactivation of the gene with genome editing technology, such as CRISPR/Cas9 technology. To assist in the development of CRISPR/Cas9 therapeutics, mice with targeted modifications in the Ttr gene were developed.

[0316] The first Ttr allele made was a complete deletion of the mouse transthyretin coding sequence and its replacement with the orthologous part of the human TTR gene. A large targeting vector comprising a 5' homology arm including 33.7 kb of sequence upstream from the mouse Ttr start codon and 34.5 kb of the sequence downstream of the mouse Ttr stop codon was generated to replace the approximately 8.3 kb region from the mouse Ttr start codon to the mouse Ttr stop codon with the approximately 7.1 kb orthologous human TTR sequence from the human TTR start codon to the end of the last human TTR exon (exon 4, including the human 3' UTR) and a self-deleting neomycin selection cassette (SDC Neo) flanked by loxP sites. See FIG. 3. The SDC Neo cassette includes the following components from 5' to 3': loxP site, mouse protamine (Prm1) promoter, Crei (Cre coding sequence optimized to include intron), polyA, human ubiquitin promoter, neomycin phosphotransferase (neo,) coding sequence, polyA, loxP. To generate the humanized allele, CRISPR/Cas9 components targeting the mouse Ttr locus were introduced into F1H4 mouse embryonic stem cells together with the large targeting vector. Loss-of-allele assays, gain-of-allele assays, retention assays,

and CRISPR assays using primers and probes set forth in FIG. 5A and in Table 3 were performed to confirm the humanization of the mouse Ttr allele. Loss-of-allele, gain-of-allele assays, and retention assays are described, for example, in US 2014/0178879; US 2016/0145646; WO 2016/081923; and Frendewey et al. (2010) *Methods Enzymol.* 476:295-307, each of which is herein incorporated by reference in its entirety for all purposes. CRISPR assays are TAQMAN® assays designed to cover the region that is disrupted by the CRISPR gRNAs. When a CRISPR gRNA cuts and creates an indel (insertion or deletion), the TAQMAN® assay will fail to amplify and thus reports CRISPR cleavage. Versions with the SDC Neo cassette and after excision of the SDC Neo cassette are shown in FIG. 3. F0 mice were then generated using the VELOCIMOUSE® method. See, e.g., U.S. Pat. Nos. 7,576,259; 7,659,442; 7,294,754; US 2008/007800; and Poueymirou et al. (2007) *Nature Biotech.* 25(1):91-99, each of which is herein incorporated by reference in its entirety for all purposes.

[0317] F0 generation mice (50% C57BL/6NTac and 50% 129S6/SvEvTac) were generated from multiple humanized ES cell clones, including clones 7576A-A5, 7576C-G7, and 7576B-F10. The sequence for the expected humanized mouse Ttr locus in the F0 generation mice is set forth in SEQ

ID NO: 14 and includes the SDC Neo cassette. F1 and F2 generation mice (75% C57BL/6NTac and 25% 129S6/SvEvTac) were then generated by breeding. The sequence for the expected humanized mouse Ttr locus in the F1 and F2 generation mice is set forth in SEQ ID NO: 15 and does not include the SDC Neo cassette.

[0318] A comparison of the human and mouse transthyretin precursor protein sequences is shown in FIG. 1A, a comparison of the human and mouse transthyretin coding sequences is shown in FIG. 1B, and a schematic showing the wild type mouse Ttr locus and the final humanized mouse Ttr locus (humanized TTR version 1 with the SDC Neo cassette deleted) is shown in FIG. 2. The endogenous mouse Ttr locus sequence from the start codon to the stop codon is provided in SEQ ID NO: 20. Sequences for the expected humanized mouse Ttr locus with the SDC Neo cassette and without the SDC Neo cassette are set forth in SEQ ID NOS: 14 and 15, respectively. The expected transthyretin precursor protein encoded by the humanized mouse Ttr locus is set forth in SEQ ID NO: 1. This allele provides the true human target of human TTR CRISPR/Cas9 therapeutics, thereby enabling testing of the efficacy and mode of action of CRISPR/Cas9 therapeutics in live animals as well as pharmacokinetic and pharmacodynamics studies in a setting where the human protein is the only version of TTR present.

TABLE 3

Primers and Probes for Loss-of-Allele Assays, Gain-of-Allele Assays, Retention Assays, and CRISPR Assays.			
Assay	Forward Primer	Reverse Primer	Probe
9090retU3	CACAGACAATCAGACGTACCAGTA (SEQ ID NO: 21)	GGGACATCTCGGTTTCCTGACTT (SEQ ID NO: 43)	TCATGTAATCTGGCTTCAGAGTGGGA (SEQ ID NO: 65)
9090retU2	CCAGCTTTGCCAGTTTACGA (SEQ ID NO: 22)	TCCACACTACTGAACTCCACAA (SEQ ID NO: 44)	TGGGAGGCAATCTTAGTTTCAATGGA (SEQ ID NO: 66)
9090retU	TTGGACGGTTGCCCTCTT (SEQ ID NO: 23)	CGGAACACTCGCTCTACGAAA (SEQ ID NO: 45)	TCCCAAAGGTGTCTGTCTGCACA (SEQ ID NO: 67)
9090mTGU	GATGGCTTCCCTTCGACTCTTC (SEQ ID NO: 24)	GGCCAGCTTCAGACACA (SEQ ID NO: 46)	CTCCTTTGCCCTCGCTGGACTGG (SEQ ID NO: 68)
7576TU	CACTGACATTTCTCTGTCTCCTCTCCAGGGTGTGGAGAATCCAA (SEQ ID NO: 25)	(SEQ ID NO: 47)	CGGACAGCATCCAGGACTT (SEQ ID NO: 69)
9090TM	GGGCTCACCACAGATGAGAAG (SEQ ID NO: 26)	GCCAAGTGTCTTCCAGTACGAT (SEQ ID NO: 48)	AGAAGGAGTGTACAGAGTAGAACTGGACA (SEQ ID NO: 70)
7576TD	CACTGTTCGCCACAGGTCTT (SEQ ID NO: 27)	GTCCCTTTCTTGGGTTTCAGA (SEQ ID NO: 49)	TGTTTGTGGGTGTCTAGTGTCTACTC (SEQ ID NO: 71)
9090mTGD	GCTCAGCCCATACTCCTACA (SEQ ID NO: 28)	GATGCTACTGCTTTGGCAAGATC (SEQ ID NO: 50)	CACCACGGCTGTCGTAGCA (SEQ ID NO: 72)
9090retD	GCCCAGGAGGACCGGAT (SEQ ID NO: 29)	CCTGAGCTGCTAACACGGTT (SEQ ID NO: 51)	CTTGCCAAAGCAGTAGCATCCCA (SEQ ID NO: 73)
9090retD2	GGCAACTTGCTTGAGGAAGA (SEQ ID NO: 30)	AGTACAGACCATGCTTAGTGTA (SEQ ID NO: 52)	AGGTCAGAAAGCAGAGTGGACCA (SEQ ID NO: 74)
9090retD3	GCAGCAACCCAGCTTCACTT (SEQ ID NO: 31)	TGCCAGTTTAGGAGGAATATGTTT (SEQ ID NO: 53)	CCCAGGCAATTCCTACCTTCCCA (SEQ ID NO: 75)
7576hTU	ACTGAGCTGGGACTTGAAC (SEQ ID NO: 32)	CTGAGGAAACAGAGGTACCAGATATTCTGAGCATTCTACCTCATTGCTTTGGT (SEQ ID NO: 54)	(SEQ ID NO: 76)
7576hTD	TGCCTCACTCTGAGAACCA (SEQ ID NO: 33)	AGTCACACAGTTCTGTCAAATCAG (SEQ ID NO: 55)	AGGCTGTCCCAGCCTGAGTCG (SEQ ID NO: 77)

TABLE 3-continued

Primers and Probes for Loss-of-Allele Assays, Gain-of-Allele Assays, Retention Assays, and CRISPR Assays.			
Assay	Forward Primer	Reverse Primer	Probe
Neo	GGTGAGAGGCTATTCCGGC (SEQ ID NO: 34)	GAACACGGCGGCATCAG (SEQ ID NO: 56)	TGGGCACAACAGACAATCGGCTG (SEQ ID NO: 78)
7655hTU	GGCCGTGCATGTGTTTCAG (SEQ ID NO: 35)	TCCTGTGGGAGGGTTCTTTG (SEQ ID NO: 57)	AAGGCTGCTGATGACACCTGGGA (SEQ ID NO: 79)
9212mTU	GGTTCCTTTGCTCTTATTTCGT (SEQ ID NO: 36)	CCCTCTCTCTGAGCCCTCTA (SEQ ID NO: 58)	AGATTCAGACACACACAACCTACCAGC (SEQ ID NO: 80)
9212mTGD	CCCACACTGCAGAAGGAAACTTG (SEQ ID NO: 37)	GCTGCCTAAGTCTTTGGAGCT (SEQ ID NO: 59)	AGACCTGCAATTCTCTAAGAGCTCCACA (SEQ ID NO: 81)
7655mTU	GGTTCCTTTGCTCTTATTTCGT (SEQ ID NO: 38)	CCCTCTCTCTGAGCCCTCTA (SEQ ID NO: 60)	AGATTCAGACACACACAACCTACCAGC (SEQ ID NO: 82)
7655TD	CCAGCTTAGCATCCTGTGAACA (SEQ ID NO: 39)	GAGAGGAGAGACAGCTAGTTCTAACTTGTCTGCAGCTCCTACCTCTGGG (SEQ ID NO: 61)	(SEQ ID NO: 83)
9204mretD	GGCAACTTGCTTGAGGAAGA (SEQ ID NO: 40)	AGCTACAGACCATGCTTAGTGTA (SEQ ID NO: 62)	AGGTGAGAAAGCAGAGTGGACCA (SEQ ID NO: 84)
9204mretU	TGTGGAGTTCAGTAGTGTGGAG (SEQ ID NO: 41)	GCCCTCTTCATACAGGAATCAC (SEQ ID NO: 63)	TTGACATGTGTGGGTGAGAGATTTACTG (SEQ ID NO: 85)
4552TU	CACTGACATTTCTCTGTCTCCTCTCGGACAGCATCCAGGACTT (SEQ ID NO: 42)	(SEQ ID NO: 64)	CCCAGGGTCTGGAGAATCCAA (SEQ ID NO: 86)

Example 2. Characterization of Mice Comprising a Humanized TTR Locus

[0319] Humanized TTR mice F0 cohorts from clones 7576A-A5 and 7576C-G7 were then characterized. As shown in FIG. 6, humanized TTR mRNA was robustly expressed in the liver of 8-week old, homozygous F0 generation humanized TTR mice. Equal mass amounts of RNA from each tissue were assayed by RT-qPCR. Five mice were assayed per genotype, with four technical replicates per sample. Each tissue had the RNA extracted. The genomic DNA was degraded so that it would not count towards the qPCR reaction. The RNA was reverse transcribed, and assays specific to human TTR and mouse Ttr were used to detect human TTR transcripts and mouse Ttr transcripts, respectively. As expected, the homozygous humanized TTR mouse showed significant expression of human TTR in liver (ct values below 30), while WT mice showed ct values of 30 and higher indicating that there was no expression of human TTR. In contrast, the wild type mouse showed significant expression of mouse Ttr in the liver, while homozygous humanized TTR mice showed ct values of 30 and higher indicating that there is no endogenous expression of mouse Ttr.

[0320] An ELISA assay was done to assess human TTR and mouse TTR protein levels in serum and cerebrospinal fluid from 8-week-old, homozygous F0 generation humanized TTR mice. A human TTR ELISA kit (Aviva Systems Biology; Cat No.: OKIA00081; 1:250 dilution for serum; 1:1000 dilution for CSF) was used to assess human TTR levels. A mouse TTR ELISA kit (Aviva Systems Biology; Cat No: OKIA00111; 1:4000 dilution for serum; 1:10000 dilution for CSF) was used to assess mouse TTR levels. Human serum and human CSF were used as positive controls for human TTR and negative controls for mouse TTR, and F1H4 serum and F1H4 CSF were used as negative controls for human TTR and positive controls for mouse TTR. As shown in FIG. 7A, human TTR was detected in the

serum from the humanized TTR mice but not in serum from wild type (F1H4) mice. As shown in FIG. 7B, mouse TTR was not detected in the serum from the humanized TTR mice but was detected in wild type mouse serum. Human and mouse TTR levels in serum were further assessed in humanized TTR mice derived from two separate humanized mouse Ttr ES cell clones: 7576C-G7 and 7576A-A5. As shown in FIG. 7C, human TTR was detected in the serum of humanized TTR mice derived from clone 7576A-A5.

[0321] Human TTR protein expression was also assessed in 8-week-old, homozygous humanized TTR mice by western blot on serum samples, liver samples, and kidney samples. The results are shown in FIGS. 8-9. Serum samples (5 μ L total volume per well) were boiled in Laemmli buffer (containing SDS and beta-mercaptoethanol) and resolved on a 4-20% denaturing gradient gel (anti-TTR antibody: 1:1000; Abcam; ab75815). Mouse IgG (anti mouse IgG-HRP: 1:7500, Jackson ImmunoResearch) was used as a loading control. Three mice per group were tested for humanized mouse Ttr clones 7576C-G7 and 7576A-A5. Five mice per group were tested for wild type mouse control (F1H4). Mouse serum and human serum were used as negative and positive controls, respectively. As shown in FIG. 8, human TTR was detected by western blot in serum from both humanized mouse Ttr clones.

[0322] Liver and kidney samples (100 μ g total protein per well) were boiled in Laemmli buffer (containing SDS and beta-mercaptoethanol) and resolved on a 4-20% denaturing gradient gel (anti-TTR antibody: 1:1000; Abcam; ab75815). GAPDH (anti-GAPDH: 1:2000, Santa Cruz) was used as a loading control. Three mice per group were tested for humanized mouse Ttr clones 7576C-G7 and 7576A-A5. Five mice per group were tested for wild type mouse control (F1H4). Mouse serum and human serum were used as negative and positive controls, respectively. As shown in

FIG. 9, human TTR was detected by western blot in serum from both homozygous humanized TTR mice generated from clone 7576A-A5.

[0323] In summary, TTR HumIn (TTR^{7576/7576}) F0 mice were found to have a detectable amount of circulating hTTR. In addition, mice from clone 7576C-A5 had detectable amounts of hTTR in liver and plasma.

[0324] We hypothesized that removal of the neomycin drug selection cassette may increase secretion of the human TTR. Human TTR levels were measured in plasma samples from non-terminal, submandibular bleeds on 5-week-old mice homozygous for the fully humanized mouse Ttr locus with the neomycin selection cassette (TTR^{7576/7576}), mice heterozygous for the fully humanized mouse Ttr locus with the neomycin selection cassette (TTR^{7576/WT}), mice heterozygous for the fully humanized mouse Ttr locus without the neomycin selection cassette (TTR^{7577/WT}), and wild type mice (F1H4). Human TTR levels were assayed with the AssayPro human TTR (hTTR) ELISA kit (cat no.: EP3010-1; 1:40000 dilution). Mouse TTR serum levels were assayed with the Aviva Systems Biology mouse TTR (mTTR) ELISA kit (cat no. OKIA00111; 1:4000 dilution). The AssayPro human TTR ELISA kit was previously determined to be specific for detecting human TTR but not mouse TTR, and the Aviva Systems Biology mouse TTR ELISA kit was previously determined to be specific for detecting mouse TTR but not human TTR (data not shown). As shown in Table 4, removal of the neomycin selection cassette resulted in a statistically significant increase in human TTR levels in the serum. MAID7576 refers to the humanized TTR locus with the neomycin selection cassette. MAID7577 refers to the humanized TTR locus with the neomycin selection cassette removed. Enhanced human TTR mRNA expression was also observed in the liver (data not shown). Mice heterozygous for hTTR and mTTR (TTR-WT^{7576/WT} and TTR-WT^{7577/WT}) had increased circulating hTTR, possibly due to increased stability from heteromeric (e.g., cross-TTR species) interaction.

TABLE 4

Circulating Human and Mouse TTR Levels.				
	TTR ^{7576/7576}	TTR ^{7576/WT}	TTR ^{7577/WT}	F1H4
mTTR, µg/mL (SD)	N.D.	207.62 (15.39)	359.9 (38.07)**	919.96 (79.73)
hTTR, µg/mL (SD)	0.61 (0.43)	28.507 (5.61)	39.93 (3.70)**	N.D.

[0325] Serum and liver TTR levels were also measured in F2 generation homozygous humanized TTR mice (three per group) that were generated from a different clone: 7576B-F10. As shown in FIG. 14 (Tris-saline sucrose (TSS) control sample), human TTR was detected in liver samples at a level of more than 1000 ng/mL. As shown in FIGS. 15A and 15B (TSS sample), human TTR was detected in serum samples at a level of 80,000 ng/mL or higher.

[0326] In another experiment, blood was collected via submandibular bleeds from TTR WT HumIn (v1.0, hTTR^{7577/7577}, clone B-F10) F2 homozygous mice at 3 months of age. hTTR levels were determined in plasma using species-specific ELISA kits (hTTR, Aviva, cat # OKIA00081; mTTR, Aviva, cat # OKIA00111). As shown in FIGS. 17A and 17B and Table 5, hTTR was secreted into circulation in TTR WT HumIn (v1.0, clone B-F10) F2 homozygous mice at 52.1±10.7 µg/mL, with no detectable levels of mTTR. mRNA levels of mTTR and hTTR in liver samples from the wild type control mice (F1H4) and WT HumIn (v1.0, hTTR^{7577/7577}, clone B-F10) mice are shown in FIG. 17C.

TABLE 5

Plasma hTTR and mTTR Levels.		
	hTTR, µg/mL (SD)	mTTR, µg/mL (SD)
TTR ^{WT/WT}	N.D.	831.5 (129.9)
TTR ^{7577/7577}	52.1 (10.7)	Not detectable
Human serum	234.5 (n.a.)	Not detectable

Example 3. Use of Mice Comprising a Humanized TTR Locus to Test Guide RNAs Targeting Human TTR Ex Vivo and In Vivo

[0327] F0 generation humanized TTR mice cohorts were then used to test guide RNAs targeting human TTR ex vivo and in vivo. As a proof of concept, human TTR guide RNAs were first tested in primary hepatocytes isolated from F0 generation humanized TTR mice produced from clone 7576C-G7. Livers from huTTR^{h/h} mice were perfused with 50 mL liver perfusion medium containing 1× PenStrep, followed by 50 mL liver digestion medium (HBSS, 100 mM CaCl₂, 500 mM HEPES, collagenase). Once livers appeared digested, they were placed into wash medium containing 1× PenStrep and L-glutamine. The livers were torn to release the hepatocytes from the liver through gentle shaking. Once cells were released, they were put through a 70 µm mesh filter and spun at 50 g for 4 minutes at 4° C. The pellets were washed 2× with wash buffer. The pellets were then re-suspended in 20 mL of 38-40% Percoll and spun at 200 g×10 min at 4° C. The pellet was washed 2× and re-suspended in plating medium (Williams E Media, 1× PenStrep, 1× L-glutamine, 5% FBS). Cells were plated at 300,000 cells per well in 24-well collagen-coated tissue culture plates. After the cells were allowed to attach for 6-18 hours, the plating medium was replaced with medium without FBS. Reagents used are shown in Table 6.

TABLE 6

Reagents for isolation of primary hepatocytes.	
Material	Catalog Number
Liver Perfusion Media	Gibco [17701-038]
HBSS (1x)	Gibco [14175-079]
Hepatocyte Wash Media	Gibco [17704-024]
Williams E media	Gibco [A12176-01]
Penstrep (100x)	Gibco [15140163]
L-glutamine (200 mM)	Gibco [25030081]
FBS supplement	Gibco [A13450]
HEPES	Gibco [15630080]
Collagen	Gibco [A1048301]
Acetic acid	Sigma [A6283]
Liberase TM	Roche [TM05401119001]
Primary Hepatocyte Thawing and Plating Supplements	Gibco [CM3000]
Primary Hepatocyte Maintenance Supplements	Gibco [CM4000]
Percoll	GE [17-0891-01]

[0328] Lipid nanoparticles (LNPs) (containing Cas9 mRNA plus a human TTR gRNA (versions 1 and 2, each targeting human TTR exon 2)) were added to the hepatocytes 24 hours post-isolation. Briefly, for each well, LNPs were added at a concentration of 500 ng in 500 μ L of hepatocyte maintenance medium containing 3% mouse serum and were incubated for 5 minutes at 37° C. Plated cells were washed and 500 μ L of incubated LNPs in medium was added to each well. After 48 hours, DNA lysates were prepared from the cells, and next-generation sequencing was performed for each guide RNA tested.

[0329] As shown in FIG. 10, editing in the humanized TTR gene was seen with both guide RNAs in primary hepatocytes isolated from humanized TTR mice. Human TTR guide RNA 1 had an editing efficiency of 20.4%, and human TTR guide RNA 2 had an editing efficiency of 7.53%. Similar results were observed for a human TTR guide RNA targeting exon 3 (editing efficiency of 17.37%; data not shown). Editing efficiency refers to the total number of insertions or deletions observed over the total number of sequences read in the PCR reaction from a pool of lysed cells as determined by next generation sequencing.

[0330] Next, human TTR guide RNAs 1 and 2 were tested in vivo in humanized TTR mice. F0 generation humanized TTR mice (Ttr^{h/h}) from clone numbers 7576A-A5 and 7576C-G7 were used. Three animal groups were targeted with fresh LNPs as shown in Table 7.

TABLE 7

Animal Groups for LNP Study.	
Group	Description
1	Ttr ^{h/h} 1M 1F A-A5 and 2M 1F C-G7: LNP(gRNA1)@ 2 mg/kg
2	Ttr ^{h/h} 1M 1F A-A5 and 2M 1F C-G7: LNP(gRNA2) @ 2 mg/kg
3	Ttr ^{h/h} 1M A-A5 and 2M C-G7: Tris-saline sucrose

[0331] LNPs were formulated with human TTR guide RNAs and mRNA encoding Cas9 with one NLS and no HA tag. The LNPs had a nitrogen-to-phosphate (N/P) ratio of 4.5 and contained cationic lipid, cholesterol, DSPC, and PEG2k-DMG in a 45:44:9:2 molar ratio. The cationic lipid used in the in vitro and in vivo LNP experiments described herein was (9Z,12Z)-3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl octadeca-9,12-dienoate, also called 3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl (9Z,12Z)-octadeca-9,12-dienoate. The (guide RNA):(Cas9 mRNA) ratio in each was 1:1 by weight. Additional LNP formulation details are provided in Table 8.

TABLE 8

Human TTR LNP Formulations.					
Human TTR Guide RNA	RNA (mg/mL)	Encapsulation	Z-avg (nm)	PDI	Number Ave (nm)
1	0.46	96%	95.22	0.053	77.51
2	0.61	97%	94.91	0.016	76.77

[0332] Mice were weighed prior to injection, and LNPs (containing Cas9 mRNA plus a human TTR gRNA) were prepared to dose at 2 mg/kg by diluting in Tris-saline sucrose so that delivery volume was 200 μ l per mouse. Delivery was intravenous through tail vein injection. After 8-14 days, mice were euthanized, and blood serum was harvested along

with liver tissues. The tissues were processed into DNA lysates that were then analyzed by next-generation sequencing (NGS).

[0333] Serum chemistry analysis for the liver enzymes ALT, AST, triglycerides, total cholesterol, HDL, LDL, non-esterified fatty acids (NEFA), and albumin showed no statistical difference between various treatment groups. See FIGS. 11A-11H. Similar results were observed for a human TTR guide RNA targeting exon 3 (data not shown).

[0334] NGS showed significant editing in liver for human TTR gRNA 1 (average 42.8%) and human TTR gRNA 2 (average 36.0%). See FIG. 12. Editing efficiency refers to the total number of insertions or deletions observed over the total number of sequences read in the PCR reaction from a pool of lysed cells. Minimal or no statistically significant levels of gene editing were observed in other tissues (data not shown).

[0335] Next, human TTR guide RNA 1 was tested in vivo in F2 generation, homozygous humanized TTR mice from clone number 7576B-F10. Animals were weighed pre-dose for dosing calculations and then monitored 24 hours post-dose for welfare. The animals were dosed intravenously at 1 mg/kg, 0.3 mg/kg, and 0.1 mg/kg with LNPs formulated with human TTR guide RNA 1 and mRNA encoding Cas9 as described above. Tris-saline sucrose was used as a control. Three mice were tested per group. At necropsy (8 days post-dose), liver and blood (for serum) was collected for analysis. The percent editing of the humanized TTR locus observed in the liver was 50.7% at a dose of 1 mg/kg of the LNP formulated with human TTR guide RNA 1 and mRNA encoding Cas9, 13.0% at a dose of 0.3 mg/kg, and 2.3% at a dose of 0.1 mg/kg, with less than 0.1% editing observed in the control mice. Human TTR levels were then measured in liver lysate and serum from the dosed mice. Livers were lysed in RIPA and protease inhibitors at 100 mg/mL. A human TTR ELISA kit (Aviva Systems Biology; Cat No.: OKIA00081; 1:100 dilution for liver lysates; 1:5000 or 1:10000 dilution for serum) was used to assess human TTR levels. As shown in FIG. 14, a level of more than 1000 ng/mL human TTR was measured in liver lysates from control animals, and these levels were decreased by more than 50% in animals dosed at 1 mg/kg of the LNP formulated with human TTR guide RNA 1 and mRNA encoding Cas9. As shown in FIGS. 15A and 15B, human TTR was measured at levels of 80,000 ng/mL or more in serum from control animals, and human TTR levels were reduced by 66% in animals dosed at 1 mg/kg of the LNP formulated with human TTR guide RNA 1 and mRNA encoding Cas9.

[0336] Next, three different human TTR guide RNAs (human TTR guide RNAs 3, 4, and 5) were tested in vivo in homozygous humanized TTR mice. The LNP formulations contained Cas9 mRNA in a 1:2 ratio by weight to the guide RNA. The LNPs contained a cationic lipid (9Z,12Z)-3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl octadeca-9,12-dienoate, also called 3-((4,4-bis(octyloxy)butanoyloxy)-2-(((3-(diethylamino)propoxy)carbonyloxy)methyl)propyl (9Z,12Z)-octadeca-9,12-dienoate), cholesterol, DSPC, and PEG2k-DMG in a 50:38:9:3 molar ratio, respectively, and had an N:P ratio of 6.

[0337] First, editing at the humanized TTR locus was assessed. Mice were weighed prior to injection, and LNPs (containing Cas9 mRNA plus a human TTR gRNA) were prepared at doses of 1 mg/kg, 0.3 mg/kg, and 0.1 mg/kg (n=5 mice per group). Delivery was intravenous through tail vein injection. As described above, mice were later euthanized, and blood serum was harvested along with liver tissues. The tissues were processed into DNA lysates that were then analyzed by next-generation sequencing (NGS). NGS showed significant editing in liver for each human TTR

gRNA at all tested doses in a dose-dependent manner. See FIG. 19. Liver editing results were determined using primers designed to amplify the region of interest for NGS analysis.

[0338] Second, serum TTR levels were assessed. Mice were weighed pre-dose for dosing calculations. The mice were dosed intravenously at 1 mg/kg, 0.3 mg/kg, and 0.1 mg/kg (n=5 mice per group) with LNPs formulated with human TTR guide RNA 3, 4, or 5 and mRNA encoding Cas9 as described above. The LNP formulations contained Cas9 mRNA in a 1:2 ratio by weight to the guide RNA. The LNPs contained a cationic lipid (9Z,12Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-(((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9,12-dienoate, also called 3-((4,4-bis(octyloxy)butanoyl)oxy)-2-(((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl (9Z,12Z)-octadeca-9,12-dienoate), cholesterol, DSPC, and PEG2k-DMG in a 50:38:9:3 molar ratio, respectively, and had an N:P ratio of 6. Tris-saline sucrose was used as a control. As described above, blood (for serum) was later collected for analysis. Human TTR levels were then measured in serum from the dosed mice. Human serum TTR levels were assessed as described above. As shown in FIG. 20, human TTR levels were significantly reduced in mice dosed with each guide RNA at all doses in a dose-dependent manner.

Example 4. Generation of Mice Comprising a Humanized TTR Locus Encoding a Chimeric Mouse/Human TTR Protein with a Mouse Signal Sequence

[0339] We hypothesized that the mouse signal sequence of TTR may enhance hTTR secretion to more robust levels. Hydrodynamic delivery (HDD) plasmids were constructed containing a cDNA insert for mouse Ttr (mTtr) signal sequence+hTTR (“m/hTTR”). HDD constructs using the pRG977 vector with the cDNA inserts listed in Table 9 were injected via HDD into male C57/BL6 mice, each 59 days old. ELISAs were performed on submandibular blood on day 4 post-HDD. F1H4 plasma and human serum were included in the ELISAs as negative and positive controls, respectively.

TABLE 9

Summary of HDD Experiment.			
HDD Construct	Number Mice	Weight; g (SD)	cDNA insert
Nanoluc	8	21.5 (2.23)	Nanoluc
Control Protein	7	22.8 (1.40)	Control protein
hTTR	8	23.4 (0.74)	hTTR signal sequence + hTTR
m/hTTR	8	22.9 (0.66)	mTTR signal sequence + hTTR

[0340] The results are shown in FIG. 13. HDD into wild type C57/BL6 mice revealed that utilizing the mouse signal sequence of TTR did in fact increase hTTR secretion into plasma when compared to human signal sequence TTR+hTTR (“hTTR”). This demonstrated that C57/BL6 mice can be used to predict TTR constructs that will result in robust hTTR secretion.

[0341] Based on these results, a second humanized TTR allele was generated comprising a deletion of the region of the mouse Ttr locus from the start of the second exon to the stop codon and its replacement with the orthologous part of the human TTR gene but also including the 3' UTR of the human TTR gene. A large targeting vector comprising a 5' homology arm including 36 kb of sequence upstream from the second exon of the mouse Ttr gene and 34.5 kb of the sequence downstream of the mouse Ttr stop codon was generated to replace the approximately 7.3 kb region from the start of the second exon in the mouse Ttr gene to the mouse Ttr stop codon with the approximately 6.1 kb orthologous human TTR sequence from the start of the second exon in the human TTR gene to the end of the last human TTR exon (exon 4, including the human 3' UTR) and a self-deleting neomycin selection cassette (SDC Neo) flanked by loxP sites. See FIG. 4. To generate the humanized allele, CRISPR/Cas9 components targeting the mouse Ttr locus were introduced into F1H4 mouse embryonic stem cells together with the large targeting vector. Loss-of-allele assays, gain-of-allele assays, and retention assays using primers and probes set forth in FIG. 5B and Table 3 were performed to confirm the humanization of the mouse Ttr allele. Versions with the SDC Neo cassette and after excision of the SDC Neo cassette are shown in FIG. 4. F0 mice were then generated using the VELOCIMOUSE® method.

[0342] A comparison of the human and mouse transthyretin precursor protein sequences is shown in FIG. 1A, a comparison of the human and mouse transthyretin coding sequences is shown in FIG. 1B, and a schematic showing the wild type mouse Ttr locus and the final humanized mouse Ttr locus (humanized TTR version 2 with the SDC Neo cassette deleted) is shown in FIG. 2. Sequences for the expected humanized mouse Ttr locus with the SDC Neo cassette and without the SDC Neo cassette are set forth in SEQ ID NOS: 16 and 17, respectively. MAID7655 refers to the humanized TTR locus (keeping mouse signal sequence) with the neomycin selection cassette. MAID7656 refers to the humanized TTR locus (keeping mouse signal sequence) with the neomycin selection cassette removed. The expected transthyretin precursor protein encoded by the humanized mouse Ttr locus (a chimeric mouse/human TTR protein) is set forth in SEQ ID NO: 2.

[0343] A human TTR ELISA kit (Aviva Systems Biology; Cat No.: OKIA00081; 1:2000 dilution) was then used to assess blood plasma human TTR levels in different versions of humanized TTR mice with ages between 1-3 months. The mice included a wild type control mouse and mice with various combinations of wild type, MAID7577, MAID7655, and MAID7656 alleles. MAID7577 refers to the humanized TTR locus with the neomycin selection cassette removed. MAID7655 refers to the humanized TTR locus (keeping mouse signal sequence) with the neomycin selection cassette. MAID7656 refers to the humanized TTR locus (keeping mouse signal sequence) with the neomycin selection cassette removed. The data are summarized in FIG. 16 and Table 10. As shown in FIG. 16, the hTTR^{7577/7577} mice (clone B-F10) had ~55 µg/mL circulating hTTR, which is significant but lower than physiological levels in wild type mice or human serum. Humanized TTR mice with the mouse TTR signal sequence (hTTR^{7655/7655}, hTTR^{7655/7656}, and hTTR^{7656/7656}) did not have increased secreted TTR levels when compared to humanized TTR mice with the human TTR signal sequence (hTTR^{7577/7577}).

TABLE 10

Plasma TTR Levels.			
Strain	Description	hTTR, µg/mL (SD)	mTTR, µg/mL (SD)
FIH4	Wild type control mouse	N.D.	920 (79.7)
V1.0 hTTR ^{7577/7577}	Humanized TTR locus, cassette deleted	54.41 (14.36)	N.D.
V2.0 hTTR ^{7655/7655}	Humanized TTR locus with mouse TTR signal sequence, cassette deleted	37.42 (2.461)	N.D.
V2.0 hTTR ^{7656/7656}	Humanized TTR locus with mouse TTR signal sequence, cassette deleted	34.88 (n.a.)	N.D.
V2.0 hTTR ^{7655/7656}	Humanized TTR locus with mouse TTR signal sequence, cassette deleted in one allele but present in other	33.86 (2.827)	N.D.
V2.0 hTTR ^{7656/WT}	Heterozygous humanized TTR locus with mouse TTR signal sequence, cassette deleted	18.36 (1.233)	57.50 (4.264)
Human serum	Human serum control	234.5 (n.a.)	N.D.

[0344] A human TTR ELISA kit (Aviva Systems Biology; Cat No.: OKIA00081; 1:2000 dilution) was then used to assess blood plasma human TTR levels in different versions of humanized TTR mice with ages between 2-3 months in another experiment. The mice included a wild type control mouse (labeled FIH4) and mice homozygous for the MAID7577 or MAID7656 alleles. MAID7577 refers to the humanized TTR locus with the neomycin selection cassette removed. MAID7656 refers to the humanized TTR locus (keeping mouse signal sequence) with the neomycin selection cassette removed. The data are summarized in FIG. 18 and Table 11. As shown in FIG. 18, the hTTR^{7577/7577} mice (hTTR v1) had ~55 µg/mL circulating hTTR, which is significant but lower than physiological levels in wild type

mice or human serum. Humanized TTR mice with the mouse TTR signal sequence (hTTR^{7656/7656}; hTTRv2) had increased secreted TTR levels when compared to humanized TTR mice with the human TTR signal sequence (hTTR^{7577/7577}).

TABLE 11

Plasma hTTR Levels.	
TTR Strain	hTTR, µg/mL (SD)
hTTR v2	88.45 (1.465)
FIH4	Not detectable

SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 94

<210> SEQ ID NO 1
<211> LENGTH: 147
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

Met Ala Ser His Arg Leu Leu Leu Leu Cys Leu Ala Gly Leu Val Phe
1          5          10          15

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

Met Val Lys Val Leu Asp Ala Val Arg Gly Ser Pro Ala Ile Asn Val
35         40         45

Ala Val His Val Phe Arg Lys Ala Ala Asp Asp Thr Trp Glu Pro Phe
50         55         60

Ala Ser Gly Lys Thr Ser Glu Ser Gly Glu Leu His Gly Leu Thr Thr
65         70         75         80

Glu Glu Glu Phe Val Glu Gly Ile Tyr Lys Val Glu Ile Asp Thr Lys
85         90         95

Ser Tyr Trp Lys Ala Leu Gly Ile Ser Pro Phe His Glu His Ala Glu
100        105        110

Val Val Phe Thr Ala Asn Asp Ser Gly Pro Arg Arg Tyr Thr Ile Ala
115        120        125

Ala Leu Leu Ser Pro Tyr Ser Tyr Ser Thr Thr Ala Val Val Thr Asn
    
```

-continued

130	135	140	
Pro Lys Glu 145			
<210> SEQ ID NO 2 <211> LENGTH: 147 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic <400> SEQUENCE: 2			
Met Ala Ser Leu Arg Leu Phe Leu Leu Cys Leu Ala Gly Leu Val Phe 1	5	10	15
Val Ser Glu Ala Gly Pro Ala Gly Thr Gly Glu Ser Lys Cys Pro Leu 20	25	30	
Met Val Lys Val Leu Asp Ala Val Arg Gly Ser Pro Ala Ile Asn Val 35	40	45	
Ala Val His Val Phe Arg Lys Ala Ala Asp Asp Thr Trp Glu Pro Phe 50	55	60	
Ala Ser Gly Lys Thr Ser Glu Ser Gly Glu Leu His Gly Leu Thr Thr 65	70	75	80
Glu Glu Glu Phe Val Glu Gly Ile Tyr Lys Val Glu Ile Asp Thr Lys 85	90	95	
Ser Tyr Trp Lys Ala Leu Gly Ile Ser Pro Phe His Glu His Ala Glu 100	105	110	
Val Val Phe Thr Ala Asn Asp Ser Gly Pro Arg Arg Tyr Thr Ile Ala 115	120	125	
Ala Leu Leu Ser Pro Tyr Ser Tyr Ser Thr Thr Ala Val Val Thr Asn 130	135	140	
Pro Lys Glu 145			
<210> SEQ ID NO 3 <211> LENGTH: 7258 <212> TYPE: DNA <213> ORGANISM: Homo sapiens <400> SEQUENCE: 3			
gttgactaag tcaataatca gaatcagcag gtttgcagtc agattggcag ggataagcag			60
cctagctcag gagaagtgag tataaaagcc ccaggctggg agcagccatc acagaagtcc			120
actcattctt ggcaggatgg cttctcatcg tctgctcctc ctctgccttg ctggactggt			180
at ttgtgtct gaggctggcc ctacggtgag tg tttctgtg acatcccatt cctacattta			240
agattcacgc taaatgaagt agaagtgact ccttccagct ttgccaacca gcttttatta			300
ctagggcaag ggtaccagc atctat tttt aatataatta attcaaactt caaaaagaat			360
gaagttccac tgagcttact gagctgggac ttgaactctg agcattctac ctcatgtctt			420
tggtgcatta gg tttgta atctggtacc tctgtttcct cagatagatg atagaaataa			480
agatatgata ttaaggaagc tgttaatact gaattttcag aaaagtatcc ctccataaaa			540
tgtat tttggg ggacaaaactg caggagatta tattctggcc ctatagttat tcaaaacgta			600
tttattgatt aatctt taaa aggccttagtg aacaatattc tagtcagata tctaattctt			660
aaatcctcta gaagaattaa ctaatactat aaaatgggtc tggatgtagt tctgacatta			720

-continued

ttttataaca	actggtaaga	gggagtgact	atagcaacaa	ctaaaatgat	ctcaggaaaa	780
cctgtttggc	cctatgtatg	gtacattaca	tcttttcagt	aattccactc	aatggagac	840
ttttaacaaa	gcaactgttc	tcaggggacc	tattttctcc	cttaaaattc	attatacaca	900
tccttggttg	atagcagtg	gtctggaggc	agaaaccatt	cttgctttgg	aaacaattac	960
gtctgtgtta	tactgagtag	ggaagctcat	taattgtcga	cacttacggt	cctgataatg	1020
ggatcagtg	gtaattcttg	tttcgctcca	gatttcta	accacaaaga	ataaatcctt	1080
tcactctgat	caattttggt	aacttctcac	gtgtctctc	tacaccagg	gcaccggtga	1140
atccaagtgt	cctctgatgg	tcaaagtct	agatgctgtc	cgaggcagtc	ctgccatcaa	1200
tgtggccgtg	catgtgttca	gaaaggctgc	tgatgacacc	tgggagccat	ttgcctctgg	1260
gtaagttgcc	aaagaacct	cccacaggac	ttggttttat	cttcccgttt	gcccctcact	1320
tggtagagag	aggctcacat	catctgctaa	agaatttaca	agtagattga	aaaacgtagg	1380
cagaggcaaa	gtatgcctc	tgaaggatgc	cctctttttg	ttttgcttag	ctaggaagtg	1440
accaggaacc	tgagcatcat	ttaggggag	acagtagaga	aaagaaggaa	tcagaactcc	1500
tctcctctag	ctgtggttg	caacccttt	gggtcacaga	acactttatg	taggtgatga	1560
aaagtaaaaa	ttctatgcc	agaaaaatg	cacagataca	cacacatata	aatcatata	1620
tgtgatttta	ggagtctcac	agattccctg	gtgtccctgg	gtaaacccaa	agctaagtgt	1680
ccttgcttta	gaattttagg	aaaaggtata	atgtgtatta	accattaac	aaaaggaaag	1740
gaattcagaa	atattattaa	ccaggcatct	gtctgtagtt	aatatggatc	accctaaacc	1800
caaggctttt	gcctaatgaa	cactttgggg	cacctactgt	gtgcaaggct	gggggctgtc	1860
aagctcagtt	aaaaaaaaa	agatagaaga	gatggatcca	tgaggcaaag	tacagcccca	1920
ggctaatccc	acgatcccc	gacttcatgt	ccaagagtg	cttctcacct	tcattagcca	1980
gttcacaatt	ttcatggagt	ttttctacct	gcactagcaa	aaacttcaag	gaaaatacat	2040
attaataaat	ctaagcaaag	tgaccagaag	acagagcaat	caggagaccc	tttgcattcca	2100
gcagaagagg	aactgctaag	tatttacctc	tccacagaga	agaatttctg	ttgggtttta	2160
attgaacccc	aagaaccaca	tgattcttca	accattatgt	ggaagatcat	tttcttaggt	2220
ctggttttaa	ctggcttttt	atttgggaat	tcatttatgt	ttatataaaa	tgccaagcat	2280
aacatgaaaa	gtggttacag	gactatttca	aggagagac	agaatggaca	ccaaaaatat	2340
tccaatgttc	ttgtgaatct	tttccttgca	ccaggacaaa	aaaaaaaaaga	agtgaaaaga	2400
agaaaggagg	aggggcataa	tcagagtcag	taaagacaac	tgctattttt	atctatcgta	2460
gctgttgag	tcaaatggga	agcaatttcc	aacattcaac	tatggagctg	gtacttacct	2520
ggaaatagaa	gttgcttagt	gtttgttgct	ggcaaagagt	tatcagagag	gttaaatata	2580
taaaagggaa	aagagtcaga	tacaggttct	tcttcctact	ttaggttttc	cactgtgtgt	2640
gcaaatgata	ctccctgggtg	gtgtgcagat	gcctcaaagc	tatcctcaca	ccacaagggg	2700
gaggagcgag	atcctgctgt	cctggagaag	tcagaggtta	gaacagctgt	ggccacttgc	2760
atccaatcat	caatcttgaa	tcacagggac	tctttcttaa	gtaaacatta	tacctggccg	2820
ggcacggtgg	ctcacgctg	taatcccagc	actttgggat	gccaaagtgg	gcatatcctc	2880
tgaggtcagg	agttcaagac	cagcctggcc	aacatggcaa	aactccgtct	ttatgaaaaa	2940
tacaaaaatt	agccagcag	ggtggcaggc	gcctgtaatc	ccagctaatt	gggaggtgta	3000

-continued

ggctggagaa tcccttgaat ctaggaggca gaggttgacg tgagctgaga tegtgccatt 3060
gcactccagc ctgggtgaca agagtaaaac tctgtctcaa aaaaaaaaa ttatacctac 3120
attctcttct tatcagagaa aaaatctac agtgagcttt tcaaaaagtt tttacaaact 3180
ttttgccatt taatttcagt taggagtttt cctacttct gacttagttg aggggaaatg 3240
ttcataacat gtttataaca tgtttatgtg tgtagttgg tgggggtgta ttactttgcc 3300
atgccatttg tttcctccat gcgtaactta atccagactt tcacacctta taggaaaacc 3360
agtgagctcg gagagctgca tgggctcaca actgaggagg aattttaga agggatatac 3420
aaagtggaaa tagacaccaa atcttactgg aaggcacttg gcatctcccc attccatgag 3480
catgcagagg tgagtataca gaccttogag ggttgttttg gttttggttt ttgcttttgg 3540
cattccagga aatgcacagt tttactcagt gtaccacaga aatgtcctaa ggaaggtgat 3600
gaatgaccaa aggttccctt tcctattata caagaaaaa ttcacaacac tctgagaagc 3660
aaatttcttt ttgactttga tgaatacca cttagtaaca tgacttgaac ttacatgaaa 3720
ctactcatag tctattcatt ccactttata tgaatattga tgatctgct gttgaaataa 3780
tagtttatga ggcagccctc cagacccac gtagagtgtg tgtaacaaga gatgcaccat 3840
tttatttctc gaaaaccctt aacattcttc attccaaaac acatctggct tctcggaggt 3900
ctggacaagt gattcttggc aacacatacc tatagagaca ataaaatcaa agtaataatg 3960
gcaacacaat agataacatt taccaagcat acaccatgtg gcagacacaa ttataagtg 4020
tttccatatt taacctactt aatcctcagg aataagccac tgaggctcagt cctattatta 4080
tccccatctt atagatgaag aaaatgagge accaggaagt caaataactt gtcaaaggte 4140
acaagactag gaaatacaca agtagaaatg tttacaatta aggccagggc tgggtttgcc 4200
ctcagttctg ctatgcctcg cattatgcc caggaaaactt tttccctgt gaaagccaag 4260
cttaaaaaa gaaaagccac atttgtaacg tgctctgttc ccctgcctat ggtgaggatc 4320
ttcaaacagt tatacatgga cccagtcccc ctgccttctc cttaatttct taagtcat 4380
gaaacagatg gctgtcatgg aaatagaatc cagacatgtt ggctcagagt aaagatcaac 4440
taattccatc aaaaatagct cggcatgaaa gggaactatt ctctggctta gtcatggatg 4500
agactttcaa ttgctataaa gtggttcctt tattagacaa tgttaccagg gaaacaacag 4560
gggtttgttt gacttctggg gcccacaagt caacaagaga gccccatcta ccaaggagca 4620
tgccctgac taccctcag ccagcagcaa gacatggacc ccagtcaggg caggagcagg 4680
gttccggcgg cccccagcac aagacattgc ccctagagtc tcagccccta ccctcgagta 4740
atagatctgc ctacctgaga ctgttgtttg cccaagagct gggctctcagc ctgatgggaa 4800
ccatataaaa aggttctactg acatactgcc cacatgttgt tctctttcat tagatcttag 4860
cttcttgtc tgctcttcat tcttgagta ttcattcaac aaacattaaa aaaaaaaaa 4920
agcattctat gtgtggaaca ctctgctaga tgctgtggat ttagaaatga aaatacatcc 4980
cgacccttg aatggaaggg aaaggactga agtaagacag attaagcagg accgtcagcc 5040
cagcttgaag cccagataaa tacgggagac aagagagagc gagtagtgag agatgagtc 5100
caatgcctca ctttgggtgac ggggtcgtgg tgggettcat gcagcttctt ctgataaatg 5160
cctcctcag aactggtcaa ctctacctg gccagtgacc caggtggtca tagtagattt 5220
accaagggaa aatggaaact tttattagga gctcttaggc ctcttcaact catggat 5280

-continued

```

tttttccttt ttttttgaga tggagttttg cctgtcacc caggctggaa tgcagtggtg 5340
caatctcagc tcaactgcaac ctccgcctcc caggttcaag caattctcct gcctcagcct 5400
cccgagtagc tgggactaca ggtgtgogcc accacaccag gctaattttt gtattttttg 5460
taaagacagg ttttcaccac gttggccagg ctggtctgaa ctccagacct caggtgattc 5520
acctgtctca gcctcccaaa gtgctgggat tacaggtgtg agccaccgtg cccggctact 5580
tcatggattt ttgattacag attatgcctc ttacaatttt taagaagaat caagtgggct 5640
gaaggtcaat gtcaccataa gacaaaagac atttttatta gttgattcta ggggaattggc 5700
cttaagggga gccctttctt cctaagagat tcttaggtga ttctcacttc ctcttgcccc 5760
agtattatth ttgtttttgg tatggctcac tcagatcctt ttttctcct atccctaagt 5820
aatccgggtt tctttttccc atatttagaa caaaatgtat ttatgcagag tgtgtccaaa 5880
cctcaaccce aggctgtat acaaaataaa tcaaatataa cacatcttta ctgtcttcta 5940
cctctttctt gacctcaata tatcccaact tgcctcactc tgagaacca ggctgtccca 6000
gcacctgagt cgcagatatt ctactgattt gacagaactg tgtgactatc tggaacagca 6060
ttttgatcca caatttgccc agttacaaag cttaaagag ctctagtgca tgcatatata 6120
tttcaaaatt ccaccatgat ctccacact ctgtattgta aatagagccc tgtaatgctt 6180
ttacttegta tttcattgct tgtatacat aaaaataac ttttctctt catgttagaa 6240
aatgcaaaga ataggaggtt gggggaatct ctgggcttgg agacaggaga ctgctctcc 6300
tactatggtt ccatacagaat gtagactggg acaatacaat aatcaagtc tggtttgctc 6360
atctgtaaat tgggaagaat gttccagct ccagaatgct aaatctctaa gtctgtggtt 6420
ggcagccact attgcagcag ctcttcaatg actcaatgca gttttgcatt ctccctacct 6480
tttttttcta aaaccaataa aatagataca gcccttaggc tttctgggat tcccttagt 6540
caagctaggg tcatcctgac tttcggcgtg aatttgcaa acaagacctg actctgtact 6600
cctgtcttaa ggactgtgca tggttccaaa ggcttagctt gccagcatat ttgagctttt 6660
tccttctggt caaactgttc caaaatataa aagaataaaa ttaattaagt tggcactgga 6720
cttccgggtg tcagtcattg gtgtcatctg tcacgttttt cgggctctgg tggaaatgga 6780
tctgtctgtc ttctctcata ggtggtattc acagccaacg actccggccc ccgocgetac 6840
accattgccc cctgctgag ccctactcc tattecca caagctgtct caccaatccc 6900
aaggaaatgag ggacttctcc tccagtggac ctgaaggacg agggatggga tttcatgtaa 6960
ccaagagtat tccattttta ctaaagcagt gttttcacct catatgctat gttagaagtc 7020
caggcagaga caataaaaca ttctgtgaa aggcactttt cattccactt taacttgatt 7080
ttttaaatcc ccttattgtc ccttccaaaa aaaagagaat caaaatttta caaagaatca 7140
aaggaaatct agaaagtatc tgggcagAAC gctaggagag atccaaattt ccattgtctt 7200
gcaagcaaag cacgtattaa atatgatctg cagccattaa aaagacacat tctgtaaa 7258

```

```

<210> SEQ ID NO 4
<211> LENGTH: 938
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 4

```

```

gttgactaag tcaataatca gaatcagcag gtttgcagtc agattggcag ggataagcag 60

```

-continued

cctagctcag gagaagtgag tataaaagcc ccaggctggg agcagccatc acagaagtcc	120
actcattctt ggcaggatgg cttctcatcg tctgctctc ctctgccttg ctggactggt	180
atthgtgtct gaggtgccc ctacgggcac cggatgaatcc aagtgtcctc tgatgggtcaa	240
agttctagat gctgtccgag gcagtcctgc catcaatgtg gccgtgcatg tgttcagaaa	300
ggctgctgat gacacctggg agccatttgc ctctgggaaa accagtgagt ctggagagct	360
gcatgggctc acaactgagg aggaatttgt agaagggata tacaagtgg aaatagacac	420
caaactctac tggaaaggcac ttggcatctc cccattccat gagcatgcag aggtggtatt	480
cacagccaac gactccggcc cccgccgcta caccattgcc gccctgctga gccctactc	540
ctattccacc acggctgtcg tcaccaatcc caaggaatga gggacttctc ctccagtgga	600
cctgaaggac gagggatggg atthcatgta accaagagta ttccattttt actaaagcag	660
tgttttcacc tcatatgcta tgttagaagt ccaggcagag acaataaac attcctgtga	720
aaggcacttt tcattccact ttaacttgat tttttaaatt cccttattgt cccttccaaa	780
aaaaagagaa tcaaatTTTT acaagaatc aaaggaatc tagaaagtat ctgggcagaa	840
cgctaggaga gatccaaatt tccattgtct tgcaagcaaa gcacgtatta aatatgatct	900
gcagccatta aaaagacaca tctgttaaaa aaaaaaaaa	938

<210> SEQ ID NO 5

<211> LENGTH: 9077

<212> TYPE: DNA

<213> ORGANISM: Mus musculus

<400> SEQUENCE: 5

ctaactctcc taggcaaggt tcatatttgt gtaggttact tattctcctt ttgttgacta	60
agtcaataat cagaatcagc aggtttggag tcagcttggc agggatcagc agcctgggtt	120
ggaaggaggg ggtataaaag ccccttcacc aggagaagcc gtcacacaga tccacaagct	180
cctgacagga tggcttcctc tcgactcttc ctcccttgcc tcgctggact ggtatttgtg	240
tctgaagctg gcccccggtt gagtgatcct gtgagcagtc cagacatggc agttagacct	300
tagataaaga agaagtgcct tcttcagat gtgagaacta gactactcag actctatatt	360
taccattaga ctccaaagag aagagctgga gtgcctctgg ctcttccttc tattgcttta	420
gcgcatggg tctgtagtgc tcagtccttg gtgctcctag ataataaaga tatgagatta	480
acatagaaat aaagatataa aagggctgga tgtatagttt agtgggtccag tgtatgctta	540
gtatgtgaaa agccttctgt tcaacctcta gcaatagaaa aacaagatat attctcgggtg	600
gggctgttaa tattgaatc tcataaaatc tttaatatat ttagtatgccc tattatgttg	660
ttatatTTTA gttcttttagc taatcaaaat gcattattga tctttctttg tctttttttg	720
gccaacactc tattccagtc tttgaaaaag tcccttaaaa gagttaatca gtataattaa	780
atgagtcagg aagtatgtga gggttatttt acaaccagag ggaattacta tagcaacagc	840
tgattagaat gatctcaaga aaaagcccat tctgtctttt tgcaccatgc acctttcagt	900
ggctccatc agatggagag gcaaacagag caatggctct cagagggcct attttccctt	960
tgaacattca ttatccatat cctggtgca cagcagtgca tctgggggca gaaactgttc	1020
ttgctttgga aacaatgctg tctatgtcat actggataaa gaagctcatt aattgtcaac	1080
acttatgtta tcataatggg atcagcatgt acttttggtt ttgttcaga gtctatcacc	1140

-continued

ggaaagaaca agccggttta ctctgacca ttctactgac atttctcttg tctcctctgt 1200
gcccaggggt ctggagaate caaatgtcct ctgatggcca aagtcctgga tgctgtccga 1260
ggcagccctg ctgtagacgt ggctgtaaaa gtgttcaaaa agacctctga gggatcctgg 1320
gagccctttg cctctgggta agcttgtaga aagcccacca tgggaccggg tccaggttcc 1380
catttgctct tattcgtggt agattcagac acacacaact taccagctag agggctcaga 1440
gagagggctc aggggcaag ggcacgtatt gctcttgtaa gagacacagg ttttaattcct 1500
agcaccagaa tggcagctca taaccatctg aaactcacag tcttaggaga tctgggtatc 1560
tgacattctc ttctaccac catgtgtgtg gtgcacaaat tcacatgcag gcatcaaatc 1620
ttataaacia caacaaaaa ccaacaaacc tggtagcaaa agaagattag aagggttaaac 1680
atatgagccg agagcttttg ttttgtttg ttttgtttg ttttgttac atttcaaatg 1740
ttatcccctt tctcggctcc cctcccaaaa cctctaccc cattctctcc tccccttctt 1800
ctatgagggt gttccccacc aaccactcc caccttctcg ctctogaatt cccctatact 1860
gggacatcaa gccttcacag aatcaagggc ctctcctccc attgatgcc gacaatgtca 1920
tctctgcta cctatgtggc tggagccatg ggtccctca tgtatcctcc ttggttggtg 1980
gtttagtctc tgggaggtct gggggatctg gttgattgat attattgttc tctcatgag 2040
attgcaaac ccttcagctc cttcggctct ttaactctc cactggggac cccgagctca 2100
gtccaatggt tggctgtgag catccaccag cagaggcctt ttttttttt ttaacaaag 2160
ctgctttatt atgttgctta gagcatgacc aggaaccaga gcacagtcca agactgaagg 2220
gaggaaaagg gggggagtca ataaccocac tgtttcatag tggtttgcaa cctttttata 2280
tcacagccca ctttaggcaa ataatgaaa ttatagtctc cagggacaga gaagatggtg 2340
caggaaatga agtgcctgct cagaaaatgg gggcttgaat gtgagttccc agactctgtg 2400
taagatgccc agcatcgaag tgcagtctta taacaccagc ctggaggtag aagcttagaa 2460
acaggggtac cctgaagtgt cttgttcacc agtgtccctg aatgggtagg tgcagtgtg 2520
gtgagagacc ctgtctcaaa aatcaaggtg taggataatt gaaaatacct agcttgagc 2580
ttagatcatg caaatgtgta cacacactca cacacaccac acacacaaaa aaatgcagag 2640
acagagagat acagagagac agagagatac agagacagag acagagagaa aaggagaaag 2700
taaaaaacia ataatttaa gacccatgac cacaagagg ctcaaagaca agcactata 2760
aaaccataca catgtaattt taggagtttt cagattccct ggtaccctg ggtgatgac 2820
aagctttgaa tcccagctct aaaatcttac gaagaacgtg tctgtgtgtg ctaatttatt 2880
gatgagagga aaggaattga caaagtgcc ttcgggagct tctgcatta cccagactca 2940
gggttttttt aaatgtacac tcagaacaga gtagctctgt gcaagggtag caaccacgaa 3000
gcttaataag aaacatatcg tgagagatct gcaaggcaaa tctaggggct gaccaatctc 3060
acagtcaccc actagcatgt caacacaact tcccactgt gctagccact tagcaatttt 3120
gtgtgttct gttttgtttt tgtttttaac aaagcaattt caaagagatt tctaattcat 3180
ctaaacaaac aaacaaaag gaaaacagca aagacgcct gagcacttag cagagcagct 3240
atgcagttat gactcctggg tggagacttt atateaggct tcaactgaat acctagaacc 3300
tactagtct cttcatcaat ccttggaag gtcattttct tttggtgctg ttttgagttt 3360
ctatttgta atgtctctat aattatacac gtgttgagca cagcatgcaa agtgattagg 3420

-continued

ggaatctagt tggagtggaa tggataccca aatattcaga ctttcttggt actcttcttt 3480
cttgtacca catcaaaaa aaaaaaatg gagatgagac atggtcagag tcaactaaac 3540
cagctgtac ttttaattac gtggggagca gtttctaaca ttgccattat tgaactgatg 3600
ctgctgggt ggaatggaa atcacttagt atttcttggt ggcaagaat tactgaatgg 3660
attaaatttc caaagggaga agtcagttac aagcttttc tttgtttatt aggetttctg 3720
ctatgataaa ttacactact tccagaagt acccttaggc catgggacac tggactatca 3780
ctctgctgtc acaagagatt acagagttag tcaaggcagc ttgtgacacc ttcagggact 3840
gtcataaact tccagcaagt cattaatcct gaatgcaata ctgtgtgtgt gtgtctatgt 3900
gtgtttgtat gtctgtgtgt gtcttatgtc tgtgtctctg tgtgtgtgtg tgtttgtgtg 3960
tgtgtgtgta tgtatgctg tgtgtgtctt atgtctgtgt ttgtgtgtct gtgtgtgtct 4020
tatgtctgtg tttgtatgtc tgtgtgtgtc tgtgtgtgtc ttatgtctgt gtctctgtgt 4080
gtgtgtgtgt gtatgtatgt atgtatgtat gtatgtgtat gtgtttgcat ctctctgtgt 4140
gtctcgctt atatatgtt gtatgtgtt atgtgtctgc ctttgtgctg tgttggggat 4200
tgaatccagg ggaatacaaa tgttaagaaa gaacgttacc actaagcttc acctgtaggc 4260
cttaaagctt ttctttcttt taaaaattgt aattaattca ttttcagtca ggatctccac 4320
acctcgtccc tgetctctc gaactcacta tttaaacaca atcgccctca aacctgcagc 4380
aacctcccg cctctacct gcgagcacta gaataataac aggtgacccc acacgcctag 4440
attaagacct ttaaggtaaa cattttacta tattttagtc tcataagaca agatgtctaca 4500
ataaagctgt acataaagtt cctctgaatt tcttctatt ttaactcaaa cataaggatt 4560
tcctccttt tgattcaggt aacagaaaa atacacaggt acatacatgt acacacatga 4620
acacacacgc atcacaacca catatgcgca cgcttgtgtg atctateatt taccatgcc 4680
ctgaactctt ctttccccat aaattcctct ggacttgtgt gccctccagg aagaccgctg 4740
agtctggaga gctgcacggg ctcaccacag atgagaagtt tgtagaagga gtgtacagag 4800
tagaactgga caccaaatcg tactggaaga cacttggcat ttccccgttc catgaattcg 4860
cggatgtaag tggacacacc aagttgtttg gattttgttt ttagtctcag gaaattccct 4920
tcgctcttgc tgtacgatgg gcatgagtgg aaagtagatt ccacagccag aatccacagt 4980
gctgggaaag caagccttct gaattttct aaaactcatt tagcaacatg gcctgaacct 5040
gttcacactg cttatggtca gctaaactata tttatgtaa tattcatttc tctgttgagg 5100
aaatgttagt atttgctttt gaggaacct ccagatacca tggagggcat gtcatagtca 5160
aagagagggc tcctatgggt atttctctaa attctggcat ttctttatt ccaaagcaca 5220
tctagtgtcc ccagaagttt gggtagacaa ttcttgcaa cacagagaat tacaacatgt 5280
tcaaaacca acagcttaat atctaaatca tcaagcaaac atcacatggc aaagggattt 5340
ctgaatcaaa actgtttcat ccttatgatc aacctatgga ggtctagcct cgacttacac 5400
ccattttacc aataagctaa gagaagctaa gttcctcacc aaggacacaa ggctagcatg 5460
tgtgagcaag tgacagagtt gccctctatg ttggttagtg tgccttagcc agtgtctcag 5520
taagaaatgg agctaaatca aaaccaagg ccaacagcca aaggcacatg agtaaccttt 5580
gcttggcact gggctcagtt tccctggctc ctctcagtc tcagttcaca gaggcagctg 5640
tcatgcaaat agaatccaag cttgttggtc agacctggag ataacaaat ccatcaaaaa 5700

-continued

tagctcctca tgtgacctag tttgctgtct gttgctatga tacacaccat gaccgaaaag 5760
caaccctggg gagagaaggg tttatttcat cttacagctt acagttcacc atggaggaaa 5820
gccagggtggg aacctggaag tggaaattga agcagagacc agaaaggaat gctgtttact 5880
ggctggctta gctccttttc ttatacagct taggtctatg tgcccagggg atgggtactgc 5940
cgagcatagg ctgagccgc ctacatcaac cattagtcaa aaaaaggtcc atagacttgc 6000
ctacaggcca atctcatgga ggcaatacc cagtgagggg tccctcttcg caggttactc 6060
tagtttgtgt caagttgaca aaacctaac acaaagcaca aacaggggtct gcccttgtgg 6120
cttagccatg gatgacactc tcagatgatg gtgttaccag acaaaccaga ggggctcacc 6180
aagagtctgc cacctaccaa ggtagtactc tactcctcac tgggcaccaa caccatatt 6240
agctgggcca gtacaggacc cttgctgttt cctgcatgaa ttgtccatag accctgggtc 6300
tcagcctgoc gggagtaoct gtaagtagtc gcctcaaaaca cattattcct gttggaagac 6360
ttgtctgatt ctctttttaga actcaatcaa caaacgtttt tattttgttt tggctttttg 6420
gagacaagat ctctcatagg ccagcctgac ttgaatgtag ctgaggatga cctgtgctgc 6480
taatcttctc gcctcttctc cccaagtggg aggataatag gcataagaca ccacagcagt 6540
tttactccat accagggtc tgaaccaga ctttaaacac tctatcaact gattcacatt 6600
cccccccat cattcaacaa acatttgaag aataaaaacc ttctgccttg agcactctgc 6660
taaatacagc ctttgagtgc ggagtatttc ctcacaacca gggccaaga tgacccatc 6720
atacatacca cggaaaatta ggagatggtt ttaggtctct ttgcttgggg taatttttat 6780
gtgtgtgtgt acacagccct gtgcgtgtgt gtgtgtgtgt gtgtgtgtgt gtacaggcac 6840
acacgtgtat gcctgtagag gctacataaa aaccttaggt gtcattctca ggcactctgt 6900
tcacccttc acacagccg aacacacaaa atttgaggca ttagcctgga gctcaccagt 6960
taggctagac tgacttgcca gcagaccca ggctgtctcc atctcccag ctctgggatt 7020
acaaactcta tcataccaga catttttata catattctga gcataaaatt catgtcttca 7080
ggctaacaag tcaagagctt aaatgactga gctctcttac gtggtggatt ttttttaaa 7140
ctacataata tctttttttt ttttttact tctggggaag aaacaaatga gcctgagtga 7200
caatgagaca gaaaagaaat tttgaggagt gtgtgtgtct gtgtgtgtgg tggcacatgc 7260
ctctcatcta atgctagagg ctacagtaga atgctcctga attagtggcc agccaaggcc 7320
aagggttagg gttgtaactc agtggcagag ggcttgcta gcattcgcag gatttgatcc 7380
atagcgtat aaataaat aaataaatac aacagtctaa gatgattctc ctttccatt 7440
atctggatgt tattttgtg ttagttttac tctgtcatcc aatcattgtt tgcctatat 7500
ttggacattt aaaaaaaaa tttattocaa gtgtgttcaa agctgtatcc aaaaactgtc 7560
cacaaatga gtccaatgac atacatctc tatattacca tctgttccag atttgctga 7620
ctcccgccac ctgggctggt gctgcacca tgtctcagat agtctagtga tttgagaagt 7680
gactagtaat tgcaaatcc agactttgtc cagaaaactc tatgagctcc aaaaacttca 7740
ttacatttc tgccagccac aaaccgcttg tgttgggag agaaccctgt gatgtcttc 7800
cacagcatct cagcctggtt tcttccctta aaatattcat cttttcacat tagaacatgc 7860
aaagggacag tgggagcga acccctggac tgggacgcac gaagccttc tttctggta 7920
ggctctcact gtagaaactt agcccggtt cagcatgcag tctgtggag aatggtcct 7980

-continued

```

gccaacattc caggctctgga agttttagt ggagttgttg ataaccactg ttcgccacag 8040
gtcttttggt tgtgggtgtc agtgtttcta ctctcctgac ttttatctga acccaagaaa 8100
gggaacaata gccttcaagc tctctgtgac tctgatctga ccagggccac ccacactgca 8160
gaaggaaact tgcaagaga gacctgcaat tctctaagag ctccacacag ctccaaagac 8220
ttaggcagca tttttaatc taattattcg tcccccaacc ccaccccaga ggacagttag 8280
acaataaaag gaagattacc agcttagcat cctgtgaaca ctttgtctgc agctcctacc 8340
tctgggctct gttagaacta gctgtctctc ctctctccta ggtggtttcc acagccaacg 8400
actctggcca tcgccactac accatcgcag cctctgctcag ccatactcc tacagcacca 8460
cggctgctgt cagcaacccc cagaattgag agactcagcc caggaggacc aggatcttgc 8520
caaagcagta gcattcccatt tgtacaaaa cagtgttctt gctctataaa ccgtgttagc 8580
agctcaggaa gatgccgtga agcattctta ttaaaccacc tgctatttca ttcaactgt 8640
gtttcttttt tatttcctca tttttctccc ctgctcctaa aacccaaaat cttctaaaga 8700
attctagaag gtatgctgac aaacttttta aagaagaaa atactttttg actcatggtt 8760
taaaggcatc ctttccatct tggggaggtc atgggtgctc ctggcaactt gcttgaggaa 8820
gataggctcag aaagcagagt ggaccaaccg ttcaatgttt tacaagcaaa acatacacta 8880
agcatggtct gtacttatta aaagcacaca atctgaaggg ctgtagatgc acagtagtgt 8940
tttcccagag catgttcaaa agccctgggt tcaatcacia tactgaaaag taggcaaaa 9000
aacattctga aaatgaaata tttgggtttt tttttataac ctttagtgac taaataaaga 9060
caaatctaag agactaa 9077
    
```

```

<210> SEQ ID NO 6
<211> LENGTH: 147
<212> TYPE: PRT
<213> ORGANISM: Mus musculus
    
```

<400> SEQUENCE: 6

```

Met Ala Ser Leu Arg Leu Phe Leu Leu Cys Leu Ala Gly Leu Val Phe
1           5           10          15
Val Ser Glu Ala Gly Pro Ala Gly Ala Gly Glu Ser Lys Cys Pro Leu
20          25          30
Met Val Lys Val Leu Asp Ala Val Arg Gly Ser Pro Ala Val Asp Val
35          40          45
Ala Val Lys Val Phe Lys Lys Thr Ser Glu Gly Ser Trp Glu Pro Phe
50          55          60
Ala Ser Gly Lys Thr Ala Glu Ser Gly Glu Leu His Gly Leu Thr Thr
65          70          75          80
Asp Glu Lys Phe Val Glu Gly Val Tyr Arg Val Glu Leu Asp Thr Lys
85          90          95
Ser Tyr Trp Lys Thr Leu Gly Ile Ser Pro Phe His Glu Phe Ala Asp
100         105         110
Val Val Phe Thr Ala Asn Asp Ser Gly His Arg His Tyr Thr Ile Ala
115        120        125
Ala Leu Leu Ser Pro Tyr Ser Tyr Ser Thr Thr Ala Val Val Ser Asn
130        135        140
Pro Gln Asn
145
    
```

-continued

<210> SEQ ID NO 7
<211> LENGTH: 1237
<212> TYPE: DNA
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 7

ctaactctccc taggcaaggt tcatatttgt gtaggttact tattctcctt ttgttgacta 60
agtcaataat cagaatcagc aggtttggag tcagcttggc agggatcagc agcctgggtt 120
ggaaggaggg ggtataaaag ccccttcacc aggagaagcc gtcacacaga tccacaagct 180
cctgacagga tggcttcocct tcgactcttc ctccctttgcc tcgctggact ggtatttgtg 240
tctgaagctg gccccgcggg tgctggagaa tccaaatgtc ctctgatggt caaagtcctg 300
gatgctgtcc gaggcagccc tgctgtagac gtggctgtaa aagtgttcaa aaagacctct 360
gagggatcct gggagccctt tgccctctggg aagaccgagg agtctggaga gctgcacggg 420
ctcaccacag atgagaagtt tgtagaagga gtgtacagag tagaactgga caccaaatcg 480
tactggaaga cacttggcat tccccgctc catgaattcg cggatgtggt tttcacagcc 540
aacgactctg gccatcgcca ctacaccatc gcagccctgc tcagcccata ctctacagc 600
accacggctg tcgtcagcaa cccccagaat tgagagactc agcccaggag gaccaggatc 660
ttgccaaaag agtagcatcc catttgtacc aaaacagtgt tcttgcctta taaacctgtt 720
tagcagctca ggaagatgcc gtgaagcatt cttattaaac cacctgctat ttcattcaaa 780
ctgtgtttct tttttatttc ctcatTTTTt tccccgctc ctaaaaccca aaatcttcta 840
aagaattcta gaaggtatgc gatcaaacct tttaaagaaa gaaaatactt tttgactcat 900
ggtttaaagg catcctttcc atcttgggga ggtcatgggt gctcctggca acttgcttga 960
ggaagatagg tcagaaagca gagtggacca accgttcaat gttttacaag caaaacatac 1020
actaagcatg gtctgtagct attaaaagca cacaatctga agggctgtag atgcacagta 1080
gtgttttccc agagcatggt caaaagccct gggttcaatc acaatactga aaagtaggcc 1140
aaaaaacatt ctgaaaatga aatatttggg ttttttttta taacctttag tgactaaata 1200
aagacaaatc taagagacta aaaaaaaaaa aaaaaaa 1237

<210> SEQ ID NO 8
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 8

guuggaacca uucaaaaacag cauagcaagu uaaaaaaagg cuaguccguu aucaacuuga 60
aaaaguggca ccgagucggu gc 82

<210> SEQ ID NO 9
<211> LENGTH: 76
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 9

guuuuagagc uagaaaagc aaguuaaaau aaggcuaguc cguuuucaac uugaaaaagu 60

-continued

ggcaccgagu cggugc 76

<210> SEQ ID NO 10
<211> LENGTH: 86
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 10

guuuuagagc uaugcuggaa acagcauagc aaguuuuuuuu aaggcuaguc cguuuaucaac 60

uugaaaaaagu ggcaccgagu cggugc 86

<210> SEQ ID NO 11
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (2)..(21)
<223> OTHER INFORMATION: n = A, T, C, or G

<400> SEQUENCE: 11

gnnnnnnnnn nnnnnnnnnn ngg 23

<210> SEQ ID NO 12
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(21)
<223> OTHER INFORMATION: n = A, T, C, or G

<400> SEQUENCE: 12

nnnnnnnnnn nnnnnnnnnn ngg 23

<210> SEQ ID NO 13
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (3)..(23)
<223> OTHER INFORMATION: n = A, T, C, or G

<400> SEQUENCE: 13

ggnnnnnnnnn nnnnnnnnnn nnnng 25

<210> SEQ ID NO 14
<211> LENGTH: 12208
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(97)
<223> OTHER INFORMATION: Mouse Sequence

-continued

```

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (98)..(7298)
<223> OTHER INFORMATION: Human Sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (7299)..(12108)
<223> OTHER INFORMATION: SDC Neo Cassette
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (12109)..(12208)
<223> OTHER INFORMATION: Mouse Sequence

<400> SEQUENCE: 14

agcttggcag ggatcagcag cctgggttgg aaggaggggg tataaaagcc ccttcaccag    60
gagaagccgt cacacagatc cacaagctcc tgacaggatg gcttctcadc gtctgtctct    120
cctctgcctt gctggactgg tattttgtgc tgaggctggc cctacgggtga gtgtttctgt    180
gacatcccat tcctacattt aagattcacg ctaaatgaag tagaagtgac tccttccagc    240
tttgccaacc agcttttatt actagggcaa gggtaaccag catctatttt taatataatt    300
aattcaaact tcaaaaagaa tgaagttcca ctgagcttac tgagctggga cttgaactct    360
gagcattcta cctcattgct ttgggtgcatt aggtttgtaa tatctggtag ctctgtttcc    420
tcagatagat gatagaaata aagatatgat attaaggaag ctgttaatac tgaattttca    480
gaaaagtatc cctccataaa atgtatttgg gggacaaaact gcaggagatt atattctggc    540
cctatagtta ttcaaaacgt atttattgat taatctttaa aaggcttagt gaacaatatt    600
ctagtcagat atctaattct taaatcctct agaagaatta actaacta taaaatgggt    660
ctggatgtag ttctgacatt attttataac aactggtaag agggagtgc tatagcaaca    720
actaaaatga tctcaggaaa acctgtttgg ccctatgtat ggtacattac atcttttcag    780
taattccact caaatggaga cttttaacaa agcaactgtt ctcaggggac ctattttctc    840
ccttaaaatt cattatacac atccctgggt gatagcagtg tgtctggagg cagaaacct    900
tcttgctttg gaaacaatta cgtctgtgtt atactgagta ggaagctca ttaattgtcg    960
acacttacgt tcctgataat gggatcagtg tgtaattctt gttctgctcc agatttctaa   1020
taccacaag aataaatcct ttcactctga tcaattttgt taacttctca cgtgtcttct   1080
ctacaccag ggcaccggtg aatccaagtg tcctctgatg gtcaaagttc tagatgctgt   1140
ccgaggcagt cctgccatca atgtggccgt gcatgtgttc agaaaggctg ctgatgacac   1200
ctgggagcca tttgcctctg ggttaagtgc caaagaacct tcccacagga cttggtttta   1260
tcttcccgtt tgcccctcac ttggtagaga gaggctcaca tcatctgcta aagaatttac   1320
aagtagattg aaaaacgtag gcagaggtca agtatgccct ctgaaggatg ccctcttttt   1380
gttttgctta gctaggaagt gaccaggaac ctgagcatca tttaggggca gacagtagag   1440
aaaagaagga atcagaactc ctctctcta gctgtggttt gcaacccttt tgggtcacag   1500
aacctttat gtagggtgat aaaagtaaac attctatgcc cagaaaaaat gcacagatac   1560
acacacatac aaaatcatat atgtgatttt aggagtttca cagattccct ggtgtccctg   1620
ggtaacacca aagctaagtg tccttgtctt agaattttag gaaaaggat aatgtgtatt   1680
aaccattaa caaaaggaaa ggaattcaga aatattatta accaggcacc tgtctgtagt   1740
taatatggat caccctaaac ccaaggcttt tgcctaatga acactttggg gcacctactg   1800
tgtgcaaggc tgggggctgt caagctcagt taaaaaaaaa aagatagaag agatggatcc   1860

```

-continued

atgaggcaaa gtacagcccc aggctaatacc cacgatcacc cgacttcatg tccaagagtg 1920
gcttctcaacc ttcattagcc agttcacaaat tttcatggag tttttctacc tgcactagca 1980
aaaacttcaa ggaataataca tattaataaa tctaagcaaa gtgaccagaa gacagagcaa 2040
tcaggagacc ctttgcaccc agcagaagag gaactgctaa gtattttacat ctccacagag 2100
aagaatttct gttgggtttt aattgaacc ccaagaaccac atgattcttc aaccattatt 2160
gggaagatca ttttcttagg tctgggttta actggctttt tatttgggaa ttcatttatg 2220
tttatataaa atgccaagca taacatgaaa agtggttaca ggactattct aaggagagaga 2280
cagaatggac accaaaaata ttccaatgtt cttgtgaatc ttttcttgc accaggacaa 2340
aaaaaaaaag aagtgaagag aagaaaggag gaggggcata atcagagtca gtaaagacaa 2400
ctgctatttt tatctatcgt agctgttgca gtcaaatggg aagcaatttc caacattcaa 2460
ctatggagct ggtacttaca tggaaataga agttgctag tgtttgtgc tggcaaagag 2520
ttatcagaga ggttaaataat ataaaaggga aaagagtcag atacaggttc ttcttctac 2580
tttaggtttt cactgtgtg tgcaaatgat actccctggg ggtgtgcaga tgcctcaaag 2640
ctatcctcac accacaaggg agaggagcga gatcctgctg tctggagaa gtgcagagt 2700
agaacagctg tggccacttg catccaatca tcaatctga atcacaggga ctctttctta 2760
agtaaacatt atacctggcc gggcacgggt gctcacgctt gtaatcccag cactttggga 2820
tgccaaagtg ggcatatcat ctgaggtcag gagttcaaga ccagcctggc caacatggca 2880
aaactccgtc tttatgaaaa atacaaaaat tagccaggca tgggtggcagg cgctgtaat 2940
cccagctaat tgggaggctg aggctggaga atcccttgaa tctaggaggc agaggttgca 3000
gtgagctgag atcgtgccat tgcactccag cctgggtgac aagagtaaaa ctctgtctca 3060
aaaaaaaaaa attataccta cattctcttc ttatcagaga aaaaaatcta cagttagctt 3120
ttcaaaaagt ttttacaaac tttttgcat ttaatttcag ttaggagttt tccctacttc 3180
tgacttagtt gaggggaaat gttcataaca tgtttataac atgtttatgt gtgttagttg 3240
gtgggggtgt attactttgc catgccattt gtttctcca tgcgtaactt aatccagact 3300
ttcacacctt ataggaaaac cagttagctt ggagagctgc atgggtcac aactgaggag 3360
gaattttag aaggatata caaagtggaa atagacacca aatcttactg gaaggcactt 3420
ggcatctccc cattccatga gcatgcagag gtgagtatac agaccttca gggttgtttt 3480
ggttttggtt tttgcttttg gcattccagg aaatgcacag ttttactcag tgtaccacag 3540
aaatgtccta aggaaggtga tgaatgacca aaggttcctt ttcctattat acaagaaaa 3600
attcacaaca ctctgagaag caaatttctt tttgactttg atgaaaatcc acttagtaac 3660
atgacttgaa cttacatgaa actactcata gtctattcat tccactttat atgaatattg 3720
atgtatctgc tgttgaaata atagtttatg aggcagcctt ccagacccca cgtagagtgt 3780
atgtaacaag agatgcacca ttttatttct cgaaaaccog taacattctt cattccaaaa 3840
cacatctggc ttctcggagg tctggacaag tgattcttgg caacacatac ctatagagac 3900
aataaaatca aagtaataat ggaacacaaa tagataacat ttaccaagca tacacatgt 3960
ggcagacaca attataagt ttttccatat ttaacctact taatcctcag gaataagcca 4020
ctgaggtcag tcctattatt atccccatct tatagatgaa gaaaatgagg caccaggaag 4080
tcaataaact tgtcaaaggt cacaagacta ggaatacac aagtagaagt gtttacaatt 4140

-continued

aaggcccagg ctgggtttgc cctcagttct gctatgcctc gcattatgcc ccaggaaact 4200
ttttcccttg tgaagccaa gcttaaaaa agaaaagcca catttgtaac gtgctctgtt 4260
cccctgccta tgggtgaggat cttcaaacag ttatacatgg acccagtecc cctgccttct 4320
ccttaatttc ttaagtcatt tgaacagat ggctgtcatg gaaatagaat ccagacatgt 4380
tggtcagagt taaagatcaa ctaattccat caaaaatagc tcggcatgaa agggaaactat 4440
tctctggcct agtcatggat gagactttca attgctataa agtgggtcct ttattagaca 4500
atgttaccag ggaacaaca ggggtttgtt tgacttctgg ggcccacaag tcaacaagag 4560
agccccatct accaaggagc atgtccctga ctaccctca gccagcagca agacatggac 4620
cccagtcagg gcaggagcag ggtttcggcg gcgcccagca caagacattg cccctagagt 4680
ctcagccctt accctcagat aatagatctg cctacctgag actgttgttt gcccagagc 4740
tgggtctcag cctgatggga accatataaa aaggttcact gacatactgc ccacatggtg 4800
ttctctttca ttagatetta gcttccttgt ctgctcttca ttcttgcagt attcattcaa 4860
caaacattaa aaaaaaaaa aagcatttca tgtgtggaac actctgctag atgctgtgga 4920
tttagaaatg aaaatacatc ccgacccttg gaatggaagg gaaaggactg aagtaagaca 4980
gattaagcag gaccgtcagc ccagcttgaa gccagataa atacggagaa caagagagag 5040
cgagtagtga gagatgagtc ccaatgcctc actttgggtg cgggtgcgtg gtgggcttca 5100
tgcagcttct tctgataaat gcctccttca gaactggtea actctacctt ggccagtgc 5160
ccaggtggtc atagtagatt taccaaggga aaatggaac ttttattagg agctcttagg 5220
cctcttcact tcattgattt ttttttctt tttttttgag atggagtttt gccctgtcac 5280
ccaggtgga atgcagtggt gcaatctcag ctactgcaa cctccgcctc ccaggttcaa 5340
gcaattctcc tgccctagcc tcccagtag ctgggactac aggtgtgcgc caccacacca 5400
ggctaatttt tgtatttttt gtaaagacag gttttcacca cgttggccag gctggctga 5460
actccagacc tcaggtgatt cacctgtctc agcctcccaa agtgcctggga ttacaggtgt 5520
gagccaccgt gcccggttac ttcattgatt tttgattaca gattatgctt cttacaattt 5580
ttaagaagaa tcaagtgggc tgaaggtaa tgtcaccata agacaaaaga catttttatt 5640
agttgattct agggaattgg ccttaagggg agccctttct tcctaagaga ttcttaggtg 5700
attctcaett cctcttgcct cagtattatt tttgtttttg gtatggctca ctcagatcct 5760
ttttctctcc tatccctaag taatccgggt ttctttttcc catatttaga acaaaatgta 5820
tttatgcaga gtgtgtccaa acctcaacct aaggcctgta taaaaataa atcaaatata 5880
acacatcttt actgtcttct acctctttcc tgacctcaat atatcccaac ttgcctcact 5940
ctgagaacca aggtgttccc agcacctgag tcgcagatat tctactgatt tgacagaact 6000
gtgtgactat ctggaacagc attttgatcc acaatttgcc cagttacaaa gcttaaatga 6060
gctctagtgc atgcataat atttcaaat tccaccatga tcttccacac tctgtattgt 6120
aaatagagcc ctgtaatgct tttacttctg atttcattgc ttgttatata taaaaatata 6180
ctttctctct tcattgtaga aaatgcaaag aataggaggg tgggggaatc tctgggcttg 6240
gagacaggag acttgcttcc ctactatggt tccatcagaa tgtagactgg gacaatacaa 6300
taattcaagt ctggtttgct catctgtaa ttgggaagaa tgtttccagc tccagaatgc 6360
taaatctcta agtctgtggt tggcagccac tattgcagca gctcttcaat gactcaatgc 6420

-continued

agttttgcat tctccctacc ttttttttct aaaaccaata aaatagatac agccttttagg 6480
ctttctggga tttcccttag tcaagctagg gtcatoctga ctttcggcgt gaatttgcaa 6540
aacaagacct gactctgtac tctctgtcta aggactgtgc atggttccaa aggcttagct 6600
tgccagcata tttgagcttt ttcctctgt tcaaactggt ccaaaatata aaagaataaa 6660
attaattaag ttggcactgg acttccggtg gtcagtcag tgtgtcatct gtcacgtttt 6720
tcgggctctg gtggaaatgg atctgtctgt cttctctcat aggtggtatt cacagccaac 6780
gactccggcc cccgccgcta caccattgcc gccctgtga gccctactc ctattccacc 6840
acggctgtcg tcaccaatcc caaggaatga gggacttctc ctccagtga cctgaaggac 6900
gagggatggg atttcatgta accaagagta ttccattttt actaaagcag tgttttcacc 6960
tcatatgcta tgttagaagt ccaggcagag acaataaaac attcctgtga aaggcacttt 7020
tcattccact ttaacttgat tttttaaatt cccttattgt cccttccaaa aaaaagagaa 7080
tcaaaatfff acaagaatc aaaggaatc tagaaagtat ctgggcagaa cgctaggaga 7140
gatccaaatt tccattgtct tgcaagcaaa gcacgtatta aatatgatct gcagccatta 7200
aaaagacaca ttctgtaaat gagagagcct tattttctg taaccttcag caaatagcaa 7260
aagacacatt ccaagggccc acttctttac tgtgggcact cgagataact tcgtataatg 7320
tatgtatac gaagtatat gcatgccagt agcagcacc acgtccact tctgtctagt 7380
aatgtccaac acctccctca gtccaaacac tgctctgcat ccatgtggct ccatttata 7440
cctgaagcac ttgatggggc ctcaatggtt tactagagcc caccocctg caactctgag 7500
acctctgga tttgtctgtc agtgcctcac tggggcgttg gataatttct taaaaggcca 7560
agttccctca gcagcattct ctgagcagtc tgaagatgtg tgcttttcac agttcaaatc 7620
catgtggctg tttcaccac ctgcctggcc ttgggttacc taccaggacc tagcctagaa 7680
gcagggtgtg ggcacttaac acctaaagct agtgactaac tgaacactca agtggatgcc 7740
atctttgtca cttcttgact gtgacacaag caactcctga tgccaaagcc ctgccaccc 7800
ctctcatgcc catatttga catggtacag gtcctcactg gccatggtct gtgaggtcct 7860
ggctctcttt gacttcataa ttctagggg ccactagtat ctataagagg aagaggggtc 7920
tggtcccag gccacagccc acaaaattcc acctgctcac aggttggtg gctogacca 7980
gggtgtgtcc cctgctctga gccagctccc ggccaagcca gcaccatggg aacccccaa 8040
aagaagagga aggtgcgtac cgatttaaat tccaatttac tgaccgtaca ccaaaattg 8100
cctgcattac cggctgatgc aacgagtgat gaggttcgca agaacctgat ggacatgttc 8160
agggatgcc aggcgttttc tgagcatacc tggaaaatgc ttctgtccgt ttgccggtcg 8220
tgggcggcat ggtgcaagtt gaataaccgg aaatggtttc ccgcagaacc tgaagatggt 8280
cgcgattatc ttctatatct tcaggcgcgc ggtctggcag taaaaactat ccagcaacat 8340
ttgggccagc taaacatgct tcatcgtcgg tccgggctgc caccaccaag tgacagcaat 8400
gctgtttcac tggttatgcg gcggatccga aaagaaaacg ttgatgccgg tgaacgtgca 8460
aaacaggtaa atataaaatt ttttaagtgt taatgatgtt aaactactga ttctaattgt 8520
ttgtgtatct taggctctag cgttcgaac cactgatctc gaccaggctc gttcactcat 8580
ggaaaatagc gatcgtgcc aggatatacg taactctgca tttctgggga ttgcttataa 8640
caccctgtta cgtatagccg aaattgccag gatcaggggtt aaagatatct cacgtactga 8700

-continued

cggtgggaga atgttaatcc atattggcag aacgaaaacg ctggttagca ccgcagggtg 8760
agagaaggca cttagcctgg gggtaactaa actggtcgag cgatggattt cegtctctgg 8820
tgtagctgat gatccgaata actacctgtt ttgccgggtc agaaaaaatg gtgttgccgc 8880
gccatctgcc accagccagc tatcaactcg cgccctggaa gggatttttg aagcaactca 8940
tcgattgatt tacggcgcta aggatgactc tggtcagaga tacctggcct ggtctggaca 9000
cagtgcccggt gtcggagccg cgcgagatat ggcccgcgct ggagtttcaa taccgagat 9060
catgcaagct ggtggctgga ccaatgtaaa tattgtcatg aactatatcc gtaacctgga 9120
tagtgaaaca ggggcaatgg tgcgcctgct ggaagatggc gattagcggg ccggccgcta 9180
atcagccata ccacatttgt agaggtttta cttgctttaa aaaacctccc acacctcccc 9240
ctgaacctga aacataaaat gaatgcaatt gttgttgta acttgttat tgcagcttat 9300
aatggttaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt ttttctactg 9360
cattctagtt gtggtttgtc caaactcatc aatgtatctt atcatgtctg gatcccccg 9420
ctagagttta aacactagaa ctagtggatc ccccgggatc atggcctccg cgccgggtt 9480
tggcgctcc cgccggcgcc cccctcctca cggcgagcgc tgccacgtca gacgaagggc 9540
gcagcgagcg tcctgatcct tccgcccgga cgctcaggac agcggcccgc tgctcataag 9600
actcggcctt agaaccccag tatcagcaga aggacatttt aggacgggac ttgggtgact 9660
ctagggcaact ggttttcttt ccagagagcg gaacaggcga ggaaaagtag tcccttctcg 9720
gcgattctgc ggagggatct ccgtggggcg gtgaacgccg atgattatat aaggacgcgc 9780
cgggtgtggc acagctagtt ccgtcgcagc cgggatttgg gtccgggttc ttgtttgtgg 9840
atcgtctgta tcgtcacttg gtgagtagcg ggctgctggg ctggccgggg ctttctgtggc 9900
cgccgggccc ctcggtggga cggaaagcgtg tggagagacc gccaaaggct gtagtctggg 9960
tccgcgagca aggttgccct gaactggggg ttggggggag cgcagcaaaa tggcggctgt 10020
tcccagctct tgaatggaag acgcttctga ggcgggctgt gaggtcgttg aaacaagggt 10080
gggggcatgg tggcgccgaa gaacccaagg tcttgaggcc ttcgctaag cgggaaagct 10140
cttattcggg tgagatgggc tggggcacca tctggggacc ctgacgtgaa gtttgtcact 10200
gactggagaa ctcggtttgt cgtctgttgc gggggcggca gttatggcgg tgcogttggg 10260
cagtgcaccc gtacctttgg gagcgcgcgc cctcgtcgtg tcgtgacgtc acccgttctg 10320
ttggttata atgcagggtg gggccaactg ccggtagggt tgcggtaggc ttttctcctg 10380
cgcaggacgc aggttccggg cctagggtag gctctcctga atcgacaggc gccggacctc 10440
tggtagggg agggataagt gaggcgtcag tttctttggt cggttttatg tacctatctt 10500
cttaagtagc tgaagctccg gttttgaact atgcgctcgg ggttggcgag tgtgtttgt 10560
gaagttttt aggcacctt tgaatgtaa tcattgggt caatatgtaa tttcagtgt 10620
tagactagta aattgtccgc taaattctgg ccgtttttgg cttttttgtt agacgtgttg 10680
acaattaatc atcggcatag tatatcggca tagtataata cgacaagggt aggaactaaa 10740
ccatgggatc ggccattgaa caagatggat tgcacgcagg ttctccggcc gcttgggtgg 10800
agaggctatt cggctatgac tgggcacaac agacaatcgg ctgctctgat gccgctgtg 10860
tccgctgtc agcgcagggg cgcccgggtc tttttgtcaa gaccgacctg tccggtgcc 10920
tgaatgaact gcaggacgag gcagcgggc tatcgtggct ggccacgacg ggcgttctt 10980

-continued

```

gcgcagctgt gctcgacgtt gtcactgaag cgggaaggga ctggctgcta ttgggcgaag 11040
tgccggggca ggatctcctg tcactcacc ttgctcctgc cgagaaagta tccatcatgg 11100
ctgatgcaat gcggcgctg catacgcttg atccggctac ctgcccattc gaccaccaag 11160
cgaaacatcg catcgagcga gcacgtactc ggatggaagc cggctcttgc gatcaggatg 11220
atctggacga agagcatcag gggctcgcgc cagccgaact gttcgcagg ctcaaggcgc 11280
gcatgcccga cggcgatgat ctcgctgga cccatggcga tgctgcttg ccgaatatca 11340
tgggtgaaaa tggcgcttt tctggattca tgaactgtgg cgggctgggt gtggcggacc 11400
gctatcagga catagcgttg gctaccgtg atattgctga agagcttggc ggcgaatggg 11460
ctgaccgctt cctcgctctt tacggtatcg ccgctcccga ttcgcagcgc atcgcttct 11520
atcgcttctt tgacgagttc tctgagggg atccgctgta agtctgcaga aattgatgat 11580
ctattaaaca ataagatgt ccactaaaat ggaagttttt cctgtcatac tttgtaaga 11640
aggggtgagaa cagagtacct acatcttgaa tgggaaggatt ggagctacgg ggggtgggggt 11700
ggggtgggat tagataaatg cctgctctt actgaaggct ctttactatt gctttatgat 11760
aatgtttcat agttggatat cataatttaa acaagcaaaa ccaaattaag ggccagctca 11820
ttcctcccac tcatgatcta tagatctata gatctctcgt gggatcattg tttttctct 11880
gattcccact ttgtggttct aagtactgtg gtttccaaat gtgtcagttt catagcctga 11940
agaacgagat cagcagcctc tgtccacat acacttcatt ctcagtattg ttttgccaag 12000
ttctaattcc atcagacctc gacctgcagc cctagataa cttcgtataa tgtatgctat 12060
acgaagttag gctaggtaac tataacggtc ctaaggtagc gagctagcga gactcagccc 12120
aggaggacca ggatcttgcc aaagcagtag catcccattt gtacaaaac agtggtcttg 12180
ctctataaac cgtgtagca gctcagga 12208

```

```

<210> SEQ ID NO 15
<211> LENGTH: 7476
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(97)
<223> OTHER INFORMATION: Mouse Sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (98)..(7298)
<223> OTHER INFORMATION: Human Sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (7299)..(7376)
<223> OTHER INFORMATION: Cassette LoxP Scar
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (7377)..(7476)
<223> OTHER INFORMATION: Mouse Sequence

```

```

<400> SEQUENCE: 15

```

```

agcttggcag ggatcagcag cctgggttgg aaggaggggg tataaaagcc ccttcaccag 60
gagaagccgt cacacagatc cacaagctcc tgacaggatg gcttctcacc gtctgctcct 120
cctctgcctt gctggactgg tatttgtgtc tgaggctggc cctacgggtga gtgtttctgt 180
gacatcccat tcctacatctt aagattcagc ctaaatgaag tagaagtgac tccttcagc 240

```

-continued

tttgccaacc agcttttatt actagggcaa gggtaeccag catctatfff taatataatt	300
aattcaaaact tcaaaaagaa tgaagttcca ctgagcttac tgagctggga cttgaactct	360
gagcattcta cctcattgct ttgggtgcatt aggtttgtaa tatctggtag ctctgtttcc	420
tcagatagat gatagaaata aagatatgat attaaggaag ctgttaatac tgaatfctca	480
gaaaagtatc cctccataaa atgtatfctg gggacaaact gcaggagatt atattctggc	540
cctatagfca tcaaaaact atfctatgat taactcttaa aaggcttagt gaacaatatt	600
ctagfcaagat atctaatfct taaactcctc agaagaatta actaacta taaaatgggt	660
ctggatgtag fctctgacatt atfctataac aactggtaag agggagfca tatagcaaca	720
actaaaatga fctcaggaaa acctgtfctg cctatgfat ggtacattac atctfctcag	780
taattccact caaatggaga ctfcttaacaa agcaactgct cfcaggggac ctatfctctc	840
ccttaaaatt cattatacac atccctggtt gatagcagfca gctctggagg cagaaacct	900
fctctgctfct gaaacaatta cgtctgctgt atactgagta gggaaagctca ftaattgctc	960
acactfcaag fctctgataat gggatcagfca gtaattctt gtfctcctc agatfctca	1020
faccacaaag aataaatcct fctcactctga fcaatfctgt taactfctca cgtgctfct	1080
ctacaccag ggcaccgggt aatccaagfca fctctctgag gtfcaagfct tagatgctgt	1140
cagaggcag cctgccatca atgtggcct gcatgcttc agaaaggctc ctgatgacac	1200
ctgggagcca fctgcctctg ggtaaagfca caaagaacct fccacagga ctfctgtfct	1260
fctfctcctg fgcctcctc atggtagaga gaggctcaca fctctctgca aagaatfct	1320
aagtagatg aaaaactgag gcagaggfca agtatccct ctgaaggatg cctctfctfct	1380
gtfctgctta gctaggaagt gaccaggaac ctgagcacta fcttagggca gacagtagag	1440
aaaagaagga atcagaactc cctcctctca cctgctgtt gcaacctfct fgggtcagac	1500
aacactfctat gtaggtgag aaaaagtaac atfctatgct cagaaaaaat gcacagatac	1560
acacacatac aaaatcatat atgtgattfct aggagfctca cagatfctc ggtgctcctg	1620
ggtaacacca aagctaagfca fctctgtctt agaatfcttag gaaaagfctat aatgtgatt	1680
aacctcattaa caaaaggaaa ggaatfca aatattata accaggcact gctctgtagt	1740
taatatggat caccacaaac ccaaggctfct gctcaatga acactfctgg gcacctactg	1800
gtgcaaggc fgggggctgt caagctcagt taaaaaaaaa aagatagaag agatggatcc	1860
atgaggcaaa gtacagcccc aggctaatcc cagcatcacc cgaactcact fccaagagfca	1920
gctfctcacc fctcattagc agfctcacaat fctcatggag fctfctctacc gcaactagca	1980
aaaactfctaa ggaatatac atfctataaa fctaaagcaaa gtfaccagaa gacagagca	2040
tcaggagacc ctfctgactc agcagaagag gaactgctaa gtafctacat ctfccacagag	2100
aagaatfctc gtfctggfctfct aatgaaacct caagaacct atgattctc aacctatfct	2160
gggaagatca fctfctctagg fctggtfctta actgctfctfct fctfctggaa fctcattatg	2220
fctatataaa atgccaagca taacatgaaa agfctgttaca ggaactatfct aaggagagag	2280
cagaatggac accaaaaata fctcaatgct ctfctgtaac fctfctcctg accaggacaa	2340
aaaaaaaaag aagfcaaaag aagaaggag gaggggcata atcagagfca gtaagacaa	2400
ctgctatfctfct fctctatcgt agctgtgca gtfcaaatgg aagcaatfct caacattca	2460
ctatggagct ggtactfca tggaaataga agfctgctag gtfctgtgct gggcaagag	2520

-continued

ttatcagaga	ggttaaatat	ataaaagga	aaagagtcag	atacaggttc	ttcttcctac	2580
tttaggtttt	ccactgtgtg	tgcaaatgat	actccctggg	gggtgcaga	tgctcaaag	2640
ctatcctcac	accacaaggg	agaggagcga	gatcctgctg	tcctggagaa	gtgcagagtt	2700
agaacagctg	tggccacttg	catccaatca	tcaatcttga	atcacagga	ctctttctta	2760
agtaaacatt	atacctggcc	gggcacgggtg	gctcacgcct	gtaatcccag	cactttggga	2820
tgccaaagtg	ggcatatcat	ctgaggtcag	gagttcaaga	ccagcctggc	caacatggca	2880
aaactccgtc	tttatgaaaa	atacaaaaat	tagccagga	tggtggcagg	cgctgtaat	2940
cccagcta	tgggagctg	aggctggaga	atcccttgaa	tctaggaggc	agaggttgca	3000
gtgagctgag	atcgtgcat	tgactccag	cctgggtgac	aagagtaaaa	ctctgtctca	3060
aaaaaaaa	attataccta	cattctcttc	ttatcagaga	aaaaatcta	cagtgcgctt	3120
ttcaaaaagt	ttttacaac	tttttgccat	ttaatttcag	ttaggagttt	tccctacttc	3180
tgacttagtt	gaggggaaat	gttcataaca	tgtttataac	atgtttatgt	gtgtagttg	3240
gtgggggtgt	attactttgc	catgccattt	gtttcctcca	tgcgtaactt	aatccagact	3300
ttcacacctt	atagaaaaac	cagtgcgctt	ggagagctgc	atgggctcac	aactgaggag	3360
gaattttag	aagggatata	caaagtggaa	atagacacca	aatcttactg	gaaggcactt	3420
ggcatctccc	cattccatga	gcctgcagag	gtgagtatac	agaccttoga	gggttgtttt	3480
ggttttggtt	tttgcttttg	gcattccagg	aaatgcacag	ttttactcag	tgtaccacag	3540
aatgtccta	aggaagtga	tgaatgacca	aaggttccct	ttcctattat	acaagaaaa	3600
attcacaaca	ctctgagaag	caaatttctt	ttgactttg	atgaaaatcc	acttagtaac	3660
atgacttgaa	cttaccatgaa	actactcata	gtctattcat	tccactttat	atgaatattg	3720
atgtatctgc	tgttgaaata	atagtttatg	aggcagccct	ccagacccca	cgtagagtgt	3780
atgtaacaag	agatgcacca	ttttatttct	cgaaaaccog	taacattctt	cattccaaaa	3840
cacatctggc	ttctcggagg	tctggacaag	tgattcttgg	caacacatac	ctatagagac	3900
aataaaatca	aagtaataat	ggcaacacaa	tagataacat	ttaccaagca	tacaccatgt	3960
ggcagacaca	attataagtg	ttttccatat	ttaacctact	taactctcag	gaataagcca	4020
ctgaggtcag	tcctattatt	atccccatct	tatagatgaa	gaaaatgagg	caccaggaag	4080
tcaataaact	tgtcaaaggt	cacaagacta	ggaaatacac	aagtagaaat	gtttacaatt	4140
aaggcccagg	ctgggtttgc	cctcagttct	gctatgcctc	gcattatgcc	ccaggaaact	4200
ttttcccttg	tgaagccaa	gcttaaaaa	agaaaagcca	catttgtaac	gtgctctgtt	4260
cccctgccta	tgggtgaggat	cttcaaacag	ttatacatgg	accagtcctc	cctgccttct	4320
ccttaatttc	ttaagtcat	tgaacagat	ggctgtcatg	gaaatagaat	ccagacatgt	4380
tggtcagagt	taaagatcaa	ctaattccat	caaaaatagc	tggcatgaa	agggaaactat	4440
tctctggctt	agtcaggtg	gagactttca	attgctataa	agtggttccct	ttattagaca	4500
atgttaccag	ggaacaaca	ggggtttgtt	tgacttctgg	ggcccacaag	tcaacaagag	4560
agcccctct	accaaggagc	atgtccctga	ctaccctca	gccagcagca	agacatggac	4620
cccagtcagg	gcaggagcag	ggtttcggcg	gcgccagca	caagacattg	cccctagagt	4680
ctcagccct	accctcaggt	aatagatctg	cctacctgag	actgttgttt	gcccagagc	4740
tgggtctcag	cctgatggga	accatataaa	aaggttcact	gacatactgc	ccacatgttg	4800

-continued

ttctctttca	ttagatctta	gcttccttgt	ctgctcttca	ttcttgca	attcattca	4860
caaacattaa	aaaaaaaaa	aagcattcta	tgtgtggaac	actctgctag	atgctgtgga	4920
tttagaaatg	aaaatacatc	ccgacccttg	gaatggaagg	gaaaggactg	aagtaagaca	4980
gattaagcag	gaccgtcagc	ccagcttgaa	gccagataa	atacggagaa	caagagagag	5040
cgagtagtga	gagatgagtc	ccaatgctc	actttggtga	cgggtgcgtg	gtgggcttca	5100
tgcagcttct	tctgataaat	gcctccttca	gaactggta	actctacctt	ggccagtgc	5160
ccagtggtc	atagtagatt	taccaaggga	aatggaac	ttttattagg	agctcttagg	5220
cctcttcact	tcattgattt	tttttctt	ttttttgag	atggagtttt	gccctgtcac	5280
ccaggctgga	atgagtggt	gcaatctcag	ctcaactgaa	cctccgctc	ccaggttcaa	5340
gcaattctcc	tgccctagcc	tcccagtag	ctgggactac	aggtgtgcgc	caccacacca	5400
ggctaatttt	tgtatttttt	gtaaagacag	gtttccacca	cgttgccag	gctggtctga	5460
actccagacc	tcagtgatt	cacctgtctc	agcctcccaa	agtgtctgga	ttacaggtgt	5520
gagccaccgt	gcccggctac	tccatggatt	tttgattaca	gattatgctt	cttacaattt	5580
ttaagaagaa	tcaagtgggc	tgaaggtcaa	tgtcaccata	agacaaaaga	catttttatt	5640
agttgattct	aggaattgg	ccttaagggg	agcccttct	tcctaagaga	ttcttaggtg	5700
attctcaact	cctcttgccc	cagtattatt	ttgtttttg	gtatggtca	ctcagatcct	5760
ttttctctcc	tatccctaag	taatccgggt	ttcttttcc	cataattaga	acaaaatgta	5820
tttatgcaga	gtgtgtccaa	acctcaacc	aaggcctgta	tacaaaataa	atcaaatata	5880
acacatcttt	actgtcttct	acctctttcc	tgacctcaat	atatcccaac	ttgctcact	5940
ctgagaacca	aggctgtccc	agcacctgag	tcgcagatat	tctactgatt	tgacagaact	6000
gtgtgactat	ctggaacagc	atttgatcc	acaatttgcc	cagttacaaa	gcttaaatga	6060
gctctagtgc	atgcataat	atttcaaat	tcccactga	tcttccacac	tctgtattgt	6120
aaatagagcc	ctgtaatgct	ttacttctg	atttcattgc	ttgttataca	taaaatata	6180
ctttctctct	tcattgtaga	aaatgcaaag	aataggagg	tgggggaatc	tctgggcttg	6240
gagacaggag	acttgcttcc	ctactatggt	tccatcagaa	tgtagactgg	gacaatacaa	6300
taattcaagt	ctggtttgct	catctgtaa	ttgggaagaa	tgtttccagc	tccagaatgc	6360
taaatctcta	agtctgtggt	tggcagccac	tattgcagca	gctcttcaat	gactcaatgc	6420
agttttgcat	tctccctacc	tttttttct	aaaaccaata	aatagatac	agcctttagg	6480
ctttctggga	ttcccttag	tcaagctagg	gtcatcctga	ctttcggcgt	gaatttgcaa	6540
aacaagacct	gactctgtac	tctctctcta	aggactgtgc	atggttccaa	aggcttagct	6600
tgccagcata	tttgagcttt	ttccttctgt	tcaaacgttt	ccaaaatata	aaagaataaa	6660
attaattaag	ttggcactgg	acttccgggt	gtcagtcctg	tgtgtcatct	gtcacgtttt	6720
tcgggctctg	gtggaaatgg	atctgtctgt	cttctctcat	aggtggtatt	cacagccaac	6780
gactccggcc	cccgcgcta	caccattgcc	gccctgctga	gcccctactc	ctattccacc	6840
acggctgtcg	tcaccaatcc	caaggaatga	gggacttctc	ctccagtgga	cctgaaggac	6900
gagggatggg	atctcatgta	accaagagta	ttccattttt	actaaagcag	tgttttcacc	6960
tcattatgta	tgttagaagt	ccaggcagag	acaataaac	attcctgtga	aaggcacttt	7020
tcattccact	ttaacttgat	tttttaaat	cccttattgt	cccttccaaa	aaaaagagaa	7080

-continued

tcaaaat	ttt acaagaatc aaaggaatc tagaaagtat ctgggcagaa cgctaggaga	7140
gatccaaatt	tccattgtct tgcaagcaaa gcacgtatta aatatgatct gcagccatta	7200
aaaagacaca	ttctgtaaat gagagagcct tattttcctg taaccttcag caaatagcaa	7260
aagacacatt	ccaagggccc acttctttac tgtgggcact cgagataact tcgtataatg	7320
tatgctatac	gaagttatgc taggtaacta taacggctct aaggtagcga gctagcgaga	7380
ctcagcccag	gaggaccagg atcttgccaa agcagtagca tcccattgtg accaaaaacag	7440
tgttcttgct	ctataaacg tgtagcagc tcagga	7476

<210> SEQ ID NO 16
 <211> LENGTH: 11218
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1)..(100)
 <223> OTHER INFORMATION: Mouse Sequence
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (101)..(6308)
 <223> OTHER INFORMATION: Human Sequence
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (6309)..(11118)
 <223> OTHER INFORMATION: SDC Neo Cassette
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (11119)..(11218)
 <223> OTHER INFORMATION: Mouse Sequence

<400> SEQUENCE: 16

atgtactttt	ggttttgttc cagagtctat caccggaag aacaagccgg tttactctga	60
cccatttcac	tgacatttct cttgtctcct ctgtgccag ggcaccggg aatccaagt	120
tcctctgatg	gtcaaagttc tagatgctgt ccgaggcagt cctgccatca atgtggcct	180
gcatgtgttc	agaaaggctg ctgatgacac ctgggagcca tttgcctctg ggtaagttgc	240
caaagaaccc	tcccacagga cttggtttta tcttcccggt tgcccctcac ttggtagaga	300
gaggctcaca	tcactctgta aagaatttac aagtagattg aaaaacgtag gcagaggcca	360
agtatgcct	ctgaaggatg cctcttttt gttttgctta gctaggaagt gaccaggaac	420
ctgagcatca	tttaggggca gacagtagag aaaagaagga atcagaactc ctctcctcta	480
gctgtggttt	gcaacccttt tgggtcacag aacactttat gtaggtgatg aaaagtaaac	540
attctatgcc	cagaaaaaat gcacagatac acacacatac aaaaatcatat atgtgatttt	600
aggagtttca	cagattccct ggtgtccctg ggtaaacacca aagctaagtg tccttgtctt	660
agaattttag	gaaaaggat aatgtgtatt aaccattaa caaaaggaaa ggaattcaga	720
aatattatta	accaggcatc tgtctgtagt taatatggat caccctaaac ccaaggcttt	780
tgcctaata	gacactttgg gcacctactg tgtgcaaggc tgggggctgt caagctcagt	840
taaaaaaaa	aagatagaag agatggatcc atgaggcaaa gtacagcccc aggctaatcc	900
cacgatcacc	cgacttcctg tccaagagtg gcttctcacc ttcattagcc agttcacaat	960
tttcatggag	ttttctacc tgcactagca aaaacttcaa ggaaaataca tattaataaa	1020
tctaagcaaa	gtgaccagaa gacagagcaa tcaggagacc ctttgcattc agcagaagag	1080

-continued

gaactgctaa	gtatttacat	ctccacagag	aagaatttct	gttgggtttt	aattgaacce	1140
caagaaccac	atgattcttc	aaccattatt	gggaagatca	ttttcttagg	tctggtttta	1200
actggctttt	tatttgggaa	ttcatttatg	tttatataaa	atgccaagca	taacatgaaa	1260
agtggttaca	ggactattct	aagggagaga	cagaatggac	acaaaaata	ttccaatgtt	1320
cttgtgaatc	ttttccttgc	accaggacaa	aaaaaaaaag	aagtgaaaag	aagaaaggag	1380
gaggggcata	atcagagtca	gtaaagacaa	ctgctatttt	tatctatcgt	agctgttgca	1440
gtcaaatggg	aagcaatttc	caacattcaa	ctatggagct	ggtacttaca	tggaaataga	1500
agttgcctag	tgtttgttgc	tggcaaagag	ttatcagaga	ggttaaata	ataaaagga	1560
aaagagtcag	atacaggttc	ttcttcttac	tttaggtttt	cactgtgtg	tgcaaatgat	1620
actcctggg	gggtgtgcaga	tgctcaaag	ctatcctcac	accacaagg	agaggagcga	1680
gatcctgctg	tcctggagaa	gtgcagagtt	agaacagctg	tggcacttg	catccaatca	1740
tcaatcttga	atcacagga	ctctttctta	agtaaacatt	atacctggcc	gggcacgggtg	1800
gctcacgcct	gtaatcccag	cactttggga	tgccaaagtg	ggcatatcat	ctgaggtcag	1860
gagttcaaga	ccagcctggc	caacatggca	aaactcctgc	tttatgaaa	atacaaaat	1920
tagccaggca	tggtggcagg	cgctgtaat	cccagctaat	tgggagctg	aggctggaga	1980
atccttgtaa	tctaggaggc	agaggttgca	gtgagctgag	atcgtgccat	tgcactccag	2040
cctgggtgac	aagagtaaaa	ctctgtctca	aaaaaaaaaa	attataccta	cattctcttc	2100
ttatcagaga	aaaaaatcta	cagttagctt	ttcaaaaagt	ttttacaaac	tttttgccat	2160
ttaatttcag	ttaggagttt	tcctacttc	tgacttagtt	gaggggaaat	gttcataaca	2220
tgtttataac	atgtttatgt	gtgttagttg	gtgggggtgt	attactttgc	catgccattt	2280
gtttcctcca	tgcgtaactt	aatccagact	ttcacacctt	ataggaaaac	cagttagtct	2340
ggagagctgc	atgggctcac	aactgaggag	gaattttag	aagggatata	caaagtggaa	2400
atagacacca	aatcttactg	gaaggcactt	ggcatctccc	cattccatga	gcatgcagag	2460
gtgagtatac	agaccttoga	gggttgtttt	ggttttggtt	tttgcttttg	gcattccagg	2520
aaatgcacag	ttttactcag	tgtaccacag	aaatgtccta	aggaaggtga	tgaatgacca	2580
aaggttcctt	ttcctattat	acaagaaaa	attcacaaca	ctctgagaag	caaatttctt	2640
tttgactttg	atgaaaatcc	acttagtaac	atgacttgaa	cttacatgaa	actactcata	2700
gtctattcat	tccactttat	atgaatattg	atgtatctgc	tgttgaaata	atagtttatg	2760
aggcagccct	ccagacccca	cgtagagtgt	atgtaacaag	agatgcacca	ttttatttct	2820
cgaaaaccgg	taacattctt	cattccaaaa	cacatctggc	ttctcggagg	tctggacaag	2880
tgattcttgg	caacacatac	ctatagagac	aataaaatca	aagtaataat	ggcaacacaa	2940
tagataacat	ttaccaagca	tacaccatgt	ggcagacaca	attataagtg	ttttccatat	3000
ttaacctact	taatcctcag	gaataagcca	ctgaggtcag	tcctattatt	atccccatct	3060
tatagatgaa	gaaaatgagg	caccaggaag	tcaataaact	tgtcaaaggt	cacaagacta	3120
ggaaatacac	aagtagaaat	gtttacaatt	aaggcccagg	ctgggtttgc	cctcagttct	3180
gctatgcctc	gcattatgcc	ccaggaaact	ttttccttgc	tgaaagccaa	gcttaaaaaa	3240
agaaaagcca	catttgtaac	gtgctctgtt	cccctgccta	tggtaggat	cttcaaacag	3300
ttatacatgg	accagctccc	cctgccttct	ccttaatttc	ttaagtcatt	tgaacagat	3360

-continued

ggctgtcatg gaaatagaat ccagacatgt tggtcagagt taaagatcaa ctaattccat	3420
caaaaatagc tcggcatgaa agggaactat tctctggctt agtcatggat gagactttca	3480
attgctataa agtggttcct ttattagaca atgttaccag ggaacaaca ggggtttgtt	3540
tgacttctgg ggcccacaag tcaacaagag agccccatct accaaggagc atgtccctga	3600
ctaccctca gccagcagca agacatggac cccagtcagg gcaggagcag ggtttcggcg	3660
gcgccagca caagacattg cccctagagt ctcagcccct accctcagat aatagatctg	3720
cctacctgag actgttgttt gcccaagagc tgggtctcag cctgatggga accatataaa	3780
aaggttcact gacatactgc ccacatgtt ttctctttca ttagatctta gcttcttgt	3840
ctgctcttca ttcttgagc attcattcaa caaacattaa aaaaaaaaa aagcattcta	3900
tgtgtggaac actctgctag atgctgtgga tttagaaatg aaaatacatc ccgacccttg	3960
gaatggaagg gaaaggactg aagtaagaca gattaagcag gaccgtcagc ccagcttgaa	4020
gccagataa atacggagaa caagagagag cgagtagtga gagatgagtc ccaatgcctc	4080
actttgggtg cgggtgctgt gtgggttca tgcagcttct tctgataaat gcctccttca	4140
gaactggtca actctacctt ggccagtgac ccaggtggtc atagtagatt taccaagga	4200
aatggaaac ttttattagg agctcttagg cctcttcact tcatggattt tttttcctt	4260
tttttttgag atggagtttt gccctgtcac ccaggctgga atgcagtggt gcaatctcag	4320
ctcactgcaa cctccgcctc ccaggttcaa gcaattctcc tgcctcagcc tcccgagtag	4380
ctgggactac aggtgtgctc caccacacca ggctaatttt tgtatttttt gtaaagacag	4440
gttttcacca cgttggccag gctggctga actccagacc tcagggtgatt cacctgtctc	4500
agcctcccaa agtgcctggga ttacaggtgt gagccaccgt gcccgctac ttcattgatt	4560
tttgattaca gattatgcct cttacaattt ttaagaagaa tcaagtgggc tgaaggtcaa	4620
tgtcaccata agacaaaaga catttttatt agttgattct agggaaattg ccttaagggg	4680
agccctttct tcctaagaga ttcttaggtg attctcactt cctcttgccc cagtattatt	4740
tttgtttttg gtatggctca ctcagatcct ttttctctcc taccctaaag taatccgggt	4800
ttctttttcc catatttaga acaaaatgta tttatgcaga gtgtgtccaa acctcaacce	4860
aaggcctgta tacaaaataa atcaaatata acacatcttt actgtcttct acctctttcc	4920
tgacctcaat atatcccac tcgctcact ctgagaacca aggctgtccc agcacctgag	4980
tcgcagatat tctactgatt tgacagaact gtgtgactat ctggaacagc attttgatcc	5040
acaatttgcc cagttacaaa gcttaaatga gctctagtgc atgcatatat atttcaaaat	5100
tccaccatga tctccacac tctgtattgt aaatagagcc ctgtaatgct tttacttctg	5160
atttcattgc ttgttatata taaaaatata cttttcttct tcatgttaga aaatgcaaag	5220
aataggaggg tgggggaatc tctgggcttg gagacaggag acttgcttct ctactatggt	5280
tccatcagaa tgtagactgg gacaatacaa taattcaagt ctggtttgct catctgtaaa	5340
ttgggaagaa tgtttccagc tccagaatgc taaatctcta agtctgtggt tggcagccac	5400
tattgcagca gctcttcaat gactcaatgc agttttgcat tctccctacc tttttttct	5460
aaaaccaata aaatagatac agcctttagg cttttctgga tttcccttag tcaagctagg	5520
gtcatcctga ctttcggcgt gaatttgcaa aacaagacct gactctgtac tctgtctcta	5580
aggactgtgc atggttccaa aggccttagct tgccagcata tttgagcttt ttcctctgt	5640

-continued

tcaaaactgtt ccaaaatata aaagaataaa attaattaag ttggcactgg acttccggtg 5700
gtcagtcacg tgtgtcatct gtcacgtttt tcgggctctg gtggaaatgg atctgtctgt 5760
cttctctcat aggtggtatt cacagccaac gactccggcc cccgccceta caccattgcc 5820
gccctgctga gccctactc ctattccacc acggetgtcg tcaccaatcc caaggaatga 5880
gggacttctc ctccagtggc cctgaaggac gagggatggg atttcatgta accaagagta 5940
ttccatTTTT actaaagcag tgttttcacc tcatatgcta tggttagaagt ccaggcagag 6000
acaataaaac attcctgtga aaggcacttt tcattccact ttaacttgat tttttaaatt 6060
cccttattgt cccttccaaa aaaaagagaa tcaaaatttt acaagaatc aaaggaattc 6120
tagaaagtat ctgggcagaa cgtaggaga gatccaaatt tccattgtct tgcaagcaaa 6180
gcacgtatta aatagatct gcagccatta aaaagacaca ttctgtaaat gagagagcct 6240
tattttcctg taacctcag caaatagcaa aagacacatt ccaagggccc acttctttac 6300
tgtgggcact cgagataact tcgtataatg tatgctatac gaagtatat gcatgccagt 6360
agcagcacc cgtccacct tctgtctagt aatgtccaac acctccctca gtccaaacac 6420
tgctctgcat ccattgtgct cccatttata cctgaagcac ttgatggggc ctcaatgttt 6480
tactagagcc cccccctg caactctgag acctctgga tttgtctgtc agtgcctcac 6540
tggggcggtg gataatttct taaaaggta agttccctca gcagcattct ctgagcagtc 6600
tgaagatgtg tgcttttcac agttcaaac catgtggctg tttcaccac ctgcctggcc 6660
ttgggttacc tatcaggacc tagcctagaa gcaggtgtgt ggcactaac acctaagctg 6720
agtactaac tgaacactca agtggatgcc atctttgtca cttcttgact gtgacacaag 6780
caactcctga tgccaaagcc ctgccacc cctctcatgcc catatttga catggtacag 6840
gtcctcactg gccatggtct gtgaggtcct ggtcctcttt gacttcataa ttctagggg 6900
ccactagtat ctataagagg aagagggtgc tggctcccag gccacagccc acaaaattcc 6960
acctgctcac aggttggctg gctcgaccca ggtggtgtcc cctgctctga gccagctccc 7020
ggccaagcca gcaccatggg aaccccccaag aagaagagga aggtgcgtac cgatttfaat 7080
tccaatttac tgaccgtaca ccaaaatttg cctgcattac cggctgatgc aacgagtgat 7140
gaggttcgca agaacctgat ggacatgttc agggatcgcc aggcgttttc tgagcatacc 7200
tggaaaaatgc ttctgtccgt ttgccggctg tgggcggcat ggtgcaagtt gaataaccgg 7260
aaatggtttc ccgcagaacc tgaagatgtt cgcgattatc ttctatatct tcaggcgcgc 7320
ggtctggcag taaaaactat ccagcaacat ttgggccagc taaacatgct tcatcgtcgg 7380
tccgggctgc cacgaccaag tgacagcaat gctgtttcac tggttatgcg gcggatccga 7440
aaagaaaaag ttgatgcccg tgaacgtgca aaacaggtaa atataaaatt ttttaagtga 7500
taatgatgtt aaactactga ttctaattgt ttgtgtattt taggctctag cgttcgaacg 7560
cactgatttc gaccaggttc gttcactcat ggaaaatagc gatcgctgcc aggatatacg 7620
taatctggca tttctgggga ttgcttataa caccctgtta cgtatagccg aaattgccag 7680
gatcagggtt aaagatatct cacgtactga cgggtgggaga atgttaatcc atattggcag 7740
aacgaaaaag ctggttagca ccgcaggtgt agagaaggca cttagcctgg gggtaactaa 7800
actggtcag cgatggatct ccgtctctgg tgtagctgat gatccgaata actacctgtt 7860
ttgccgggtc agaaaaaatg gtgttgccgc gccatctgcc accagccagc tatcaactcg 7920

-continued

cgccctggaa	gggatttttg	aagcaactca	tcgattgatt	tacggcgcta	aggatgactc	7980
tggtcagaga	tacctggcct	ggtctggaca	cagtgccctg	gtcggagccg	cgcgagatat	8040
ggcccgcgct	ggagtttcaa	taccggagat	catgcaagct	ggtggctgga	ccaatgtaaa	8100
tattgtcatg	aactatatcc	gtaacctgga	tagtgaaaca	ggggcaatgg	tgcgcctgct	8160
ggaagatggc	gattaggcgg	ccggccgcta	atcagccata	ccacatttgt	agaggtttta	8220
cttgctttaa	aaaacctccc	acacctcccc	ctgaacctga	aacataaaat	gaatgcaatt	8280
gttgttgta	acttgtttat	tgcagcttat	aatggttaca	aataaagcaa	tagcatcaca	8340
aatttcacaa	ataaagcatt	ttttcactg	cattctagtt	gtggtttgtc	caaactcatc	8400
aatgtatctt	atcatgtctg	gatccccccg	ctagagtta	aacactagaa	ctagtggatc	8460
ccccgggatc	atggcctccg	cgccggggtt	tggcgccctc	cgcgggcgcc	cccctcctca	8520
cggcgagcgc	tgccacgtca	gacgaagggc	gcagcgagcg	tctgatcct	tccgcccgga	8580
cgctcaggac	agcggcccgc	tgctcataag	actcggcctt	agaaccccag	tatcagcaga	8640
aggacatttt	aggacgggac	ttgggtgact	ctagggcact	ggttttcttt	ccagagagcg	8700
gaacaggcga	gaaaaagtag	tcccttctcg	gcgattctgc	ggagggatct	ccgtggggcg	8760
gtgaacgcgc	atgattatat	aaggacgcgc	cgggtgtggc	acagctagtt	ccgtcgcagc	8820
cgggatattg	gtcgcggttc	ttgtttgtgg	atcgtctgta	tcgtcaactg	gtgagtagcg	8880
ggctgctggg	ctggccgggg	ctttcgtggc	cgccgggccc	ctcggtgga	cggaagcgtg	8940
tggagagacc	gccaaaggct	gtagtctggg	tcccgagca	aggttgccct	gaactggggg	9000
ttggggggag	cgcagcaaaa	tggcggctgt	tcccagctct	tgaatggaag	acgcttgtga	9060
ggcgggctgt	gaggtcgttg	aaacaaggtg	gggggcatgg	tggcgggcaa	gaaccaagg	9120
tcttgaggcc	tctcctaagt	cgggaaagct	cttattcggg	tgagatgggc	tggggcacca	9180
tctggggacc	ctgacgtgaa	gtttgtcact	gactggagaa	ctcggtttgt	cgctctgttc	9240
ggggcgggca	gttatggcgg	tgcggtggg	cagtgcaccc	gtacctttgg	gagcgcgcgc	9300
cctcgtcgtg	tcgtgacgtc	acccttctg	ttggcttata	atgcagggtg	gggccacctg	9360
ccggtagggtg	tgcggtaggc	ttttctccgt	cgcaggacgc	agggttcggg	cctagggtag	9420
gctctcctga	atcgacaggc	gccggaacctc	tggtaggggg	agggataagt	gaggcgtcag	9480
tttctttggt	cggttttatg	tacctatctt	cttaagtagc	tgaagctccg	gttttgaact	9540
atgcgctcgg	gggtggcgag	tgtgttttgt	gaagtttttt	aggcaccttt	tgaatgtaa	9600
tcatttgggt	caatatgtaa	ttttcagtgt	tagactagta	aattgtccgc	taaattctgg	9660
ccgtttttgg	cttttttgtt	agacgtgttg	acaattaatc	atcggcatag	tatatcggca	9720
tagtataata	cgacaagggtg	aggaactaaa	ccatgggatc	ggccattgaa	caagatggat	9780
tgcacgcagg	ttctccggcc	gcttgggtgg	agaggctatt	cggctatgac	tgggcacaac	9840
agacaatecg	ctgctctgat	gccgcctgt	tccgctgtc	agcgcagggg	cgccccgttc	9900
tttttgcata	gaccgaacctg	tccggtgcc	tgaatgaact	gcaggacgag	gcagcgcggc	9960
tategtggct	ggccacgacg	ggcgttcctt	gcgcagctgt	gctcgaactt	gtcactgaag	10020
cgggaaggga	ctggctgcta	ttgggcgaag	tgcggggca	ggatctcctg	tcactctacc	10080
ttgctcctgc	cgagaaagta	tccatcatgg	ctgatgcaat	gcgcggtctg	catacgcttg	10140
atccggctac	ctgcccattc	gaccaccaag	cgaaacatcg	catcgaagca	gcacgtactc	10200

-continued

```

ggatggaagc cggctctgtc gatcaggatg atctggacga agagcatcag gggctcgcgc 10260
cagccgaact gttcgcagg ctcaaggcgc gcatgcccgga cggcgatgat ctctcgtga 10320
cccatggcga tgccctgctt ccaaatatca tggtggaaaa tggccgcttt tctggattca 10380
tcgactgtgg ccggctgggt gtggcggacc gctatcagga catagcgttg gctaccctgt 10440
atattgctga agagcttggc ggcgaatggg ctgaccgctt cctcgtgctt tacggtatcg 10500
ccgctcccga ttcgcagcgc atcgccttct atcgccttct tgacgagttc ttctgagggg 10560
atccgctgta agtctgcaga aattgatgat ctattaaaca ataaagatgt ccactaaat 10620
ggaagttttt cctgtcatac tttgttaaga agggtgagaa cagagtacct acattttgaa 10680
tggaaggatt ggagctacgg ggggtgggggt ggggtgggat tagataaatg cctgctcttt 10740
actgaaggct ctttactatt gctttatgat aatgtttcat agttggatat cataatttaa 10800
acaagcaaaa ccaaatatag ggccagctca ttcctcccac tcatgatcta tagatctata 10860
gatctctcgt gggatcattg tttttctctt gattcccact ttgtggttct aagtaactgtg 10920
gtttccaaat gtgtcagttt catagcctga agaacgagat cagcagcctc tgttccacat 10980
acacttcatt ctcagtattg ttttgccaag ttctaattcc atcagacctc gacctgcagc 11040
ccctagataa cttcgtataa tgtatgctat acgaagttat gctaggtaac tataacggtc 11100
ctaaggtagc gagctagcga gactcagccc aggaggacca ggatccttgc aaagcagtag 11160
catcccattt gtacaaaac agtgttcttg ctctataaac cgtgttagca gctcagga 11218

```

```

<210> SEQ ID NO 17
<211> LENGTH: 6486
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(100)
<223> OTHER INFORMATION: Mouse Sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (101)..(6308)
<223> OTHER INFORMATION: Human Sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (6309)..(6386)
<223> OTHER INFORMATION: Cassette LoxP Scar
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (6387)..(6486)
<223> OTHER INFORMATION: Mouse Sequence

```

```

<400> SEQUENCE: 17

```

```

atgtactttt ggttttgttc cagagtctat caccggaaag aacaagccgg tttactctga 60
cccatttcac tgacatttct cttgtctcct ctgtgccag ggcaccgggt aatccaagtg 120
tcctctgatg gtcaaagttc tagatgctgt ccgaggcagt cctgccatca atgtggccgt 180
gcatgtgttc agaaaggctg ctgatgacac ctgggagcca tttgcctctg ggtaagttgc 240
caaagaaccc tcccacagga cttggtttta tcttcccgtt tgcccctcac ttggtagaga 300
gaggctcaca tcactctgcta aagaatttac aagtagattg aaaaacgtag gcagaggtea 360
agtatgcct ctgaaggatg ccctcttttt gttttgctta gctaggaagt gaccaggaac 420
ctgagcatca tttaggggca gacagtagag aaaagaagga atcagaactc ctctcctcta 480

```

-continued

gctgtggttt gcaacccttt tgggtcacag aacactttat gtaggtgatg aaaagtaaac	540
attctatgcc cagaaaaaat gcacagatac acacacatac aaaatcatat atgtgatttt	600
aggagtttca cagattccct ggtgtccctg ggtaacacca aagctaagtg tccttgtcct	660
agaattttag gaaaaggat aatgtgtatt aaccattaa caaaaggaaa ggaattcaga	720
aatattatta accaggcatc tgtctgtagt taatatgat caccctaaac ccaaggcttt	780
tgcctaatga acactttggg gcacctactg tgtgcaaggc tgggggctgt caagctcagt	840
taaaaaaaaa aagatagaag agatggatcc atgaggcaaa gtacagcccc aggctaatec	900
cacgatcacc cgacttcctg tccaagagtg gcttctcacc ttcattagcc agttcacaat	960
tttcatggag tttttctacc tgcactagca aaaacttcaa ggaaaatata tattaataaa	1020
tctaagcaaa gtgaccagaa gacagagcaa tcaggagacc ctttgcctcc agcagaagag	1080
gaactgctaa gtatttaccat ctccacagag aagaatttct gttgggtttt aattgaacct	1140
caagaaccac atgattcttc aaccattatt gggaagatca ttttcttagg tctggtttta	1200
actggctttt tatttgggaa ttcatttatg tttatataaa atgccaagca taacatgaaa	1260
agtggttaca ggactattct aaggagaga cagaatggac accaaaaata ttccaatgtt	1320
cttgtgaatc ttttccttgc accaggacaa aaaaaaaaaa aagtgaagg aagaaaggag	1380
gaggggcata atcagagtca gtaaaagaaa ctgctatttt tatctatcgt agctgttgca	1440
gtcaaatggg aagcaatttc caacattcaa ctatggagct ggtacttaca tggaaataga	1500
agttgcctag tgtttgttgc tggcaagag ttatcagaga ggtaaatat ataaaaggga	1560
aaagagtcag atacaggttc ttcttctac tttaggtttt ccactgtgtg tgcaaatgat	1620
actccctggg ggtgtgcaga tgcctcaaag ctatcctcac accacaaggg agaggagcga	1680
gatcctgctg tcctggagaa gtgcagagt tgaacagctg tggccaactg catccaatca	1740
tcaatcttga atcacagga ctctttctta agtaaacatt atacctggcc gggcacggtg	1800
gctcacgcct gtaatcccag cactttggga tgccaaagtg ggcataatcat ctgaggtcag	1860
gagttcaaga ccagcctggc caacatggca aaactccgtc tttatgaaa atacaaaaat	1920
tagccaggca tgggtggcagg cgctgtaat cccagctaat tgggaggctg aggctggaga	1980
atcccttgaa tctaggaggc agaggttgca gtgagctgag atcgtgccat tgcactccag	2040
cctgggtgac aagagtaaaa ctctgtctca aaaaaaaaaa attataccta cattctcttc	2100
ttatcagaga aaaaaatcta cagttagctt ttcaaaaagt ttttacaac tttttgccat	2160
ttaatttcag ttaggagttt tccctacttc tgacttagtt gaggggaaat gttcataaca	2220
tgtttataac atgtttatgt gtgttagttg gtgggggtgt attactttgc catgccattt	2280
gtttcctcca tgcgtaactt aatccagact ttcacacctt ataggaaaac cagttagctt	2340
ggagagctgc atgggctcac aactgaggag gaattttag aaggatata caaagtggaa	2400
atagacacca aatcttactg gaaggcactt ggcattctcc cattccatga gcatgcagag	2460
gtgagtatac agacctoga gggttgtttt ggttttggtt tttgcttttg gcattccag	2520
aatgcacag ttttactcag tgtaccacag aaatgtccta aggaaggatga tgaatgacca	2580
aaggttccct ttctattat acaagaaaa attcacaaca ctctgagaag caaatttctt	2640
tttgactttg atgaaaatcc acttagtaac atgacttgaa cttacatgaa actactcata	2700
gtctattcat tccactttat atgaatattg atgtatctgc tgttgaaata atagtttatg	2760

-continued

aggcagccct ccagaccoca cgtagagtgt atgtaacaag agatgcacca ttttatttct	2820
cgaaaaccog taacattott cattccaaaa cacatctggc ttctcggagg tctggacaag	2880
tgattcttgg caacacatac ctatagagac aataaaatca aagtaataat ggcaacacaa	2940
tagataacat ttaccaagca tacaccatgt ggcagacaca attataagtg ttttccatat	3000
ttaacctact taatcctcag gaataagcca ctgaggtcag tcctattatt atccccatct	3060
tatagatgaa gaaaatgagg caccaggaag tcaaataact tgtcaaaggt cacaagacta	3120
ggaaatacac aagtagaaat gtttacaatt aaggcccagg ctgggtttgc cctcagttct	3180
gctatgcctc gcattatgcc ccaggaaact ttttccttg tgaaagccaa gcttaaaaaa	3240
agaaaagcca catttgtaac gtgctctgtt ccctgccta tggtagaggat cttcaaacag	3300
ttatacatgg acccagtooc cctgccttct ccttaatttc ttaagtcatt tgaacagat	3360
ggctgtcatg gaaatagaat ccagacatgt tggtcagagt taaagatcaa ctaattccat	3420
caaaaatagc tcggcatgaa agggaaactat tctctggctt agtcatggat gagactttca	3480
attgctataa agtggttcct ttattagaca atggtaccag ggaaacaaca ggggtttgtt	3540
tgacttctgg ggcccacaag tcaacaagag agccccatct accaaggagc atgtccctga	3600
ctaccctca gccagcagca agacatggac ccagtcagg gcaggagcag ggtttcggcg	3660
gogcccagca caagacattg ccctagagt ctcagccct acctcagagt aatagatctg	3720
cctacctgag actgttgttt gcccaagagc tgggtctcag cctgatggga accatataa	3780
aaggttact gacatactgc ccacatgttg ttctcttca ttagatetta gcttccttgt	3840
ctgctcttca ttcttgacgt attcattcaa caaacattaa aaaaaaaaaa aagcatttca	3900
tgtgtggaac actctgctag atgctgtgga tttagaaatg aaaatacatc ccgacccttg	3960
gaatggaagg gaaaggactg aagtaagaca gattaagcag gaccgtcagc ccagcttga	4020
gcccagataa atacggagaa caagagagag cgagtagtga gagatgagtc ccaatgctc	4080
actttggtga cgggtgcgtg gtgggcttca tgcagcttct tctgataaat gcctccttca	4140
gaactggtca actctacctt ggccagtgac ccaggtggtc atagtagatt taccaagga	4200
aaatggaaac ttttattagg agctcttagg cctcttcact tcatggattt ttttttctt	4260
ttttttgag atggagtttt gcctgtcac ccaggctgga atgcagtggt gcaatctcag	4320
ctcactgcaa cctccgctc ccaggttcaa gcaattctcc tgcctcagcc tcccagtag	4380
ctgggactac aggtgtgcgc caccacacca ggctaatttt tgtatttttt gtaaagacag	4440
gttttcacca cgttggccag gctggtctga actccagacc tcaggtgatt cacctgtctc	4500
agcctcccaa agtgcctggga ttacaggtgt gagccaccgt gcccggtac ttcattggatt	4560
tttgattaca gattatgctt cttacaattt ttaagaagaa tcaagtgggc tgaaggtcaa	4620
tgtcaccata agacaaaaga catttttatt agttgattct agggaattgg ccttaagggg	4680
agccctttct tcctaagaga ttcttagtg atctcactt cctctgccc cagtattatt	4740
tttgttttg gtatggctca ctcagatcct ttttctctcc taccctaaag taatccgggt	4800
ttctttttcc catatttaga acaaaatgta tttatgcaga gtgtgtccaa acctcaacct	4860
aaggcctgta tacaaaataa atcaaatata acacatcttt actgtcttct acctctttcc	4920
tgacctcaat atatcccaac ttgcctcact ctgagaacca aggctgtccc agcacctgag	4980
tcgcagatat tctactgatt tgacagaact gtgtgactat ctggaacagc attttgatcc	5040

-continued

```

acaatttgcc cagttacaaa gcttaaatga gctctagtgc atgcatatat atttcaaaat 5100
tccaccatga tcttccacac tctgtattgt aaatagagcc ctgtaatgct tttacttcgt 5160
atttcattgc ttgtatata taaaaatata cttttcttct tcatgttaga aaatgcaaag 5220
aataggaggg tgggggaatc tctgggcttg gagacaggag acttgccttc ctactatggt 5280
tccatcagaa tgtagactgg gacaatacaa taattcaagt ctggtttgcct catctgtaaa 5340
tgggaagaa tgtttccagc tccagaatgc taaatctcta agtctgtggt tggcagccac 5400
tattgcagca gctcttcaat gactcaatgc agttttgcat tctccctacc ttttttttct 5460
aaaaccaata aaatagatac agcctttagg ctttctggga tttcccttag tcaagctagg 5520
gtcatcctga ctttcggcgt gaatttgcaa aacaagacct gactctgtac tcctgctcta 5580
aggactgtgc atggttccaa aggcttagct tggcagcata tttgagcttt ttccttctgt 5640
tcaaactgtt ccaaaatata aaagaataaa attaattaag ttggcactgg acttccgggtg 5700
gtcagtcatg tgtgtcatct gtcacgtttt tccggctctg gtggaaatgg atctgtctgt 5760
cttctctcat aggtggtatt cacagccaac gactccggcc cccgccccta caccattgcc 5820
gccctgctga gccctactc ctattccacc acggctgtcg tcaccaatcc caaggaatga 5880
gggacttctc ctccagtga cctgaaggac gagggatggg atttcatgta accaagagta 5940
ttccattttt actaaagcag tgttttcacc tcatatgcta tgttagaagt ccaggcagag 6000
acaataaaac attcctgtga aaggcacttt tcattccact ttaacttgat tttttaaatt 6060
cccttattgt cccttccaaa aaaagagaa tcaaaatttt acaagaatc aaaggaattc 6120
tagaaagtat ctgggcagaa cgctaggaga gatccaaatt tccattgtct tgcaagcaaa 6180
gcacgtatta aatgatgctc gcagccatta aaaagacaca ttctgtaaat gagagagcct 6240
tattttctcg taacctcag caaatagcaa aagacacatt ccaagggccc acttctttac 6300
tgtgggcact cgagataact tcgtataatg tatgctatac gaagttatgc taggtaacta 6360
taacggctcct aaggtagcga gctagcgaga ctcagcccag gaggaccagg atcttgccaa 6420
agcagtagca tcccatttgt accaaaacag tgttcttgcct ctataaacccg tgttagcagc 6480
tcagga 6486

```

<210> SEQ ID NO 18

<211> LENGTH: 7201

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

```

atggcttctc atcgtctgct cctcctctgc cttgctggac tggatattgt gtctgaggct 60
ggccctacgg tgagtgttct tgtgacatcc cattcctaca ttttaagattc acgctaaatg 120
aagtagaagt gactccttcc agctttgcca accagctttt attactaggg caaggggtacc 180
cagcatctat ttttaatata attaattcaa acttcaaaaa gaatgaagtt cactgagct 240
tactgagctg ggacttgaac tctgagcatt ctacctcatt gctttggtgc attaggtttg 300
taatatctgg tacctctggt tcctcagata gatgatagaa ataaagatat gatattaagg 360
aagctgttaa tactgaatct tcagaaaagt atccctccat aaaatgtatt tgggggacaa 420
actgcaggag attatattct ggccctatag ttattcaaaa cgtattttatt gattaatctt 480
taaaaggctt agtgaacaat attctagtca gatatcfaat tcttaaatcc tctagaagaa 540

```

-continued

ttaactaata ctataaaatg ggtctggatg tagttctgac attattttat aacaactggt 600
aagagggagt gactatagca acaactaaaa tgatctcagg aaaacctggt tggccctatg 660
tatggtacat tacatctttt cagtaattcc actcaaatgg agacttttaa caaagcaact 720
gttctcaggg gacctatfff ctccttataa attcattata cacatccctg gttgatagca 780
gtgtgtctgg aggcagaaac cattcttget ttggaacaa ttacgtctgt gttatactga 840
gtagggaaagc tcattaattg tgcacactta cgttctgat aatgggatca gtgtgtaatt 900
cttgtttcgc tccagatttc taataccaca aagaataaat cctttcactc tgatcaattt 960
tgtaacttc tcacgtgtct tctctacacc cagggcaccg gtgaatccaa gtgtcctctg 1020
atggtcaaaag ttctagatgc tgtccgaggc agtcctgccca tcaatgtggc cgtgcatggt 1080
ttcagaaaag ctgctgatga cacctgggag ccatttgcct ctgggtaagt tgccaaagaa 1140
cctcccaca ggacttggtt ttatcttccc gtttgcccct cacttggtag agagaggctc 1200
acatcatctg ctaagaatt tacaagtaga ttgaaaaacg taggcagagg tcaagtatgc 1260
cctctgaagg atgccctctt tttgttttgc ttagctagga agtgaccagg aacctgagca 1320
tcatttaggg gcagacagta gagaaaagaa ggaatcagaa ctctctcct ctagctgtgg 1380
tttgcaacc ttttgggtca cagaacactt tatgtagggt atgaaaagta aacattctat 1440
gccagaaaa aatgcacaga tacacacaca taaaaatca tatatgtgat tttaggagt 1500
tcacagattc cctggtgtcc ctgggtaaca ccaaagctaa gtgtccttgt cttagaattt 1560
taggaaaagg tataatgtgt attaacccat taacaaaagg aaaggaattc agaaatatta 1620
ttaaccaggc atctgtctgt agttaatatg gatcaoccaa aaccoaaggc ttttgccata 1680
tgaacacttt gggcaccta ctgtgtgcaa ggctgggggc tgtcaagctc agttaaaaa 1740
aaaaagatag aagagatgga tccatgaggc aaagtacagc cccaggctaa tcccacgatc 1800
accgacttc atgtccaaga gtggtcttc accttcatta gccagttcac aattttcatg 1860
gagttttct acctgcacta gcaaaaactt caaggaaaat acatattaat aaatctaagc 1920
aaagtgacca gaagacagag caatcaggag accctttgca tccagcagaa gaggaactgc 1980
taagtattta catctccaca gagaagaatt tctgttgggt ttttaattgaa ccccaagaac 2040
cacatgattc ttcaaccatt attgggaaga tcattttctt aggtctggtt ttaactggct 2100
ttttatttgg gaattcattt atgtttatat aaaatgcca gcataacatg aaaagtgggt 2160
acaggactat tctaagggag agacagaatg gacacaaaa atattccaat gttcttgtga 2220
atctttctc tgcaccagga caaaaaaaaa aagaagtga aagaagaaag gaggaggggc 2280
ataatcagag tcagtaaaga caactgctat ttttatctat cgtagctggt gcagtcaaat 2340
gggaagcaat ttccaacatt caactatgga gctggtactt acatggaat agaagttgcc 2400
tagtgtttgt tgctggcaaa gagttatcag agaggttaaa tatataaaag ggaaaagagt 2460
cagatacagg ttcttcttcc tactttaggt tttccactgt gtgtgcaaat gatactcct 2520
ggtggtgtgc agatgcctca aagctatcct cacaccacaa gggagaggag cgagatcctg 2580
ctgtcctgga gaagtgcaga gttagaacag ctgtggccac ttgcatccaa tcatcaatct 2640
tgaatcacag ggactcttcc ttaagtaaac attatacctg gccgggcacg gtggctcacg 2700
cctgtaatcc cagcactttg ggatgccaaa gtgggcatat catctgaggt caggagtcca 2760
agaccagcct ggccaacatg gcaaaaactc gtccttatga aaaatacaaa aattagccag 2820

-continued

gcatggtggc aggcgcctgt aatcccagct aattgggagg ctgaggctgg agaatccctt	2880
gaatctagga ggcagagggt gcagtgagct gagatcgtgc cattgcactc cagcctgggt	2940
gacaagagta aaactctgtc tcaaaaaaaaa aaaattatac ctacattctc ttcttatcag	3000
agaaaaaaaa ctacagtgag cttttcaaaa agtttttaca aactttttgc catttaattt	3060
cagttaggag tttccctac ttctgactta gttgagggga aatgttcata acatgtttat	3120
aacatgttta tgtgtgttag ttggtggggg tgtattactt tgccatgcca ttgtttcct	3180
ccatgcgtaa cttaatccag actttcacac cttataggaa aaccagtgag tctggagagc	3240
tgcattggct cacaactgag gaggaatttg tagaagggat atacaaagtg gaaatagaca	3300
ccaaatctta ctggaaggca cttggcatct cccattcca tgagcatgca gaggtgagta	3360
tacagacctt cgagggttgt tttggtttg gtttttgett ttggcattcc aggaaatgca	3420
cagttttact cagtgtacca cagaaatgtc ctaaggaagg tgatgaatga ccaaaggttc	3480
cctttcctat tatacaagaa aaaattcaca acaactctgag aagcaaatct ctttttgact	3540
ttgatgaaaa tccacttagt aacatgactt gaacttacct gaaactactc atagtctatt	3600
cattccactt tatatgaata ttgatgtatc tgctgttgaa ataatagttt atgaggcagc	3660
cctccagacc ccacgtagag tgtatgtaac aagagatgca ccattttatt tctcgaaaac	3720
cogtaacatt cttcattoca aaacacatct ggcttctcgg aggtctggac aagtgattct	3780
tggcaacaca tacctataga gacaataaaa tcaaagtaat aatggcaaca caatagataa	3840
catttaccaa gcatacacca tgtggcagac acaattataa gtgttttcca tatttaacct	3900
acttaactct caggaataag ccaactgaggt cagtcctatt attatcccca tcttatagat	3960
gaagaaaatg aggcaccagg aagtcaaata acttgtcaaa ggtcacaaga ctaggaaata	4020
cacaagtaga aatgtttaca attaaggccc aggctggggtt tgccctcagt tctgctatgc	4080
ctgcattat gccccaggaa actttttccc ttgtgaaagc caagcttaaa aaaagaaaag	4140
ccacatttgt aacgtgctct gttccctgc ctatggtgag gatcttcaaa cagttatata	4200
tggaccagct cccctgcct tctccttaat ttcttaagtc atttgaaca gatggctgtc	4260
atggaaatag aatccagaca tgttggtcag agttaaagat caactaatc catcaaaaat	4320
agctcggcat gaaaggaac tattctctgg cttagtcatg gatgagactt tcaattgcta	4380
taaagtgggt cctttattag acaatgttac cagggaaca acaggggttt gtttgacttc	4440
tggggccccc aagtcaaca gagagcccca tctaccaagg agcatgtccc tgactacccc	4500
tcagccagca gcaagacatg gacccagtc agggcaggag cagggtttcg gcggcgccca	4560
gcacaagaca ttgcccctag agtctcagcc cctaccctcg agtaatagat ctgcctacct	4620
gagactgttg tttgcccag agctgggtct cagcctgatg ggaaccatat aaaaaggttc	4680
actgacatac tgcccacatg ttgttctctt tcattagatc ttagcttctt tgtctgctct	4740
tcattcttgc agtattcatt caacaacat taaaaaaaa aaaaagcatt ctatgtgtgg	4800
aacactctgc tagatgctgt ggatttagaa atgaaaatc atccccaccc ttggaatgga	4860
agggaaaagga ctgaagtaag acagattaag caggaccgtc agcccagctt gaagcccaga	4920
taaatcagga gaacaagaga gagcgagtag tgagagatga gtcccattgc ctcactttgg	4980
tgacgggtgc gtggtgggct tcatgcagct tcttctgata aatgcctcct tcagaactgg	5040
tcaactctac cttggccagt gacccaggtg gtcatagtag atttaccag ggaaaatgga	5100

-continued

```

aacttttatt aggagctott aggcctcttc acttcatgga ttttttttc ctttttttt 5160
gagatggagt tttgccctgt caccaggct ggaatgcagt ggtgcaatct cagctcactg 5220
caacctccgc ctcccagggt caagcaatc tctgcctca gcctcccag tagctgggac 5280
tacaggtgtg cgccaccaca ccaggctaatt tttgtattt tttgtaaaga caggttttca 5340
ccacgttggc caggctggc tgaactccag acctcagggtg attcacctgt ctcagcctcc 5400
caaagtgtg ggattacagg tgtgagccac cgtgcccgc tacttcatgg atttttgatt 5460
acagattatg cctcttaca ttttaagaa gaatcaagtg ggctgaaggt caatgtcacc 5520
ataagacaaa agacatttt attagttgat tctagggaa tggccttaag gggagcctt 5580
tcttccaaag agattcttag gtgattctca cttcctctg cccagctatt atttttgtt 5640
ttggtatggc tcaactcagat ctttttttc tctatccct aagtaatccg ggtttctttt 5700
tcccatattt agaacaaaat gtatttatgc agagtgtgtc caaacctcaa cccaaggcct 5760
gtatacaaaa taaatcaaat taaacacatc tttactgtct tctacctct tcttgacctc 5820
aatatatccc aacttgctc actctgagaa ccaaggctgt ccagcacct gagtcgcaga 5880
tattctactg atttgacaga actgtgtgac tatctggaac agcattttga tccacaattt 5940
gccagttac aaagcttaaa tgagctctag tgcattgata tatatttcaa aattccacca 6000
tgatcttcca cactctgtat tgtaaataga gccctgtaat gcttttactt cgtatttcat 6060
tgcttgttat acataaaaat atacttttct tcttcatgtt agaaaatgca aagaatagga 6120
gggtggggga atctctgggc ttggagacag gagacttgcc ttcctactat ggttccatca 6180
gaatgtagac tgggacaata caataattca agtctggtt gctcatctgt aaattgggaa 6240
gaatgtttcc agctccagaa tgctaaatct ctaagtctgt ggttggcagc cactattgca 6300
gcagctcttc aatgactcaa tgcagttttg cattctcctt acctttttt tctaaaacca 6360
ataaaataga tacagccttt aggcctttct ggatttccct tagtcaagct aggggtcatcc 6420
tgactttcgg cgtgaatttg caaaacaaga cctgactctg tactcctgct ctaaggactg 6480
tgcattggtc caaaggctta gcttgccagc atatttgagc tttttcctt tgttcaaact 6540
gttccaaaat ataaaagaat aaaattaatt aagttggcac tggacttccg gtggtcagtc 6600
atgtgtgtca tctgtcaagt ttttcgggct ctggtggaaa tggatctgtc tgtcttctct 6660
cataggtggt attcacagcc aacgactccg gccccgcg ctacaccatt gccgcctgc 6720
tgagccccta ctctattcc accacggctg tegtaccaa tccaaggaa tgagggactt 6780
ctcctccagt ggacctgaag gacgaggat gggatttcat gtaaccaaga gtattccatt 6840
tttactaaag cagtgttttc acctcatatg ctatgttaga agtccaggca gagacaataa 6900
aacattctg tgaaggcacc ttttcattcc actttaactt gattttttaa attcccttat 6960
tgtcccttcc aaaaaaaga gaatcaaat tttacaaaga atcaaaggaa ttctagaaag 7020
tatctgggca gaacgctagg agagatccaa atttccattg tcttgcaagc aaagcacgta 7080
ttaaataatga tctgcagcca ttaaaaagac acattctgta aatgagagag ccttattttc 7140
ctgtaacctt cagcaaatag caaaagacac attccaaggg cccacttctt tactgtgggc 7200
a 7201

```

<210> SEQ ID NO 19

<211> LENGTH: 6208

-continued

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

```

ggcaccgggtg aatccaagtg tcctctgatg gtcaaagttc tagatgctgt ccgaggcagt      60
cctgccatca atgtggccgt gcatgtgttc agaaaggctg ctgatgacac ctgggagcca      120
tttgccctctg ggtaagtgc caaagaacct tcccacagga cttggtttta tcttcccgtt      180
tgcccctcac ttggtagaga gaggctcaca tcatctgcta aagaatttac aagtagattg      240
aaaaacgtag gcagaggta agtatgccct ctgaaggatg ccctcttttt gttttgctta      300
gctaggaagt gaccaggaac ctgagcatca tttaggggca gacagtagag aaaagaagga      360
atcagaactc ctctcctcta gctgtggttt gcaacccttt tgggtcacag aacactttat      420
gtaggtgatg aaaagtaaac attctatgcc cagaaaaaat gcacagatac acacacatac      480
aaaatcatat atgtgatttt aggagtttca cagattccct ggtgtccctg ggtaaacacca      540
aagctaagtg tccttgtcctt agaattttag gaaaaggat aatgtgtatt aaccattaa      600
caaaaggaaa ggaattcaga aatattatta accaggcatc tgtctgtagt taatatggat      660
caccctaaac ccaaggcttt tgcctaata gaactttggg gcacctactg tgtgcaaggc      720
tgggggctgt caagctcagt taaaaaaaa aagatagaag agatggatcc atgaggcaaa      780
gtacagcccc aggtaatcc cacgatcacc cgacttcatg tccaagagtg gcttctcacc      840
ttcattagcc agttcacaat tttcatggag tttttctacc tgcactagca aaaacttcaa      900
ggaaaataca tattaataaa tctaagcaaa gtgaccagaa gacagagcaa tcaggagacc      960
ctttgcatcc agcagaagag gaactgctaa gtatttacct ctccacagag aagaatttct     1020
gttgggtttt aattgaacct caagaacct atgattcttc aaccattatt ggggaagatca     1080
ttttcttagg tctggtttta actggctttt tatttgggaa ttcatttatg tttatataaa     1140
atgccaaagca taacatgaaa agtgggttaca ggactattct aaggggagaga cagaatggac     1200
accaaaaata ttccaatggt cttgtgaatc ttttccttgc accaggacaa aaaaaaaaaag     1260
aagtgaagaag aagaaggag gaggggcata atcagagtca gtaaagacaa ctgctatttt     1320
tatctatcgt agctgttga gtcacaatggg aagcaatttc caacattcaa ctatggagct     1380
ggtaacttaca tggaaataga agttgcctag tgtttgttgc tggcaaagag ttatcagaga     1440
ggttaaatat ataaaaggga aaagagtcag atacaggttc ttcttcctac tttaggtttt     1500
ccactgtgtg tgcaaatgat actccctggt ggtgtgcaga tgctcaaag ctatcctcac     1560
accacaaggg agaggagcga gatcctgctg tcctggagaa gtgcagagtt agaacagctg     1620
tggccacttg catccaatca tcaatcttga atcacagga ctctttctta agtaaacatt     1680
atacctggcc gggcacgggt gctcacgcct gtaatccag cactttggga tgccaaagtg     1740
ggcatatcat ctgaggtcag gagttcaaga ccagcctggc caacatggca aaactccgtc     1800
tttatgaaaa atacaaaaat tagccaggca tgggggcagg cgcctgtaat cccagctaat     1860
tgggaggctg aggctggaga atcccttgaa tctaggaggc agaggttgca gtgagctgag     1920
atcgtgccat tgcactccag cctgggtgac aagagtaaaa ctctgtctca aaaaaaaaaa     1980
attataccta cattctcttc ttatcagaga aaaaaatcta cagtgagctt ttcaaaaagt     2040
ttttacaaac tttttgccat ttaatttcag ttaggagttt tccctacttc tgacttagtt     2100
gaggggaaat gttcataaca tgtttataac atgtttatgt gtgttagttg gtgggggtgt     2160

```

-continued

attactttgc catgccattt gtttctcca tgcgtaactt aatccagact ttcacacctt 2220
ataggaaaac cagtgagtct ggagagctgc atgggctcac aactgaggag gaattttagt 2280
aagggatata caaagtggaa atagacacca aatcttactg gaaggcactt ggcatctccc 2340
cattccatga gcatgcagag gtgagtatac agaccttcca gggttgtttt ggttttggtt 2400
tttgcttttg gcattccagg aaatgcacag ttttactcag tgtaccacag aaatgtccta 2460
aggaaggtga tgaatgacca aaggttocct ttcctattat acaagaaaaa attcacaaca 2520
ctctgagaag caaatttctt tttgactttg atgaaaatcc acttagtaac atgacttgaa 2580
cttacctgaa actactcata gtctattcat tccactttat atgaatattg atgtatctgc 2640
tgttgaaata atagtttatg aggcagocct ccagacccca cgtagagtgt atgtaacaag 2700
agatgcacca ttttatttct cgaaaacccg taacattctt cattccaaaa cacatctggc 2760
ttctcggagg tctggacaag tgattcttgg caacacatac ctatagagac aataaaatca 2820
aagtaataat ggcaacacaa tagataacat ttaccaagca tacaccatgt ggagacaca 2880
attataagtg ttttccatat ttaacctact taatcctcag gaataagcca ctgaggtcag 2940
tcctattatt atccccatct tatagatgaa gaaaatgagg caccaggaag tcaaataact 3000
tgtcaaaggc cacaagacta ggaatatcac aagtagaaat gtttacaatt aaggcccagg 3060
ctgggtttgc cctcagttct gctatgctc gcattatgcc ccaggaaact ttttcccttg 3120
tgaaagccaa gcttaaaaaa agaaaagcca catttgtaac gtgctctggt ccctgctcta 3180
tggtaggat cttcaaacag ttatacatgg acccagtcct cctgccttct ccttaatttc 3240
ttaagtcatt tgaacacagat ggtgtcctg gaaatagaat ccagacatgt tggtcagagt 3300
taaagatcaa ctaattccat caaaaatagc tcggcatgaa agggactat tctctggctt 3360
agtcatggat gagactttca attgctataa agtggttcct ttattagaca atgttaccag 3420
ggaaacaaca ggggtttggt tgacttctgg ggcccacaag tcaacaagag agccccatct 3480
accaaggagc atgtccctga ctaccctca gccagcagca agacatggac cccagtcagg 3540
gcaggagcag ggtttcggcg gcgcccagca caagacattg cccctagagt ctgagccct 3600
accctcgagt aatagatctg cctacctgag actgttgttt gcccagagc tgggtctcag 3660
cctgatggga accatataaa aaggttact gacatactgc ccacatgtt tctctttca 3720
ttagatctta gcttccctgt ctgctcttca ttcttgagc attcattcaa caaacattaa 3780
aaaaaaaaa aagcattcta tgtgtggaac actctgctag atgctgtgga tttagaatg 3840
aaaatacatc ccgacccttg gaatggaagg gaaaggactg aagtaagaca gattaagcag 3900
gaccgtcagc ccagcttgaa gccagataa atacggagaa caagagagag cgagttagta 3960
gagatgagtc ccaatgcctc actttgggtga cgggtgcgtg gtgggcttca tgcagcttct 4020
tctgataaat gcctccttca gaactggta actctacctt ggccagtgc ccaggtggtc 4080
atagtagatt taccaaggga aaatggaac ttttattagg agctcttagg cctcttact 4140
tcatggattt ttttttccct tttttttgag atggagtttt gccctgtcac ccaggctgga 4200
atgcagtggt gcaatctcag ctactgcaa cctccgcctc ccaggttcaa gcaattctcc 4260
tgcctcagcc tcccagtag ctgggactac aggtgtgcgc caccacacca ggctaatttt 4320
tgtattttt gtaaagacag gttttacca cgttggccag gctggctga actccagacc 4380
tcagtgatt cacctgtctc agcctccaa agtgctggga ttacaggtgt gagccacctg 4440

-continued

```

gcccggctac ttcattgatt tttgattaca gattatgcct cttacaattt ttaagaagaa 4500
tcaagtgggc tgaaggtcaa tgtcaccata agacaaaaga catttttatt agttgattct 4560
agggaaattgg ccttaagggg agccctttct tccaaagaga ttcttaggtg attctcactt 4620
cctcttgccc cagtattatt tttgtttttg gtatggctca ctcagatcct tttttctccc 4680
tatccctaag taatccgggt ttctttttcc catatttaga acaaaatgta tttatgcaga 4740
gtgtgtccaa acctcaaccc aaggcctgta taaaaataa atcaaatata acacatcttt 4800
actgtctctt acctctttcc tgacctcaat atateccaac ttgcctcact ctgagaacca 4860
aggctgtccc agcacctgag tcgcagatat tctactgatt tgacagaact gtgtgactat 4920
ctggaacagc attttgatcc acaatttgcc cagttacaaa gcttaaatga gctctagtgc 4980
atgcatatat atttcaaaat tccaccatga tcttccacac tctgtattgt aaatagagcc 5040
ctgtaatgct tttacttctg atttcattgc ttgttataca taaaaatata cttttcttct 5100
tcatgttaga aaatgcaaag aataggaggg tgggggaatc tctgggcttg gagacaggag 5160
acttgccctc ctactatggt tccatcagaa tgtagactgg gacaatacaa taattcaagt 5220
ctggtttgct catctgtaaa ttgggaagaa tgtttccagc tccagaatgc taaatctcta 5280
agtctgtggg tggcagccac tattgcagca gctcttcaat gactcaatgc agttttgcat 5340
tctccctacc ttttttttct aaaaccaata aaatagatac agcctttagg ctttctggga 5400
tttcccttag tcaagctagg gtcacacctg ctttcggcgt gaatttgcaa aacaagacct 5460
gactctgtac tcctgctcta aggactgtgc atggttccaa aggcttagct tgccagcata 5520
tttgagcttt ttctctctgt tcaaaactgt ccaaaatata aaagaataaa attaattaag 5580
ttggcactgg acttccgggt gtcagtcatg tgtgtcatct gtcacgtttt tcgggctctg 5640
gtggaatagg atctgtctgt cttctctcat aggtgggtatt cacagccaac gactccggcc 5700
cccgcgcta caccattgcc gccctgctga gccctactc ctattccacc acggctgtcg 5760
tcaccaatcc caaggaatga gggacttctc ctccagtggg cctgaaggac gagggatggg 5820
atctcatgta accaagagta ttccattttt actaaagcag tgttttcacc tcatatgcta 5880
tgttagaagt ccaggcagag acaataaaac attcctgtga aaggcacttt tcattccact 5940
ttaacttgat tttttaaatt cccttattgt cccttccaaa aaaaagagaa tcaaaatfff 6000
acaaagaatc aaaggaatc tagaaagtat ctgggcagaa cgctaggaga gatccaaatt 6060
tccattgtct tgcaagcaaa gcacgtatta aatatgatct gcagccatta aaaagacaca 6120
ttctgtaaat gagagagcct tattttctctg taaccttcag caaatagcaa aagacacatt 6180
ccaagggccc acttctttac tgtgggca 6208

```

<210> SEQ ID NO 20

<211> LENGTH: 8300

<212> TYPE: DNA

<213> ORGANISM: Mus musculus

<400> SEQUENCE: 20

```

atggcttccc ttgactctt cctcctttgc ctgctggac tggattttgt gtctgaagct 60
ggccccgagg tgagtgatec tgtgagcgat ccagacatgg cagttagacc ttagataaag 120
aagaagtgcc ttcttcagaa tgtgagaact agagtactca gactctatat ttaccattag 180
actccaaaga gaagagctgg agtgcctctg gctcttctct ctattgcttt agcgcattgg 240

```

-continued

gtctgtagtg ctcagtctct ggtgtcctta gataataaag atatgagatt aacatagaaa 300
taaagatata aaagggtctg atgtatagtt tagtggcca gtgtatgctt agtatgtgaa 360
aagccttctg ttcaacctct agcaatagaa aaacaagata tattctcggg ggggctgtta 420
atattgaatt ctcataaaat ctttaataata tttagatgct ctattatggt gttatatttt 480
agttctttag ctaatcaaaa tgcattattg atctttcttt gtcttttttt ggccaacact 540
ctattccagt ctttgaaaaa gtcctttaa agagttaatc agtataatta aatgagtcag 600
gaagtatgtg agggttattt tacaaccaga ggggaattact atagcaacag ctgattagaa 660
tgatctcaag aaaaagccca ttctgtcttt ttgcaccatg cacctttcag tggetccatt 720
cagatggaga ggcaaacaga gcaatggctc tcagagggcc tattttccct ttgaacattc 780
attatccata tccttggctg acagcagctg atctgggggc agaaactgtt cttgctttgg 840
aaacaatgct gtctatgtca tactggataa agaagctcat taattgtcaa cacttatggt 900
atcataatgg gatcagcatg tacttttggg tttgttccag agtctatcac cggaaagaac 960
aagccggttt actctgacct atttactga catttctctt gtctcctctg tgcccagggt 1020
gctggagaat ccaaatgtcc tctgatggct aaagtccctg atgctgtccg aggcagccct 1080
gctgtagacg tggctgtaaa agtgttcaaa aagacctctg agggatcctg ggagcccttt 1140
gcctctgggt aagctttag aaagcccacc atgggaccgg ttccaggttc ccatttgctc 1200
ttattcgtgt tagattcaga cacacacaac ttaccagcta gagggctcag agagagggtc 1260
caggggcaaa gggcacgtat tgcctctgta agagacacag gtttaattcc tagcaccaga 1320
atggcagctc ataaccatct gaaactcaca gtcttaggag atctgggtat ctgacattct 1380
cttctacca ccatgtgtgt ggtgcacaaa ttcacatgca ggcacaaat cttataaaca 1440
acaacaaaaa accaacaaac ctggtagcaa aagaagatta gaagggtaaa catatgagcc 1500
gagagctttt gttttgtttt gttttgtttt gttttgttta catttcaaat gttatccctt 1560
ttctcggctc ccctcccaa accctctacc ccattctctc ctccctctt tctatgaggg 1620
tgttccccac caaccctc ccacctctc gctctcgaat tcccctatac tgggacatca 1680
agccttcaca gaatcaaggg cctctcctcc cattgatgct cgacaatgct atcctctgct 1740
acctatgtgg ctggagccat gggcccttc atgtatcctc cttggttggg ggtttagtct 1800
ctgggaggtc tgggggatct ggttgattga tattattggt cttcctatga gattgcaaac 1860
ccctcagct ccttgggtcc ttttaactct ccactgggga ccccgagctc agtccaatgg 1920
ttggctgtga gcattccacca gcagaggcct tttttttttt ttttaacaaa gctgctttat 1980
tatgttgctt agagcatgac caggaaccag agcacagctc aagactgaag ggaggaaaag 2040
ggggggagtc aataaccoca ctgtttcata gtggtttgca acccttttat atcacagccc 2100
actttaggca aataatgaaa attatagtct ccagggacag agaagatggt gcaggaagtg 2160
aagtgcctgc tcagaaaaat ggggcttgaa tgtgagttcc cagactctgt gtaagatgct 2220
cagcatcgaa gtgcagctt ataacaccag cctggaggta gaagcttaga aacaggggta 2280
ccctgaagtt gcttggctcag cagtgtccct gaatgggtag gtgcattgtt ggtgagagac 2340
cctgtctcaa aaatcaaggt gtaggataat tgaaaatacc tagctttgag cttagatcat 2400
gcaaatgtgt acacacactc acacacacca cacacacaaa aaaatgcaga gacagagaga 2460
tacagagaga cagagagata cagagacaga gacagagaga aaaggagaaa gtaaaaaaca 2520

-continued

aataatTTaa agaccatgg ccacaaagag gctcaaagac aagcacgtat aaaaccatac 2580
acatgtaatt ttaggagttt tcagattccc tggtaaccgt gggatgatgca caagctttga 2640
atcccagtct taaaatctta cgaagaacgt gttcgtgtgt gctaatttat tgatgagagg 2700
aaaggaattg acaaaagtgc cttccggagc ttcctgcatt acccagactc aggggttttt 2760
taaatgtaca ctcagaacag agtagctctg tgcaagggtg gcaaccacga agcttaataa 2820
gaaacatac gtgagagatc tgcaaggcaa atctaggggc tgaccaatct cacagtcacc 2880
cactagcatg tcaacacaac tccccactg tgctagccac ttagcaattt tgtgtgttc 2940
tgttttgttt ttgtttttaa caaagcaatt tcaaagagat ttctaattca tctaaacaaa 3000
caaaccacaa ggaaaacagc aaagacgcc tgagcactta gcagagcagc tatgcagtta 3060
tgactcctgg gtggagactt tatatcaggc ttcaactgaa tacctagaac ctactagtgc 3120
tcttcatcaa tccttgggaa ggtcattttc ttttgggtct gttttgagtt tctatttgtt 3180
aatgtcttca taattataca cgtgttgagc acagcatgca aagtgattag gggaatctag 3240
ttggagtggg atggataccc aaatattcag actttcttgt gactcttctt tctgttacc 3300
acatcaaaaa aaaaaaaaaat ggagatgaga catggtcaga gtcactaaaa ccagctgcta 3360
cttttaatta cgtggggagc agtttctaac attgccatta ttgaactgat gctgcctggg 3420
tgaaaatgga aatcacttag tatttcttgt tggcaaagaa ttactgaatg gattaaattt 3480
ccaaaggag aagtcagtta caagtcttt ctttgtttat taggctttct gctatgataa 3540
attacactac ttccagaagt tacccttagg ccatgggaca ctggactatc actctgctgt 3600
cacaagagat tacagagtta gtcaaggcag cttgtgacac cttcagggac tgtcataaac 3660
ttccagcaag tcattaatcc tgaatgcaat actgtgtgtg tgtgtctatg tgtgtttgta 3720
tgtctgtgtg tgtcttatgt ctgtgtctct gtgtgtgtgt gtgtttgtgt gtgtgtgtgt 3780
atgtatgctt gtgtgtgtct tatgtctgtg tttgtgtgtc tgtgtgtgtc ttatgtctgt 3840
gtttgtatgt ctgtgtgtgt ctgtgtgtgt cttatgtctg tgtctctgtg tgtgtgtgtg 3900
tgtatgtatg tatgtatgta tgtatgtgta tgtgtttgca tctctctgtg tgtctgcgct 3960
tatataattg tgtatgtgtt tatgtgttcg cctttgtgag ttgttgggga ttgaatccag 4020
gggaatacaa atgttaagaa agaacgttac cactaagctt cacctgtagg ccttaaagct 4080
tttctttctt ttaaaaattg taattaatc attttcagtc aggatctcca cacctcgtcc 4140
ctgctgctct agaactcact atttaaacac aatcgccctc aaacctgcag caacctccc 4200
gcctctacc tgcgagcact agaataataa caggtgacct cacacgccta gattaagacc 4260
tttaaggtaa acattttact atattttagt ctcataagac aagatgctac aataaagctg 4320
tacataaagt tcctcgaat ttcttctgat tttaactcaa acataaggat ttctctcttt 4380
ttgattcagg taacagaaaa aatacacagg tacatacatg tacacacatg aacacacacg 4440
catcacaacc acatattgct acgcttgtgt gatctatcat ttaccatgcc actgaactct 4500
tctttcccca taaattctc tggacttgtg tgcctccag gaagaccgag gactctggag 4560
agctgcacgg gctcaccaca gatgagaagt ttgtagaagg agtgtacaga gtagaactgg 4620
acacaaaatc gtactggaag acacttgcca tttccccgtt ccatgaaatc gcggatgtaa 4680
gtggacacac caagttgttt ggattttgtt tttagtctca ggaaattccc ttcgctcttg 4740
ctgtacgatg ggcagtagtg gaaagtagat tccacagcca gaatccacag tgctgggaaa 4800

-continued

gcaagccttc tgaatttttc taaaactcat ttagcaacat ggectgaacc tgttcacact 4860
gcttatggtc agctaactat atttatgtaa atattcattt ctctgttgag gaaatgttag 4920
tatttgcttt tgaggcaacc tccagatacc atggagggca tgtcatagtc aaagagaggg 4980
ctccctatgg tattttctota aattctggca ttccctttat tccaaagcac atctagtgtc 5040
cccagaagtt tgggtagaca attcctggca acacagagaa ttacaacatg ttcaaaacc 5100
aacagcctaa tatctaaatc atcaagcaaa catcacatgg caaagggatt tctgaatcaa 5160
aactgtttca tccttatgat caacctatgg aggtetagcc tgcacttaca cccattttac 5220
caataagcta agagaagcta agttcctcat caaggacaca aggctagcat gtgtgagcaa 5280
gtgacagagt tgccctctat gttggttagt gtgccttagc cagtgtctca gtaagaaatg 5340
gagctaaatc aaaacccaag gccaacagcc aaaggcacaat gagtaacctt tgettggcac 5400
tgggctcagt ttccctggct cctctcagtc ctccagttcac agaggcagct gtcattgcaaa 5460
tagaatccaa gcttggttgg cagacctgga gataacaaat tccatcaaaa atagctcctc 5520
atgtgacctc gtttgcctgc tgttgcctatg atacacacca tgaccgaaaa gcaaccttgg 5580
ggagagaagg gtttatttca tcttacagct tacagttcac catggaggaa agccaggtgg 5640
gaacctggaa gtgaaaattg aagcagagac cagaaaaggaa tgctgtttac tggctggctt 5700
agctcctttt cttatacagc ttaggtctat gtgccagggg gatggtagctg ccgagcatag 5760
gctgagcccg cctacatcaa ccattagtca aaaaaaggtc catagacttg cctacaggcc 5820
aatctcatgg aggcaatacc ccagtgaggg gtccctcttc gcagggtact ctagtttgtg 5880
tcaagttgac aaaacctaac cacaaagcac aaacagggtc tgcccttggt gcttagccat 5940
ggatgacctc ctcagatgat ggtgttacca gacaaaccag aggggctcac caagagtctg 6000
ccacctacca aggtagtact ctactcctca ctgggcacca acaccatag tagctgggcc 6060
agtacaggac ccttgcctgt tctgcctatg attgtccata gacctgggt ctcagcctgc 6120
cgggagtacc tgtaagtagt cgctcaaac acattattcc tgggtgaaga cttgtctgat 6180
tctcttttag aactcaatca acaaacgttt ttattttgtt ttggcttttt ggagacaaga 6240
tctctcatag gccagcctga cttgaaatga gctgaggatg acctgtgctg ctaattctct 6300
cgctcttcc tcccaagtgg taggataata ggcataagac accacagcag tttactcca 6360
taccagggct ctgaaccag actttaaaca ctctatcaac tgattcacat tcccaccca 6420
tcattcaaca aacattttaa aaataaaacc cttctgcctt gagcactctg ctaaatacag 6480
cctttgagtg cggagtattt cctcacaacc agggccaag atgacccat catacatacc 6540
acggaaaatt aggagatgtt tttaggtctc tttgcttggg gtaattttta tgtgtgtgtg 6600
tacacagccc tgtgcgtgtg tgtgtgtgtg tgtgtgtgtg tgtacaggca cacacgtgta 6660
tgcattgtaga ggctacataa aaaccttagg tgcattctc aggcactctg ttcaccctt 6720
cacacagccc gaacacacaa aatttgaggc attagcctgg agctcaccag ttaggctaga 6780
ctgacttgcc agcagacccc aggctgtctc catctcccca gctctgggat tacaaactct 6840
atcataccag acatttttat acatattctg agcataaaat tcatgtcttc aggctaacaa 6900
gtcaagagct taaatgactg agctctotta cgtgggtggat tttttttaa actacataat 6960
atcttttttt ttttttccac ttctggggaa gaaacaaatg agcctgagtg acaatgagc 7020
agaaaagaaa ttttgaggag tgtgtgtgtc tgtgtgtgtg gtggcacatg cctctcatct 7080

-continued

```

aatgctagag gctacagtag aatgctcctg aattagtggc cagccaaggc caagggctag 7140
ggttgtaact cagtggcaga gggcttgctt agcattcgca ggatttgatc catagcgcta 7200
taaataataa taaataaata caacagtcta agatgattct ccctttcatt tatctggatg 7260
ttatttttgt gttagtttta ctctgtcacc caatcattgt ttgcccataa tttggacatt 7320
taaaaaaaaa ctttattcca agtgtgttca aagctgtatc caaacctgt ccaccaaatg 7380
agtccaatga catacatctt ctatattacc atctgttcca gatttggtct actcccggca 7440
cctgggctgt tgtgtcacc atgtctcaga tagtctagtg atttgagaag tgactagtaa 7500
ttgcaaaatc cagactttgt ccagaaactt ctatgagctc caaaactttc atttacattt 7560
ctgccagcca caaacctgtt gtgtttgtga gagaaccctg tgatgtcttc ccacagcatc 7620
tcagccttgt ttcttccctt aaaatattca tcttttcaca ttagaacatg caaagggaca 7680
gtgggagcga aacccttggc ctgggagcga cgaagccttc ctttctggtc aggetctcac 7740
tgtagaaact taggccggtt tcagcatgca gtctgtctga gaatggctcc tgccaacatt 7800
ccaggtctgg aagttttagt tggagtgtt gataaaccct gttcggcaca ggtcttttgt 7860
ttgtgggtgt cagtgtttct actctcctga cttttatctg aaccaagaa agggaacaat 7920
agccttcaag ctctctgtga ctctgatctg accagggcca cccacactgc agaaggaac 7980
ttgcaaagag agacctgcaa ttctetaaga gtccacaca gctccaaaga cttaggcagc 8040
atattttaat ctaattattc gtccccaac cccaccccag aggacagtta gacaataaaa 8100
ggaagattac cagcttagca tctgtgaac actttgtctg cagctcctac ctctgggctc 8160
tgttagaact agctgtctct cctctctctt aggtggtttt cacagccaac gactctggcc 8220
atgccacta caccatcgca gccctgctca gcccatactc ctacagcacc acggctgtcg 8280
tcagcaaccc ccagaattga 8300

```

```

<210> SEQ ID NO 21
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 21

```

```

cacagacaat cagacgtacc agta 24

```

```

<210> SEQ ID NO 22
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 22

```

```

ccagctttgc cagtttacga 20

```

```

<210> SEQ ID NO 23
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 23

```

-continued

ttggacgggtt gccctctt 18

<210> SEQ ID NO 24
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 24

gatggcttcc cttegactct tc 22

<210> SEQ ID NO 25
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 25

cactgacatt tctcttctct cctct 25

<210> SEQ ID NO 26
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 26

gggctcacca cagatgagaa g 21

<210> SEQ ID NO 27
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 27

cactgttcgc cacaggtctt 20

<210> SEQ ID NO 28
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 28

gctcagccca tactcctaca 20

<210> SEQ ID NO 29
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 29

gcccaggagg accaggat 18

-continued

<210> SEQ ID NO 30
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 30

ggcaacttgc ttgaggaaga 20

<210> SEQ ID NO 31
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 31

gcagcaaccc agcttcactt 20

<210> SEQ ID NO 32
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 32

actgagctgg gacttgaac 19

<210> SEQ ID NO 33
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 33

tgcctcactc tgagaacca 19

<210> SEQ ID NO 34
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 34

ggtggagagg ctattcggc 19

<210> SEQ ID NO 35
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 35

ggcctgtgat gtgttcag 18

<210> SEQ ID NO 36
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 36

ggttccatt tgctcttatt cgt 23

<210> SEQ ID NO 37
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 37

cccacactgc agaaggaaac ttg 23

<210> SEQ ID NO 38
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 38

ggttccatt tgctcttatt cgt 23

<210> SEQ ID NO 39
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 39

ccagcttagc atcctgtgaa ca 22

<210> SEQ ID NO 40
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 40

ggcaacttgc ttgaggaaga 20

<210> SEQ ID NO 41
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 41

tgtggagttc agtagtgtgg ag 22

<210> SEQ ID NO 42
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 42

-continued

cactgacatt tctctgtct cctct 25

<210> SEQ ID NO 43
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 43

gggacatctc ggttctctga ctt 23

<210> SEQ ID NO 44
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 44

tccacactac tgaactccac aa 22

<210> SEQ ID NO 45
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 45

cggaacactc gctctacgaa a 21

<210> SEQ ID NO 46
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 46

gggccagctt cagacaca 18

<210> SEQ ID NO 47
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 47

cccagggtgc tggagaatcc aa 22

<210> SEQ ID NO 48
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 48

gccaagtgtc ttccagtacg at 22

-continued

<210> SEQ ID NO 49
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 49

gttccctttc ttgggttcag a 21

<210> SEQ ID NO 50
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 50

gatgctactg ctttggcaag atc 23

<210> SEQ ID NO 51
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 51

cctgagctgc taacacggtt 20

<210> SEQ ID NO 52
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 52

agctacagac catgcttagt gta 23

<210> SEQ ID NO 53
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 53

tgccagttta ggaggaatat gttc 24

<210> SEQ ID NO 54
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 54

ctgaggaaac agaggtacca gatat 25

<210> SEQ ID NO 55
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 55

agtcacacag ttctgtcaaa tcag 24

<210> SEQ ID NO 56
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 56

gaacacggcg gcatcag 17

<210> SEQ ID NO 57
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 57

tcctgtggga gggttctttg 20

<210> SEQ ID NO 58
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 58

ccctctctct gagccctcta 20

<210> SEQ ID NO 59
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 59

gctgcctaag tctttggagc t 21

<210> SEQ ID NO 60
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 60

ccctctctct gagccctcta 20

<210> SEQ ID NO 61
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 61

-continued

gagaggagag acagctagtt ctaac 25

<210> SEQ ID NO 62
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 62

agctacagac catgcttagt gta 23

<210> SEQ ID NO 63
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 63

gccctcttca tacaggaatc ac 22

<210> SEQ ID NO 64
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 64

cggacagcat ccaggactt 19

<210> SEQ ID NO 65
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 65

tcatgtaatc tggcttcaga gtggga 26

<210> SEQ ID NO 66
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 66

tgggaggcaa ttcttagttt caatgga 27

<210> SEQ ID NO 67
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 67

tcccaaaggt gtctgtctgc aca 23

-continued

<210> SEQ ID NO 68
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 68

ctcctttgcc tcgctggact gg 22

<210> SEQ ID NO 69
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 69

cggacagcat ccaggactt 19

<210> SEQ ID NO 70
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 70

agaaggagtg tacagagtag aactggaca 29

<210> SEQ ID NO 71
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 71

tgtttgggg tgtcagtgtt tctactc 27

<210> SEQ ID NO 72
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 72

caccacggct gtcgtcagca a 21

<210> SEQ ID NO 73
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 73

cttgccaaag cagtagcatc cca 23

<210> SEQ ID NO 74
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 74

aggtcagaaa gcagagtgga cca 23

<210> SEQ ID NO 75
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 75

cccaggcaat tcctaccttc cca 23

<210> SEQ ID NO 76
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 76

tctgagcatt ctacctcatt gctttggt 28

<210> SEQ ID NO 77
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 77

aggctgtccc agcacctgag tcg 23

<210> SEQ ID NO 78
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 78

tgggcacaac agacaatcg ctg 23

<210> SEQ ID NO 79
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 79

aaggctgctg atgacacctg gga 23

<210> SEQ ID NO 80
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 80

-continued

agattcagac acacacaact taccagc 27

<210> SEQ ID NO 81
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 81

agacctgcaa ttctctaaga gctccaca 28

<210> SEQ ID NO 82
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 82

agattcagac acacacaact taccagc 27

<210> SEQ ID NO 83
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 83

ttgtctgcag ctctacctc tggg 24

<210> SEQ ID NO 84
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 84

aggtcagaaa gcagagtgga cca 23

<210> SEQ ID NO 85
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 85

ttgacatgtg tgggtgagag attttactg 29

<210> SEQ ID NO 86
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 86

cccagggtgc tggagaatcc aa 22

-continued

<210> SEQ ID NO 87
 <211> LENGTH: 16
 <212> TYPE: RNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic

 <400> SEQUENCE: 87

 guuuuagagc uaugcu 16

<210> SEQ ID NO 88
 <211> LENGTH: 67
 <212> TYPE: RNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic

 <400> SEQUENCE: 88

 agcauagcaa guuaaaauaa ggcuaguccg uuaucaacu gaaaaagugg caccgagucg 60
 gugcuuu 67

<210> SEQ ID NO 89
 <211> LENGTH: 77
 <212> TYPE: RNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic

 <400> SEQUENCE: 89

 guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60
 ggcaccgagu cggugcu 77

<210> SEQ ID NO 90
 <211> LENGTH: 444
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic

 <400> SEQUENCE: 90

 atggtttctc atcgtctgct cctcctctgc cttgctggac tggattttgt gtctgagget 60
 ggccctacgg gcaccggtga atccaagtgt cctctgatgg tcaaagttct agatgctgtc 120
 cgaggcagtc ctgccatcaa tgtggccgtg catgtgttca gaaaggctgc tgatgacacc 180
 tgggagccat ttgctctg gaaaaccagt gagtctggag agctgcatgg gctcacaact 240
 gaggaggaat ttgtagaagg gatatacaa gtggaaatag acaccaaadc ttactggaag 300
 gcacttgca tctccccatt ccatgagcat gcagagggtg tattcacagc caacgactcc 360
 ggccccccgc gctacaccat tggccgctg ctgagcccct actcctattc caccacgget 420
 gtcgtcacca atcccaagga atga 444

<210> SEQ ID NO 91
 <211> LENGTH: 444
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic

 <400> SEQUENCE: 91

 atggtctccc ttcgactctt cctcctttgc ctcgctggac tggattttgt gtctgaagct 60

-continued

```

ggccccgcgg gcaccggtga atccaagtgt cctctgatgg tcaaagttct agatgctgtc 120
cgaggcagtc ctgccatcaa tgtggccgtg catgtgttca gaaaggctgc tgatgacacc 180
tgggagccat ttgctctgga gaaaaccagt gagtctggag agctgcatgg gctcacaact 240
gaggaggaat ttgtagaagg gatatacaaa gtggaatag acaccaaatac ttactggaag 300
gcacttgga tctccccatt ccatgagcat gcagaggtgg tattcacagc caacgactcc 360
ggccccgcgg gctacacatc tgccgcccctg ctgagcccct actcctattc caccacggct 420
gtcgtcacca atccaagga atga 444

```

```

<210> SEQ ID NO 92
<211> LENGTH: 444
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 92

```

```

atggettccc ttcgactcct cctcctttgc ctcgctggac tggattttgt gtctgaagct 60
ggccccgcgg gtgctggaga atccaaatgt cctctgatgg tcaaagtctt ggatgctgtc 120
cgaggcagcc ctgctgtaga cgtggctgta aaagtgttca aaaagacctc tgagggatcc 180
tgggagccct ttgctctgga gaagaccgag gagtctggag agctgcaagg gctcaccaca 240
gatgagaagt ttgtagaagg agtgtacaga gtagaactgg acaccaaatac gtactggaag 300
acacttgga tttccccggt ccatgaatc gcggatgtgg ttttcacagc caacgactct 360
ggccatgcc actacacatc cgcagcccctg ctcagcccct actcctacag caccacggct 420
gtcgtcagca acccccagaa ttga 444

```

```

<210> SEQ ID NO 93
<211> LENGTH: 4176
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(3)
<223> OTHER INFORMATION: Start Codon
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (10)..(30)
<223> OTHER INFORMATION: 5' NLS
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (4126)..(4173)
<223> OTHER INFORMATION: 3' NLS
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (4174)..(4176)
<223> OTHER INFORMATION: Stop Codon

```

```

<400> SEQUENCE: 93

```

```

atggacaagc ccaagaaaaa gcggaaagtg aagtacagca tcggcctgga catcggcacc 60
aactctgtgg gctggggcgt gatcaccgac gagtacaagg tgcccagcaa gaaattcaag 120
gtgctgggca acaccgacag gcacagcacc aagaagaacc tgatcggcgc cctgctgttc 180
gacagcggcg aaacagccga ggccaccaga ctgaagagaa ccgccagaag aagatacacc 240
aggcgggaaga acaggatctg ctatctgcaa gagatcttca gcaacgagat ggccaaggtg 300

```

-continued

gacgacagct tcttccacag actggaagag tccttctctgg tggaagagga caagaagcac	360
gagagacaacc ccattctctgg caacatctgt gacgaggtgg cctaccacga gaagtacccc	420
accatctacc acctgagaaa gaaactggtg gacagcaccg acaaggccga cctgagactg	480
atctacctgg ccctggccca catgatcaag ttcagaggcc acttctctgat cgaggcgac	540
ctgaaccccg acaacagcga cgtggacaag ctgttcatcc agctggtgca gacctacaac	600
cagctgtctcg aggaaaaccc catcaacgcc agcggcgtgg acgccaaggc tatcctgtct	660
gccagactga gcaagagcag aaggctggaa aatctgatcg cccagctgcc cggcgagaag	720
aagaacggcc tgttcggcaa cctgattgcc ctgagcctgg gcctgacccc caacttcaag	780
agcaacttctg acctggccga ggatgcaaaa ctgcagctga gcaaggacac ctacgacgac	840
gacctggaca acctgctggc ccagatcggc gaccagtagc ccgacctgtt cctggccgcc	900
aagaacctgt ctgacgccat cctgctgagc gacatcctga gagtgaacac cgagatcacc	960
aaggccccc tgagcgctc tatgatcaag agatacgacg agcaccacca ggacctgacc	1020
ctgctgaaag ctctcgtgcg gcagcagctg cctgagaagt acaagaaat cttctctgac	1080
cagagcaaga acggtctcgc cggctacatc gatggcggcg ctagccagga agagttctac	1140
aagttcatca agcccatcct ggaagatg gacggcaccg aggaactgct cgtgaagctg	1200
aacagagagg acctgctgag aaagcagaga accttcgaca acggcagcat cccccaccg	1260
atccacctgg gagagctgca cgctatcctg agaaggcagg aagatttcta cccattcctg	1320
aaggacaacc gggaaaagat cgagaagatc ctgacctca ggatecccta ctacgtgggc	1380
cccctggcca gaggcaacag cagattcggc tggatgacca gaaagagcga ggaaccatc	1440
accctctgga acttcgagga agtgggtgac aaggcgcca gcgcccagag ctctcatcgag	1500
agaatgacaa acttcgataa gaacctgcc aacgagaagg tgctgcccga gcacagcctg	1560
ctgtacgagt acttcacctg gtacaacgag ctgaccaaag tgaataacgt gaccgagggg	1620
atgagaaagc ccgcttctct gagcggcgag cagaaaaagg ccatcgtgga cctgctgttc	1680
aagaccaaca gaaaagtgac cgtgaagcag ctgaaagagg actacttcaa gaaaatcgag	1740
tgcttcgact ccgtggaaat ctccggcgtg gaagatagat tcaacgcctc cctgggcaca	1800
taccacgac tgctgaaaat tatcaaggac aaggacttcc tggataacga agagaacgag	1860
gacattctgg aagatatcgt gctgacctg acactgtttg aggaccgcga gatgatcgag	1920
gaaaggctga aaacctacgc tcacctgttc gacgacaaag tgatgaagca gctgaagaga	1980
aggcggta ca cggctgggg caggctgagc agaaagctga tcaacggcat cagagacaag	2040
cagagcggca agacaatcct ggatttctctg aagtccgacg gcttcgcca cgggaacttc	2100
atgcagctga tccacgacga cagcctgaca ttcaaagagg acatccagaa agcccaggtg	2160
tccggccagg gcgactctct gcacgagcat atcgctaacc tggccggcag ccccgctatc	2220
aagaagggca tcctgcagac agtgaaggtg gtggacgagc tcgtgaaagt gatgggcaga	2280
cacaagccc agaacatcgt gatcgagatg gctagagaga accagaccac ccagaaggga	2340
cagaagaact ccccgagag gatgaagaga atcgaagagg gcatcaaaga gctgggcagc	2400
cagatcctga aagaacaccc cgtggaaaac acccagctgc agaacgagaa gctgtacctg	2460
tactacctgc agaatggccg ggatatgtac gtggaccagg aactggacat caacagactg	2520
tccgactaog atgtggacca tatcgtgctc cagagcttcc tgaaggacga ctccatgat	2580

-continued

aacaaagtgc	tgactcggag	cgacaagaac	agaggcaaga	gcgacaacgt	gccctccgaa	2640
gaggtcgtga	agaagatgaa	gaactactgg	cgacagctgc	tgaacgcca	gctgattacc	2700
cagaggaagt	tcgataacct	gaccaaggcc	gagagaggcg	gcctgagcga	gctggataag	2760
gccggcttca	tcaagaggca	gctggtgga	accagacaga	tcacaaagca	cgtggcacag	2820
atcctggact	cccggatgaa	cactaagtac	gacgaaaacg	ataagctgat	ccgggaagtg	2880
aaagtgatca	ccctgaagtc	caagctggtg	tccgatttcc	ggaaggattt	ccagttttac	2940
aaagtgcgcg	agatcaacaa	ctaccaccac	gcccacgacg	cctacctgaa	cgccgtcgtg	3000
ggaaccgccc	tgatcaaaaa	gtaccctaag	ctggaaagcg	agttcgtgta	cgcgactac	3060
aaggtgtacg	acgtgcggaa	gatgatcgcc	aagagcagc	aggaatcgg	caaggctacc	3120
gccaagtact	tcttctacag	caacatcatg	aactttttca	agaccgaaat	caccctggcc	3180
aacggcgaga	tcagaaagcg	ccctctgatc	gagacaaaacg	gcgaaaccgg	ggagatcgtg	3240
tgggataaag	gcagagactt	cgccacagtg	cgaaagggtc	tgagcatgcc	ccaagtgaat	3300
atcgtgaaaa	agaccgaggt	gcagacaggc	ggcttcagca	aagagtctat	cctgccaag	3360
aggaacagcg	acaagctgat	cgccagaaaag	aaggactggg	acccaagaa	gtacggcggc	3420
ttcgacagcc	ctaccgtggc	ctactctgtg	ctgggtgggg	ctaaggtgga	aaagggcaag	3480
tccaagaaac	tgaagagtgt	gaaagagtgt	ctggggatca	ccatcatgga	aagaagcagc	3540
tttgagaaga	accctatcga	ctttctggaa	gccaaaggct	acaagaagt	gaaaaggac	3600
ctgatcatca	agctgcctaa	gtactccctg	ttcgagctgg	aaaacggcag	aaagagaatg	3660
ctggcctctg	ccggcgaact	gcagaaggga	aacgagctgg	ccctgcctag	caaatatgtg	3720
aacttcctgt	acctggcctc	ccactatgag	aagctgaagg	gcagccctga	ggacaacgaa	3780
cagaaacagc	tgtttgtgga	acagcataag	cactacctgg	acgagatcat	cgagcagatc	3840
agcgagttct	ccaagagagt	gatcctggcc	gacgccaatc	tggacaaggt	gctgtctgcc	3900
tacaacaagc	acagggacaa	gcctatcaga	gagcaggccg	agaatatcat	ccacctgttc	3960
accctgacaa	acctgggctg	tctgcccgc	ttcaagtact	ttgacaccac	catcgaccgg	4020
aagaggtaca	ccagcaccaa	agaggtgctg	gacgccccc	tgatccacca	gagcatcacc	4080
ggcctgtacg	agacaagaat	cgacctgtct	cagctgggag	gcgacaagag	acctgcccgc	4140
actaagaagg	ccggacaggc	caaaaagaag	aagtga			4176

<210> SEQ ID NO 94
 <211> LENGTH: 1391
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic
 <220> FEATURE:
 <221> NAME/KEY: MISC_FEATURE
 <222> LOCATION: (4)..(10)
 <223> OTHER INFORMATION: 5' NLS
 <220> FEATURE:
 <221> NAME/KEY: MISC_FEATURE
 <222> LOCATION: (1376)..(1391)
 <223> OTHER INFORMATION: 3' NLS

<400> SEQUENCE: 94

Met Asp Lys Pro Lys Lys Lys Arg Lys Val Lys Tyr Ser Ile Gly Leu
 1 5 10 15

-continued

Asp	Ile	Gly	Thr	Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr
			20					25					30		
Lys	Val	Pro	Ser	Lys	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His
		35				40					45				
Ser	Ile	Lys	Lys	Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu
	50				55						60				
Thr	Ala	Glu	Ala	Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr
	65				70					75					80
Arg	Arg	Lys	Asn	Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu
				85					90					95	
Met	Ala	Lys	Val	Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe
			100					105					110		
Leu	Val	Glu	Glu	Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn
		115					120					125			
Ile	Val	Asp	Glu	Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His
	130					135					140				
Leu	Arg	Lys	Lys	Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu
	145				150					155					160
Ile	Tyr	Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu
				165					170					175	
Ile	Glu	Gly	Asp	Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe
			180					185					190		
Ile	Gln	Leu	Val	Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile
		195					200					205			
Asn	Ala	Ser	Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser
	210					215					220				
Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys
	225				230					235					240
Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr
				245					250					255	
Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln
			260					265					270		
Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln
		275					280					285			
Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser
	290					295					300				
Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr
	305				310					315					320
Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His
				325					330					335	
Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu
			340					345					350		
Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly
		355					360					365			
Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys
	370					375					380				
Pro	Ile	Leu	Glu	Lys	Met	Asp	Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu
	385				390					395					400
Asn	Arg	Glu	Asp	Leu	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser
				405					410					415	
Ile	Pro	His	Gln	Ile	His	Leu	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg

-continued

420					425					430					
Gln	Glu	Asp	Phe	Tyr	Pro	Phe	Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu
	435						440					445			
Lys	Ile	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg
	450					455					460				
Gly	Asn	Ser	Arg	Phe	Ala	Trp	Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile
	465					470					475				480
Thr	Pro	Trp	Asn	Phe	Glu	Glu	Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln
				485					490					495	
Ser	Phe	Ile	Glu	Arg	Met	Thr	Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu
			500					505					510		
Lys	Val	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr
		515					520					525			
Asn	Glu	Leu	Thr	Lys	Val	Lys	Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro
	530					535					540				
Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe
	545					550					555				560
Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe
				565					570					575	
Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp
		580						585					590		
Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile
		595					600						605		
Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu
	610					615					620				
Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	Glu
	625					630					635				640
Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys
				645					650					655	
Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys
			660					665						670	
Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp
		675					680						685		
Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile
	690					695					700				
His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val
	705					710					715				720
Ser	Gly	Gln	Gly	Asp	Ser	Leu	His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly
				725					730					735	
Ser	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp
			740					745					750		
Glu	Leu	Val	Lys	Val	Met	Gly	Arg	His	Lys	Pro	Glu	Asn	Ile	Val	Ile
		755					760						765		
Glu	Met	Ala	Arg	Glu	Asn	Gln	Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser
	770					775					780				
Arg	Glu	Arg	Met	Lys	Arg	Ile	Glu	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser
	785					790					795				800
Gln	Ile	Leu	Lys	Glu	His	Pro	Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu
				805					810					815	
Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp
			820					825					830		

-continued

Gln	Glu	Leu	Asp	Ile	Asn	Arg	Leu	Ser	Asp	Tyr	Asp	Val	Asp	His	Ile
	835						840					845			
Val	Pro	Gln	Ser	Phe	Leu	Lys	Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu
	850					855					860				
Thr	Arg	Ser	Asp	Lys	Asn	Arg	Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu
865					870					875					880
Glu	Val	Val	Lys	Lys	Met	Lys	Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala
					885					890					895
Lys	Leu	Ile	Thr	Gln	Arg	Lys	Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg
			900					905					910		
Gly	Gly	Leu	Ser	Glu	Leu	Asp	Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu
		915					920					925			
Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser
	930					935					940				
Arg	Met	Asn	Thr	Lys	Tyr	Asp	Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val
945					950					955					960
Lys	Val	Ile	Thr	Leu	Lys	Ser	Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp
				965					970						975
Phe	Gln	Phe	Tyr	Lys	Val	Arg	Glu	Ile	Asn	Asn	Tyr	His	His	Ala	His
			980						985					990	
Asp	Ala	Tyr	Leu	Asn	Ala	Val	Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr
		995					1000						1005		
Pro	Lys	Leu	Glu	Ser	Glu	Phe	Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	
	1010					1015						1020			
Asp	Val	Arg	Lys	Met	Ile	Ala	Lys	Ser	Glu	Gln	Glu	Ile	Gly	Lys	
	1025					1030						1035			
Ala	Thr	Ala	Lys	Tyr	Phe	Phe	Tyr	Ser	Asn	Ile	Met	Asn	Phe	Phe	
	1040					1045						1050			
Lys	Thr	Glu	Ile	Thr	Leu	Ala	Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	
	1055					1060						1065			
Leu	Ile	Glu	Thr	Asn	Gly	Glu	Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	
	1070					1075						1080			
Gly	Arg	Asp	Phe	Ala	Thr	Val	Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	
	1085					1090						1095			
Val	Asn	Ile	Val	Lys	Lys	Thr	Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	
	1100					1105						1110			
Lys	Glu	Ser	Ile	Leu	Pro	Lys	Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	
	1115					1120						1125			
Arg	Lys	Lys	Asp	Trp	Asp	Pro	Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	
	1130					1135						1140			
Pro	Thr	Val	Ala	Tyr	Ser	Val	Leu	Val	Val	Ala	Lys	Val	Glu	Lys	
	1145					1150						1155			
Gly	Lys	Ser	Lys	Lys	Leu	Lys	Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	
	1160					1165						1170			
Thr	Ile	Met	Glu	Arg	Ser	Ser	Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	
	1175					1180						1185			
Leu	Glu	Ala	Lys	Gly	Tyr	Lys	Glu	Val	Lys	Lys	Asp	Leu	Ile	Ile	
	1190					1195						1200			
Lys	Leu	Pro	Lys	Tyr	Ser	Leu	Phe	Glu	Leu	Glu	Asn	Gly	Arg	Lys	
	1205					1210						1215			

-continued

Arg	Met	Leu	Ala	Ser	Ala	Gly	Glu	Leu	Gln	Lys	Gly	Asn	Glu	Leu
	1220					1225					1230			
Ala	Leu	Pro	Ser	Lys	Tyr	Val	Asn	Phe	Leu	Tyr	Leu	Ala	Ser	His
	1235					1240					1245			
Tyr	Glu	Lys	Leu	Lys	Gly	Ser	Pro	Glu	Asp	Asn	Glu	Gln	Lys	Gln
	1250					1255					1260			
Leu	Phe	Val	Glu	Gln	His	Lys	His	Tyr	Leu	Asp	Glu	Ile	Ile	Glu
	1265					1270					1275			
Gln	Ile	Ser	Glu	Phe	Ser	Lys	Arg	Val	Ile	Leu	Ala	Asp	Ala	Asn
	1280					1285					1290			
Leu	Asp	Lys	Val	Leu	Ser	Ala	Tyr	Asn	Lys	His	Arg	Asp	Lys	Pro
	1295					1300					1305			
Ile	Arg	Glu	Gln	Ala	Glu	Asn	Ile	Ile	His	Leu	Phe	Thr	Leu	Thr
	1310					1315					1320			
Asn	Leu	Gly	Ala	Pro	Ala	Ala	Phe	Lys	Tyr	Phe	Asp	Thr	Thr	Ile
	1325					1330					1335			
Asp	Arg	Lys	Arg	Tyr	Thr	Ser	Thr	Lys	Glu	Val	Leu	Asp	Ala	Thr
	1340					1345					1350			
Leu	Ile	His	Gln	Ser	Ile	Thr	Gly	Leu	Tyr	Glu	Thr	Arg	Ile	Asp
	1355					1360					1365			
Leu	Ser	Gln	Leu	Gly	Gly	Asp	Lys	Arg	Pro	Ala	Ala	Thr	Lys	Lys
	1370					1375					1380			
Ala	Gly	Gln	Ala	Lys	Lys	Lys	Lys							
	1385					1390								

1. A non-human animal comprising in its genome a genetically modified endogenous Ttr locus comprising a human TTR sequence comprising both TTR coding sequence and non-coding sequence.

2. The non-human animal of claim 1, wherein a region of the endogenous Ttr locus comprising both Ttr coding sequence and non-coding sequence has been deleted and replaced with a corresponding human TTR sequence comprising both TTR coding sequence and non-coding sequence.

3. The non-human animal of claim 1, wherein the genetically modified endogenous Ttr locus comprises the endogenous Ttr promoter, wherein the human TTR sequence is operably linked to the endogenous Ttr promoter.

4. The non-human animal of claim 1, wherein at least one intron and at least one exon of the endogenous Ttr locus have been deleted and replaced with the corresponding human TTR sequence.

5. The non-human animal of claim 1, wherein the entire Ttr coding sequence of the endogenous Ttr locus has been deleted and replaced with the corresponding human TTR sequence.

6. The non-human animal of claim 5, wherein the region of the endogenous Ttr locus from the Ttr start codon to the Ttr stop codon has been deleted and replaced with the corresponding human TTR sequence.

7. The non-human animal of claim 1, wherein the genetically modified endogenous Ttr locus comprises a human TTR 3' untranslated region.

8. The non-human animal of claim 1, wherein the endogenous Ttr 5' untranslated region has not been deleted and replaced with the corresponding human TTR sequence.

9. The non-human animal of claim 1, wherein the region of the endogenous Ttr locus from the Ttr start codon to the Ttr stop codon has been deleted and replaced with a human TTR sequence comprising the corresponding human TTR sequence and a human TTR 3' untranslated region, and

wherein the endogenous Ttr 5' untranslated region has not been deleted and replaced with the corresponding human TTR sequence, and

wherein the endogenous Ttr promoter has not been deleted and replaced with the corresponding human TTR sequence.

10. The non-human animal of claim 9, wherein:

(i) the human TTR sequence at the genetically modified endogenous Ttr locus comprises a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 18; or

(ii) the genetically modified endogenous Ttr locus encodes a protein comprising a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 1;

(iii) the genetically modified endogenous Ttr locus comprises a coding sequence comprising a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 90; or

(iv) the genetically modified endogenous Ttr locus comprises a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 14 or 15.

11. The non-human animal of claim **1**, wherein the genetically modified endogenous Ttr locus encodes a transthyretin precursor protein comprising a signal peptide, and the region of the endogenous Ttr locus encoding the signal peptide has not been deleted and replaced with the corresponding human TTR sequence.

12. The non-human animal of claim **11**, wherein the first exon of the endogenous Ttr locus has not been deleted and replaced with the corresponding human TTR sequence.

13. The non-human animal of claim **12**, wherein the first exon and first intron of the endogenous Ttr locus have not been deleted and replaced with the corresponding human TTR sequence.

14. The non-human animal of claim **11**, wherein the region of the endogenous Ttr locus from the start of the second Ttr exon to the Ttr stop codon has been deleted and replaced with the corresponding human TTR sequence.

15. The non-human animal of claim **11**, wherein the genetically modified endogenous Ttr locus comprises a human TTR 3' untranslated region.

16. The non-human animal of claim **11**, wherein the region of the endogenous Ttr locus from the second Ttr exon to the Ttr stop codon has been deleted and replaced with a human TTR sequence comprising the corresponding human TTR sequence and a human TTR 3' untranslated region, and wherein the endogenous Ttr 5' untranslated region has not been deleted and replaced with the corresponding human TTR sequence, and wherein the endogenous Ttr promoter has not been deleted and replaced with the corresponding human TTR sequence.

17. The non-human animal of claim **16**, wherein:

- (i) the human TTR sequence at the genetically modified endogenous Ttr locus comprises a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 19; or
- (ii) the genetically modified endogenous Ttr locus encodes a protein comprising a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 2;
- (iii) the genetically modified endogenous Ttr locus comprises a coding sequence comprising a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 91; or
- (iv) the genetically modified endogenous Ttr locus comprises a sequence at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence set forth in SEQ ID NO: 16 or 17.

18. The non-human animal of claim **1**, wherein the genetically modified endogenous Ttr locus does not comprise a selection cassette or a reporter gene.

19. The non-human animal of claim **1**, wherein the non-human animal is homozygous for the genetically modified endogenous Ttr locus.

20. The non-human animal of claim **1**, wherein the non-human animal is a mammal.

21. The non-human animal of claim **20**, wherein the non-human animal is a rat or mouse.

22. The non-human animal of claim **21**, wherein the non-human animal is a mouse.

23. A method of assessing the activity of a human-TTR-targeting reagent in vivo, comprising:

- (a) administering the human-TTR-targeting reagent to the non-human animal of claim **1**; and
- (b) assessing the activity of the human-TTR-targeting reagent in the non-human animal.

24.-40. (canceled)

41. A method of optimizing the activity of a human-TTR-targeting reagent in vivo, comprising:

- (I) performing the method of claim **23a** first time in a first non-human animal comprising in its genome a genetically modified endogenous Ttr locus comprising a human TTR sequence comprising both TTR coding sequence and non-coding sequence;
- (II) changing a variable and performing the method of step (I) a second time with the changed variable in a second non-human animal comprising in its genome the genetically modified endogenous Ttr locus comprising the human TTR sequence comprising both TTR coding sequence and non-coding sequence; and
- (III) comparing the activity of the human-TTR-targeting reagent in step (I) with the activity of the human-TTR-targeting reagent in step (II), and selecting the method resulting in the higher activity.

42.-51. (canceled)

52. A method of making the non-human animal of claim **1**, comprising:

- (a) introducing into a non-human animal embryonic stem (ES) cell:
 - (i) a nuclease agent that targets a target sequence in the endogenous Ttr locus; and
 - (ii) a targeting vector comprising a nucleic acid insert comprising the human TTR sequence flanked by a 5' homology arm corresponding to a 5' target sequence in the endogenous Ttr locus and a 3' homology arm corresponding to a 3' target sequence in the endogenous Ttr locus,

wherein the targeting vector recombines with the endogenous Ttr locus to produce a genetically modified non-human ES cell comprising in its genome the genetically modified endogenous Ttr locus comprising the human TTR sequence;

- (b) introducing the genetically modified non-human ES cell into a non-human animal host embryo; and
- (c) gestating the non-human animal host embryo in a surrogate mother, wherein the surrogate mother produces an F0 progeny genetically modified non-human animal comprising in its genome the genetically modified endogenous Ttr locus comprising the human TTR sequence.

53.-57. (canceled)

* * * * *