

**(12) STANDARD PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

(11) Application No. **AU 2019336793 B2**

(54) Title  
**RNA and DNA base editing via engineered ADAR recruitment**

(51) International Patent Classification(s)  
**C12N 15/11** (2006.01) **C12N 15/113** (2010.01)

(21) Application No: **2019336793** (22) Date of Filing: **2019.09.06**

(87) WIPO No: **WO20/051555**

(30) Priority Data

| (31) Number       | (32) Date         | (33) Country |
|-------------------|-------------------|--------------|
| <b>62/766,433</b> | <b>2018.10.17</b> | <b>US</b>    |
| <b>62/780,241</b> | <b>2018.12.15</b> | <b>US</b>    |
| <b>62/728,007</b> | <b>2018.09.06</b> | <b>US</b>    |
| <b>62/773,146</b> | <b>2018.11.29</b> | <b>US</b>    |
| <b>62/773,150</b> | <b>2018.11.29</b> | <b>US</b>    |

(43) Publication Date: **2020.03.12**

(44) Accepted Journal Date: **2024.11.21**

(71) Applicant(s)  
**The Regents of the University of California**

(72) Inventor(s)  
**MALI, Prashant;KATREKAR, Dhruva;MELUZZI, Dario;CHEN, Genghao;FORD, Kyle M.**

(74) Agent / Attorney  
**Spruson & Ferguson, GPO Box 3898, Sydney, NSW, 2001, AU**

(56) Related Art  
**WO 2017/220751 A1**



## (51) International Patent Classification:

C12N 15/11 (2006.01) C12N 15/113 (2010.01)

## (21) International Application Number:

PCT/US2019/050095

## (22) International Filing Date:

06 September 2019 (06.09.2019)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

62/728,007 06 September 2018 (06.09.2018) US  
 62/766,433 17 October 2018 (17.10.2018) US  
 62/773,146 29 November 2018 (29.11.2018) US  
 62/773,150 29 November 2018 (29.11.2018) US  
 62/780,241 15 December 2018 (15.12.2018) US

## (71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 1111 Franklin Street, 5th Floor, Oakland, CA 94607-52000 (US).

## (72) Inventors: MALI, Prashant; c/o UC San Diego, 9500 Gilman Drive, Mail Code:0910, La Jolla, CA 92093 (US). KATREKAR, Dhruva; c/o UC San Diego, 9500 Gilman Drive, Mail Code:0910, La Jolla, CA 92093 (US). MELUZZI, Dario; c/o UC San Diego, 9500 Gilman Drive, Mail Code:0910, La Jolla, CA 92093 (US). CHEN, Genghao; c/o UC San Diego, 9500 Gilman Drive, Mail Code:0910, La Jolla, CA 92093 (US). FORD, Kyle M.; c/o UC San Diego, 9500 Gilman Drive, Mail Code:0910, La Jolla, CA 92093 (US).

## (74) Agent: BAKER, Joseph R., Jr.; Gavrilovich, Dodd &amp; Lindsay LLP, 4370 La Jolla Village Drive, Suite 303, San Diego, CA 92122 (US).

## (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

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

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

## Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

## (54) Title: RNA AND DNA BASE EDITING VIA ENGINEERED ADAR RECRUITMENT

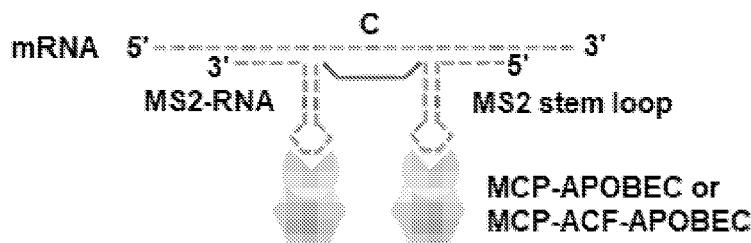


FIG. 4

% edited mRNA (RAB7A)

(57) Abstract: Disclosed herein is a system to recruit ADARs to catalyze therapeutic editing of point mutations via the use of engineered RNA scaffolds, engineered DNA scaffolds or DNA-RNA hybrid scaffolds. The system comprises an engineered ADAR2 guide RNA (adRNA) that bears a 20-100bp complementarity with the target RNA and ADAR2 recruiting domain from the GluR2 mRNA at either or both the 5' end or the 3' end.



# RNA AND DNA BASE EDITING VIA ENGINEERED ADAR RECRUITMENT

## CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 62/728,007, filed September 6, 2018, U.S. Provisional Application No. 62/766,433, filed October 17, 2018, and U.S. Provisional Application No. 62/780,241, filed December 15, 2018, which are incorporated by reference herein in their entirety.

## STATEMENT REGARDING GOVERNMENT SUPPORT

[0002] This disclosure was made with government support under grant numbers CA222826, GM123313, and HG009285 awarded by the National Institutes of Health. The government has certain rights in the invention.

## SEQUENCE LISTING

[0002.1] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on October 18, 2019, is named 00015-365WO1\_SL.txt and is 163,723 bytes in size.

## SUMMARY

[0002.2] In a first aspect, the present invention provides a vector comprising a nucleic acid encoding an engineered guide RNA that comprises a targeting domain that hybridizes to a target RNA and forms a double stranded RNA comprising a mismatch formed by a target adenosine of the target RNA and a cytosine of the engineered guide RNA, wherein the double stranded RNA binds to an ADAR1 or an ADAR2 enzyme present in a cell, resulting in a targeted deamination of the target adenosine in the target RNA; and wherein the engineered guide RNA does not have a GluR2 ADAR recruiting domain.

[0002.3] In a second aspect, the present invention provides a pharmaceutical composition comprising the vector of the first aspect, and a pharmaceutically acceptable excipient, diluent, or carrier.

[0002.4] In a third aspect, the present invention provides a method of treating a disease or condition in a subject in need thereof, the method comprising administering to the subject the pharmaceutical composition of the second aspect in an amount sufficient to treat the disease or condition.

[0002.5] In a fourth aspect, the present invention provides use of the vector of the first aspect in the manufacture of a medicament for the treatment of a disease or condition selected from

the group consisting of a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, and any combination thereof.

**[0003]** An aspect of the disclosure provides a vector. In some cases, the vector can comprise a nucleic acid with a polynucleotide sequence encoding for at least one RNA editing entity recruiting domain, wherein: (a) the polynucleotide sequence encoding for the at least one RNA editing entity recruiting domain does not form a secondary structure comprising a stem-loop, or (b) wherein the polynucleotide sequence encoding for the at least one RNA editing entity recruiting domain comprises at least about 80% sequence identity to at least one sequence selected from: an Alu domain encoding sequence, an Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC) recruiting domain encoding sequence, and any combination thereof. In some cases, the polynucleotide sequence encoding for the RNA editing entity recruiting domain can comprise at least about 80% sequence identity to the Alu domain sequence. In some cases, the polynucleotide sequence encoding for the RNA editing entity recruiting domain can comprise at least about 80% sequence identity to the APOBEC recruiting domain encoding sequence. In some cases, the vector can be a viral vector. In some cases, the vector can be a liposome. In some cases, the vector can be a nanoparticle. In some cases, the RNA editing entity recruiting domain can be configured to recruit an ADAR protein. In some cases, the ADAR protein can be an ADAR1, ADAR2, or ADAR3 protein. In some cases, the RNA editing entity recruiting domain can be configured to recruit an ADAR protein. In some cases, the ADAR protein can be an ADAR1, ADAR2, or ADAR3 protein. In some cases, the ADAR protein can be a human ADAR protein. In some cases, the ADAR protein can be a recombinant ADAR protein. In some cases, the ADAR protein can be a modified ADAR protein. In some cases, the RNA editing entity recruiting domain can be configured to recruit an APOBEC protein. In some cases, the APOBEC protein can be an APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3E, APOBEC3F



APOBEC3G, APOBEC3H, or APOBEC4 protein. In some cases, the ADAR protein can be a human ADAR protein. In some cases, the ADAR protein can be a recombinant ADAR protein. In some cases, the ADAR protein can be a modified ADAR protein. In some cases, the at least one RNA editing entity recruiting domain may not form a secondary structure comprising a stem-loop. In some cases, the polynucleotide sequence can encode for at least two RNA editing recruiting domains. In some cases, at least one of the at least two RNA editing recruiting domains can be an Alu domain. In some cases, the Alu domain sequence can form a secondary structure that comprises at least one stem-loop. In some cases, the Alu domain encoding sequence can comprise a plurality of Alu repeats. In some cases, the Alu domain encoding sequence can be at least partially single stranded. In some cases, at least one of the at least two RNA editing recruiting domains can be an APOBEC recruiting domain. In some cases, at least one of the at least two RNA editing recruiting domain encoding sequences can comprise at least about 80% sequence identity to a GluR2 domain encoding sequence. In some cases, at least one of the at least two RNA editing recruiting domains can be a GluR2 domain. In some cases, at least one of the at least two RNA editing recruiting domains can be a Cas13 domain. In some cases, the at least two RNA editing recruiting domains can be the Alu domain and the APOBEC recruiting domain. In some cases, the vector can further comprise a nucleic acid encoding for an RNA that can be complementary to at least a portion of a target RNA. In some cases, the nucleic acid encoding for the RNA that can be complementary to at least the portion of the target RNA can be from about 10 base pairs (bp) to about 1000 bp in length. In some cases, the nucleic acid encoding for the at least one RNA editing entity recruiting domain and the nucleic acid encoding for the RNA that can be complementary to at least the portion of the target RNA can comprise a contiguous nucleic acid of at least about 200 bp in length. In some cases, the nucleic acid can be chemically synthesized. In some cases, the nucleic acid can be genetically encoded. In some cases, the vector can comprise DNA. In some cases, the DNA can be double stranded. In some cases, the DNA can be single stranded. In some cases, the vector can comprise RNA. In some cases, the RNA can comprise a base modification. In some cases, the vector can be an adeno-associated virus (AAV) vector. In some cases, the AAV can be a recombinant AAV (rAAV). In some cases, the AAV can be selected from the group consisting of an AAV1 serotype, an AAV2 serotype, an AAV3 serotype, an AAV4 serotype, an AAV5 serotype, an AAV6 serotype, an AAV7 serotype, an AAV8 serotype, an AAV9 serotype, a derivative of any these, and a combination of any of these. In some cases, the AAV can be the AAV5

serotype or a derivative thereof. In some cases, the derivative of the AAV can comprise a modified VP1 protein. In some cases, the APOBEC recruiting domain can be selected from the group consisting of: an APOBEC1 recruiting domain, an APOBEC2 recruiting domain, an APOBEC3A recruiting domain, an APOBEC3B recruiting domain, an APOBEC3C recruiting domain, an APOBEC3E recruiting domain, an APOBEC3F recruiting domain, an APOBEC3G recruiting domain, an APOBEC3H recruiting domain, an APOBEC4 recruiting domain, and any combination thereof. In some cases, the RNA editing entity recruiting domain can recruit at least two RNA editing entities, and wherein at least one of the at least two polynucleotide sequences encoding for the RNA editing entities comprises at least about 80% identity to an APOBEC protein encoding sequence. In some cases, the RNA editing entity recruiting domain can recruit at least two RNA editing entities, and wherein at least one of the at least two polynucleotide sequences encoding for the RNA editing entities comprises at least about 80% identity to an ADAR protein encoding sequence. In some cases, the RNA recruiting domain encoded by the nucleic acid can comprise at least one stem loop. In some cases, the polynucleotide sequence encoding for the RNA editing recruiting domain can comprise a secondary structure that can be substantially a cruciform. In some cases, the polynucleotide sequence encoding for the RNA editing recruiting domain can comprise at least two secondary structures that are substantially cruciforms. In some cases, the polynucleotide sequence encoding for the RNA editing entity recruiting domain can be positioned between a polynucleotide sequence that forms the at least two secondary structures that are substantially cruciforms. In some cases, the cruciform secondary structure can comprise a stem-loop adjoining at least one pair of at least partially complementary strands of the cruciform secondary structure. In some cases, the polynucleotide sequence encoding for the RNA editing recruiting domain can comprise a secondary structure that can be substantially a toehold. In some cases, a non-naturally occurring RNA can be encoded by the vector. In some cases, a kit can comprise the vector in a container. In some cases, the kit can further comprise a syringe. In some cases, the container can be the syringe. In some cases, an isolated cell can comprise the vector. In some cases, a pharmaceutical composition can comprise the vector in unit dose form. In some cases, the pharmaceutical composition can further comprise a pharmaceutically acceptable excipient, diluent, or carrier. In some cases, the pharmaceutical composition can comprise a second active ingredient. In some cases, a method of treating a disease or condition in a subject comprising administering to the subject the vector. In some cases, the administering can be by intravenous injection, intramuscular

injection, an intrathecal injection, an intraorbital injection, a subcutaneous injection, or any combination thereof. In some cases, the method can further comprise administering a second therapy to the subject. In some cases, the disease or condition can be selected from the group consisting of: a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, a cell proliferative disorder (*e.g.*, a neoplasm) and any combination thereof. In some cases, the disease or condition can be Alzheimer's disease. In some cases, the disease or condition can be muscular dystrophy. In some cases, the disease or condition can be retinitis pigmentosa. In some cases, the disease or condition can be Parkinson disease. In some cases, the disease or condition can be pain. In some cases, the disease or condition can be Stargardt macular dystrophy. In some cases, the disease or condition can be Charcot-Marie-Tooth disease. In some cases, the disease or condition can be Rett syndrome. In some cases, the administering can be sufficient to decrease expression of a gene relative to prior to the administering. In some cases, the administering can be sufficient to edit at least one point mutation in the subject. In some cases, the administering can be sufficient to edit at least one stop codon in the subject, thereby producing a readthrough of the stop codon. In some cases, the administering can be sufficient to produce an exon skip in the subject.

**[0004]** Another aspect of the disclosure provides a vector. In some cases, the vector can comprise a nucleic acid encoding for RNA with a two dimensional shape that can be substantially a cruciform, wherein the RNA comprises at least one sequence encoding an RNA editing entity recruiting domain. In some cases, the vector can further comprise a nucleic acid encoding for RNA with at least one targeting domain encoding sequence that can be complementary to at least a portion of a target RNA sequence. In some cases, the nucleic acid encoding for the RNA with the at least one targeting domain that can be complementary to at least the portion of the target RNA sequence further can comprise a substantially linear two dimensional structure. In some cases, a non-naturally occurring RNA can be encoded by the vector. In some cases, a kit can comprise the vector in a container. In some cases, the kit can further comprise a syringe. In some cases, the container can be the syringe. In some cases, an isolated cell can comprise the vector. In some cases, a pharmaceutical composition can comprise the vector in unit dose form. In some cases, the pharmaceutical composition can further comprise a pharmaceutically acceptable excipient, diluent, or carrier. In some cases, the pharmaceutical composition can comprise a second active ingredient. In some cases, a method of treating a disease or condition in a subject comprising administering to the subject the vector. In some cases, the administering can be by intravenous injection, intramuscular

injection, an intrathecal injection, an intraorbital injection, a subcutaneous injection, or any combination thereof. In some cases, the method can further comprise administering a second therapy to the subject. In some cases, the disease or condition can be selected from the group consisting of: a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, a cell proliferative disorder (*e.g.*, a neoplasm) and any combination thereof. In some cases, the disease or condition can be Alzheimer's disease. In some cases, the disease or condition can be muscular dystrophy. In some cases, the disease or condition can be retinitis pigmentosa. In some cases, the disease or condition can be Parkinson disease. In some cases, the disease or condition can be pain. In some cases, the disease or condition can be Stargardt macular dystrophy. In some cases, the disease or condition can be Charcot-Marie-Tooth disease. In some cases, the disease or condition can be Rett syndrome. In some cases, the administering can be sufficient to decrease expression of a gene relative to prior to the administering. In some cases, the administering can be sufficient to edit at least one point mutation in the subject. In some cases, the administering can be sufficient to edit at least one stop codon in the subject, thereby producing a read-through of the stop codon. In some cases, the administering can be sufficient to produce an exon skip in the subject.

**[0005]** Another aspect of the disclosure provides for a non-naturally occurring RNA. In some cases, the non-naturally occurring RNA can comprise a first domain sequence comprising a two dimensional shape that can be substantially a cruciform and a second domain sequence that has a substantially linear two dimensional structure connected to the first domain sequence, wherein the first domain sequence encodes for an RNA editing entity recruiting domain and the second domain sequence encodes for a targeting domain, wherein the second domain sequence can be complementary to at least a portion of a target RNA. In some cases, the non-naturally occurring RNA can further comprise a third domain sequence attached to the second domain sequence. In some cases, the third domain sequence can comprise an RNA editing entity recruiting domain encoding sequence that forms a secondary structure having a two dimensional shape that can be substantially a cruciform. In some cases, at least one base of the non-naturally occurring RNA can comprise a chemical modification. In some cases, at least one sugar of the non-naturally occurring RNA can comprise a chemical modification. In some cases, a kit can comprise the non-naturally occurring RNA in a container. In some cases, the kit can further comprise a syringe. In some cases, the container can be the syringe. In some cases, an isolated cell can comprise the non-naturally occurring RNA. In some cases, a pharmaceutical composition can comprise the non-naturally occurring RNA in unit dose

form. In some cases, the pharmaceutical composition can further comprise a pharmaceutically acceptable excipient, diluent, or carrier. In some cases, the pharmaceutical composition can comprise a second active ingredient. In some cases, a method of treating a disease or condition in a subject comprising administering to the subject the non-naturally occurring RNA. In some cases, the administering can be by intravenous injection, intramuscular injection, an intrathecal injection, an intraorbital injection, a subcutaneous injection, or any combination thereof. In some cases, the method can further comprise administering a second therapy to the subject. In some cases, the disease or condition can be selected from the group consisting of: a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, a cell proliferative disorder (*e.g.*, a neoplasm) and any combination thereof. In some cases, the disease or condition can be Alzheimer's disease. In some cases, the disease or condition can be muscular dystrophy. In some cases, the disease or condition can be retinitis pigmentosa. In some cases, the disease or condition can be Parkinson disease. In some cases, the disease or condition can be pain. In some cases, the disease or condition can be Stargardt macular dystrophy. In some cases, the disease or condition can be Charcot-Marie-Tooth disease. In some cases, the disease or condition can be Rett syndrome. In some cases, the administering can be sufficient to decrease expression of a gene relative to prior to the administering. In some cases, the administering can be sufficient to edit at least one point mutation in the subject. In some cases, the administering can be sufficient to edit at least one stop codon in the subject, thereby producing a readthrough of the stop codon. In some cases, the administering can be sufficient to produce an exon skip in the subject.

**[0006]** Another aspect of the disclosure provides for a nucleic acid. In some cases, the nucleic acid can comprise an RNA editing entity recruiting domain and an antisense domain sequence, wherein when the nucleic acid can be contacted with an RNA editing entity and a target nucleic acid complementary to at least a portion of the antisense domain, modifies at least one base pair of the target nucleic acid at an efficiency of at least about 4 times greater than a comparable nucleic acid complexed with a Cas13b protein or an active fragment thereof, as determined by Sanger Method sequencing of the target nucleic acid.

**[0007]** Another aspect of the disclosure provides for a nucleic acid. In some cases, the nucleic acid can comprise an RNA editing entity recruiting domain and an antisense domain, wherein the nucleic acid when contacted with an RNA editing entity and a target nucleic acid complementary to at least a portion of the antisense domain, modifies at least one base pair of

the target nucleic acid at an efficiency of at least about 4 times greater than a comparable nucleic acid complexed with a GluR2 domain and the antisense domain, as determined by Sanger Method sequencing of the target nucleic acid. In some cases, the nucleic acid can comprise RNA. In some cases, the target nucleic acid can comprise RNA. In some cases, the RNA can be mRNA. In some cases, the mRNA can encode a protein or a portion thereof. In some cases, a dysfunction of the protein or portion thereof can be implicated in a disease or condition. In some cases, the disease or condition can be selected from the group consisting of: a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, a cell proliferative disorder (*e.g.*, a neoplasm) and any combination thereof. In some cases, the RNA can be small interfering RNA (siRNA). In some cases, the RNA editing entity recruiting domain can comprise at least about 80% identity to a GluR2 domain. In some cases, the RNA editing entity recruiting domain can comprise at least about 80% identity to an Alu domain. In some cases, the RNA editing entity recruiting domain can comprise at least about 80% identity to an APOBEC recruiting domain. In some cases, the RNA editing entity recruiting domain can be configured to recruit an ADAR protein. In some cases, the ADAR protein can be an ADAR1, ADAR2, or ADAR3 protein. In some cases, the ADAR protein can be a human ADAR protein. In some cases, the ADAR protein can be a recombinant ADAR protein. In some cases, the ADAR protein can be a modified ADAR protein. In some cases, the RNA editing entity recruiting domain can be configured to recruit an APOBEC protein. In some cases, the APOBEC protein can be an APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3E, APOBEC3F, APOBEC3G, APOBEC3H, or APOBEC4 protein. In some cases, the ADAR protein can be a human ADAR protein. In some cases, the ADAR protein can be a recombinant ADAR protein. In some cases, the ADAR protein can be a modified ADAR protein. In some cases, the nucleic acid can be chemically synthesized. In some cases, the nucleic acid can be genetically encoded. In some cases, a kit can comprise the nucleic acid in a container. In some cases, the kit can further comprise a syringe. In some cases, the container can be the syringe. In some cases, an isolated cell can comprise the nucleic acid. In some cases, a pharmaceutical composition can comprise the nucleic acid in unit dose form. In some cases, the pharmaceutical composition can further comprise a pharmaceutically acceptable excipient, diluent, or carrier. In some cases, the pharmaceutical composition can comprise a second active ingredient. In some cases, a method of treating a disease or condition in a subject comprising administering to the subject the nucleic acid. In some cases, the administering can

be by intravenous injection, intramuscular injection, an intrathecal injection, an intraorbital injection, a subcutaneous injection, or any combination thereof. In some cases, the method can further comprise administering a second therapy to the subject. In some cases, the disease or condition can be selected from the group consisting of: a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, a cell proliferative disorder (*e.g.*, a neoplasm) and any combination thereof. In some cases, the disease or condition can be Alzheimer's disease. In some cases, the disease or condition can be muscular dystrophy. In some cases, the disease or condition can be retinitis pigmentosa. In some cases, the disease or condition can be Parkinson disease. In some cases, the disease or condition can be pain. In some cases, the disease or condition can be Stargardt macular dystrophy. In some cases, the disease or condition can be Charcot-Marie-Tooth disease. In some cases, the disease or condition can be Rett syndrome. In some cases, the administering can be sufficient to decrease expression of a gene relative to prior to the administering. In some cases, the administering can be sufficient to edit at least one point mutation in the subject. In some cases, the administering can be sufficient to edit at least one stop codon in the subject, thereby producing a read-through of the stop codon. In some cases, the administering can be sufficient to produce an exon skip in the subject.

**[0008]** Another aspect of the disclosure can provide for a nucleic acid. In some cases, the nucleic acid can comprise sequences comprising an antisense domain, a first stem-loop forming sequence, and a second stem-loop forming sequence, wherein the nucleic acid when contacted with (a) a first polypeptide comprising a first portion of an RNA editing entity and a first polynucleotide binding domain configured to bind to the first stem-loop forming sequence, and (b) a second polypeptide comprising a second portion of an RNA editing entity and a second polynucleotide binding domain configured to bind to the second stem-loop forming sequence, and (c) a target nucleic acid complementary to at least a portion of the antisense domain, modifies at least one base pair of the target nucleic acid. In some cases, the first stem-loop or the second stem-loop can be an MS2 stem loop. In some cases, the first stem loop or the second stem-loop can be a BoxB stem-loop. In some cases, the first stem-loop or the second stem-loop can be a U1A stem-loop. In some cases, the first portion of the RNA editing entity or the second portion of the RNA editing entity can comprise an N-terminal fragment of an ADAR deaminase domain encoding sequence. In some cases, the first portion of the RNA editing entity or the second portion of the RNA editing entity can comprise a C-terminal fragment of an ADAR deaminase domain encoding sequence. In

some cases, the first polynucleotide binding domain or the second polynucleotide binding domain can comprise an MS2 coat protein. In some cases, the first polynucleotide binding domain or the second polynucleotide binding domain can comprise a Lambda N peptide. In some cases, the first polynucleotide binding domain or the second polynucleotide binding domain can comprise a human nucleic acid binding protein. In some cases, the human nucleic acid binding protein can be a U1A protein, a TBP6.7 protein, a human histone stem-loop binding protein, or a DNA binding domain of a glucocorticoid receptor. In some cases, the RNA editing entity can be capable of performing an adenosine to inosine mutation on the target nucleic acid. In some cases, the RNA editing entity can be capable of performing a cytosine to thymine mutation on the target nucleic acid. In some cases, a kit can comprise the nucleic acid in a container. In some cases, the kit can further comprise a syringe. In some cases, the container can be the syringe. In some cases, an isolated cell can comprise the nucleic acid. In some cases, a pharmaceutical composition can comprise the nucleic acid in unit dose form. In some cases, the pharmaceutical composition can further comprise a pharmaceutically acceptable excipient, diluent, or carrier. In some cases, the pharmaceutical composition can comprise a second active ingredient. In some cases, a method of treating a disease or condition in a subject comprising administering to the subject the nucleic acid. In some cases, the administering can be by intravenous injection, intramuscular injection, an intrathecal injection, an intraorbital injection, a subcutaneous injection, or any combination thereof. In some cases, the method can further comprise administering a second therapy to the subject. In some cases, the disease or condition can be selected from the group consisting of: a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, a cell proliferative disorder (*e.g.*, a neoplasm) and any combination thereof. In some cases, the disease or condition can be Alzheimer's disease. In some cases, the disease or condition can be muscular dystrophy. In some cases, the disease or condition can be retinitis pigmentosa. In some cases, the disease or condition can be Parkinson disease. In some cases, the disease or condition can be pain. In some cases, the disease or condition can be Stargardt macular dystrophy. In some cases, the disease or condition can be Charcot-Marie-Tooth disease. In some cases, the disease or condition can be Rett syndrome. In some cases, the administering can be sufficient to decrease expression of a gene relative to prior to the administering. In some cases, the administering can be sufficient to edit at least one point mutation in the subject. In some cases, the administering can be sufficient to edit at least one



stop codon in the subject, thereby producing a read-through of the stop codon. In some cases, the administering can be sufficient to produce an exon skip in the subject.

**[0009]** Another aspect of the disclosure provides a method of treating muscular dystrophy in a subject. In some cases, the method can comprise: administering to the subject a pharmaceutical composition comprising an adeno-associated virus (AAV) vector that comprises a first nucleic acid encoding a second nucleic acid, wherein the second nucleic acid comprises (a) an antisense region that can be at least partially complementary to an RNA sequence implicated in muscular dystrophy, and (b) at least one RNA editing entity recruiting domain, wherein the at least one RNA editing entity recruiting domain does not comprise a stem-loop, or wherein the at least one RNA editing entity recruiting domain comprises at least about 80% sequence identity to at least one of: an Alu domain, an Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC) recruiting domain, and any combination thereof. In some cases, the pharmaceutical composition can be in unit dose form. In some cases, the administering can be at least once a week. In some cases, the administering can be at least once a month. In some cases, the administering can be by injection. In some cases, the injection can be subcutaneous, intravenous, infusion, intramuscular, intrathecal, or intraperitoneal injection. In some cases, the administering can be transdermal, transmucosal, oral, or pulmonary. In some cases, the method can further comprise administering a second therapy to the subject.

**[0010]** Another aspect of the disclosure can provide a method of making a vector. In some cases, the method can comprise: cloning at least one copy of a nucleic acid into the vector, wherein the nucleic acid encodes for at least one RNA editing entity recruiting domain, and wherein a sequence encoding the at least one RNA editing entity recruiting domain does not form a secondary structure that comprises a stem-loop, or wherein the nucleic acid that encodes the at least one RNA editing entity recruiting domain comprises at least about 80% sequence identity to a sequence selected from: an Alu domain encoding sequence, an Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC) recruiting domain encoding sequence, and any combination thereof. In some cases, the vector can be a viral vector. In some cases, the viral vector can be an AAV vector. In some cases, the viral vector can comprise a modified VP1 protein. In some cases, the vector can be a liposome. In some cases, the vector can be a nanoparticle. In some cases, the method further comprises transfecting or transducing the vector into an isolated human cell.

[0011] Aspects of this disclosure relate to an engineered ADAR1 or ADAR2 guide RNA (“adRNA”) comprising, or alternatively consisting essentially of, or yet further consisting of, a sequence complementary to a target RNA, that optionally comprises, or consists essentially of, or yet further consists of the engineered ADAR2 and an ADAR2 recruiting domain derived from GluR2 mRNA. In some embodiments, the sequence complementary to the target RNA comprises, or consists essentially of, or yet further consists of, between about 15 to 200 or alternatively from 20 to 100 base pairs. In one aspect, the engineered adRNA comprises, or consists essentially of, or yet further consists of no ADAR recruiting domains. In some embodiments, the ADAR recruiting domains comprise, or alternatively consist essentially of, or yet further consist of GluR2 mRNA, Alu repeat elements or other RNA motifs to which ADAR binds. In one aspect, the engineered adRNA comprises, or consists essentially of, or yet further consists of between about 1 to 10 ADAR recruiting domains. In another aspect, the ADAR2 recruiting domain can be derived from GluR2 mRNA and can be located at the 5’ end or the 3’ end of the engineered adRNA. In still further embodiments, the engineered adRNA comprises, or consists essentially of, or yet further consists of, the GluR2 mRNA at both the 5’ end and the 3’ end. In a further aspect, the engineered adRNA of this disclosure further comprises, or alternatively consists essentially of, or yet further consists of two MS2 hairpins flanking the sequence complementary to a target RNA.

[0012] In some embodiments, the target RNA can be ornithine transcarbamylase. Also provided herein is a complex comprising, or alternatively consisting essentially of, or yet further consisting of, an AdRNA as disclosed herein hybridized to a complementary polynucleotide under conditions of high stringency. In one aspect, the polynucleotide can be DNA. In another aspect, the polynucleotide can be RNA.

[0013] In one aspect, the engineered adRNA of this disclosure, further comprises, or alternatively consists essentially of, or yet further consists of an editing inducer element.

[0014] Further aspects relate to an engineered ADAR2 guide RNA (“adRNA”) encoded by a sequence selected from the group of sequences provided in **TABLE 1** or **FIG. 2**. The adRNAs can be combined with a carrier, such as a pharmaceutically acceptable carrier, examples of such are provided herein.

[0015] Also disclosed herein is an engineered adRNA-sRNA (small nuclear RNA) fusion. In one aspect, the engineered adRNA further comprises, or alternatively consists essentially of, or yet further consists of an N-terminal mitochondrial targeting sequence (MTS) to facilitate localization of the engineered adRNA to the mitochondria. In another aspect,

provided herein can be an engineered further comprising, or alternatively consisting essentially of, or yet further consisting of a cis-acting zipcode to facilitate localization of the engineered adRNA into peroxisomes, endosomes and exosomes.

**[0016]** Further provided herein is small molecule regulatable engineered adRNA. In one aspect, disclosed herein are engineered adRNA-aptamer fusions. Non-limiting examples of aptamers that can be used for this purpose include aptamers that bind flavin mononucleotide, guanine, other natural metabolites, or sugars. Also disclosed herein is a U1A-ADAR fusion, entirely of human origin.

**[0017]** Also disclosed herein is a complex comprising, or alternatively consisting essentially of, or yet further consisting of an engineered adRNA of this disclosure hybridized to a complementary polynucleotide under conditions of high stringency.

**[0018]** Also provided herein is a vector comprising, or alternatively consisting essentially of, or yet further consisting of one or more of the isolated polynucleotide sequence encoding the engineered adRNA of this disclosure and optionally regulatory sequences operatively linked to the isolated polynucleotide. Non-limiting examples of a vector include a plasmid or a viral vector such as a retroviral vector, a lentiviral vector, an adenoviral vector, or an adeno-associated viral vector.

**[0019]** Further disclosed herein is a recombinant cell further comprising or alternatively consisting essentially of, or yet further consisting of the vector described above, wherein the engineered adRNA can be recombinantly expressed.

**[0020]** Compositions comprising one or more of the above-noted compounds and a carrier are provided. In one embodiment, the composition can be a pharmaceutical composition and therefore further comprises at least a pharmaceutically acceptable carrier or a pharmaceutically acceptable excipient. The compositions are formulated for various delivery modes, *e.g.*, systemic (oral) or local.

**[0021]** Also provided herein is a method of modifying protein expression comprising, or alternatively consisting essentially of, or yet further consisting of contacting a polynucleotide encoding the protein, the expression of which is to be modified, with the engineered adRNA of this disclosure.

**[0022]** Still further aspects relate to methods of treating a disease or disorder associated with aberrant protein expression comprising, or alternatively consisting essentially of, or yet further consisting of, administering an effective amount of any one or more of the engineered adRNA disclosed herein and/or uses for an effective amount of any one or more of the

engineered adRNA disclosed herein to a subject in need thereof and for treating a disease or disorder associated with aberrant protein expression. In one particular aspect, provided herein is a method of treating Duchenne Muscular Dystrophy comprising, or alternatively consisting essentially of, or yet further consisting of administering to a subject in need of such treatment an effective amount of one or more of the engineered adRNA of this disclosure.

**[0023]** The disclosure demonstrates validation of this approach *in vivo* in the spf-ash mouse model of ornithine transcarbamylase deficiency. This model bears a G->A point mutation in the last nucleotide of exon 4. Upon delivery of only adRNA via AAVs, up to 1% correction of the point mutation in the absence of the overexpression of the ADAR enzymes was observed.

**[0024]** Additional aspects relate to the same or similar structures comprising, or alternatively consisting essentially of, or yet further consisting of, DNA or a combination of DNA and RNA. Further aspects relate to kits comprising any one or more of the embodiments above and instructions for use *in vitro* and/or *in vivo*.

**[0025]** Aspects of the disclosure can relate to ADAR and APOBEC systems for gene editing. Some aspects relate to an ADAR system for exon skipping comprising an adRNA targeting a splice acceptor and/or a branch point in an intron and, optionally, an ADAR enzyme. In some embodiments, the ADAR enzyme can be ADAR1, ADAR2, or a mutant or variant each thereof. In some embodiments, the mutant or variant can be selected from ADAR1 (E1008Q) and ADAR2 (E488Q). In some embodiments, the intron can be comprised in a gene selected from dystrophin, SCN9A, or ornithine transcarbamylase. In some cases, the adRNA can be selected from **SEQUENCE SET 1**. Further aspects can relate to a method of treating a disease, disorder, or condition characterized by aberrant gene expression comprising, administering the disclosed ADAR system. In some embodiments, the disease, disorder, or condition can be selected from Duchenne muscular dystrophy or ornithine transcarbamylase deficiency. In some embodiments, the disease, disorder, or condition can be associated with pain.

**[0026]** Additional aspects relate to an APOBEC system for cytosine to thymine editing comprising a pair of gRNA that create alipoprotein B mRNA like structure and, optionally, an APOBEC enzyme. In some embodiments, the pair of gRNA can be the pair of sequences provided in **SEQUENCE SET 2**.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0027] **FIG. 1** is a schematic showing endogenous recruitment of ADARs.

[0028] **FIG. 2** provides a listing of stabilized scaffolds being evaluated for (1) improved efficiency and (2) ability to recruit ADAR1s (SEQ ID NOs:2-9).

[0029] **FIG. 3** shows the results of *in vitro* and *in vivo* screening of exemplary adRNAs.

[0030] **FIG. 4** is a schematic of how MCP-APOBEC fusions or MCP-ACF-APOBEC fusions are recruited via MS2-RNAs bearing two MS2 stem loops. As shown in the Figure, the target cytosine can be kept single stranded so as to be accessible for APOBEC mediated editing via creation of a bulge. A bulge was created by using either the exact target sequence (20-30 base pairs) or the sequence ACATATATGATACAATTTGATCAGTATATT (SEQ ID NO:175) in the MS2-RNA between the MS2 stem loops (blue) along with complementary sequences (20-30 base pairs) on either side of the MS2 stem loops (green) that bind to the mRNA of interest. The sequence ACATATATGATACAATTTGATCAGTATATT (SEQ ID NO:176) is taken from the naturally occurring apoB substrate that the APOBEC edits. The MS2 stem loop sequence used in the designs is aACATGAGGATCACCCATGTc (SEQ ID NO:177).

[0031] **FIG. 5** shows percent mRNA editing by a construct comprising an Alu domain in comparison to editing by a construct comprising a GluR2 domain.

[0032] **FIG. 6** shows an example of a construct (SEQ ID NO:10) comprising an Alu domain that was used to generate the data of **FIG. 5**.

[0033] **FIG. 7** shows an example of a construct comprising two cruciforms linked by an antisense domain and comprising an Alu domain. Both exemplary structure (SEQ ID NO:11) and exemplary sequence (SEQ ID NO:1) are provided.

[0034] **FIG. 8** shows various construct designs (including those that comprise a GluR2 domain) (SEQ ID NOs: 15, 252, and 16-34) and percent mRNA editing of each.

[0035] **FIG. 9** shows mapping enzyme targeting preferences for both ADAR1 and ADAR2, for each base (*e.g.*, A, C, G, and T).

[0036] **FIG. 10** shows engineering next-gen adRNAs with enhanced ADAR1 and ADAR2 recruitment potential. The first series of columns show relative activity employing no adRNA. The second series of columns show relative activity employing a construct comprising a GluR2 domain. The third series of columns show relative activity employing a construct comprising an Alu domain associated with two cruciform structures. Figure 10 discloses SEQ ID NO: 11.

[0037] **FIG. 11A-D** shows different construct designs. **FIG. 11A** exemplifies an antisense domain linked to two GluR2 domains. **FIG. 11B** exemplifies an anti-sense domain only. **FIG. 11C** exemplifies an antisense domain linked to two cruciforms. **FIG. 11D** exemplifies a toe-hold

[0038] **FIG. 12** shows percent editing yield between different construct designs.

[0039] **FIG. 13** shows a comparison of a short antisense oligonucleotide (AON) with mismatched bulges as compared to a longer construct comprising a hairpin structure.

[0040] **FIG. 14** shows exemplary adRNA designs (SEQ ID NOs:250 and 251).

[0041] **FIG. 15** shows exemplary adRNA structures (SEQ ID NOs: 35-37) having parameters (d, l, m), wherein d = number of GluR2 domains, l = length of antisense domain, m = position of mismatch.

[0042] **FIG. 16** shows a schematic of RNA editing via recruitment of endogenous ADARs in the presence of adRNA.

[0043] **FIG. 17** shows a U6 promoter transcribed adRNAs with progressively longer antisense domain lengths, in combination with zero, one or two GluR2 domains that were evaluated for their ability to induce targeted RNA editing with or without exogenous ADAR2 expression. Values represent mean  $\pm$  SEM (n=3). Long adRNA can recruit endogenous ADARs for RNA editing.

[0044] **FIG. 18** shows chemically synthesized adRNAs versions tested against a panel of mRNAs with or without exogenous ADAR2 expression. Chemical modifications are identified along with the source of adRNA. Values represent mean  $\pm$  SEM (n=3).

[0045] **FIG. 19** shows *in vivo* RNA correction efficiencies in correctly spliced OTC mRNA in the livers of treated adult *spf<sup>ash</sup>* mice (reto-orbital injections). RNA editing levels of 0.6% are seen in mice injected with U6 transcribed short adRNA.

[0046] **FIG. 20** shows a design of an Alu adRNA. Left: a structure of an Alu element. Middle: a design as described herein that comprises a locus-specific antisense sequence with a C mismatch opposite a target A. Right: Recruitment of an RNA editing enzyme ADAR to the target.

[0047] **FIG. 21** shows exemplary Alu guide sequences (SEQ ID NOs:38-41).

[0048] **FIG. 22** shows a schematic of the split-ADAR2 DD system.

[0049] **FIG. 23** shows an exemplary sequence (SEQ ID NO:42) of the split-ADAR2 DD with potential sites for splitting highlighted.

[0050] **FIG. 24** shows pairs of fragments 1-16 assayed via a cypridina luciferase reporter (Cluc W85X).

[0051] **FIG. 25** shows fragments 9 and 10 assayed against a Cluc reporter.

[0052] **FIG. 26** shows exemplary sequences (SEQ ID NOs:43-55).

[0053] **FIG. 27** shows a schematic of ADAR recruitment via U1A (SEQ ID NO:56).

[0054] **FIG. 28** shows exemplary sequences (SEQ ID NOs:57-68) for several fusion constructs or one or more APOBEC family members (SEQ ID NOs:57-68).

[0055] **FIG. 29** shows exemplary sequences (SEQ ID NOs:69-73) of engineered apRNAs configured to recruit APOBEC3A.

[0056] **FIG. 30** shows exemplary sequences (SEQ ID NOs:74-78) of engineered MS2-apRNAs configured to recruit MCP-APOBEC3A.

[0057] **FIG. 31** shows two different scenarios of no ADAR recruitment and ADAR recruitment to permit ribosomal read-through that results in normal luciferase expression.

[0058] **FIG. 32A-C** shows engineering programmable RNA editing and characterizing specificity profiles: **(FIG. 32A)** Schematics of RNA editing via constructs utilizing the full length ADAR2 and an engineered adRNA derived from the GluR2 transcript, or MS2 Coat Protein (MCP) fusions to the ADAR1/2 deaminase domains and the corresponding MS2 hairpin bearing adRNA. **(FIG. 32B)** Comparison of RNA editing efficiency of the endogenous RAB7A transcript by different RNA editing constructs quantified by Sanger sequencing (efficiency calculated as a ratio of Sanger peak heights G/(A+G)). Experiments were carried out in HEK 293T cells. Values represent mean +/- SEM (n=3). **(FIG. 32C)** Violin plots representing distributions of A->G editing yields observed at reference sites where at least one treatment sample was found to have a significant change (Fisher's exact test, FDR = 1%) in editing yield relative to the control sample. Blue circles indicate editing yields at the target A-site within the RAB7A transcript. Black dots represent median off-target editing yields. To better visualize the shapes of the distributions, their maximum extent along the y-axis was equalized across all plots, and were truncated at 60% yield..

[0059] **FIG. 33A-E** shows *in vivo* RNA editing in mouse models of human disease. **(FIG. 33A)** Schematic of the DNA and RNA targeting approaches to restore dystrophin expression in the *mdx* mouse model of Duchenne Muscular Dystrophy: (i) a dual gRNA-CRISPR based approach leading to in frame excision of exon 23 and (ii) ADAR2 and MCP-ADAR1 based editing of the ochre codon. **(FIG. 33B)** Immunofluorescence staining for dystrophin in the TA muscle shows partial restoration of expression in treated samples (intra-muscular

injections of AAV8-ADAR2, AAV8-ADAR2 (E488Q), and AAV8-CRISPR). Partial restoration of nNOS localization is also seen in treated samples (scale bar: 250 $\mu$ m). (FIG. 33C) In vivo TAA->TGG/TAG/TGA RNA editing efficiencies in corresponding treated adult mdx mice. Values represent mean $\pm$  SEM (n=4, 3, 7, 3, 3, 10, 3, 4 independent TA muscles respectively). (FIG. 33D) Schematic of the OTC locus in the spfsh mouse model of Ornithine Transcarbamylase deficiency which have a G->A point mutation at a donor splice site in the last nucleotide of exon 4, and approach for correction of mutant OTC mRNA via ADAR2 mediated RNA editing. (FIG. 33E) In vivo RNA correction efficiencies in the correctly spliced OTC mRNA in the livers of treated adult spfsh mice (retro-orbital injections of AAV8-ADAR2 and AAV8-ADAR2 (E488Q)). Values represent mean $\pm$  SEM (n=4, 4, 3, 3, 4, 5 independent animals respectively).

**[0060]** FIG. 34A-C shows antisense domain engineering. (FIG. 34A) Optimization of adRNA antisense region using adRNA scaffold 2: length and distance from the ADAR2 recruiting region were systematically varied. Values represent mean $\pm$  SEM (n=3)(SEQ ID NOs 253 and 79-102). (FIG. 34B) U6 promoter transcribed adRNAs with progressively longer antisense domain lengths, in combination with zero, one or two GluR2 domains were evaluated for their ability to induce targeted RNA editing with or without exogenous ADAR2 expression. Values represent mean $\pm$  SEM (n=3). All the above experiments were carried out in HEK 293T cells. (FIG. 34C) Experimental confirmation of expression of endogenous ADAR1 and ADAR2 (relative to GAPDH) in HEK 293T and HeLa cell lines. Observed levels were similar to those documented in The Human Protein Atlas (see world wide web (www) at [proteatlas.org](http://proteatlas.org)).

**[0061]** FIG. 35A-B shows engineering MS2 adRNAs. (FIG. 35A) Systematic evaluation of antisense RNA targeting domain of the MS2 adRNA (SEQ ID NO:103-110). Values represent mean $\pm$  SEM (n=3). (FIG. 35B) On-target RNA editing by MCP-ADAR2 DD-NLS requires co-expression of the MS2 adRNA. Values represent mean $\pm$  SEM (n=3). All experiments were carried out in HEK 293T cells.



**[0062]** FIG. 36A-C shows analysis of RNA editing yields across a panel of targets. (FIG. 36A) Comparison of RNA editing efficiency of the OTC reporter transcript by GluR2 adRNA and MS2 adRNA guided RNA editing constructs as well as the Cas13b based REPAIR construct. Values represent mean $\pm$  SEM (n=6 for reporter and Cas 13b based constructs, n=3 for all other constructs). (FIG. 36B) Chemically synthesized adRNAs versions were tested against a panel of mRNAs with or without exogenous ADAR2

expression. The exact chemical modifications are stated in the figure along with the source of adRNA. Values represent mean $\pm$  SEM (n=3). (FIG. 36C) Analysis of RNA editing yields across a spectrum of endogenous targets chosen to cover a range of expression levels. U6 transcribed long adRNAs with none or two GluR2 domains were also evaluated against multiple endogenous mRNA targets with or without exogenous ADAR2 expression. Editing is observed at all tested loci even in the absence of exogenous ADAR2 expression. Values represent mean $\pm$  SEM (n=3). All experiments were carried out in HEK 293T cells.

**[0063]** FIG. 37A-D shows ADAR2 variants and their impact on editing and specificity. (FIG. 37A) Comparison of on target RNA editing and editing in flanking adenosines of the RAB7 A transcript by GluR2 adRNA and MS2 adRNA guided RNA editing constructs as well as the Cas 13b based REP AIR construct. Mean (n=3) editing yields are depicted (SEQ ID NO: 111). All experiments were carried out in HEK 293T cells and editing efficiency was calculated as a ratio of Sanger peak heights G/(A+G). (FIG. 37B) ADAR2 (E488Q) exhibits higher efficiency than the ADAR2 in the in vitro editing of the spfash OTC reporter transcript (p=0.037, unpaired t-test, two-tailed); values represent mean $\pm$  SEM (n=3), and (FIG. 37C) mdx DMD reporter transcript (p=0.048, p=0.012 respectively, unpaired t-test, two-tailed); values represent mean $\pm$  SEM (n=3). (FIG. 37D) Comparison of the editing efficiency and specificity profiles of the ADAR2, ADAR2 (E488Q) and the ADAR2 (iil-138) for the OTC reporter transcript (upper panel) and endogenous RAB7 A transcript (lower panel). Heatmap indicates the A->G edits in the vicinity of the target (arrow). Values represent mean $\pm$  SEM (n=3). All experiments were carried out in HEK 293T cells and editing efficiency was calculated as a ratio of Sanger peak heights G/(A+G). Figure 37D discloses SEQ ID NOs 254-255, respectively, in order of appearance.

**[0064]** FIG. 38 shows transcriptome scale specificity profiles of RNA editing approaches (Cas13b-ADAR REPAIR $\pm$  gRNA).

**[0065]** FIG. 39 shows transcriptome scale specificity profiles of RNA editing approaches (ADAR2  $\pm$  adRNA). The version used for these studies is GluR2 adRNA (1,20,6).

**[0066]** FIG. 40 shows transcriptome scale specificity profiles of RNA editing approaches (MCP-ADAR1 DD $\pm$  adRNA).

**[0067]** FIG. 41 shows transcriptome scale specificity profiles of RNA editing approaches (MCP-ADAR2 DD $\pm$  adRNA).

[0068] FIG. 42A-B shows variation of transcriptome scale editing specificity with construct features. (FIG. 42A) Each point in the box plots corresponds to the fraction of edited sites for one of the MCP-ADAR constructs listed in FIG. 32. The fraction of edited sites for each

construct was calculated by dividing the number of reference sites with significant changes in A-to-G editing yield (see Table 3) by the total number 8,729,464 of reference sites considered. Construct features indicated on the horizontal axes were compared using the Mann-Whitney U test, yielding p-values of 0.16 for NLS vs. NES, 0.0070 for ADAR1 vs. ADAR2, 0.72 for “- adRNA” vs. “+ adRNA”, and 0.038 for “ADAR WT” vs. “ADAR E>Q” (n=8 for all conditions). **(FIG. 42B)** 2D histograms comparing the transcriptome-wide A->G editing yields observed with each construct (y-axis) to the yields observed with the control sample (x-axis). Inset shows violin plots representing distributions of A->G editing yields observed at reference sites where at least one treatment sample was found to have a significant change (Fisher’s exact test, FDR = 1%) in editing yield relative to the control sample. Blue circles indicate editing yields at the target A-site within the RAB7A transcript. To better visualize the shapes of the distributions, their maximum extent along the y-axis was equalized across all plots, and were truncated at 60% yield. Samples here correspond to 293Ts transfected with long antisense domain bearing adRNAs that can enable RNA editing via exogenous and/or endogenous ADAR recruitment.

**[0069] FIG. 43A-E** shows optimization and evaluation of dystrophin editing experiments *in vitro* and *in vivo* in *mdx* mice. **(FIG. 43A)** Schematic of RNA editing utilizing the full length ADAR2 along with an engineered adRNA or a reverse oriented adRNA (radRNA); (ii) RNA editing efficiencies of amber and ochre stop codons, in one-step and two-steps. Experiments were carried out in HEK 293T cells. Values represent mean +/- SEM (n=3). **(FIG. 43B)** RNA editing of ochre codons requires two cytosine mismatches in the antisense RNA targeting domains of adRNA or radRNA (SEQ ID NOs:112-116) to restore GFP expression. Experiments were carried out in HEK 293T cells. Values represent mean +/- SEM (n=3). **(FIG. 43C)** Schematic of the AAV vectors utilized for *in vivo* delivery of adRNA and ADAR2, and *in vitro* optimization of RNA editing of amber and ochre stop codons in the presence of one or two copies of the adRNA, delivered via an AAV vector (p=0.0003, p=0.0001, p=0.0015 respectively, unpaired t-test, two-tailed). Experiments were carried out in HEK 293T cells. Values represent mean +/- SEM (n=3 for reporters, n=6 for all other conditions). **(FIG. 43D)** Representative Sanger sequencing plot showing editing of the ochre stop codon (TAA->TGG) in the *mdx* DMD reporter transcript (quantified by NGS)(SEQ ID NO:117-118). Experiments were carried out in HEK 293T cells (n=3). **(FIG. 43E)** Representative example of *in vivo* RNA editing analyses of treated *mdx* mice (quantified using NGS) (SEQ ID NOs:119-130).

[0070] **FIG. 44A-C** shows immunofluorescence and western blot analyses of *in vivo* dystrophin RNA editing experiments in *mdx* mice. (**FIG. 44A**) Immunofluorescence staining for dystrophin in the TA muscle shows partial restoration of expression in treated samples (intra-muscular injections of AAV8-ADAR2, AAV8-ADAR2 (E488Q), AAV8-MCP-ADAR1 (E1008Q) NLS). Partial restoration of nNOS is localization also seen in treated samples (scale bar: 250µm). (**FIG. 44B**) Western blots showing partial recovery of dystrophin expression (1-2.5%) in TA muscles of *mdx* mice injected with both components of the editing machinery, the enzyme and adRNA, and stable ADAR2 expression in injected TA muscles up to 8 weeks post injections. (**FIG. 44C**) Western blot showing partial restoration of dystrophin expression (10%) using AAV8-CRISPR.

[0071] **FIG. 45A-E** shows optimization and evaluation of OTC RNA editing experiments *in vitro* and *in vivo* in *spf<sup>ash</sup>* mice. (**FIG. 45A**) Representative Sanger sequencing plot showing correction of the point mutation in the *spfash* OTC reporter transcript (quantified using NGS) (SEQ ID NO:131-132). Experiments were carried out in HEK 293T cells (n=3). (**FIG. 45B**) Representative example of *in vivo* RNA editing analyses of treated *spfash* mice showing correction of the point mutation in the correctly spliced OTC mRNA (quantified using NGS)(SEQ ID NO:133-139). (**FIG. 45C**) *In vivo* RNA correction efficiencies in the OTC pre-mRNA in the livers of treated adult *spfash* mice (retro-orbital injections of AAV8-ADAR2 and AAV8-ADAR2 (E488Q). Values represent mean +/- SEM (n=4, 4, 3, 3, 4, 5 independent animals respectively). (**FIG. 45D**) PCR products showing the correctly and incorrectly spliced OTC mRNA. The incorrectly spliced mRNA is elongated by 48 base pairs. Fraction of incorrectly spliced mRNA is reduced in mice treated with adRNA+ADAR2 (E488Q). (**FIG. 45E**) Western blot for OTC shows partial restoration (2.5%-5%) of expression in treated adult *spfash* mice and stable ADAR2 (E488Q) expression three weeks post injections.

[0072] **FIG. 46** shows toxicity analyses of *in vivo* RNA editing experiments.

[0073] **FIG. 47** is a schematic showing exon skipping via creation of a splice acceptor and/or branch point mutation.

[0074] **FIG. 48** is a schematic of C → T editing via APOBECs.

[0075] **FIG. 49A-D** shows schematics of editing DNA and both strands of DNA/RNA hybrids.

[0076] **FIG. 50A-D** shows the results of a study in a model for ornithine transcarbamylase deficiency. **FIG. 50A** depicts *in vivo* RNA correction efficiencies in the livers of treated

adult  $spf^{ash}$  mice (retro-orbital injections of AAV8-ADAR2 and AAV8-ADAR2 (E488Q)). Each data point represents an independent animal. Editing efficiencies measured in the spliced OTC mRNA. Error bars represent  $\pm$  SEM. **FIG. 50B** depicts *in vivo* RNA correction efficiencies in the OTC pre-mRNA in the livers of treated adult  $spf^{ash}$  mice (retro-orbital injections of AAV8-ADAR2 and AAV8-ADAR2 (E488Q)). Each data point represents an independent animal. **FIG. 50C** shows PCR products showing the correctly and incorrectly spliced OTC mRNA. The incorrectly spliced mRNA is elongated by 48 base pairs. The fraction of incorrectly spliced mRNA is reduced in mice treated with adRNA+ADAR2 (E488Q). **Fig. 50D** is a Western blot for *OTC* shows partial restoration (2.5%-5%) of expression in treated adult  $spf^{ash}$  mice.

[0077] **FIG. 51A-B** shows the results of a study in a model of Duchenne muscular dystrophy. **FIG. 51A** depicts *in vivo* TAA->TGG/TAG/TGA RNA editing efficiencies in corresponding treated adult *mdx* mice. Each data point represents an independent TA muscle. Error bars represent  $\pm$  SEM. **FIG. 51B** is a Western blot for dystrophin shows partial restoration (1-2.5%) of expression in corresponding treated adult *mdx* mice.

[0078] **FIG. 52** provides further information about potential branch point locations.

#### DETAILED DESCRIPTION

[0079] As aspect of the disclosure provides for nucleic acids, non-naturally occurring RNAs, vectors comprising nucleic acids, compositions, and pharmaceutical compositions for RNA editing. Any of the above or as described herein can be configured for an A (adenosine) to I (inosine) edit, a C (cytosine) to T (thymine) edit, or a combination thereof. In some cases, an A to I edit can be interpreted or read as a C to U mutation. In some cases, an A to I edit can be interpreted or read as an A to G mutation. Nucleic acids, non-naturally occurring RNAs, vectors comprising nucleic acids, compositions, and pharmaceutical compositions as described herein can provide enhanced editing efficiencies as compared to native systems, reduced off-target editing, enhanced stability or *in vivo* half-lives, or any combination thereof.

[0080] An aspect of the disclosure provides for a vector. The vector can comprise a nucleic acid with a polynucleotide sequence encoding (i) an RNA editing entity recruiting domain, (ii) a targeting domain complementary to at least a portion of a target RNA, (iii) more than one of either domain, or (iv) any combination thereof. In some cases, the vector can be administered to a subject, such as a subject in need thereof. In some cases, the vector can be

administered as part of a pharmaceutical composition to a subject, such as a subject in need thereof.

**[0081]** An aspect of the disclosure provides for a non-naturally occurring RNA. The non-naturally occurring RNA can comprise (i) an RNA editing entity recruiting domain, (ii) a targeting domain complementary to at least a portion of a target RNA, (iii) more than one of either domain, or (iv) any combination thereof. In some cases, the non-naturally occurring RNA can be administered to a subject, such as a subject in need thereof. In some cases, the non-naturally occurring RNA can be administered as part of a pharmaceutical composition to a subject, such as a subject in need thereof. In some cases, the non-naturally occurring RNA can be formulated in a vector for administration. The vector can comprise a viral vector, a liposome, a nanoparticle, or any combination thereof. In some cases, the non-naturally occurring RNA can comprise at least one base, at least one sugar, more than one of either, or a combination thereof having a modification, such as a chemical modification.

**[0082]** An aspect of the disclosure provides for a nucleic acid. The nucleic acid can comprise (i) an RNA editing entity recruiting domain, (ii) a targeting domain complementary to at least a portion of a target RNA, (iii) more than one of either domain, or (iv) any combination thereof. In some cases, the nucleic acid can be administered to a subject, such as a subject in need thereof. In some cases, the nucleic acid can be administered as part of a pharmaceutical composition to a subject, such as a subject in need thereof. In some cases, the nucleic acid can be formulated in a vector for administration. The vector can comprise a viral vector, a liposome, a nanoparticle, or any combination thereof. The nucleic acid can be genetically encoded. The nucleic acid can be chemically synthesized.

**[0083]** A nucleic acid can comprise one or more domains, such as 1, 2, 3, 4, 5 or more domains. In some cases, a nucleic acid can comprise a recruiting domain, a targeting domain, more than one of either, or a combination thereof. In some cases, a nucleic acid can comprise a targeting domain and a recruiting domain. In some cases, a nucleic acid can comprise a targeting domain and two recruiting domains.

**[0084]** A domain can form a two dimensional shape or secondary structure. For example, a targeting domain, a recruiting domain or a combination thereof can form a secondary structure that can comprise a linear region, a cruciform or portion thereof, a toe hold, a stem loop, or any combination thereof. The domain itself can form a substantially linear two dimensional structure. The domain can form a secondary structure that can comprise a

cruciform. The domain can form a secondary structure that can comprise a stem loop. The domain can form a secondary structure that can comprise a toehold.

**[0085]** In some cases, a targeting domain can be positioned adjacent to a recruiting domain, including immediately adjacent or adjacent to but separated by a number of nucleotides. In some cases, a targeting domain can be flanked by two recruiting domains. In some cases, two or more recruiting domains can be adjacent one another.

**[0086]** An aspect of the disclosure includes reducing off target editing. One approach as described herein includes restricting catalytic activity of an ADAR or APOBEC by a split reassembly approach. In such a design, a first domain (such as a recruiting domain) can be catalytically inactive by itself and a second domain can be catalytically inactive by itself but when brought together in a reassembly the two domains together provide catalytic activity to recruit an ADAR or APOBEC. A nucleic acid comprising two domains can be split at any number of locations, such as a location between the two domains. In some cases, a first domain or second domain can comprise an MS2 stem loop, a BoxB stem-loop, a U1A stem-loop, a modified version of any of these, or any combination thereof.

**[0087]** Two dimensional shape or secondary structure of a domain can influence efficiency of editing, off target effects, or a combination thereof as compared to a nucleic acid that can form a different two dimension shape or secondary structure. Therefore, an aspect of the disclosure includes modifying nucleic acids such that two dimensional shapes can be advantageously designed to enhance efficiency of editing and reduce off target effects.

Modifications to a sequence comprising a naturally occurring recruiting domains can also enhance editing efficiency and reduce off target effects. Therefore, an aspect of the disclosure includes modifying nucleic acids such that a sequence (such as a synthetic sequence) can be advantageously designed to enhance efficiency of editing and reduce off target effects.

Modifications can include altering a length of a domain (such as extending a length), altering a native sequence that results in a change in secondary structure, adding a chemical modification, or any combination thereof. Nucleic acids as described herein can provide these advantages.

**[0088]** In some cases, a nucleic acid as described herein can modify at least one base pair of a target nucleic acid at an efficiency of at least about: 3, 4, or 5 times greater than a comparable nucleic acid complexed with a native recruiting domain and an antisense domain (complementary to the target nucleic acid). In some cases, a nucleic acid as described herein can modify at least one base pair of a target nucleic acid at an efficiency of at least about: 3,



4, or 5 times greater than a comparable nucleic acid complex with a GluR2 domain and an antisense domain (complementary to the target nucleic acid). In some cases, a nucleic acid as described herein can modify at least one base pair of a target nucleic acid at an efficiency of at least about: 3, 4, or 5 times greater than a comparable nucleic acid complex with a Cas13b protein or active fragment thereof and an antisense domain (complementary to the target nucleic acid.) An improvement in efficiency can be measured by a sequencing method, such as Sanger Method.

**[0089]** An aspect of the disclosure provides for a vector. The vector can comprise a nucleic acid with a polynucleotide sequence encoding for at least one RNA editing entity recruiting domain. In some cases, the polynucleotide sequence may not form a secondary structure comprising a stem-loop. In some cases, the polynucleotide sequence can form one or more stem-loops. In some cases, the polynucleotide sequence can form a secondary structure comprising a cruciform. In some cases, the polynucleotide sequence can form a secondary structure that can be substantially linear. In some cases, the polynucleotide sequence can comprise at least about 80% sequence identity to one or more sequences comprising: an Alu domain encoding sequence, an Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC) recruiting domain encoding sequence, and any combination thereof. In some cases, the nucleic acid can be genetically encoded. In some cases, the nucleic acid can be chemically synthesized.

**[0090]** In some cases, a polynucleotide sequence can comprise at least about 80% sequence identity to an Alu domain encoding sequence. In some cases, a polynucleotide sequence can comprise at least about 85% sequence identity to an Alu domain encoding sequence. In some cases, a polynucleotide sequence can comprise at least about 90% sequence identity to an Alu domain encoding sequence. In some cases, a polynucleotide sequence can comprise at least about 95% sequence identity to an Alu domain encoding sequence. In some cases, the Alu domain encoding sequence can be a non-naturally occurring sequence. In some cases, the Alu domain encoding sequence can comprise a modified portion. In some cases, the Alu domain encoding sequence can comprise a portion of a naturally occurring Alu domain sequence.

**[0091]** In some cases, a polynucleotide sequence can comprise at least about 80% sequence identity to an APOBEC domain encoding sequence. In some cases, a polynucleotide sequence can comprise at least about 85% sequence identity to an APOBEC domain encoding sequence. In some cases, a polynucleotide sequence can comprise at least about 90% sequence identity to an APOBEC domain encoding sequence. In some cases, a

polynucleotide sequence can comprise at least about 95% sequence identity to an APOBEC domain encoding sequence. In some cases, the APOBEC domain encoding sequence can be a non-naturally occurring sequence. In some cases, the APOBEC domain encoding sequence can comprise a modified portion. In some cases, the APOBEC domain encoding sequence can comprise a portion of a naturally occurring APOBEC domain sequence.

**[0092]** In some cases, a polynucleotide sequence can comprise at least about 80% sequence identity to a GluR2 domain encoding sequence. In some cases, a polynucleotide sequence can comprise at least about 85% sequence identity to a GluR2 domain encoding sequence. In some cases, a polynucleotide sequence can comprise at least about 90% sequence identity to a GluR2 domain encoding sequence. In some cases, a polynucleotide sequence can comprise at least about 95% sequence identity to a GluR2 domain encoding sequence. In some cases, the GluR2 domain encoding sequence can be a non-naturally occurring sequence. In some cases, the GluR2 domain encoding sequence can comprise a modified portion. In some cases, the GluR2 domain encoding sequence can comprise a portion of a naturally occurring GluR2 domain sequence.

**[0093]** In some cases, a polynucleotide sequence can comprise at least about 80% sequence identity to an encoding sequence that recruits an ADAR. A polynucleotide sequence encoding for at least one RNA editing entity recruiting domain can be isolated and purified or can be synthesized. Such a polynucleotide sequence can be configured specifically to recruit an ADAR to a target site. The recruitment can include exogenous ADAR recruitment (that can be co delivered or separately delivered), endogenous ADAR recruitment, or a combination thereof. In some cases, a polynucleotide sequence can be configured specifically to enhance recruitment of ADAR or enhance specificity of ADAR recruitment to a particular site as compared to a naturally occurring recruiting domain. In some cases, the encoding sequence can be non-naturally occurring sequence. In some cases, the encoding sequence can comprise a modified portion. In some cases, the encoding sequence can comprise a portion of a naturally occurring ADAR recruiting domain sequence. Any sequence, either natural or synthetic, that recruits ADAR can be envisioned to be included in the polynucleotide sequence. In some cases, a polynucleotide sequence can comprise an exemplary sequence as described herein. **FIG. 2, FIG. 6, FIG. 7, FIG. 8, FIG. 14, FIG. 15, FIG. 21, FIG. 26, FIG. 28, FIG. 29, FIG. 30, and Table 1** include exemplary sequences. The sequences provided herein include sequences having at least a portion that can encode for at least one RNA editing entity recruiting domain.

[0094] **Table 1:** DNA encoding adRNA sequences. The adRNA sequence produced from these sequences is identical to the DNA sequence but T is replaced with U.:

|                                 |                                                                                                                                                                                                                                                      |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dual_<br>V2_R<br>ab7a_<br>20_6  | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>GCCGCCAGCTGGATTTCCTCAATTCTGAGTGTGGAATAGTATAACAAT<br>ATGCTAAATGTTGTTATAGTATCCCACT (SEQ ID NO:178)                                                                                                   |
| Dual_<br>V2_R<br>ab7a_<br>40_6  | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>GCCGCCAGCTGGATTTCCTCAATTCTGAGTAACACTCTGCGTGGGAATAG<br>TATAACAATATGCTAAATGTTGTTATAGTATCCCACT (SEQ ID NO:179)                                                                                        |
| Dual_<br>V2_R<br>ab7a_<br>60_6  | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>GCCGCCAGCTGGATTTCCTCAATTCTGAGTAACACTCTGCAATCCAAAC<br>AGGGTTCAACCGTGGGAATAGTATAACAATATGCTAAATGTTGTTATA<br>GTATCCCACT (SEQ ID NO:180)                                                                |
| Dual_<br>V2_R<br>ab7a_<br>80_6  | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>GCCGCCAGCTGGATTTCCTCAATTCTGAGTAACACTCTGCAATCCAAAC<br>AGGGTTCAACCGTCCACCTTACAGGCCTGCAGTGGGAATAGTATAACA<br>ATATGCTAAATGTTGTTATAGTATCCCACT (SEQ ID NO:181)                                            |
| Dual_<br>V2_R<br>ab7a_<br>100_6 | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>GCCGCCAGCTGGATTTCCTCAATTCTGAGTAACACTCTGCAATCCAAAC<br>AGGGTTCAACCGTCCACCTTACAGGCCTGCATTACAGGACTTAAACAC<br>ATAGTGGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>AC (SEQ ID NO:182)                  |
| Dual_<br>V2_R<br>ab7a_<br>120_6 | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>GCCGCCAGCTGGATTTCCTCAATTCTGAGTAACACTCTGCAATCCAAAC<br>AGGGTTCAACCGTCCACCTTACAGGCCTGCATTACAGGACTTAAACAC<br>ATAATCCAAGAATTTCTTACACTGTGGAATAGTATAACAATATGCTA<br>AATGTTGTTATAGTATCCCACT (SEQ ID NO:183) |
| Dual_<br>V2_R<br>ab7a_<br>20_10 | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>TACTGCCGCCAGCTGGATTGTGGAATAGTATAACAATATGCTAAATG<br>TTGTTATAGTATCCCACT (SEQ ID NO:184)                                                                                                              |
| Dual_<br>V2_R<br>ab7a_<br>40_20 | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>CTGTACAGAATACTGCCGCCAGCTGGATTTCCTCAATTCTGTGGAATAG<br>TATAACAATATGCTAAATGTTGTTATAGTATCCCACT (SEQ ID<br>NO:185)                                                                                      |
| Dual_<br>V2_R<br>ab7a_<br>60_30 | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACT<br>CTTGTGTCTACTGTACAGAATACTGCCGCCAGCTGGATTTCCTCAATTCT<br>GAGTAACACTGTGGAATAGTATAACAATATGCTAAATGTTGTTATA<br>GTATCCCACT (SEQ ID NO:186)                                                                 |

|                                      |                                                                                                                                                                                                                                                        |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dual_<br>V2_R<br>ab7a_<br>80_40      | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACC<br>GTACATAATTCTTGTGTCTACTGTACAGAATACTGCCGCCAGCTGGATT<br>TCCCAATTCTGAGTAACACTCTGCAATCCAGTGGAATAGTATAACAAT<br>ATGCTAAATGTTGTTATAGTATCCCAC (SEQ ID NO:187)                                                 |
| Dual_<br>V2_R<br>ab7a_<br>100_5<br>0 | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACC<br>GATAAAAGGCGTACATAATTCTTGTGTCTACTGTACAGAATACTGCCG<br>CCAGCTGGATTTCCCAATTCTGAGTAACACTCTGCAATCCAAACAGGG<br>TTCGTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCC<br>AC (SEQ ID NO:188)                         |
| Dual_<br>V2_R<br>ab7a_<br>120_6<br>0 | GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACC<br>TTAAGTCTTTGATAAAAGGCGTACATAATTCTTGTGTCTACTGTACAGA<br>ATACTGCCGCCAGCTGGATTTCCCAATTCTGAGTAACACTCTGCAATCC<br>AAACAGGGTTCAACCCTCCACGTGGAATAGTATAACAATATGCTAAA<br>TGTTGTTATAGTATCCCAC (SEQ ID NO:189)     |
| Dual_<br>V7_Ra<br>b7a_20<br>6        | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CTGCCGCCAGCTGGATTTCCCAATTCTGAGTGTGGAAGAGGAGAAAC<br>AATAGGCTAAACGTTGTTCTCGTCTCCCAC(SEQ ID NO:190)                                                                                                     |
| Dual_V<br>7_Rab7<br>a 40 6           | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CTGCCGCCAGCTGGATTTCCCAATTCTGAGTAACACTCTGCGTGGAA<br>GAGGAGAAACAATAGGCTAAACGTTGTTCTCGTCTCCCAC(SEQ ID<br>NO:191)                                                                                        |
| Dual_V<br>7_Rab7<br>a 60 6           | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CTGCCGCCAGCTGGATTTCCCAATTCTGAGTAACACTCTGCAATCCA<br>AACAGGGTTCAACCCTCCACCTTACAGGCCTGCAGTGGAAGAGGAG<br>TTCTCGTCTCCCAC(SEQ ID NO:192)                                                                   |
| Dual_V<br>7_Rab7<br>a 80 6           | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CTGCCGCCAGCTGGATTTCCCAATTCTGAGTAACACTCTGCAATCCA<br>AACAGGGTTCAACCCTCCACCTTACAGGCCTGCAGTGGAAGAGGAG<br>AACAATAGGCTAAACGTTGTTCTCGTCTCCCAC(SEQ ID NO:193)                                                |
| Dual_V<br>7_Rab7<br>a_100_<br>6      | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CTGCCGCCAGCTGGATTTCCCAATTCTGAGTAACACTCTGCAATCCA<br>AACAGGGTTCAACCCTCCACCTTACAGGCCTGCATTACAGGACTTAA<br>ACACATAGTGGAAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGT<br>CTCCCAC (SEQ ID NO:194)                    |
| Dual_V<br>7_Rab7<br>a_120_<br>6      | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CTGCCGCCAGCTGGATTTCCCAATTCTGAGTAACACTCTGCAATCCA<br>AACAGGGTTCAACCCTCCACCTTACAGGCCTGCATTACAGGACTTAA<br>ACACATAATCCAAGAATTTCTTACACTGTGGAAGAGGAGAGAACAATA<br>GGCTAAACGTTGTTCTCGTCTCCCAC (SEQ ID NO:195) |

|                                 |                                                                                                                                                                                                                                                                  |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dual_V<br>7_Rab7<br>a_20_1<br>0 | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CATACTGCCGCC <u>C</u> AGCTGGATTGTGGAAGAGGAGAGAACAATAGGCTA<br>AACGTTGTTCTCGTCTCCAC (SEQ ID NO:196)                                                                                                              |
| Dual_V<br>7_Rab7<br>a_40_2<br>0 | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CACTGTACAGAATACTGCCGCC <u>C</u> AGCTGGATTTCCTCAATTCTGTGGAA<br>GAGGAGAACAAATAGGCTAAACGTTGTTCTCGTCTCCAC (SEQ ID<br>NO:197)                                                                                       |
| Dual_V<br>7_Rab7<br>a_60_3<br>0 | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CTCTTGTGTCTACTGTACAGAATACTGCCGCC <u>C</u> AGCTGGATTTCCTCAAT<br>TCTGAGTAACACTGTGGAAGAGGAGAGAACAATAGGCTAAACGTTGT<br>TCTCGTCTCCAC (SEQ ID NO:198)                                                                 |
| Dual_V<br>7_Rab7<br>a_80_4<br>0 | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CCGTACATAATTCTTGTGTCTACTGTACAGAATACTGCCGCC <u>C</u> AGCTGG<br>ATTTCCCAATTCTGAGTAACACTCTGCAATCCAGTGGAAGAGGAGA<br>ACAATAGGCTAAACGTTGTTCTCGTCTCCAC (SEQ ID NO:199)                                                |
| Dual_V<br>7_Rab7<br>a_100_50    | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CTGATAAAAGGCGTACATAATTCTTGTGTCTACTGTACAGAATACTG<br>CCGCC <u>C</u> AGCTGGATTTCCTCAATTCTGAGTAACACTCTGCAATCCAAAC<br>AGGGTTTCGTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGT<br>CTCCAC (SEQ ID NO:200)                    |
| Dual_V<br>7_Rab7<br>a_120_60    | GTGGAAGAGGAGAGAACAATAGGCTAAACGTTGTTCTCGTCTCCCA<br>CCTTAAGTCTTTGATAAAAGGCGTACATAATTCTTGTGTCTACTGTAC<br>AGAATACTGCCGCC <u>C</u> AGCTGGATTTCCTCAATTCTGAGTAACACTCTGC<br>AATCCAAACAGGGTTCAACCCTCCACGTGGAAGAGGAGAGAACAATA<br>GGCTAAACGTTGTTCTCGTCTCCAC (SEQ ID NO:201) |

[0095] A polynucleotide sequence encoding for an RNA editing entity recruiting domain, can include recruitment of any ADAR protein (such as ADAR1, ADAR2, ADAR3 or any combination thereof), any APOBEC protein (such as APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3E, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, or any combination thereof), or a combination thereof. In some cases, the ADAR or APOBEC protein recruited can be mammalian. In some cases, the ADAR or APOBEC protein recruited can be human. In some cases, the ADAR or APOBEC protein recruited can be recombinant (such as an exogenously delivered ADAR or APOBEC), modified (such as an exogenously delivered ADAR or APOBEC), endogenous (such as an endogenous ADAR or APOBEC), or any combination thereof.

**[0096]** In some cases, the at least one RNA editing entity recruiting domain does not form a second structure comprising a stem-loop. In some cases, the at least one RNA editing entity recruiting domain forms a second structure comprising a stem-loop. In some cases, the at least one RNA editing entity recruiting domain forms a second structure that does not comprise a stem-loop. In some cases, the at least one RNA editing entity recruiting domain forms a secondary structure comprising a linear portion. In some cases, the at least one RNA editing entity recruiting domain forms a secondary structure comprising a cruciform or portion thereof.

**[0097]** A polynucleotide sequence can encode for more than one RNA editing recruiting domains. In some cases, a polynucleotide sequence can encode for a plurality of recruiting domains. In some cases, a polynucleotide sequence can encode for 2, 3, 4, 5, 6 or more recruiting domains. A recruiting domain of a plurality can include an Alu domain, an APOBEC domain, a GluR2 domain, Cas13 domain, or any combination thereof. In some case, the Alu domain, APOBEC domain, Cas13 domain, or GluR2 domain can be a naturally occurring recruiting domain. In some cases, the Alu domain, the APOBEC domain, Cas13 domain, or the GluR2 domain can be non-naturally occurring, can be modified from a native sequence, or can be recombinant. At least one of the plurality of recruiting domains can comprise a single stranded sequence. At least one of the plurality of recruiting domains can comprise a plurality of Alu repeats. At least one of the plurality of recruiting domains can form a secondary structure comprising a stem-loop. At least one of the plurality of recruiting domains can form a secondary structure that does not comprise a stem-loop. At least one of the plurality of recruiting domains can form a secondary structure that comprises a cruciform or portion thereof. At least one of the plurality of recruiting domains can form a secondary structure that comprises a toe hold.

**[0098]** In some cases, a nucleic acid can encode for at least one RNA editing entity recruiting domain. In some cases, the nucleic acid can encode for an RNA that is complementary to at least a portion of a target RNA. In some cases, the nucleic acid can encode for a recruiting domain and a targeting domain. In some cases, the nucleic acid can encode for a recruiting domain and a nucleic acid can encode for a targeting domain. The portion of the target RNA can comprise a single base. The portion of the target RNA can comprise a plurality of bases. The portion of the target RNA can comprise about: 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 100, 200, 300, 400, 500, 600, 700, 800, 900, 100 base pairs or more. In some cases, the target RNA can comprise from about 1 bps to about 10 bps. In some cases, the target RNA

can comprise from about 10 bps to about 100 bps. In some cases, the target RNA can comprise from about 10 bps to about 500 bps. In some cases, the target RNA can comprise from about 10 bps to about 1000 bps. A nucleic acid comprising a targeting domain and a recruiting domain can comprise a contiguous sequence of at least about 200bp in length. A nucleic acid comprising a targeting domain and a recruiting domain can comprise a contiguous sequence of at least about 150bp in length. A nucleic acid comprising a targeting domain and a recruiting domain can comprise a contiguous sequence of at least about 250bp in length. A nucleic acid comprising a targeting domain and a recruiting domain can comprise a contiguous sequence of at least about 275bp in length. A nucleic acid comprising a targeting domain and a recruiting domain can comprise a contiguous sequence of at least about 300bp in length. A nucleic acid comprising a targeting domain and a recruiting domain can comprise a contiguous sequence of at least about 400bp in length. A nucleic acid comprising a targeting domain and a recruiting domain can comprise a contiguous sequence of at least about 500bp in length.

**[0099]** A vector can be employed to deliver a nucleic acid. A vector can comprise DNA, such as double stranded DNA or single stranded DNA. A vector can comprise RNA. In some cases, the RNA can comprise a base modification. The vector can comprise a recombinant vector. The vector can be a vector that is modified from a naturally occurring vector. The vector can comprise at least a portion of a non-naturally occurring vector. Any vector can be utilized. In some cases, the vector can comprise a viral vector, a liposome, a nanoparticle, an exosome, an extracellular vesicle, or any combination thereof. In some cases, a viral vector can comprise an adenoviral vector, an adeno-associated viral vector (AAV), a lentiviral vector, a retroviral vector, a portion of any of these, or any combination thereof. In some cases, a nanoparticle vector can comprise a polymeric-based nanoparticle, an aminolipid based nanoparticle, a metallic nanoparticle (such as gold-based nanoparticle), a portion of any of these, or any combination thereof. In some cases, a vector can comprise an AAV vector. A vector can be modified to include a modified VP1 protein (such as an AAV vector modified to include a VP1 protein). An AAV can comprise a serotype – such as an AAV1 serotype, an AAV2 serotype, AAV3 serotype, an AAV4 serotype, AAV5 serotype, an AAV6 serotype, AAV7 serotype, an AAV8 serotype, an AAV9 serotype, a derivative of any of these, or any combination thereof.

**[0100]** Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this disclosure

belongs. All nucleotide sequences provided herein are presented in the 5' to 3' direction unless identified otherwise. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the disclosure, the preferred methods, devices, and materials are now described. All technical and patent publications cited herein are incorporated herein by reference in their entirety. Nothing herein is to be construed as an admission that the disclosure is not entitled to antedate such disclosure by virtue of prior disclosure.

**[0101]** The practice of the technology will employ, unless otherwise indicated, conventional techniques of tissue culture, immunology, molecular biology, microbiology, cell biology, and recombinant DNA, which are within the skill of the art. See, *e.g.*, Green and Sambrook eds. (2012) *Molecular Cloning: A Laboratory Manual*, 4th edition; the series Ausubel *et al.* eds. (2015) *Current Protocols in Molecular Biology*; the series Methods in Enzymology (Academic Press, Inc., N.Y.); MacPherson *et al.* (2015) *PCR 1: A Practical Approach* (IRL Press at Oxford University Press); MacPherson *et al.* (1995) *PCR 2: A Practical Approach*; McPherson *et al.* (2006) *PCR: The Basics* (Garland Science); Harlow and Lane eds. (1999) *Antibodies, A Laboratory Manual*; Greenfield ed. (2014) *Antibodies, A Laboratory Manual*; Freshney (2010) *Culture of Animal Cells: A Manual of Basic Technique*, 6th edition; Gait ed. (1984) *Oligonucleotide Synthesis*; U.S. Pat. No. 4,683,195; Hames and Higgins eds. (1984) *Nucleic Acid Hybridization*; Anderson (1999) *Nucleic Acid Hybridization*; Herdewijn ed. (2005) *Oligonucleotide Synthesis: Methods and Applications*; Hames and Higgins eds. (1984) *Transcription and Translation*; Buzdin and Lukyanov ed. (2007) *Nucleic Acids Hybridization: Modern Applications*; Immobilized Cells and Enzymes (IRL Press (1986)); Grandi ed. (2007) *In vitro Transcription and Translation Protocols*, 2nd edition; Guisan ed. (2006) *Immobilization of Enzymes and Cells*; Perbal (1988) *A Practical Guide to Molecular Cloning*, 2nd edition; Miller and Calos eds, (1987) *Gene Transfer Vectors for Mammalian Cells* (Cold Spring Harbor Laboratory); Makrides ed. (2003) *Gene Transfer and Expression in Mammalian Cells*; Mayer and Walker eds. (1987) *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Lundblad and Macdonald eds. (2010) *Handbook of Biochemistry and Molecular Biology*, 4th edition; Herzenberg *et al.* eds (1996) *Weir's Handbook of Experimental Immunology*, 5th edition; and/or more recent editions thereof.

**[0102]** The terminology used in the description herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the disclosure.



[0103] All numerical designations, *e.g.*, pH, temperature, time, concentration, and molecular weight, including ranges, are approximations which are varied ( + ) or ( - ) by increments of 1.0 or 0.1, as appropriate or alternatively by a variation of +/- 15 %, or alternatively 10% or alternatively 5% or alternatively 2%. It is to be understood, although not always explicitly stated, that all numerical designations are preceded by the term “about”. It also is to be understood, although not always explicitly stated, that the reagents described herein are merely exemplary and that equivalents of such are known in the art.

[0104] Unless the context indicates otherwise, it is specifically intended that the various features of the disclosure described herein can be used in any combination. Moreover, the disclosure also contemplates that in some embodiments, any feature or combination of features set forth herein can be excluded or omitted. To illustrate, if the specification states that a complex comprises components A, B and C, it is specifically intended that any of A, B or C, or a combination thereof, can be omitted and disclaimed singularly or in any combination.

[0105] Unless explicitly indicated otherwise, all specified embodiments, features, and terms intend to include both the recited embodiment, feature, or term and biological equivalents thereof.

### Definitions

[0106] As used in the specification and claims, the singular form “a”, “an” and “the” include plural references unless the context clearly dictates otherwise. For example, the term “a polypeptide” includes a plurality of polypeptides, including mixtures thereof. Accordingly, unless the contrary is indicated, the numerical parameters set forth in this application are approximations that can vary depending upon the desired properties sought to be obtained by the disclosure.

[0107] The term “about,” as used herein can mean within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which can depend in part on how the value is measured or determined, *e.g.*, the limitations of the measurement system. For example, “about” can mean plus or minus 10%, per the practice in the art. Alternatively, “about” can mean a range of plus or minus 20%, plus or minus 10%, plus or minus 5%, or plus or minus 1% of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, within 5-fold, or within 2-fold, of a value. Where particular values are described in the application and claims, unless otherwise stated the term “about” meaning within an acceptable error range for the

particular value can be assumed. Also, where ranges and/or subranges of values are provided, the ranges and/or subranges can include the endpoints of the ranges and/or subranges. In some cases, variations can include an amount or concentration of 20%, 10%, 5%, 1 %, 0.5%, or even 0.1 % of the specified amount.

**[0108]** For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

**[0109]** The terms “adenine”, “guanine”, “cytosine”, “thymine”, “uracil” and “hypoxanthine” (the nucleobase in inosine) as used herein refer to the nucleobases as such.

**[0110]** The terms “adenosine”, “guanosine”, “cytidine”, “thymidine”, “uridine” and “inosine”, refer to the nucleobases linked to the (deoxy)ribose sugar.

**[0111]** The term “adeno-associated virus” or “AAV” as used herein refers to a member of the class of viruses associated with this name and belonging to the genus dependoparvovirus, family Parvoviridae. Multiple serotypes of this virus are known to be suitable for gene delivery; all known serotypes can infect cells from various tissue types. At least 11, sequentially numbered, are disclosed in the prior art. Non-limiting exemplary serotypes useful for the purposes disclosed herein include any of the 11 serotypes, *e.g.*, AAV2 and AAV8. The term “lentivirus” as used herein refers to a member of the class of viruses associated with this name and belonging to the genus lentivirus, family Retroviridae. While some lentiviruses are known to cause diseases, other lentivirus are known to be suitable for gene delivery. *See, e.g.*, Tomás *et al.* (2013) Biochemistry, Genetics and Molecular Biology: “Gene Therapy – Tools and Potential Applications,” ISBN 978-953-51-1014-9, DOI: 10.5772/52534.

**[0112]** The term “adenosine deaminases acting on RNA” or “ADAR” as used herein can refer to an adenosine deaminase that can convert adenosines (A) to inosines (I) in an RNA sequence. ADAR1 and ADAR2 are two exemplary species of ADAR that are involved in mRNA editing *in vivo*. Non-limiting exemplary sequences for ADAR1 can be found under the following reference numbers: HGNC: 225; Entrez Gene: 103; Ensembl: ENSG00000160710; OMIM: 146920; UniProtKB: P55265; and GeneCards: GC01M154554, as well as biological equivalents thereof. Non-limiting exemplary sequences for ADAR2 can be found under the following reference numbers: HGNC: 226; Entrez Gene: 104; Ensembl: ENSG00000197381; OMIM: 601218; UniProtKB: P78563; and GeneCards: GC21P045073,

as well as biological equivalents thereof. Further non-limited exemplary sequences of the catalytic domain are provided hereinabove. The forward and reverse RNA used to direct site-specific ADAR editing are known as “adRNA” and “radRNA,” respectively. The catalytic domains of ADAR1 and ADAR2 are comprised in the sequences provided herein below.

**[0113]** ADAR1 catalytic domain:

KAERMGFTEVTPVTGASLRRTMLLSRSPEAQPKTLPLTGSTFHDQIAMLSHRCFNTL  
TNSFQPSLLGRKILAAIIMKKDSEDMGVVVS LGTGNRCVKGDSLSLKGETVNDCHAE  
IISRRGFIRFLYSELMKYNSQTAKDSIFEPAGGGEKLQIKKTVSFHLYISTAPCGDGALF  
DKSCSDRAMESTESRHYPVFENPKQGKLRTKVENEGGTIPVESSDIVPTWDGIRLGER  
LRTMSCSDKILRWNVLGLQGALLTHFLQPIYLKSVTLGYLFSQGH LTRAICCRVTRD  
GSAFEDGLRHPFIVNHPKVGRVSIYDSKRQSGKTKETSVNWCLADGYDLEILDGTRG  
TVDGPRNELSRVSKKNIFLLFKKLCSEFRYRRDLLRLSYGEAKKAARDYETAKNYFKK  
GLKDMGYGNWISKPQEEKNFYLCPV (SEQ ID NO:140)

**[0114]** ADAR2 catalytic domain:

**[0115]** QLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVL AGVVM TTTGTDVKDA  
KVISVSTGTCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRS  
IFQKSERGGFRLKENVQFHL YISTSPCGDARIFSPHEPILEEPADRHPNRKARGQLRTKI  
ESGEGTIPVRSNASIQTW DGV LQGERLLT MSCSDKIARWNVVG IQGSLLSIFVEPIYFS  
SIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNFSVNW  
TVGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPSHLLRSKITKPNVY  
HESKLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLT (SEQ ID NO:141)

**[0116]** The double stranded RNA binding domains (dsRBD) of an ADAR is comprised in the sequence provided herein below.

**[0117]** ADAR dsRBD:

**[0118]** MDIEDEENMSSSSTDVKENRNLDNVSPKDGSTPGPGEGS QLSNGGGGGPGRK  
RPLEEGSNGH SKYRLKKRRKTPGPVLPKNALMQLNEIKPGLQYTLTSQTGPVHAPLF  
VMSVEVNGQVFEGSGPTKKKAKLHAAEKALRSFVQFPNASEAHLAMGRTLSVNTD  
FTSDQADFPDTL FNGFETPDKAEP PFYVGSNGDDSFSSSGDLSLSASPVPASLAQPPLP  
VLPPFPFPPSGKNPVMILNELRPGLKYDFLSESGESHAKSFVMSVVVDGQFFEGSGR NK  
KLAKARAAQSALAAIFN (SEQ ID NO:142)

**[0119]** It is appreciated that further mutations can be made to the sequence of the ADAR and/or its various domains. For example, the disclosure provides E488Q and E1008Q mutants of both ADAR1 and ADAR2, as well as a “promiscuous” variant of ADAR2 –

resulting from a C-terminal deletion. This “promiscuous” variant is known as such because it demonstrated promiscuity in edited reads with several A’s close to a target sequence showing an A to G conversion (verified across 2 different loci). The sequence of this variant is provided herein below.

**[0120]** “Promiscuous” ADAR2 variant:

**[0121]** MLRSFVQFPNASEAHLAMGRTLSVNTDFTSDQADFPDTLFGFETPDKAEPFF  
YVGSNGDDSFSSSGDLSLSASPVPASLAQPPLPVLPPFPPPSGKNPVMILNELRPGLKY  
DFLSESGESHAKSFVMSVVVDGQFFEGSGRNKKLAKARAAQSALAAIFNLHLDQTPS  
RQPIPSEGLQLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVLAGVVMTTGTD  
VKDAKVISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDD  
QKRSIFQKSERGGFRLKENVQFHLIYSTPCGDARIFSPHEPILEEPADRHPNRKARGQ  
LRTKIESGEGTIPVRSNASIQTWGVLQGERLLTMSCDKIARWNVVGIQGSLLSIFVE  
PIYFSSIILGSLYHGDHLSTRAMYQRISNIEDLPPLYTLNKPLLSGISNAEARQPGKAPNF  
SVNWTVGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVP SHLLRSKITK  
PNVYHESKLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLTP\* (SEQ ID  
NO:143)

**[0122]** Not to be bound by theory, a C-terminal deletion in ADAR1 can produce the same or similar effect.

**[0123]** The term “Alu domain” can refer to a sequence obtained from the Alu transposable element (“Alu element”). Typically the Alu element is about 300 base pairs in length. An Alu element typically comprise a structure: cruciform-polyA5-TAC-polyA6-cruciform-polyA tail, wherein both cruciform domains are similar in nucleotide sequence. An “Alu domain” can comprise a cruciform portion of the Alu element. In some embodiments, two Alu domains comprising cruciform structures are linked by a sequence complementary to a target RNA sequence.

**[0124]** The term “APOBEC” as used herein can refer to any protein that falls within the family of evolutionarily conserved cytidine deaminases involved in mRNA editing – catalyzing a C to T edit, which can be interpreted as a C to U conversion – and equivalents thereof. In some aspects, the term APOBEC can refer to any one of APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3E, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4, or equivalents each thereof. Non-limiting exemplary sequences of fusion proteins comprising one or more APOBEC domains are provided herein both fused to an ADAR domain or fused to alternative domains to render them suitable for use in an RNA

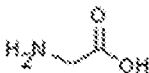
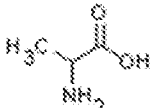
editing system. To this end, APOBECs can be considered an equivalent of ADAR – catalyzing editing albeit by a different conversion. Thus, not to be bound by theory, it is believed that all embodiments contemplated herein for use with an ADAR based editing system can be adapted for use in an APOBEC based RNA editing system. In some cases, use of APOBEC can involve certain modifications, such as but not limited to the use of particular guide RNA or “gRNA” to recruit the enzyme.

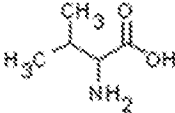
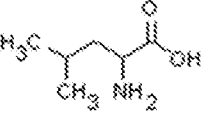
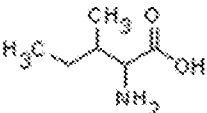
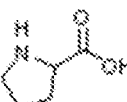
**[0125]** An “aptamer” can refer to a short single-stranded oligonucleotide capable of binding various molecules with high affinity and specificity. Non-limiting examples of aptamers are described in Lakhin, A. V. *et al.* (2013). *Acta naturae*, 5(4), 34-43.

**[0126]** As used herein, the term “comprising” is intended to mean that the compositions and methods include the recited elements, but do not exclude others. Unless otherwise indicated, open terms for example “contain,” “containing,” “include,” “including,” and the like mean comprising. “Consisting essentially of” when used to define compositions and methods, shall mean excluding other elements of any essential significance to the combination for the intended use. Thus, a composition consisting essentially of the elements as defined herein may not exclude trace contaminants from the isolation and purification method and pharmaceutically acceptable carriers, such as phosphate buffered saline, preservatives, and the like. “Consisting of” shall mean excluding more than trace elements of other ingredients and substantial method steps for administering the compositions of this disclosure. Embodiments defined by each of these transition terms are within the scope of this disclosure.

**[0127]** “Canonical amino acids” refer to those 20 amino acids found naturally in the human body shown in the table below with each of their three letter abbreviations, one letter abbreviations, structures, and corresponding codons:

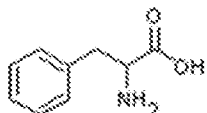
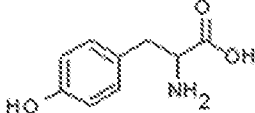
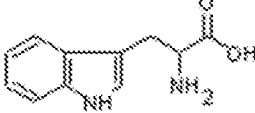
**non-polar, aliphatic residues**

|         |     |   |                                                                                     |                 |
|---------|-----|---|-------------------------------------------------------------------------------------|-----------------|
| Glycine | Gly | G |  | GGU GGC GGA GGG |
| Alanine | Ala | A |  | GCU GCC GCA GCG |

|            |     |   |                                                                                   |                            |
|------------|-----|---|-----------------------------------------------------------------------------------|----------------------------|
| Valine     | Val | V |  | GUU GUC GUA GUG            |
| Leucine    | Leu | L |  | UUA UUG CUU CUC<br>CUA CUG |
| Isoleucine | Ile | I |  | AUU AUC AUA                |
| Proline    | Pro | P |  | CCU CCC CCA CCG            |

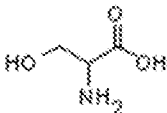
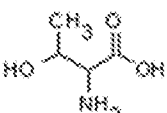
---

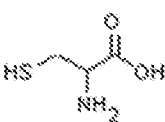
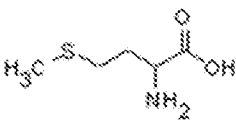
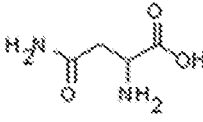
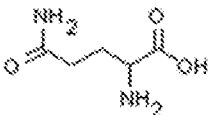
**aromatic residues**

|               |     |   |                                                                                     |         |
|---------------|-----|---|-------------------------------------------------------------------------------------|---------|
| Phenylalanine | Phe | F |   | UUU UUC |
| Tyrosine      | Tyr | Y |  | UAU UAC |
| Tryptophan    | Trp | W |  | UGG     |

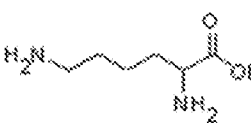
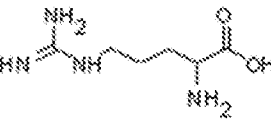
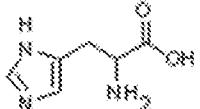
---

**polar, non-charged residues**

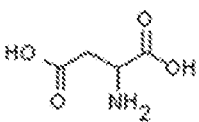
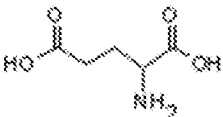
|           |     |   |                                                                                     |                            |
|-----------|-----|---|-------------------------------------------------------------------------------------|----------------------------|
| Serine    | Ser | S |  | UCU UCC UCA UCG<br>AGU AGC |
| Threonine | Thr | T |  | ACU ACC ACA ACG            |

|            |     |   |                                                                                   |         |
|------------|-----|---|-----------------------------------------------------------------------------------|---------|
| Cysteine   | Cys | C |  | UGU UGC |
| Methionine | Met | M |  | AUG     |
| Asparagine | Asn | N |  | AAU AAC |
| Glutamine  | Gln | Q |  | CAA CAG |

### positively charged residues

|           |       |                                                                                     |                         |
|-----------|-------|-------------------------------------------------------------------------------------|-------------------------|
| Lysine    | Lys K |  | AAA AAG                 |
| Arginine  | Arg R |  | CGU CGC CGA CGG AGA AGG |
| Histidine | His H |  | CAU CAC                 |

### negatively charged residues

|           |       |                                                                                     |         |
|-----------|-------|-------------------------------------------------------------------------------------|---------|
| Aspartate | Asp D |  | GAU GAC |
| Glutamate | Glu E |  | GAA GAG |

**[0128]** The term “Cas9” can refer to a CRISPR associated endonuclease referred to by this name. Non-limiting exemplary Cas9s include *Staphylococcus aureus* Cas9, nuclease dead Cas9, and orthologs and biological equivalents each thereof. Orthologs include but are not limited to *Streptococcus pyogenes* Cas9 (“spCas9”), Cas 9 from *Streptococcus thermophiles*, *Legionella pneumophila*, *Neisseria lactamica*, *Neisseria meningitides*, *Francisella novicida*; and Cpf1 (which performs cutting functions analogous to Cas9) from various bacterial species including *Acidaminococcus* spp. and *Francisella novicida* U112. For example, UniProtKB G3ECR1 (CAS9\_STRTR)) as well as dead Cas9 or dCas9, which lacks endonuclease activity (*e.g.*, with mutations in both the RuvC and HNH domain) can be used. The term “Cas9” may further refer to equivalents of the referenced Cas9 having at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto, including but not limited to other large Cas9 proteins. In some embodiments, the Cas9 is derived from *Campylobacter jejuni* or another Cas9 orthologs 1000 amino acids or less in length.

**[0129]** As used herein, the term “CRISPR” can refer to a technique of sequence specific genetic manipulation relying on the clustered regularly interspaced short palindromic repeats pathway. CRISPR can be used to perform gene editing and/or gene regulation, as well as to simply target proteins to a specific genomic location. “Gene editing” can refer to a type of genetic engineering in which the nucleotide sequence of a target polynucleotide is changed through introduction of deletions, insertions, single stranded or double stranded breaks, or base substitutions to the polynucleotide sequence. In some aspect, CRISPR-mediated gene editing utilizes the pathways of nonhomologous end-joining (NHEJ) or homologous recombination to perform the edits. Gene regulation can refer to increasing or decreasing the production of specific gene products such as protein or RNA.

**[0130]** The term “deficiency” as used herein can refer to lower than normal (physiologically acceptable) levels of a particular agent. In context of a protein, a deficiency can refer to lower than normal levels of the full-length protein.

**[0131]** As used herein, the term “detectable marker” can refer to at least one marker capable of directly or indirectly, producing a detectable signal. A non-exhaustive list of this marker includes enzymes which produce a detectable signal, for example by colorimetry, fluorescence, luminescence, such as horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, glucose-6-phosphate dehydrogenase, chromophores such as fluorescent, luminescent dyes, groups with electron density detected by electron microscopy or by their electrical property such as conductivity, amperometry, voltammetry, impedance, detectable



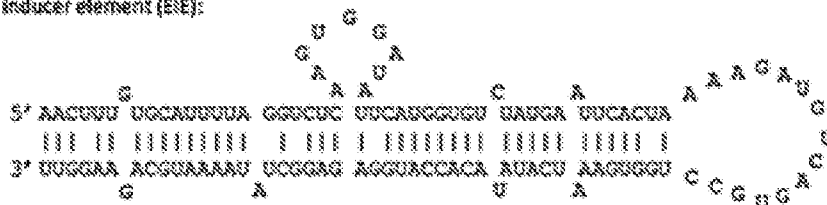
groups, for example whose molecules are of sufficient size to induce detectable modifications in their physical and/or chemical properties, such detection can be accomplished by optical methods such as diffraction, surface plasmon resonance, surface variation, the contact angle change or physical methods such as atomic force spectroscopy, tunnel effect, or radioactive molecules such as  $^{32}\text{P}$ ,  $^{35}\text{S}$  or  $^{125}\text{I}$ .

[0132] As used herein, the term “domain” can refer to a particular region of a protein or polypeptide and is associated with a particular function. For example, “a domain which associates with an RNA hairpin motif” can refer to the domain of a protein that binds one or more RNA hairpin. This binding can optionally be specific to a particular hairpin.

[0133] The term “dystrophin” as used herein refers to the protein corresponding with that name and encoded by the gene *Dmd*; a non-limiting example of which is found under UniProt Reference Number P11532 (for humans) and P11531 (for mice).

[0134] An “editing inducer element” can refer to a structure that is largely a double-stranded RNA, which is necessary for efficient RNA editing. Non-limiting examples of editing inducer elements are described in Daniel, C. *et al.* (2017) *Genome Biol.* 18, 195. A further non-limiting example of an editing inducer element is provided by the structure below (SEQ ID NO:15):

editing inducer element (EE):



### Modes of Carrying Out the Disclosure

[0135] ADARs are naturally occurring RNA editing enzymes that catalyze the hydrolytic deamination of adenosine to inosine that is biochemically recognized as guanosine.

APOBECs are enzymes, described herein above, that can perform a similar function but for cytosine to thymine.

[0136] The term “encode” as it is applied to polynucleotides can refer to a polynucleotide which is said to “encode” a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for the polypeptide and/or a fragment thereof. The antisense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

[0137] The terms “equivalent” or “biological equivalent” are used interchangeably when referring to a particular molecule, biological, or cellular material and intend those having minimal homology while still maintaining desired structure or functionality.

[0138] “Eukaryotic cells” comprise all of the life kingdoms except monera. They can be easily distinguished through a membrane-bound nucleus. Animals, plants, fungi, and protists are eukaryotes or organisms whose cells are organized into complex structures by internal membranes and a cytoskeleton. The most characteristic membrane-bound structure is the nucleus. Unless specifically recited, the term “host” includes a eukaryotic host, including, for example, yeast, higher plant, insect and mammalian cells. Non-limiting examples of eukaryotic cells or hosts include simian, bovine, porcine, murine, rat, avian, reptilian and human.

[0139] As used herein, “expression” can refer to the process by which polynucleotides are transcribed into mRNA and/or the process by which the transcribed mRNA is subsequently being translated into peptides, polypeptides, or proteins. If the polynucleotide is derived from genomic DNA, expression can include splicing of the mRNA in an eukaryotic cell.

[0140] As used herein, the term “functional” can be used to modify any molecule, biological, or cellular material to intend that it accomplishes a particular, specified effect.

[0141] The term “Glur2 mRNA” as used herein can refer to the mRNA encoding ionotropic AMPA glutamate receptor 2 (“Glur2”) which undergoes adenosine to inosine (A → I) editing. This mRNA recruits ADARs in a site specific manner.

[0142] The term “gRNA” or “guide RNA” as used herein can refer to guide RNA sequences used to target specific polynucleotide sequences for gene editing employing the CRISPR technique. Techniques of designing gRNAs and donor therapeutic polynucleotides for target specificity are well known in the art. For example, Doench, J., *et al.* Nature biotechnology 2014; 32(12):1262-7, Mohr, S. *et al.* (2016) FEBS Journal 283: 3232-38, and Graham, D., *et al.* Genome Biol. 2015; 16: 260. gRNA comprises or alternatively consists essentially of, or yet further consists of a fusion polynucleotide comprising CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA); or a polynucleotide comprising CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA). In some aspect, a gRNA is synthetic (Kelley, M. *et al.* (2016) J of Biotechnology 233 (2016) 74-83).

[0143] The terms “hairpin,” “hairpin loop,” “stem loop,” and/or “loop” used alone or in combination with “motif” is used in context of an oligonucleotide to refer to a structure formed in single stranded oligonucleotide when sequences within the single strand which are

complementary when read in opposite directions base pair to form a region whose conformation resembles a hairpin or loop.

**[0144]** “Homology” or “identity” or “similarity” can refer to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which can be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An “unrelated” or “non-homologous” sequence shares less than 40% identity, or alternatively less than 25% identity, with one of the sequences of the disclosure.

**[0145]** Homology refer to a % identity of a sequence to a reference sequence. As a practical matter, whether any particular sequence can be at least 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to any sequence described herein (which can correspond with a particular nucleic acid sequence described herein), such particular polypeptide sequence can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, Wis. 53711). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters can be set such that the percentage of identity is calculated over the full length of the reference sequence and that gaps in homology of up to 5% of the total reference sequence are allowed.

**[0146]** For example, in a specific embodiment the identity between a reference sequence (query sequence, *i.e.*, a sequence of the disclosure) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag *et al.* (Comp. App. Biosci. 6:237-245 (1990)). In some cases, parameters for a particular embodiment in which identity is narrowly construed, used in a FASTDB amino acid alignment, can include: Scoring Scheme=PAM (Percent Accepted Mutations) 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject sequence, whichever is shorter. According to this embodiment, if the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction can be made to the results to take into consideration the fact that the FASTDB program does not

account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity can be corrected by calculating the number of residues of the query sequence that are lateral to the N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. A determination of whether a residue is matched/aligned can be determined by results of the FASTDB sequence alignment. This percentage can be then subtracted from the percent identity, calculated by the FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score can be used for the purposes of this embodiment. In some cases, only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence are considered for this manual correction. For example, a 90 residue subject sequence can be aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C-termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity can be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for.

**[0147]** “Hybridization” can refer to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding can occur by Watson-Crick base pairing, Hoogsteen binding, or in any other sequence-specific manner. The complex can comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction can constitute a step

in a more extensive process, such as the initiation of a PC reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme.

**[0148]** Examples of stringent hybridization conditions include: incubation temperatures of about 25°C to about 37°C; hybridization buffer concentrations of about 6x SSC to about 10x SSC; formamide concentrations of about 0% to about 25%; and wash solutions from about 4x SSC to about 8x SSC. Examples of moderate hybridization conditions include: incubation temperatures of about 40°C to about 50°C; buffer concentrations of about 9x SSC to about 2x SSC; formamide concentrations of about 30% to about 50%; and wash solutions of about 5x SSC to about 2x SSC. Examples of high stringency conditions include: incubation temperatures of about 55°C to about 68°C; buffer concentrations of about 1x SSC to about 0.1x SSC; formamide concentrations of about 55% to about 75%; and wash solutions of about 1x SSC, 0.1x SSC, or deionized water. In general, hybridization incubation times are from 5 minutes to 24 hours, with 1, 2, or more washing steps, and wash incubation times are about 1, 2, or 15 minutes. SSC is 0.15 M NaCl and 15 mM citrate buffer. It is understood that equivalents of SSC using other buffer systems can be employed.

**[0149]** “Inhibit” as used herein refers to the ability to substantially antagonize, prohibit, prevent, restrain, slow, disrupt, alter, eliminate, stop, or reverse the progression or severity of the activity of a particular agent (*e.g.*, infectious agent) or disease.

**[0150]** As used herein, the term “interferon” can refer to a group of signaling proteins known to be associated with the immune response. In context of this application, the interferons of interest are those that result in enhanced expression of an ADAR. The correlation between interferon  $\alpha$  and ADAR1 is well known, and, thus, the disclosure contemplates use of interferon  $\alpha$  as a means of increasing endogenous ADAR1 expression. Commercial sources of isolated or recombinant interferon  $\alpha$  include but are not limited to Sigma-Aldrich, R&D Systems, Abcam, and Thermo Fisher Scientific. Alternatively, interferon  $\alpha$  can be produced using a known vector and given protein sequence, *e.g.*, Q6QNB6 (human IFNA).

**[0151]** The term “isolated” as used herein can refer to molecules or biologicals or cellular materials being substantially free from other materials. In one aspect, the term “isolated” can refer to nucleic acid, such as DNA or RNA, or protein or polypeptide (*e.g.*, an antibody or derivative thereof), or cell or cellular organelle, or tissue or organ, separated from other DNAs or RNAs, or proteins or polypeptides, or cells or cellular organelles, or tissues or organs, respectively, that are present in the natural source. The term “isolated” also can refer to a nucleic acid or peptide that is substantially free of cellular material, viral material, or

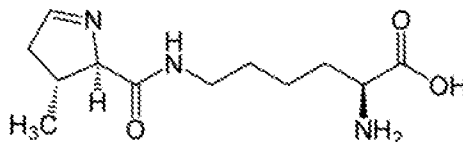
culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an “isolated nucleic acid” is meant to include nucleic acid fragments which are not naturally occurring as fragments and may not be found in the natural state. The term “isolated” is also used herein to refer to polypeptides which are isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides. The term “isolated” is also used herein to refer to cells or tissues that are isolated from other cells or tissues and is meant to encompass both cultured and engineered cells or tissues.

**[0152]** “Messenger RNA” or “mRNA” is a nucleic acid molecule that is transcribed from DNA and then processed to remove non-coding sections known as introns. The resulting mRNA is exported from the nucleus (or another locus where the DNA is present) and translated into a protein. The term “pre-mRNA” can refer to the strand prior to processing to remove non-coding sections.

**[0153]** The term “mutation” as used herein, can refer to an alteration to a nucleic acid sequence encoding a protein relative to the consensus sequence of said protein. “Missense” mutations result in the substitution of one codon for another; “nonsense” mutations change a codon from one encoding a particular amino acid to a stop codon. Nonsense mutations often result in truncated translation of proteins. “Silent” mutations are those which have no effect on the resulting protein. As used herein the term “point mutation” can refer to a mutation affecting only one nucleotide in a gene sequence. “Splice site mutations” are those mutations present pre-mRNA (prior to processing to remove introns) resulting in mistranslation and often truncation of proteins from incorrect delineation of the splice site. A mutation can comprise a single nucleotide variation (SNV). A mutation can comprise a sequence variant, a sequence variation, a sequence alteration, or an allelic variant. The reference DNA sequence can be obtained from a reference database. A mutation can affect function. A mutation may not affect function. A mutation can occur at the DNA level in one or more nucleotides, at the ribonucleic acid (RNA) level in one or more nucleotides, at the protein level in one or more amino acids, or any combination thereof. The reference sequence can be obtained from a database such as the NCBI Reference Sequence Database (RefSeq) database. Specific changes that can constitute a mutation can include a substitution, a deletion, an insertion, an inversion, or a conversion in one or more nucleotides or one or more amino acids. A mutation can be a point mutation. A mutation can be a fusion gene. A fusion pair or a fusion gene can result from a mutation, such as a translocation, an interstitial deletion, a chromosomal

inversion, or any combination thereof. A mutation can constitute variability in the number of repeated sequences, such as triplications, quadruplications, or others. For example, a mutation can be an increase or a decrease in a copy number associated with a given sequence (*i.e.*, copy number variation, or CNV). A mutation can include two or more sequence changes in different alleles or two or more sequence changes in one allele. A mutation can include two different nucleotides at one position in one allele, such as a mosaic. A mutation can include two different nucleotides at one position in one allele, such as a chimeric. A mutation can be present in a malignant tissue. A presence or an absence of a mutation can indicate an increased risk to develop a disease or condition. A presence or an absence of a mutation can indicate a presence of a disease or condition. A mutation can be present in a benign tissue. Absence of a mutation can indicate that a tissue or sample is benign. As an alternative, absence of a mutation may not indicate that a tissue or sample is benign. Methods as described herein can comprise identifying a presence of a mutation in a sample.

**[0154]** The term “non-canonical amino acids” can refer to those synthetic or otherwise modified amino acids that fall outside this group, typically generated by chemical synthesis or modification of canonical amino acids (*e.g.* amino acid analogs). The disclosure employs proteinogenic non-canonical amino acids in some of the methods and vectors disclosed herein. A non-limiting exemplary non-canonical amino acid is pyrrolysine (Pyl or O), the chemical structure of which is provided below:



**[0155]** Inosine (I) is another exemplary non-canonical amino acid, which can be found in tRNA and is essential for proper translation according to “wobble base pairing.” The structure of inosine is provided above.

**[0156]** The term “ornithine transcarbamylase” or “OTC” as used herein can refer to the protein corresponding with that name and encoded by the gene *Otc*; a non-limiting example of which is found under UniProt Reference Number P00480 (for humans) and P11725 (for mice). OTC deficiency is an X-linked genetic condition resulting in high concentrations of ammonia in blood. In some cases, OTC deficiency is caused by a G->A splice site mutation in the donor splice site of exon 4 that results in mis-splicing of the pre-mRNA. This mutation results in the formation of a protein that either is elongated or bears a point mutation. There is

a 15-20 fold reduction in the OTC protein levels. See, *e.g.*, Hodges, P. E. & Rosenberg, L. E. The *spf<sup>ash</sup>* mouse: a missense mutation in the ornithine transcarbamylase gene also causes aberrant mRNA splicing. *Proc. Natl. Acad. Sci. U. S. A.* 86, 4142–4146 (1989)) (showing the alternative forms of OTC produced). The sequences thereof are provided below:

*OTC pre-mRNA (wild type):*

.....CTCACAGACACCGCTCCGTTTGTAAAACCTTTTCTTC..... (SEQ ID NO:144)

*OTC pre-mRNA (mutant):*

.....CTCACAGACACCGCTCAGTTTGTAAAACCTTTTCTTC..... (SEQ ID NO:145)

*OTC mRNA (incorrectly spliced, mutant):*

.....CTCACAGACACCGCTCAGTTTGTAAAACCTTTTCTTC..... (SEQ ID NO:146)

*OTC mRNA (correctly spliced, mutant):*

.....CTCACAGACACCGCTCATGTCTTATCTAGCATGACA..... (SEQ ID NO:147)

*OTC mRNA (correctly spliced, wild type):*

.....CTCACAGACACCGCTCCGGTGTCTTATCTAGCATGACA..... (SEQ ID NO:148)

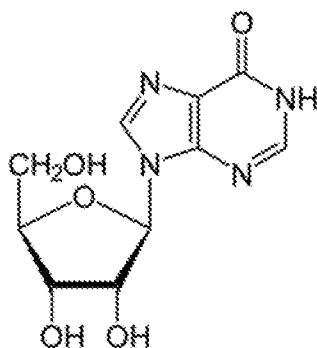
**[0157]** As shown above, a correct splice variant can be produced when the mutation is present; however, such production results in a missense mutation, which also can contribute to OTC deficiency.

**[0158]** The term “protein”, “peptide” and “polypeptide” are used interchangeably and in their broadest sense to refer to a compound of two or more subunit amino acids, amino acid analogs or peptidomimetics. The subunits can be linked by peptide bonds. In another embodiment, the subunit can be linked by other bonds, *e.g.*, ester, ether, *etc.* A protein or peptide can contain at least two amino acids and no limitation is placed on the maximum number of amino acids which can comprise a protein’s or peptide’s sequence. As used herein the term “amino acid” can refer to either natural and/or unnatural or synthetic amino acids, including glycine and both the D and L optical isomers, amino acid analogs and peptidomimetics. As used herein, the term “fusion protein” can refer to a protein comprised of domains from more than one naturally occurring or recombinantly produced protein, where generally each domain serves a different function. In this regard, the term “linker” can refer to a protein fragment that is used to link these domains together – optionally to preserve the conformation of the fused protein domains and/or prevent unfavorable interactions between the fused protein domains which can compromise their respective functions.



**[0159]** The terms “polynucleotide” and “oligonucleotide” are used interchangeably and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides or analogs thereof. Polynucleotides can have any three-dimensional structure and can perform any function, known or unknown. The following are non-limiting examples of polynucleotides: a gene or gene fragment (for example, a probe, primer, EST or SAGE tag), exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, RNAi, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A polynucleotide can comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure can be imparted before or after assembly of the polynucleotide. The sequence of nucleotides can be interrupted by non-nucleotide components. A polynucleotide can be further modified after polymerization, such as by conjugation with a labeling component. The term also can refer to both double- and single-stranded molecules. Unless otherwise specified or required, any embodiment of this disclosure that is a polynucleotide encompasses both the double-stranded form and each of two complementary single-stranded forms known or predicted to make up the double-stranded form.

**[0160]** A polynucleotide is composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); thymine (T); and uracil (U) for thymine when the polynucleotide is RNA. In some embodiments, the polynucleotide can comprise one or more other nucleotide bases, , such as inosine (I), a nucleoside formed when hypoxanthine is attached to ribofuranose via a  $\beta$ -N9-glycosidic bond, resulting in the chemical structure:



**[0161]** Inosine is read by the translation machinery as guanine (G).

**[0162]** The term “polynucleotide sequence” is the alphabetical representation of a polynucleotide molecule. This alphabetical representation can be input into databases in a

computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

**[0163]** As used herein, the term “purification marker” can refer to at least one marker useful for purification or identification. A non-exhaustive list of this marker includes His, lacZ, GST, maltose-binding protein, NusA, BCCP, c-myc, CaM, FLAG, GFP, YFP, cherry, thioredoxin, poly (NANP), V5, Snap, HA, chitin-binding protein, Softag 1, Softag 3, Strep, or S-protein. Suitable direct or indirect fluorescence marker comprise FLAG, GFP, YFP, RFP, dTomato, cherry, Cy3, Cy 5, Cy 5.5, Cy 7, DNP, AMCA, Biotin, Digoxigenin, Tamra, Texas Red, rhodamine, Alexa fluors, FITC, TRITC or any other fluorescent dye or hapten.

**[0164]** As used herein, the term “recombinant expression system” refers to a genetic construct or constructs for the expression of certain genetic material formed by recombination; the term “construct” in this regard is interchangeable with the term “vector” as defined herein.

**[0165]** As used herein, the term “recombinant protein” can refer to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding the polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein.

**[0166]** As used herein the term “restoring” in relation to expression of a protein can refer to the ability to establish expression of full length protein where previously protein expression was truncated due to mutation. In the context of “restoring activity” the term includes effecting the expression of a protein to its normal, non-mutated levels where a mutation resulted in aberrant expression (*e.g.*, too low or too high).

**[0167]** The term “sample” as used herein, generally refers to any sample of a subject (such as a blood sample or a tissue sample). A sample or portion thereof can comprise a stem cell. A portion of a sample can be enriched for the stem cell. The stem cell can be isolated from the sample. A sample can comprise a tissue, a cell, serum, plasma, exosomes, a bodily fluid, or any combination thereof. A bodily fluid can comprise urine, blood, serum, plasma, saliva, mucus, spinal fluid, tears, semen, bile, amniotic fluid, or any combination thereof. A sample or portion thereof can comprise an extracellular fluid obtained from a subject. A sample or portion thereof can comprise cell-free nucleic acid, DNA or RNA. A sample or portion thereof can be analyzed for a presence or absence or one or more mutations. Genomic data can be obtained from the sample or portion thereof. A sample can be a sample suspected or

confirmed of having a disease or condition. A sample can be a sample removed from a subject via a non-invasive technique, a minimally invasive technique, or an invasive technique. A sample or portion thereof can be obtained by a tissue brushing, a swabbing, a tissue biopsy, an excised tissue, a fine needle aspirate, a tissue washing, a cytology specimen, a surgical excision, or any combination thereof. A sample or portion thereof can comprise tissues or cells from a tissue type. For example, a sample can comprise a nasal tissue, a trachea tissue, a lung tissue, a pharynx tissue, a larynx tissue, a bronchus tissue, a pleura tissue, an alveoli tissue, breast tissue, bladder tissue, kidney tissue, liver tissue, colon tissue, thyroid tissue, cervical tissue, prostate tissue, heart tissue, muscle tissue, pancreas tissue, anal tissue, bile duct tissue, a bone tissue, brain tissue, spinal tissue, kidney tissue, uterine tissue, ovarian tissue, endometrial tissue, vaginal tissue, vulvar tissue, uterine tissue, stomach tissue, ocular tissue, sinus tissue, penile tissue, salivary gland tissue, gut tissue, gallbladder tissue, gastrointestinal tissue, bladder tissue, brain tissue, spinal tissue, a blood sample, or any combination thereof.

**[0168]** The term “sequencing” as used herein, can comprise bisulfite-free sequencing, bisulfite sequencing, TET-assisted bisulfite (TAB) sequencing, ACE-sequencing, high-throughput sequencing, Maxam-Gilbert sequencing, massively parallel signature sequencing, Polony sequencing, 454 pyrosequencing, Sanger sequencing, Illumina sequencing, SOLiD sequencing, Ion Torrent semiconductor sequencing, DNA nanoball sequencing, Heliscope single molecule sequencing, single molecule real time (SMRT) sequencing, nanopore sequencing, shot gun sequencing, RNA sequencing, Enigma sequencing, or any combination thereof.

**[0169]** The term “stop codon” intends a three nucleotide contiguous sequence within messenger RNA that signals a termination of translation. Non-limiting examples include in RNA, UAG, UAA, UGA and in DNA TAG, TAA or TGA. Unless otherwise noted, the term also includes nonsense mutations within DNA or RNA that introduce a premature stop codon, causing any resulting protein to be abnormally shortened. tRNA that correspond to the various stop codons are known by specific names: amber (UAG), ochre (UAA), and opal (UGA).

**[0170]** “Transfer ribonucleic acid” or “tRNA” is a nucleic acid molecule that helps translate mRNA to protein. tRNA have a distinctive folded structure, comprising three hairpin loops; one of these loops comprises a “stem” portion that encodes an anticodon. The anticodon recognizes the corresponding codon on the mRNA. Each tRNA is “charged with” an amino

acid corresponding to the mRNA codon; this “charging” is accomplished by the enzyme tRNA synthetase. Upon tRNA recognition of the codon corresponding to its anticodon, the tRNA transfers the amino acid with which it is charged to the growing amino acid chain to form a polypeptide or protein. Endogenous tRNA can be charged by endogenous tRNA synthetase. Accordingly, endogenous tRNA are typically charged with canonical amino acids. Orthogonal tRNA, derived from an external source, require a corresponding orthogonal tRNA synthetase. Such orthogonal tRNAs may be charged with both canonical and non-canonical amino acids. In some embodiments, the amino acid with which the tRNA is charged may be detectably labeled to enable detection *in vivo*. Techniques for labeling are known in the art and include, but are not limited to, click chemistry wherein an azide/alkyne containing unnatural amino acid is added by the orthogonal tRNA/synthetase pair and, thus, can be detected using alkyne/azide comprising fluorophore or other such molecule.

[0171] As used herein, the terms “treating,” “treatment” and the like are used herein to mean obtaining a desired pharmacologic and/or physiologic effect. The effect can be prophylactic in terms of completely or partially preventing a disease, disorder, or condition or sign or symptom thereof, and/or can be therapeutic in terms of a partial or complete cure for a disorder and/or adverse effect attributable to the disorder.

[0172] As used herein, the term “vector” can refer to a nucleic acid construct designed for transfer between different hosts, including but not limited to a plasmid, a virus, a cosmid, a phage, a BAC, a YAC, *etc.* A “viral vector” is defined as a recombinantly produced virus or viral particle that comprises a polynucleotide to be delivered into a host cell, either *in vivo*, *ex vivo* or *in vitro*. In some embodiments, plasmid vectors can be prepared from commercially available vectors. In other embodiments, viral vectors can be produced from baculoviruses, retroviruses, adenoviruses, AAVs, *etc.* according to techniques known in the art. In one embodiment, the viral vector is a lentiviral vector. Examples of viral vectors include retroviral vectors, adenovirus vectors, adeno-associated virus vectors, alphavirus vectors and the like. Infectious tobacco mosaic virus (TMV)-based vectors can be used to manufacture proteins and have been reported to express Griffithsin in tobacco leaves (O'Keefe *et al.* (2009) *Proc. Nat. Acad. Sci. USA* 106(15):6099-6104). Alphavirus vectors, such as Semliki Forest virus-based vectors and Sindbis virus-based vectors, have also been developed for use in gene therapy and immunotherapy. See, Schlesinger & Dubensky (1999) *Curr. Opin. Biotechnol.* 5:434-439 and Ying *et al.* (1999) *Nat. Med.* 5(7):823-827. In aspects where gene transfer is mediated by a retroviral vector, a vector construct can refer to the polynucleotide

comprising the retroviral genome or part thereof, and a gene of interest. Further details as to modern methods of vectors for use in gene transfer can be found in, for example, Kotterman *et al.* (2015) *Viral Vectors for Gene Therapy: Translational and Clinical Outlook Annual Review of Biomedical Engineering* 17. Vectors that contain both a promoter and a cloning site into which a polynucleotide can be operatively linked are well known in the art. Such vectors are capable of transcribing RNA *in vitro* or *in vivo* and are commercially available from sources such as Agilent Technologies (Santa Clara, Calif.) and Promega Biotech (Madison, Wis.). In one aspect, the promoter is a pol III promoter.

[0173] The pharmaceutical compositions for the administration of the AdRNA can be conveniently presented in dosage unit form and can be prepared by any of the methods well known in the art of pharmacy. The pharmaceutical compositions can be, for example, prepared by uniformly and intimately bringing the compounds provided herein into association with a liquid carrier, a finely divided solid carrier or both, and then, if necessary, shaping the product into the desired formulation. In the pharmaceutical composition the compound provided herein is included in an amount sufficient to produce the desired therapeutic effect. For example, pharmaceutical compositions of the technology can take a form suitable for virtually any mode of administration, including, for example, topical, ocular, oral, buccal, systemic, nasal, injection, infusion, transdermal, rectal, and vaginal, or a form suitable for administration by inhalation or insufflation.

[0174] For topical administration, the compounds can be formulated as solutions, gels, ointments, creams, suspensions, *etc.*, as is well-known in the art.

[0175] Systemic formulations include those designed for administration by injection (*e.g.*, subcutaneous, intravenous, infusion, intramuscular, intrathecal, or intraperitoneal injection) as well as those designed for transdermal, transmucosal, oral, or pulmonary administration.

[0176] Useful injectable preparations include sterile suspensions, solutions, or emulsions of the compounds provided herein in aqueous or oily vehicles. The compositions can also contain formulating agents, such as suspending, stabilizing, and/or dispersing agents. The formulations for injection can be presented in unit dosage form, *e.g.*, in ampules or in multidose containers, and can contain added preservatives.

[0177] Alternatively, the injectable formulation can be provided in powder form for reconstitution with a suitable vehicle, including but not limited to sterile pyrogen free water, buffer, and dextrose solution, before use. To this end, the compounds provided herein can be dried by any art-known technique, such as lyophilization, and reconstituted prior to use.

[0178] For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are known in the art.

[0179] For oral administration, the pharmaceutical compositions can take the form of, for example, lozenges, tablets, or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (*e.g.*, pregelatinised maize starch, polyvinylpyrrolidone, or hydroxypropyl methylcellulose); fillers (*e.g.*, lactose, microcrystalline cellulose, or calcium hydrogen phosphate); lubricants (*e.g.*, magnesium stearate, talc, or silica); disintegrants (*e.g.*, potato starch or sodium starch glycolate); or wetting agents (*e.g.*, sodium lauryl sulfate). The tablets can be coated by methods well known in the art with, for example, sugars, films, or enteric coatings.

[0180] Compositions intended for oral use can be prepared according to any method known to the art for the manufacture of pharmaceutical compositions, and such compositions can contain one or more agents selected from the group consisting of sweetening agents, flavoring agents, coloring agents, and preserving agents in order to provide pharmaceutically elegant and palatable preparations. Tablets contain the compounds provided herein in admixture with non-toxic pharmaceutically acceptable excipients which are suitable for the manufacture of tablets. These excipients can be for example, inert diluents, such as calcium carbonate, sodium carbonate, lactose, calcium phosphate or sodium phosphate; granulating and disintegrating agents (*e.g.*, corn starch or alginic acid); binding agents (*e.g.* starch, gelatin, or acacia); and lubricating agents (*e.g.*, magnesium stearate, stearic acid, or talc). The tablets can be left uncoated or they can be coated by known techniques to delay disintegration and absorption in the gastrointestinal tract and thereby provide a sustained action over a longer period. For example, a time delay material such as glyceryl monostearate or glyceryl distearate can be employed. They can also be coated by the techniques well known to the skilled artisan. The pharmaceutical compositions of the technology can also be in the form of oil-in-water emulsions.

[0181] Liquid preparations for oral administration can take the form of, for example, elixirs, solutions, syrups, or suspensions, or they can be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations can be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives, or hydrogenated edible fats); emulsifying agents (*e.g.*, lecithin, or acacia); non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol, cremophore™, or fractionated vegetable oils); and preservatives (*e.g.*, methyl or

propyl-p-hydroxybenzoates or sorbic acid). The preparations can also contain buffer salts, preservatives, flavoring, coloring, and sweetening agents as appropriate.

**[0182]** “Administration” can be effected in one dose, continuously or intermittently throughout the course of treatment. Methods of determining the most effective means and dosage of administration are known to those of skill in the art and can vary with the composition used for therapy, the purpose of the therapy, the target cell being treated, and the subject being treated. Single or multiple administrations can be carried out with the dose level and pattern being selected by the treating physician. Suitable dosage formulations and methods of administering the agents are known in the art. Route of administration can also be determined and method of determining the most effective route of administration are known to those of skill in the art and can vary with the composition used for treatment, the purpose of the treatment, the health condition or disease stage of the subject being treated, and target cell or tissue. Non-limiting examples of route of administration include oral administration, nasal administration, injection, and topical application.

**[0183]** Administration can refer to methods that can be used to enable delivery of compounds or compositions to the desired site of biological action (such as DNA constructs, viral vectors, or others). These methods can include topical administration (such as a lotion, a cream, an ointment) to an external surface of a surface, such as a skin. These methods can include parenteral administration (including intravenous, subcutaneous, intrathecal, intraperitoneal, intramuscular, intravascular or infusion), oral administration, inhalation administration, intraduodenal administration, rectal administration. In some instances, a subject can administer the composition in the absence of supervision. In some instances, a subject can administer the composition under the supervision of a medical professional (*e.g.*, a physician, nurse, physician’s assistant, orderly, hospice worker, *etc.*). In some cases, a medical professional can administer the composition. In some cases, a cosmetic professional can administer the composition.

**[0184]** Administration or application of a composition disclosed herein can be performed for a treatment duration of at least about 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, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 days consecutive or nonconsecutive days. In some cases, a treatment duration can be from about 1 to about 30 days, from about 2 to about

30 days, from about 3 to about 30 days, from about 4 to about 30 days, from about 5 to about 30 days, from about 6 to about 30 days, from about 7 to about 30 days, from about 8 to about 30 days, from about 9 to about 30 days, from about 10 to about 30 days, from about 11 to about 30 days, from about 12 to about 30 days, from about 13 to about 30 days, from about 14 to about 30 days, from about 15 to about 30 days, from about 16 to about 30 days, from about 17 to about 30 days, from about 18 to about 30 days, from about 19 to about 30 days, from about 20 to about 30 days, from about 21 to about 30 days, from about 22 to about 30 days, from about 23 to about 30 days, from about 24 to about 30 days, from about 25 to about 30 days, from about 26 to about 30 days, from about 27 to about 30 days, from about 28 to about 30 days, or from about 29 to about 30 days.

**[0185]** Administration or application of composition disclosed herein can be performed at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 times a day. In some cases, administration or application of composition disclosed herein can be performed at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 21 times a week. In some cases, administration or application of composition disclosed herein can be performed at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90 times a month.

**[0186]** In some cases, a composition can be administered/applied as a single dose or as divided doses. In some cases, the compositions described herein can be administered at a first time point and a second time point. In some cases, a composition can be administered such that a first administration is administered before the other with a difference in administration time of 1 hour, 2 hours, 4 hours, 8 hours, 12 hours, 16 hours, 20 hours, 1 day, 2 days, 4 days, 7 days, 2 weeks, 4 weeks, 2 months, 3 months, 4 months, 5 months, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 1 year or more.

**[0187]** The term “effective amount” can refer to a quantity sufficient to achieve a desired effect. In the context of therapeutic or prophylactic applications, the effective amount will depend on the type and severity of the condition at issue and the characteristics of the individual subject, such as general health, age, sex, body weight, and tolerance to pharmaceutical compositions. In the context of an immunogenic composition, in some embodiments the effective amount is the amount sufficient to result in a protective response against a pathogen. In other embodiments, the effective amount of an immunogenic



composition is the amount sufficient to result in antibody generation against the antigen. In some embodiments, the effective amount is the amount required to confer passive immunity on a subject in need thereof. With respect to immunogenic compositions, in some embodiments the effective amount can depend on the intended use, the degree of immunogenicity of a particular antigenic compound, and the health/responsiveness of the subject's immune system, in addition to the factors described above. The skilled artisan can determine appropriate amounts depending on these and other factors.

**[0188]** In the case of an *in vitro* application, in some embodiments the effective amount can depend on the size and nature of the application in question. It can also depend on the nature and sensitivity of the *in vitro* target and the methods in use. The skilled artisan can determine the effective amount based on these and other considerations. The effective amount can comprise one or more administrations of a composition depending on the embodiment.

**[0189]** It is to be inferred without explicit recitation and unless otherwise intended, that when the disclosure relates to a polypeptide, protein, polynucleotide or antibody, an equivalent or a biologically equivalent of such is intended within the scope of this disclosure. As used herein, the term "biological equivalent thereof" is intended to be synonymous with "equivalent thereof" when referring to a reference protein, antibody, polypeptide or nucleic acid, intends those having minimal homology while still maintaining desired structure or functionality. Unless specifically recited herein, it is contemplated that any polynucleotide, polypeptide or protein mentioned herein also includes equivalents thereof. For example, an equivalent intends at least about 70% homology or identity, or at least 80 % homology or identity and alternatively, or at least about 85 %, or alternatively at least about 90 %, or alternatively at least about 95 %, or alternatively 98 % percent homology or identity and exhibits substantially equivalent biological activity to the reference protein, polypeptide or nucleic acid. Alternatively, when referring to polynucleotides, an equivalent thereof is a polynucleotide that hybridizes under stringent conditions to the reference polynucleotide or its complement.

**[0190]** The disclosure provides polypeptide and/or polynucleotide sequences for use in gene and protein editing techniques described below. It should be understood, although not always explicitly stated that the sequences provided herein can be used to provide the expression product as well as substantially identical sequences that produce a protein that has the same biological properties. These "biologically equivalent" or "biologically active" polypeptides are encoded by equivalent polynucleotides as described herein. They can possess at least

60%, or alternatively, at least 65%, or alternatively, at least 70%, or alternatively, at least 75%, or alternatively, at least 80%, or alternatively at least 85%, or alternatively at least 90%, or alternatively at least 95% or alternatively at least 98%, identical primary amino acid sequence to the reference polypeptide when compared using sequence identity methods run under default conditions. Specific polypeptide sequences are provided as examples of particular embodiments. Modifications to the sequences to amino acids with alternate amino acids that have similar charge. Additionally, an equivalent polynucleotide is one that hybridizes under stringent conditions to the reference polynucleotide or its complement or in reference to a polypeptide, a polypeptide encoded by a polynucleotide that hybridizes to the reference encoding polynucleotide under stringent conditions or its complementary strand. Alternatively, an equivalent polypeptide or protein is one that is expressed from an equivalent polynucleotide.

**[0191]** A “composition” typically intends a combination of the active agent, *e.g.*, an adRNA of this disclosure, a compound or composition, and a naturally-occurring or non-naturally-occurring carrier, inert (for example, a detectable agent or label) or active, such as an adjuvant, diluent, binder, stabilizer, buffers, salts, lipophilic solvents, preservative, adjuvant or the like and include pharmaceutically acceptable carriers. Carriers also include pharmaceutical excipients and additives proteins, peptides, amino acids, lipids, and carbohydrates (*e.g.*, sugars, including monosaccharides, di-, tri-, tetra-oligosaccharides, and oligosaccharides; derivatized sugars such as alditols, aldonic acids, esterified sugars and the like; and polysaccharides or sugar polymers), which can be present singly or in combination, comprising alone or in combination 1-99.99% by weight or volume. Exemplary protein excipients include serum albumin such as human serum albumin (HSA), recombinant human albumin (rHA), gelatin, casein, and the like. Representative amino acid/antibody components, which can also function in a buffering capacity, include alanine, arginine, glycine, arginine, betaine, histidine, glutamic acid, aspartic acid, cysteine, lysine, leucine, isoleucine, valine, methionine, phenylalanine, aspartame, and the like. Carbohydrate excipients are also intended within the scope of this technology, examples of which include but are not limited to monosaccharides such as fructose, maltose, galactose, glucose, D-mannose, sorbose, and the like; disaccharides, such as lactose, sucrose, trehalose, cellobiose, and the like; polysaccharides, such as raffinose, melezitose, maltodextrins, dextrans, starches, and the like; and alditols, such as mannitol, xylitol, maltitol, lactitol, xylitol sorbitol (glucitol) and myoinositol.

**[0192]** The compositions used in accordance with the disclosure, including cells, treatments, therapies, agents, drugs and pharmaceutical formulations can be packaged in dosage unit form for ease of administration and uniformity of dosage. The term "unit dose" or "dosage" can refer to physically discrete units suitable for use in a subject, each unit containing a predetermined quantity of the composition calculated to produce the desired responses in association with its administration, *i.e.*, the appropriate route and regimen. The quantity to be administered, both according to number of treatments and unit dose, depends on the result and/or protection desired. Precise amounts of the composition also depend on the judgment of the practitioner and are peculiar to each individual. Factors affecting dose include physical and clinical state of the subject, route of administration, intended goal of treatment (alleviation of symptoms versus cure), and potency, stability, and toxicity of the particular composition. Upon formulation, solutions can be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically or prophylactically effective. The formulations are easily administered in a variety of dosage forms, such as the type of injectable solutions described herein.

**[0193]** As used herein, the term "reduce or eliminate expression and/or function of" can refer to reducing or eliminating the transcription of said polynucleotides into mRNA, or alternatively reducing or eliminating the translation of said mRNA into peptides, polypeptides, or proteins, or reducing or eliminating the functioning of said peptides, polypeptides, or proteins. In a non-limiting example, the transcription of polynucleotides into mRNA is reduced to at least half of its normal level found in wild type cells.

**[0194]** The phrase "first line" or "second line" or "third line" can refer to the order of treatment received by a patient. First line therapy regimens are treatments given first, whereas second or third line therapy are given after the first line therapy or after the second line therapy, respectively. The National Cancer Institute defines first line therapy as "the first treatment for a disease or condition. In patients with cancer, primary treatment can be surgery, chemotherapy, radiation therapy, or a combination of these therapies. First line therapy is also referred to those skilled in the art as "primary therapy and primary treatment." See National Cancer Institute website at [cancer.gov](http://cancer.gov), last visited November 15, 2017. Typically, a patient is given a subsequent chemotherapy regimen because the patient did not show a positive clinical or sub-clinical response to the first line therapy or the first line therapy has stopped.

[0195] The term “contacting” means direct or indirect binding or interaction between two or more entities. A particular example of direct interaction is binding. A particular example of an indirect interaction is where one entity acts upon an intermediary molecule, which in turn acts upon the second referenced entity. Contacting as used herein includes in solution, in solid phase, *in vitro*, *ex vivo*, in a cell and *in vivo*. Contacting *in vivo* can be referred to as administering, or administration.

[0196] “Cryoprotectants” are known in the art and include without limitation, *e.g.*, sucrose, trehalose, and glycerol. A cryoprotectant exhibiting low toxicity in biological systems is generally used.

[0197] Disclosed herein are adRNAs for site-specific editing of RNA in the absence of overexpression of the ADAR enzymes. Further provided herein is engineering A -> G editing of DNA. In addition, provided herein is screening for ADAR2 mutants that enable site-specific C ->T editing of RNA and DNA. Still further provided herein is engineering C->T edits of RNA via the use of APOBEC1 expressed along with ACF.

[0198] Compared to other ADAR2 systems, the disclosure is unique as it presents a novel method of recruitment of endogenous ADARs to catalyze therapeutic RNA editing. In addition, none of the prior art systems offer a means to use ADAR enzymes for engineering C -> T edits. Lastly, they do not disclose the use of APOBEC for programmable site-specific RNA editing.

[0199] Disclosed herein is an exemplary adRNA comprises an RNA targeting domain, complementary to the target RNA and one or more ADAR recruiting domain. When bound to its target, the adRNA is able to recruit the ADAR enzyme to the target RNA. This ADAR enzyme is then able to catalyze the conversion of a target adenosine to inosine. Not to be bound by theory, it is believed that adRNA can be used analogously to recruit one of the ADAR2 mutants or APOBEC1 to affect C -> T RNA editing.

[0200] Also disclosed herein, both *in vitro* and *in vivo* experiments have been carried out using the engineered adRNA to recruit the endogenous ADAR enzymes. Also disclosed herein are experiments showing C->T editing efficiencies of ADAR mutants as well as the APOBEC1/ACF constructs.

[0201] A viral vector as described herein can comprise a nucleic acid sequence encoding for at least one RNA editing entity recruiting domain. In some cases, a nucleic acid sequence can encode for more than one RNA editing entity recruiting domain, such as 2, 3, 4 or more. An RNA editing entity recruiting domain can comprise at least about 80% sequence identity to at

least one of: an Alu domain, an Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC) recruiting domain, a Cas13 domain, a GluR2 domain, or any combination thereof. The recruiting domain can comprise one or more GluR2 domains, one or more Alu domains, one or more APOBEC domains, or any combination thereof. The recruiting domain can comprise more than one GluR2 domain, more than one Alu domain, more than one APOBEC domain, Cas13 domain, or any combination thereof. The recruiting domain may not comprise an Alu domain. The recruiting domain may not comprise an GluR2 domain. The recruiting domain may not comprise an APOBEC domain. The recruiting domain may not comprise a Cas13 domain.

**[0202]** An APOBEC recruiting domain can comprise an APOBEC1 recruiting domain, APOBEC2 recruiting domain, APOBEC3A recruiting domain, APOBEC3B recruiting domain, APOBEC3C recruiting domain, APOBEC3D recruiting domain, APOBEC3E recruiting domain, APOBEC3F recruiting domain, APOBEC3G recruiting domain, APOBEC3H recruiting domain, APOBEC4 recruiting domain, a derivative of any of these, or any combination thereof.

**[0203]** A recruiting domain can comprise at least about 80% sequence identity to any one of the Alu domains as described herein. In some cases, the recruiting domain can comprise at least about: 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to any one of the Alu domains as described herein.

**[0204]** A recruiting domain can comprise at least about 80% sequence identity to any one of the APOBEC domains as described herein. In some cases, the recruiting domain can comprise at least about: 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to any one of the APOBEC domains as described herein.

**[0205]** A recruiting domain can comprise at least about 80% sequence identity to any one of the GluR2 domains as described herein. In some cases, the recruiting domain can comprise at least about: 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to any one of the GluR2 domains as described herein.

**[0206]** A recruiting domain can comprise at least about 80% sequence identity to any one of the Cas13 domains as described herein. In some cases, the recruiting domain can comprise at least about: 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to any one of the Cas13 domains as described herein.

**[0207]** A nucleic acid sequence can encode for at least 1, 2, 3, 4, 5, RNA editing recruiting domains. A nucleic acid sequence can encode for at least 2 RNA editing recruiting domains,

wherein one is an Alu domain. A nucleic acid sequence can encode for at least 2 RNA editing recruiting domains, wherein one is an APOBEC domain. A nucleic acid sequence can encode for at least 2 RNA editing recruiting domains, wherein one is a GluR2 domain. A nucleic acid sequence can encode for at least 2 RNA editing recruiting domains, wherein one is a Cas13 domain.

**[0208]** A recruiting domain can comprise one or more stem loop structures. A recruiting domain can comprise at least 2 stem loop structures. A recruiting domain can comprise at least 3 stem loop structures. A recruiting domain may not comprise a stem loop structure. A recruiting domain that comprises at least one stem loop structure can be an Alu domain, an APOBEC domain, a GluR2 domain, Cas13 domain, or any combination thereof.

**[0209]** At least a portion of a recruiting domain can be single stranded. In some cases, an Alu domain can be at least partially single stranded. In some cases, an APOBEC domain can be at least partially single stranded. In some cases, an GluR2 domain can be at least partially single stranded. In some cases, a Cas13 domain can be at least partially single stranded.

**[0210]** A recruiting domain can comprise a plurality of repeats. A recruiting domain can comprise a plurality of Alu repeats.

**[0211]** In some cases, a viral vector can comprise one or more RNA editing recruiting domains. In some cases, a viral vector can comprise more than one RNA editing recruiting domain. In some cases, a viral vector can comprise 2, 3, 4, 5 or more RNA editing recruiting domains. A nucleic acid sequence can encode for one or more RNA editing recruiting domains. A nucleic acid sequence can encode for more than one RNA editing recruiting domain. A nucleic acid sequence can encode for 2, 3, 4, 5 or more RNA editing recruiting domains. A nucleic acid sequence can encode for at least an Alu domain and a GluR2 domain. A nucleic acid sequence can encode for at least an Alu domain and a Cas13 domain. A nucleic acid sequence can encode for at least an Alu domain and an APOBEC domain. A nucleic acid sequence can encode for at least a GluR2 domain and an APOBEC domain. A nucleic acid sequence can encode for at least a GluR2 domain and a Cas13 domain. A nucleic acid sequence can encode for at least a Cas13 domain and an APOBEC domain.

**[0212]** A nucleic acid sequence can encode for a target RNA that can be complementary to at least a portion of a target RNA. It can be complementary to at least a portion of that target RNA. The portion that can be complementary can be from about 50 basepairs (bp) to about 200 bp in length. The portion that can be complementary can be from about 20 bp to about 100 bp in length. The portion that can be complementary can be from about 10 bp to about 50

bp in length. The portion that can be complementary can be from about 50 bp to about 300 bp in length. Modifying a length of the portion that is complementary can enhance efficiency of editing. In some cases, longer lengths of the portion can enhance efficiency of editing as compared to shorter lengths.

**[0213]** A nucleic acid sequence can encode for at least one RNA editing entity recruiting domain and a nucleic acid sequence encoding for an RNA that can be complementary to at least a portion of a target RNA and comprise a contiguous nucleic acid sequence of at least about 200 bp in length. The contiguous nucleic acid sequence can comprise a length from about 100 bp to about 300 bp in length. The contiguous nucleic acid sequence can comprise a length from about 150 bp to about 400 bp in length. The contiguous nucleic acid sequence can comprise a length from about 200 bp to about 500 bp in length. The contiguous nucleic acid sequence can comprise a length from about 50 bp to about 300 bp in length. Modifying a length of the contiguous nucleic acid sequence can enhance efficiency of editing. In some cases, longer lengths of the contiguous sequence can enhance efficiency of editing as compared to shorter lengths.

**[0214]** A nucleic acid can comprise a linker sequence, such as a linker sequence positioned between a targeting domain and a recruiting domain. In some cases, a nucleic acid can comprise a sequence such as 5'-X-(Y-X')<sub>n</sub>-L-Z-3', wherein X is complementary to the target RNA sequence downstream of the specific position, X' is complementary to the target RNA sequence upstream of the specific position, Y comprises one or more nucleotides which may not be complementary to the target RNA sequence, n can be an integer from 1 to 10, L can be a linker sequence comprising any number of nucleotides (including zero), and Z can be a sequence that is recognized by and binds to the RNA editing entity. L can also consist of a different chemical linkage, such as a (oligo)peptide linkage, or PEG linkage.

**[0215]** A nucleic acid can comprise between 20 and several hundred nucleotides. In some cases, longer targeting portions provide more specificity for the target site of the RNA sequence to be edited, less off-target effects due to unintentional (off-target) binding as well as more room to create secondary structures, such as stem-loop structures, cruciforms, toe hold structures, within the targeting portion itself, mismatches or wobble-bases (due to mismatches with one or more of the complementary base(s) in the targeted RNA sequence at or near the site to be edited), and so forth. In some cases, targeting portions can be complementary to the target RNA sequence over the entire length of the targeting portion

except for the mismatch opposite the nucleotide to be edited, and optionally one or two wobble bases.

**[0216]** Nucleic acids can be modified using various chemistries and modifications. In some cases, regular internucleosidic linkages between nucleotides can be altered by mono- or dithioation of the phosphodiester bonds to yield phosphorothioate esters or phosphorodithioate esters, respectively. Other modifications of the internucleosidic linkages can include amidation or peptide linkers. A ribose sugar can be modified by substitution of the 2'-O moiety with a lower alkyl (C1 -4, such as 2'- O-Me), alkenyl (C2-4), alkynyl (C2-4), methoxyethyl (2'-MOE), or other substituent. In some cases, substituents of the 2' OH group can comprise a methyl, methoxyethyl or 3,3'-dimethylallyl group. In some cases, locked nucleic acid sequences (LNAs), comprising a 2'-4' intramolecular bridge (such as a methylene bridge between the 2' oxygen and 4' carbon) linkage inside the ribose ring, can be applied. Purine nucleobases and/or pyrimidine nucleobases can be modified to alter their properties, for example by amination or deamination of the heterocyclic rings.

**[0217]** A viral vector can be an adeno-associated virus (AAV) vector. An AAV can be a recombinant AAV. An AAV can comprise an AAV1 serotype, an AAV2 serotype, an AAV3 serotype, an AAV4 serotype, an AAV5 serotype, an AAV6 serotype, an AAV7 serotype, an AAV8 serotype, an AAV9 serotype, a derivative of any of these, or any combination thereof. An AAV can be selected from the group consisting of: an AAV1 serotype, an AAV2 serotype, an AAV3 serotype, an AAV4 serotype, an AAV5 serotype, an AAV6 serotype, an AAV7 serotype, an AAV8 serotype, an AAV9 serotype, a derivative of any of these, and any combination thereof. A viral vector can be a modified viral vector. A viral vector can be modified to include a modified protein. In some cases, a viral vector can comprise a modified VP1 protein.

**[0218]** A nucleic acid sequence, that encodes for an RNA editing entity recruiting domain, a targeting domain, or a combination thereof, can comprise a structure (such as a secondary structure) that can be substantially a cruciform. A nucleic acid sequence can comprise at least two structures that can be substantially cruciforms. A recruiting domain can comprise a structure that can be substantially a cruciform. A recruiting domain can comprise at least two structures that can be substantially cruciforms. A secondary structure of a nucleic acid sequence (such as a portion encoding a recruiting domain) can be modified to enhance recruitment or binding of an ADAR. Modification of structure to enhance recruitment or binding of an ADAR can include forming cruciform structures.



[0219] An RNA editing entity recruiting domain can be positioned between at least two structures that can be substantially cruciforms. A targeting domain can be positioned between at least two structures that can be substantially cruciforms. An RNA editing entity recruiting domain can be positioned flanked by at least one structure that can be substantially a cruciform. A targeting domain can be positioned flanked by at least one structure that can be substantially a cruciform.

[0220] A cruciform structure can comprise a stem loop adjoining at least one pair of at least partially complementary strands of a cruciform structure. A cruciform structure can be substantially a cruciform. A cruciform structure can comprise less than substantially a cruciform, such as 3 of 4 stem loops, or 2 of 4 stem loops. One or more stem loops that can form a cruciform can comprise a different length. One or more stem loops that can form a cruciform can comprise a same length. One or more stem loops that can form a cruciform can comprise one or more mismatch bulges.

[0221] An RNA editing entity recruiting domain can comprise a structure that can be substantially a toehold. An RNA editing entity can comprise one or more mismatch bulges. An RNA editing entity may not comprise a mismatch bulge. An RNA editing entity recruiting domain can comprise a substantially toehold structure, a substantially cruciform structure, a substantially linear structure, a stem loop structure, a double stem loop structure, or a combination thereof.

[0222] A viral vector can comprise a nucleic acid sequence encoding for an RNA with a two dimensional shape. The two dimensional shape can convey superior recruitment or binding of ADAR as compared to an RNA with a different two dimensional shape. A sequence of a nucleic acid sequence that encodes for an RNA can be modified such that the RNA comprises a two dimensional shape that conveys superior recruitment or binding of ADAR. The two dimensional shape can be substantially a cruciform, a toehold, a stem loop, or any combination thereof. The two dimensional shape can comprise the substantially a cruciform. The two dimensional shape can comprise the toehold. The two dimensional shape can comprise the stem loop. The two dimensional shape can be linear.

[0223] An RNA encoded by a nucleic acid sequence can comprise a first domain and a second domain. A first domain can comprise a cruciform and a second domain can comprise a linear structure. The first and second domain can be directly or indirectly connected. The first domain can be a recruiting domain and the second domain can be a targeting domain. The RNA can comprise a third domain. The third domain can be directly or indirectly

connected to the first or second domains. The third domain can be a recruiting domain. The third domain can comprise a cruciform structure.

**[0224]** An RNA encoded by a nucleic acid sequence can be a non-naturally occurring RNA. An RNA encoded by a nucleic acid sequence can comprise at least one base or at least one sugar comprises a chemical modification. An RNA can comprise two or more chemical modifications. A chemical modification can increase a stability of an RNA, such as a bioactive half-life of the RNA *in vivo*.

**[0225]** A nucleic acid can comprise one or more recruiting domains and one or more antisense domains. When the nucleic acid is contacted with an RNA editing entity and a target nucleic acid complementary to at least a portion of the antisense domain, it can modify at least one base pair of the target nucleic acid at an efficiency of at least about: 2x, 2.5x, 3x, 3.5x, 4x, 4.5x, 5x, 5.5x, or 6x greater than a comparable nucleic acid complexed with a Cas13b protein or an active fragment thereof, as determined by a sequence methods (such as Sanger method). The efficiency can be at least about 3x greater. The efficiency can be at least about 4x greater. The efficiency can be at least about 5x greater.

**[0226]** A nucleic acid can comprise one or more recruiting domains and one or more antisense domains. When the nucleic acid is contacted with an RNA editing entity and a target nucleic acid complementary to at least a portion of the antisense domain, it can modify at least one base pair of the target nucleic acid at an efficiency of at least about: 2x, 2.5x, 3x, 3.5x, 4x, 4.5x, 5x, 5.5x, or 6x greater than a comparable nucleic acid complexed with a GluR2 protein or an active fragment thereof, as determined by a sequence methods (such as Sanger method). The efficiency can be at least about 3x greater. The efficiency can be at least about 4x greater. The efficiency can be at least about 5x greater.

**[0227]** Nucleic acids as described herein can provide greater editing efficiencies than at least a portion of a native recruiting domain, such as a GluR2 domain. Nucleic acids can provide greater editing efficiencies than at least a portion of a modified recruiting domain, such as a modified GluR2 domain.

**[0228]** A target nucleic acid can comprise RNA. The RNA can be mRNA. The RNA can encode a protein or a portion thereof. A dysfunction of the protein or portion thereof can be implicated in a disease or condition. Administration of a composition, a vector, a nucleic acid, a non-naturally occurring RNA as described herein can treat, eliminate, cure, or reduce one or more symptoms of the disease or condition.

**[0229]** A disease or condition can comprise a neurodegenerative disease, a muscular disorder, a metabolic disorder, an ocular disorder, or any combination thereof. The disease or condition can comprise cystic fibrosis, albinism, alpha-1 -antitrypsin deficiency, Alzheimer disease, Amyotrophic lateral sclerosis, Asthma,  $\beta$ -thalassemia, Cadasil syndrome, Charcot-Marie-Tooth disease, Chronic Obstructive Pulmonary Disease (COPD), Distal Spinal Muscular Atrophy (DSMA), Duchenne/Becker muscular dystrophy, Dystrophic Epidermolysis bullosa, Epidermylosis bullosa, Fabry disease, Factor V Leiden associated disorders, Familial Adenomatous, Polyposis, Galactosemia, Gaucher's Disease, Glucose-6-phosphate dehydrogenase, Haemophilia, Hereditary Hematochromatosis, Hunter Syndrome, Huntington's disease, Hurler Syndrome, Inflammatory Bowel Disease (IBD), Inherited polyagglutination syndrome, Leber congenital amaurosis, Lesch-Nyhan syndrome, Lynch syndrome, Marfan syndrome, Mucopolysaccharidosis, Muscular Dystrophy, Myotonic dystrophy types I and II, neurofibromatosis, Niemann-Pick disease type A, B and C, NY-eso1 related cancer, Parkinson's disease, Peutz-Jeghers Syndrome, Phenylketonuria, Pompe's disease, Primary Ciliary Disease, Prothrombin mutation related disorders, such as the Prothrombin G20210A mutation, Pulmonary Hypertension, Retinitis Pigmentosa, Sandhoff Disease, Severe Combined Immune Deficiency Syndrome (SCID), Sickle Cell Anemia, Spinal Muscular Atrophy, Stargardt's Disease, Tay-Sachs Disease, Usher syndrome, X-linked immunodeficiency, various forms of cancer (*e.g.* BRCA1 and 2 linked breast cancer and ovarian cancer). The disease or condition can comprise a muscular dystrophy, an ornithine transcarbamylase deficiency, a retinitis pigmentosa, a breast cancer, an ovarian cancer, Alzheimer's disease, pain, Stargardt macular dystrophy, Charcot-Marie-Tooth disease, Rett syndrome, or any combination thereof. Administration of a composition can be sufficient to: (a) decrease expression of a gene relative to an expression of the gene prior to administration; (b) edit at least one point mutation in a subject, such as a subject in need thereof; (c) edit at least one stop codon in the subject to produce a readthrough of a stop codon; (d) produce an exon skip in the subject, or (e) any combination thereof.

**[0230]** A pharmaceutical composition can comprise a first active ingredient. The first active ingredient can comprise a viral vector as described herein, a non-naturally occurring RNA as described herein, or a nucleic acid as described herein. The pharmaceutical composition can be formulated in unit dose form. The pharmaceutical composition can comprise a pharmaceutically acceptable excipient, diluent, or carrier. The pharmaceutical composition can comprise a second, third, or fourth active ingredient.

**[0231]** A composition described herein can comprise an excipient. An excipient can be added to a stem cell or can be co-isolated with the stem cell from its source. An excipient can comprise a cryo-preserved, such as DMSO, glycerol, polyvinylpyrrolidone (PVP), or any combination thereof. An excipient can comprise a cryo-preserved, such as a sucrose, a trehalose, a starch, a salt of any of these, a derivative of any of these, or any combination thereof. An excipient can comprise a pH agent (to minimize oxidation or degradation of a component of the composition), a stabilizing agent (to prevent modification or degradation of a component of the composition), a buffering agent (to enhance temperature stability), a solubilizing agent (to increase protein solubility), or any combination thereof. An excipient can comprise a surfactant, a sugar, an amino acid, an antioxidant, a salt, a non-ionic surfactant, a solubilizer, a triglyceride, an alcohol, or any combination thereof. An excipient can comprise sodium carbonate, acetate, citrate, phosphate, poly-ethylene glycol (PEG), human serum albumin (HSA), sorbitol, sucrose, trehalose, polysorbate 80, sodium phosphate, sucrose, disodium phosphate, mannitol, polysorbate 20, histidine, citrate, albumin, sodium hydroxide, glycine, sodium citrate, trehalose, arginine, sodium acetate, acetate, HCl, disodium edetate, lecithin, glycerine, xanthan rubber, soy isoflavones, polysorbate 80, ethyl alcohol, water, teprenone, or any combination thereof. An excipient can be an excipient described in the Handbook of Pharmaceutical Excipients, American Pharmaceutical Association (1986).

**[0232]** Non-limiting examples of suitable excipients can include a buffering agent, a preservative, a stabilizer, a binder, a compaction agent, a lubricant, a chelator, a dispersion enhancer, a disintegration agent, a flavoring agent, a sweetener, a coloring agent.

**[0233]** In some cases, an excipient can be a buffering agent. Non-limiting examples of suitable buffering agents can include sodium citrate, magnesium carbonate, magnesium bicarbonate, calcium carbonate, and calcium bicarbonate. As a buffering agent, sodium bicarbonate, potassium bicarbonate, magnesium hydroxide, magnesium lactate, magnesium glucomate, aluminium hydroxide, sodium citrate, sodium tartrate, sodium acetate, sodium carbonate, sodium polyphosphate, potassium polyphosphate, sodium pyrophosphate, potassium pyrophosphate, disodium hydrogen phosphate, dipotassium hydrogen phosphate, trisodium phosphate, tripotassium phosphate, potassium metaphosphate, magnesium oxide, magnesium hydroxide, magnesium carbonate, magnesium silicate, calcium acetate, calcium glycerophosphate, calcium chloride, calcium hydroxide and other calcium salts or combinations thereof can be used in a pharmaceutical formulation.

**[0234]** In some cases, an excipient can comprise a preservative. Non-limiting examples of suitable preservatives can include antioxidants, such as alpha-tocopherol and ascorbate, and antimicrobials, such as parabens, chlorobutanol, and phenol. Antioxidants can further include but not limited to EDTA, citric acid, ascorbic acid, butylated hydroxytoluene (BHT), butylated hydroxy anisole (BHA), sodium sulfite, p-amino benzoic acid, glutathione, propyl gallate, cysteine, methionine, ethanol and N- acetyl cysteine. In some instances a preservatives can include validamycin A, TL-3, sodium ortho vanadate, sodium fluoride, N-a-tosyl-Phe- chloromethylketone, N-a-tosyl-Lys-chloromethylketone, aprotinin, phenylmethylsulfonyl fluoride, diisopropylfluorophosphate, kinase inhibitor, phosphatase inhibitor, caspase inhibitor, granzyme inhibitor, cell adhesion inhibitor, cell division inhibitor, cell cycle inhibitor, lipid signaling inhibitor, protease inhibitor, reducing agent, alkylating agent, antimicrobial agent, oxidase inhibitor, or other inhibitor.

**[0235]** In some cases, a pharmaceutical formulation can comprise a binder as an excipient. Non-limiting examples of suitable binders can include starches, pregelatinized starches, gelatin, polyvinylpyrrolidone, cellulose, methylcellulose, sodium carboxymethylcellulose, ethylcellulose, polyacrylamides, polyvinylloxazolidone, polyvinylalcohols, C12-C18 fatty acid alcohol, polyethylene glycol, polyols, saccharides, oligosaccharides, and combinations thereof.

**[0236]** The binders that can be used in a pharmaceutical formulation can be selected from starches such as potato starch, corn starch, wheat starch; sugars such as sucrose, glucose, dextrose, lactose, maltodextrin; natural and synthetic gums; gelatine; cellulose derivatives such as microcrystalline cellulose, hydroxypropyl cellulose, hydroxyethyl cellulose, hydroxypropyl methyl cellulose, carboxymethyl cellulose, methyl cellulose, ethyl cellulose; polyvinylpyrrolidone (povidone); polyethylene glycol (PEG); waxes; calcium carbonate; calcium phosphate; alcohols such as sorbitol, xylitol, mannitol and water or a combination thereof.

**[0237]** In some cases, a pharmaceutical formulation can comprise a lubricant as an excipient. Non-limiting examples of suitable lubricants can include magnesium stearate, calcium stearate, zinc stearate, hydrogenated vegetable oils, sterotex, polyoxyethylene monostearate, talc, polyethyleneglycol, sodium benzoate, sodium lauryl sulfate, magnesium lauryl sulfate, and light mineral oil. The lubricants that can be used in a pharmaceutical formulation can be selected from metallic stearates (such as magnesium stearate, calcium stearate, aluminium stearate), fatty acid esters (such as sodium stearyl fumarate), fatty acids (such as stearic acid),

fatty alcohols, glyceryl behenate, mineral oil, paraffins, hydrogenated vegetable oils, leucine, polyethylene glycols (PEG), metallic lauryl sulphates (such as sodium lauryl sulphate, magnesium lauryl sulphate), sodium chloride, sodium benzoate, sodium acetate and talc or a combination thereof.

**[0238]** In some cases, a pharmaceutical formulation can comprise a dispersion enhancer as an excipient. Non-limiting examples of suitable dispersants can include starch, alginic acid, polyvinylpyrrolidones, guar gum, kaolin, bentonite, purified wood cellulose, sodium starch glycolate, isoamorphous silicate, and microcrystalline cellulose as high HLB emulsifier surfactants.

**[0239]** In some cases, a pharmaceutical formulation can comprise a disintegrant as an excipient. In some cases, a disintegrant can be a non-effervescent disintegrant. Non-limiting examples of suitable non-effervescent disintegrants can include starches such as corn starch, potato starch, pregelatinized and modified starches thereof, sweeteners, clays, such as bentonite, micro-crystalline cellulose, alginates, sodium starch glycolate, gums such as agar, guar, locust bean, karaya, pectin, and tragacanth. In some cases, a disintegrant can be an effervescent disintegrant. Non-limiting examples of suitable effervescent disintegrants can include sodium bicarbonate in combination with citric acid, and sodium bicarbonate in combination with tartaric acid.

**[0240]** In some cases, an excipient can comprise a flavoring agent. Flavoring agents incorporated into an outer layer can be chosen from synthetic flavor oils and flavoring aromatics; natural oils; extracts from plants, leaves, flowers, and fruits; and combinations thereof. In some cases, a flavoring agent can be selected from the group consisting of cinnamon oils; oil of wintergreen; peppermint oils; clover oil; hay oil; anise oil; eucalyptus; vanilla; citrus oil such as lemon oil, orange oil, grape and grapefruit oil; and fruit essences including apple, peach, pear, strawberry, raspberry, cherry, plum, pineapple, and apricot.

**[0241]** In some cases, an excipient can comprise a sweetener. Non-limiting examples of suitable sweeteners can include glucose (corn syrup), dextrose, invert sugar, fructose, and mixtures thereof (when not used as a carrier); saccharin and its various salts such as a sodium salt; dipeptide sweeteners such as aspartame; dihydrochalcone compounds, glycyrrhizin; Stevia Rebaudiana (Stevioside); chloro derivatives of sucrose such as sucralose; and sugar alcohols such as sorbitol, mannitol, xylitol, and the like.

**[0242]** In one aspect, this disclosure, the adRNA, helps recruit endogenous ADAR enzymes to a target mRNA and bring about site specific A-to-G editing. This has immense potential

for gene therapy wherein the delivery of a single adRNA can potentially correct G-to-A point mutations. It also enables target RNA editing without having to overexpress RNA editing enzymes such as the ADAR2. The disclosure demonstrates the applicability of this technology both *in vitro* and *in vivo*. This disclosure also demonstrates that by the creation of a long double stranded RNA, it is possible to recruit endogenous ADARs even in the absence of the ADAR recruiting domains. In addition, using these engineered adRNAs, it is possible to create multiple A-to-G edits in the mRNA in a target region. In one aspect, the system uses a U6 promoter (polIII) transcribed adRNAs as well as chemically synthesized adRNAs and there was shown to be efficient RNA editing. Thus, in one aspect, the constructs further comprise a promoter, such as a polII promoter, to transcribe adRNAs. Transcription from promoter such as a polIII promoter can improve target RNA editing efficiencies. Alu transcripts from a polIII promoter are preferentially edited. Also provided herein are engineered adRNA from the structure of Alu repeats that are targets for the endogenous ADARs.

**[0243]** The constructs of this disclosure can be used to localize adRNA to specific cellular compartments. For example, for nuclear localization, one can use adRNA-snRNA fusions. Similarly, by adding the N-terminal mitochondrial targeting sequence (MTS) it is possible to localize the adRNA to the mitochondria. Thus, in one aspect, the constructs further comprise the N-terminal mitochondrial targeting sequence (MTS). By the addition of an appropriate cis-acting zipcode, it is possible to localize adRNA into peroxisomes, endosomes and exosomes. Thus, in a further aspect, the constructs further comprise the appropriate cis-acting zipcode for localizing adRNA into peroxisomes, endosomes and exosomes. Localization of adRNA into endosomes can likely enable their transport across long distances in the case of neurons. Localization in exosomes can likely potentially help propagate adRNA to neighboring cells. Tethering moieties such as cholesterol to the adRNA can help in cellular uptake. Thus, in one aspect the constructs further comprise targeting moieties such as cholesterol.

**[0244]** In one aspect, the disclosure demonstrates that to create small molecule regulatable adRNAs, adRNA-aptamer are disclosed to be used in one aspect, to enable temporal control of RNA editing *e.g.* aptamers that bind flavin mononucleotide, guanine and other natural metabolites. Aptamers that bind sugars can also be used for this purpose.

[0245] In one aspect, the creation of a U1A-ADAR fusion is entirely of human origin. The N-terminal RNA recognition motif of the splicesomal U1A protein binds to its cognate U1 hairpin II RNA with a dissociation constant of 63 nM.

[0246] The disclosure also provides constructs that further comprise a toehold.

[0247] The constructs of this disclosure can, in one aspect, be used in the absence of overexpression of the ADAR enzyme.

[0248] Thus, in certain aspects, the adRNA of this disclosure have certain components: a RNA targeting domain, from about 15 to about 200 base pairs in length (and ranges therebetween), which is complementary to the target RNA; 0-10 ADAR recruiting domains which can be derived from the GluR2 mRNA, Alu repeat elements or other RNA motifs that the ADAR binds to; and a cytosine mismatch required to direct the ADAR to the target adenosine which might be present anywhere in the targeting domain. When this adRNA binds to its target RNA, it recruits the ADAR enzyme to the target RNA. This ADAR enzyme now can catalyze the conversion of a target adenosine to inosine. For adRNAs of lengths over 50 base pairs, when expressed in HEK 293T and HeLa, the adRNAs cells can recruit ADARs even in the absence of the ADAR recruiting domain and enable significant levels of target RNA editing. A single adRNA can also be used to create multiple A-to-G edits in the target mRNA. In addition, by utilizing multiple adRNA it is possible to edit multiple different adRNA in the same cell. For example, in the *mdxmdx* mouse model of Duchenne muscular dystrophy, it is possible to not only correct the mutation in dystrophin, but also disrupt the mRNA sequences of genes coding for proteins involved in nonsense mediated decay. Another application is the use of this technology to create loss of function, gain of function and dominant negative mutations and in one aspect, can be used for cancer screens, tumor progression as well as immunoediting studies.

#### **Engineered adRNA**

[0249] Provided herein is an engineered ADAR1 or ADAR2 guide RNA ("adRNA") comprising, or alternatively consisting essentially of, or yet further consisting of: a sequence complementary to a target RNA. In one particular aspect, the engineered adRNA of this disclosure further comprises, or alternatively consists essentially of, or yet further consists of a sequence complementary to ornithine transcarbamylase.

[0250] In one aspect, the engineered adRNA of this disclosure further comprises, or alternatively consists essentially of, or yet further consists of an ADAR2 recruiting domain derived from GluR2 mRNA. In another aspect, the engineered adRNA of this disclosure



further comprises, or alternatively consists essentially of, or yet further consists of ADAR1 recruiting domain derived from Alu repeats. In a further aspect, the engineered adRNA of this disclosure further comprises, or alternatively consists essentially of, or yet further consists of two MS2 hairpins flanking the sequence complementary to a target RNA. In some embodiments, the sequence complementary to a target RNA in the engineered adRNA of this disclosure comprises, or alternatively consists essentially of, or yet further consists of about 15 to 30 base pairs, or about 30 to 45 base pairs, or about 45 to 60 base pairs, or about 60 to 75 base pairs, or about 75 to 90 base pairs, or about 90 to 105 base pairs, or about 105 to 120 base pairs, or about 120 to 135 base pairs, or about 135 to 150 base pairs, or about 150 to 165 base pairs, or about 165 to 180 base pairs, or about 180 to 200 base pairs. In a further aspect, it is from about 40 to about 200, or about 50 to about 200, or from about 60 to about 200, or from about 70 to about 200, or from about 80 to about 200, or from about 90 to about 200, or from about 100 to about 200, base pairs.

**[0251]** Disclosed herein is an engineered adRNA comprising, or alternatively consisting essentially of, or yet further consisting of no ADAR recruiting domains, or about 1-2 ADAR recruiting domains, or about 2-3 ADAR recruiting domains, or about 3-4 ADAR recruiting domains, or about 4-5 ADAR recruiting domains, or about 5-6 ADAR recruiting domains, or about 6-7 ADAR recruiting domains, or about 7-8 ADAR recruiting domains, or about 8-9 ADAR recruiting domains, or about 9-10 ADAR recruiting domains. In some embodiments, the ADAR recruiting domains comprise, or alternatively consist essentially of, or yet further consist of GluR2 mRNA, Alu repeat elements or other RNA motifs to which ADAR binds. Also, provided herein is an engineered adRNA, wherein the ADAR2 recruiting domain of the engineered adRNA derived from GluR2 mRNA is located at the 5' end or the 3' end of the engineered adRNA. In some embodiments, the GluR2 mRNA is located at both the 5' end and the 3' end of the engineered adRNA.

**[0252]** In one aspect, the engineered adRNA of this disclosure, further comprises, or alternatively consists essentially of, or yet further consists of an editing inducer element. An "editing inducer element" can refer to a structure that is largely a double-stranded RNA, which is necessary for efficient RNA editing. Non-limiting examples of editing inducer elements are described in Daniel, C. *et al.* (2017) *Genome Biol.* 18, 195.

**[0253]** In one particular aspect, the engineered adRNA of this disclosure is encoded by a polynucleotide sequence selected from the group of sequences provided in **TABLE 1** or **FIG. 2**, or an equivalent of each thereof.

[0254] Also disclosed herein is a complex comprising, or alternatively consisting essentially of, or yet further consisting of an engineered adRNA of this disclosure hybridized to a complementary polynucleotide under conditions of high stringency.

[0255] The disclosure also provides polypeptide and/or polynucleotide sequences for use in gene and protein editing techniques described below. It should be understood, although not always explicitly stated that the sequences provided herein can be used to provide the expression product as well as substantially identical sequences that produce a protein that has the same biological properties. These “biologically equivalent” or “biologically active” polypeptides are encoded by equivalent polynucleotides as described herein. They can possess at least 60%, or alternatively, at least 65%, or alternatively, at least 70%, or alternatively, at least 75%, or alternatively, at least 80%, or alternatively at least 85%, or alternatively at least 90%, or alternatively at least 95% or alternatively at least 98%, identical primary amino acid sequence to the reference polypeptide when compared using sequence identity methods run under default conditions. Specific polypeptide sequences are provided as examples of particular embodiments. Modifications to the sequences to amino acids with alternate amino acids that have similar charge. Additionally, an equivalent polynucleotide is one that hybridizes under stringent conditions to the reference polynucleotide or its complement or in reference to a polypeptide, a polypeptide encoded by a polynucleotide that hybridizes to the reference encoding polynucleotide under stringent conditions or its complementary strand. Alternatively, an equivalent polypeptide or protein is one that is expressed from an equivalent polynucleotide.

[0256] Also disclosed herein is an engineered adRNA-sRNA (small nuclear RNA) fusion. In one aspect, the engineered adRNA further comprises, or alternatively consists essentially of, or yet further consists of an N-terminal mitochondrial targeting sequence (MTS) to facilitate localization of the engineered adRNA to the mitochondria. In another aspect, provided herein is an engineered further comprising, or alternatively consisting essentially of, or yet further consisting of a cis-acting zipcode to facilitate localization of the engineered adRNA into peroxisomes, endosomes and exosomes. Localization of adRNA into endosomes can potentially enable their transport across long distances in the case of neurons. Localization in exosomes can potentially help propagate adRNA to neighboring cells. Tethering moieties such as cholesterol to the adRNA can help in cellular uptake.

[0257] Further provided herein is small molecule regulatable engineered adRNA. In one aspect, disclosed herein are engineered adRNA-aptamer fusions. Non-limiting examples of

aptamers that can be used for this purpose include aptamers that bind flavin mononucleotide, guanine, other natural metabolites, or sugars. An “aptamer” can refer to a short single-stranded oligonucleotide capable of binding various molecules with high affinity and specificity. Non-limiting examples of aptamers are described in Lakhin, A. V. *et al.* (2013). *Acta naturae*, 5(4), 34-43.

[0258] Also disclosed herein is a U1A-ADAR fusion, entirely of human origin. The N-terminal RNA recognition motif of the splicesomal U1A protein binds to its cognate U1 hairpin II RNA with a dissociation constant of 63 nM.

#### **Vectors and Recombinant Cells Expressing the Engineered adRNA**

[0259] Provided herein is a vector comprising, or alternatively consisting essentially of, or yet further consisting of one or more of the isolated polynucleotide sequence encoding the engineered adRNA of this disclosure and optionally regulatory sequences operatively linked to the isolated polynucleotide. Non-limiting examples of a vector include a plasmid or a viral vector such as a retroviral vector, a lentiviral vector, an adenoviral vector, or an adeno-associated viral vector. The vectors can further comprise targeting sequences, zip codes or toeholds, as known in the art.

[0260] In one aspect, the regulatory sequences comprise, or alternatively consist essentially of, or yet further consist of a promoter, an enhancer element and/or a reporter. In some embodiments, the promoter is a human U6, a mouse U6 promoter, a CMV promoter, or a polIII promoter, or a polII promoter. In one aspect, the vector further comprises, or alternatively consists essentially of, or yet further consists of a detectable marker or a purification marker.

[0261] Further disclosed herein is a recombinant cell further comprising or alternatively consisting essentially of, or yet further consisting of the vector described above, wherein the engineered adRNA is recombinantly expressed.

#### **Compositions of the Engineered adRNA**

[0262] Disclosed herein is a composition comprising, or alternatively consisting essentially of, or yet further consisting of a carrier and one or more of the engineered adRNA of this disclosure, the isolated polynucleotide encoding the engineered adRNA of this disclosure, the vector expressing the engineered adRNA of this disclosure, or the recombinant cell expressing the engineered adRNA of this disclosure. In one aspect, the carrier is a pharmaceutically acceptable carrier or a solid support. In a further aspect, the composition

further comprises, or alternatively consists essentially of, or yet further consists of a chemotherapeutic agent or drug.

#### **Methods of Using the Engineered adRNAs**

[0263] Provided herein is a method of modifying protein expression comprising, or alternatively consisting essentially of, or yet further consisting of contacting a polynucleotide encoding the protein, the expression of which is to be modified, with the engineered adRNA of this disclosure.

[0264] Also provided herein is a method of treating a disease or disorder associated with aberrant protein expression comprising, or alternatively consisting essentially of, or yet further consisting of administering to a subject in need of such treatment an effective amount of one or more of the engineered adRNA of this disclosure. In one particular aspect, provided herein is a method of treating Duchenne Muscular Dystrophy comprising, or alternatively consisting essentially of, or yet further consisting of administering to a subject in need of such treatment an effective amount of one or more of the engineered adRNA of this disclosure.

[0265] In the case of an *in vitro* application, in some embodiments the effective amount can depend on the size and nature of the application in question. It can also depend on the nature and sensitivity of the *in vitro* target and the methods in use. The skilled artisan can determine the effective amount based on these and other considerations. The effective amount can comprise one or more administrations of a composition depending on the embodiment.

[0266] The term “subject,” “host,” “individual,” and “patient” are as used interchangeably herein to refer to animals, typically mammalian animals. Any suitable mammal can be treated by a method, cell or composition described herein. Non-limiting examples of mammals include humans, non-human primates (*e.g.*, apes, gibbons, chimpanzees, orangutans, monkeys, macaques, and the like), domestic animals (*e.g.*, dogs and cats), farm animals (*e.g.*, horses, cows, goats, sheep, pigs) and experimental animals (*e.g.*, mouse, rat, rabbit, guinea pig). In some embodiments a mammal is a human. A mammal can be any age or at any stage of development (*e.g.*, an adult, teen, child, infant, or a mammal in utero). A mammal can be male or female. A mammal can be a pregnant female. In some embodiments a subject is a human. In some embodiments, a subject has or is suspected of having a cancer or neoplastic disorder. In other embodiments, a subject has or is suspected of having a disease or disorder associated with aberrant protein expression.

[0267] Referring to **FIG. 13**, in contrast to antisense oligonucleotide (AON) designs (left side of **FIG. 13**) which can comprise short ssRNA (such as about 35 nt in length), exemplary constructs of the disclosure (right side of **FIG. 13**) can be long ssRNA having a length from about 60 bp to about 100 bp with superior target specificity and a total sequence length of from about 150 nt to about 250 nt in length. Constructs as described herein can include optimized and true hairpin structures – as opposed to mismatched bases that can be added to create hairpin-like RNA bulges, as shown on the left side of **FIG. 13** for the AON design. Advantages of true hairpin designs of the constructs include optimal and superior ADAR recruiting efficiency, resulting in higher on target editing yield – as compared to an AON construct. Hairpin designs can be completely independent from the mRNA target sequence and can be easily deployable for any new mRNA target. The target site for deamination can be unique, precise, and without need for undesired chemical modification, and without risk of undesired deamination at other sites. In contrast, AON designs utilizing a mismatch hairpin-like RNA bulge (such as show on the left side of **FIG. 13**), can be (a) limited efficacy in recruiting ADAR (not a true hairpin, structure can be too short); (b) prevented from reuse of sequence and bulges (need a unique design for each new target mRNA); (c) decreased specificity to a cell mRNA target site due to mismatches (increased risk of target damages), (d) required for 2'OMe modification in bulges to protect other Adenines from ADAR deamination activity (can cause undesired mutation at the wrong ribonucleotide).

**[0268] Recruitment of exogenous and endogenous ADARs via long-antisense-adRNAs**

[0269] The CRISPR/Cas9 system is widely used in research but concerns over its *in vivo* applications exist. The two main concerns are the permanent edits that are made on the genome, and the immune response that can likely result from introducing a system that is bacterial in origin. As a result, RNA editing has gained interest as a potential solution for both challenges. The first challenge can be easily overcome by the application of RNA editing; RNAs transcribed from genes can be transient and edits will not permanently alter the cells. However, this can create a new problem in the form of decreased editing efficiencies, since many RNAs can be edited to achieve a phenotypic effect. Additionally, the second challenge of immune response may not be necessarily overcome. The ADAR family of enzymes that edit RNA are human in origin. ADAR1 in particular is expressed nearly ubiquitously in many cell types; there can be many potential therapeutic applications by harnessing its natural preference for RNA A-to-I (A-to-G) editing. This strategy also overcomes the significant hurdle of delivery, since a small guide RNA can be much simpler

to transport than a large bulky enzyme. Therefore, the goal of the methods and compositions as described herein can include engineering a guide (referred to as adRNA) to recruit endogenous ADARs. ADARs can prefer to edit regions of double stranded RNA, particularly near sequences with secondary structures. Thus, engineering guides with increasing length of the antisense domain can be advantageous. The adRNA can have three components: (1) a RNA targeting domain, (such as from about 15 base pairs to about 200 base pairs in length), which can be complementary to the target RNA; (2) from about 0 to about 10 ADAR recruiting domains (which can be derived from the GluR2 mRNA, Alu repeat elements or other RNA motifs) to which the ADAR binds; and (3) a cytosine mismatch which can be required to direct the ADAR to the target adenosine which can be present anywhere in the targeting domain.

**[0270]** When this adRNA binds to its target RNA, it can recruit the ADAR enzyme to the target RNA. This ADAR enzyme can catalyze the conversion of a target adenosine to inosine. Interestingly, long adRNAs of lengths over 50 base pairs, when expressed in HEK 293T and HeLa cells can recruit ADARs even in the absence of the ADAR recruiting domain and can enable significant levels of target RNA editing. A single adRNA can also be used to create multiple A-to-G edits in the target mRNA. In addition, by utilizing multiple adRNA it can be possible to edit multiple different mRNA in the same cell. For example, in the *mdx* mouse model of Duchenne muscular dystrophy, it can be possible to not only correct the mutation in dystrophin, but also can disrupt the mRNA sequences of genes coding for proteins involved in nonsense mediated decay. Another application can be the use of this technology to create loss of function, gain of function, dominant negative mutations, or any combination thereof. This can be of paramount importance in cancer screens - tumor progression as well as immunoediting studies.

**[0271]** Referring to **FIG. 16**, a schematic shows RNA editing via recruitment of endogenous ADARs in the presence of adRNA. These adRNA can be delivered either as chemically modified RNA or as U6 transcribed RNA. Referring to **FIG. 17**, a U6 promoter transcribed adRNAs with progressively longer antisense domain lengths, in combination with zero, one or two GluR2 domains are evaluated for their ability to induce targeted RNA editing with or without exogenous ADAR2 expression. Values represent mean  $\pm$  SEM (n=3). Long adRNA can recruit endogenous ADARs for RNA editing. Referring to **FIG. 18**, chemically synthesized adRNAs versions are tested against a panel of mRNAs with or without exogenous ADAR2 expression. The exact chemical modifications are stated in the figure

along with the source of adRNA. Values represent mean  $\pm$  SEM (n=3). Referring to **FIG. 19**, *in vivo* RNA correction efficiencies in the correctly spliced OTC mRNA in the livers of treated adult *spf<sup>ash</sup>* mice (retro-orbital injections). RNA editing levels of 0.6% are seen in mice injected with U6 transcribed short adRNA.

**[0272] Recruitment of exogenous and endogenous ADARs via Alu-adRNAs**

**[0273]** Alu genes are a transposable element in the genome and are a natural target of RNA editing by ADARs. An adRNA can be designed based on an Alu element structure to enable editing by endogenously recruited ADARs. Various positions on the Alu element structure can be tested where the native sequence can be replaced with the antisense sequence complementary to the target. Single stranded linker region can be selected, such as shown in **FIG. 20**. Length of the antisense guide can be optimized, varying the length from about 20 to about 100 bases, targeting the RAB7A locus. Each guide can be designed to include one or more mismatches. A mismatch can be positioned between the cytosine of the antisense and the target adenosine base to be edited. This mismatch can be positioned in the middle of each antisense length. Editing efficiencies can be compared with an 100-50 antisense guide which can recruit both ADAR1 and ADAR2 and the GluR2 20-6 guide that can only recruits ADAR2.

**[0274]** Referring to **FIG. 20**, a design of the Alu adRNA is shown. Left: a structure of an Alu element. Middle: a design incorporating a locus-specific antisense sequence with a C mismatch opposite the target A. Right: recruitment of the RNA editing enzyme ADAR to the target.

**[0275]** These guides are tested in 293FT cells by transfection with lipofectamine. Each guide is tested in cells overexpressing either ADAR1p110, ADAR1p150, ADAR2, or no enzyme overexpression to test the adRNA's ability to recruit endogenous ADARs. After 48 hours post-transfection, the cells are harvested, RNA is extracted, is converted to cDNA, and the RAB7A locus is amplified for Sanger sequencing. Editing efficiencies are calculated as a ratio of peak height. **FIG. 5** shows the results of the Alu guide length experiment.

**[0276]** Referring to **FIG. 5**, the long Alu-v2-100-50 guide shows improved editing in cells where no ADAR enzyme is overexpressed. The overexpression of ADAR1p150 results in significantly higher editing rates for the Alu constructs while editing rates are similar for the linear 100-50 guide. ADAR1p150 can preferentially binds Z-RNA due to its extra ds-RNA binding domain than the shorter isoform, ADAR1p110, lacks. The Alu elements with their

high GC content are known to form Z-DNA and Z-RNA, aiding in the recruitment of ADAR1p150.

**[0277] Split-ADAR2 deaminase domain (DD)**

**[0278]** Overexpression of ADARs can lead to several transcriptome wide off-target edits. The ability to restrict the catalytic activity of the ADAR2 DD only to the target mRNA can reduce the number of off-targets. Creation of a split-ADAR2 DD can be one potential approach to reduce the number of off-targets. Split-protein reassembly or protein fragment complementation can be a widely used approach to study protein-protein interactions. Splitting the ADAR2 DD can be designed in such a way that each fragment of the split-ADAR2 DD can be catalytically inactive by itself. However, in the presence of the adRNA, the split halves can dimerize to form a catalytically active enzyme at the intended mRNA target.

**[0279]** Regions for splitting a protein can be identified by studying the crystal structure of the ADAR2 DD in complex with its naturally occurring substrates, understanding solvent accessibility scores, using predictive software(s), or any combination thereof. MS2-MCP systems and boxB-lambda N systems (that can efficiently recruit ADARs) can be utilized alone or in combination to recruit the N and C terminals of the split ADAR2 DD respectively. The adRNA can comprise one MS2 stem loop and one boxB hairpin along with an antisense domain complementary to the target. This can enable recruitment of the N and C terminals of the split ADAR2 DD at the target and thereby can constitute a catalytically active DD.

**[0280]** Referring to **FIG. 22** and **FIG. 23**, a schematic of the split-ADAR2 DD system is shown and an exemplary sequence of the ADAR2 DD with sites for splitting highlighted.

**[0281]** Referring to **FIG. 24**, pairs of fragments 1-16 can be assayed via a cypridina luciferase reporter (cluc W85X). Fragments 9 and 10 show the highest activity. The split positions corresponding to fragments 9 and 10 are circled in blue in **FIG. 23**.

**[0282]** Referring to **FIG. 25**, fragments 9 and 10 assayed against the Cluc reporter. Further, a NES-MCP-AD2-C-U-variant can also be tested for A-G editing. N: N-terminal fragment, C: C-terminal fragment, M-M: MS2-MS2 adRNA, M-B: MS2-BoxB adRNA, B-B: BoxB-BoxB adRNA

**[0283]** Further completely humanized versions of these constructs can be created by harnessing human RNA binding proteins, such as (a) U1A or (b) its evolved variant TBP6.7 which has no known endogenous human hairpin targets or (c) the human histone stem loop binding protein (SLBP) or (d) the DNA binding domain of glucocorticoid receptor, or (e) any



combination thereof. These proteins can be fused to the N and C terminal fragments of the ADAR2 to create a completely human and programmable RNA editing toolset that can edit adenosines with exquisite specificity. Further, chimeric RNA bearing two of the corresponding RNA hairpins can be utilized to recruit the ADAR2 fragments. Sequences of the RNA hairpins are provided herein.

**[0284]** A C-U RNA editing enzyme can be created by making one or more mutations (such as 16 mutations) in the ADAR2 deaminase domain. Even with one or more mutations, this editing enzyme can still show several transcriptome wide A-G and C-U off-targets. To improve the specificity of this enzyme, it can be split at certain residues (such as those residues identified in previous screens) and thus develop a split ADAR system for C-U editing.

**[0285] Recruitment of exogenous and endogenous APOBECs for C-U editing**

**[0286]** APOBECs (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) are RNA editing enzymes that convert cytidines to uracil and create diversity at the mRNA level. These entities can be recruited for C-U editing of RNA by a guide RNA, such as an engineered APOBEC recruiting guide RNAs (apRNAs). In some cases, a fusion construct can be created comprising one or more APOBEC family members. For example, a fusion construct can comprise ADAT1 (Adenosine Deaminase TRNA Specific 1) and AID (Activation-induced cytidine deaminase) to the MCP (MS2 Coat Protein). Engineered MS2-apRNA can be utilized to recruit one or more MCP fusions. The protein sequences for the constructs as well as MS2-apRNA sequences are shown in **FIG. 28- FIG. 30**. A protein sequence for a fusion construct or a sequence utilized in any of the methods as described herein can comprise at least about: 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% sequence identity or more to at least a portion of any sequence of **FIG. 28**.

**[0287]** In some cases, methods can include constructs configured for recruitment of an APOBEC, such as APOBEC3A. Recruitment can be endogenous recruitment. Constructs configured for recruitment of APOBECs, can be designed by targeting preferences of primary sequence, secondary structure or a combination thereof. Designs for one or more apRNA can include those sequences or a portion thereof as show in **FIG. 29**. In some cases, a sequence that can recruit an APOBEC, such as APOBEC3A or a sequence utilized in any of the methods as described herein can comprise at least about: 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% sequence identity or more to at least a portion of any sequence of **FIG. 29**.

[0288] To recruit MCP-APOBEC3A, MS2-apRNA can be designed and their sequence can comprise any one or more of the sequences of **FIG. 30**. In some cases, a sequence that can recruit MCP-APOBEC3A or a sequence utilized in any of the methods as described herein can comprise at least about: 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% sequence identity or more to at least a portion of any sequence of **FIG. 30**.

**[0289] Luciferase Assay**

[0290] When the ADAR converts the target A in TAG to an I (read as a G) the ribosome can be able to fully translate the protein as diagrammed in **FIG. 31**. Cells can be transfected and after 48 hours can be visualized with a luciferase reporter assay. The light readout from the cells can indicate restoration of luciferase activity.

[0291] Referring to **FIG. 31**, a first scenario demonstrates an example in which a ribosome can reach a premature TAG stop codon in a luciferase gene, and can stop translation resulting in a truncated non-functional luciferase enzyme. In a second scenario, an ADAR can be recruited to a site by an adRNA where it can edit a TAG stop codon to a TGG codon for tryptophan that can allow for ribosomal read-through that can result in normal luciferase expression. Such a system can permit evaluation of appropriate ADAR recruitment by an adRNA. When luciferase expression is detected, ADAR recruitment by an adRNA can have occurred. When luciferase expression is not detected, ADAR recruitment may not have occurred.

**Kits**

[0292] Also disclosed herein is a kit comprising, or alternatively consisting essentially of, or yet further consisting of the engineered adRNA of this disclosure, the isolated polynucleotide encoding the engineered adRNA of this disclosure, the vector expressing the engineered adRNA of this disclosure, the recombinant cell expressing the engineered adRNA of this disclosure, or the compositions disclosed herein and instructions for use. In one aspect, the instructions recite the methods of using the engineered adRNA disclosed herein.

[0293] A kit can comprise a vector. The vector can be packaged in a container. The kit can comprise a non-naturally occurring RNA. The non-naturally occurring RNA can be packaged in a container. The kit can comprise a syringe. A syringe can be the container in which the vector, nucleic acid, or non-naturally occurring RNA can be packaged. The kit can comprise a pharmaceutical composition as described herein. The kit can comprise instructions for administration to a subject in need thereof of a viral vector, a non-naturally occurring RNA, a pharmaceutical composition as described herein.

## EXAMPLES

[0294] The following examples are non-limiting and illustrative of procedures which can be used in various instances in carrying the disclosure into effect. Additionally, all reference disclosed herein are incorporated by reference in their entirety.

### Example 1

[0295] Not to be bound by theory, it is believed that ADARs, which may be found in mammals, can be recruited to catalyze therapeutic editing of point mutations. ADAR1 or ADAR2 can be recruited to the target RNA or potentially DNA via the use of engineered RNA scaffolds, engineered DNA scaffolds or DNA-RNA hybrid scaffolds. Tissues that can be potentially targeted using this approach include, but are not limited to, the central and peripheral nervous system, lungs, liver, gastrointestinal tract, pancreas, cardiac muscle, kidneys and skin.

[0296] An exemplary embodiment proposed herein is an engineered ADAR2 guide RNA (adRNA) that bears a 20-100 bp complementarity with the target RNA. This engineered adRNA also contains an ADAR2 recruiting domain from the GluR2 mRNA either at the 5' end or the 3' end, or both ends.

[0297] This was tested *in vivo* in the *spf<sup>ash</sup>* mouse model of ornithine transcarbamylase deficiency. This model bears a G->A point mutation in the last nucleotide of exon 4. Upon delivery of only adRNA via AAVs, up to 1% correction of the point mutation in the absence of the overexpression of the ADAR enzymes was observed. The disclosure also shows this efficacy *in vitro* in HeLa cells that are known to express ADARs. RNA editing was observed in these cells upon delivery of only the adRNA.

[0298] Further, this efficiency can be applied to other recruiting domains. Accordingly, further aspects relate to an engineered single stranded ADAR2 guide DNA (adDNA) as well as adDNA-RNA hybrid with potential greater stability than adRNA. Methods of use of these enzymes as disclosed herein are further provided.

[0299] Not to be bound by theory, since the ADAR family of enzymes catalyze the hydrolytic deamination of adenosine to inosine, it is believed that these enzymes can be used to catalyze the hydrolytic deamination of cytosine to thymine by mutating three specific residues of the ADAR2 – V351, E396 and C451 that interact with the target adenosine. Thus, methods of providing this catalytic potential is provided by mutation of these sites.

[0300] All possible amino acid substitutions at the three residues mentioned have been created and are being screened to test the hypothesis.

[0301] In addition experiments are being performed to explore the roles of other amino acids such as S486 that might enable elimination of the ADARs intrinsic preference for a UAG editing site.

[0302] In order to engineer C->T edits, the roles of hAPOBEC1 and rAPOBEC1 along with the overexpression of the Apobec1 complementation factor (ACF) are determined. In addition, a MCP-(h/r)APOBEC1-ACF fusion protein is further provided herein.

[0303] Also provided to the current adenine base editing approach to Cas9 (or Cpf1)-ADAR deaminase domain fusions (ADAR1, ADAR2 and their catalytically active mutants E1008Q and E488Q), compositions targeting the ssDNA displaced strand by current base editors are further provided herein. To accomplish such, the gRNA bound strand with a A-C bulge, ideally in the first 10bp close to the 5' end of the gRNA is targeted.

[0304] Additional embodiments are exemplified in the appended documents, incorporated herein by reference.

### Example 2

[0305] Referring to **FIG. 6**, Alu elements can be a primary target of endogenous ADAR based RNA editing. Therefore, for the purpose of programmable RNA editing, the goal is to design an ADAR recruiting RNA (adRNA) based on the Alu elements as these can potentially enable efficient recruitment of endogenous ADARs. The Alu-adRNA is expressed from a human U6 promoter and the linker sequence between the Alu repeats is replaced by antisense domains of a variety of lengths, targeting the RAB7A transcript. Each antisense domain has a mismatched nucleotide in the middle of the antisense region, a cytosine across from the target adenosine.

[0306] The Alu-adRNA are tested out *in vitro* by transfection of 293FT cells, either along with ADAR1p110, ADAR1p150, ADAR2, or without an overexpressed enzyme, to demonstrate recruitment of endogenous ADARs. Cells are harvested 48 hours post transfection, RNA is extracted and converted to cDNA via the use of either random hexamers or oligo-dT primers. The RAB7A locus is then amplified and sent for Sanger sequencing. Editing efficiencies are calculated as the ratio of Sanger peak heights G/(A + G).

### Example 3

[0307] **Vector design and construction**

[0308] Zero, one or two copies of the GluR2 adRNAs were cloned into an AAV vector containing a human U6 and mouse U6 promoter along with a CMV promoter driving the expression of GFP or the full length human ADAR2 enzyme or its hyperactive mutant

ADAR2 (E488Q). Similarly, one or two copies of the MS2 adRNAs were cloned into an AAV vectors bearing the MCP-ADAR1 or MCP-ADAR2 deaminase domain fusions and their hyperactive mutants. To construct the GFP reporters – GFP-Amber, GFP-Ochre and GFP-Opal, three gene blocks were synthesized with ‘TAG’, ‘TAA’ and ‘TGA’ respectively replacing the Y39 residue of the wild type GFP and were cloned downstream of a CAG promoter. To construct the OTC and DMD reporters, 200 bp fragments of the *spf<sup>ash</sup>* OTC and *mdx* DMD transcript bearing the target adenosine(s) to be edited were cloned downstream of the CAG promoter.

#### **[0309] Mammalian cell culture and transfection**

**[0310]** All HEK 293T cells were grown in Dulbecco’s Modified Eagle Medium supplemented with 10% FBS and 1% Antibiotic-Antimycotic (Thermo Fisher) in an incubator at 37 °C and 5% CO<sub>2</sub> atmosphere. All *in vitro* transfection experiments were carried out in HEK 293T cells using the commercial transfection reagent Lipofectamine 2000 (Thermo Fisher). All *in vitro* RNA editing experiments involving a reporter were carried out in 24 well plates using 400ng of reporter plasmid and 800ng of the adRNA+enzyme plasmid. All *in vitro* RNA editing experiments targeting an endogenous transcript were carried out in 24 well plates using 800ng of the adRNA/Enzyme plasmid. dCas13b-ADAR2DDE488Q based RNA editing experiments were carried out using 800ng of the enzyme plasmid (Addgene #103864) as well 800 ng of the gRNA plasmid. Cells were transfected at 25-30% confluence and harvested 60 hours post transfection for quantification of editing. Chemically synthesized adRNAs (synthesized via IDT or Synthego) were transfected using Lipofectamine 3000 (Thermo Fisher) at an amount of 20 pmol/well.

#### **[0311] Production of AAV vectors**

**[0312]** AAV8 particles were produced using HEK 293T cells via the triple transfection method and purified via an iodixanol gradient. Confluency at transfection was about 80%. Two hours prior to transfection, DMEM supplemented with 10% FBS was added to the HEK 293T cells. Each virus was produced in 5 x 15 cm plates, where each plate was transfected with 7.5 ug of pXR-8, 7.5 of ug recombinant transfer vector, 7.5 ug of pHelper vector using PEI (1ug/uL linear PEI in 1xDPBS pH 4.5, using HCl) at a PEI:DNA mass ratio of 4:1. The mixture was incubated for 10 minutes at RT and then applied dropwise onto the cell media. The virus was harvested after 72 hours and purified using an iodixanol density gradient ultracentrifugation method. The virus was then dialyzed with 1 x PBS (pH 7.2) supplemented with 50 mM NaCl and 0.0001% of Pluronic F68 (Thermo Fisher) using 50kDA filters

(Millipore), to a final volume of ~1 mL and quantified by qPCR using primers specific to the ITR region, against a standard (ATCC VR-1616).

[0313] AAV-ITR-F: 5'-CGGCCTCAGTGAGCGA-3' (SEQ ID NO:149) and

[0314] AAV-ITR-R: 5'-GGAACCCCTAGTGATGGAGTT-3' (SEQ ID NO:150).

**Example 4 – *in vivo* RNA editing of point mutations via RNA-guided adenosine deaminases.**

[0315] A system for sequence-specific RNA base editing via Adenosine Deaminases acting on RNA (ADAR) enzymes with associated ADAR guide RNAs (adRNAs) was designed. The system was systematically engineered to harness ADARs, and comprehensively evaluated its specificity and activity *in vitro* and *in vivo* via two mouse models of human disease. In some cases, this platform can enable tunable and reversible engineering of RNAs for diverse applications.

[0316] Adenosine to inosine RNA editing, a post-transcriptional RNA modification, is catalyzed by Adenosine Deaminases acting on RNA (ADAR) enzymes. Inosine is a deaminated form of adenosine that is biochemically recognized as guanine. Recently, multiple studies have demonstrated ADAR mediated targeted RNA editing. Building on these, two orthogonal toolsets were engineered for sequence-specific programmable RNA base editing *in vitro* and *in vivo*. Specifically, a system for targeted RNA editing via ADAR1/2 with associated ADAR guide RNAs (adRNAs) was utilized (**FIG. 32A**). The adRNAs comprise in part a programmable antisense region that is complementary to the target RNA sequence with a mismatched cytidine opposite the target adenosine. Additionally, they bear in one version, zero, one, or two ADAR-recruiting domains engineered from the naturally occurring ADAR substrate GluR2 pre-mRNA (referred hereon as GluR2 adRNA); and in a second format, two MS2 hairpins flanking the antisense region (referred hereon as MS2 adRNA). The GluR2 adRNA was systematically optimized to enhance recruitment of exogenous and/or endogenous ADARs by evaluating multiple scaffold variants, including mutagenized scaffolds based on G-C versus A-U pairing, addition of editing inducer elements, and antisense domain length and mis-match position modifications (**FIG. 8, FIG. 34A-C**). The latter MS2 adRNA version was in turn optimized to harness synthetic proteins comprising the deaminase domains (DD) of ADAR1 or ADAR2 fused to the MS2 Coat Protein (MCP), via systematic antisense domain length and mis-match position modifications, coupled with use of hyper-active versions of the deaminase domains, and

versions bearing nuclear localization (NLS) versus export (NES) signals (**FIG. 32B, FIG. 35A, 41B**).

[0317] The activity of the above two systems were comprehensively evaluated and benchmarked with the recently developed RNA editing system based on Cas13b. These *in vitro* experiments revealed that: (1) the engineered constructs were active in their ability to effect targeted RNA editing with yields comparable to the Cas13b based system (**FIG. 32B, FIG. 36A, Tables 2, 3**), and U6 transcribed adRNAs and chemically synthesized adRNAs were both effective formats (**FIG. 36B**); (2) adRNAs bearing long antisense domains, both with and without GluR2 domains, suffice to recruit exogenously expressed ADARs, and to a degree endogenous ADARs too to enable efficient RNA editing (**FIG. 32B, FIG. 34B, 40C, 42C**); (3) the constructs based on the MS2 adRNAs and corresponding MCP-ADAR1/2 fusions showed the highest and most robust activity, including across a large panel of endogenous genes chosen across a spectrum of different expression levels (**FIG. 32B, FIG. 36C**); (4) use of a NES and/or hyper-active deaminase domains in the MCP-ADAR1/2 fusions consistently yielded higher RNA editing yields at the target adenosine, but also led to a higher propensity of editing at non-targeted adenosines in the flanking sequences (**FIG. 32B, FIG. 37A**). To further validate this, a similar promiscuity ensued from deletion of the native NLS domain in ADAR2 ( $\Delta$ 1-138) (**FIG. 37B - FIG. 37D**); and 5) these two toolsets were operationally orthogonal: specifically, the editing efficiency of the MCP-ADAR2 deaminase domain fusion with a co-expressed MS2 adRNA or GluR2 adRNA was evaluated and displayed on-target editing only via the former. Conversely, full-length ADAR2 was observed to be recruited by the GluR2 adRNA and not the MS2 adRNAs (**FIG. 35B**).

[0318] Having demonstrated robust activity of this toolset, its specificity profiles were investigated via analysis of the transcriptome-wide off-target A->G editing effected by this system (**FIG. 32C**). To this end, HEK 293T cells were transfected with each construct and analyzed by RNA-seq. Untransfected cells were included as controls. From each sample, ~40 million uniquely aligned sequencing reads were collected. Fisher's exact test was used to quantify significant changes in A->G editing yields, relative to untransfected cells, at each reference adenosine site having sufficient read coverage. The number of sites with at least one A->G editing event detected in any of the samples was computed. Of these, the number of sites with statistically significant A->G edits, at a false discovery rate (FDR) of 1%, and with fold change of at least 1.1, was found to vary over a wide range, from lowest for the MCP-ADAR2 DD-NLS construct, to highest for the MCP-ADAR1 DD (E1008Q)-NES

(**FIG. 38 – FIG. 41, Tables 4, 5**). To investigate the distribution of editing yields, violin plots were generated considering the A-sites whose editing yields changed significantly in at least one sample (**FIG. 32**). Taken together, the RNA-seq experiments revealed that transcriptome-wide off-target edits were: 1) less prevalent in MCP-ADAR constructs with NLS than constructs with NES; 2) less prevalent in MCP-ADAR2 constructs than MCP-ADAR1 constructs; 3) less prevalent in the wild-type MCP-ADAR constructs than the E>Q hyperactive mutants (**FIG. 42A, Table 5**); and 4) the off-targets were primarily due to ADAR overexpression and use of adRNAs alone resulted in least number of off-targets (**FIG. 42B**).

[0319] **Table 2:** List of adRNA and gRNA antisense sequences:



| Name                               | adRNA/gRNA antisense sequence (5' to 3')                                                                     |
|------------------------------------|--------------------------------------------------------------------------------------------------------------|
| mOTC <sup>2</sup>                  | ACAAACCGAGCGGTGTCTGT (SEQ ID NO:202)                                                                         |
| mDMD <sup>2</sup>                  | GCCATTCCATTGCTCTTTCA (SEQ ID NO:203)                                                                         |
| RAB7A (20, 6)                      | TGCCGCGCAGCTGGATTTC (SEQ ID NO:204)                                                                          |
| CCNB1 (20, 6)                      | CTGTACCCAGCCAGTCAATTA (SEQ ID NO:205)                                                                        |
| DAXX (20, 6)                       | CTTCTCCACAGCCCGAAGCA (SEQ ID NO:206)                                                                         |
| CKDN2 (20, 6)                      | CTGCTCCACCCGACCCCGGG (SEQ ID NO:207)                                                                         |
| GAPDH (20, 6)                      | GGGTGCCAAGCAGTTGGTGG (SEQ ID NO:208)                                                                         |
| ALDOA (20, 6)                      | CTTGTCCACCTTGATGCCCA (SEQ ID NO:209)                                                                         |
| ARHGAP8 (20, 6)                    | TTCATCCAAATGGCTGGTTAT (SEQ ID NO:210)                                                                        |
| CKB (20, 6)                        | CAAGGCCAAGGGCTCGCCAG (SEQ ID NO:211)                                                                         |
| KRAS (20, 6)                       | TCCAACCACCACAAGTTTAT (SEQ ID NO:212)                                                                         |
| Cas13b_RAB7A                       | TACAGAATACTGCCGCGCAGCTGGATTTCCTCAATTCTGAGTAACACTCTGC (SEQ ID NO:213)                                         |
| Cas13b_mOTC                        | GAAAAAGTTTTACAAACCGAGCGGTGTCTGTGAGACTTTCATTACACCCA (SEQ ID NO:214)                                           |
| Cas9_mDMD_1                        | ATAATTTCTATTATATTACA (SEQ ID NO:215)                                                                         |
| Cas9_mDMD_2                        | ATTTCAGGTAAGCCGAGGTT (SEQ ID NO:216)                                                                         |
| RAB7A (20, 10)                     | ATACTGCCGCGCAGCTGGATT (SEQ ID NO:217)                                                                        |
| RAB7A (60, 6)                      | TGCCGCGCAGCTGGATTTCCTCAATTCTGAGTAACACTCTGCAATCCAAACAGGG<br>TTCAACC (SEQ ID NO:218)                           |
| RAB7A (60, 30)                     | TCTTGTGTCTACTGTACAGAATACTGCCGCGCAGCTGGATTTCCTCAATTCTGAGT<br>AACACT (SEQ ID NO:219)                           |
| RAB7A (100, 6)<br>(SEQ ID NO:220)  | TGCCGCGCAGCTGGATTTCCTCAATTCTGAGTAACACTCTGCAATCCAAACAGGG<br>TTCAACCCCTCCACCTTACAGGCCCTGCATTACAGGACTTAAACACATA |
| RAB7A (100, 50)<br>(SEQ ID NO:221) | TGATAAAAGGCGTACATAATTCTTGTGTCTACTGTACAGAATACTGCCGCGCAGC<br>TGGATTTCCTCAATTCTGAGTAACACTCTGCAATCCAAACAGGGTTC   |
| KRAS (100, 50)<br>(SEQ ID NO:222)  | TGAATTAGCTGTATCGTCAAGGCACTCTTGCCCTACGCCACCAGCTCCAACCCAC<br>CACAAAGTTTATATTCAGTCATTTTCAGCAGGCCCTCTCTCCCGCACCT |
| CKB (100, 50)<br>(SEQ ID NO:223)   | ATCAAAAAATAAACTCTACCAAGGGTGACGGAAGTCTCTACAGCAAGGCCAA<br>GGGCTCGCCAGACGCGCAACATCAGGGGTGCATGGTGGGCACTGCC       |

[0320] **Table 3:** List of primers for next generation sequencing (NGS) analyses.

| Name                  | Sequence (5' to 3')                              |
|-----------------------|--------------------------------------------------|
| mDMD_NGS_F1           | CTCTCTGTACCTTATCTTAGTGTACTGA (SEQ ID NO:224)     |
| mDMD_NGS_R1           | ATTTCTGGCATATTTCTGAAGGTG (SEQ ID NO:225)         |
| mOTC_NGS_F1           | ACCCCTTCCTTTCTTACCACACA (SEQ ID NO:226)          |
| mOTC_spliced_NGS_R1   | CAGGGTGTCCAGATCTGATTGTT (SEQ ID NO:227)          |
| mOTC_unspliced_NGS_R1 | CTTCTCTTTTAACTAACCCATCAGAGTT (SEQ ID NO:228)     |
| CCNB1_NGS_F           | CAAGCAGTCAGACCAAAATACCTACTG (SEQ ID NO:229)      |
| CCNB1_NGS_R           | TCTTAGCATGCTTCGATGTGGCATA (SEQ ID NO:230)        |
| DAXX_NGS_F            | CATCAACAAGCCAGGGCCTG (SEQ ID NO:231)             |
| DAXX_NGS_R            | GAAGAGGAAATGTCCGTCTCCAC (SEQ ID NO:232)          |
| RAB7A_NGS_F           | AGGCCTGTAAGGTGGAGGG (SEQ ID NO:233)              |
| RAB7A_NGS_R           | (SEQ ID NO:234) TGAAATAACGGCAATTTATCCATTGCACATAC |
| CDKN2A_NGS_F          | GGGAGCAGCATGGAGCCTT (SEQ ID NO:234)              |
| CDKN2A_NGS_R          | TCCGACCGTAACTATTCGGTGC (SEQ ID NO:236)           |
| GAPDH_NGS_F           | TGGGTGTGAACCATGAGAAGTAT (SEQ ID NO:237)          |
| GAPDH_NGS_R           | TGGCATGGACTGTGGTCATG (SEQ ID NO:238)             |
| CKB_NGS_F             | CCTAACTTATTGCCTGGGCAGTG (SEQ ID NO:238)          |
| CKB_NGS_R             | GCATCAGCAGTATCTTAGCCATCAA (SEQ ID NO:240)        |
| NGS_KRAS_F            | CAGAGGCTCAGCGGCTCC (SEQ ID NO:241)               |
| NGS_KRAS_R            | TAGCTGTATCGTCAAGGCACTC (SEQ ID NO:242)           |
| ARHGAP8_NGS_F         | CACACCTGTCTGTGCACTTGTA (SEQ ID NO:243)           |
| ARHGAP8_NGS_R         | CGGTCCACAGCTCAGGAACC (SEQ ID NO:244)             |
| ALDOA_NGS_F           | ACCAGAAGGCGGATGATGGG (SEQ ID NO:245)             |
| ALDOA_NGS_R           | CTCAGACAGCCCATCCAACC (SEQ ID NO:246)             |
| KRAS_NGS_R2           | TACTACTTGCTTCCTGTAGGAATCCTC (SEQ ID NO:247)      |
| CKB_NGS_F2            | AGCCCTGCTGCTTCCTAACTT (SEQ ID NO:248)            |
| CKB_NGS_R2            | ACCCTAGTTTATTTAGCATCAGCAG (SEQ ID NO:249)        |

**[0321] Table 4.** Tallies of RNA-seq reads from high-throughput sequencing experiments.

Tallies of RNA-seq reads from high-throughput sequencing experiments. The given counts represent read mates, not read pairs, from paired end sequencing. Columns are: sn, sample name; nt, total number of raw reads after demultiplexing; nu, number of reads in pairs

uniquely aligned to the reference genome; nd, number of reads in duplicated pairs; nr, number of remaining reads; df, down-sampling fraction. Samples named “293T”, “293T L2”, “293T L8”, and “293T L4” were taken from the same control library but were sequenced on different lanes of the Illumina instrument:

| sn                                   | nt         | nu         | nd        | nr         | df       |
|--------------------------------------|------------|------------|-----------|------------|----------|
| 293T L1 (0)                          | 108434072  | 86792552   | 21376362  | 65416190   | 0.474401 |
| 293T (1)                             | 107737550  | 86652062   | 21362310  | 65289752   | 0.47532  |
| MCP-ADAR1 DD-NES -adRNA (2)          | 75469304   | 57715184   | 11907804  | 45807380   | 0.677479 |
| MCP-ADAR1 DD-NES +adRNA (3)          | 76113978   | 55714058   | 11591434  | 44122624   | 0.703347 |
| MCP-ADAR1 DD-NLS -adRNA (4)          | 96485146   | 79023222   | 17911404  | 61111818   | 0.507815 |
| MCP-ADAR1 DD-NLS +adRNA (5)          | 70684382   | 56425658   | 8656076   | 47769582   | 0.64965  |
| MCP-ADAR1 DD (E1008Q)-NES -adRNA (6) | 73073334   | 54369950   | 10484274  | 43885676   | 0.707145 |
| MCP-ADAR1 DD (E1008Q)-NES +adRNA (7) | 95946852   | 71154174   | 18408668  | 52745506   | 0.588363 |
| MCP-ADAR1 DD (E1008Q)-NLS -adRNA (8) | 54654264   | 43441234   | 6314734   | 37126500   | 0.835886 |
| MCP-ADAR1 DD (E1008Q)-NLS +adRNA (9) | 78346400   | 59725272   | 11074654  | 48650618   | 0.637886 |
| MCP-ADAR2 DD-NES -adRNA (10)         | 89534306   | 74166552   | 14629650  | 59536902   | 0.521249 |
| MCP-ADAR2 DD-NES +adRNA (11)         | 80859886   | 66904932   | 12911706  | 53993226   | 0.574767 |
| MCP-ADAR2 DD (E488Q)-NES -adRNA (12) | 79769278   | 65570598   | 11792520  | 53778078   | 0.577066 |
| MCP-ADAR2 DD (E488Q)-NES +adRNA (13) | 98084994   | 80214200   | 20639602  | 59574598   | 0.520919 |
| MCP-ADAR2 DD-NLS -adRNA (14)         | 75862216   | 60748320   | 14040780  | 46707540   | 0.664422 |
| MCP-ADAR2 DD-NLS +adRNA (15)         | 80473694   | 66106146   | 12359830  | 53746316   | 0.577407 |
| MCP-ADAR2 DD (E488Q)-NLS -adRNA (16) | 48576248   | 40372732   | 6032488   | 34340244   | 0.903707 |
| MCP-ADAR2 DD (E488Q)-NLS +adRNA (17) | 72617598   | 58732422   | 12388120  | 46344302   | 0.66963  |
| 293T L8 (18)                         | 68191034   | 54705760   | 10254952  | 44450808   | 0.686154 |
| 293T+GFP (19)                        | 88146982   | 64109834   | 15072908  | 49036926   | 0.63286  |
| ADAR2 -adRNA (20)                    | 86641852   | 69462198   | 18251916  | 51210282   | 0.606002 |
| ADAR2 +adRNA (21)                    | 74048950   | 59010378   | 16071512  | 42938866   | 0.722737 |
| ADAR2 (E488Q) -adRNA (22)            | 81927154   | 65842936   | 16273572  | 49569364   | 0.626063 |
| ADAR2 (E488Q) +adRNA (23)            | 74616248   | 56997066   | 17856714  | 39140352   | 0.792878 |
| Cas13b-ADAR2 DD (E488Q) -gRNA (24)   | 72072754   | 54678392   | 12500074  | 42178318   | 0.73577  |
| Cas13b-ADAR2 DD (E488Q) +gRNA (25)   | 116274658  | 91188024   | 29488244  | 61699780   | 0.502976 |
| 293T L4 (26)                         | 70234868   | 56894104   | 11595590  | 45296514   | 0.685089 |
| MS2 adRNA (27)                       | 78457766   | 51176442   | 19161354  | 32015088   | 0.969341 |
| GluR2 adRNA (28)                     | 65855024   | 41791326   | 10757802  | 31033524   | 1        |
| gRNA (29)                            | 89132098   | 59288978   | 19226242  | 40062736   | 0.774623 |
| min                                  | 48576248   | 40372732   | 6032488   | 31033524   | 0.474401 |
| max                                  | 116274658  | 91188024   | 29488244  | 65416190   | 1        |
| total                                | 2588377968 | 2005961258 | 481864884 | 1524096374 | 21.68535 |

**[0322] Table 5.** Results of A->G editing yield quantification from aligned RNA-seq reads.

Results of A->G editing yield quantification from aligned RNA-seq reads. Columns are: sample name; total sites, the total number of reference sites with a significant change in A->G editing yield in at least one comparison between treatment and control sample; changed sites, the number of reference A-sites found to have a significant change in A->G editing yield when comparing the treatment to the control sample, which is the first sample in Table 4; on-target editing yield, the editing yield observed at the intended target A-site within the RAB7A mRNA; median editing yield, the median yield at all sites considered except the target site:

| Sample name                       | Total sites | Changed sites | On-target editing yield | Median editing yield |
|-----------------------------------|-------------|---------------|-------------------------|----------------------|
| 293T L2                           | 382978      | 0             | 0                       |                      |
| 293T + GFP                        | 382978      | 6             | 0                       | 0.295913155          |
| gRNA                              | 382978      | 32            | 0.013333333             | 0.142102619          |
| Cas13b-ADAR2 DD (E488Q) - gRNA    | 382978      | 112853        | 0.025316458             | 0.105263158          |
| Cas13b-ADAR2 DD (E488Q) + gRNA    | 382978      | 49432         | 0.11637931              | 0.092205807          |
| GluR2 adRNA                       | 382978      | 23            | 0                       | 0.144542773          |
| ADAR2 - adRNA                     | 382978      | 5769          | 0                       | 0.157894737          |
| ADAR2 + adRNA                     | 382978      | 18573         | 0.27638191              | 0.14                 |
| ADAR2 (E488Q) - adRNA             | 382978      | 25732         | 0                       | 0.131578947          |
| ADAR2 (E488Q) + adRNA             | 382978      | 125409        | 0.268398268             | 0.150943396          |
| MS2 adRNA                         | 382978      | 19            | 0.006666667             | 0.169230769          |
| MCP-ADAR1 DD-NLS - adRNA          | 382978      | 20481         | 0.006849315             | 0.079069767          |
| MCP-ADAR1 DD-NLS + adRNA          | 382978      | 28537         | 0.159763314             | 0.084745763          |
| MCP-ADAR1 DD (E1008Q)-NLS - adRNA | 382978      | 90182         | 0                       | 0.112                |
| MCP-ADAR1 DD (E1008Q)-NLS + adRNA | 382978      | 110565        | 0.261627907             | 0.118081181          |
| MCP-ADAR1 DD-NES - adRNA          | 382978      | 116185        | 0.017142857             | 0.097222222          |
| MCP-ADAR1 DD-NES + adRNA          | 382978      | 101183        | 0.366459627             | 0.096618357          |
| MCP-ADAR1 DD (E1008Q)-NES - adRNA | 382978      | 226634        | 0.010416667             | 0.123076923          |
| MCP-ADAR1 DD (E1008Q)-NES + adRNA | 382978      | 195533        | 0.418604651             | 0.12244898           |
| MCP-ADAR2 DD-NLS - adRNA          | 382978      | 3760          | 0.006756757             | 0.056173674          |
| MCP-ADAR2 DD-NLS + adRNA          | 382978      | 4740          | 0.07                    | 0.076555024          |
| MCP-ADAR2 DD (E488Q)-NLS - adRNA  | 382978      | 28026         | 0.014778325             | 0.095238095          |
| MCP-ADAR2 DD (E488Q)-NLS + adRNA  | 382978      | 38087         | 0.113122172             | 0.098591549          |
| MCP-ADAR2 DD-NES - adRNA          | 382978      | 9489          | 0.021276596             | 0.09375              |
| MCP-ADAR2 DD-NES + adRNA          | 382978      | 20249         | 0.416216216             | 0.102564103          |
| MCP-ADAR2 DD (E488Q)-NES - adRNA  | 382978      | 35287         | 0.004672897             | 0.09929078           |
| MCP-ADAR2 DD (E488Q)-NES + adRNA  | 382978      | 42715         | 0.278350515             | 0.101351351          |

**[0323]** Following these *in vitro* studies, the system was evaluated for *in vivo* RNA targeting in gene therapy applications, utilizing the adRNA cum exogenous ADAR expression construct versions, as those consistently enabled the highest *in vitro* RNA editing yields. The

*mdx* mouse model for Duchenne muscular dystrophy (DMD) was first evaluated, which bears an ochre stop site in exon 23 of the dystrophin gene. This choice was additionally motivated by the fact that nonsense mutations in general may be responsible for nearly 11% of described gene lesions causing inheritable human disease, and close to 20% of disease-associated single base substitutions that affect the coding regions of genes. Thus, validation of an RNA editing strategy here can have broad therapeutic application. Towards this, the RNA editing of stop codons was first optimized *in vitro* (**FIG. 43**). Notably, it was observed that addition of a second copy of the adRNA significantly improved the targeting efficiencies (**FIG. 43C**), and thus in the *in vivo* studies a dual-adRNA delivery approach was utilized. The constructs were then packaged into AAV8, and injected 2E+12 vector genomes (vg)/muscle into the tibialis anterior (TA) or gastrocnemius of *mdx* mice. To further benchmark the approach, the *mdx* mice were concurrently targeted via CRISPR-Cas9 based excision of exon 23 (**FIG. 33A**). Four or eight weeks post injection, TA and gastrocnemius muscles were collected from *mdx* mice, wild type mice, mice treated with adRNA targeting and non-targeting controls, and CRISPR-Cas9. Immunofluorescence staining revealed clear restoration of dystrophin expression via targeted RNA editing (**FIG. 33B, FIG. 44A**). In addition, nNOS activity was also restored at the sarcolemma (**FIG. 33B, FIG. 44A**). RNA editing yields (TAA->TGG/TAG/TGA) of up to 3.6%, and TAA->TGG up to 2.4% were observed in treated mice (**FIG. 33C, FIG. 43E**). Western blots of the treated muscles confirmed the immunofluorescence observations, demonstrating 1-2.5% protein restoration. (**FIG. 44B**). As benchmark, muscles injected with vectors bearing CRISPR-Cas9 also expectedly led to restoration of dystrophin expression in a subset of the muscle cells (**FIG. 33B**), with Western blots of the treated muscles confirming up to 10% protein restoration. (**FIG. 44C**).

[0324] To further confirm the efficacy of this approach, ADAR mediated RNA editing was next evaluated in an independent mouse model of human disease, the male sparse fur ash (*spf<sup>ash</sup>*) mouse model of ornithine transcarbamylase (OTC) deficiency. The *spf<sup>ash</sup>* mice harbor a G->A point mutation in the last nucleotide of the fourth exon of the OTC gene, which leads to OTC mRNA deficiency and production of a mutant protein. Recent studies have demonstrated the use of CRISPR-Cas9 and homologous recombination based strategies for robust correction of this mutation in neonatal mice. To test the effectiveness of the system in editing the point mutation in *spf<sup>ash</sup>* OTC mRNA (**FIG. 33D**), the constructs were evaluated *in vitro* (**FIG. 45A**). The constructs were packaged into AAV8, which has high liver tropism,

and injected  $2.5 \times 10^{12}$  vg/mouse in 10-12 week old  $spf^{ash}$  mice. Three to four weeks post injection, liver samples were collected from  $spf^{ash}$  mice, wild-type litter mates, and  $spf^{ash}$  mice treated with the ADAR2 targeting and non-targeting vectors and evaluated corresponding editing efficiency via NGS. Notably, upon delivery of the adRNA and the ADAR2, 0.8-4.7% edited mRNA was observed amongst the correctly spliced OTC mRNA, and interestingly adRNA alone resulted in low but significant RNA editing yields (**FIG. 33E**). Moreover, upon the delivery of the hyper-active ADAR2 mutant (E488Q), a high edited fraction (4.6-33.8%) was observed in the correctly spliced OTC mRNA (**FIG. 33e, FIG. 45B**), 4.6-8.2% in the OTC pre-mRNA (**FIG. 45C**), and confirmed a reduction in the incorrectly spliced product (**FIG. 45D**). Western blots of the treated liver samples confirmed partial (2.5-5%) restoration of OTC protein (**FIG. 45E**).

[0325] Taken together, the results establish the utility of RNA-guided ADARs for *in vivo* RNA editing of point mutations. In some cases, sequence preferences of the ADAR enzymes, RNA folding, intrinsic half-life, localization, translation machinery, and resident RNA binding proteins can potentially impact accessibility and editability of target sites in the RNA, and can be important design parameters to consider for enabling efficacious targeting. For instance, in the *mdx* model, ADAR based RNA editing approaches can have to compete with nonsense mediated decay of mutant dystrophin mRNA, and also the requirement for effecting two A->I substitutions in the context of non-ideal flanking nucleotides to eliminate the premature stop codon and potential impact on RNA stability and function. Furthermore, in the  $spf^{ash}$  model, the need to target the transient OTC pre-mRNA can entail rapid target engagement and editing. Further progress can also be needed addressing important limitations of the system such as the off-targets induced by intrinsic enzyme-RNA binding, processivity, promiscuity, stimulation of the interferon response by the delivery modalities themselves (such as lipid, nanoparticles or viral) leading in turn to increased endogenous ADAR expression, potential of adRNAs to induce RNAi, and also off-target hybridization of the antisense domain of the adRNA which can potentially have deleterious effects. In this regard, the studies revealed toxicity in mice systemically injected with the hyperactive ADAR mutants (**FIG. 46**). These studies can be critical to aid systematic improvement of the specificity and safety of this approach. Another important consideration while considering RNA targeting for gene therapy, especially via the use of non-integrating vectors, can be the necessity for periodic re-administration of the effector constructs, owing to the limited half-life of edited mRNAs and effectors. In this regard, compared to the CRISPR based RNA

editing approaches, the RNA-guided ADAR strategy can be directly human therapeutics relevant, as versions of the same solely utilize effector RNAs and human proteins. Additionally, as ADARs are widely expressed, for instance, ADAR1 across most human tissues and ADAR2 in particular in the lung and brain, endogenous recruitment of these via adRNAs bearing long-antisense domains (as demonstrated in **FIG. 32**, **FIG. 33E** and **FIG. 34**, **FIG. 36**) presents a very attractive strategy for efficacious RNA editing. With progressive improvements, this toolset can have broad implications for diverse basic science and therapeutic applications.

**[0326] FIG. 32: Engineering programmable RNA editing and characterizing specificity profiles:** (A) Schematics of RNA editing via constructs utilizing the full length ADAR2 and an engineered adRNA derived from the GluR2 transcript, or MS2 Coat Protein (MCP) fusions to the ADAR1/2 deaminase domains and the corresponding MS2 hairpin bearing adRNA. (B) Comparison of RNA editing efficiency of the endogenous RAB7A transcript by different RNA editing constructs quantified by Sanger sequencing (efficiency calculated as a ratio of Sanger peak heights G/(A+G)). Experiments were carried out in HEK 293T cells. Values represent mean  $\pm$  SEM (n=3). (C) Violin plots representing distributions of A->G editing yields observed at reference sites where at least one treatment sample was found to have a significant change (Fisher's exact test, FDR = 1%) in editing yield relative to the control sample. Blue circles indicate editing yields at the target A-site within the RAB7A transcript. Black dots represent median off-target editing yields. To better visualize the shapes of the distributions, their maximum extent along the y-axis was equalized across plots, and were truncated at 60% yield.

**[0327] FIG. 33: *In vivo* RNA editing in mouse models of human disease:** (A) Schematic of the DNA and RNA targeting approaches to restore dystrophin expression in the *mdx* mouse model of Duchenne Muscular Dystrophy: (i) a dual gRNA-CRISPR based approach leading to in frame excision of exon 23 and (ii) ADAR2 and MCP-ADAR1 based editing of the ochre codon. (B) Immunofluorescence staining for dystrophin in the TA muscle shows partial restoration of expression in treated samples (intra-muscular injections of AAV8-ADAR2, AAV8-ADAR2 (E488Q), and AAV8-CRISPR). Partial restoration of nNOS localization is also seen in treated samples (scale bar: 250 $\mu$ m). (C) *In vivo* TAA->TGG/TAG/TGA RNA editing efficiencies in corresponding treated adult *mdx* mice. Values represent mean  $\pm$  SEM (n=4, 3, 7, 3, 3, 10, 3, 4 independent TA muscles respectively). (D) Schematic of the OTC locus in the *spf<sup>ash</sup>* mouse model of Ornithine Transcarbamylase

deficiency which have a G->A point mutation at a donor splice site in the last nucleotide of exon 4, and approach for correction of mutant OTC mRNA via ADAR2 mediated RNA editing. **(E)** *In vivo* RNA correction efficiencies in the correctly spliced OTC mRNA in the livers of treated adult *spf<sup>ash</sup>* mice (retro-orbital injections of AAV8-ADAR2 and AAV8-ADAR2 (E488Q)). Values represent mean +/- SEM (n=4, 4, 3, 3, 4, 5 independent animals respectively).

**[0328] Vector design and construction**

**[0329]** One or two copies of the adRNAs were cloned into an AAV vector containing a human U6 and mouse U6 promoter along with a CMV promoter driving the expression of the enzyme. To construct the GFP reporters – GFP-Amber, GFP-Ochre and GFP-Opal, three gene blocks were synthesized with ‘TAG’, ‘TAA’ and ‘TGA’ respectively replacing the Y39 residue of the wild type GFP and were cloned downstream of a CAG promoter. To construct the OTC and DMD reporters, 200 bp fragments of the *spf<sup>ash</sup>* OTC and *mdx* DMD transcript bearing the target adenosine(s) to be edited were cloned downstream of the CAG promoter.

**[0330] Mammalian cell culture and transfection**

**[0331]** All HEK 293T cells were grown in Dulbecco’s Modified Eagle Medium supplemented with 10% FBS and 1% Antibiotic-Antimycotic (Thermo Fisher) in an incubator at 37 °C and 5% CO<sub>2</sub> atmosphere. All *in vitro* transfection experiments were carried out in HEK 293T cells using the commercial transfection reagent Lipofectamine 2000 (Thermo Fisher). All *in vitro* RNA editing experiments involving a reporter were carried out in 24 well plates using 400ng of reporter plasmid and 800ng of the adRNA+enzyme plasmid. All *in vitro* RNA editing experiments targeting an endogenous transcript were carried out in 24 well plates using 800ng of the adRNA/Enzyme plasmid. dCas13b-ADAR2<sub>DD</sub>E488Q based RNA editing experiments were carried out using 800ng of the enzyme plasmid (Addgene #103864) as well 800 ng of the gRNA plasmid. Cells were transfected at 25-30% confluence and harvested 60 hours post transfection for quantification of editing. Chemically synthesized adRNAs (synthesized via IDT or Synthego) were transfected using Lipofectamine 3000 (Thermo Fisher) at an amount of 20 pmol/well.

**[0332] Production of AAV vectors**

**[0333]** AAV8 particles were produced using HEK 293T cells via the triple transfection method and purified via an iodixanol gradient. Confluency at transfection was about 80%. Two hours prior to transfection, DMEM supplemented with 10% FBS was added to the HEK 293T cells. Each virus was produced in 5 x 15 cm plates, where each plate was transfected



with 7.5 ug of pXR-8, 7.5 of ug recombinant transfer vector, 7.5 ug of pHelper vector using PEI (1ug/uL linear PEI in 1xDPBS pH 4.5, using HCl) at a PEI:DNA mass ratio of 4:1. The mixture was incubated for 10 minutes at RT and then applied dropwise onto the cell media. The virus was harvested after 72 hours and purified using an iodixanol density gradient ultracentrifugation method. The virus was then dialyzed with 1 x PBS (pH 7.2) supplemented with 50 mM NaCl and 0.0001% of Pluronic F68 (Thermo Fisher) using 50kDA filters (Millipore), to a final volume of ~1 mL and quantified by qPCR using primers specific to the ITR region, against a standard (ATCC VR-1616).

[0334] AAV-ITR-F: 5'-CGGCCTCAGTGAGCGA-3' (SEQ ID NO:149) and

[0335] AAV-ITR-R: 5'-GGAACCCCTAGTGATGGAGTT-3' (SEQ ID NO:150).

[0336] ***RNA isolation and Next Generation Sequencing library preparation***

[0337] RNA from animal tissue was extracted using the RNeasy Plus Universal Mini Kit (Qiagen), according to the manufacturer's protocol. RNA from cells was extracted using the RNeasy Mini Kit (Qiagen). cDNA was synthesized from 500ng RNA using the Protoscript II First Strand cDNA synthesis Kit (NEB). Next generation sequencing libraries were prepared as follows. Briefly, 1ul of cDNA prepared above was amplified by PCR with primers that amplify about 150 bp surrounding the sites of interest using KAPA Hifi HotStart PCR Mix (Kapa Biosystems). PCR products were purified (Qiagen PCR Purification Kit/ Gel Extraction Kit) to eliminate byproducts. Libraries were constructed with NEBNext Multiplex Oligos for Illumina kit (NEB). 10 ng of input DNA was amplified with indexing primers. Samples were then pooled and loaded on an Illumina Miseq (150bp single-end run) or Hiseq (100bp paired-end run). Data analysis was performed using CRISPResso (Pinello, L. *et al.* 2016). A minimum of 100,000 reads were analyzed for all *in vivo* experiments. RNA-seq libraries were prepared from 300ng of RNA, using the NEBNext Poly(A) mRNA magnetic isolation module and NEBNext Ultra RNA Library Prep Kit for Illumina. Samples were pooled and loaded on an Illumina Hiseq (100bp paired-end run).

[0338] ***Quantification of OTC mRNA editing yields in the *spf<sup>ash</sup>* mice***

[0339] The *spf<sup>ash</sup>* mice bear three forms of OTC RNA: the pre-mRNA, the correctly spliced mRNA and an incorrectly spliced, elongated mRNA formed due to the use of a cryptic splice site 48 base pairs into intron 4. Let the total number of the correctly spliced mRNA be X, incorrectly spliced variant be Y and the pre-mRNA be Z. Xe, Ye and Ze denote the A->G edited mRNA in the three forms. The mRNA editing yield ideally can be calculated as  $(Xe+Ye+Ze)/(X+Y+Z)$ . However, since it is not possible to amplify the spliced and pre-mRNA variants using the same primers, **FIG. 33E** shows the fraction of edited transcripts in

the correctly spliced mRNA (Xe/X) which will in turn be translated to produce the OTC protein. In addition, **FIG. 45C** shows the fraction of edited transcripts in the pre-mRNA (Ze/Z). This fraction, upon correct splicing will contribute to formation of OTC protein. Finally, the incorrectly spliced mRNA results in the production of a protein elongated by 16 amino acids which is selectively degraded. In **FIG. 45D**, bands corresponding to X and Y are shown.

#### [0340] *Animal experiments*

[0341] All animal procedures were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee (IACUC) of the University of California, San Diego. Mice were acquired from Jackson labs. AAVs were injected into the gastrocnemius or TA muscle of *mdx* mice (C57BL/10ScSn-*Dmd*<sup>*mdx*</sup>/J) using 2E+12 vg/muscle. AAVs were injected into *spf*<sup>ash</sup> mice (B6EiC3Sn a/*A-Otc*<sup>*spf-ash*</sup>/J) via retro-orbital injections using 2.5E+12 vg/mouse. Mice that appeared to have a rough hair coat, moved slowly and appeared slightly hunched were termed as sick mice and euthanized.

#### [0342] *Immunofluorescence*

[0343] Harvested gastrocnemius or TA muscles were placed in molds containing OCT compound (VWR) and flash frozen in liquid nitrogen. 10 µm sections were cut onto pre-treated histological slides. Slides were fixed using 4% Paraformaldehyde. Dystrophin and nNOS were detected with rabbit polyclonal antibodies against the C-terminal domain of dystrophin (1:200, Abcam 15277) and N-terminal domain of nNOS (1:100, Immunostar 24431) respectively, followed by a donkey anti-rabbit Alexa 546 secondary antibody (1:400, Thermo Fisher).

#### [0344] *Western Blots*

Muscle biopsies from *mdx* mice and liver biopsies from *spf*<sup>ash</sup> mice were fragmented in RIPA buffer (Sigma) with a proteinase inhibitor cocktail (Roche) and incubated for 1 hour on ice with intermittent vortexing. Samples were centrifuged at 15500xg for 30 min at 4°C and the supernatant was isolated and quantified with a Pierce Coomassie Plus (Bradford) assay kit (Thermo Fisher). Protein isolate was mixed with 4x Laemmli Loading buffer (Biorad) and 2-Mercaptoethanol (Biorad) and boiled at 100°C for 10 min. 100 µg total protein from muscle biopsies or 60 µg from liver biopsies was loaded into each well of a 4-15% Mini Protean TGX gel (Biorad) with Tris-Glycine-SDS buffer (Biorad) and electrophoresed for 60 min at 100 V. Protein from muscle biopsies was transferred to

nitrocellulose membranes overnight at 34V while that from liver biopsies was transferred at 65V for 1 hour 30 minutes in a 1X tris-glycine transfer buffer containing 10% methanol and 0.1% SDS at 4°C. The blot was blocked for 1 hour in 5% milk-TBST. Blots were probed with rabbit anti-dystrophin (1:200, Abcam 15277), rabbit anti-GAPDH (1:4000, Cell Signaling 2118S), rabbit anti-OTC (1:800, Abcam 203859) and mouse anti-ADAR2 (1:150, Santa Cruz Biotechnology 73409) overnight at 4°C in 5% milk-TBST. Blots were washed with TBST and then incubated with anti-rabbit or anti-mouse horseradish peroxidase-conjugated secondary antibodies (Cell Signaling) for 1 hour in 5% milk-TBST. After washing with TBST, blots were visualized using SuperSignal West Femto Chemiluminescent Substrate (Thermo Fisher) and X-Ray films.

**[0346] *Statistics and Reproducibility***

**[0347]** *In vitro* experiments: *In vitro* experiments were carried out once with a minimum of 3 independent replicates. *In vivo* experiments: For the *mdx* mouse model, ADAR2 and MCP-ADAR1 (E1008Q) NLS based experiments were carried out twice. Both rounds of experiments yielded consistent RNA editing efficiencies, dystrophin immunofluorescence and dystrophin restoration as seen by western blots. ADAR2 (E488Q) and CRISPR-Cas9 based experiments were carried out once. For the *spf<sup>ash</sup>* mouse model, experiments were carried out twice, based on the availability of mice. RNA editing efficiencies of the OTC transcript, both the spliced and pre-mRNA were consistent in both rounds of experiments. RT-PCR and Western blots were carried out on animals in experimental set 1.

**[0348]** *1. Quantification of RNA A->G editing*

**[0349]** *(a) RNA-seq read alignment*

**[0350]** RNA-seq read pairs with 100 bases per read mate were aligned to the GRCh38 reference genome using STAR aligner version 2.6.0c (Dobin A et al 2013). The genome index was built using primary assembly annotations from GENCODE release 28 (GRCh38.p12). Default parameters were used to run STAR, except for the following relevant settings: readMapNumber=-1, alignSJoverhangMin=5, alignSJDBoverhangMin=1, alignEndsType=EndToEnd, outFilterMismatchNmax=10, outFilterMultimapNmax=1, outSAMunmapped=None, outSAMmultNmax=1. The reads of the resulting uniquely aligned pairs were sorted by genomic coordinate using samtools sort (Li H. et al 2009). Duplicated read pairs were marked using samtools markdup and were removed from subsequent analysis. Tallies of total, aligned, duplicated, and remaining reads (not pairs) are reported for each sample in **Table 4**.

[0351] (b) Selection of reference sites for quantification of editing yields

[0352] The assessment of sites with significant changes in A-to-G editing yields (see below) is sensitive to the number of uniquely aligned reads available for each sample. To minimize potential biases when comparing different samples in terms of significantly edited sites, the uniquely aligned reads for each HEK293T sample were down-sampled using samtools view with option -s and the down-sampling fractions reported in **Table 4**. These fractions were calculated by dividing the smallest number of uniquely aligned reads among samples by the number of uniquely aligned reads available for the sample being down-sampled. Down-sampling was not performed on the reads of the control sample, the first in **Table 4**. The down-sampled reads were then processed using samtools mpileup. The output of this tool was parsed to extract the counts of each base found in the aligned reads at each A-site and T-site in the GRCh38 reference genome sequence. Insertions and deletions were ignored. Reference sites with read coverage less than 10 were omitted from downstream analysis. The number of remaining reference A- and T-sites with read coverage of at least 10 varied by ~15% across the samples listed in **Table 4**. Without down-sampling, such number was found to vary by ~50%. From the reference A- and T-sites with read coverage of at least 10, a final list of total sites (A-sites and T-sites) was selected by choosing those sites that were common to all samples and for which at least one G or C was observed at a reference A- or T-site, respectively, in the aligned reads of at least one sample. The other sites, those not common to all samples or with zero observed editing events in all samples, were discarded.

[0353] (c) Assessment of significant changes in A-to-G editing yields

[0354] To uncover significant changes in A-to-G editing yields, several pairs of control and treatment samples were considered. For each pair, the control sample was the first sample listed in **Table 4**, while the treatment sample was one of the samples shown in **FIG. 32**. For each pair of compared samples, and for each reference A-site selected as described above, a Fisher exact test was carried out using a 2x2 contingency matrix  $C$  with entries defined as follows:  $C_{1,1}$  = count of bases other than G observed in the control sample,  $C_{2,1}$  = count of G bases observed in the control sample,  $C_{1,2}$  = count of bases other than G observed in the test sample,  $C_{2,2}$  = count of G bases observed in the test sample. A similar contingency matrix was used for each selected reference T-site, except that G was replaced with C in the above definitions. The p-values calculated for all selected reference sites and for a given comparison of samples were adjusted for multiple testing using the Benjamini-Hochberg method. A-sites and T-sites with adjusted p-values less than a false discovery rate (FDR) of

1% and with a fold change of at least 1.1 in editing yield were deemed to have a significant change in A-to-G editing yield on forward and reverse transcripts, respectively. The counts of these sites for each comparison of samples are shown as  $N_{\text{sig}}$  in **FIG. 38 – FIG. 42**, and are reported under the column “changed sites” in **Table 5**. The total number of reference sites with a significant change in A-to-G editing yield was computed. The editing yields at these sites were used to construct the distributions shown in **FIG. 32**. The on-target A-to-G editing yields shown as blue circles in **FIG. 32** and **FIG. 38 – FIG. 42** were estimated for each sample as  $C_{2,2} / (C_{1,2} + C_{2,2})$  using counts observed at the intended target A-site in the RAB7A transcript. These values are reported under the column “editing yield” in **Table 5**. The 1-based genomic coordinate of the intended target A-site was found to be chr3:128814202 by submitting the following sequence to BLAT after selecting reference assembly hg38:

AGCGGCAGTATTCTGTACAGTAGACACAAGAATTATGTACGCCTTTTATCA (SEQ ID NO:151).

**[0355] FIG. 8** – Engineering GluR2 adRNAs: scaffold domain engineering. Sequence information of adRNA scaffolds: ADAR recruiting domain, antisense RNA targeting domain and the cytosine mismatch highlighted. Base pairs mutated to create stabilized scaffolds are numbered and highlighted in red, and the editing inducer element motif is shown in green. Quantification of editing efficiency of thus generated scaffolds for the OTC reporter transcript quantified by Sanger sequencing is shown. Values represent mean +/- SEM (n=3). Experiments were carried out in HEK 293T cells.

**[0356] FIG. 34** – Engineering GluR2 adRNAs: antisense domain engineering. (a) Optimization of adRNA antisense region using adRNA scaffold 2: length and distance from the ADAR2 recruiting region were systematically varied. Values represent mean +/- SEM (n=3). (b) U6 promoter transcribed adRNAs with progressively longer antisense domain lengths, in combination with zero, one or two GluR2 domains were evaluated for their ability to induce targeted RNA editing with or without exogenous ADAR2 expression. Values represent mean +/- SEM (n=3). A portion of this data is reused in **FIG. 32B**. All the above experiments were carried out in HEK 293T cells. (c) Experimental confirmation of expression of endogenous ADAR1 and ADAR2 (relative to GAPDH) in HEK 293T and HeLa cell lines. Observed levels were similar to those documented in The Human Protein Atlas (see world-wide-web at [proteatlas.org](http://proteatlas.org)).

**[0357] FIG. 35** – Engineering MS2 adRNAs. (a) Systematic evaluation of antisense RNA targeting domain of the MS2 adRNA. Values represent mean  $\pm$  SEM (n=3). (b) On-target RNA editing by MCP-ADAR2 DD-NLS requires co-expression of the MS2 adRNA. Values represent mean  $\pm$  SEM (n=3). Experiments were carried out in HEK 293T cells.

**[0358] FIG. 36** – Analysis of RNA editing yields across a panel of targets. **(A)** Comparison of RNA editing efficiency of the OTC reporter transcript by GluR2 adRNA and MS2 adRNA guided RNA editing constructs as well as the Cas13b based REPAIR construct. Values represent mean  $\pm$  SEM (n=6 for reporter and Cas13b based constructs, n=3 for other constructs). **(B)** Chemically synthesized adRNAs versions were tested against a panel of mRNAs with or without exogenous ADAR2 expression. The exact chemical modifications are stated in the figure along with the source of adRNA. Values represent mean  $\pm$  SEM (n=3). **(C)** Analysis of RNA editing yields across a spectrum of endogenous targets chosen to cover a range of expression levels. U6 transcribed long adRNAs with none or two GluR2 domains were also evaluated against multiple endogenous mRNA targets with or without exogenous ADAR2 expression. Editing is observed at tested loci even in the absence of exogenous ADAR2 expression. Values represent mean  $\pm$  SEM (n=3). Experiments were carried out in HEK 293T cells.

**[0359] FIG. 37** – ADAR2 variants and their impact on editing and specificity. **(A)** Comparison of on target RNA editing and editing in flanking adenosines of the RAB7A transcript by GluR2 adRNA and MS2 adRNA guided RNA editing constructs as well as the Cas13b based REPAIR construct. Mean (n=3) editing yields are depicted. Experiments were carried out in HEK 293T cells and editing efficiency was calculated as a ratio of Sanger peak heights G/(A+G). **(B)** ADAR2 (E488Q) exhibits higher efficiency than the ADAR2 in the *in vitro* editing of the spf<sup>ash</sup> OTC reporter transcript (p=0.037, unpaired t-test, two-tailed); values represent mean  $\pm$  SEM (n=3), and **(C)** *mdx* DMD reporter transcript (p=0.048, p=0.012 respectively, unpaired t-test, two-tailed); values represent mean  $\pm$  SEM (n=3). **(d)** Comparison of the editing efficiency and specificity profiles of the ADAR2, ADAR2 (E488Q) and the ADAR2 ( $\Delta$ 1-138) for the OTC reporter transcript (upper panel) and endogenous RAB7A transcript (lower panel). Heatmap indicates the A->G edits in the vicinity of the target (red arrow). Values represent mean  $\pm$  SEM (n=3). Experiments were carried out in HEK 293T cells and editing efficiency was calculated as a ratio of Sanger peak heights G/(A+G).

**[0360] FIG. 38** – Transcriptome scale specificity profiles of RNA editing approaches (Cas13b-ADAR REPAIR +/- gRNA). 2D histograms comparing the transcriptome-wide A->G editing yields observed with each Cas13b-ADAR2 construct (y-axis) to the yields observed with the control sample (x-axis). Each histogram represents the same set of 8,729,464 reference sites, where read coverage was at least 10 and at least one putative editing event was detected in at least one sample. Bins highlighted in red contain sites with significant changes in A->G editing yields when comparing treatment to control sample. Red crosses in each plot indicate the 100 sites with the smallest adjusted p-values. Blue circles indicate the intended target A-site within the RAB7A transcript. Large counts in bins near the lower-left corner likely correspond not only to low editing yields in both test and control samples, but also to sequencing errors and alignment errors. Large counts in bins near the upper-right corner of each plot likely correspond to homozygous single nucleotide polymorphisms (SNPs), as well as other differences between the reference genome and the genome of the HEK293T cell line used in the experiments.

**[0361] FIG. 39** – Transcriptome scale specificity profiles of RNA editing approaches (ADAR2 +/- adRNA). The version used for these studies is GluR2 adRNA(1,20,6). 2D histograms comparing the transcriptome-wide A->G editing yields observed with each ADAR construct (y-axis) to the yields observed with the control sample (x-axis). More details are provided in **FIG. 38**.

**[0362] FIG. 40** - Transcriptome scale specificity profiles of RNA editing approaches (MCP-ADAR1 DD +/- adRNA). 2D histograms comparing the transcriptome-wide A->G editing yields observed with each ADAR construct (y-axis) to the yields observed with the control sample (x-axis). More details are provided in **FIG. 38**.

**[0363] FIG. 41** - Transcriptome scale specificity profiles of RNA editing approaches (MCP-ADAR2 DD +/- adRNA). 2D histograms comparing the transcriptome-wide A->G editing yields observed with each ADAR construct (y-axis) to the yields observed with the control sample (x-axis). More details are provided in **FIG. 38**.

**[0364] FIG. 42** - Variation of transcriptome scale editing specificity with construct features. (A) Each point in the box plots corresponds to the fraction of edited sites for one of the MCP-ADAR constructs listed in **FIG. 32**. The fraction of edited sites for each construct was calculated by dividing the number of reference sites with significant changes in A-to-G editing yield (see Table 3) by the total number 8,729,464 of reference sites considered. Construct features indicated on the horizontal axes were compared using the Mann-Whitney

U test, yielding p-values of 0.16 for NLS vs. NES, 0.0070 for ADAR1 vs. ADAR2, 0.72 for “- adRNA” vs. “+ adRNA”, and 0.038 for “ADAR WT” vs. “ADAR E>Q” (n=8 for all conditions). **(B)** 2D histograms comparing the transcriptome-wide A->G editing yields observed with each construct (y-axis) to the yields observed with the control sample (x-axis). More details are provided in **FIG. 38**. Inset shows violin plots representing distributions of A->G editing yields observed at reference sites where at least one treatment sample was found to have a significant change (Fisher’s exact test, FDR = 1%) in editing yield relative to the control sample. Blue circles indicate editing yields at the target A-site within the RAB7A transcript. To better visualize the shapes of the distributions, their maximum extent along the y-axis was equalized across plots, and were truncated at 60% yield. Samples here correspond to 293Ts transfected with long antisense domain bearing adRNAs that can enable RNA editing via exogenous and/or endogenous ADAR recruitment.

**[0365] FIG. 43 - Optimization and evaluation of dystrophin RNA editing experiments *in vitro* and *in vivo* in *mdx* mice. (A)** Schematic of RNA editing utilizing the full length ADAR2 along with an engineered adRNA or a reverse oriented adRNA (radRNA); (ii) RNA editing efficiencies of amber and ochre stop codons, in one-step and two-steps. Experiments were carried out in HEK 293T cells. Values represent mean +/- SEM (n=3). **(B)** RNA editing of ochre codons requires two cytosine mismatches in the antisense RNA targeting domains of adRNA or radRNA to restore GFP expression. Experiments were carried out in HEK 293T cells. Values represent mean +/- SEM (n=3). **(C)** Schematic of the AAV vectors utilized for *in vivo* delivery of adRNA and ADAR2, and *in vitro* optimization of RNA editing of amber and ochre stop codons in the presence of one or two copies of the adRNA, delivered via an AAV vector (p=0.0003, p=0.0001, p=0.0015 respectively, unpaired t-test, two-tailed). Experiments were carried out in HEK 293T cells. Values represent mean +/- SEM (n=3 for reporters, n=6 for other conditions). **(D)** Representative Sanger sequencing plot showing editing of the ochre stop codon (TAA->TGG) in the *mdx* DMD reporter transcript (quantified by NGS). Experiments were carried out in HEK 293T cells (n=3). **(E)** Representative example of *in vivo* RNA editing analyses of treated *mdx* mice (quantified using NGS).

**[0366] FIG. 44 - Immunofluorescence and Western blot analyses of *in vivo* dystrophin RNA editing experiments in *mdx* mice. (A)** Immunofluorescence staining for dystrophin in the TA muscle shows partial restoration of expression in treated samples (intra-muscular injections of AAV8-ADAR2, AAV8-ADAR2 (E488Q), AAV8-MCP-ADAR1 (E1008Q) NLS). Partial restoration of nNOS is localization also seen in treated samples (scale bar: 250µm). **(B)**



Western blots showing partial recovery of dystrophin expression (1-2.5%) in TA muscles of *mdx* mice injected with both components of the editing machinery, the enzyme and adRNA, and stable ADAR2 expression in injected TA muscles up to 8 weeks post injections. (C)

Western blot showing partial restoration of dystrophin expression (10%) using AAV8-CRISPR.

**[0367] FIG. 45** - Optimization and evaluation of OTC RNA editing experiments *in vitro* and *in vivo* in *spf<sup>ash</sup>* mice. (A) Representative Sanger sequencing plot showing correction of the point mutation in the *spf<sup>ash</sup>* OTC reporter transcript (quantified using NGS). Experiments were carried out in HEK 293T cells (n=3). (B) Representative example of *in vivo* RNA editing analyses of treated *spf<sup>ash</sup>* mice showing correction of the point mutation in the correctly spliced OTC mRNA (quantified using NGS). (C) *In vivo* RNA correction efficiencies in the OTC pre-mRNA in the livers of treated adult *spf<sup>ash</sup>* mice (retro-orbital injections of AAV8-ADAR2 and AAV8-ADAR2 (E488Q). Values represent mean +/- SEM (n=4, 4, 3, 3, 4, 5 independent animals respectively). (D) PCR products showing the correctly and incorrectly spliced OTC mRNA. The incorrectly spliced mRNA is elongated by 48 base pairs. Fraction of incorrectly spliced mRNA is reduced in mice treated with adRNA+ADAR2 (E488Q). (E) Western blot for OTC shows partial restoration (2.5%-5%) of expression in treated adult *spf<sup>ash</sup>* mice and stable ADAR2 (E488Q) expression three weeks post injections.

**[0368] FIG. 46** - Toxicity analyses of *in vivo* RNA editing experiments. Summary of animal experiments documenting the route of AAV administration, construct delivered, and health of injected mice 3 weeks post injections.

### Example 5

**[0369]** Disclosed herein are the results of experiments in an *mdx* mouse model of Duchenne muscular dystrophy using an E100Q mutant of ADAR1 (comprised in MCP-ADAR1 (E100Q)). In some cases, this mutant can improve editing yields *in vivo* compared to ADAR2 and an E488Q mutant of ADAR2. Further disclosed herein is an application of the ADAR system in a manner to alter splicing patterns by editing a splice acceptor site or branch point in an intron, thereby resulting in exon skipping. In addition, methods of using APOBECs are contemplated and exemplified. Demonstrated herein are examples in which the creation of local structure of alipoprotein B mRNA that ACF-APOBEC complex binds to, at an mRNA of interest, can enable C → T RNA editing. Further disclosed herein are method of utilizing ADAR enzymes for programmable editing of both RNA and DNA.

[0370] To utilize the ADAR editing system, with or without exogenous ADAR1 or ADAR2, adRNAs can be generated. The examples provided specific adRNAs of interest for a given target. Further, provided herein are exemplary chemically synthesized adRNAs with 2'-O-methyl 3' phosphothiorate modifications in the first and last 3, 6, 9, or 12 nucleotides, and/or 2'-O-methyl modifications throughout the antisense region except the 3 nucleotides centered around the mis-match site (underlined below), coupled with targeting moieties such as GalNc or cholesterol or cell-penetrating peptides can be used to engineer targeted RNA editing in cells or *in vivo* (especially in liver, lung and brain) with or without exogenous ADAR1/2 over-expression.

**General Design:**

[0371] **adRNA22:** GTGGAAGAGgAgAACAATATGCTAAATGTTGTTcTcGTcTCCC  
ACNNNNNNCNNNNNNNNNNNNNNNN (SEQ ID NO:152)

[0372] **adRNA32:**  
GGTGTGCGAGAAgAGgAgAACAATATGCTAAATGTTGTTcTcGTcTCCTCGACACCNN  
NNNNCNNNNNNNNNNNNNNNN (SEQ ID NO:153)

[0373] specific examples:

[0374] **adRNA22\_RAB7A:** GTGGAAGAGgAgAACAATATGCTAAATGTTGTTcTcGTcT  
CCCACTGCCGCCAGCTGGATTTC (SEQ ID NO:154)

[0375] **adRNA32\_RAB7A:** GGTGTGCGAGAAgAGgAgAACAATATGCTAAATGTTGTTc  
TcGTcTCCTCGACACCTGCCGCCAGCTGGATTTC (SEQ ID NO:155)

[0376] **adRNA22\_CKDN2A:**  
GTGGAAGAGgAgAACAATATGCTAAATGTTGTTcTcGTcTCCCACCTCCTCCACCCG  
ACCCCGGG (SEQ ID NO:156)

[0377] **adRNA32\_CKDN2A:**  
GGTGTGCGAGAAgAGgAgAACAATATGCTAAATGTTGTTcTcGTcTCCTCGACACCCT  
CCTCCACCCGACCCCGGG (SEQ ID NO:157)

[0378] The examples provided herein below describe a multiplicity of steps taken by to further elucidate the embodiments disclosed herein, such as:

- (1) delivering MCP-ADAR constructs with MS2-adRNA targeting the premature stop codon in the dystrophin transcript of *mdx* mice;
- (2) testing out exon skipping via creation of point mutations in the *mdx* mouse model of Duchenne muscular dystrophy, with adRNAs delivered along with ADAR2/ADAR2(E488Q) or MS-adRNAs delivered with MCP-ADAR1/ADAR (E1008Q) or ms-adRNAs delivered

with MCP-ADAR2/ADAR2 (E488Q) – utilized to create A→G substitutions at the splice site;

- (3) testing editing of splice site mutations in the OTC transcript of spc-ash mice;
- (4) testing out editing of a mutant splice site to alter splicing patterns in the spf-ash mouse model of ornithine transcarbamylase deficiency;
- (5) targeting the mouse SCN9A transcript to engineer insensitivity to pain by creating A→G substitutions at the splice sites to knockdown the SCN9A transcript;
- (6) engineering APOBEC gRNAs to create the local structure of the alipoprotein B mRNA at a target mRNA of interest; and
- (7) testing out ADARs in editing RNA and DNA by utilizing RNA, DNA, and RNA-DNA hybrids to recruit the ADAR.

[0379] In the examples, it was observed that enhanced RNA yields and protein expression were achieved via the use of MS2-adRNAs and MCP-ADAR1 (E1008Q) while editing the premature stop codon in a dystrophin transcript of *mdx* mice. Upon delivery of the ADAR2 (E488Q) with an adRNA, a reduction in incorrectly spliced OTC mRNA in an spf-ash mouse model was also confirmed.

[0380] The disclosure thus describes a system that can enable site-specific A->G editing of RNA. Such an approach can be used to edit splice acceptor sites and branch points to alter splicing. This can also enable exon skipping. Further embodiments contemplate use of A->G editing of DNA. In addition, the disclosure provides the potential for C->T edits in RNA via the use of APOBEC1 expressed along with ACF1.

[0381] The disclosure describes the first site-specific RNA editing *in vivo*. In fact, utilization of MCP-ADAR1 (E1008Q) demonstrates higher editing efficiencies than prior constructs in the *mdx* mouse model of muscular dystrophy. Further test, both *in vitro* and *in vivo*, examiner (1) RNA editing for altering splicing, as well as (2) efficacy of C->T editing via APOBECs along with the overexpression of ACF and (3) ADARs for editing DNA. [SEP]

[0382] Compared to other ADAR2 systems (*e.g.* Stafforst, Zhang, and Rosenthal labs) and Cas13d based inhibition of splicing, the present ADAR system is unique.

[0383] As described herein, the ADAR system comprise, or alternatively consist essentially of, or yet further consist of: a RNA targeting domain complementary to the target RNA and one or more ADAR recruiting domains that enable recruitment of ADARs. Upon introduction of these components, the ADAR enzyme can catalyze the conversion of a target adenosine to inosine and thereby repair point mutations. The disclosure describes the use of

ADAR2 or MCP-ADAR1/2-NLS or their hyperactive mutants to create a A->G edits at splice acceptor sites and/or branch point in introns to engineer exon skipping. Skipping symmetric exons results in formation of a truncated protein. This strategy can be used to skip exon 23 in the dystrophin transcript of the *mdx* mouse model of DMD which bears a premature stop codon in this exon. Skipping exon 23 results in the translation of a truncated dystrophin protein which is functional. In addition, skipping asymmetric exons leads to frameshift mutations. Thus, by skipping essential exons or asymmetric exons it can be possible to engineer gene knockdowns. In addition, editing of the start codon ATG and Kozak/Shine-Dalgarno sequences, can also help alter translation efficiencies and result in knockdown of genes. Exon skipping strategy can be used to target the SCN9A transcript to engineer insensitivity to pain. The local structure of the apolipoprotein B mRNA that the ACF-APOBEC complex binds to, at the mRNA target of interest, can be created for carefully positioning the C to be edited at a position analogous to the naturally occurring site in the apolipoprotein B mRNA. This is achieved by overexpression of a pair of adRNAs to create an apolipoprotein B mRNA like structure that can be edited by overexpression of MCP-ACF1 and APOBEC1.

**Example 6 – Exon skipping via creation of splice acceptor and/or branch point mutations**

[0384] The disclosure demonstrates the use of ADAR2 or MCP-ADAR1/2-NLS or their hyperactive mutants to create a A->G edits at splice acceptor sites and/or branch point in introns to engineer exon skipping. Skipping symmetric exons results in formation of a truncated protein. In the present example, this strategy can be used to skip exon 23 in the dystrophin transcript of the *mdx* mouse model of DMD which bears a premature stop codon in this exon. Skipping exon 23 results in the translation of a truncated dystrophin protein which is functional. In addition, skipping asymmetric exons leads to frameshift mutations. Thus, by skipping essential exons or asymmetric exons it can be possible to engineer gene knockdowns. In addition, editing of the start codon ATG and Kozak/Shine-Dalgarno sequences, helps alter translation efficiencies and result in knockdown of genes. This strategy can be used to target the SCN9A transcript to engineer insensitivity to pain.

[0385] Exemplary sequences for achieving this end are provided herein below.

**[0386] SEQUENCE SET 1: Exemplary adRNA Sequences**

[0387] adRNA sequences for exon 23 skipping in the dystrophin transcript in *mdx* mice (SEQ ID NO:158-159)

|                           |                                                                                     |
|---------------------------|-------------------------------------------------------------------------------------|
| adDM<br>D-<br>I22-<br>E23 | <b>GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACG<br/>AGCC<u>C</u>AAAATTAAATAGA</b> |
| adDM<br>D-<br>E23-<br>I23 | <b>GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCAC<br/>TACCC<u>G</u>AAATTTTCGAAGT</b> |

[0388] MS2-adRNA sequences for exon 23 skipping in the dystrophin transcript in *mdx* mice (SEQ ID NO:160-161)

|                             |                                                                                   |
|-----------------------------|-----------------------------------------------------------------------------------|
| MS2-<br>DMD-<br>I22-<br>E23 | <b>aACATGAGGATCACCCATGTcGAGCC<u>C</u>AAAATTAAATAGAAaACAT<br/>GAGGATCACCCATGTc</b> |
| MS2-<br>DMD-<br>E23-<br>I23 | <b>aACATGAGGATCACCCATGTcTTACCC<u>G</u>AAATTTTCGAAGTaACATG<br/>AGGATCACCCATGTc</b> |

[0389] MS2-adRNA sequence for editing the premature stop codon in *mdx* mice (SEQ ID 162)

|             |                                                                                 |
|-------------|---------------------------------------------------------------------------------|
| MS2_D<br>MD | <b>aACATGAGGATCACCCATGTcCCATT<u>C</u>ATTGCTCTTCAAaACATG<br/>AGGATCACCCATGTc</b> |
|-------------|---------------------------------------------------------------------------------|

[0390] adRNA sequences for exon skipping in the SCN9A transcript in mice (SEQ ID NO:163-167)

|                           |                                                                               |
|---------------------------|-------------------------------------------------------------------------------|
| adRNA_SCN9A_<br>I8 E9 F   | <b>GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATC<br/>CCACACTGCCCACAGATGAACAAG</b> |
| adRNA_SCN9A_<br>I4 E5 F   | <b>GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATC<br/>CCACTGTACCCGAAGGAGAGAATA</b> |
| adRNA_SCN9A_<br>I2 E3 F   | <b>GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATC<br/>CCACAAAGTCCGAGGAGGAAAAAG</b> |
| adRNA_SCN9A_<br>I5 E6 F   | <b>GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATC<br/>CCACAATACCCGTAGGATTAAATC</b> |
| adRNA_SCN9A_<br>I13 E14 F | <b>GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATC<br/>CCACAAGTTCCGAAACAAACAA</b>   |

[0391] MS2-adRNA sequences for exon skipping in the SCN9A transcript in mice (SEQ ID NO:168-172)

|                             |                                                                            |
|-----------------------------|----------------------------------------------------------------------------|
| MS2_adRNA_SCN9<br>A_I8_E9_F | <b>aACATGAGGATCACCCATGTcACTGCCCACAGATGAACAA<br/>GaACATGAGGATCACCCATGTc</b> |
| MS2_adRNA_SCN9<br>A_I4_E5_F | <b>aACATGAGGATCACCCATGTcTGTACCCGAAGGAGAGAAT<br/>AaACATGAGGATCACCCATGTc</b> |
| MS2_adRNA_SCN9<br>A_I2_E3_F | <b>aACATGAGGATCACCCATGTcAAAGTCCGAGGAGGAAAA<br/>AGaACATGAGGATCACCCATGTc</b> |

|                               |                                                                    |
|-------------------------------|--------------------------------------------------------------------|
| MS2_adRNA_SCN9<br>A_15_E6_F   | aACATGAGGATCACCCATGTcAATACCCGTAGGATTAAAT<br>CaACATGAGGATCACCCATGTc |
| MS2_adRNA_SCN9<br>A_113_E14_F | aACATGAGGATCACCCATGTcAAGTTCCGGAAAACAAACA<br>AaACATGAGGATCACCCATGTc |

**Example 7 – C → T editing via APOBECs**

[0392] The local structure of the apolipoprotein B mRNA that the ACF-APOBEC complex binds to, at the mRNA target of interest, can be created for carefully positioning the C to be edited at a position analogous to the naturally occurring site in the apolipoprotein B mRNA. This is achieved by overexpression of a pair of gRNAs to create a apolipoprotein B mRNA like structure that can be edited by overexpression of one or more of the following combinations:

1. MCP-ACF1 and APOBEC1
2. MCP-ACF1 and MCP-APOBEC1
3. MCP-linker-ACF-linker-APOBEC1
4. MCP-APOBEC1 and ACF1

**[0393] SEQUENCE SET 2: Exemplary C → T Editing Sequences**

[0394] MS2-adRNA sequences for C-T editing (SEQ ID NO:173-174)

|                |                                                                          |
|----------------|--------------------------------------------------------------------------|
| MS2-gRN<br>A-1 | acatatatgataNNNNNNNNNNNaACATGAGGATCACCCATGTcNNNNNNNNNN                   |
| MS2-gRN<br>A-2 | NNNNNNNNNNNNNNNNNNNNNaACATGAGGATCACCCATGTcNNNNNN<br>NNNNNttgatcagtatatta |

**Example 8 – Creation of point mutations relevant for cancers**

[0395] Several genes involved in cancer pathways harbor single amino acid substitution. Creation of dominant negative mutants, constitutionally active mutants and catalytically inactive mutants is possible by creating A->G substitutions in the mRNA sequences of these genes. Some of these genes include KRAS, HRAS, JAK2, GSK3 $\beta$ ,  $\beta$ -catenin, SmoM2, Caspase3, Caspase 8, TGF- $\beta$ , p53.

**Example 9 – Editing DNA and both strands of DNA/RNA hybrids**

[0396] Since ADARs have been shown to edit double stranded RNA as well as both strands of a DNA-RNA hybrid, it is possible to recruit ADARs via single stranded DNA or DNA-RNA hybrids to edit both DNA and RNA. This can be used to modify the current adenine base editing approach to Cas9 (or Cpf1)-ADAR-deaminase domain fusions (ADAR1, ADAR2 and their catalytically active mutants (E1008Q) and E488Q), and instead of targeting

the ssDNA displaced strand by current base editors, the gRNA bound strand with a A-C bulge can be targeted, ideally in the first 10bp close to the 5' end of the gRNA.

#### **Example 10 – Ornithine transcarbamylase deficiency**

[0397] ADAR2 (E488Q) along with an adRNA was delivered in *spf<sup>ash</sup>* mice. In addition to the correctly spliced mRNA, *spf<sup>ash</sup>* mice harbor an incorrectly spliced, elongated mRNA variant which is formed due to the use of a cryptic splice site, 48 base pairs into exon 4. Upon delivery of the ADAR2 (E488Q) with an adRNA, a reduction in the incorrectly spliced product was confirmed. Highly efficient RNA editing yields of up to 33.9% in the spliced mRNA was also observed. In the pre-mRNA yields of up to 8% were observed. Protein restoration of 2.5-5%, within 3 weeks of injections was also observed. This demonstrates the utility of RNA editing in the correction of splice site mutations and altering splicing patterns. (FIG. 50A-D)

#### **Example 11 – Duchenne muscular dystrophy**

[0398] MCP-ADAR1(E1008Q) was tested with MS2-adRNA in *mdx* mice. Enhanced RNA editing yields and restoration of dystrophin expression over the ADAR2 and ADAR2 (E488Q) enzymes were observed. (FIG. 51A-B)

#### **Equivalents**

[0399] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this technology belongs.

[0400] The technology illustratively described herein can suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms “comprising,” “including,” “containing,” *etc.* shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the technology claimed.

[0401] Thus, it should be understood that the materials, methods, and examples provided here are representative of preferred aspects, are exemplary, and are not intended as limitations on the scope of the technology.

[0402] The technology has been described broadly and generically herein. Each of the narrower species and sub-generic groupings falling within the generic disclosure also form

part of the technology. This includes the generic description of the technology with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0403] In addition, where features or aspects of the technology are described in terms of Markush groups, those skilled in the art will recognize that the technology is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0404] All publications, patent applications, patents, and other references mentioned herein are expressly incorporated by reference in their entirety, to the same extent as if each were incorporated by reference individually. In case of conflict, the specification, including definitions, will control.

### **Specific Embodiments**

[0405] A number of compositions, methods and systems are disclosed herein. Specific exemplary embodiments of these compositions, methods and systems are disclosed below.

#### **[0406] PART 1**

[0407] *Embodiment 1.* An engineered ADAR1 or ADAR2 guide RNA (“adRNA”) comprising: a sequence complementary to a target RNA.

[0408] *Embodiment 2.* The engineered adRNA of embodiment 1, further comprising an ADAR2 recruiting domain derived from GluR2 mRNA.

[0409] *Embodiment 3.* The engineered adRNA of embodiment 1, further comprising two MS2 hairpins flanking the sequence complementary to a target RNA.

[0410] *Embodiment 4.* The engineered adRNA of any one of embodiments 1-3, wherein the sequence complementary to the target RNA comprises between about 20 to 100 base pairs.

[0411] *Embodiment 5.* The engineered adRNA of embodiment 2 or 4, wherein the ADAR2 recruiting domain derived from GluR2 mRNA is located at the 5' end or the 3' end of the engineered adRNA.

[0412] *Embodiment 6.* The engineered adRNA of embodiment 5, comprising a GluR2 mRNA at both the 5' end and the 3' end of the engineered adRNA.

[0413] *Embodiment 7.* The engineered adRNA of embodiment 5 or 6, further comprising an editing inducer element.

[0414] *Embodiment 8.* The engineered adRNA of any of the preceding embodiments, wherein the target RNA is ornithine transcarbamylase.



[0415] *Embodiment 9.* An engineered ADAR2 guide RNA (“adRNA”) encoded by a polynucleotide sequence selected from the group of sequences provided in **TABLE 1** or **FIG. 2**, or an equivalent of each thereof.

[0416] *Embodiment 10.* An isolated polynucleotide encoding the engineered adRNA of any one of embodiments 1-9, or an equivalent of each thereof.

[0417] *Embodiment 11.* A vector comprising one or more of the isolated polynucleotide of embodiment 10 or the polynucleotide sequence encoding the engineered adRNA of embodiment 9 and optionally regulatory sequences operatively linked to the isolated polynucleotide.

[0418] *Embodiment 12.* The vector of embodiment 11, wherein the regulatory sequences comprise a promoter, an enhancer element and/or a reporter.

[0419] *Embodiment 13.* The vector of embodiment 12, wherein the promoter is a human U6, a mouse U6 promoter or a CMV promoter.

[0420] *Embodiment 14.* The vector of any one of embodiments 11-13, further comprising a detectable marker or a purification marker.

[0421] *Embodiment 15.* The vector of embodiment 14, wherein the vector is a plasmid or a viral vector.

[0422] *Embodiment 16.* The vector of embodiment 15, wherein the vector is selected from a group consisting of a retroviral vector, a lentiviral vector, an adenoviral vector, and an adeno-associated viral vector.

[0423] *Embodiment 17.* A recombinant cell further comprising the vector of any one of embodiments 11-16, wherein the engineered adRNA is recombinantly expressed.

[0424] *Embodiment 18.* A composition comprising a carrier and one or more of the engineered adRNA of any one of embodiments 1-9, the isolated polynucleotide of embodiment 10, the vector of any one of embodiments 11-16 or the recombinant cell of embodiment 17.

[0425] *Embodiment 19.* The composition of embodiment 18, further comprising a chemotherapeutic agent or drug.

[0426] *Embodiment 20.* The composition of embodiment 18 or 19, wherein the carrier is a pharmaceutically acceptable carrier or a solid support.

[0427] *Embodiment 21.* A method of modifying protein expression comprising contacting a polynucleotide encoding the protein with the engineered adRNA of any one of embodiments 1-9.

[0428] *Embodiment 22.* The method of embodiment 21, wherein the contacting is *in vitro* or *in vivo*.

[0429] *Embodiment 23.* A method of treating a disease or disorder associated with aberrant protein expression comprising administering to a subject in need of such treatment an effective amount of one or more of the engineered adRNA of any one of embodiments 1-9.

[0430] *Embodiment 24.* The method of embodiment 23, wherein the disease or disorder is Duchenne Muscular Dystrophy.

[0431] *Embodiment 25.* The method of embodiment 23 or 24, wherein the subject is an animal.

[0432] *Embodiment 26.* The method of embodiment 23 or 24, wherein the animal is a mammal.

[0433] *Embodiment 27.* Use of an effective amount of one or more of the engineered adRNA of any one of embodiments 1-9 for treating a disease or disorder associated with aberrant protein expression.

[0434] *Embodiment 28.* The use of embodiment 27, wherein the disease or disorder is Duchenne Muscular Dystrophy.

[0435] *Embodiment 29.* A kit comprising the engineered adRNA of any one of embodiments 1-9, the isolated polynucleotide of embodiment 10, the vector of any one of embodiments 11-16, the recombinant cell of embodiment 17, or the composition of any one of embodiments 18-20 and instructions for use.

[0436] *Embodiment 30.* The kit of embodiment 19, wherein the instructions recite the method of any one of embodiments 21-26.

[0437] *Embodiment 31.* A complex comprising an adRNA of any one of embodiments 1-9, hybridized to a complementary polynucleotide under conditions of high stringency.

[0438] **PART 2**

[0439] *Embodiment 1.* An engineered ADAR2 guide RNA (“adRNA”) comprising: a sequence complementary to a target RNA, and an ADAR2 recruiting domain derived from GluR2 mRNA.

[0440] *Embodiment 2.* The engineered adRNA of embodiment 1, wherein the sequence complementary to the target RNA comprises between about 20 to 100 base pairs.

[0441] *Embodiment 3.* The engineered adRNA of embodiment 1 or embodiment 2, wherein the ADAR2 recruiting domain derived from GluR2 mRNA is located at the 5' end or the 3' end of the engineered adRNA.

[0442] *Embodiment 4.* The engineered adRNA of embodiment 3, comprising a GluR2 mRNA at both the 5' end and the 3' end of the engineered adRNA.

[0443] *Embodiment 5.* The engineered adRNA of any of the preceding embodiments, wherein the target RNA is ornithine transcarbamylase.

[0444] *Embodiment 6.* An engineered ADAR2 guide RNA ("adRNA") encoded by a sequence selected from the group of sequences provided in **TABLE 1** or **FIG. 2**.

[0445] *Embodiment 7.* A method of modifying protein expression comprising contacting a polynucleotide encoding the protein with the engineered adRNA of any one of embodiments 1-6.

[0446] *Embodiment 8.* The method of embodiment 7, wherein the contacting is *in vitro* or *in vivo*.

[0447] *Embodiment 9.* A method of treating a disease or disorder associated with aberrant protein expression comprising administering to a subject in need of such treatment an effective amount of one or more of the engineered adRNA of any one of embodiments 1 to 6.

[0448] *Embodiment 10.* The method of embodiment 9, wherein the subject is an animal.

[0449] *Embodiment 11.* The method of embodiment 9, wherein the animal is a mammal.

[0450] *Embodiment 12.* Use of an effective amount of one or more of the engineered adRNA of any one of embodiments 1 to 6 for treating a disease or disorder associated with aberrant protein expression.

[0451] *Embodiment 13.* A kit comprising the engineered adRNA of any one of embodiments 1 to 6 and instructions for use.

[0452] *Embodiment 14.* The kit of embodiment 13, wherein the instructions recite the method of embodiment 7 or embodiment 9.

[0453] *Embodiment 15.* A composition comprising the AdRNA of any one of embodiments 1-6, and a carrier.

[0454] *Embodiment 16.* The composition of embodiment 15, wherein the carrier is a pharmaceutically acceptable carrier.

[0455] *Embodiment 17.* A complex comprising an AdRNA of any one of embodiments 1-6, hybridized to a complementary polynucleotide under conditions of high stringency.

[0456] **PART 3**

[0457] *Embodiment 1.* An ADAR system for exon skipping comprising an adRNA targeting a splice acceptor and/or a branch point in an intron and, optionally, an ADAR enzyme.

- [0458] *Embodiment 2.* The ADAR system of claim 1, wherein the ADAR enzyme is ADAR1, ADAR2, or a mutant or variant each thereof.
- [0459] *Embodiment 3.* The ADAR system of claim 2, wherein the mutant or variant is selected from ADAR1 (E1008Q) and ADAR2 (E488Q).
- [0460] *Embodiment 4.* The ADAR system of any one of claims 1 to 3, wherein the intron is comprised in a gene selected from dystrophin, SCN9A, or ornithine transcarbamylase.
- [0461] *Embodiment 5.* The ADAR system of any one of claims 1 to 4, wherein the adRNA is selected from **SEQUENCE SET 1**.
- [0462] *Embodiment 6.* A method of treating a disease, disorder, or condition characterized by aberrant gene expression comprising administering the ADAR system of any one of claims 1 to 5.
- [0463] *Embodiment 7.* The method of claim 6, wherein the disease, disorder, or condition is selected from Duchenne muscular dystrophy or ornithine transcarbamylase deficiency.
- [0464] *Embodiment 8.* The method of claim 6 or claim 7, wherein the disease, disorder, or condition is associated with pain.
- [0465] *Embodiment 9.* An APOBEC system for cytosine to thymine editing comprising a pair of gRNA that creates an alipoprotein B mRNA like structure and, optionally, an APOBEC enzyme.
- [0466] *Embodiment 10.* The APOBEC system of claim 9, wherein the pair of gRNA is the pair of sequences provided in **SEQUENCE SET 2**.

#### REFERENCES

- [0467] Montiel-Gonzalez, M. F., Vallecillo-Viejo, I., Yudowski, G. A. & Rosenthal, J. J. C. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18285–90 (2013).
- [0468] Wettengel, J., Reautschnig, P., Geisler, S., Kahle, P. J. & Stafforst, T. Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res.* 45, 2797–2808 (2017).
- [0469] Fukuda, M. *et al.* Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. *Sci. Rep.* 7, 41478 (2017).
- [0470] Heep, M., Mach, P., Reautschnig, P., Wettengel, J. & Stafforst, T. Applying Human ADAR1p110 and ADAR1p150 for Site-Directed RNA Editing—G/C Substitution Stabilizes GuideRNAs against Editing. *Genes (Basel).* 8, 34 (2017).

- [0471] Vallecillo-Viejo, I. C., Liscovitch-Brauer, N., Montiel-Gonzalez, M. F., Eisenberg, E. & Rosenthal, J. J. C. Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol.* 15, 104–114 (2018).
- [0472] Sinnamon, J. R. *et al.* Site-directed RNA repair of endogenous Mecp2 RNA in neurons. *Proc. Natl. Acad. Sci. U. S. A.* 114, E9395–E9402 (2017).
- [0473] Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* eaaq0180 (2017). doi:10.1126/science.aaq0180
- [0474] Montiel-González, M. F., Vallecillo-Viejo, I. C. & Rosenthal, J. J. C. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res.* 44, gkw738 (2016).
- [0475] Hanswillemenke, A., Kuzdere, T., Vogel, P., Jeely, G. & Stafforst, T. Site-Directed RNA Editing *in vivo* Can Be Triggered by the Light- Driven Assembly of an Artificial Riboprotein. doi:10.1021/jacs.5b10216
- [0476] Woolf, T. M., Chase, J. M. & Stinchcomb, D. T. Toward the therapeutic editing of mutated RNA sequences. *Biochemistry* 92, 8298–8302 (1995).
- [0477] Stafforst, T. & Schneider, M. F. An RNA-Deaminase Conjugate Selectively Repairs Point Mutations. *Angew. Chemie Int. Ed.* 51, 11166–11169 (2012).
- [0478] Schneider, M. F., Wettengel, J., Hoffmann, P. C. & Stafforst, T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. *Nucleic Acids Res.* 42, e87 (2014).
- [0479] Azad, M. T. A., Bhakta, S. & Tsukahara, T. Site-directed RNA editing by adenosine deaminase acting on RNA for correction of the genetic code in gene therapy. *Gene Ther.* 24, 779–786 (2017).
- [0480] Vogel, P. *et al.* Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods* 15, 535–538 (2018).
- [0481] Smalley, E. First AAV gene therapy poised for landmark approval. *Nat. Biotechnol.* 2017 3511 (2017).
- [0482] Calcedo, R. & Wilson, J. M. AAV Natural Infection Induces Broad Cross-Neutralizing Antibody Responses to Multiple AAV Serotypes in Chimpanzees. *Hum. Gene Ther. Clin. Dev.* 27, 79–82 (2016).
- [0483] Harbison, C. E. *et al.* Examining the cross-reactivity and neutralization mechanisms of a panel of mabs against adeno-associated virus serotypes 1 and 5. *J. Gen. Virol.* 93, (2012).

- [0484] Riviere, C., Danos, O. & Douar, A. M. Long-term expression and repeated administration of AAV type 1, 2 and 5 vectors in skeletal muscle of immunocompetent adult mice. *Gene Ther.* 13, 1300–1308 (2006).
- [0485] Mays, L. E. & Wilson, J. M. The Complex and Evolving Story of T cell Activation to AAV Vector-encoded Transgene Products. *Mol. Ther.* 19, 16–27 (2011).
- [0486] Chen, J., Wu, Q., Yang, P., Hsu, H. C. & Mountz, J. D. Determination of specific CD4 and CD8 T cell epitopes after AAV2- and AAV8-hF.IX gene therapy. *Mol Ther* 13, 260–269 (2006).
- [0487] Ertl, H. C. J. & High, K. A. Impact of AAV Capsid-Specific T-Cell Responses on Design and Outcome of Clinical Gene Transfer Trials with Recombinant Adeno-Associated Viral Vectors: An Evolving Controversy. *Hum. Gene Ther.* 28, 328–337 (2017).
- [0488] Hinderer, C. *et al.* Severe toxicity in nonhuman primates and piglets following high-dose intravenous administration of an AAV vector expressing human SMN. *Hum. Gene Ther. hum.2018.015* (2018). doi:10.1089/hum.2018.015
- [0489] Mingozzi, F. *et al.* Prevalence and pharmacological modulation of humoral immunity to AAV vectors in gene transfer to synovial tissue. *Gene Ther* 20, 417–424 (2013).
- [0490] Basner-Tschakarjan, E., Bijjiga, E. & Martino, A. T. Pre-clinical assessment of immune responses to adeno-associated virus (AAV) vectors. *Frontiers in Immunology* 5, (2014).
- [0491] Harding, F. A., Stickler, M. M., Razo, J. & DuBridge, R. B. The immunogenicity of humanized and fully human antibodies: Residual immunogenicity resides in the CDR regions. *MAbs* 2, 256–265 (2010).
- [0492] De Groot, a S., Knopp, P. M. & Martin, W. De-immunization of therapeutic proteins by T-cell epitope modification. *Dev. Biol. (Basel)*. 122, 171–194 (2005).
- [0493] Ferdosi, S. R. *et al.* Multifunctional CRISPR/Cas9 with engineered immunosilenced human T cell epitopes. *bioRxiv* (2018).
- [0494] Veronese, F. M. & Mero, A. The impact of PEGylation on biological therapies. *BioDrugs* 22, 315–329 (2008).
- [0495] Armstrong, J. K. *et al.* Antibody against poly(ethylene glycol) adversely affects PEG-asparaginase therapy in acute lymphoblastic leukemia patients. *Cancer* 110, 103–111 (2007).
- [0496] Kurosaki, T., Kometani, K. & Ise, W. Memory B cells. *Nat. Rev. Immunol.* 15, 149–159 (2015).

- [0497] Zabel, F. *et al.* Distinct T helper cell dependence of memory B-cell proliferation versus plasma cell differentiation. *Immunology* 150, 329–342 (2017).
- [0498] Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research* 42, 6091–6105 (2014).
- [0499] Burstein, D. *et al.* New CRISPR–Cas systems from uncultivated microbes. *Nat. Publ. Gr.* 542, (2017).
- [0500] Makarova, K. S., Zhang, F. & Koonin, E. V. SnapShot: Class 2 CRISPR-Cas Systems. *Cell* 168, 328–328.e1 (2017).
- [0501] Ellis, B. L. *et al.* A survey of *ex vivo/in vitro* transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype. *Virol. J.* 10, 74 (2013).
- [0502] Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 32, 511–517 (2015).
- [0503] Andreatta, M. *et al.* Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 67, 641–650 (2015).
- [0504] Truong, D.-J. J. *et al.* Development of an intein-mediated split–Cas9 system for gene therapy. *Nucleic Acids Res.* 43, 6450–6458 (2015).
- [0505] Moreno, A. M. *et al.* In Situ Gene Therapy via AAV-CRISPR-Cas9-Mediated Targeted Gene Regulation. *Mol. Ther.* 0, (2018).
- [0506] Benveniste, O. *et al.* Prevalence of Serum IgG and Neutralizing Factors Against Adeno-Associated Virus (AAV) Types 1,2,5,6,8, and 9 in the Healthy Population: Implications for Gene Therapy Using AAV Vectors. *Hum. Gene Ther.* 21, 704–712 (2010).
- [0507] Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278 (2014).
- [0508] Ran, F. A. *et al.* *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186–190 (2015).
- [0509] Clemente, T., Dominguez, M. R., Vieira, N. J., Rodrigues, M. M. & Amarante-Mendes, G. P. *In vivo* assessment of specific cytotoxic T lymphocyte killing. *Methods* 61, 105–109 (2013).
- [0510] Weinmann, J. & Grimm, D. Next-generation AAV vectors for clinical use: an ever-accelerating race. *Virus Genes* 53, 707–713 (2017).

- [0511] Gabriel, N. *et al.* Bioengineering of AAV2 Capsid at Specific Serine, Threonine, or Lysine Residues Improves Its Transduction Efficiency *in vitro* and *in vivo*. *Hum. Gene Ther. Methods* 24, 80–93 (2013).
- [0512] Zhong, L. *et al.* Next generation of adeno-associated virus 2 vectors: Point mutations in tyrosines lead to high-efficiency transduction at lower doses. *Proc. Natl. Acad. Sci.* 105, 7827–7832 (2008).
- [0513] Shen, S. *et al.* Engraftment of a Galactose Receptor Footprint onto Adeno-associated Viral Capsids Improves Transduction Efficiency. *J. Biol. Chem.* 288, 28814–28823 (2013).
- [0514] Zinn, E. *et al.* In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Rep.* 12, 1056–1068 (2017).
- [0515] Deverman, B. E. *et al.* Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat. Biotechnol.* 34, 204–209 (2016).
- [0516] Smith, J. K. & Agbandje-McKenna, M. Creating an arsenal of Adeno-associated virus (AAV) gene delivery stealth vehicles. *PLOS Pathog.* 14, e1006929 (2018).
- [0517] Bartel, M., Schaffer, D. & Büning, H. Enhancing the clinical potential of AAV vectors by capsid engineering to evade pre-existing immunity. *Front. Microbiol.* 2, 204 (2011).
- [0518] Hui, D. J. *et al.* AAV capsid CD8<sup>+</sup> T-cell epitopes are highly conserved across AAV serotypes. *Mol. Ther. - Methods Clin. Dev.* 2, 15029 (2015).
- [0519] Gao, G., Vandenberghe, L. & Wilson, J. New Recombinant Serotypes of AAV Vectors. *Curr. Gene Ther.* 5, 285–297 (2005).
- [0520] Grieger, J. C., Choi, V. W. & Samulski, R. J. Production and characterization of adeno-associated viral vectors. *Nat. Protoc.* 1, 1412–1428 (2006).
- [0521] Mingozzi, F. & Büning, H. Adeno-Associated Viral Vectors at the Frontier between Tolerance and Immunity. *Front. Immunol.* 6, 120 (2015).
- [0522] Mingozzi, F. & High, K. A. Therapeutic *in vivo* gene transfer for genetic disease using AAV: progress and challenges. *Nat. Rev. Genet.* 12, 341–355 (2011).
- [0523] Wagner, D. L. *et al.* High prevalence of *S. pyogenes* Cas9-specific T cell sensitization within the adult human population - A balanced effector/regulatory T cell response. *bioRxiv* 295139 (2018). doi:10.1101/295139
- [0524] Gernoux, G., Wilson, J. M. & Mueller, C. Regulatory and Exhausted T Cell Responses to AAV Capsid. *Hum. Gene Ther.* 28, 338–349 (2017).



- [0525] Zhu, J., Huang, X. & Yang, Y. The TLR9-MyD88 pathway is critical for adaptive immune responses to adeno-associated virus gene therapy vectors in mice. *J. Clin. Invest.* 119, 2388–2398 (2009).
- [0526] Rogers, G. L. & Herzog, R. W. TLR9 and Dendritic Cells Are Required for CD8+ T Cell Responses to the AAV Capsid. *Blood* 124, (2014).
- [0527] Faust, S. M. *et al.* CpG-depleted adeno-associated virus vectors evade immune detection. *J. Clin. Invest.* 123, 2994–3001 (2013).
- [0528] Hirsch, M. & Wu, J. 319. Rationally Designed AAV Inverted Terminal Repeats Enhance Gene Targeting. *Mol. Ther.* 24, S129 (2016).
- [0529] Léger, A. *et al.* Adeno-Associated Viral Vector-Mediated Transgene Expression Is Independent of DNA Methylation in Primate Liver and Skeletal Muscle. *PLoS One* 6, e20881 (2011).
- [0530] Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–22 (2011).
- [0531] Shao, W. *et al.* Double-stranded RNA innate immune response activation from long-term adeno-associated virus vector transduction. *JCI Insight* 3, (2018).
- [0532] Peisley, A. *et al.* Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition. *Proc. Natl. Acad. Sci. U. S. A.* 108, 21010–5 (2011).
- [0533] Borel, F., Kay, M. A. & Mueller, C. Recombinant AAV as a platform for translating the therapeutic potential of RNA interference. *Mol. Ther.* 22, 692–701 (2014).
- [0534] Xie, J., Burt, D. R. & Gao, G. Adeno-associated virus-mediated microRNA delivery and therapeutics. *Semin. Liver Dis.* 35, 81–8 (2015).
- [0535] Li, J., Kim, S. G. & Blenis, J. Rapamycin: one drug, many effects. *Cell Metab.* 19, 373–9 (2014).
- [0536] Meliani, A. *et al.* Antigen-selective modulation of AAV immunogenicity with tolerogenic rapamycin nanoparticles enables successful vector re-administration. *Nat. Commun.* 9, 4098 (2018).
- [0537] Battaglia, M. *et al.* Rapamycin promotes expansion of functional CD4+CD25+FOXP3+ regulatory T cells of both healthy subjects and type 1 diabetic patients. *J. Immunol.* 177, 8338–47 (2006).
- [0538] Lee, E. J., Guenther, C. M. & Suh, J. Adeno-associated virus (AAV) vectors: Rational design strategies for capsid engineering. *Curr. Opin. Biomed. Eng.* 7, 58–63 (2018).

- [0539] Xie, Q. *et al.* The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10405–10 (2002).
- [0540] Walters, R. W. *et al.* Structure of Adeno-Associated Virus Serotype 5. *J. Virol.* 78, 3361–3371 (2004).
- [0541] Kotterman, M. A., Chalberg, T. W. & Schaffer, D. V. Viral Vectors for Gene Therapy: Translational and Clinical Outlook. *Annu. Rev. Biomed. Eng.* 17, 63–89 (2015).
- [0542] Grimm, D. *et al.* *In vitro* and *In vivo* Gene Therapy Vector Evolution via Multispecies Interbreeding and Retargeting of Adeno-Associated Viruses. *J. Virol.* 82, 5887–5911 (2008).
- [0543] Chew, W. L. *et al.* A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat. Methods* 13, 868–874 (2016).
- [0544] Katrekar, D. *et al.* *In vivo* RNA editing of point mutations via RNA-guided adenosine deaminases. *Nat. Methods* (2019). doi:10.1038/s41592-019-0323-0
- [0545] Kim, D. D. Y. *et al.* Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.* 14, 1719–25 (2004).
- [0546] Herbert, A. Z-DNA and Z-RNA in human disease. *Commun. Biol.* 2, 7 (2019).
- [0547] Koeris, M., Funke, L., Shrestha, J., Rich, A. & Maas, S. Modulation of ADAR1 editing activity by Z-RNA *in vitro*. *Nucleic Acids Res.* 33, 5362–70 (2005).
- [0548] Yang, J.-H. *et al.* Intracellular localization of differentially regulated RNA-specific adenosine deaminase isoforms in inflammation. *J. Biol. Chem.* 278, 45833–42 (2003).
- [0549] Abudayyeh, O. O. *et al.* A cytosine deaminase for programmable single-base RNA editing. *Science* 365, 382–386 (2019).
- [0550] Point Mutation - Definition, Types and Examples | Biology Dictionary.
- [0551] Lueck, J. D. *et al.* Engineered transfer RNAs for suppression of premature termination codons. *Nat. Commun.* 10, 822 (2019).
- [0552] Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488, 96–99 (2012).
- [0553] Sun, J. & Roy, S. The physical approximation of APP and BACE-1: A key event in alzheimer's disease pathogenesis. *Dev. Neurobiol.* 78, 340–347 (2018).
- [0554] Sun, J. *et al.* CRISPR/Cas9 editing of APP C-terminus attenuates  $\beta$ -cleavage and promotes  $\alpha$ -cleavage. *Nat. Commun.* 10, 53 (2019).
- [0555] Melcher, T. *et al.* A mammalian RNA editing enzyme. *Nature* 379, 460–464 (1996).

- [0556] Montiel-Gonzalez, M. F., Vallecillo-Viejo, I., Yudowski, G. A. & Rosenthal, J. J. C. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18285–90 (2013).
- [0557] Wettengel, J., Reautschnig, P., Geisler, S., Kahle, P. J. & Stafforst, T. Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res.* 45, 2797–2808 (2017).
- [0558] Fukuda, M. *et al.* Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. *Sci. Rep.* 7, 41478 (2017).
- [0559] Vallecillo-Viejo, I. C., Liscovitch-Brauer, N., Montiel-Gonzalez, M. F., Eisenberg, E. & Rosenthal, J. J. C. Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol.* 15, 104–114 (2018).
- [0560] Sinnamon, J. R. *et al.* Site-directed RNA repair of endogenous Mecp2 RNA in neurons. *Proc. Natl. Acad. Sci. U. S. A.* 114, E9395–E9402 (2017).
- [0561] Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027 (2017).
- [0562] Azad, M. T. A., Bhakta, S. & Tsukahara, T. Site-directed RNA editing by adenosine deaminase acting on RNA for correction of the genetic code in gene therapy. *Gene Ther.* 24, 779–786 (2017).
- [0563] Vogel, P. *et al.* Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods* 15, 535–538 (2018).
- [0564] Daniel, C., Widmark, A., Rigardt, D. & Öhman, M. Editing inducer elements increases A-to-I editing efficiency in the mammalian transcriptome. *Genome Biol.* 18, 195 (2017).
- [0565] Woolf, T.M., Chase J.M., & Stinchcomn D.T. Toward the therapeutic editing of mutated RNA sequences. *Proc. Natl. Acad. Sci. USA* 92, 8298-8302 (1995).
- [0566] Chung, H. *et al.* Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown. *Cell* 172, 811–824.e14 (2018).
- [0567] Desterro, J. M. P. *et al.* Dynamic association of RNA-editing enzymes with the nucleolus. *J. Cell Sci.* 116, 1805–1818 (2003).
- [0568] Mort, M., Ivanov, D., Cooper, D. N. & Chuzhanova, N. A. A meta-analysis of nonsense mutations causing human genetic disease. *Hum. Mutat.* 29, 1037–1047 (2008).
- [0569] Nelson, C. E. *et al.* *In vivo* genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* 351, 403–7 (2016).

- [0570] Tabebordbar, M. *et al.* *In vivo* gene editing in dystrophic mouse muscle and muscle stem cells. *Science* (80-. ). 351, 407–411 (2016).
- [0571] Long, C. *et al.* Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* (80-. ). 351, 400–403 (2016).
- [0572] Hodges, P. E. & Rosenberg, L. E. The *spf<sup>ash</sup>* mouse: a missense mutation in the ornithine transcarbamylase gene also causes aberrant mRNA splicing. *Proc. Natl. Acad. Sci. U. S. A.* 86, 4142–4146 (1989).
- [0573] Yang, Y. *et al.* A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* 34, 334–338 (2016).
- [0574] Chen, L. *et al.* Recoding RNA editing of *AZIN1* predisposes to hepatocellular carcinoma. *Nat. Med.* 19, 209–216 (2013).
- [0575] Mann CJ, Honeyman K, Cheng AJ, *et al.* Antisense-induced exon skipping and synthesis of dystrophin in the *mdx* mouse. *Proceedings of the National Academy of Sciences of the United States of America.* 2001;98(1):42-47.
- [0576] S. Konermann, P. Lotfy, N. J. Brideau, J. Oki, M. N. Shokhirev, P. D. Hsu, Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* 173, 665–676.e14 (2018).

**What is claimed is:**

1. A vector comprising a nucleic acid encoding an engineered guide RNA that comprises a targeting domain that hybridizes to a target RNA and forms a double stranded RNA comprising a mismatch formed by a target adenosine of the target RNA and a cytosine of the engineered guide RNA,  
wherein the double stranded RNA binds to an ADAR1 or an ADAR2 enzyme present in a cell, resulting in a targeted deamination of the target adenosine in the target RNA; and  
wherein the engineered guide RNA does not have a GluR2 ADAR recruiting domain.
2. The vector of claim 1, wherein the vector further comprises a U6 promoter.
3. The vector of claim 2, wherein the U6 promoter is a human U6 promoter or a mouse U6 promoter.
4. The vector of claim 1, wherein the targeting domain comprises a base that upon hybridization of the targeting domain to the target RNA forms a wobble base pair with a base of the target RNA.
5. The vector of claim 4, wherein the targeting domain comprises two bases that upon hybridization of the targeting domain to the target RNA each form two wobble bases pairs.
6. The vector of claim 1, wherein the targeted deamination of the adenosine of the target RNA results in an increase of an expression of a polypeptide encoded by the target RNA, relative to an otherwise comparable target RNA that is not deaminated.
7. The vector of claim 1, wherein the targeted deamination of the adenosine of the target RNA results in a decrease of an expression of a polypeptide encoded by the target RNA, relative to an otherwise comparable target RNA that is not deaminated.
8. The vector of any one of claims 1-7, wherein the targeting domain is from 50 to 200 base pairs in length.

9. The vector of claim 8, wherein the targeting domain is from 100 to 200 base pairs in length.
10. The vector of any one of claims 1-9, wherein the vector comprises a viral vector.
11. The vector of claim 10, wherein the viral vector comprises an adeno-associated virus (AAV) vector.
12. The vector of any one of claims 1-11, wherein the nucleic acid further encodes for a second engineered guide RNA that comprises a second targeting domain sequence complementary to the target RNA.
13. A pharmaceutical composition comprising the vector of any one of claims 1-12, and a pharmaceutically acceptable excipient, diluent, or carrier.
14. The pharmaceutical composition of claim 13, wherein the pharmaceutical composition is in unit dose form.
15. A method of treating a disease or condition in a subject in need thereof, the method comprising administering to the subject the pharmaceutical composition of any one of claims 13 or 14 in an amount sufficient to treat the disease or condition.
16. The method of claim 15, wherein the disease or condition is selected from the group consisting of a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, and any combination thereof.
17. Use of the vector of any one of claims 1-12 in the manufacture of a medicament for the treatment of a disease or condition selected from the group consisting of a neurodegenerative disorder, a muscular disorder, a metabolic disorder, an ocular disorder, and any combination thereof.
18. The method of claim 16, or the use according to claim 17, wherein the disease or condition is Alzheimer's disease.

19. The method of any one of claims 15, 16 or 18, or the use of claim 17 or 18, wherein the subject is a human.

**The Regents of the University of California Office of Technology Transfer  
The Regents of the University of California**

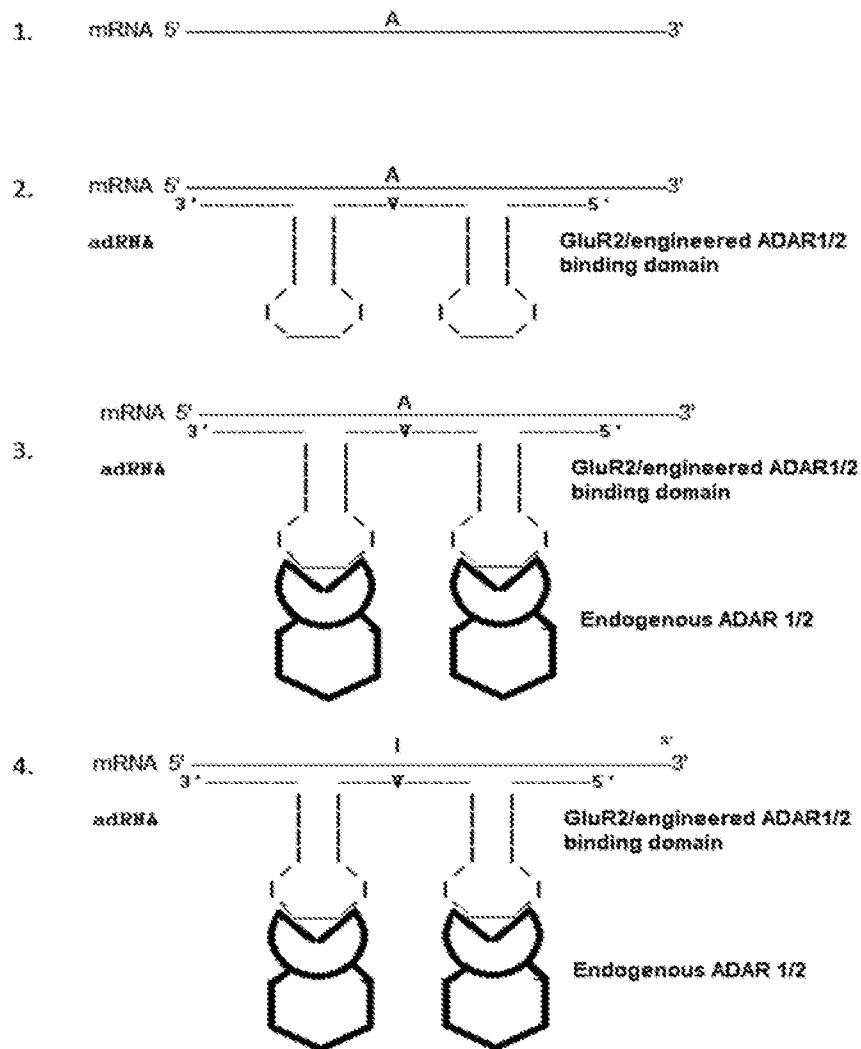
**Patent Attorneys for the Applicant/Nominated Person**

**SPRUSON & FERGUSON**

1 / 50

## Endogenous recruitment of ADARs

Steps:



## Cas9/Cpf1 - ADAR1/2 deaminase domain fusions

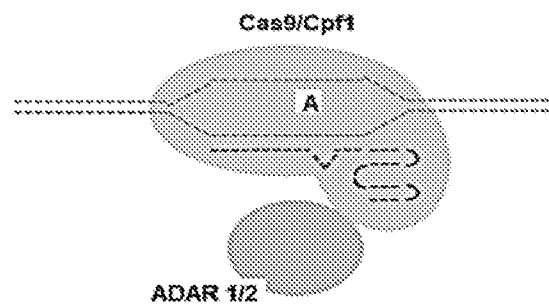


FIG. 1

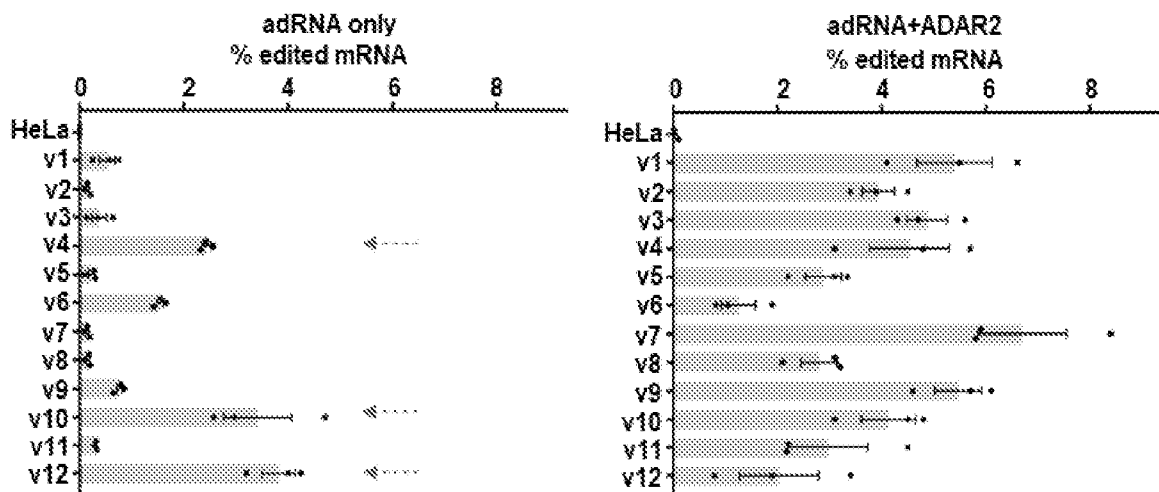


**2 / 50**

```
v3 : GTGGAAGAGGAGAAACAATATGCTAAATGTTGTTCTCGTCTCCAC NNNNNNCCCCCCCCCCCCCCCCCC
v4 : GGGTGAAGAGGAGAAACAATATGCTAAATGTTGTTCTCGTCTCCACCT NNNCCCCCCCCCCCCCCCCCCCCC
v5 : GGTGAAGAGGAGAAACAATATGCTAAATGTTGTTCTCGTCTCCACC NNNNNNCCCCCCCCCCCCCCCCCC
v6 : GGTGAAGAGGAGAAACAATATGCTAAATGTTGTTCTCGTCTCCACC NNNNNNNCCCCCCCCCCCCCCCCCC
v7 : GTGGAAGAGGAGAAACAATAGGCTAAACGTTGTTCTCGTCTCCCAC NNNNNCCCCCCCCCCCCCCCCCC
v8 : GGGTGAAGAGGAGAAACAATAGGCTAAACGTTGTTCTCGTCTCCACCT NNNCCCCCCCCCCCCCCCCCCCCC
v9 : GGTGAAGAGGAGAAACAATAGGCTAAACGTTGTTCTCGTCTCCACC NNNNNCCCCCCCCCCCCCCCCCC
v10: GGTGAAGAGGAGAAACAATAGGCTAAACGTTGTTCTCGTCTCCACC NNNNNNNCCCCCCCCCCCCCCCCCC
```

**FIG. 2**

***In vitro*: screening adRNA designs for RNA editing via recruitment of endogenous ADARs**



### *In vivo*: RNA editing via recruitment of endogenous ADARs

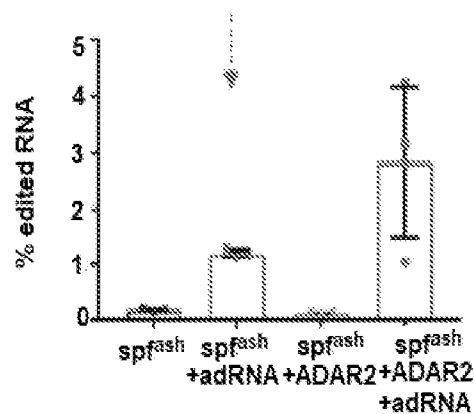


**FIG. 3**

3 / 50

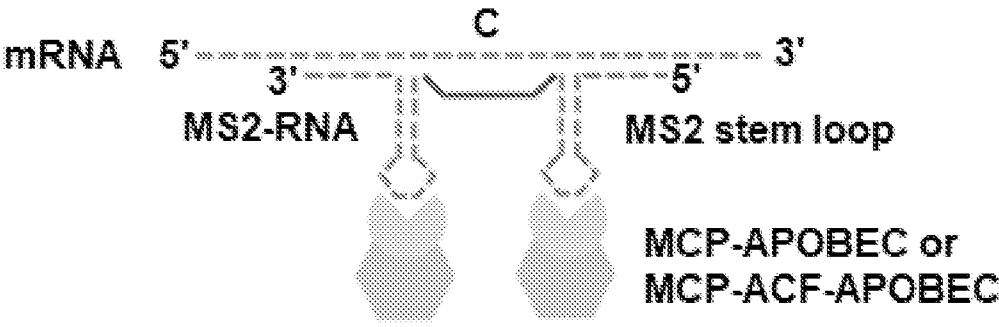


FIG. 4

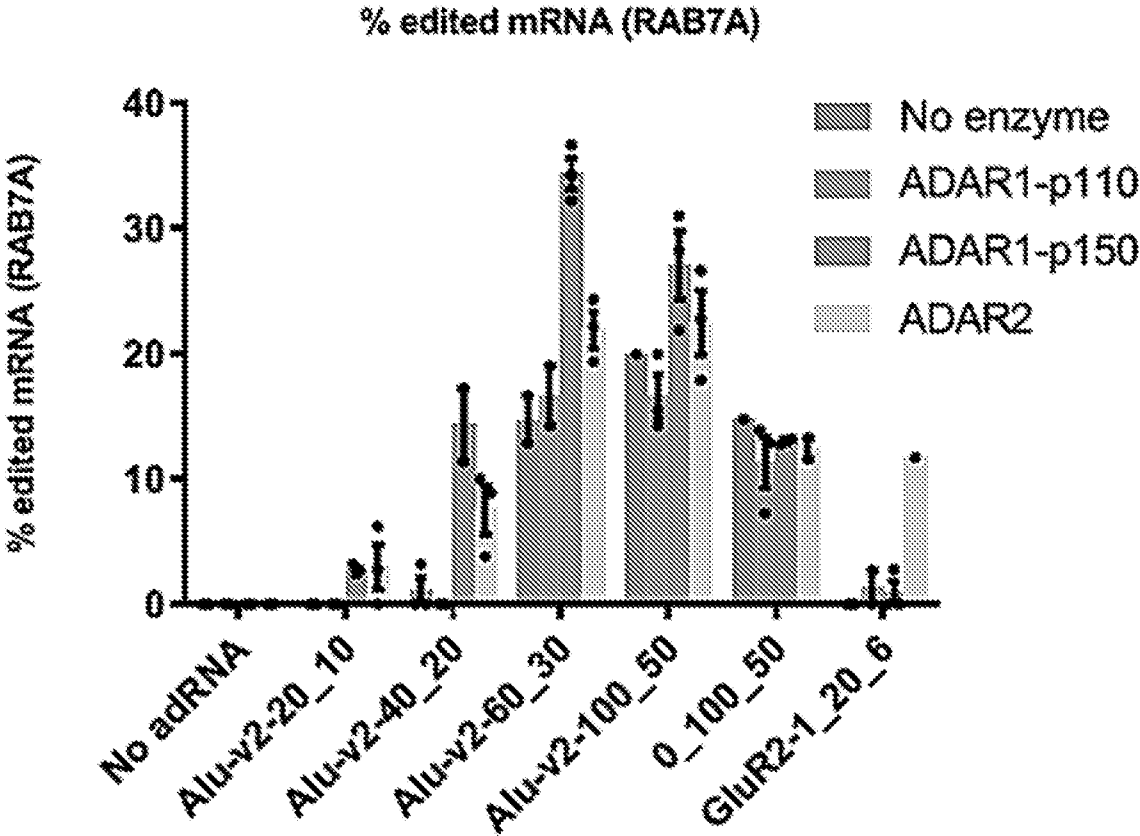


FIG. 5

**4 / 50**

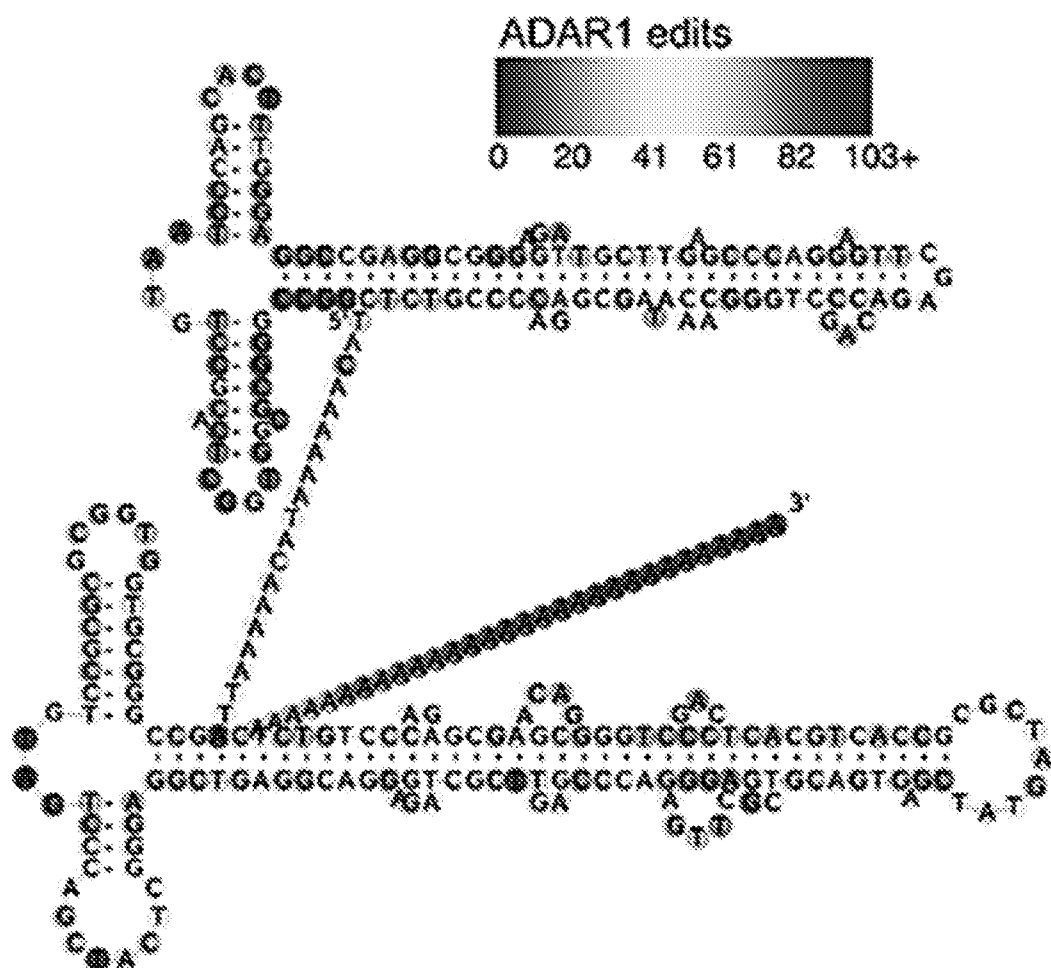
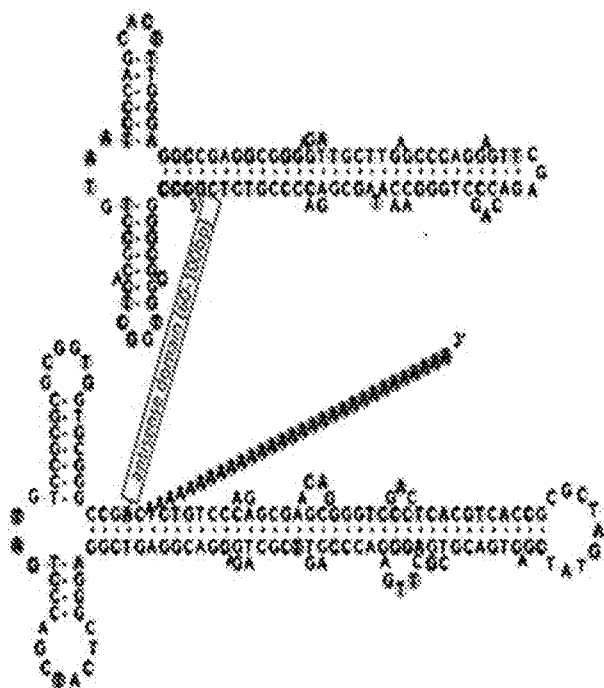


**FIG. 6**

**5 / 50**



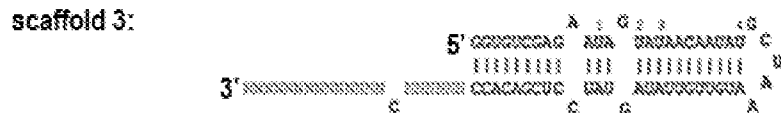
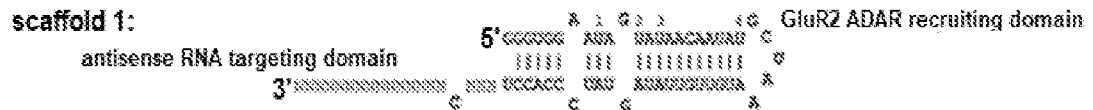
SEQ ID NO: 1

### Example 5:

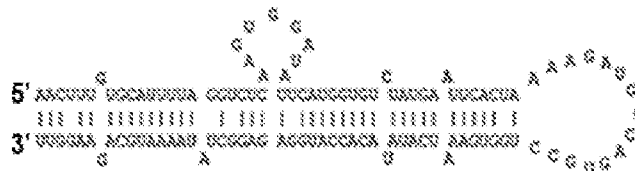
[illegible]

**FIG. 7**

6 / 50



editing inducer element (EiE):



| Var | TARGET: 3' ~ TTCTTTTC AAAATGTTTCCTCGCCACAGACACTCTGAAAGTAAGT ~ 5' |
|-----|------------------------------------------------------------------|
| 1.1 | 1:U-A 2:U-A 3:U-A 4:U-A AACCGAGCGGTGTCTGTGA                      |
| 1.2 | 1:G-C 2:G-C 3:G-C 4:U-A AACCGAGCGGTGTCTGTGA                      |
| 1.3 | 1:G-C 2:G-C 3:G-C 4:G-C AACCGAGCGGTGTCTGTGA                      |
| 2.1 | 1:U-A 2:U-A 3:U-A 4:U-A 5:U-A 6:G-C ACAAAACCGAGCGGTGTCTGTG       |
| 2.2 | 1:G-C 2:G-C 3:G-C 4:U-A 5:U-A 6:G-C ACAAAACCGAGCGGTGTCTGTG       |
| 2.3 | 1:G-C 2:G-C 3:G-C 4:G-C 5:U-A 6:G-C ACAAAACCGAGCGGTGTCTGTG       |
| 2.4 | 1:G-C 2:G-C 3:G-C 4:U-A 5:G-C 6:U-A ACAAAACCGAGCGGTGTCTGTG       |
| 2.5 | 1:G-C 2:G-C 3:G-C 4:U-A 5:G-C 6:U-A TACAAACCGAGCGGTGTCTGT        |
| 2.6 | 1:G-C 2:G-C 3:G-C 4:G-C 5:G-C 6:U-A ACAAAACCGAGCGGTGTCTGTG       |
| 2.7 | 1:G-C 2:G-C 3:G-C 4:G-C 5:G-C 6:U-A TACAAACCGAGCGGTGTCTGT        |
| 3.1 | 1:U-A 2:U-A 3:U-A 4:U-A TACAAACCGAGCGGTGTCT                      |
| 3.2 | 1:G-C 2:G-C 3:G-C 4:U-A TACAAACCGAGCGGTGTCT                      |
| 3.2 | 1:G-C 2:G-C 3:G-C 4:G-C TACAAACCGAGCGGTGTCT                      |
| 4.1 | 1:U-A 2:U-A 3:U-A 4:U-A 5:U-A 6:G-C CAAACCGAGCGGTGTCTGTG         |
| 4.2 | 1:U-A 2:U-A 3:U-A 4:U-A 5:U-A 6:G-C TACAAACCGAGCGGTGTCTGT        |
| 4.3 | 1:U-A 2:U-A 3:U-A 4:U-A 5:U-A 6:G-C TTTACAAACCGAGCGGTGTCT        |
| 4.4 | 1:U-A 2:U-A 3:U-A 4:U-A 5:U-A 6:G-C ACAAAACCGAGCGGTGTCTGTG       |
| 4.5 | 1:U-A 2:U-A 3:U-A 4:U-A 5:U-A 6:G-C ACAAAACCGAGCGGTGTCTGTG       |
| 4.6 | 1:U-A 2:U-A 3:U-A 4:U-A 5:U-A 6:G-C ACAAAACCGAGCGGTGTCTGTG       |

FIG. 8

7 / 50

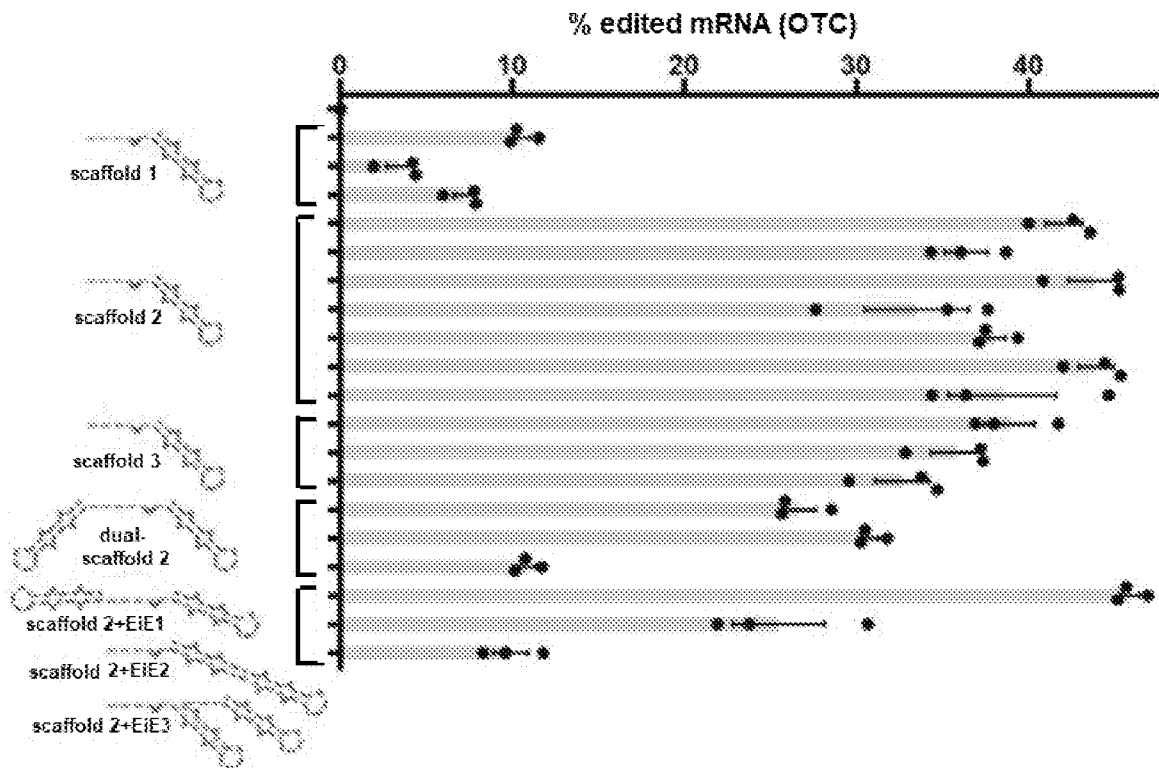


FIG. 8 (continued)

8 / 50

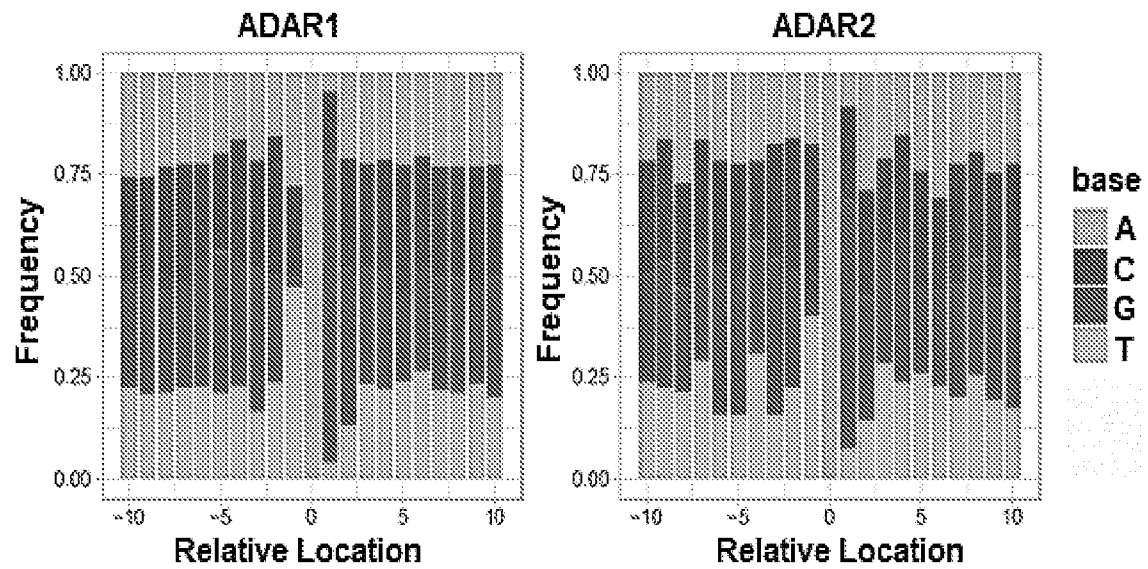


FIG. 9

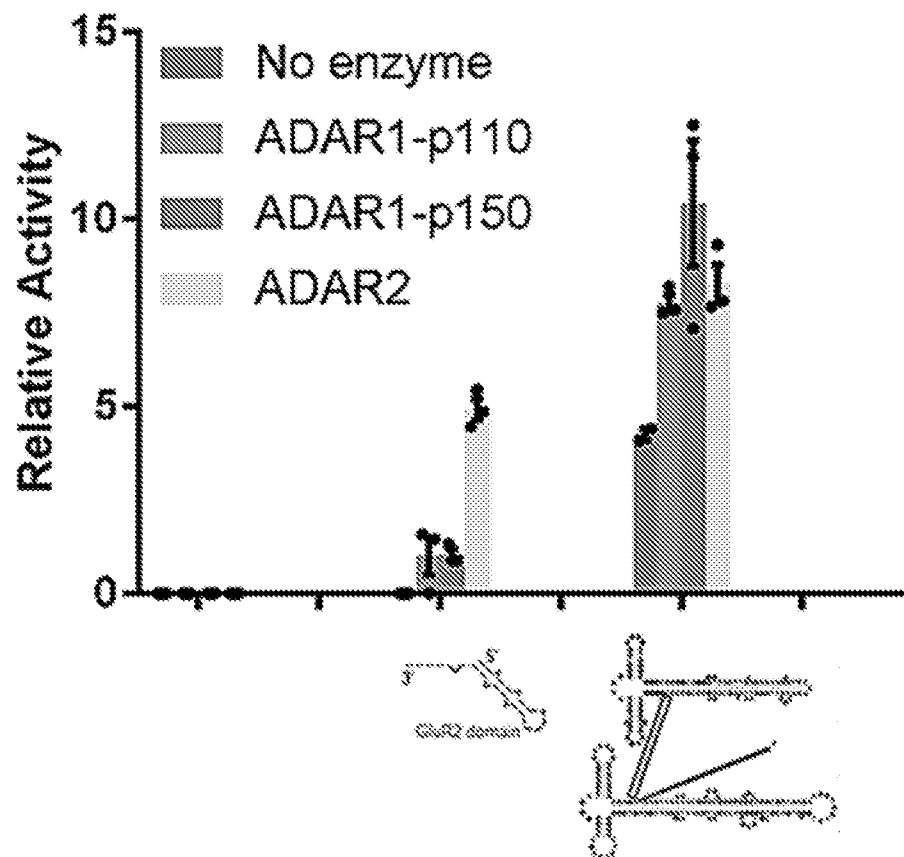


FIG. 10

9 / 50

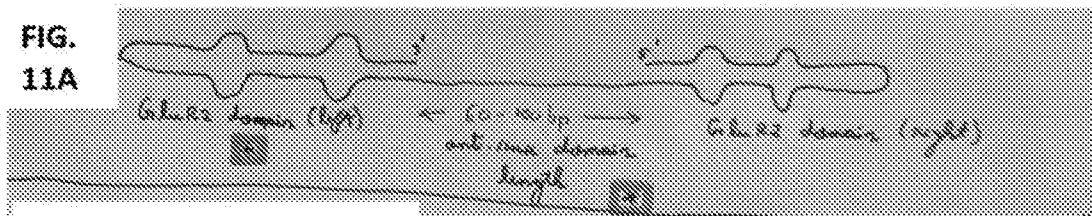
FIG.  
11A

FIG. 11B

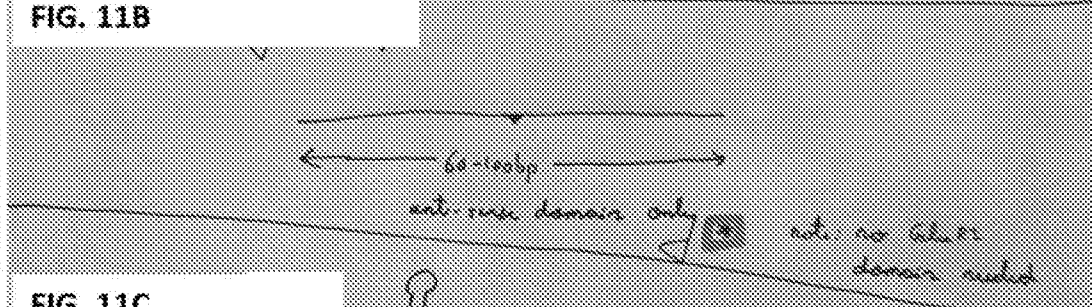
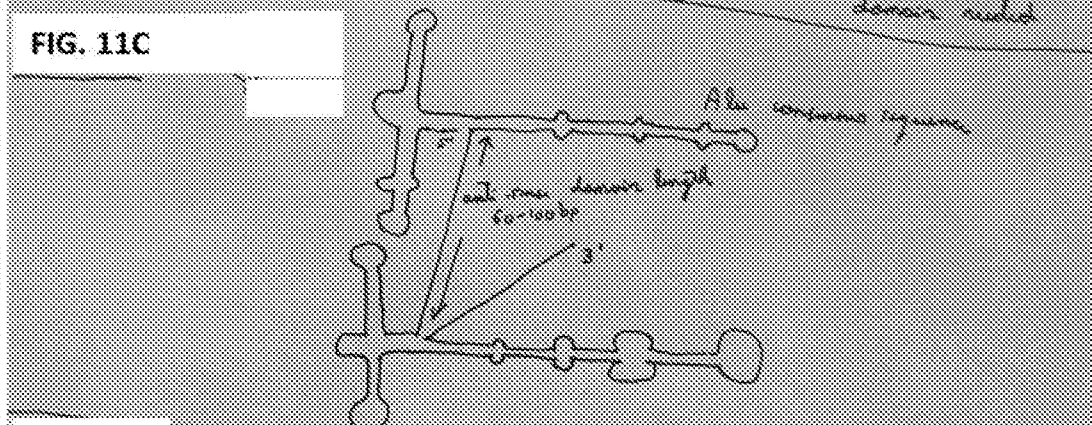
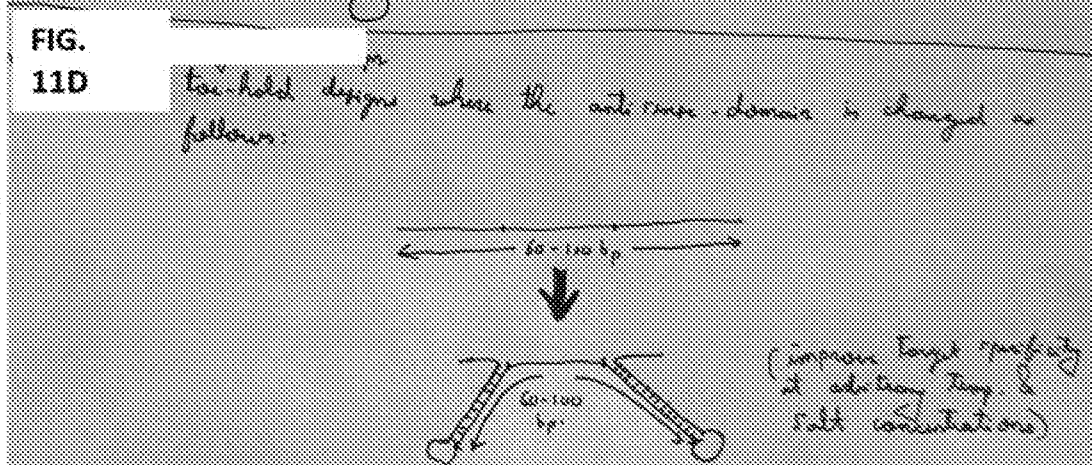


FIG. 11C

FIG.  
11D



RNAfix shows superior editing efficiency, and negligible off-target editing, when compared to competing platform.

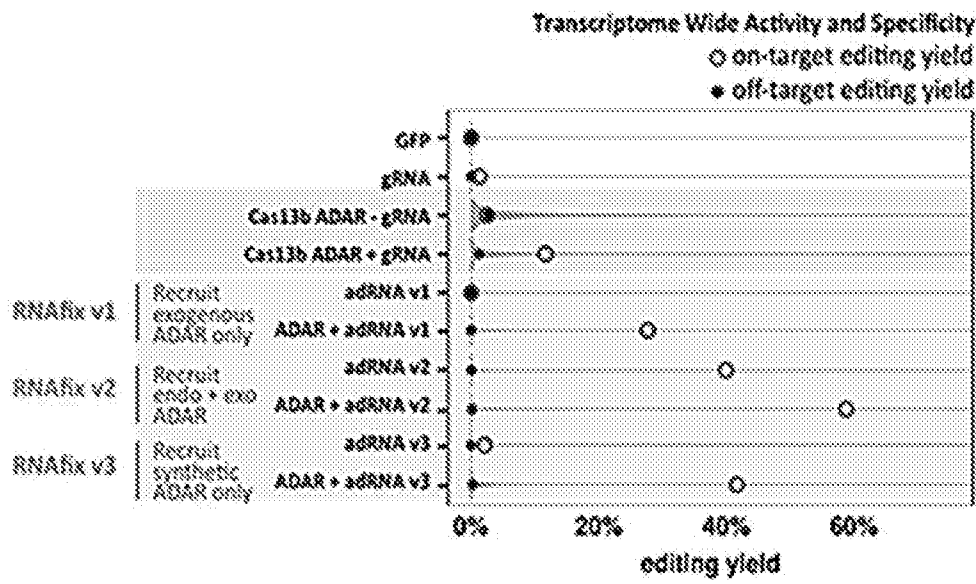


FIG. 12

11 / 50

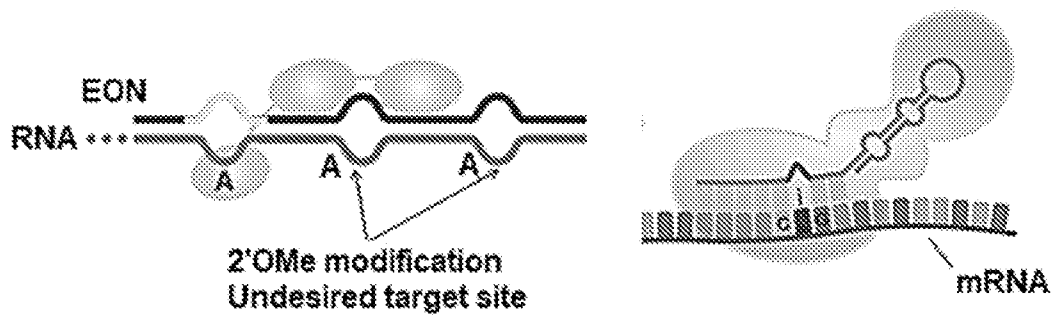


FIG. 13

General design of adRNA\_0\_100\_50:

```
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

General design of adRNA\_2\_100\_50:

[illegible]

FIG. 14

**12 / 50**

### Representative adRNA structures

## Parameters

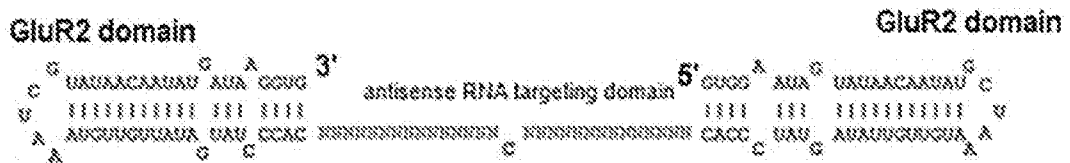
(d.l.m)

d: number of GluR2 domains

l: length of antisense domain

m: position of mismatch

adRNA (2,30,15)



adRNA (0,30,15)

antisense RNA targeting domain



acRNA (1,20,6)

**GluR2 domain**

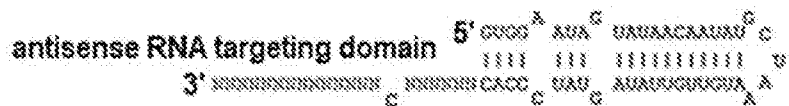


FIG. 15

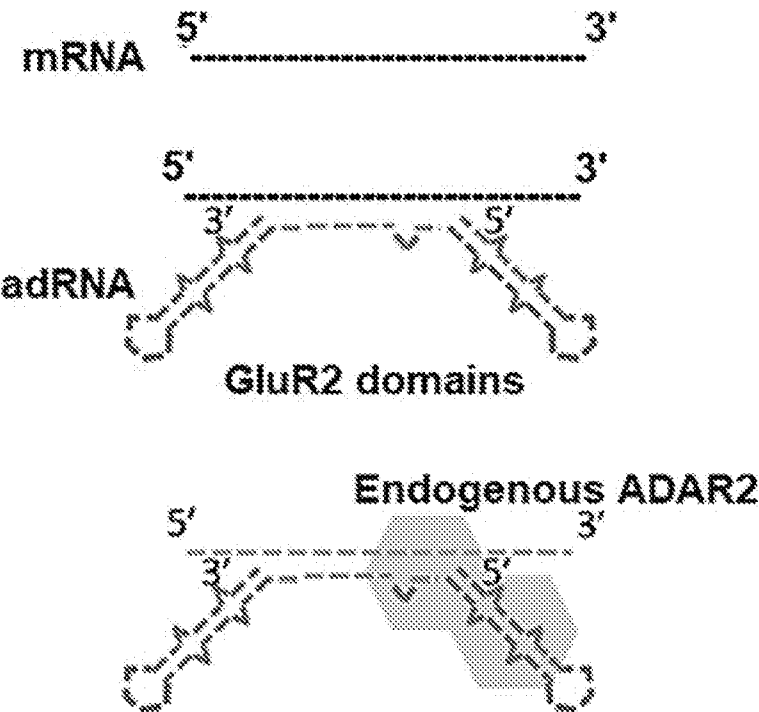


FIG. 16

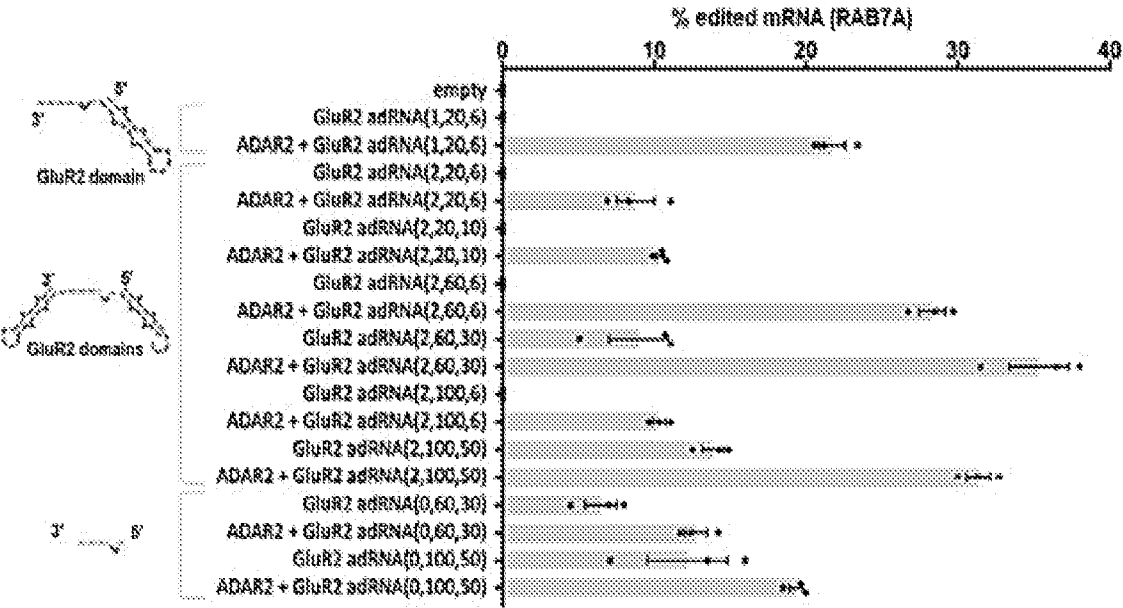


FIG. 17

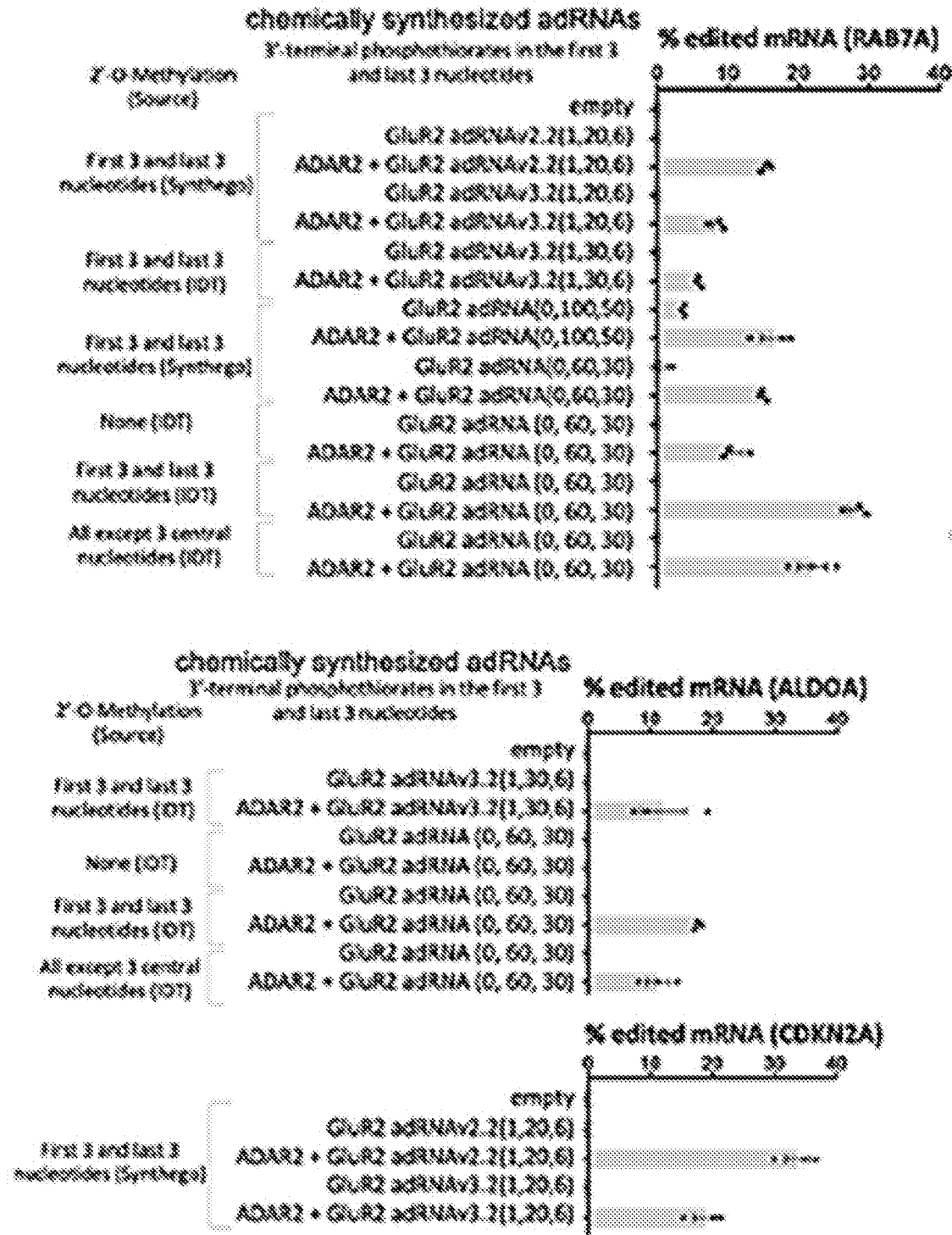


FIG. 18

15 / 50

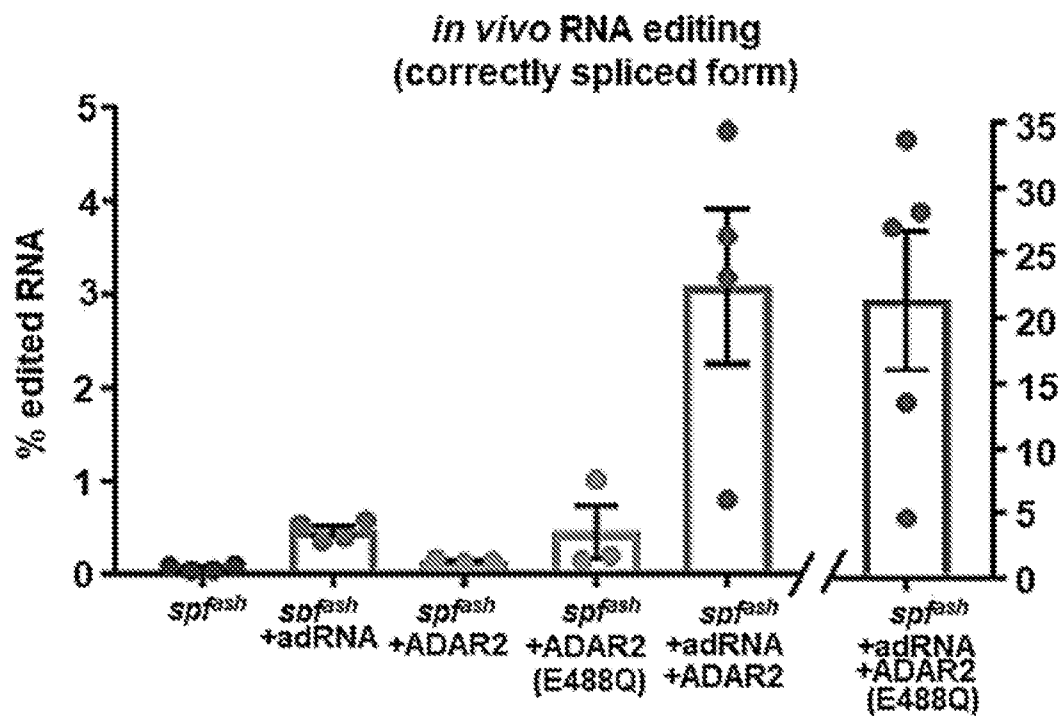


FIG. 19

16 / 50

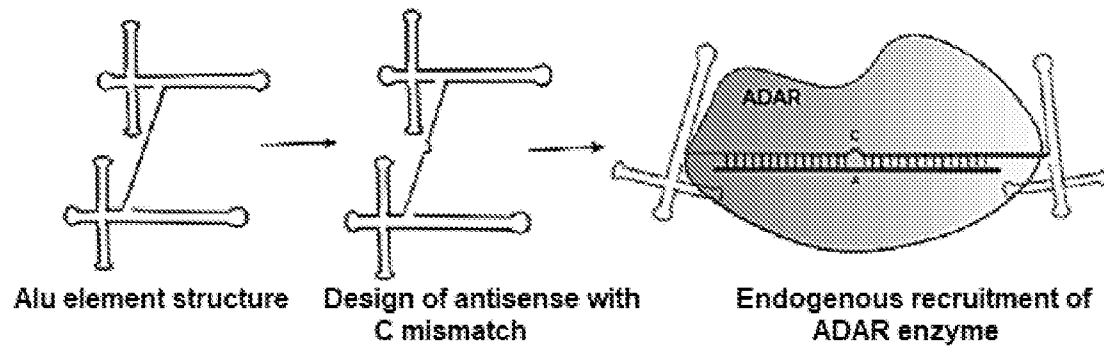


FIG. 20

Alu v2 Guide Sequences:

Alu Element

Antisense Guide

**Target C Mismatch**

Alu-v2-20-10:

GGCCGGGCGCGGTGGCTCACGCCTGTAATCCACGACACTTTTGGGAGGCGCGAGGGAGATTGC  
 TTGAGCCCAGGAGTTTCGAGACCCAGCCTGGGCAACATAGCGAGACCCCGTCTCATACTGCCGCCA  
 GCTGGATTAGCCGCGCGTGGTGGCGCGCCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAGG  
 AGGATCGCTTGAGCCCAGGAGTTTCGAGGCTGCAGTGAGCTATGATCGCGCCACTGCACTCCAGC  
 CTGGGCGACAGAGCGAGACCCTGTCTC

Alu-v2-40-20:

GGCCGGGCGCGGTGGCTCACGCCTGTAATCCACGACACTTTTGGGAGGCGCGAGGGAGATTGC  
 TTGAGCCCAGGAGTTTCGAGACCCAGCCTGGGCAACATAGCGAGACCCCGTCTCACTGTACAGAAT  
 ACTGCCGCCAGCTGGATTTCCTCAATTCTAGCCGGCGTGGTGGCGCGCCCTGTAGTCCAGCT  
 ACTCGGAGGCTGAGGCAGGAGGATCGCTTGAGCCCAGGAGTTTCGAGGCTGCAGTGAGCTATGA  
 TCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACCCTGTCTC

Alu-v2-60-30:

GGCCGGGCGCGGTGGCTCACGCCTGTAATCCACGACACTTTTGGGAGGCGCGAGGGAGATTGC  
 TTGAGCCCAGGAGTTTCGAGACCCAGCCTGGGCAACATAGCGAGACCCCGTCTCTCTTGTGTCTAC  
 TGTACAGAATACTGCCGCCAGCTGGATTTCCTCAATTCTGAGTAACACAGCCGGCGTGGTGGCG  
 CGCGCCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAGGAGGATCGCTTGAGCCCAGGAGTTTCG  
 AGGCTGCAGTGAGCTATGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACCCTGTCTC  
 TC

Alu-v2-100-50:

GGCCGGGCGCGGTGGCTCACGCCTGTAATCCACGACACTTTTGGGAGGCGCGAGGGAGATTGCCTTGGC  
 CCAGGAGTTTCGAGACCCAGCCTGGGCAACATAGCGAGACCCCGTCTCTGATAAAAGGCGTACATAATTCTT  
 GTGTCTACTGTACAGAATACTGCCGCCAGCTGGATTTCCTCAATTCTGAGTAACACTCTCCAATCCAAACA  
 GGGTTCCAGCCGGGCGTGGTGGCGCGCCCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAGGAGGATCGC  
 TTGAGCCCAGGAGTTTCGAGGCTGCAGTGAGCTATGATCGCGCCACTGCACTCCAGCCTGGGCGACAGAGC  
 GAGACCCTGTCTC

FIG. 21

17 / 50

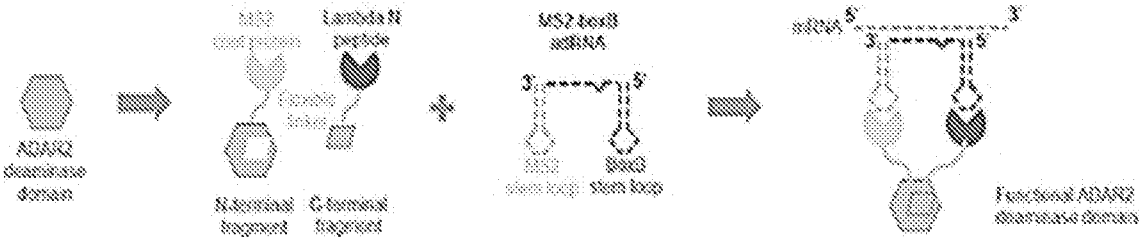


FIG. 22

316

QLHLPOVLADAVSRVLGKFGDLT FSSPHARRKVLGVVHTG Y AKVISVSTGKTCINGEYMS  
GLALNDCHAEITSRRLRFLYTQLELYLNN DC SIFQKS GGFRLLK VOFHLYISTSPCGDAR  
IFSPHEPTL RH F RGOLETKIES ST VRNASI SDGVLOE LLTM SDKIARWNV  
VGIOGSLLSIFVEPIYFSSIIIGSLYHGDHLSKAMYQRISNIEDLPPLYTLNKULLSGISNAEARQPGKA  
PNFSVNWTVGUSAIEVINATTGKDELGRASRLCKNALYCRWVRVHGKVPSHLLRSKITKPNVYHESKLAA  
KEYQAARLFTAFIKAGLGAWVERPTEQDOFSLTP

701

FIG. 23

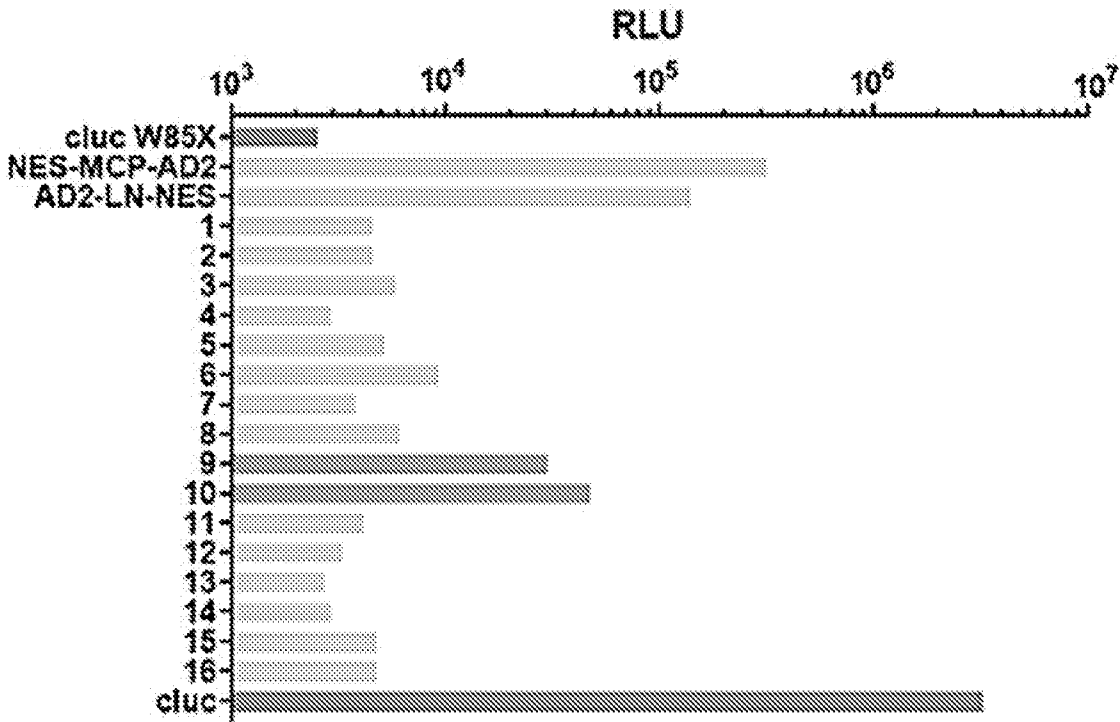


FIG. 24



18 / 50

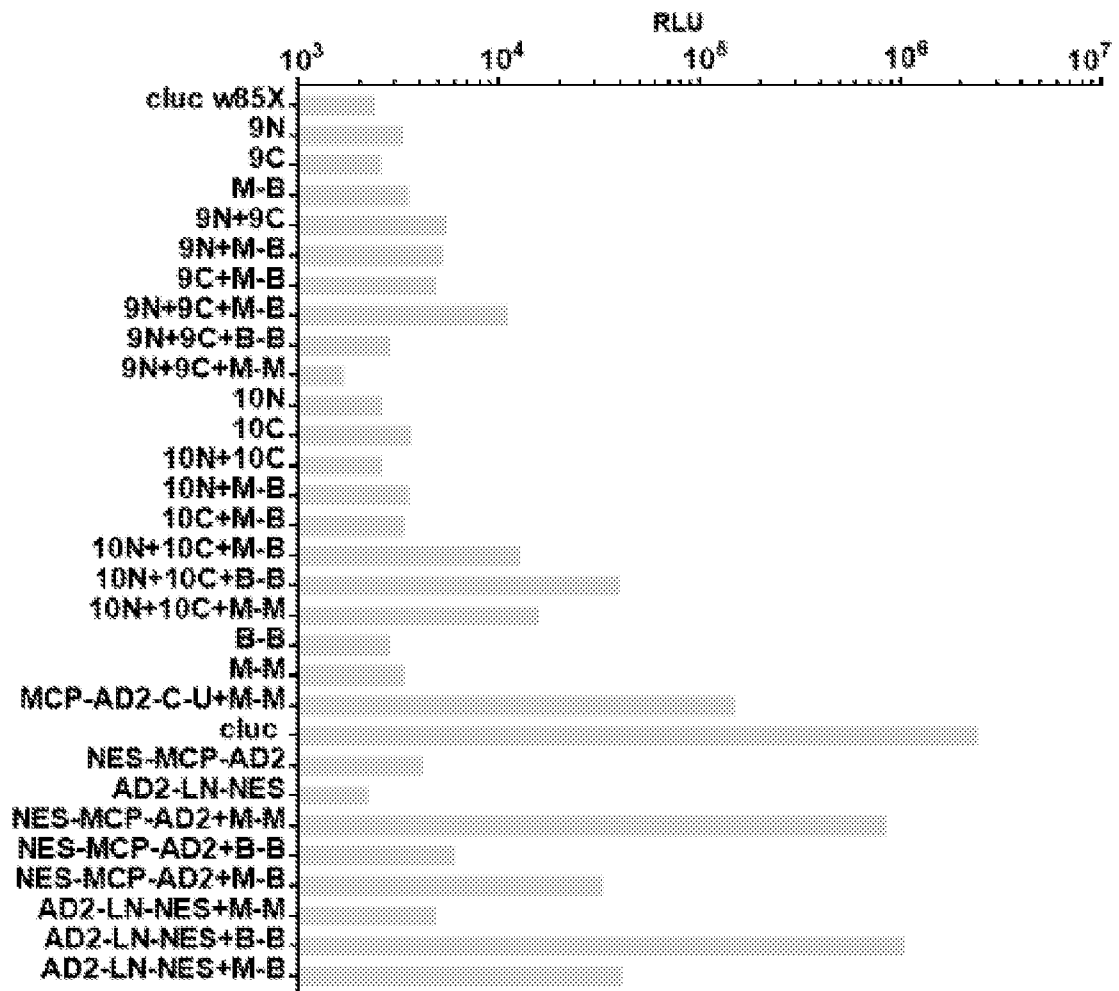


FIG. 25

19 / 50

**Cluc W85X reporter:**

ATGAAGACCTTAATTCTTGCCGTTGCATTAGTCTACTGCGCCACTGTTTCATTGCCAGGACTGTCTTACG  
AACCTGATCCACCAAAACACAGTTCCAACTTCCTGTGAAGCTAAAGAAGGAGAATGTATTGATAGCAGCTG  
TGGCACCTGCACGAGAGACATACTATCAGATGGACTGTGTGAAAATAAACAGGAAAAACATGTTGCCGA  
ATGTGTCACTATGTAAATTGAATGCAGAGTAGAGGCCCGCAGGATAGTPTTAGAACATTCTATGGAAAGAGAT  
TCCAGTTCCAGGAACCTGGTACATACGTGTTGGGTCAAGGAACCAAGGGCCGGCGACTGGAAGGTGTCCAT  
CACCCTGGAGAACCTGGATGGAACCAAGGGGGCTGTGTGTGACCAAGACAAGACTGGAAGTGGCTGGAGAC  
ATCATTGACATCGCTCAAGCTACTGAGAATCCCATCACTGTAAACGGTGGAGCTGACCCCTATCATCGCCA  
ACCCGTACACCATCGGCGAGGTCAACATCGCTGTTGTTGAGATGCCAGGCTTCAACATCACCCTCATTGA  
GTTCTTCAAACCTGATCGGTGATCGACATCCTCGGAGGAAGATCTGTAAAGAAATCGCCCCAGACACAGCAAAC  
AAAGGAATGATCTCTGGCCTCTGTGGAGATCTTAAAATGATGGAAGATACAGACTTCACTTCAGATCCAG  
AACAACTCGCTATTCAGCCTAAGATCAACCAGGAGTTTGACGGTTGTCCACTCTATGGAAATCCTGATGA  
CGTTGCATACTGCAAAGGTCTTCTGGAGCCGTACAAGGACAGCTGCCGCAACCCCATCAACTTCTACTAC  
TACACCATCTCCTGCGCCTTCGCCCCGCTGTATGGGTGGAGACGAGCGAGCCTCACACGTGCTGCTTGA  
ACAGGGAGACGTGCGCTGCTCCCGAACTAGAGGAACCTGCGTTTTGTCTGGACATACTTTCTACGATAC  
ATTTGACAAAGCAAGATACCAATTCAGGGTCCCTGCAAGGAGATTCTTATGGCCGCCGACTGTTTCTGG  
AACACTTGGGATGTGAAGGTTTCACACAGGAATGTTGACTCTTACACTGAAGTAGAGAAAGTACGAATCA  
GGAAACAATCGACTGTAGTAGAATCATTTGTTGATGGAAAACAGATTCTGGTTGGAGGAGAAGCCGTGTC  
CGTCCCGTACAGCTCTCAGAACACTTCCATCTACTGGCAAGATGGTGACATACTGACTACAGCCATCCTA  
CCTGAAGCTCTGGTGGTCAAGTTCAACTTCAAGCAACTGCTCGTCTGATATATTAGAGATCCATTGATG  
GTAAGACTTGGCGTATTTGCGGTAACTACAACCAGGATTTGAGTGATGATTCTTTTGTGCTGAAGGAGC  
CTGTGATCTGACCCCCAACCCACCGGGATGCACCGAAGAACAGAAACCTGAAGCTGAACGACTCTGCAAT  
AGTCTCTTCGCCGGTCAAAGTGATCTTGATCAGAAATGTAACTGTGCCCACAAGCCTGACCGTGTGGAAC  
GATGCATGTACGAGTATTCCTGAGGGGACAACAGGGTTTCTGTGACCACGCATGGGAGTTCAAGAAAGA  
ATGCTACATAAAGCATGGAGACACCCTAGAAGTACCAGATGAATGCAAATAA

**MS2/BoxB-adRNA sequences:**

|                     |                                                                     |
|---------------------|---------------------------------------------------------------------|
| BoxB_Luc-TAG_MS2_F  | ggccctgaaaaagggcctGTCTTAAACCACTCCTGCGGCCCTCaACATGAGGATCACCCTATGc    |
| MS2_Luc-TAG_BoxB_F  | aACATGAGGATCACCCTATGCTcTAAACCACTCCTGCGGCCCTCTACTCTggccctgaaaaagggcc |
| MS2_Luc-TAG_MS2_F   | aACATGAGGATCACCCTATGCTcTAAACCATCCTGCGGCCCTCaACATGAGGATCACCCTATGCTc  |
| BoxB_Luc-TAG_BoxB_F | ggccctgaaaaagggcctGTCTTAAACCATCCTGCGGCCCTCTACTCTggccctgaaaaagggcc   |

**NES-FLAG-MCP-N-terminal-AD2-9:**

MLPPLERLTLGSDYKDDDDKSGSGSGSMASNFTQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYK  
VTCVSRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELVKAMQGLLKDGNPIPSAIAANS  
GIYGGSGSGAGSGSPAGGGAPGSGGGSQLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVLAVVMT  
TGTDVKDAKVISVSTGPKCINGEYMSDRGLALNDCHAEIISRRSLRLFLYTQLELYLNNKDDQKRSIFQK  
SERGGFRLKENVQFHLIYSTSPCGDARIFSPHEPILE

**FIG. 26**

20 / 50

**C-terminal-AD2-9-LN-HA-NES:**

MEPADRRPNRKARGQLRTKIESGEGTIPVRSNASIQTWDCVLOGERLLTMSCSDKIAEWNVVGIOGSLLS  
 IFVEPIYFSSIIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPILLSGISNAEARQPGKAPNFSVNWTV  
 GDSAEIVINATTGKDELGRASRLCKHALYCRWMEVRGKVP SHLLRSKITKPNVYHESKLAKEYQAQAKAR  
 LFTAFIKAGLGAWVEKPTEQDQFSLTPGGSGSGAGSGSPAGGGAPGSGGGSNARTRRRERRAEKQAQWKA  
 ANGGGGSGGGSGGGGSNARTRRRERRAEKQAQWKAANGGGGGSGGGSGGGGSNARTRRRERRAEKQAQW  
 KAANGGGGGSGGGSGGGGSNARTRRRERRAEKQAQWKAANGSYPDVDPDYAGSLPFLERLTL

**NES-FLAG-MCP-N-terminal-AD2-10:**

MLPFLERLTLGSDYKDDDDKSGSGSGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYK  
 VTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMEITPIFATNSDCELIKAMQGLLDGNPIPSAIAANS  
 GIYGGSGSGAGSGSPAGGGAPGSGGGSQLHLPQVLADAVSRLVLGKFGDLTDNFSSPHARRKVLAVVMT  
 TGTDVKDAKVISVSTGTCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRSIFQK  
 SERGGFRLKENVQFHLYISTSPCGDARIFSPHEPILEEPA

**C-terminal-AD2-10-LN-HA-NES:**

MDRHFNPKARGQLRTKIESGEGTIPVRSNASIQTWDCVLOGERLLTMSCSDKIAEWNVVGIOGSLLSIFV  
 EPIYFSSIIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPILLSGISNAEARQPGKAPNFSVNWTVGDS  
 AIEVINATTGKDELGRASRLCKHALYCRWMEVRGKVP SHLLRSKITKPNVYHESKLAKEYQAQAKARLFT  
 AFIKAGLGAWVEKPTEQDQFSLTPGGSGSGAGSGSPAGGGAPGSGGGSNARTRRRERRAEKQAQWKAANG  
 GGGSGGGSGGGGSNARTRRRERRAEKQAQWKAANGGGGGSGGGSGGGGSNARTRRRERRAEKQAQWKA  
 NGGGSGGGSGGGGSNARTRRRERRAEKQAQWKAANGSYPDVDPDYAGSLPFLERLTL

**NES-FLAG-MCP-N-terminal-AD2-C-U-variant-9:**

MLPFLERLTLGSDYKDDDDKSGSGSGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYK  
 VTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMEITPIFATNSDCELIKAMQGLLDGNPIPSAIAANS  
 GIYGGSGSGAGSGSPAGGGAPGSGGGSQLHLPQVLADAVSRLVIGKFGDLTDNFSSPHARRIGLAGVVM  
 TGTDVKDAKVICVSTGSKCINGEYLSDRGLALNDCHAEIVSRRSLLRFLYTQLELYLNNEDDQKRSIFQK  
 SERGGFRLKENIQFHLYISTSPCGDARIFSPHEAILE

**C-terminal-AD2-C-U-variant-9-LN-HA-NES:**

MEPADRRPNRKARGQLRTKIEAGQGTIPVRNNASIQTWDCVLOGERLLTMSCSDKIAEWNVVGIOGSLLS  
 IFVEPIYFSSIIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPILLTGISNAEARQPGKAPIFSVNWTV  
 GDSAEIVINATTGKDELGRASRLCKHALYCRWMEVRGKVP SHLLRSKITKPNVYHETKLAKEYQAQAKAR  
 LFTAFIKAGLGAWVEKPTEQDQFSLTPGGSGSGAGSGSPAGGGAPGSGGGSNARTRRRERRAEKQAQWKA  
 ANGGGGSGGGSGGGGSNARTRRRERRAEKQAQWKAANGGGGGSGGGSGGGGSNARTRRRERRAEKQAQW  
 KAANGGGGGSGGGSGGGGSNARTRRRERRAEKQAQWKAANGSYPDVDPDYAGSLPFLERLTL

**NES-FLAG-MCP-N-terminal-AD2-C-U-variant-10:**

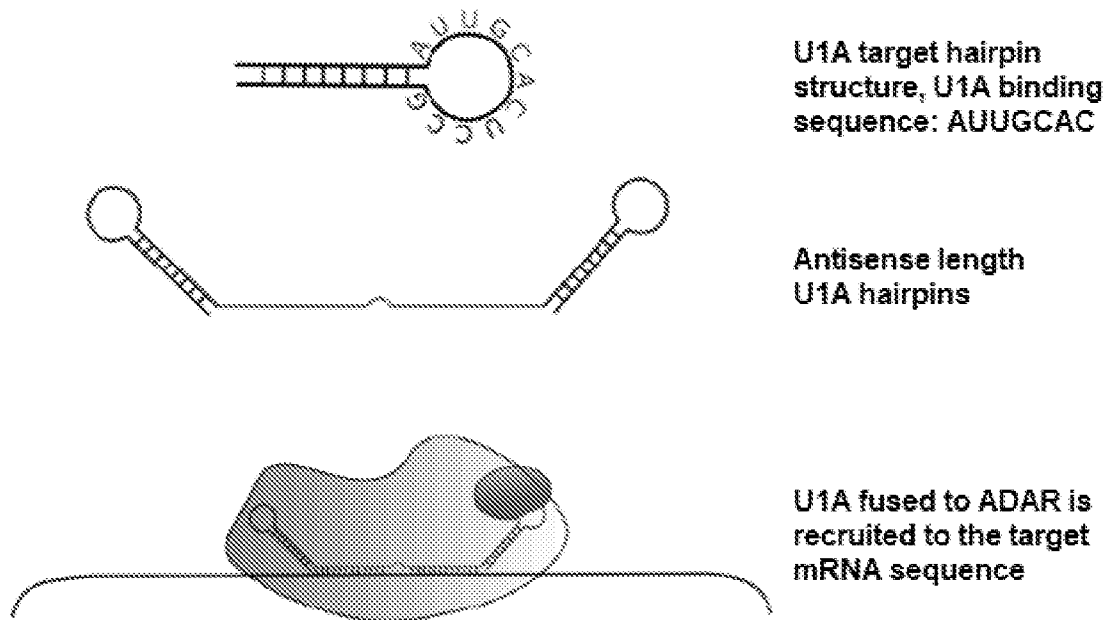
MLPFLERLTLGSDYKDDDDKSGSGSGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYK  
 VTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMEITPIFATNSDCELIKAMQGLLDGNPIPSAIAANS  
 GIYGGSGSGAGSGSPAGGGAPGSGGGSQLHLPQVLADAVSRLVIGKFGDLTDNFSSPHARRIGLAGVVM  
 TGTDVKDAKVICVSTGSKCINGEYLSDRGLALNDCHAEIVSRRSLLRFLYTQLELYLNNEDDQKRSIFQK  
 SERGGFRLKENIQFHLYISTSPCGDARIFSPHEAILEEPA

FIG. 26 continued

21 / 50

**C-terminal-AD2-C-U-variant-10-LN-HA-NES:**

MDRHPNRKARGQLRTKIEAGGGTIPVRNNASIQTWGVLQGERLLTMSCSDKIARWNVVGIQCSLLSIFV  
 EPIYFSSIIILGSLYHGDLHLSRAMYQRISNIEDLPPLYTLNKPLLGTGISNAEARQPGKAPIFSVNWTVGDS  
 AIEVINATTGKGEIGRASRLCKHALYCRWMRVHGKVPShLLRSKITKPNVYHETKLAKEYQAAKARLET  
 AFIKAGLGAWVEKPTEQDQFSLTPGGSGSGAGSGSPAGGGAPGSGGGSNARTRRRERRAEKQAQWKAANG  
 GGGSGGGSGGGGSNARTRRRERRAEKQAQWKAANGGGSGGGSGGGGSNARTRRRERRAEKQAQWKA  
 NGGGSGGGSGGGGSNARTRRRERRAEKQAQWKAANGSYPDVDPDYAGSLPPLERLTL

**FIG. 26 continued****FIG. 27**

22 / 50

**MCP-hAPOBEC1**

MASNETQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNI PPSAIAANS GIYGGSGGSGAGSGSPAGGGAPGSGGG  
 SSGSETPGTSESATPESMTSEKGPSTGDP TLRRIEPWEFDVFDPRELRKEACLLYEIKWGM SRKIWRS  
 SGKNTTNHVEVNFIIKKTSE RDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHM  
 DQQNRQGLRDLVNSGV TIQIMRASEYYHCWRNFVNYPPGDEAHWPQY PPLNMMLYALELHCTILSLPPCL  
 KISR RWQNHLTFFRLHLQ NCHYQTIPPHILLATGLIHPSVAVNR

**MCP-hAPOBEC2**

MASNETQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNI PPSAIAANS GIYGGSGGSGAGSGSPAGGGAPGSGGG  
 SMAQKEEA AVATEAASQNGEDLENLDDPEKLKELI ELPPEI VTGERLPANFFKFQFRNVEYSSGRNKTF  
 LCYVVEAQGGKGGQVQASRGYLEDEHAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPCAACADRI IKT  
 SKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDPFEYVWQNFVEQEEGESKAFQPWEDIQ  
 ENFLYYEEKLADILK

**MCP-hAPOBEC3A**

MASNETQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNI PPSAIAANS GIYGGSGGSGAGSGSPAGGGAPGSGGG  
 SMEASPASGPRHLM DPHIFTSN FNNGIGRHKTYLCYEVE RL DNGTSVKMDQHRGFLHNQAKNLLCGFYGR  
 HAELRFLDLVPSLQLDPAQIYRV TWFI SWSPCF SWGCAGEVRAFLQENTHVLRIFAARIYDYDPLYKEA  
 LQMLRDAGA QVSIMTYDEFEKHCWDTFVDHQGCPFPQPDGLDEHSQALSGLRLRAILQNGN

**MCP-hAPOBEC3B**

MASNETQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNI PPSAIAANS GIYGGSGGSGAGSGSPAGGGAPGSGGG  
 SMNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAEM  
 CFLSWFCGNQLPAYKCFQITW FVSWTPCPDCVAKLA EFLSEHPNVTLTISAARLYYYWERDYRRALCRLS  
 QAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKF DENYAF LHR TLKEILRYLMDPD TTFNFNNDPLVL  
 RRRQTYLCYEVE RL DNGTWV LMDQHM GFLCNEAKNLLCGFYGRHAE LRFDLVPSLQLDPAQIYRV TWFI  
 SWSPCF SWGCAGEVRAFLQENTHVLRIFAARIYDYDPLYKEALQMLRDAGA QVSIMTYDEFEYCWDTFV  
 YRQGCFFQPDGL EEHSQALSGLRLRAILQNGN

**MCP-hAPOBEC3C**

MASNETQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNI PPSAIAANS GIYGGSGGSGAGSGSPAGGGAPGSGGG  
 SMNPQIRNPMKAMYPGTFYFQFKNLWEANDRN ETWLCTTVEGIKRBSVVS WKTVFRNQVDSETHCHAE R  
 CFLSWFCDDILSPNTRYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGRLSL S  
 QEGVAVEIMDYEDFKYCWENFVYNDNEFFKPKWGLKTNPRLKRRRLRESLQ

**FIG. 28**

23 / 50

**MCP-hAPOBEC3D**

MASNFTQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELIIVKAMQGLLKDGNPIPSAIAANSIYGGSGSGAGSGSPAGGGAPGSGGG  
 SMNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEYVKIKRGRSNLLWDTGVFRGPPVLPKRQSNHRQ  
 EVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFSWNPCLPCVVKVTKFLAEHPNVTLTISAARLYYYR  
 DRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMMEAMY  
 PHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSVAFRRKRGVERNQVDPETHCHAERCFLSWFCDDILSPN  
 TNYEVTWYTSWSPCECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGCLCSLSQEGASVKIMGYKDF  
 VSCWKNFVYSDDEFFRPPWKGLQTNFRLLKRRLEILQ

**MCP-hAPOBEC3F**

MASNFTQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELIIVKAMQGLLKDGNPIPSAIAANSIYGGSGSGAGSGSPAGGGAPGSGGG  
 SMKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEYVTKGFSRPRDLAKIFRGQVYSQPEHHAEMC  
 FLSWFCGNQLPAYKCFQITWFSWTPCPDCVAKLAEPFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQ  
 AGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMMEAMYPHIFYFHFKNLRK  
 AYGRNESWLCFTMEVVKHHSVPVSWKRGVERNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSP  
 CPECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGRLSLSQEGASVEIMGYKDFKYCWENFVYNDDE  
 PFKPWKGLKYNFLFLDSKLQEILE

**MCP-hAPOBEC3G**

MASNFTQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELIIVKAMQGLLKDGNPIPSAIAANSIYGGSGSGAGSGSPAGGGAPGSGGG  
 SMKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEYVTKGFSRPRDLAKIFRGQVYSELKYHPME  
 FFHWFSKWRKLHRDQEYEVWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEAALRSLC  
 QKRDGPRATMKIMNYDEFQHCWSKFVYSQRELEFPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNE  
 FWVRGRHETYLCEYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLEVPFWKLDLDQDYRV  
 TCFTSWSPCFSCAQEMAKFISKNHVSLCIFTARIYDDQGRQCQGLRTLAEAGAKISIMTYSEFKHCWDT  
 FVDHQGCFFQPWDGLDEHSQDLSCRLRAILQNQEN

**MCP-hAPOBEC3H**

MASNFTQFVLVDNNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCVSRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELIIVKAMQGLLKDGNPIPSAIAANSIYGGSGSGAGSGSPAGGGAPGSGGG  
 SMALLTAETFRLOFNKPRRLRFPYPRKALLCYQLTFCNGSTPTRGYFENKKKCHAEICFINEIKSMGLD  
 ETQCYQVTCYLTWSPCSSCAWELVDFIKAHDLNLGIFASRLYYHWCKPQQKGLRLLCGSQVFVEVMGFP  
 EFADCWENFVDHEKPLSFNPKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV

**FIG. 28 continued**

24 / 50

**MCP-hAPOBEC4**

MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELVKAMQGLLKDGNIPIPSAIAANSGLYGGSGGAGSGSPAGGGAPGSGGG  
 SMEPIYEEYLANHGTTIVKPYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTFPQTKHLTFY

ELKTSSGSLVQKGRASSCTGNYIHFESMLFEMNGYLDSAIYNNDSIRHIILYSNNSPCNEANHCCISKMY  
 NFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSPISGVSGSH  
 VFQPIILTGRALADRHNAYEINAITGVKPYFTDVLQTKRNPNTKAGEALESYPLNNAFFGQFFQMPSPGQL  
 QPNLFFDLRAFPVFLVFLRDLFPFMMHGQNPENKPRNIVRHLNMPQMSFQETKDLGLELPTGRSVEIVEITE  
 QFASSKEADEKKGKKGK

**MCP-AID**

MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELVKAMQGLLKDGNIPIPSAIAANSGLYGGSGGAGSGSPAGGGAPGSGGG  
 SMDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRDSATSFSLDFGYLRNKNKGCHVELLFLRYISDWD  
 LDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFCEDRKAEPFGLRRLHRAGVQIAIM  
 TFKDYFYCWNTFVENHERTFKAWEGHENSVRLSRQLRRILLPLYEVDDLRLDAFRTLGL

**MCP-hADAT1**

MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW  
 RSYLNMELTIPIFATNSDCELVKAMQGLLKDGNIPIPSAIAANSGLYGGSGGAGSGSPAGGGAPGSGGG  
 SMWTADEIAQLCYEHYGIKRLPKKGGKPEPNHEWTLLAAVVKIQSPADKACDTPDKFVQVTKEVVSMTGTGK  
 CIGQSKMRKNGDILNDSHAETIARRSFQRYLLHQLQLAATLKEDSIFVPGTQKGVWKLRRDLIFVVFSSH  
 TPCGDASIIPMLEFEDQPCCPVFRNWAHNSSEASSNLEAPGNERKCEDPDSFVTKKMRLEPGTAAREVT  
 NGAHRHQSFGKQKSGPISPGIHSCDLTVEGLATVTRIAPGSAKVIDVYRTGAKCVPGEAGDSGKPGAAFH  
 QVGLLRVKPGRGDRTSMSCSDKMARWNVLGCGALLMHLLLEPIYLSAVVIGKCPYSQEAMQALIGRC  
 QNVSAALPKGFGVQELKILQSDLLFEQSRSAVQAKRADSPGRLVPCGAISWSAVPEQPLDVTANGFPQGT  
 TTKTIGSLQARSQISKVELFRSFQKLLSRIARDEKWPMSLRVQKLDITYQEYKEAASSYQEAWSLTKKQVFG  
 SWIRNPPDYHQFK

**FIG. 28 continued**

25 / 50

Target locus: GFP L202S

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTTCGAGCTGGACGGCGACGTAA  
 ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTT  
 CATCTGCACCAACGGCAAGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGACCTACGGCGTGCAG  
 TGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG  
 TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGG  
 CGACACCCCTGGTGAACCGCATCGAGCTGAAGGECATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC  
 AAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGG  
 TGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC  
 CCCCATCGGGCGACGGCCCGTGTCTGCTGCCCCACAATCATTTATTCGAGCACTCAGTCCGCCCTGAGCAAA  
 GACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCA  
 TGGACGAGCTGTACAAGTACTCAGATCTCGAGCTCAAGTAG

apRNA\_v1:

gcgcctccatcgaagtgtaaagccatctatcgatgggaaaGTCTTTGCTCTCCCCGGACACTCTGCTCCT  
 TAAATGAAACACGGGCTCGTGCACGGGGCCaaagcgcctccatcgaagtgtaaagccatctatcgatggg

apRNA\_v2:

gccatctatcgatgggaaaGTCTTTGCTCTCCCCGGACACTCTGCTCCTTAAATGAAACACGGGCTCGTG  
 CACGGGGCCaaagccatctatcgatggg

apRNA\_v3:

gccatctatcgatgggaaaCGGACTGAGTGCTCaaAATGATTGTGGGGCaaagccatctatcgatggg

apRNA\_v4:

gccatctatcgatgggaaaTGCTCAGGGCGGACTGAGaaaaTTGTGGGGCAGCAGCACGaaagccatcta  
 tcgatggg

FIG. 29



26 / 50

MS2-apRNA\_v1:

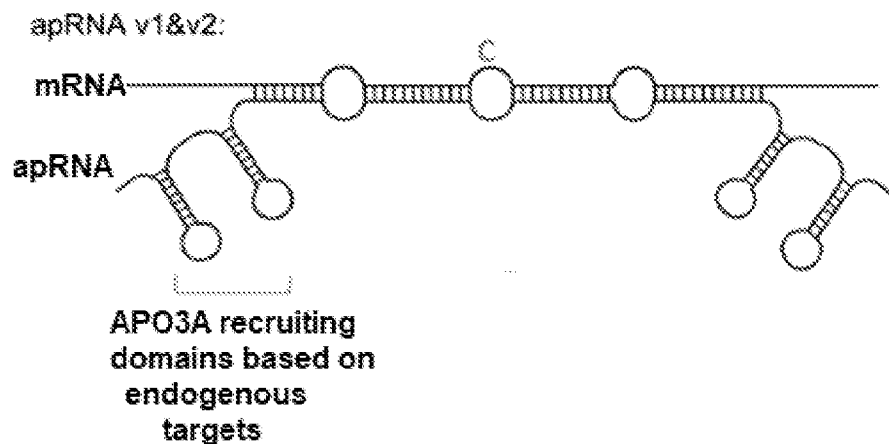
aACATGAGGATCAGCCATGTcGTCTTTGCTCTCCCCGGACACTCTGCTCCTTAAATGAAACACGGGCTCG  
TGCACGGGGCCaACATGAGGATCAGCCATGTc

MS2-apRNA\_v2:

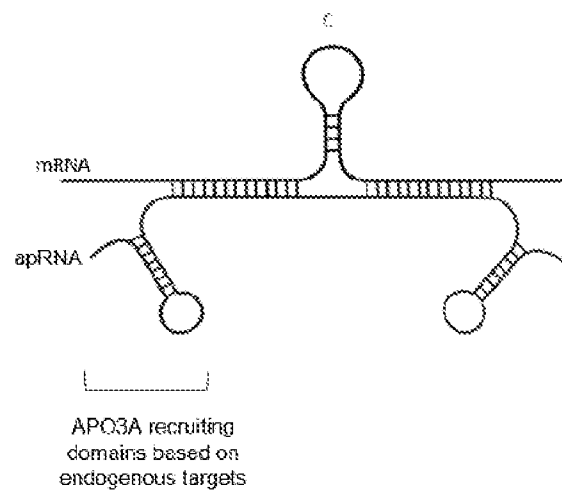
aACATGAGGATCAGCCATGTcCGGACTGAGTGCTCaAATGATTGTGGGCaACATGAGGATCAGCCATG  
Tc

MS2-apRNA\_v3:

aACATGAGGATCAGCCATGTcTGCTCAGGGCGGACTGAGaaaaTTGTCGGGCAGCAGCAGCaACATGAGG  
ATCAGCCATGTc



apRNA v3&amp;v4:



APO3A recruiting domain sequences:

1. ASCC2: ggcctccatcgaagtgt
2. SDHB: gccatctatcgatggg

FIG. 30

27 / 50

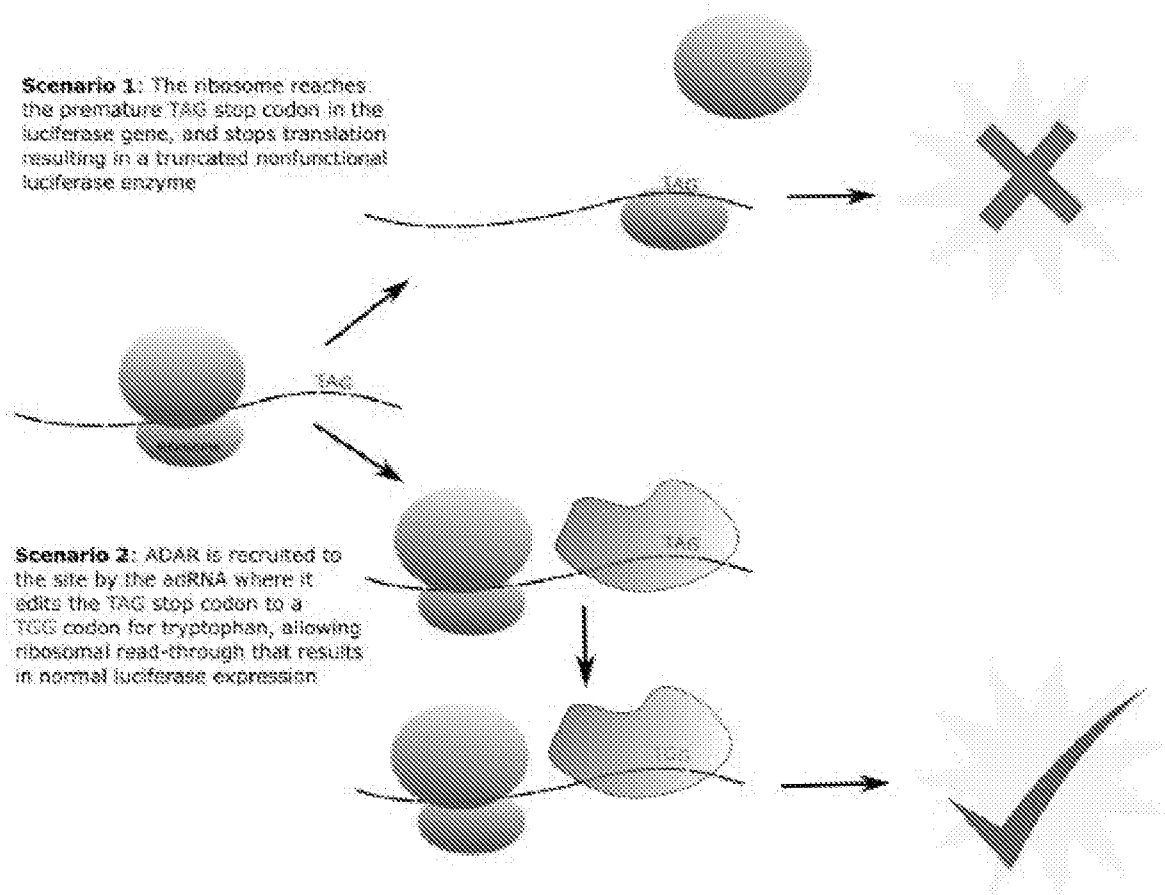


FIG. 31

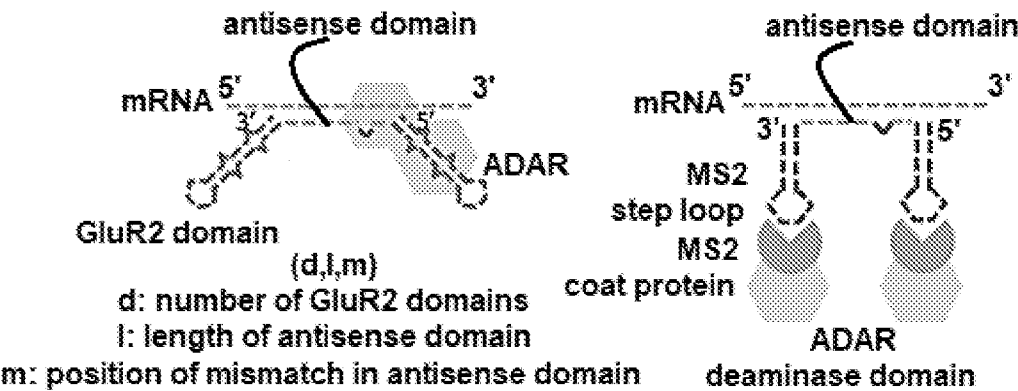


FIG. 32A

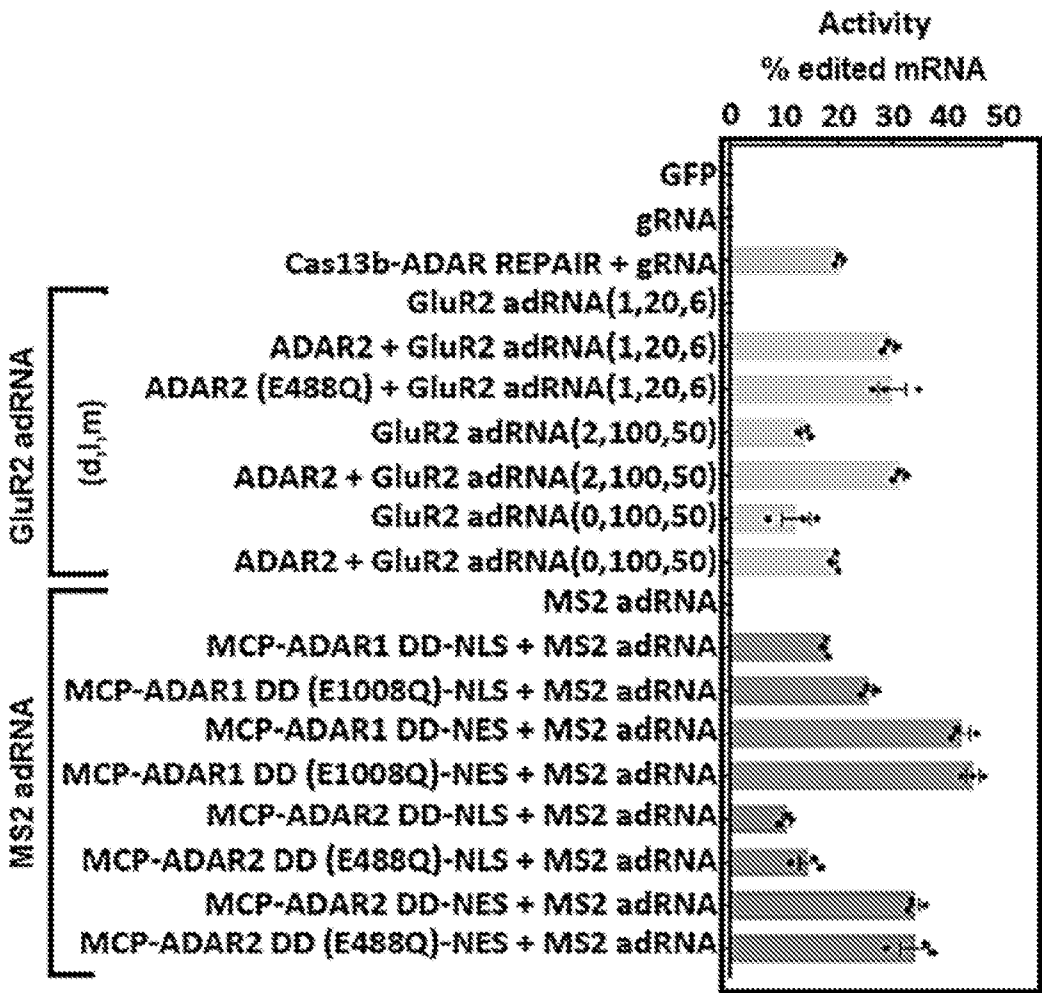


FIG. 32B

29 / 50

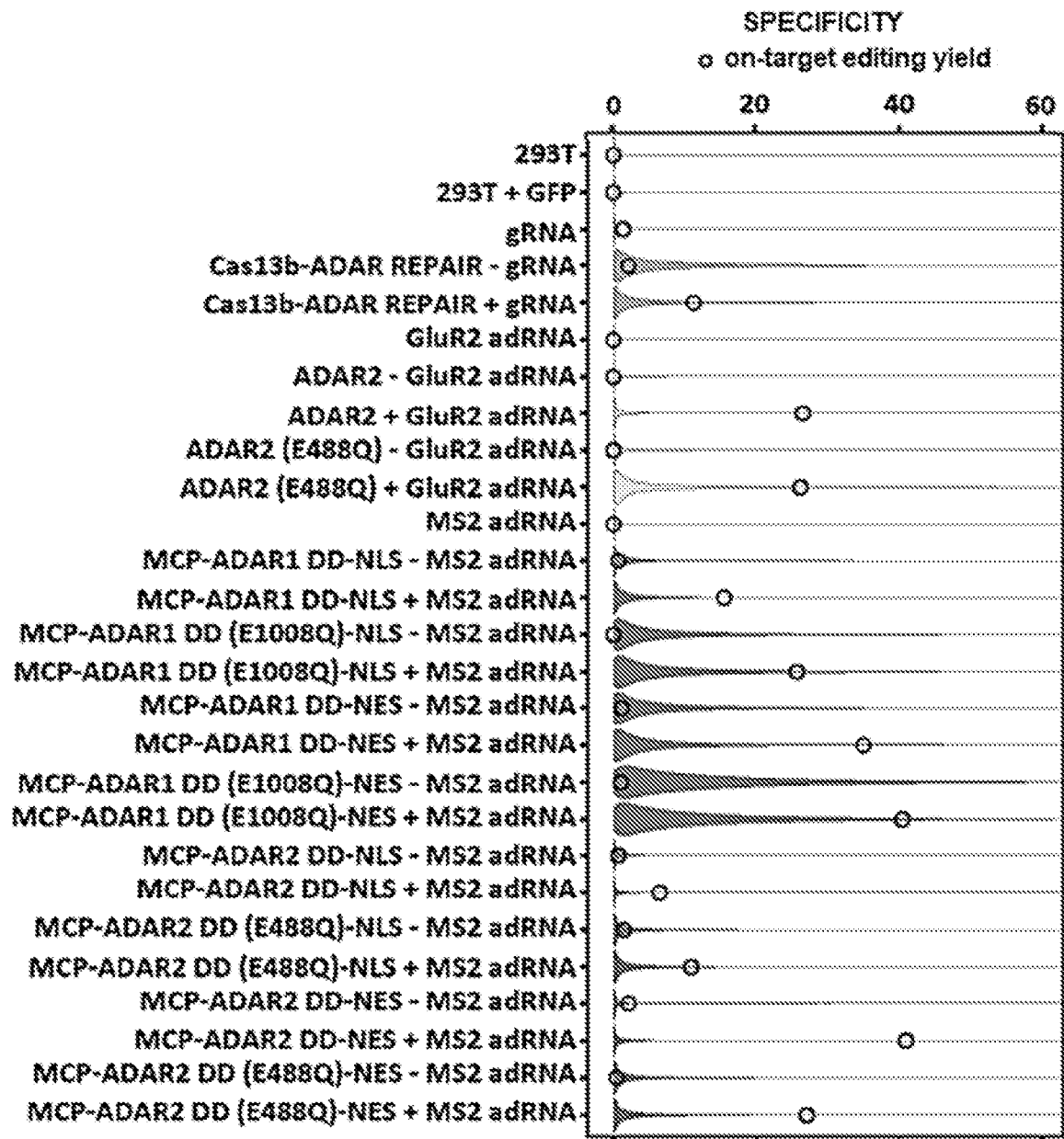


FIG. 32C

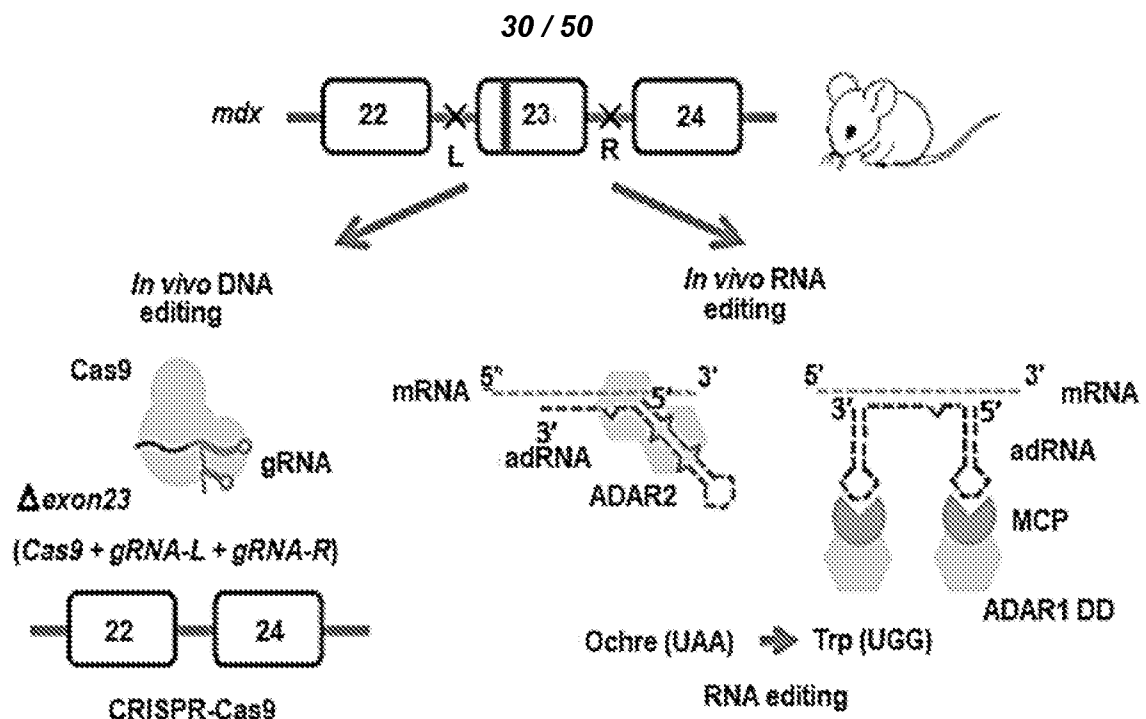


FIG. 33A

31 / 50

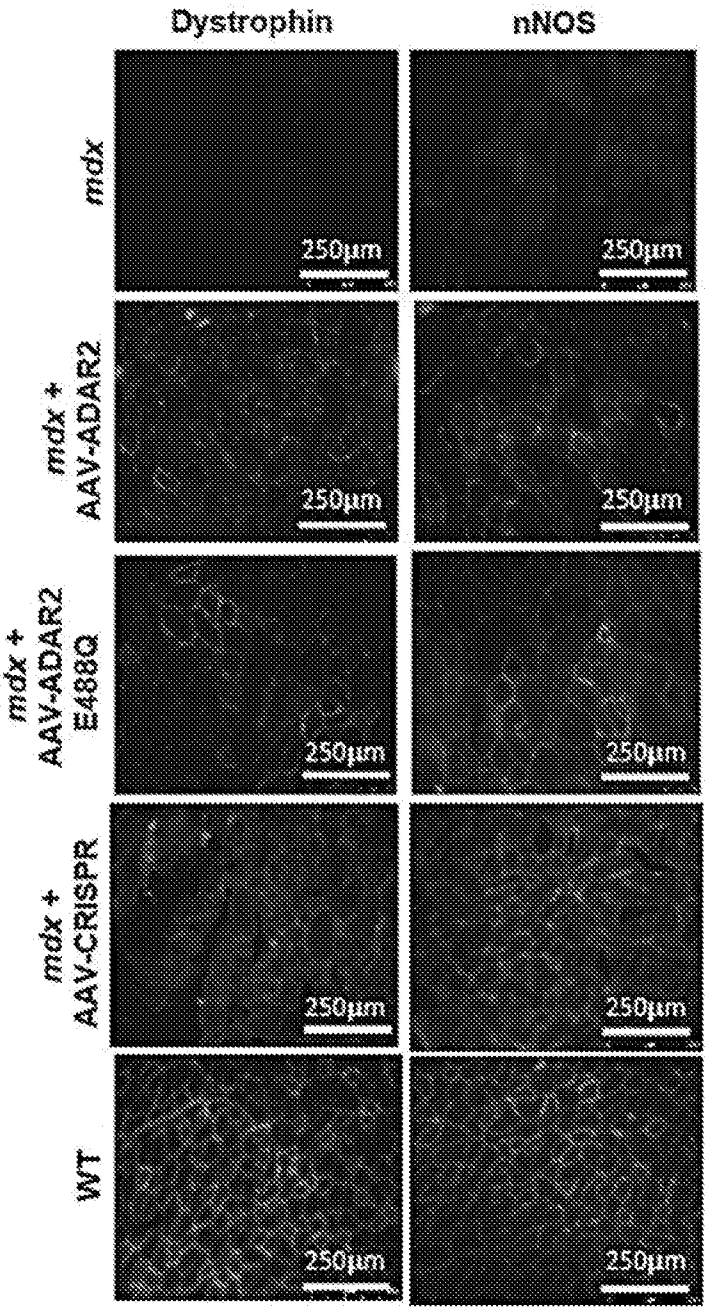


FIG. 33B

32 / 50

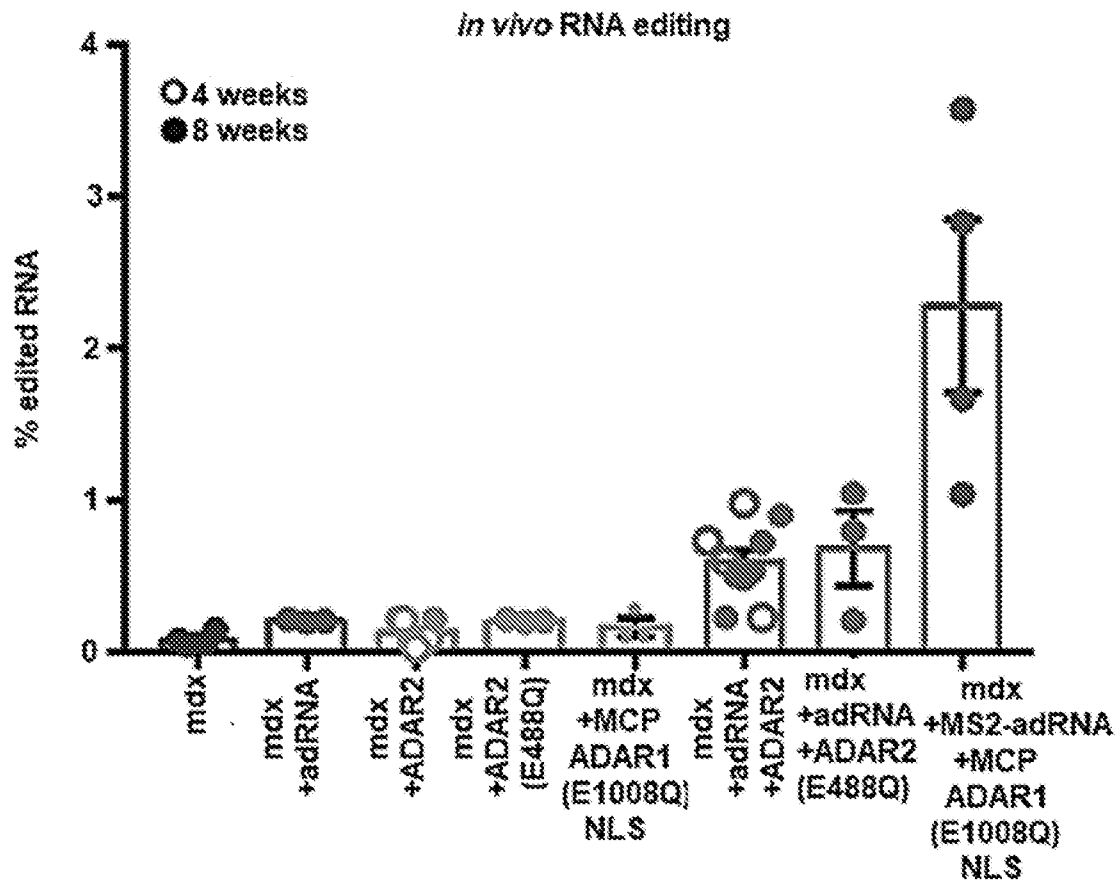


FIG. 33C

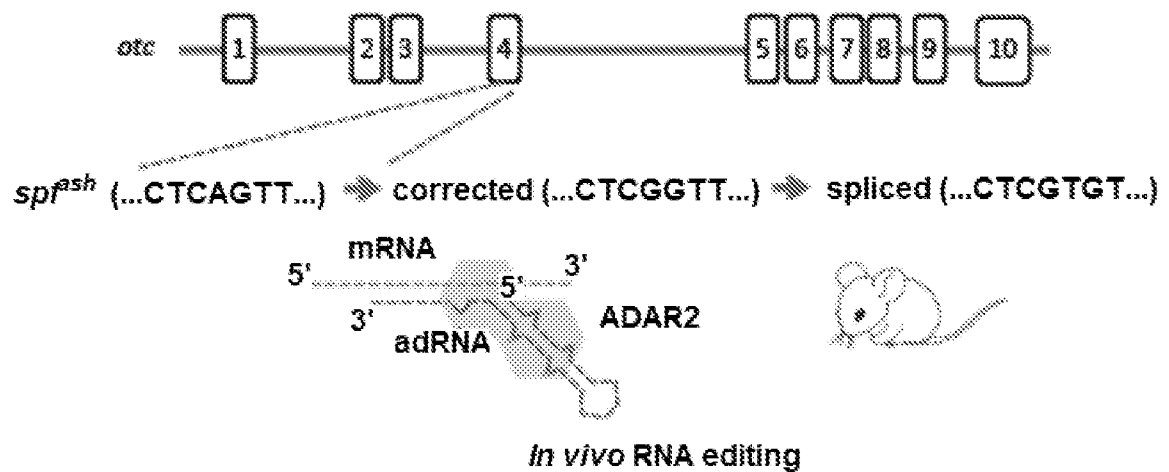


FIG. 33D

33 / 50

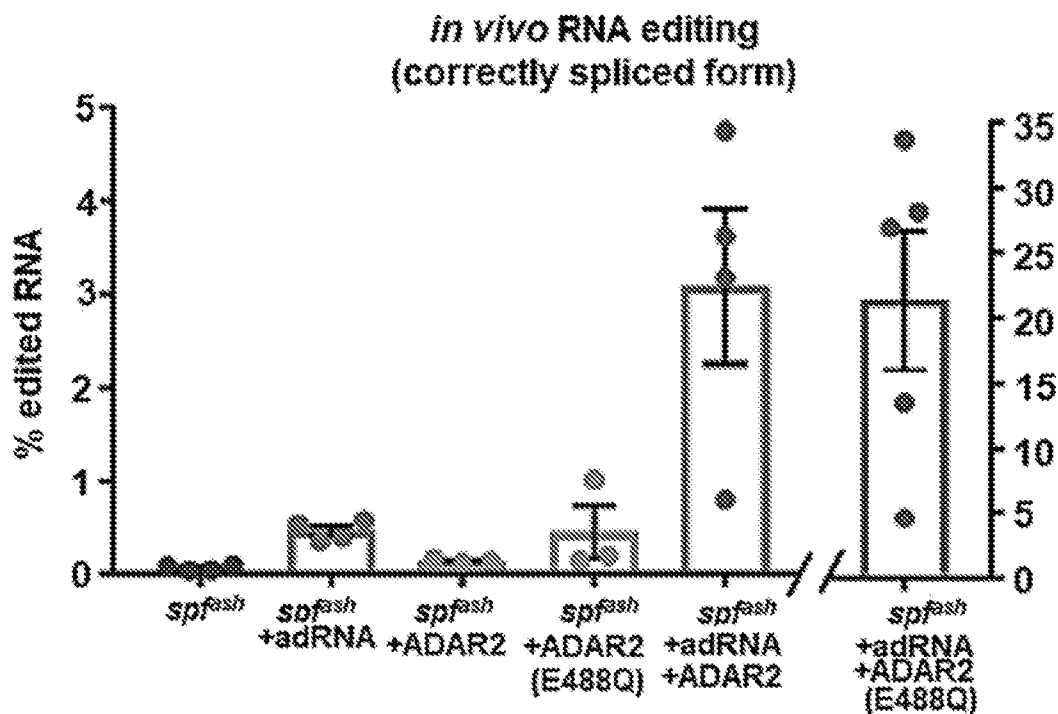


FIG. 33E

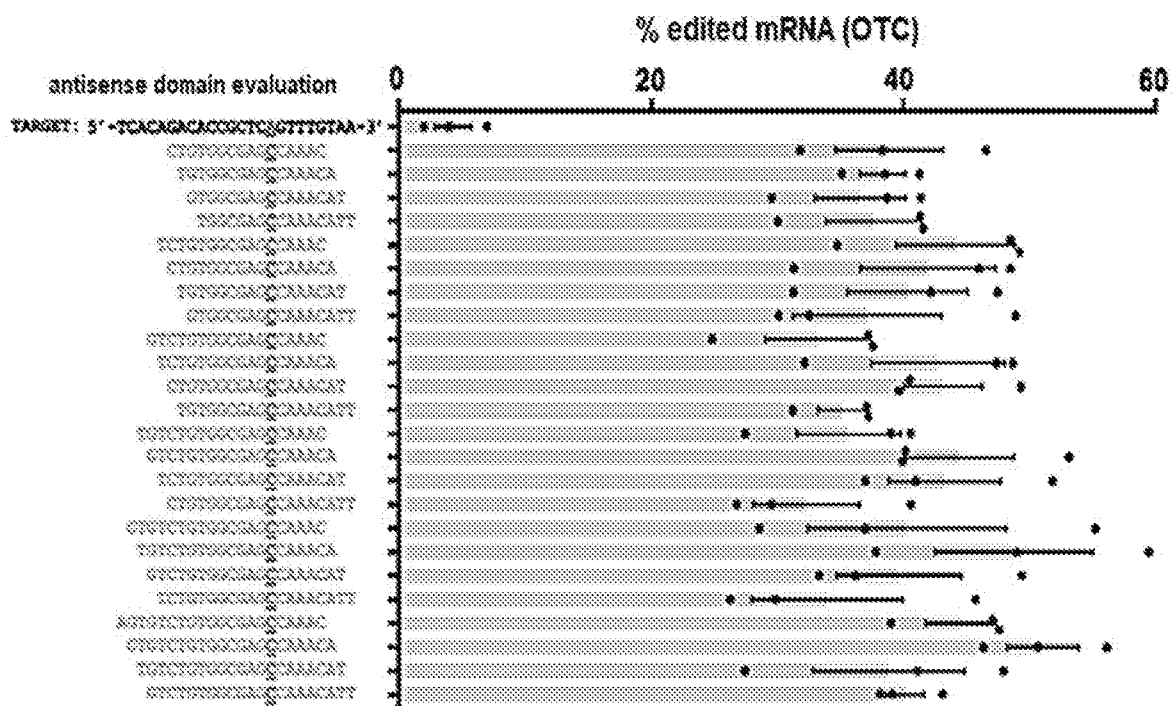


FIG. 34A



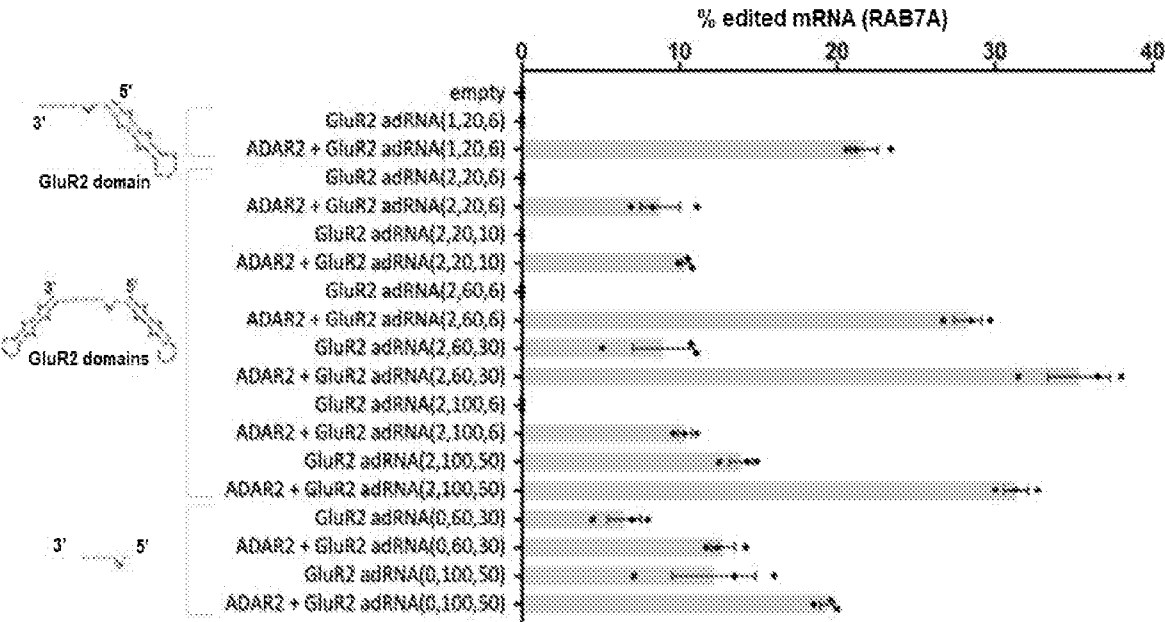


FIG. 34B

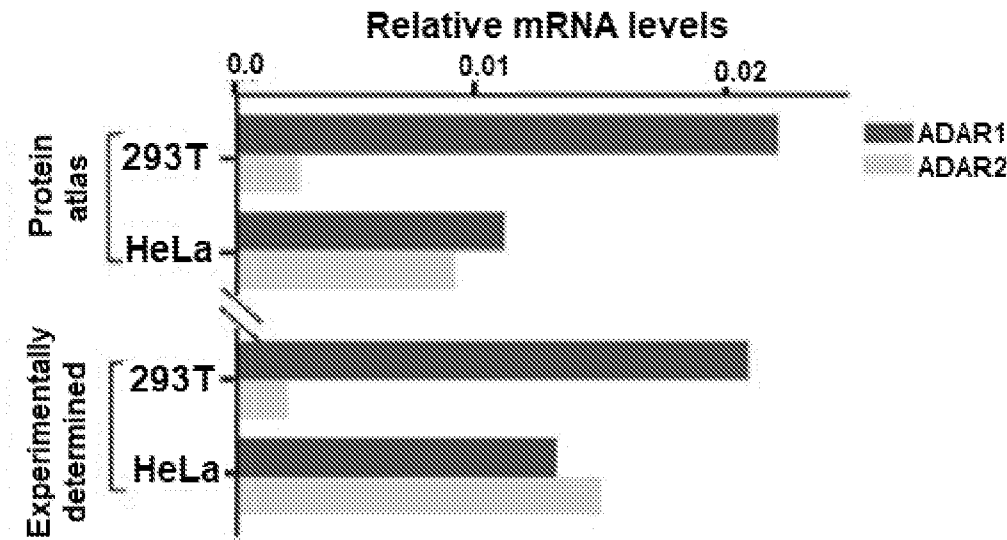


FIG. 34C

**35 / 50**

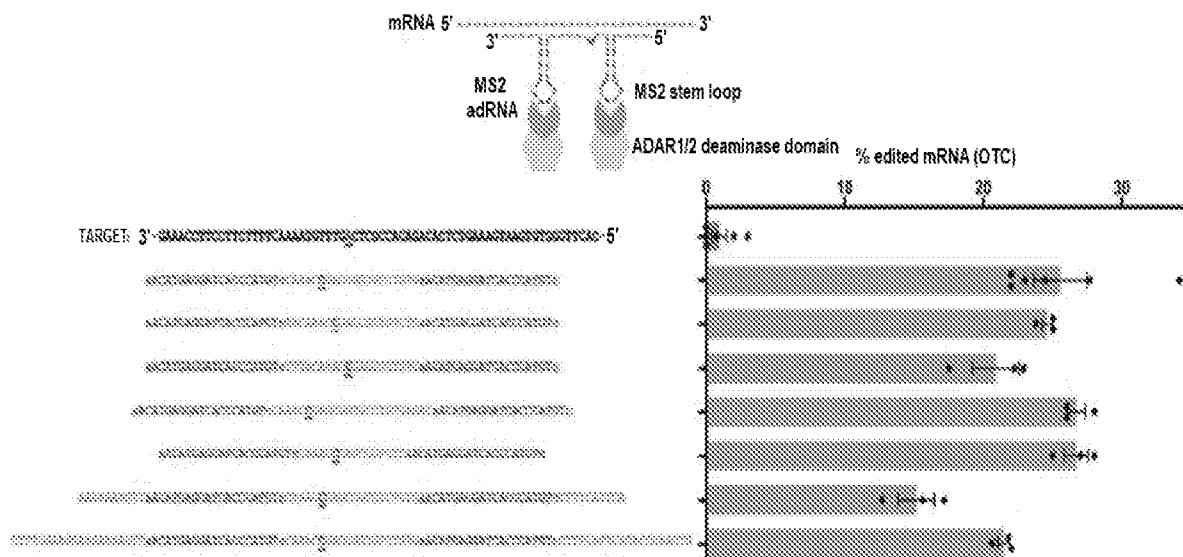


**FIG. 35A**

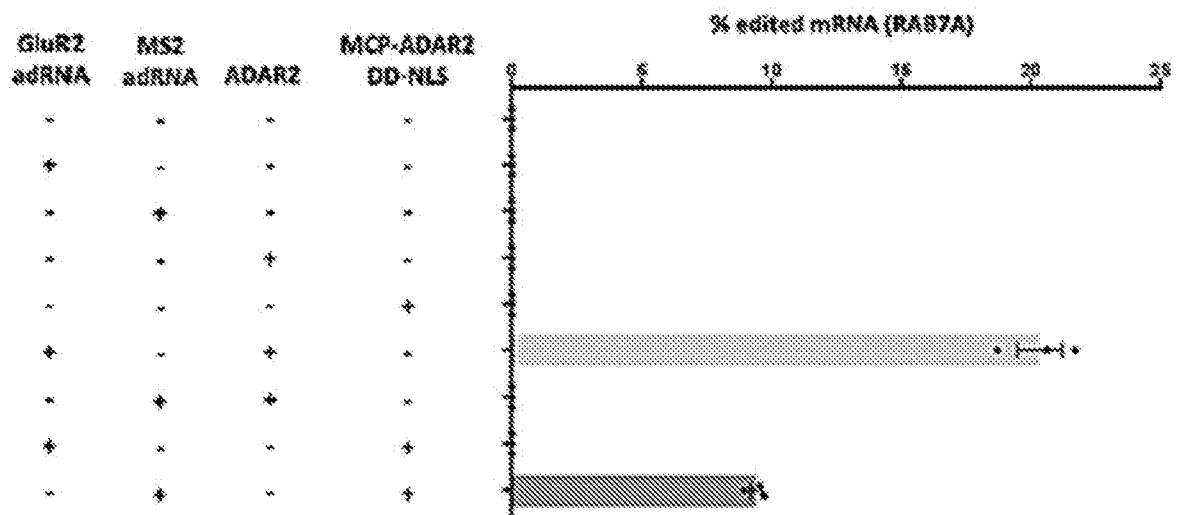
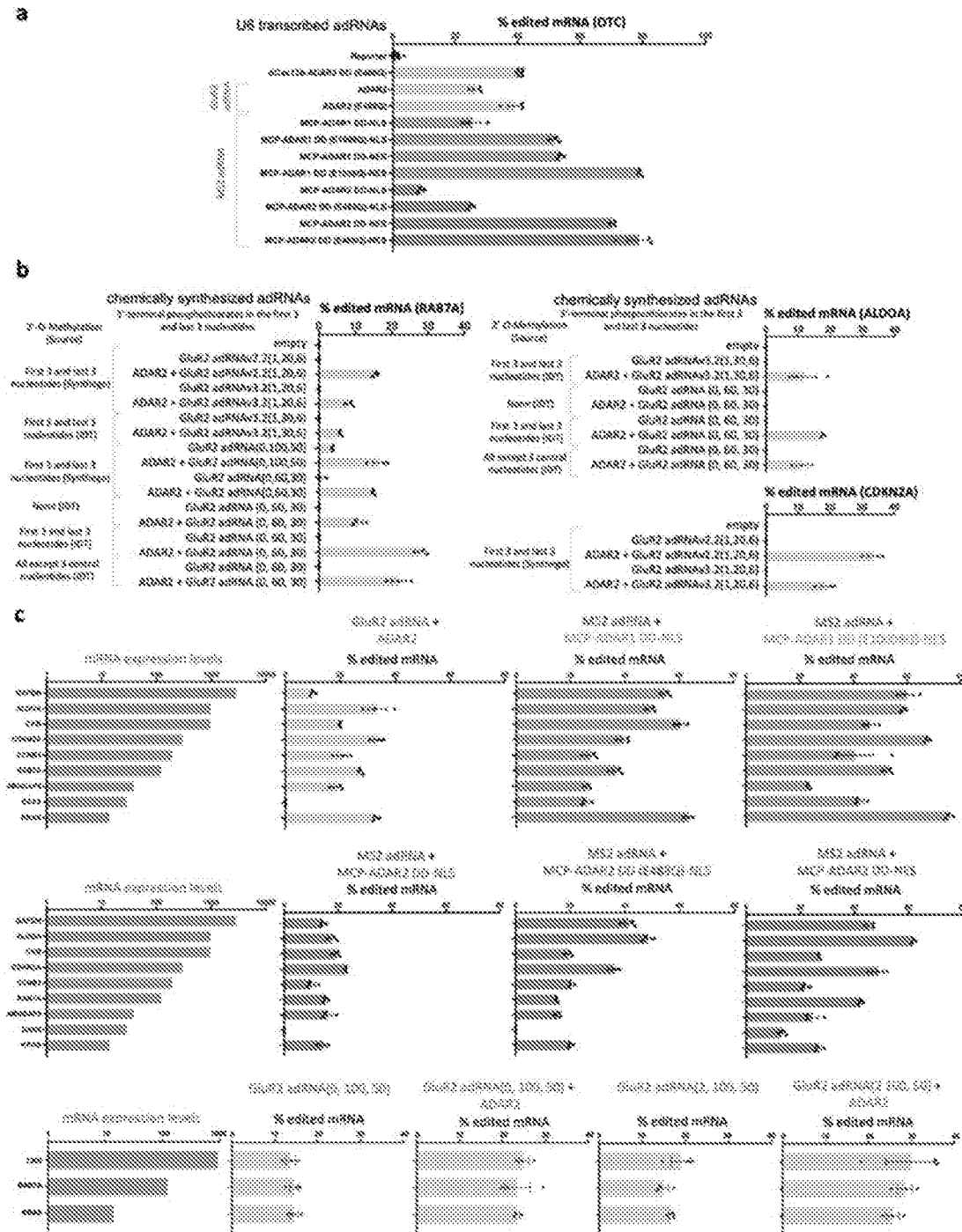


**FIG. 35B**

**36 / 50**



**Supplementary Figure 4: Characterization of editing rates across a spectrum of genes with varying expression levels**

**FIG. 36A-C**

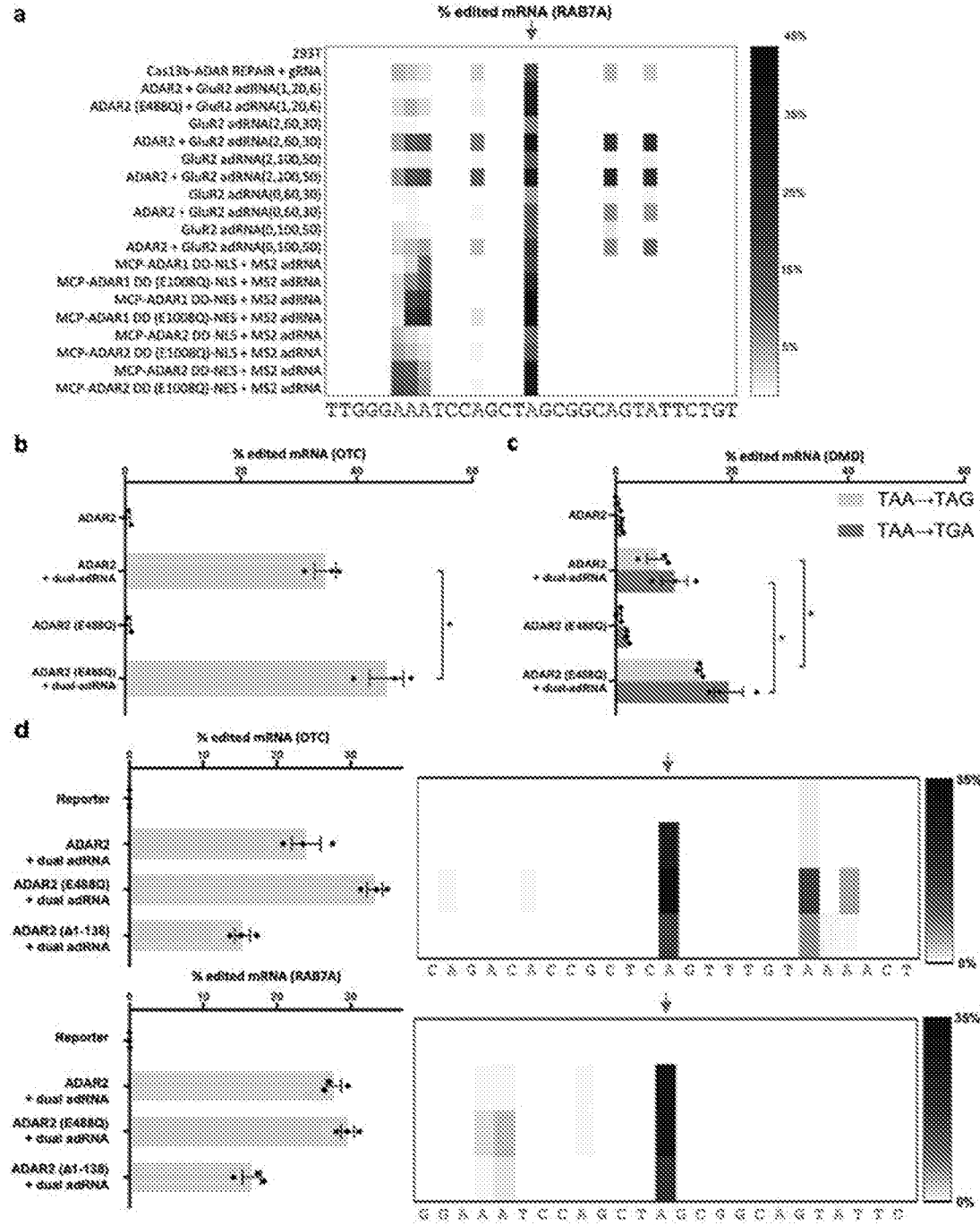


FIG. 37A-D

38 / 50

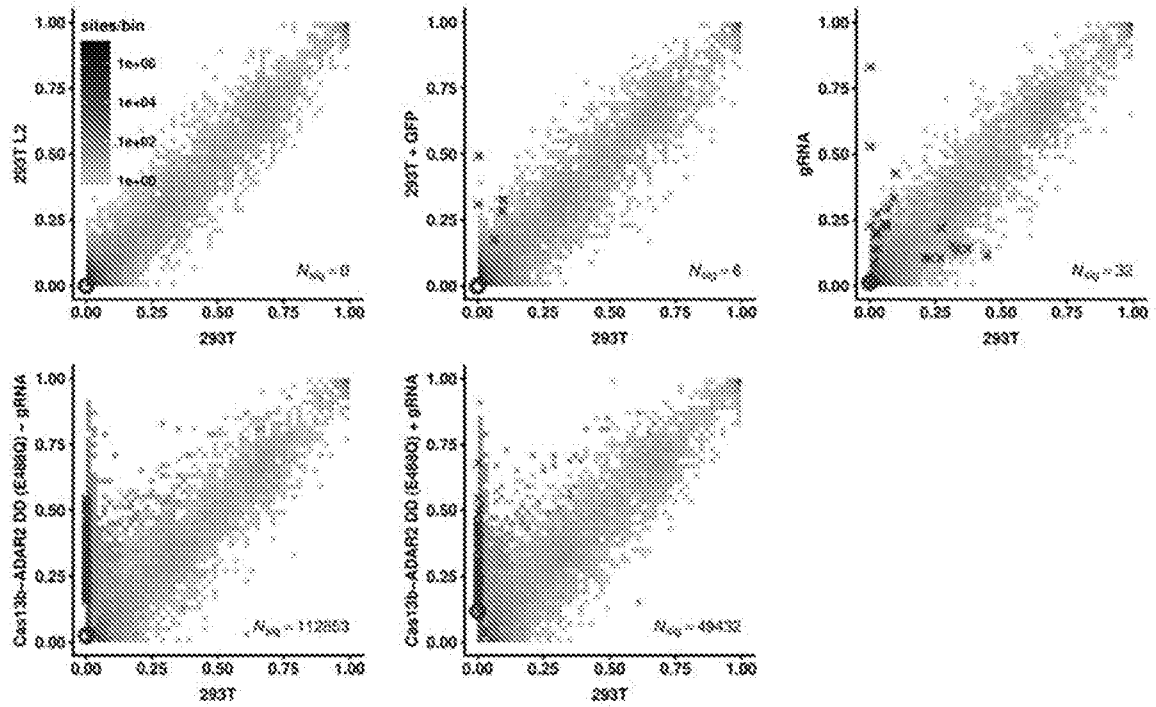


FIG. 38

39 / 50

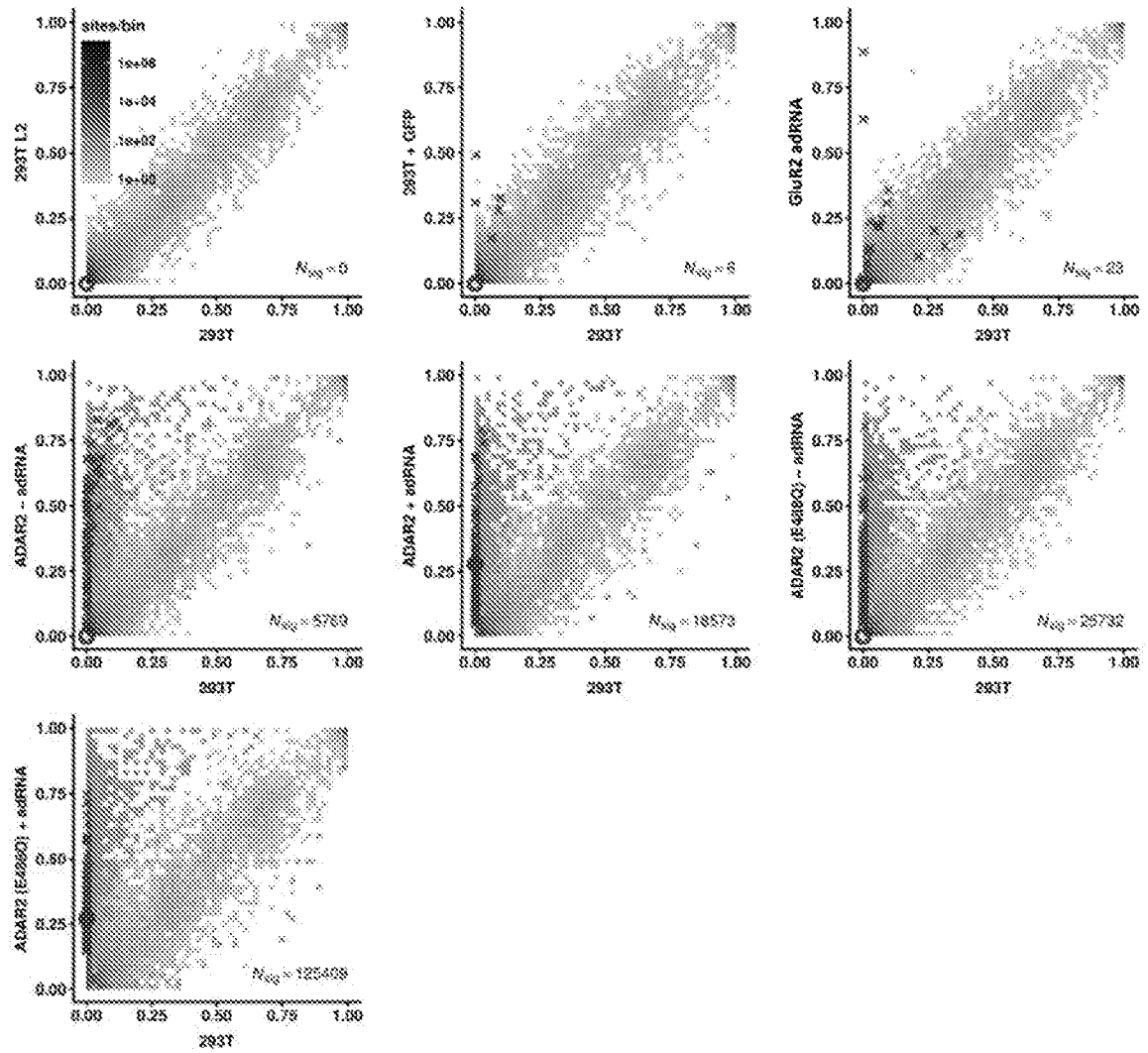


FIG. 39

40 / 50

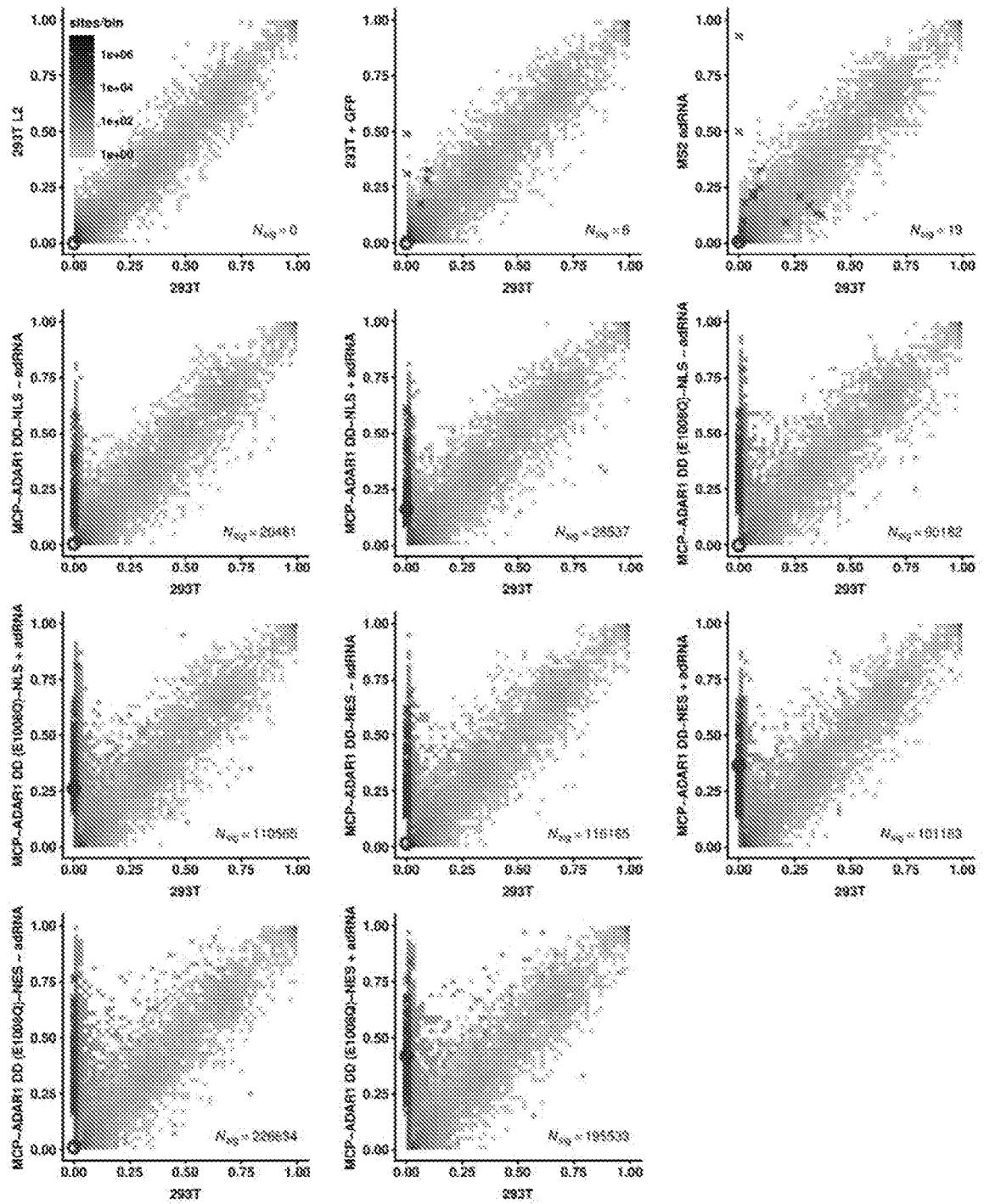


FIG. 40

41 / 50

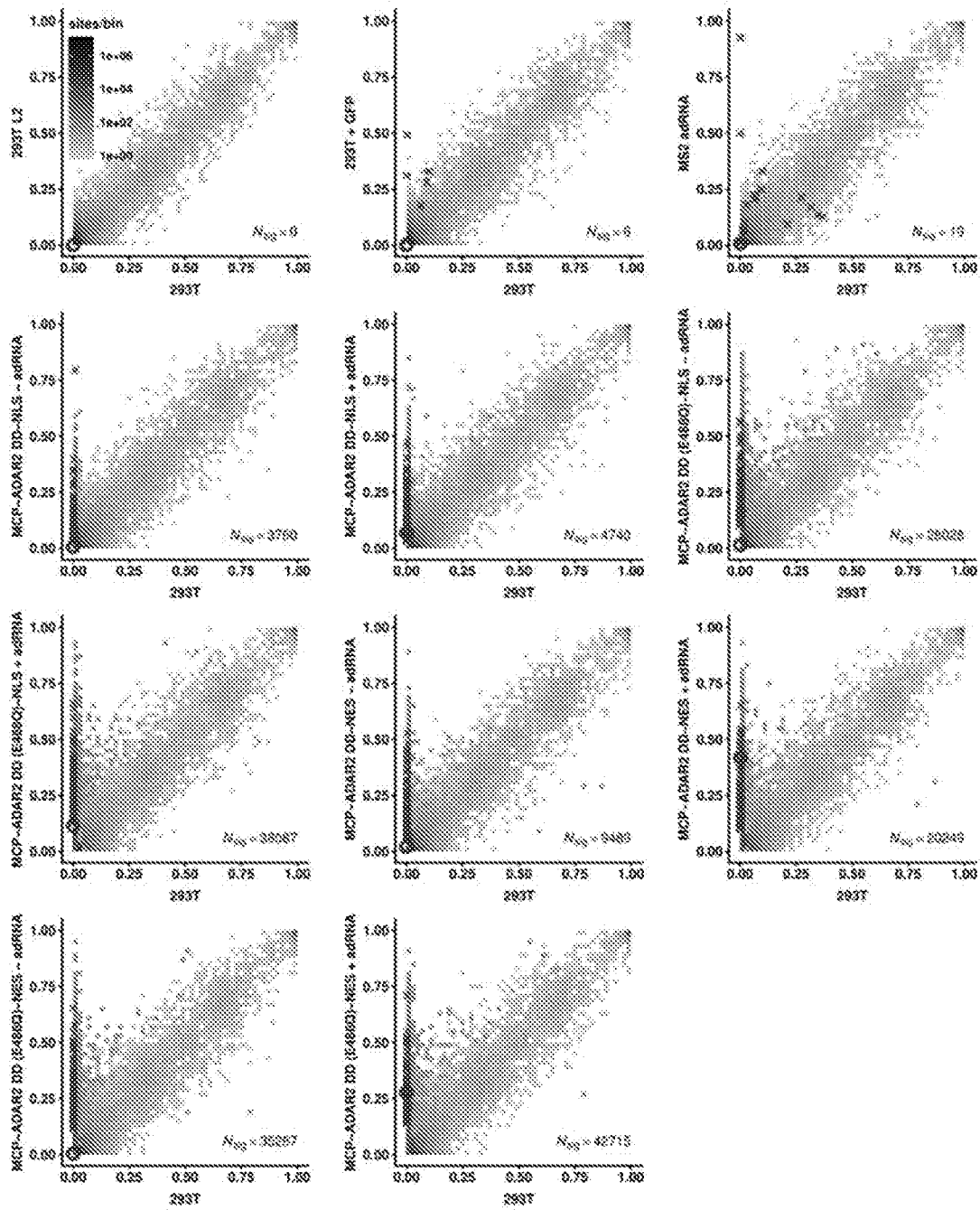


FIG. 41



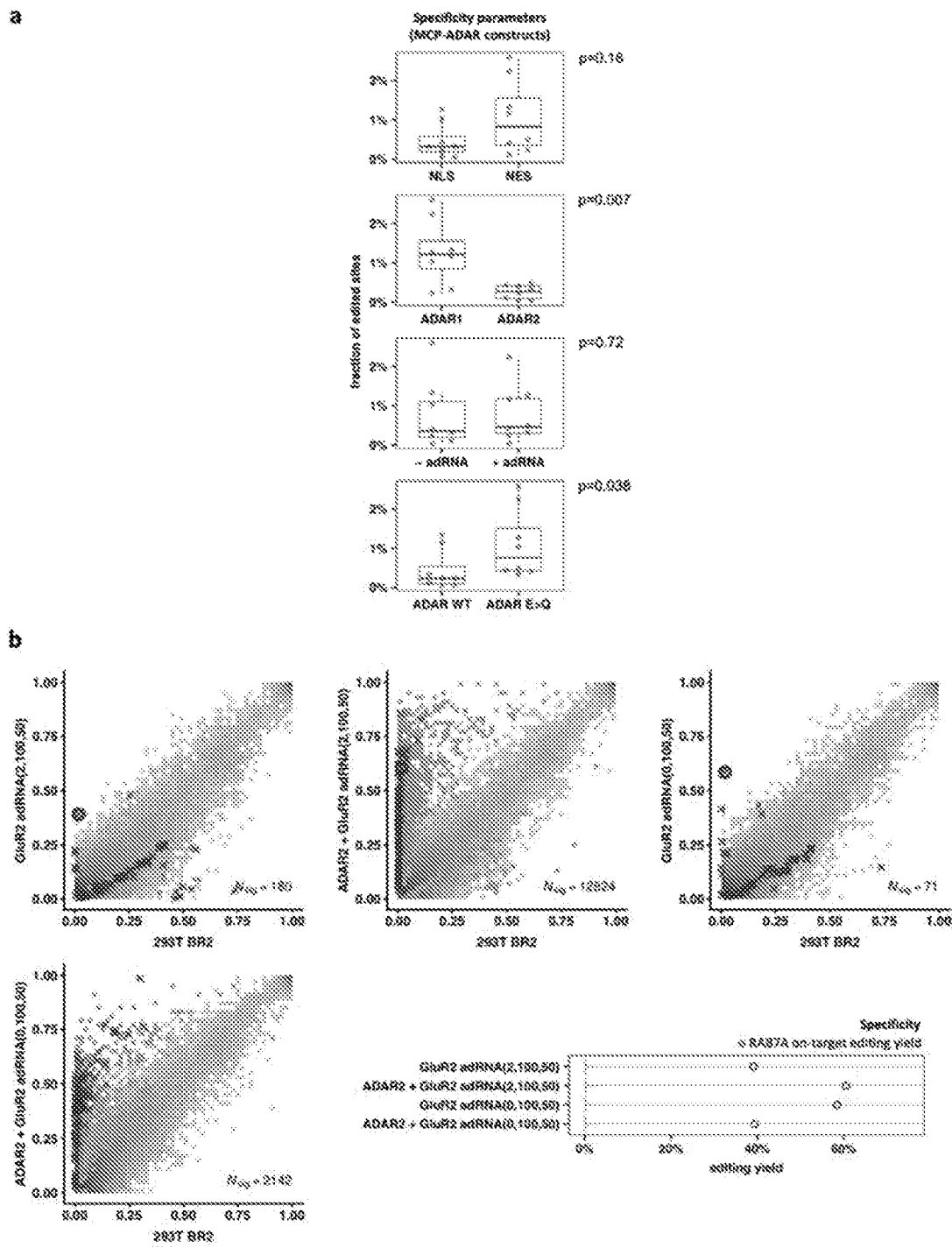


FIG. 42A-B

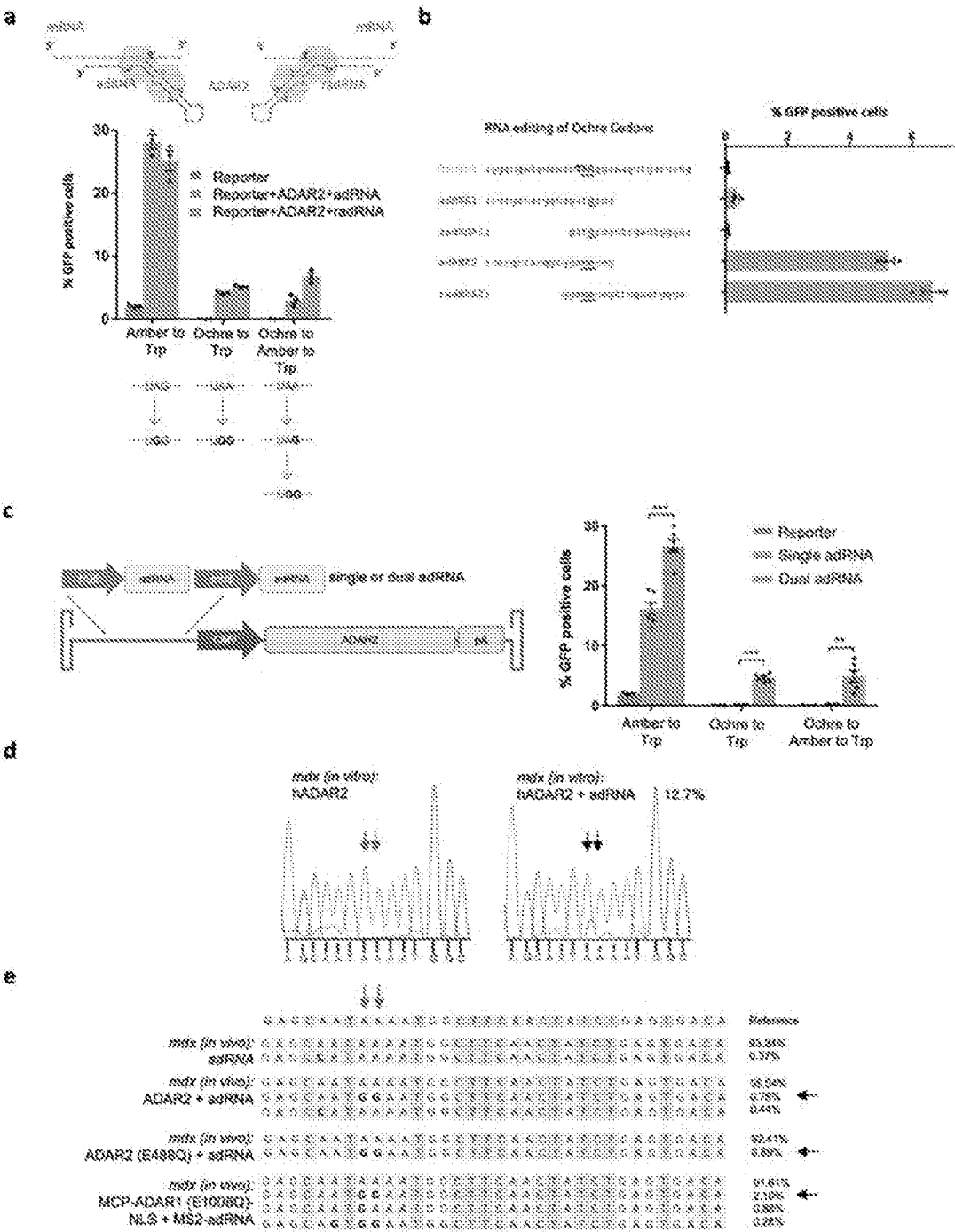


FIG. 43A-E

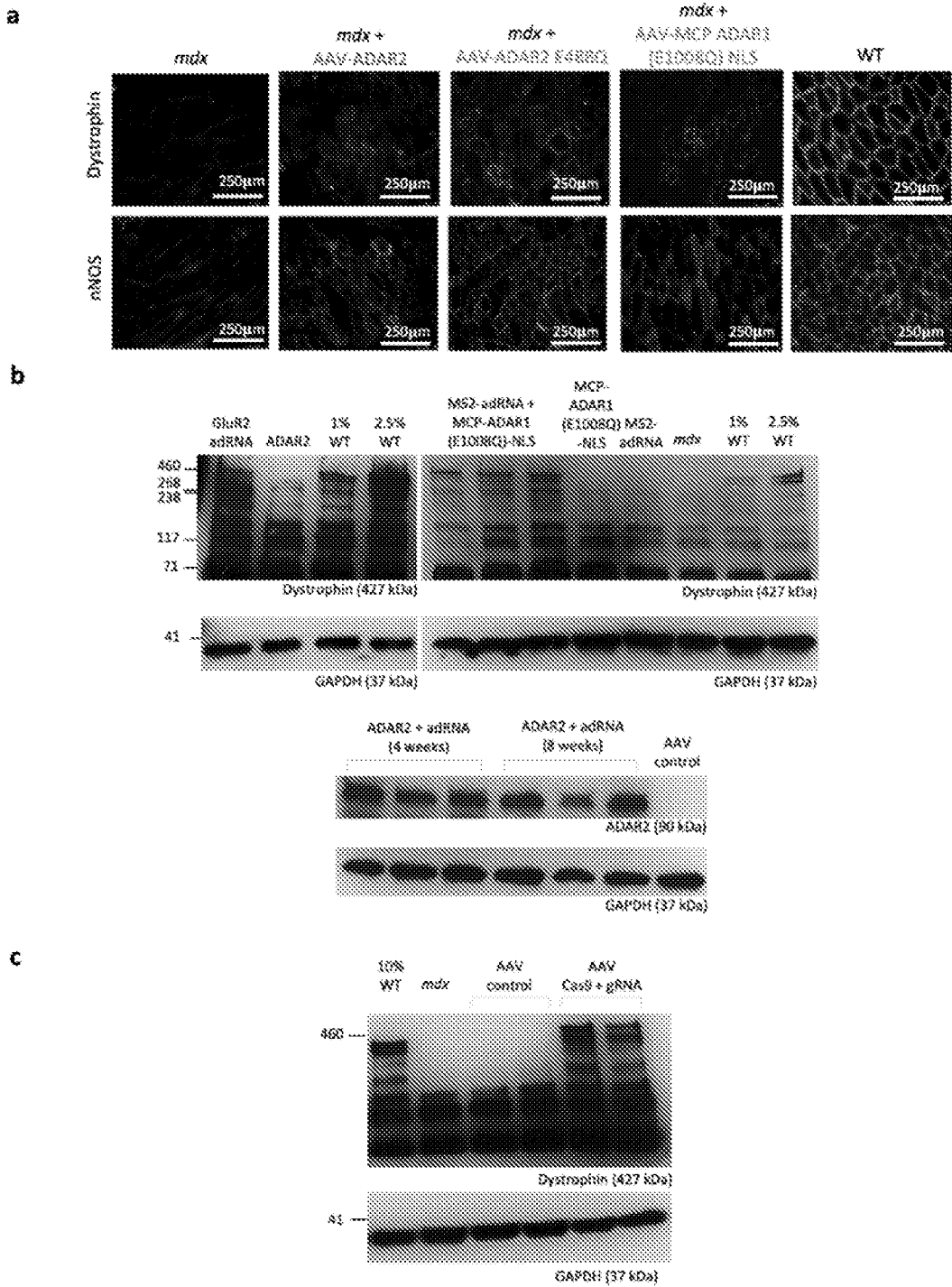


FIG. 44A-C

45 / 50

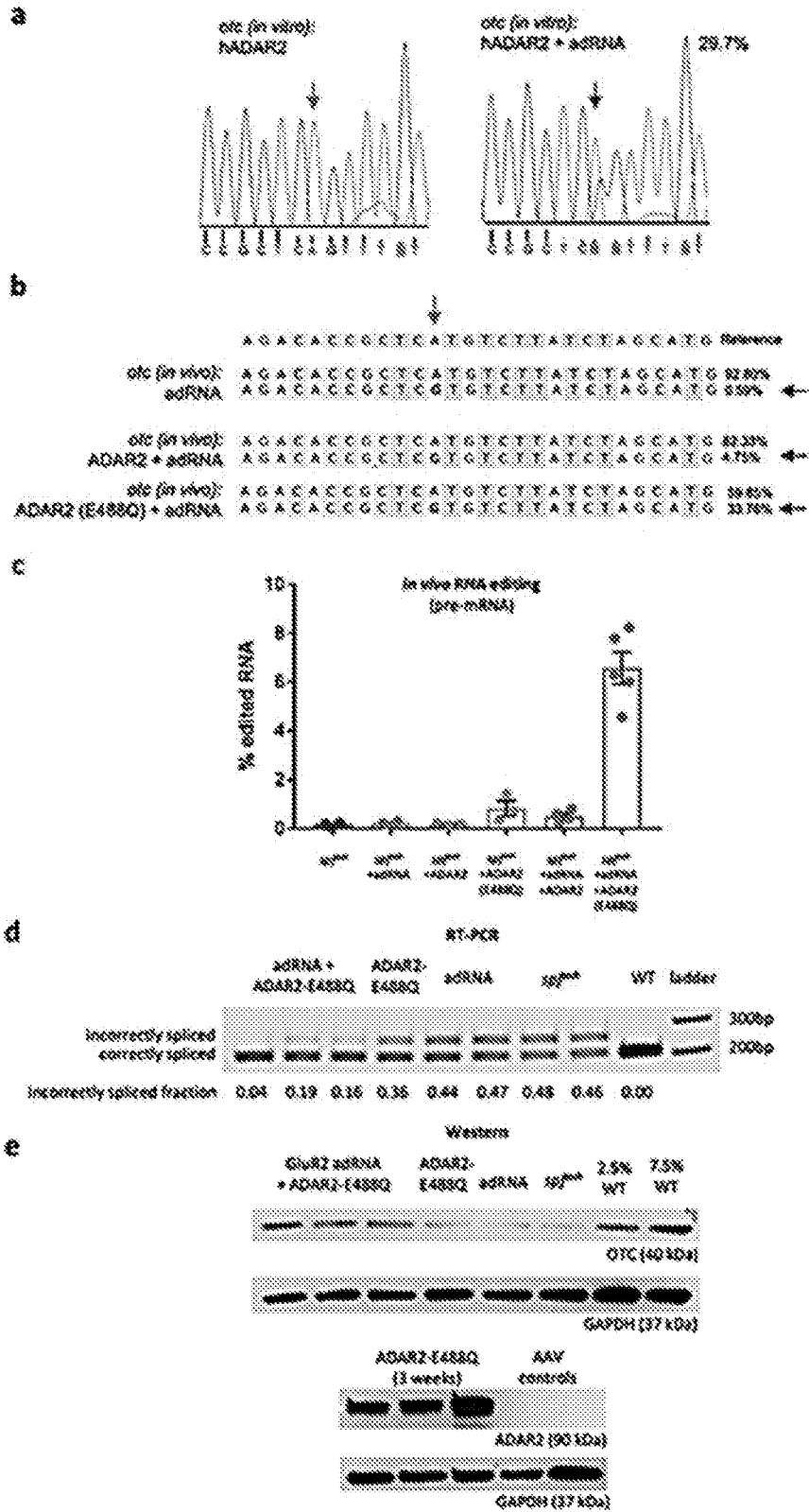


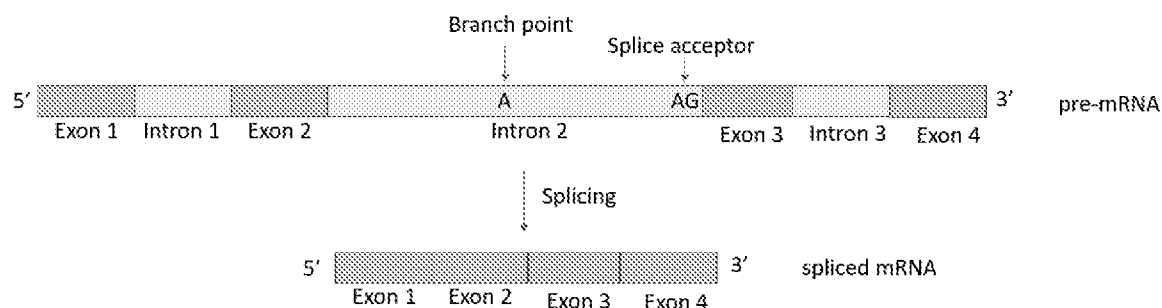
FIG. 45A-E

46 / 50

| Perturbation | Delivery route | % sick mice<br>(n=total mice; assay = 3 weeks) |
|--------------|----------------|------------------------------------------------|
| Saline       | retro-orbital  | 0% (n=8)                                       |
| Saline       | retro-orbital  | 0% (n=13)                                      |
| Saline       | retro-orbital  | 50% (n=14)                                     |
| Saline       | intra-muscular | 0% (n=9)                                       |
| Saline       | intra-muscular | 6% (n=18)                                      |
| Saline       | intra-muscular | 17% (n=6)                                      |
| Saline       | intra-muscular | 9% (n=11)                                      |
| Saline       | intra-muscular | 0% (n=3)                                       |

FIG. 46

Correct splicing:



Exon skipping via creation of A->G mutations at the splice acceptor site and/or the branch point:

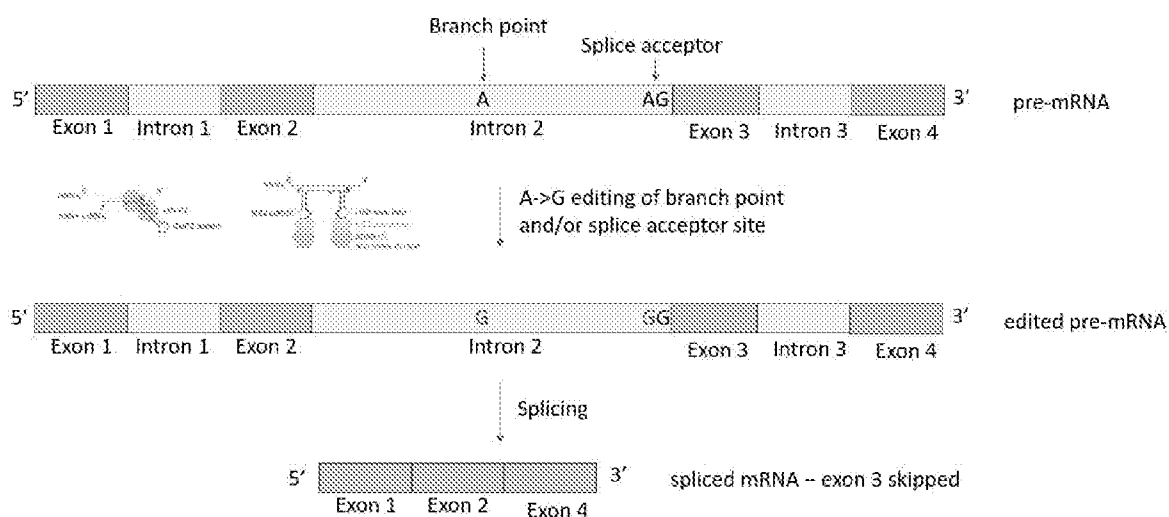


FIG. 47

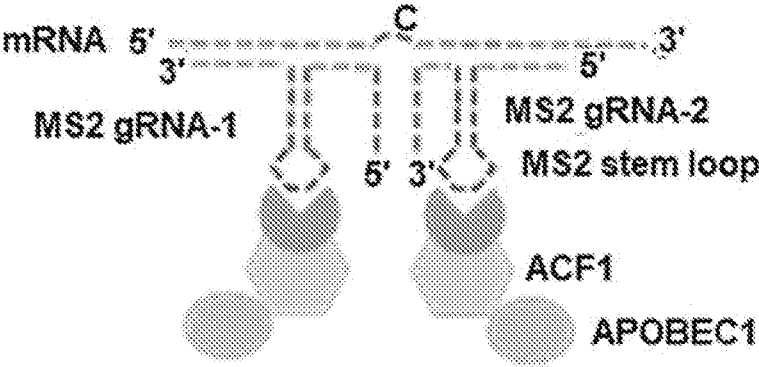


FIG. 48



FIG. 49A



FIG. 49B



FIG. 49C

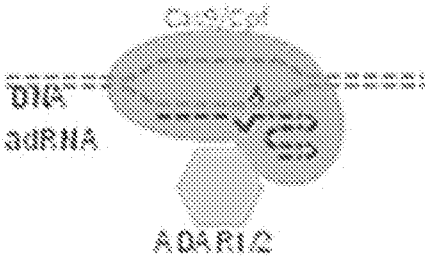
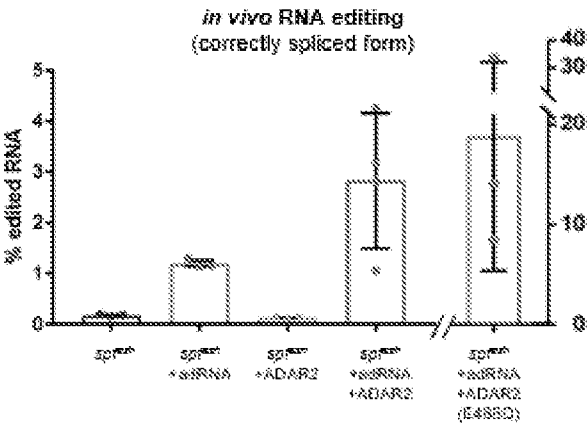
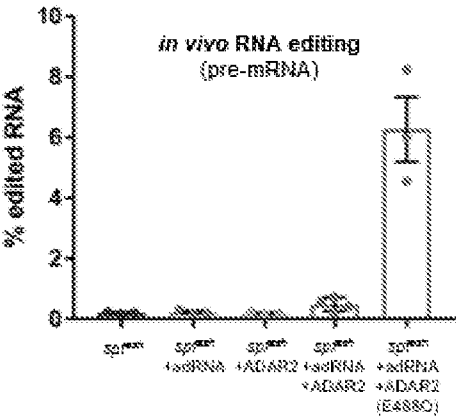


FIG. 49D

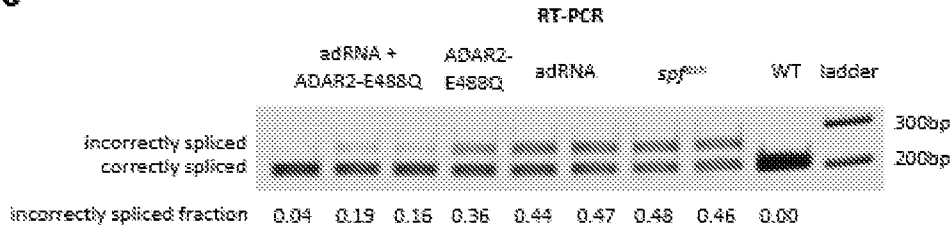
a



b



c



d

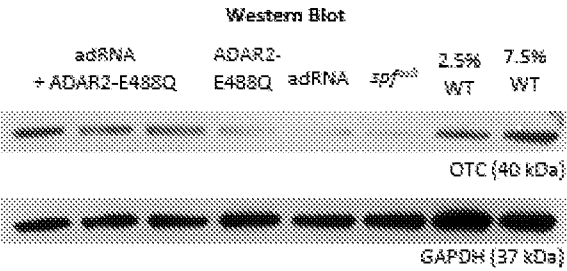


FIG. 50A-D

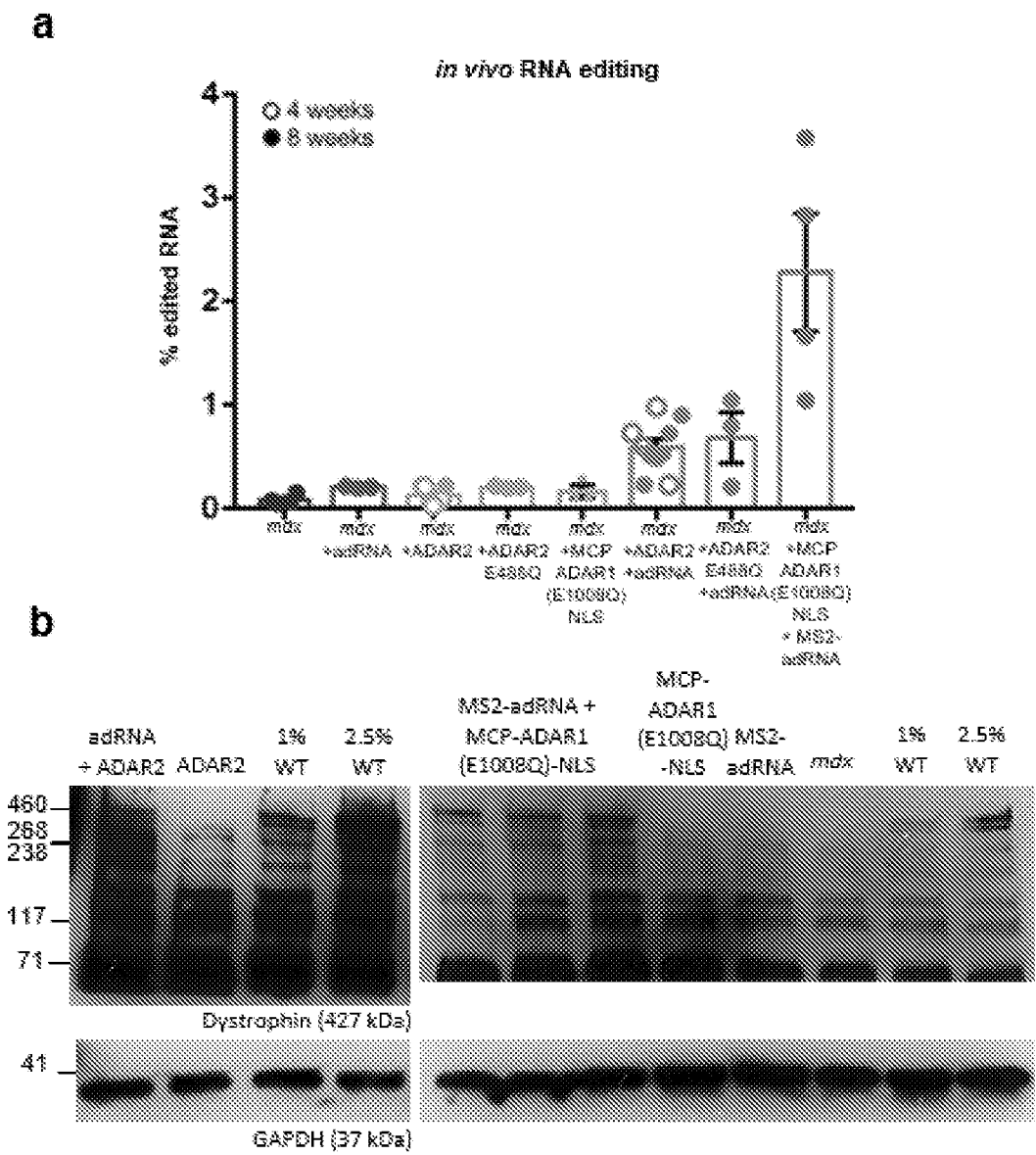


FIG. 51A-B



50 / 50

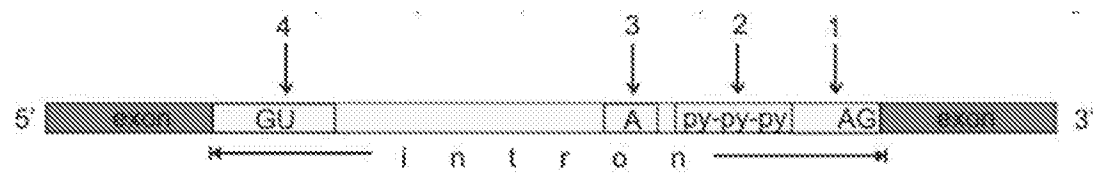


FIG. 52

## SEQUENCE LISTING

<110> THE REGENTS OF THE UNIVERSITY OF CALIFORNIA

<120> RNA AND DNA BASE EDITING VIA ENGINEERED ADAR RECRUITMENT

<130> 00015-365W01

<140> PCT/US2019/050095

<141> 2019-09-06

<150> 62/780,241

<151> 2018-12-15

<150> 62/766,433

<151> 2018-10-17

<150> 62/728,007

<151> 2018-09-06

<160> 255

<170> PatentIn version 3.5

<210> 1

<211> 323

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<220>

<221> modified\_base

<222> (117)..(146)

<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (148)..(176)

<223> a, c, t, g, unknown or other

<400> 1

ggccggggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggcggggaga  
 60

ttgcttgagc ccaggagttc gagaccagcc tgggcaacat agcgagaccc cgtctcnnnn  
 120

nnnnnnnnnn nnnnnnnnnn nnnnnncnnn nnnnnnnnnn nnnnnnnnnn nnnnnnagcc  
 180

gggcgtggtg gcgcgcgcct gtagtcccag ctactcggga ggctgaggca ggaggatcgc  
 240

ttgagcccag gagttcgagg ctgcagtgag ctatgatcgc gccactgcac tccagcctgg  
 300

g c g a c a g a g c                      g a g a c c c t g t                      c t c  
 323

<210> 2  
 <211> 66  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: Synthetic  
 oligonucleotide

<220>  
 <221> modified\_base  
 <222> (46)..(51)  
 <223> a, c, t, g, unknown or other

<220>  
 <221> modified\_base  
 <222> (53)..(66)  
 <223> a, c, t, g, unknown or other

<400> 2  
 gtggaagagg agaacaatat gctaaatgtt gttctcgtct cccacnnnnn ncnnnnnnnn  
 60

n                      n                      n                      n                      n  
 66

<210> 3  
<211> 68  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>  
<221> modified\_base  
<222> (50)..(52)  
<223> a, c, t, g, unknown or other

<220>  
<221> modified\_base  
<222> (54)..(68)  
<223> a, c, t, g, unknown or other

<400> 3  
gggtggaaga ggagaacaat atgctaaatg ttgttctcgt ctcccacctn nncnnnnnnn  
60

n            n            n            n            n            n            n  
68

<210> 4  
<211> 66  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>  
<221> modified\_base  
<222> (46)..(51)  
<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (53)..(66)

<223> a, c, t, g, unknown or other

<400> 4

ggtgaagagg agaacaatat gctaaatggt gttctcgtct ccaccnnnnn ncnnnnnnnn  
60

n                    n                    n                    n                    n                    n  
66

<210> 5

<211> 66

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>

<221> modified\_base

<222> (46)..(52)

<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (54)..(66)

<223> a, c, t, g, unknown or other

<400> 5

ggtgaagagg agaacaatat gctaaatggt gttctcgtct ccaccnnnnn nncnnnnnnn  
60

n                    n                    n                    n                    n                    n  
66

<210> 6

<211> 66

<212> DNA

<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>  
<221> modified\_base  
<222> (46)..(51)  
<223> a, c, t, g, unknown or other

<220>  
<221> modified\_base  
<222> (53)..(66)  
<223> a, c, t, g, unknown or other

<400> 6  
gtggaagagg agaacaatag gctaaacgtt gttctcgtct cccacnnnnn ncnnnnnnnn  
60

n                    n                    n                    n                    n                    n  
66

<210> 7  
<211> 68  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>  
<221> modified\_base  
<222> (50)..(52)  
<223> a, c, t, g, unknown or other

<220>  
<221> modified\_base  
<222> (54)..(68)  
<223> a, c, t, g, unknown or other

<400> 7

gggtggaaga ggagaacaat aggctaaacg ttgttctcgt ctcccacctn nncnnnnnnn  
60

n n n n n n n n  
68

<210> 8

<211> 66

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>

<221> modified\_base

<222> (46)..(51)

<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (53)..(66)

<223> a, c, t, g, unknown or other

<400> 8

ggtgaagagg agaacaatag gctaaacggtt gttctcgtct ccaccnnnnn ncnnnnnnnn  
60

n n n n n n n  
66

<210> 9

<211> 66

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>  
<221> modified\_base  
<222> (46)..(52)  
<223> a, c, t, g, unknown or other

<220>  
<221> modified\_base  
<222> (54)..(66)  
<223> a, c, t, g, unknown or other

<400> 9  
ggtgaagagg agaacaatag gctaaacgtt gttctcgtct ccaccnnnnn nncnnnnnnn  
60

n n n n n n  
66

<210> 10  
<211> 312  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 10  
ggccggggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggcgggggaga  
60

ttgcttgagc ccaggagttc gagaccagcc tgggcaacat agcgagaccc cgtctctaca  
120

aaaaatacaa aaattagccg ggcgtggtgg cgcgcgcctg tagtcccagc tactcgggag  
180

gctgaggcag gaggatcgct tgagcccagg agttcgaggc tgcagtgagc tatgatcgcg  
240

ccactgcact ccagcctggg cgacagagcg agaccctgtc tcaaaaaaaaa aaaaaaaaaa  
300

a a a a a a a a a a a a a a a a  
312



<210> 11  
<211> 393  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic polynucleotide

<220>  
<221> modified\_base  
<222> (117)..(216)  
<223> a, c, t, g, unknown or other

<400> 11  
ggccggggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggcggggaga  
60  
ttgcttgagc ccaggagttc gagaccagcc tgggcaacat agcgagaccc cgtctcnnnn  
120  
nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn  
180  
nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnagcc gggcgtggtg gcgcgcgcct  
240  
gtagtcccag ctactcggga ggctgaggca ggaggatcgc ttgagcccag gagttcgagg  
300  
ctgcagtgag ctatgatcgc gccactgcac tccagcctgg gcgacagagc gagaccctgt  
360  
c t c a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a a  
393

<210> 12  
<211> 68  
<212> RNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>

<221> modified\_base

<222> (50)..(52)

<223> a, c, u, g, unknown or other

<220>

<221> modified\_base

<222> (54)..(68)

<223> a, c, u, g, unknown or other

<400> 12

ggguggaaua guauaacaau augcuaaaug uuguuauagu aucccaccun nncnnnnnnn  
60

n            n            n            n            n            n            n  
68

<210> 13

<211> 66

<212> RNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>

<221> modified\_base

<222> (46)..(51)

<223> a, c, u, g, unknown or other

<220>

<221> modified\_base

<222> (53)..(66)

<223> a, c, u, g, unknown or other

<400> 13

guggaauagu auaacaauau gcuaaauguu guuauaguau cccacnnnnn ncnnnnnnnn  
60

n                    n                    n                    n                    n                    n  
66

<210> 14  
<211> 76  
<212> RNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
         oligonucleotide

<220>  
<221> modified\_base  
<222> (56)..(61)  
<223> a, c, u, g, unknown or other

<220>  
<221> modified\_base  
<222> (63)..(76)  
<223> a, c, u, g, unknown or other

<400> 14  
ggugucgaga auaguauaac aaauaugcuaa auguuguuau aguauccucg acaccnnnnn  
60

n c n n n n n n n n                    n n n n n n  
76

<210> 15  
<211> 117  
<212> RNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
         polynucleotide

<400> 15  
aacuuugugc auuuuagguc ucaaguggau auucauggug ucuaugaauu cacuaaaaga  
60

ugucagugcc uggugaauc auauacacca uggagaggcu auaaaaugca gaagguu  
117

<210> 16

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 16

a a c c g a g c g g  
19

t g t c t g t g a

<210> 17

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 17

a a c c g a g c g g  
19

t g t c t g t g a

<210> 18

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 18

a a c c g a g c g g  
19

t g t c t g t g a

<210> 19  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 19  
a c a a a c c g a g c g g t g t c t g t g  
21

<210> 20  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 20  
a c a a a c c g a g c g g t g t c t g t g  
21

<210> 21  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 21  
a c a a a c c g a g c g g t g t c t g t g  
21

<210> 22

<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 22

a c a a a c c g a g  
21

c g g t g t c t g t

g

<210> 23

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 23

t a c a a a c c g a  
21

g c g g t g t c t g

t

<210> 24

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 24

a c a a a c c g a g  
21

c g g t g t c t g t

g

<210> 25

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 25

t a c a a a c c g a

g c g g t g t c t g

t

21

<210> 26

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 26

t a c a a a c c g a

g c g g t g t c

18

<210> 27

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 27

t a c a a a c c g a

g c g g t g t c

18

<210> 28

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 28

t a c a a a c c g a  
18

g c g g t g t c

<210> 29

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 29

c a a a c c g a g c  
20

g g t g t c t g t g

<210> 30

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 30

t a c a a a c c g a  
20

g c g g t g t c t g

<210> 31

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 31



t t t a c a a a c c  
20

g a g c g g t g t c

<210> 32  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 32  
a c a a a c c g a g c g g t g t c t g t g  
21

<210> 33  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 33  
a c a a a c c g a g c g g t g t c t g t g  
21

<210> 34  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 34  
a c a a a c c g a g c g g t g t c t g t g  
21

<210> 35  
<211> 120  
<212> RNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic polynucleotide

<220>  
<221> modified\_base  
<222> (46)..(59)  
<223> a, c, u, g, unknown or other

<220>  
<221> modified\_base  
<222> (61)..(75)  
<223> a, c, u, g, unknown or other

<400> 35  
guggaauagu auaacaauau gcuaaauguu guuauaguau cccacnnnnn nnnnnnnnnc  
60  
  
nnnnnnnnnn nnnnncaccc uaugauauug uuguaaaucg uauaacaaua ugauaaggug  
120

<210> 36  
<211> 30  
<212> RNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<220>  
<221> modified\_base  
<222> (1)..(14)  
<223> a, c, u, g, unknown or other

<220>

<221> modified\_base  
<222> (16)..(30)  
<223> a, c, u, g, unknown or other

<400> 36  
n n n n n n n n n n n n n n c n n n n n n n n n n n n n n n n n n n n  
30

<210> 37  
<211> 66  
<212> RNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>  
<221> modified\_base  
<222> (46)..(51)  
<223> a, c, u, g, unknown or other

<220>  
<221> modified\_base  
<222> (53)..(66)  
<223> a, c, u, g, unknown or other

<400> 37  
guggaauagu auaacaauau gcuaaauguu guuauaguau cccacnnnnn ncnnnnnnnn  
60

n n n n n n n  
66

<210> 38  
<211> 283  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 38

ggccggggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggcggggaga  
60

ttgcttgagc ccaggagttc gagaccagcc tgggcaacat agcgagaccc cgtctcatac  
120

tgccgccagc tggattagcc gggcgtggtg gcgcgcgcct gtagtcccag ctactcggga  
180

ggctgaggca ggaggatcgc ttgagcccag gagttcgagg ctgcagtgag ctatgatcgc  
240

gccactgcac tccagcctgg gcgacagagc gagaccctgt ctc  
283

<210> 39

<211> 303

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 39

ggccggggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggcggggaga  
60

ttgcttgagc ccaggagttc gagaccagcc tgggcaacat agcgagaccc cgtctcactg  
120

tacagaatac tgccgccagc tggatttccc aattctagcc gggcgtggtg gcgcgcgcct  
180

gtagtcccag ctactcggga ggctgaggca ggaggatcgc ttgagcccag gagttcgagg  
240

ctgcagtgag ctatgatcgc gccactgcac tccagcctgg gcgacagagc gagaccctgt  
300

c t c  
303

<210> 40  
<211> 322  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 40  
ggccggggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggcggggaga  
60  
ttgcttgagc ccaggagttc gagaccagcc tgggcaacat agcgagaccc cgtctctctt  
120  
gtgtctactg tacagaatac tgccgccagc tggatttccc aattctgagt aacacagccg  
180  
ggcgtggtgg cgcgcgcctg tagtcccagc tactcgggag gctgaggcag gaggatcgct  
240  
tgagcccagg agttcgaggc tgcagtgagc tatgatcgcg ccactgcact ccagcctggg  
300  
c g a c a g a g c g a g a c c c t g t c t c  
322

<210> 41  
<211> 363  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 41  
ggccggggcgc ggtggctcac gcctgtaatc ccagcacttt gggaggccga ggcggggaga  
60  
ttgcttgagc ccaggagttc gagaccagcc tgggcaacat agcgagaccc cgtctctgat  
120

aaaaggcgta cataattctt gtgtctactg tacagaatac tgccgccagc tggatttccc  
180

aattctgagt aacactctgc aatccaaaca gggttcagcc gggcgtgggtg gcgcgcgcct  
240

gtagtcccag ctactcggga ggctgaggca ggaggatcgc ttgagcccag gagttcgagg  
300

ctgcagtgag ctatgatcgc gccactgcac tccagcctgg gcgacagagc gagaccctgt  
360

c t c  
363

<210> 42

<211> 386

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 42

Gln Leu His Leu Pro Gln Val Leu Ala Asp Ala Val Ser Arg Leu Val  
1 5 10 15

Leu Gly Lys Phe Gly Asp Leu Thr Asp Asn Phe Ser Ser Pro His Ala  
20 25 30

Arg Arg Lys Val Leu Ala Gly Val Val Met Thr Thr Gly Thr Asp Val  
35 40 45

Lys Asp Ala Lys Val Ile Ser Val Ser Thr Gly Thr Lys Cys Ile Asn  
50 55 60

Gly Glu Tyr Met Ser Asp Arg Gly Leu Ala Leu Asn Asp Cys His Ala  
65 70 75 80

Glu Ile Ile Ser Arg Arg Ser Leu Leu Arg Phe Leu Tyr Thr Gln Leu  
85 90 95

Glu Leu Tyr Leu Asn Asn Lys Asp Asp Gln Lys Arg Ser Ile Phe Gln  
100 105 110

Lys Ser Glu Arg Gly Gly Phe Arg Leu Lys Glu Asn Val Gln Phe His  
115 120 125

Leu Tyr Ile Ser Thr Ser Pro Cys Gly Asp Ala Arg Ile Phe Ser Pro  
130 135 140

His Glu Pro Ile Ile Glu Glu Pro Ala Asp Arg His Pro Asn Arg Lys  
145 150 155 160

Ala Arg Gly Gln Leu Arg Thr Lys Ile Glu Ser Gly Glu Gly Thr Ile  
165 170 175

Pro Val Arg Ser Asn Ala Ser Ile Gln Thr Trp Asp Gly Val Leu Gln  
180 185 190

Gly Glu Arg Leu Leu Thr Met Ser Cys Ser Asp Lys Ile Ala Arg Trp  
195 200 205

Asn Val Val Gly Ile Gln Gly Ser Leu Leu Ser Ile Phe Val Glu Pro  
210 215 220

Ile Tyr Phe Ser Ser Ile Ile Leu Gly Ser Leu Tyr His Gly Asp His  
225 230 235 240

Leu Ser Arg Ala Met Tyr Gln Arg Ile Ser Asn Ile Glu Asp Leu Pro  
245 250 255

Pro Leu Tyr Thr Leu Asn Lys Pro Leu Leu Ser Gly Ile Ser Asn Ala  
260 265 270

Glu Ala Arg Gln Pro Gly Lys Ala Pro Asn Phe Ser Val Asn Trp Thr  
275 280 285

Val Gly Asp Ser Ala Ile Glu Val Ile Asn Ala Thr Thr Gly Lys Asp  
290 295 300

Glu Leu Gly Arg Ala Ser Arg Leu Cys Lys His Ala Leu Tyr Cys Arg  
305 310 315 320

Trp Met Arg Val His Gly Lys Val Pro Ser His Leu Leu Arg Ser Lys  
325 330 335

Ile Thr Lys Pro Asn Val Tyr His Glu Ser Lys Leu Ala Ala Lys Glu  
340 345 350

Tyr Gln Ala Ala Lys Ala Arg Leu Phe Thr Ala Phe Ile Lys Ala Gly  
355 360 365

Leu Gly Ala Trp Val Glu Lys Pro Thr Glu Gln Asp Gln Phe Ser Leu  
370 375 380

Thr Pro  
385

<210> 43

<211> 1662

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide



<400> 43

atgaagacct taattcttgc cgttgcatta gtctactgcg ccactgttca ttgccaggac  
60

tgtccttacg aacctgatcc accaaacaca gttccaactt cctgtgaagc taaagaagga  
120

gaatgtattg atagcagctg tggcacctgc acgagagaca tactatcaga tggactgtgt  
180

gaaaataaac caggaaaaac atgttgccga atgtgtcagt atgtaattga atgcagagta  
240

gaggccgcag gatagtttag aacattctat ggaaagagat tccagttcca ggaacctggt  
300

acatacgtgt tgggtcaagg aaccaagggc ggcgactgga aggtgtccat caccctggag  
360

aacctggatg gaaccaaggg ggctgtgctg accaagacaa gactggaagt ggctggagac  
420

atcattgaca tcgctcaagc tactgagaat cccatcactg taaacggtgg agctgaccct  
480

atcatcgcca acccgtagac catcggcgag gtcaccatcg ctgttggtga gatgccaggc  
540

ttcaacatca ccgtcattga gttcttcaaa ctgatcgtga tcgacatcct cggaggaaga  
600

tctgtaagaa tcgccccaga cacagcaaac aaaggaatga tctctggcct ctgtggagat  
660

cttaaaatga tggaagatac agacttcact tcagatccag aacaactcgc tattcagcct  
720

aagatcaacc aggagtttga cggttgtcca ctctatggaa atcctgatga cgttgcatac  
780

tgcaaaggtc ttctggagcc gtacaaggac agctgccgca accccatcaa cttctactac  
840

tacaccatct cctgcgcctt cgcccgtgt atgggtggag acgagcgagc ctcacacgtg  
900

ctgcttgact acagggagac gtgcgctgct cccgaaacta gaggaacctg cgttttgtct  
960

ggacatactt tctacgatac atttgacaaa gcaagatacc aattccaggg tccctgcaag  
1020

gagattctta tggccgccga ctgtttctgg aacacttggg atgtgaaggt ttcacacagg  
1080

aatgttgact cttacactga agtagagaaa gtacgaatca ggaaacaatc gactgtagta  
1140

gaactcattg ttgatggaaa acagattctg gttggaggag aagccgtgtc cgtcccgtac  
1200

agctctcaga acacttccat ctactggcaa gatggtgaca tactgactac agccatccta  
1260

cctgaagctc tgggtggtcaa gttcaacttc aagcaactgc tcgtcgtaca tattagagat  
1320

ccattcgatg gtaagacttg cggtatattgc ggtaactaca accaggattt cagtgatgat  
1380

tcttttgatg ctgaaggagc ctgtgatctg accccaacc caccgggatg caccgaagaa  
1440

cagaaacctg aagctgaacg actctgcaat agtctcttcg ccggtcaaag tgatcttgat  
1500

cagaaatgta acgtgtgccca caagcctgac cgtgtcgaac gatgcatgta cgagtattgc  
1560

ctgaggggac aacagggttt ctgtgaccac gcatgggagt tcaagaaaga atgctacata  
1620

aagcatggag acaccctaga agtaccagat gaatgcaaat aa  
1662

<210> 44

<211> 61

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 44

ggccctgaaa aagggcctgt tctaaacat cctgcggcct caacatgagg atcacccatg  
60

c  
61

<210> 45

<211> 64

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 45

aacatgagga tcacccatgt cctaaacat cctgcggcct ctactctggc cctgaaaaag  
60

g g c c  
64

<210> 46

<211> 62

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 46

aacatgagga tcacccatgt cctaaacat cctgcggcct caacatgagg atcacccatg  
60

t c  
62

<210> 47  
<211> 64  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 47  
ggccctgaaa aagggcctgt tctaaaccat cctgcggcct ctactctggc cctgaaaaag  
60

g g c c  
64

<210> 48  
<211> 317  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 48  
Met Leu Pro Pro Leu Glu Arg Leu Thr Leu Gly Ser Asp Tyr Lys Asp  
1 5 10 15

Asp Asp Asp Lys Gly Ser Gly Ser Gly Ser Met Ala Ser Asn Phe Thr  
20 25 30

Gln Phe Val Leu Val Asp Asn Gly Gly Thr Gly Asp Val Thr Val Ala  
35 40 45

Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu Trp Ile Ser Ser Asn Ser  
50 55 60

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | Ser | Gln | Ala | Tyr | Lys | Val | Thr | Cys | Ser | Val | Arg | Gln | Ser | Ser | Ala | 65  | 70  | 75  | 80  |
| Gln | Asn | Arg | Lys | Tyr | Thr | Ile | Lys | Val | Glu | Val | Pro | Lys | Gly | Ala | Trp | 85  | 90  | 95  |     |
| Arg | Ser | Tyr | Leu | Asn | Met | Glu | Leu | Thr | Ile | Pro | Ile | Phe | Ala | Thr | Asn | 100 | 105 | 110 |     |
| Ser | Asp | Cys | Glu | Leu | Ile | Val | Lys | Ala | Met | Gln | Gly | Leu | Leu | Lys | Asp | 115 | 120 | 125 |     |
| Gly | Asn | Pro | Ile | Pro | Ser | Ala | Ile | Ala | Ala | Asn | Ser | Gly | Ile | Tyr | Gly | 130 | 135 | 140 |     |
| Gly | Ser | Gly | Ser | Gly | Ala | Gly | Ser | Gly | Ser | Pro | Ala | Gly | Gly | Gly | Ala | 145 | 150 | 155 | 160 |
| Pro | Gly | Ser | Gly | Gly | Gly | Ser | Gln | Leu | His | Leu | Pro | Gln | Val | Leu | Ala | 165 | 170 | 175 |     |
| Asp | Ala | Val | Ser | Arg | Leu | Val | Leu | Gly | Lys | Phe | Gly | Asp | Leu | Thr | Asp | 180 | 185 | 190 |     |
| Asn | Phe | Ser | Ser | Pro | His | Ala | Arg | Arg | Lys | Val | Leu | Ala | Gly | Val | Val | 195 | 200 | 205 |     |
| Met | Thr | Thr | Gly | Thr | Asp | Val | Lys | Asp | Ala | Lys | Val | Ile | Ser | Val | Ser | 210 | 215 | 220 |     |
| Thr | Gly | Thr | Lys | Cys | Ile | Asn | Gly | Glu | Tyr | Met | Ser | Asp | Arg | Gly | Leu | 225 | 230 | 235 | 240 |
| Ala | Leu | Asn | Asp | Cys | His | Ala | Glu | Ile | Ile | Ser | Arg | Arg | Ser | Leu | Leu |     |     |     |     |

255

Thr Met Ser Cys Ser Asp Lys Ile Ala Arg Trp Asn Val Val Gly Ile  
50 55 60

Gln Gly Ser Leu Leu Ser Ile Phe Val Glu Pro Ile Tyr Phe Ser Ser  
65 70 75 80

Ile Ile Leu Gly Ser Leu Tyr His Gly Asp His Leu Ser Arg Ala Met  
85 90 95

Tyr Gln Arg Ile Ser Asn Ile Glu Asp Leu Pro Pro Leu Tyr Thr Leu  
100 105 110

Asn Lys Pro Leu Leu Ser Gly Ile Ser Asn Ala Glu Ala Arg Gln Pro  
115 120 125

Gly Lys Ala Pro Asn Phe Ser Val Asn Trp Thr Val Gly Asp Ser Ala  
130 135 140

Ile Glu Val Ile Asn Ala Thr Thr Gly Lys Asp Glu Leu Gly Arg Ala  
145 150 155 160

Ser Arg Leu Cys Lys His Ala Leu Tyr Cys Arg Trp Met Arg Val His  
165 170 175

Gly Lys Val Pro Ser His Leu Leu Arg Ser Lys Ile Thr Lys Pro Asn  
180 185 190

Val Tyr His Glu Ser Lys Leu Ala Ala Lys Glu Tyr Gln Ala Ala Lys  
195 200 205

Ala Arg Leu Phe Thr Ala Phe Ile Lys Ala Gly Leu Gly Ala Trp Val  
210 215 220

Glu Lys Pro Thr Glu Gln Asp Gln Phe Ser Leu Thr Pro Gly Gly Ser  
225 230 235 240

Gly Ser Gly Ala Gly Ser Gly Ser Pro Ala Gly Gly Gly Ala Pro Gly

245

250

255

Ser Gly Gly Gly Ser Asn Ala Arg Thr Arg Arg Arg Glu Arg Arg Ala  
260 265 270

Glu Lys Gln Ala Gln Trp Lys Ala Ala Asn Gly Gly Gly Gly Ser Gly  
275 280 285

Gly Gly Gly Ser Gly Gly Gly Gly Ser Asn Ala Arg Thr Arg Arg Arg  
290 295 300

Glu Arg Arg Ala Glu Lys Gln Ala Gln Trp Lys Ala Ala Asn Gly Gly  
305 310 315 320

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asn Ala Arg  
325 330 335

Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala Gln Trp Lys Ala  
340 345 350

Ala Asn Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly  
355 360 365

Ser Asn Ala Arg Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala  
370 375 380

Gln Trp Lys Ala Ala Asn Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr  
385 390 395 400

Ala Gly Ser Leu Pro Pro Leu Glu Arg Leu Thr Leu  
405 410

&lt;210&gt; 50

&lt;211&gt; 320



<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 50

Met Leu Pro Pro Leu Glu Arg Leu Thr Leu Gly Ser Asp Tyr Lys Asp  
1 5 10 15

Asp Asp Asp Lys Gly Ser Gly Ser Gly Ser Met Ala Ser Asn Phe Thr  
20 25 30

Gln Phe Val Leu Val Asp Asn Gly Gly Thr Gly Asp Val Thr Val Ala  
35 40 45

Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu Trp Ile Ser Ser Asn Ser  
50 55 60

Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser Val Arg Gln Ser Ser Ala  
65 70 75 80

Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu Val Pro Lys Gly Ala Trp  
85 90 95

Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe Ala Thr Asn  
100 105 110

Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu Leu Lys Asp  
115 120 125

Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly Ile Tyr Gly  
130 135 140

Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser Pro Ala Gly Gly Gly Ala

|                                                                 |     |     |  |     |  |     |
|-----------------------------------------------------------------|-----|-----|--|-----|--|-----|
| 145                                                             |     | 150 |  | 155 |  | 160 |
| Pro Gly Ser Gly Gly Gly Ser Gln Leu His Leu Pro Gln Val Leu Ala |     |     |  |     |  |     |
|                                                                 | 165 |     |  | 170 |  | 175 |
| Asp Ala Val Ser Arg Leu Val Leu Gly Lys Phe Gly Asp Leu Thr Asp |     |     |  |     |  |     |
|                                                                 | 180 |     |  | 185 |  | 190 |
| Asn Phe Ser Ser Pro His Ala Arg Arg Lys Val Leu Ala Gly Val Val |     |     |  |     |  |     |
|                                                                 | 195 |     |  | 200 |  | 205 |
| Met Thr Thr Gly Thr Asp Val Lys Asp Ala Lys Val Ile Ser Val Ser |     |     |  |     |  |     |
|                                                                 | 210 |     |  | 215 |  | 220 |
| Thr Gly Thr Lys Cys Ile Asn Gly Glu Tyr Met Ser Asp Arg Gly Leu |     |     |  |     |  |     |
|                                                                 | 225 |     |  | 230 |  | 240 |
| Ala Leu Asn Asp Cys His Ala Glu Ile Ile Ser Arg Arg Ser Leu Leu |     |     |  |     |  |     |
|                                                                 | 245 |     |  | 250 |  | 255 |
| Arg Phe Leu Tyr Thr Gln Leu Glu Leu Tyr Leu Asn Asn Lys Asp Asp |     |     |  |     |  |     |
|                                                                 | 260 |     |  | 265 |  | 270 |
| Gln Lys Arg Ser Ile Phe Gln Lys Ser Glu Arg Gly Gly Phe Arg Leu |     |     |  |     |  |     |
|                                                                 | 275 |     |  | 280 |  | 285 |
| Lys Glu Asn Val Gln Phe His Leu Tyr Ile Ser Thr Ser Pro Cys Gly |     |     |  |     |  |     |
|                                                                 | 290 |     |  | 295 |  | 300 |
| Asp Ala Arg Ile Phe Ser Pro His Glu Pro Ile Leu Glu Glu Pro Ala |     |     |  |     |  |     |
|                                                                 | 305 |     |  | 310 |  | 320 |

<210> 51  
 <211> 409

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 51

Met Asp Arg His Pro Asn Arg Lys Ala Arg Gly Gln Leu Arg Thr Lys  
1 5 10 15

Ile Glu Ser Gly Glu Gly Thr Ile Pro Val Arg Ser Asn Ala Ser Ile  
20 25 30

Gln Thr Trp Asp Gly Val Leu Gln Gly Glu Arg Leu Leu Thr Met Ser  
35 40 45

Cys Ser Asp Lys Ile Ala Arg Trp Asn Val Val Gly Ile Gln Gly Ser  
50 55 60

Leu Leu Ser Ile Phe Val Glu Pro Ile Tyr Phe Ser Ser Ile Ile Leu  
65 70 75 80

Gly Ser Leu Tyr His Gly Asp His Leu Ser Arg Ala Met Tyr Gln Arg  
85 90 95

Ile Ser Asn Ile Glu Asp Leu Pro Pro Leu Tyr Thr Leu Asn Lys Pro  
100 105 110

Leu Leu Ser Gly Ile Ser Asn Ala Glu Ala Arg Gln Pro Gly Lys Ala  
115 120 125

Pro Asn Phe Ser Val Asn Trp Thr Val Gly Asp Ser Ala Ile Glu Val  
130 135 140

Ile Asn Ala Thr Thr Gly Lys Asp Glu Leu Gly Arg Ala Ser Arg Leu

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 145 |     | 150 |     | 155 |     | 160 |     |     |     |     |     |     |     |     |     |
| Cys | Lys | His | Ala | Leu | Tyr | Cys | Arg | Trp | Met | Arg | Val | His | Gly | Lys | Val |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Pro | Ser | His | Leu | Leu | Arg | Ser | Lys | Ile | Thr | Lys | Pro | Asn | Val | Tyr | His |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Glu | Ser | Lys | Leu | Ala | Ala | Lys | Glu | Tyr | Gln | Ala | Ala | Lys | Ala | Arg | Leu |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Phe | Thr | Ala | Phe | Ile | Lys | Ala | Gly | Leu | Gly | Ala | Trp | Val | Glu | Lys | Pro |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Thr | Glu | Gln | Asp | Gln | Phe | Ser | Leu | Thr | Pro | Gly | Gly | Ser | Gly | Ser | Gly |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Ala | Gly | Ser | Gly | Ser | Pro | Ala | Gly | Gly | Gly | Ala | Pro | Gly | Ser | Gly | Gly |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Gly | Ser | Asn | Ala | Arg | Thr | Arg | Arg | Arg | Glu | Arg | Arg | Ala | Glu | Lys | Gln |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Ala | Gln | Trp | Lys | Ala | Ala | Asn | Gly | Gly | Gly | Gly | Ser | Gly | Gly | Gly | Gly |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Ser | Gly | Gly | Gly | Gly | Ser | Asn | Ala | Arg | Thr | Arg | Arg | Arg | Glu | Arg | Arg |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Ala | Glu | Lys | Gln | Ala | Gln | Trp | Lys | Ala | Ala | Asn | Gly | Gly | Gly | Gly | Ser |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Gly | Gly | Gly | Gly | Ser | Gly | Gly | Gly | Gly | Ser | Asn | Ala | Arg | Thr | Arg | Arg |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |

Arg Glu Arg Arg Ala Glu Lys Gln Ala Gln Trp Lys Ala Ala Asn Gly  
340 345 350

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asn Ala  
355 360 365

Arg Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala Gln Trp Lys  
370 375 380

Ala Ala Asn Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Gly Ser  
385 390 395 400

Leu Pro Pro Leu Glu Arg Leu Thr Leu  
405

<210> 52

<211> 317

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 52

Met Leu Pro Pro Leu Glu Arg Leu Thr Leu Gly Ser Asp Tyr Lys Asp  
1 5 10 15

Asp Asp Asp Lys Gly Ser Gly Ser Gly Ser Met Ala Ser Asn Phe Thr  
20 25 30

Gln Phe Val Leu Val Asp Asn Gly Gly Thr Gly Asp Val Thr Val Ala  
35 40 45

Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu Trp Ile Ser Ser Asn Ser

50

55

60

Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser Val Arg Gln Ser Ser Ala  
65 70 75 80

Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu Val Pro Lys Gly Ala Trp  
85 90 95

Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe Ala Thr Asn  
100 105 110

Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu Leu Lys Asp  
115 120 125

Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly Ile Tyr Gly  
130 135 140

Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser Pro Ala Gly Gly Gly Ala  
145 150 155 160

Pro Gly Ser Gly Gly Gly Ser Gln Leu His Leu Pro Gln Val Leu Ala  
165 170 175

Asp Ala Val Ser Arg Leu Val Ile Gly Lys Phe Gly Asp Leu Thr Asp  
180 185 190

Asn Phe Ser Ser Pro His Ala Arg Arg Ile Gly Leu Ala Gly Val Val  
195 200 205

Met Thr Thr Gly Thr Asp Val Lys Asp Ala Lys Val Ile Cys Val Ser  
210 215 220

Thr Gly Ser Lys Cys Ile Asn Gly Glu Tyr Leu Ser Asp Arg Gly Leu  
225 230 235 240

Ala Leu Asn Asp Cys His Ala Glu Ile Val Ser Arg Arg Ser Leu Leu  
245 250 255

Arg Phe Leu Tyr Thr Gln Leu Glu Leu Tyr Leu Asn Asn Glu Asp Asp  
260 265 270

Gln Lys Arg Ser Ile Phe Gln Lys Ser Glu Arg Gly Gly Phe Arg Leu  
275 280 285

Lys Glu Asn Ile Gln Phe His Leu Tyr Ile Ser Thr Ser Pro Cys Gly  
290 295 300

Asp Ala Arg Ile Phe Ser Pro His Glu Ala Ile Leu Glu  
305 310 315

<210> 53

<211> 412

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 53

Met Glu Pro Ala Asp Arg His Pro Asn Arg Lys Ala Arg Gly Gln Leu  
1 5 10 15

Arg Thr Lys Ile Glu Ala Gly Gln Gly Thr Ile Pro Val Arg Asn Asn  
20 25 30

Ala Ser Ile Gln Thr Trp Asp Gly Val Leu Gln Gly Glu Arg Leu Leu  
35 40 45

Thr Met Ser Cys Ser Asp Lys Ile Ala Arg Trp Asn Val Val Gly Ile

50

55

60

Gln Gly Ser Leu Leu Ser Ile Phe Val Glu Pro Ile Tyr Phe Ser Ser  
65 70 75 80

Ile Ile Leu Gly Ser Leu Tyr His Gly Asp His Leu Ser Arg Ala Met  
85 90 95

Tyr Gln Arg Ile Ser Asn Ile Glu Asp Leu Pro Pro Leu Tyr Thr Leu  
100 105 110

Asn Lys Pro Leu Leu Thr Gly Ile Ser Asn Ala Glu Ala Arg Gln Pro  
115 120 125

Gly Lys Ala Pro Ile Phe Ser Val Asn Trp Thr Val Gly Asp Ser Ala  
130 135 140

Ile Glu Val Ile Asn Ala Thr Thr Gly Lys Gly Glu Leu Gly Arg Ala  
145 150 155 160

Ser Arg Leu Cys Lys His Ala Leu Tyr Cys Arg Trp Met Arg Val His  
165 170 175

Gly Lys Val Pro Ser His Leu Leu Arg Ser Lys Ile Thr Lys Pro Asn  
180 185 190

Val Tyr His Glu Thr Lys Leu Ala Ala Lys Glu Tyr Gln Ala Ala Lys  
195 200 205

Ala Arg Leu Phe Thr Ala Phe Ile Lys Ala Gly Leu Gly Ala Trp Val  
210 215 220

Glu Lys Pro Thr Glu Gln Asp Gln Phe Ser Leu Thr Pro Gly Gly Ser  
225 230 235 240



Gly Ser Gly Ala Gly Ser Gly Ser Pro Ala Gly Gly Gly Ala Pro Gly  
245 250 255

Ser Gly Gly Gly Ser Asn Ala Arg Thr Arg Arg Arg Glu Arg Arg Ala  
260 265 270

Glu Lys Gln Ala Gln Trp Lys Ala Ala Asn Gly Gly Gly Gly Ser Gly  
275 280 285

Gly Gly Gly Ser Gly Gly Gly Gly Ser Asn Ala Arg Thr Arg Arg Arg  
290 295 300

Glu Arg Arg Ala Glu Lys Gln Ala Gln Trp Lys Ala Ala Asn Gly Gly  
305 310 315 320

Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asn Ala Arg  
325 330 335

Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala Gln Trp Lys Ala  
340 345 350

Ala Asn Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly  
355 360 365

Ser Asn Ala Arg Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala  
370 375 380

Gln Trp Lys Ala Ala Asn Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr  
385 390 395 400

Ala Gly Ser Leu Pro Pro Leu Glu Arg Leu Thr Leu  
405 410

<210> 54  
<211> 320  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 54  
Met Leu Pro Pro Leu Glu Arg Leu Thr Leu Gly Ser Asp Tyr Lys Asp  
1 5 10 15  
  
Asp Asp Asp Lys Gly Ser Gly Ser Gly Ser Met Ala Ser Asn Phe Thr  
20 25 30  
  
Gln Phe Val Leu Val Asp Asn Gly Gly Thr Gly Asp Val Thr Val Ala  
35 40 45  
  
Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu Trp Ile Ser Ser Asn Ser  
50 55 60  
  
Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser Val Arg Gln Ser Ser Ala  
65 70 75 80  
  
Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu Val Pro Lys Gly Ala Trp  
85 90 95  
  
Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe Ala Thr Asn  
100 105 110  
  
Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu Leu Lys Asp  
115 120 125  
  
Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly Ile Tyr Gly  
130 135 140

Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser Pro Ala Gly Gly Gly Ala  
145 150 155 160

Pro Gly Ser Gly Gly Gly Ser Gln Leu His Leu Pro Gln Val Leu Ala  
165 170 175

Asp Ala Val Ser Arg Leu Val Ile Gly Lys Phe Gly Asp Leu Thr Asp  
180 185 190

Asn Phe Ser Ser Pro His Ala Arg Arg Ile Gly Leu Ala Gly Val Val  
195 200 205

Met Thr Thr Gly Thr Asp Val Lys Asp Ala Lys Val Ile Cys Val Ser  
210 215 220

Thr Gly Ser Lys Cys Ile Asn Gly Glu Tyr Leu Ser Asp Arg Gly Leu  
225 230 235 240

Ala Leu Asn Asp Cys His Ala Glu Ile Val Ser Arg Arg Ser Leu Leu  
245 250 255

Arg Phe Leu Tyr Thr Gln Leu Glu Leu Tyr Leu Asn Asn Glu Asp Asp  
260 265 270

Gln Lys Arg Ser Ile Phe Gln Lys Ser Glu Arg Gly Gly Phe Arg Leu  
275 280 285

Lys Glu Asn Ile Gln Phe His Leu Tyr Ile Ser Thr Ser Pro Cys Gly  
290 295 300

Asp Ala Arg Ile Phe Ser Pro His Glu Ala Ile Leu Glu Glu Pro Ala  
305 310 315 320

<210> 55  
<211> 409  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 55  
Met Asp Arg His Pro Asn Arg Lys Ala Arg Gly Gln Leu Arg Thr Lys  
1 5 10 15

Ile Glu Ala Gly Gln Gly Thr Ile Pro Val Arg Asn Asn Ala Ser Ile  
20 25 30

Gln Thr Trp Asp Gly Val Leu Gln Gly Glu Arg Leu Leu Thr Met Ser  
35 40 45

Cys Ser Asp Lys Ile Ala Arg Trp Asn Val Val Gly Ile Gln Gly Ser  
50 55 60

Leu Leu Ser Ile Phe Val Glu Pro Ile Tyr Phe Ser Ser Ile Ile Leu  
65 70 75 80

Gly Ser Leu Tyr His Gly Asp His Leu Ser Arg Ala Met Tyr Gln Arg  
85 90 95

Ile Ser Asn Ile Glu Asp Leu Pro Pro Leu Tyr Thr Leu Asn Lys Pro  
100 105 110

Leu Leu Thr Gly Ile Ser Asn Ala Glu Ala Arg Gln Pro Gly Lys Ala  
115 120 125

Pro Ile Phe Ser Val Asn Trp Thr Val Gly Asp Ser Ala Ile Glu Val  
130 135 140

Ile Asn Ala Thr Thr Gly Lys Gly Glu Leu Gly Arg Ala Ser Arg Leu  
145 150 155 160

Cys Lys His Ala Leu Tyr Cys Arg Trp Met Arg Val His Gly Lys Val  
165 170 175

Pro Ser His Leu Leu Arg Ser Lys Ile Thr Lys Pro Asn Val Tyr His  
180 185 190

Glu Thr Lys Leu Ala Ala Lys Glu Tyr Gln Ala Ala Lys Ala Arg Leu  
195 200 205

Phe Thr Ala Phe Ile Lys Ala Gly Leu Gly Ala Trp Val Glu Lys Pro  
210 215 220

Thr Glu Gln Asp Gln Phe Ser Leu Thr Pro Gly Gly Ser Gly Ser Gly  
225 230 235 240

Ala Gly Ser Gly Ser Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly  
245 250 255

Gly Ser Asn Ala Arg Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln  
260 265 270

Ala Gln Trp Lys Ala Ala Asn Gly Gly Gly Gly Ser Gly Gly Gly Gly  
275 280 285

Ser Gly Gly Gly Gly Ser Asn Ala Arg Thr Arg Arg Arg Glu Arg Arg  
290 295 300

Ala Glu Lys Gln Ala Gln Trp Lys Ala Ala Asn Gly Gly Gly Gly Ser  
305 310 315 320

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asn Ala Arg Thr Arg Arg  
325 330 335

Arg Glu Arg Arg Ala Glu Lys Gln Ala Gln Trp Lys Ala Ala Asn Gly  
340 345 350

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Asn Ala  
355 360 365

Arg Thr Arg Arg Arg Glu Arg Arg Ala Glu Lys Gln Ala Gln Trp Lys  
370 375 380

Ala Ala Asn Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Gly Ser  
385 390 395 400

Leu Pro Pro Leu Glu Arg Leu Thr Leu  
405

<210> 56

<211> 11

<212> RNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 56

a u u g c a c u c c  
11

g

<210> 57

<211> 393

<212> PRT

<213> Artificial Sequence

<220>

**<223> Description of Artificial Sequence: Synthetic polypeptide**

**<400> 57**

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Ser | Asn | Phe | Thr | Gln | Phe | Val | Leu | Val | Asp | Asn | Gly | Gly | Thr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Asp | Val | Thr | Val | Ala | Pro | Ser | Asn | Phe | Ala | Asn | Gly | Ile | Ala | Glu |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Trp | Ile | Ser | Ser | Asn | Ser | Arg | Ser | Gln | Ala | Tyr | Lys | Val | Thr | Cys | Ser |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Arg | Gln | Ser | Ser | Ala | Gln | Asn | Arg | Lys | Tyr | Thr | Ile | Lys | Val | Glu |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Pro | Lys | Gly | Ala | Trp | Arg | Ser | Tyr | Leu | Asn | Met | Glu | Leu | Thr | Ile |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Pro | Ile | Phe | Ala | Thr | Asn | Ser | Asp | Cys | Glu | Leu | Ile | Val | Lys | Ala | Met |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gln | Gly | Leu | Leu | Lys | Asp | Gly | Asn | Pro | Ile | Pro | Ser | Ala | Ile | Ala | Ala |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asn | Ser | Gly | Ile | Tyr | Gly | Gly | Ser | Gly | Ser | Gly | Ala | Gly | Ser | Gly | Ser |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Pro | Ala | Gly | Gly | Gly | Ala | Pro | Gly | Ser | Gly | Gly | Gly | Ser | Ser | Gly | Ser |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Glu | Thr | Pro | Gly | Thr | Ser | Glu | Ser | Ala | Thr | Pro | Glu | Ser | Met | Thr | Ser |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Glu | Lys | Gly | Pro | Ser | Thr | Gly | Asp | Pro | Thr | Leu | Arg | Arg | Arg | Ile | Glu |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

165

170

175

Pro Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu Arg Lys Glu  
 180 185 190

Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg Lys Ile Trp  
 195 200 205

Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val Asn Phe Ile  
 210 215 220

Lys Lys Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met Ser Cys Ser  
 225 230 235 240

Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser Gln Ala  
 245 250 255

Ile Arg Glu Phe Leu Ser Arg His Pro Gly Val Thr Leu Val Ile Tyr  
 260 265 270

Val Ala Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg Gln Gly Leu  
 275 280 285

Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met Arg Ala Ser  
 290 295 300

Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro Pro Gly Asp  
 305 310 315 320

Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met Leu Tyr Ala  
 325 330 335

Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys Leu Lys Ile  
 340 345 350



Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu His Leu Gln  
355 360 365

Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu Ala Thr Gly  
370 375 380

Leu Ile His Pro Ser Val Ala Trp Arg  
385 390

<210> 58

<211> 365

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 58

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met

85

90

95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
 100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
 115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Ala Gln  
 130 135 140

Lys Glu Glu Ala Ala Val Ala Thr Glu Ala Ala Ser Gln Asn Gly Glu  
 145 150 155 160

Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu Leu Ile Glu  
 165 170 175

Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro Ala Asn Phe  
 180 185 190

Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly Arg Asn Lys  
 195 200 205

Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly Gly Gln Val  
 210 215 220

Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala Ala His Ala  
 225 230 235 240

Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp Pro Ala Leu  
 245 250 255

Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys Ala Ala Cys  
 260 265 270

Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn Leu Arg Leu  
275 280 285

Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro Glu Ile Gln  
290 295 300

Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu Arg Ile Met  
305 310 315 320

Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val Glu Gln Glu  
325 330 335

Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile Gln Glu Asn  
340 345 350

Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys  
355 360 365

<210> 59

<211> 340

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 59

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Trp  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser

35

40

45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met  
85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Glu Ala  
130 135 140

Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His Ile Phe Thr  
145 150 155 160

Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Thr Tyr Leu Cys Tyr  
165 170 175

Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met Asp Gln His  
180 185 190

Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys Gly Phe Tyr  
195 200 205

Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro Ser Leu Gln  
210 215 220

Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser  
225 230 235 240

Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala Phe Leu Gln  
245 250 255

Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp  
260 265 270

Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg Asp Ala Gly  
275 280 285

Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His Cys Trp Asp  
290 295 300

Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu  
305 310 315 320

Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln  
325 330 335

Asn Gln Gly Asn  
340

<210> 60

<211> 523

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 60

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| 1   |     | 5   |     |     |     |     |     | 10  |     |     |     |     | 15  |     |     |  |
| Gly | Asp | Val | Thr | Val | Ala | Pro | Ser | Asn | Phe | Ala | Asn | Gly | Ile | Ala | Glu |  |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |  |
| Trp | Ile | Ser | Ser | Asn | Ser | Arg | Ser | Gln | Ala | Tyr | Lys | Val | Thr | Cys | Ser |  |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |  |
| Val | Arg | Gln | Ser | Ser | Ala | Gln | Asn | Arg | Lys | Tyr | Thr | Ile | Lys | Val | Glu |  |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |  |
| Val | Pro | Lys | Gly | Ala | Trp | Arg | Ser | Tyr | Leu | Asn | Met | Glu | Leu | Thr | Ile |  |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |  |
| Pro | Ile | Phe | Ala | Thr | Asn | Ser | Asp | Cys | Glu | Leu | Ile | Val | Lys | Ala | Met |  |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |  |
| Gln | Gly | Leu | Leu | Lys | Asp | Gly | Asn | Pro | Ile | Pro | Ser | Ala | Ile | Ala | Ala |  |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |  |
| Asn | Ser | Gly | Ile | Tyr | Gly | Gly | Ser | Gly | Ser | Gly | Ala | Gly | Ser | Gly | Ser |  |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |  |
| Pro | Ala | Gly | Gly | Gly | Ala | Pro | Gly | Ser | Gly | Gly | Gly | Ser | Met | Asn | Pro |  |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |  |
| Gln | Ile | Arg | Asn | Pro | Met | Glu | Arg | Met | Tyr | Arg | Asp | Thr | Phe | Tyr | Asp |  |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |  |
| Asn | Phe | Glu | Asn | Glu | Pro | Ile | Leu | Tyr | Gly | Arg | Ser | Tyr | Thr | Trp | Leu |  |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |  |
| Cys | Tyr | Glu | Val | Lys | Ile | Lys | Arg | Gly | Arg | Ser | Asn | Leu | Leu | Trp | Asp |  |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |  |

Thr Gly Val Phe Arg Gly Gln Val Tyr Phe Lys Pro Gln Tyr His Ala  
195 200 205

Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu Pro Ala Tyr  
210 215 220

Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro Cys Pro Asp  
225 230 235 240

Cys Val Ala Lys Leu Ala Glu Phe Leu Ser Glu His Pro Asn Val Thr  
245 250 255

Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg Asp Tyr  
260 265 270

Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg Val Thr Ile  
275 280 285

Met Asp Tyr Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe Val Tyr Asn  
290 295 300

Glu Gly Gln Gln Phe Met Pro Trp Tyr Lys Phe Asp Glu Asn Tyr Ala  
305 310 315 320

Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Tyr Leu Met Asp Pro  
325 330 335

Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val Leu Arg Arg  
340 345 350

Arg Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr  
355 360 365

Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn Glu Ala Lys  
370 375 380

Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu  
385 390 395 400

Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val  
405 410 415

Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly  
420 425 430

Glu Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile  
435 440 445

Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu  
450 455 460

Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp  
465 470 475 480

Glu Phe Glu Tyr Cys Trp Phe Thr Phe Val Tyr Arg Gln Gly Cys Pro  
485 490 495

Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala Leu Ser Gly  
500 505 510

Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn  
515 520

<210> 61

<211> 331

<212> PRT

<213> Artificial Sequence



<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 61

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met  
85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Asn Pro  
130 135 140

Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr Phe Tyr Phe  
145 150 155 160

Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu Thr Trp Leu  
165 170 175

Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val Ser Trp Lys  
180 185 190

Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His Cys His Ala  
195 200 205

Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn  
210 215 220

Thr Lys Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro Cys Pro Asp  
225 230 235 240

Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser Asn Val Asn  
245 250 255

Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Tyr Pro Cys Tyr  
260 265 270

Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala Val Glu Ile  
275 280 285

Met Asp Tyr Glu Asp Phe Lys Tyr Cys Trp Glu Asn Phe Val Tyr Asn  
290 295 300

Asp Asn Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Thr Asn Phe Arg  
305 310 315 320

Leu Leu Lys Arg Arg Leu Arg Glu Ser Leu Gln  
325 330

<210> 62  
<211> 527  
<212> PRT  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 62

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met  
85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Asn Pro  
130 135 140

Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr Phe Tyr Asp  
145 150 155 160

Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr Thr Trp Leu  
165 170 175

Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu Leu Trp Asp  
180 185 190

Thr Gly Val Phe Arg Gly Pro Val Leu Pro Lys Arg Gln Ser Asn His  
195 200 205

Arg Gln Glu Val Tyr Phe Arg Phe Glu Asn His Ala Glu Met Cys Phe  
210 215 220

Leu Ser Trp Phe Cys Gly Asn Arg Leu Pro Ala Asn Arg Arg Phe Gln  
225 230 235 240

Ile Thr Trp Phe Val Ser Trp Asn Pro Cys Leu Pro Cys Val Val Lys  
245 250 255

Val Thr Lys Phe Leu Ala Glu His Pro Asn Val Thr Leu Thr Ile Ser  
260 265 270

Ala Ala Arg Leu Tyr Tyr Tyr Arg Asp Arg Asp Trp Arg Trp Val Leu  
275 280 285

Leu Arg Leu His Lys Ala Gly Ala Arg Val Lys Ile Met Asp Tyr Glu  
290 295 300

Asp Phe Ala Tyr Cys Trp Glu Asn Phe Val Cys Asn Glu Gly Gln Pro  
305 310 315 320

Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn Tyr Ala Ser Leu His Arg  
325 330 335

Thr Leu Lys Glu Ile Leu Arg Asn Pro Met Glu Ala Met Tyr Pro His  
340 345 350

Ile Phe Tyr Phe His Phe Lys Asn Leu Leu Lys Ala Cys Gly Arg Asn  
355 360 365

Glu Ser Trp Leu Cys Phe Thr Met Glu Val Thr Lys His His Ser Ala  
370 375 380

Val Phe Arg Lys Arg Gly Val Phe Arg Asn Gln Val Asp Pro Glu Thr  
385 390 395 400

His Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile  
405 410 415

Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr Trp Tyr Thr Ser Trp Ser  
420 425 430

Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His  
435 440 445

Ser Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Cys Tyr Phe Trp  
450 455 460

Asp Thr Asp Tyr Gln Glu Gly Leu Cys Ser Leu Ser Gln Glu Gly Ala  
465 470 475 480

Ser Val Lys Ile Met Gly Tyr Lys Asp Phe Val Ser Cys Trp Lys Asn  
485 490 495

Phe Val Tyr Ser Asp Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Gln

500

505

510

Thr Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ile Leu Gln  
 515 520 525

&lt;210&gt; 63

&lt;211&gt; 514

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: Synthetic  
 polypeptide

&lt;400&gt; 63

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
 1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
 20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
 35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
 50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
 65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met  
 85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
 100 105 110

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asn | Ser | Gly | Ile | Tyr | Gly | Gly | Ser | Gly | Ser | Gly | Ala | Gly | Ser | Gly | Ser | 115 | 120 | 125 |     |
| Pro | Ala | Gly | Gly | Gly | Ala | Pro | Gly | Ser | Gly | Gly | Gly | Ser | Met | Lys | Pro | 130 | 135 | 140 |     |
| His | Phe | Arg | Asn | Thr | Val | Glu | Arg | Met | Tyr | Arg | Asp | Thr | Phe | Ser | Tyr | 145 | 150 | 155 | 160 |
| Asn | Phe | Tyr | Asn | Arg | Pro | Ile | Leu | Ser | Arg | Arg | Asn | Thr | Val | Trp | Leu | 165 | 170 | 175 |     |
| Cys | Tyr | Glu | Val | Lys | Thr | Lys | Gly | Pro | Ser | Arg | Pro | Arg | Leu | Asp | Ala | 180 | 185 | 190 |     |
| Lys | Ile | Phe | Arg | Gly | Gln | Val | Tyr | Ser | Gln | Pro | Glu | His | His | Ala | Glu | 195 | 200 | 205 |     |
| Met | Cys | Phe | Leu | Ser | Trp | Phe | Cys | Gly | Asn | Gln | Leu | Pro | Ala | Tyr | Lys | 210 | 215 | 220 |     |
| Cys | Phe | Gln | Ile | Thr | Trp | Phe | Val | Ser | Trp | Thr | Pro | Cys | Pro | Asp | Cys | 225 | 230 | 235 | 240 |
| Val | Ala | Lys | Leu | Ala | Glu | Phe | Leu | Ala | Glu | His | Pro | Asn | Val | Thr | Leu | 245 | 250 | 255 |     |
| Thr | Ile | Ser | Ala | Ala | Arg | Leu | Tyr | Tyr | Tyr | Trp | Glu | Arg | Asp | Tyr | Arg | 260 | 265 | 270 |     |
| Arg | Ala | Leu | Cys | Arg | Leu | Ser | Gln | Ala | Gly | Ala | Arg | Val | Lys | Ile | Met | 275 | 280 | 285 |     |
| Asp | Asp | Glu | Glu | Phe | Ala | Tyr | Cys | Trp | Glu | Asn | Phe | Val | Tyr | Ser | Glu |     |     |     |     |

290

295

300

Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn Tyr Ala Phe  
305 310 315 320

Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met Glu Ala Met  
325 330 335

Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Arg Lys Ala Tyr  
340 345 350

Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val Val Lys His  
355 360 365

His Ser Pro Val Ser Trp Lys Arg Gly Val Phe Arg Asn Gln Val Asp  
370 375 380

Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys  
385 390 395 400

Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr Trp Tyr Thr  
405 410 415

Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu  
420 425 430

Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr  
435 440 445

Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln  
450 455 460

Glu Gly Ala Ser Val Glu Ile Met Gly Tyr Lys Asp Phe Lys Tyr Cys  
465 470 475 480



Trp Glu Asn Phe Val Tyr Asn Asp Asp Glu Pro Phe Lys Pro Trp Lys  
485 490 495

Gly Leu Lys Tyr Asn Phe Leu Phe Leu Asp Ser Lys Leu Gln Glu Ile  
500 505 510

Leu Glu

<210> 64

<211> 525

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 64

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met

85

90

95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
 100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
 115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Lys Pro  
 130 135 140

His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr Phe Ser Tyr  
 145 150 155 160

Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr Val Trp Leu  
 165 170 175

Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro Leu Asp Ala  
 180 185 190

Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr His Pro Glu  
 195 200 205

Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu His Arg Asp  
 210 215 220

Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro Cys Thr Lys  
 225 230 235 240

Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro Lys Val Thr  
 245 250 255

Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp Pro Asp Tyr  
 260 265 270

Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly Pro Arg Ala  
275 280 285

Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys Trp Ser Lys  
290 295 300

Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn Asn Leu Pro  
305 310 315 320

Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile Leu Arg His  
325 330 335

Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp  
340 345 350

Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met  
355 360 365

His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys  
370 375 380

Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu  
385 390 395 400

Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln  
405 410 415

Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys  
420 425 430

Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu  
435 440 445

Cys Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu  
450 455 460

Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Ile Met Thr  
465 470 475 480

Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly  
485 490 495

Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu  
500 505 510

Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn  
515 520 525

<210> 65

<211> 341

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 65

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met  
85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Ala Leu  
130 135 140

Leu Thr Ala Glu Thr Phe Arg Leu Gln Phe Asn Asn Lys Arg Arg Leu  
145 150 155 160

Arg Arg Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr Gln Leu Thr  
165 170 175

Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly Tyr Phe Glu Asn Lys Lys  
180 185 190

Lys Cys His Ala Glu Ile Cys Phe Ile Asn Glu Ile Lys Ser Met Gly  
195 200 205

Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu Thr Trp Ser  
210 215 220

Pro Cys Ser Ser Cys Ala Trp Glu Leu Val Asp Phe Ile Lys Ala His  
225 230 235 240

Asp His Leu Asn Leu Gly Ile Phe Ala Ser Arg Leu Tyr Tyr His Trp  
245 250 255

Cys Lys Pro Gln Gln Lys Gly Leu Arg Leu Leu Cys Gly Ser Gln Val  
260 265 270

Pro Val Glu Val Met Gly Phe Pro Glu Phe Ala Asp Cys Trp Glu Asn  
275 280 285

Phe Val Asp His Glu Lys Pro Leu Ser Phe Asn Pro Tyr Lys Met Leu  
290 295 300

Glu Glu Leu Asp Lys Asn Ser Arg Ala Ile Lys Arg Arg Leu Glu Arg  
305 310 315 320

Ile Lys Ile Pro Gly Val Arg Ala Gln Gly Arg Tyr Met Asp Ile Leu  
325 330 335

Cys Asp Ala Glu Val  
340

<210> 66

<211> 508

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 66

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met  
85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Glu Pro  
130 135 140

Ile Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val Lys Pro Tyr  
145 150 155 160

Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro Tyr His Ile  
165 170 175

Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe Cys Gln Ile  
180 185 190

Phe Gly Phe Pro Tyr Gly Thr Thr Phe Pro Gln Thr Lys His Leu Thr  
195 200 205

Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln Lys Gly His  
210 215 220

Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser Met Leu Phe  
225 230 235 240

Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn Asp Ser Ile  
245 250 255

Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asn Glu Ala Asn  
260 265 270

His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Ile Thr Tyr Pro Gly  
275 280 285

Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr Glu Met Asp  
290 295 300

Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser Leu Ala Ser  
305 310 315 320

Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly Ile Trp His  
325 330 335

Ser Val Leu His Ser Phe Ile Ser Gly Val Ser Gly Ser His Val Phe  
340 345 350

Gln Pro Ile Leu Thr Gly Arg Ala Leu Ala Asp Arg His Asn Ala Tyr  
355 360 365

Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Tyr Phe Thr Asp Val Leu  
370 375 380



Leu Gln Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Glu Ala Leu Glu  
385 390 395 400

Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Phe Phe Gln Met Pro  
405 410 415

Ser Gly Gln Leu Gln Pro Asn Leu Pro Pro Asp Leu Arg Ala Pro Val  
420 425 430

Val Phe Val Leu Val Pro Leu Arg Asp Leu Pro Pro Met His Met Gly  
435 440 445

Gln Asn Pro Asn Lys Pro Arg Asn Ile Val Arg His Leu Asn Met Pro  
450 455 460

Gln Met Ser Phe Gln Glu Thr Lys Asp Leu Gly Arg Leu Pro Thr Gly  
465 470 475 480

Arg Ser Val Glu Ile Val Glu Ile Thr Glu Gln Phe Ala Ser Ser Lys  
485 490 495

Glu Ala Asp Glu Lys Lys Lys Lys Lys Gly Lys Lys  
500 505

<210> 67

<211> 339

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polypeptide

<400> 67

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met  
85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Asp Ser  
130 135 140

Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys Asn Val Arg  
145 150 155 160

Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val Val Lys Arg  
165 170 175

Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr Leu Arg Asn  
180 185 190

Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr Ile Ser Asp  
195 200 205

Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp Phe Thr Ser  
210 215 220

Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp Phe Leu Arg  
225 230 235 240

Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg Leu Tyr Phe  
245 250 255

Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg Leu His Arg  
260 265 270

Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr Phe Tyr Cys  
275 280 285

Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu  
290 295 300

Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu Arg Arg Ile  
305 310 315 320

Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr  
325 330 335

Leu Gly Leu

<210> 68

<211> 643

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 68

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr  
1 5 10 15

Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu  
20 25 30

Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser  
35 40 45

Val Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu  
50 55 60

Val Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile  
65 70 75 80

Pro Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met  
85 90 95

Gln Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala  
100 105 110

Asn Ser Gly Ile Tyr Gly Gly Ser Gly Ser Gly Ala Gly Ser Gly Ser  
115 120 125

Pro Ala Gly Gly Gly Ala Pro Gly Ser Gly Gly Gly Ser Met Trp Thr  
130 135 140

Ala Asp Glu Ile Ala Gln Leu Cys Tyr Glu His Tyr Gly Ile Arg Leu  
145 150 155 160

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| Pro | Lys | Lys | Gly | Lys | Pro | Glu | Pro | Asn | His | Glu | Trp | Thr | Leu | Leu | Ala |  |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Ala | Val | Val | Lys | Ile | Gln | Ser | Pro | Ala | Asp | Lys | Ala | Cys | Asp | Thr | Pro |  |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Asp | Lys | Pro | Val | Gln | Val | Thr | Lys | Glu | Val | Val | Ser | Met | Gly | Thr | Gly |  |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Thr | Lys | Cys | Ile | Gly | Gln | Ser | Lys | Met | Arg | Lys | Asn | Gly | Asp | Ile | Leu |  |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Asn | Asp | Ser | His | Ala | Glu | Val | Ile | Ala | Arg | Arg | Ser | Phe | Gln | Arg | Tyr |  |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Leu | Leu | His | Gln | Leu | Gln | Leu | Ala | Ala | Thr | Leu | Lys | Glu | Asp | Ser | Ile |  |
|     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Phe | Val | Pro | Gly | Thr | Gln | Lys | Gly | Val | Trp | Lys | Leu | Arg | Arg | Asp | Leu |  |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Ile | Phe | Val | Phe | Phe | Ser | Ser | His | Thr | Pro | Cys | Gly | Asp | Ala | Ser | Ile |  |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Ile | Pro | Met | Leu | Glu | Phe | Glu | Asp | Gln | Pro | Cys | Cys | Pro | Val | Phe | Arg |  |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Asn | Trp | Ala | His | Asn | Ser | Ser | Val | Glu | Ala | Ser | Ser | Asn | Leu | Glu | Ala |  |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Pro | Gly | Asn | Glu | Arg | Lys | Cys | Glu | Asp | Pro | Asp | Ser | Pro | Val | Thr | Lys |  |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |  |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
| Lys | Met | Arg | Leu | Glu | Pro | Gly | Thr | Ala | Ala | Arg | Glu | Val | Thr | Asn | Gly |  |

340

345

350

Ala Ala His His Gln Ser Phe Gly Lys Gln Lys Ser Gly Pro Ile Ser  
 355 360 365

Pro Gly Ile His Ser Cys Asp Leu Thr Val Glu Gly Leu Ala Thr Val  
 370 375 380

Thr Arg Ile Ala Pro Gly Ser Ala Lys Val Ile Asp Val Tyr Arg Thr  
 385 390 395 400

Gly Ala Lys Cys Val Pro Gly Glu Ala Gly Asp Ser Gly Lys Pro Gly  
 405 410 415

Ala Ala Phe His Gln Val Gly Leu Leu Arg Val Lys Pro Gly Arg Gly  
 420 425 430

Asp Arg Thr Arg Ser Met Ser Cys Ser Asp Lys Met Ala Arg Trp Asn  
 435 440 445

Val Leu Gly Cys Gln Gly Ala Leu Leu Met His Leu Leu Glu Glu Pro  
 450 455 460

Ile Tyr Leu Ser Ala Val Val Ile Gly Lys Cys Pro Tyr Ser Gln Glu  
 465 470 475 480

Ala Met Gln Arg Ala Leu Ile Gly Arg Cys Gln Asn Val Ser Ala Leu  
 485 490 495

Pro Lys Gly Phe Gly Val Gln Glu Leu Lys Ile Leu Gln Ser Asp Leu  
 500 505 510

Leu Phe Glu Gln Ser Arg Ser Ala Val Gln Ala Lys Arg Ala Asp Ser  
 515 520 525

Pro Gly Arg Leu Val Pro Cys Gly Ala Ala Ile Ser Trp Ser Ala Val  
530 535 540

Pro Glu Gln Pro Leu Asp Val Thr Ala Asn Gly Phe Pro Gln Gly Thr  
545 550 555 560

Thr Lys Lys Thr Ile Gly Ser Leu Gln Ala Arg Ser Gln Ile Ser Lys  
565 570 575

Val Glu Leu Phe Arg Ser Phe Gln Lys Leu Leu Ser Arg Ile Ala Arg  
580 585 590

Asp Lys Trp Pro His Ser Leu Arg Val Gln Lys Leu Asp Thr Tyr Gln  
595 600 605

Glu Tyr Lys Glu Ala Ala Ser Ser Tyr Gln Glu Ala Trp Ser Thr Leu  
610 615 620

Arg Lys Gln Val Phe Gly Ser Trp Ile Arg Asn Pro Pro Asp Tyr His  
625 630 635 640

Gln Phe Lys

<210> 69

<211> 741

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 69

atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac  
60

ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac  
120

ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc  
180

ctcgtgacca ccctgaccta cggcgtgcag tgcttcagcc gctaccccga ccacatgaag  
240

cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc  
300

ttcaaggacg acggcaacta caagaccgc gccgaggtga agttcgaggg cgacaccctg  
360

gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac  
420

aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac  
480

ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc  
540

gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaatcat  
600

tattcgagca ctcagtccgc cctgagcaaa gaccccaacg agaagcgcga tcacatggtc  
660

ctgctggagt tcgtgaccgc cgccggggtc actctcggca tggacgagct gtacaagtac  
720

t c a g a t c t c g a g c t c a a g t a g  
741

<210> 70

<211> 140

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic



## polynucleotide

<400> 70

gcgcctccat cgaagtgtaa agccatctat cgatgggaaa gtctttgctc tccccggaca  
60

ctctgctcct taaatgaaac acgggctcgt gcacggggcc aaagcgcctc catcgaagtg  
120

t a a a g c c a t c  
140

t a t c g a t g g g

<210> 71

<211> 98

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 71

gccatctatc gatgggaaag tctttgctct ccccggacac tctgctcctt aaatgaaaca  
60

c g g g c t c g t g            c a c g g g g c c a            a a g c c a t c t a            t c g a t g g g  
98

<210> 72

<211> 68

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 72

gccatctatc gatgggaaac ggactgagtg ctcaaaatga ttgtcggggca aagccatcta  
60

t            c            g            a            t            g            g            g  
68

<210> 73  
<211> 78  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 73  
gccatctatc gatgggaaat gctcagggcg gactgagaaa attgtcgggc agcagcacga  
60

a a g c c a t c t a t c g a t g g g  
78

<210> 74  
<211> 102  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 74  
aacatgagga tcacccatgt cgtctttgct ctccccggac actctgctcc ttaaataaaa  
60

cacgggctcg tgcacggggc caacatgagg atcacccatg tc  
102

<210> 75  
<211> 72  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 75

aacatgagga tcacccatgt ccggactgag tgctcaaaat gattgtcggg caacatgagg  
60

a t c a c c c a t g t c  
72

<210> 76

<211> 82

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 76

aacatgagga tcacccatgt ctgctcaggg cggactgaga aaattgtcgg gcagcagcac  
60

g a a c a t g a g g a t c a c c c a t g t c  
82

<210> 77

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 77

g c g c c t c c a t c g a a g t g t  
18

<210> 78

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 78

g c c a t c t a t c  
16

g a t g g g

<210> 79

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 79

c a a a c c g a g c  
16

g g t g t c

<210> 80

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 80

a c a a a c c g a g  
16

c g g t g t

<210> 81

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 81

t a c a a a c c g a  
16

g c g g t g

<210> 82

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 82

t t a c a a a c c g  
16

a g c g g t

<210> 83

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 83

c a a a c c g a g c  
17

g g t g t c t

<210> 84

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 84

a c a a a c c g a g  
17

c g g t g t c

<210> 85  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 85  
t a c a a a c c g a g c g g t g t  
17

<210> 86  
<211> 17  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 86  
t t a c a a a c c g a g c g g t g  
17

<210> 87  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 87  
c a a a c c g a g c g g t g t c t g  
18

<210> 88

<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 88

a c a a a c c g a g  
18

c g g t g t c t

<210> 89

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 89

t a c a a a c c g a  
18

g c g g t g t c

<210> 90

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 90

t t a c a a a c c g  
18

a g c g g t g t

<210> 91

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 91

c a a a c c g a g c  
19

g g t g t c t g t

<210> 92

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 92

a c a a a c c g a g  
19

c g g t g t c t g

<210> 93

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 93

t a c a a a c c g a  
19

g c g g t g t c t

<210> 94

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic



oligonucleotide

<400> 94

t t a c a a a c c g  
19

a g c g g t g t c

<210> 95

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 95

c a a a c c g a g c  
20

g g t g t c t g t g

<210> 96

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 96

a c a a a c c g a g  
20

c g g t g t c t g t

<210> 97

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 97

t a c a a a c c g a

20

g c g g t g t c t g

<210> 98

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 98

t t a c a a a c c g

20

a g c g g t g t c t

<210> 99

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 99

c a a a c c g a g c

21

g g t g t c t g t g

a

<210> 100

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 100

a c a a a c c g a g

21

c g g t g t c t g t

g

<210> 101  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 101  
t a c a a a c c g a g c g g t g t c t g t  
21

<210> 102  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 102  
t t a c a a a c c g a g c g g t g t c t g  
21

<210> 103  
<211> 66  
<212> DNA  
<213> Unknown

<220>  
<223> Description of Unknown:  
target sequence

<400> 103  
cacttgggtg tgaatgaaag tctcacagac accgctcagt ttgtaaaact tttcttcctt  
60  
c c a a a g  
66

<210> 104  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 104  
aacatgagga tcacccatgt ccaaaccgag cgggtgtctgt gaacatgagg atcacccatg  
60

t c  
62

<210> 105  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 105  
aacatgagga tcacccatgt ctacaaaccg agcggtgtct gaacatgagg atcacccatg  
60

t c  
62

<210> 106  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 106

aacatgagga tcacccatgt ctttaciaaac cgagcgggtgt caacatgagg atcacccatg  
60

t c  
62

<210> 107  
<211> 66  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 107  
aacatgagga tcacccatgt ccaaaccgag cgggtgtctgt gagacaacat gaggatcacc  
60

c a t g t c  
66

<210> 108  
<211> 58  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 108  
aacatgagga tcacccatgt ccaaaccgag cgggtgtcaac atgaggatca cccatgtc  
58

<210> 109  
<211> 82  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 109

aagaaaagtt aacatgagga tcacccatgt ccaaaccgag cgggtgtctgt gaacatgagg  
60

a t c a c c c a t g                      t c c t t t c a t t                      c a  
82

<210> 110

<211> 102

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 110

ctttggaagg aagaaaagtt aacatgagga tcacccatgt ccaaaccgag cgggtgtctgt  
60

gaacatgagg      atcacccatg      tccttttcatt      cacacccaag      tg  
102

<210> 111

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 111

t t g g g a a a t c                      c a g c t a g c g g                      c a g t a t t c t g                      t  
31

<210> 112

<211> 31

<212> DNA

<213> Unknown

**<220>**

**<223> Description of Unknown:**  
target sequence

**<400> 112**

g g g c g a t g c c                    a c c t a a g g c a                    a g c t g a c c c t                    g  
31

**<210> 113**

**<211> 19**

**<212> DNA**

### <213> Artificial Sequence

**<220>**

**<223> Description of Artificial Sequence: Synthetic oligonucleotide**

**<400> 113**

g c c c t a g g t g g c a t c g c c c

**<210> 114**

**<211> 19**

## <212> DNA

### <213> Artificial Sequence

**<220>**

**<223> Description of Artificial Sequence: Synthetic oligonucleotide**

**<400> 114**

c a g g g t c a g c                      t t g c c c t a g

19

**<210> 115**

**<211> 19**

## <212> DNA

### <213> Artificial Sequence

**<220>**

### <223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 115

g c c c c a g g t g  
19

g c a t c g c c c

<210> 116

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 116

a g g g t c a g c t  
19

t g c c c c a g g

<210> 117

<211> 14

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 117

a g c a a t a a a a  
14

t g g c

<210> 118

<211> 14

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 118



a g c a a t a a a a  
14

t g g c

<210> 119  
<211> 34  
<212> DNA  
<213> Mus sp.

<400> 119  
g a g c a a t a a a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 120  
<211> 34  
<212> DNA  
<213> Mus sp.

<400> 120  
g a g c a a t a a a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 121  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 121  
g a g c c a t a a a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 122  
<211> 34  
<212> DNA  
<213> Mus sp.

<400> 122  
g a g c a a t a a a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 123  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 123  
g a g c a a t g g a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 124  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 124  
g a g c c a t a a a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 125  
<211> 34  
<212> DNA  
<213> Mus sp.

<400> 125  
g a g c a a t a a a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 126  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 126

g a g c a a t g g a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 127

<211> 34

<212> DNA

<213> Mus sp.

<400> 127

g a g c a a t a a a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 128

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 128

g a g c a a t g g a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 129

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 129

g a g c a a t g a a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 130  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 130  
g a g c a g t g g a            a t g g c t t c a a            c t a t c t g a g t            g a c a  
34

<210> 131  
<211> 13  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>  
<221> modified\_base  
<222> (12)..(12)  
<223> a, c, t, g, unknown or other

<400> 131  
c c g c t c a g t t                                            t n t  
13

<210> 132  
<211> 13  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>  
<221> modified\_base  
<222> (7)..(8)  
<223> a, c, t, g, unknown or other

<220>  
<221> modified\_base  
<222> (12)..(12)  
<223> a, c, t, g, unknown or other

<400> 132  
c c g c t c n n t t t n t  
13

<210> 133  
<211> 28  
<212> DNA  
<213> Mus sp.

<400> 133  
a g a c a c c g c t c a t g t c t t a t c t a g c a t g  
28

<210> 134  
<211> 28  
<212> DNA  
<213> Mus sp.

<400> 134  
a g a c a c c g c t c a t g t c t t a t c t a g c a t g  
28

<210> 135  
<211> 28  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 135

a g a c a c c g c t  
28

c g t g t c t t a t

c t a g c a t g

<210> 136

<211> 28

<212> DNA

<213> Mus sp.

<400> 136

a g a c a c c g c t  
28

c a t g t c t t a t

c t a g c a t g

<210> 137

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 137

a g a c a c c g c t  
28

c g t g t c t t a t

c t a g c a t g

<210> 138

<211> 28

<212> DNA

<213> Mus sp.

<400> 138

a g a c a c c g c t  
28

c a t g t c t t a t

c t a g c a t g

<210> 139

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 139

a g a c a c c g c t  
28

c g t g t c t t a t

c t a g c a t g

<210> 140

<211> 429

<212> PRT

<213> Unknown

<220>

<223> Description of Unknown:

ADAR1 catalytic domain sequence

<400> 140

Lys Ala Glu Arg Met Gly Phe Thr Glu Val Thr Pro Val Thr Gly Ala  
1 5 10 15

Ser Leu Arg Arg Thr Met Leu Leu Leu Ser Arg Ser Pro Glu Ala Gln  
20 25 30

Pro Lys Thr Leu Pro Leu Thr Gly Ser Thr Phe His Asp Gln Ile Ala  
35 40 45

Met Leu Ser His Arg Cys Phe Asn Thr Leu Thr Asn Ser Phe Gln Pro  
50 55 60

Ser Leu Leu Gly Arg Lys Ile Leu Ala Ala Ile Ile Met Lys Lys Asp  
65 70 75 80

Ser Glu Asp Met Gly Val Val Val Ser Leu Gly Thr Gly Asn Arg Cys  
85 90 95

Val Lys Gly Asp Ser Leu Ser Leu Lys Gly Glu Thr Val Asn Asp Cys  
100 105 110

His Ala Glu Ile Ile Ser Arg Arg Gly Phe Ile Arg Phe Leu Tyr Ser  
115 120 125

Glu Leu Met Lys Tyr Asn Ser Gln Thr Ala Lys Asp Ser Ile Phe Glu  
130 135 140

Pro Ala Lys Gly Gly Glu Lys Leu Gln Ile Lys Lys Thr Val Ser Phe  
145 150 155 160

His Leu Tyr Ile Ser Thr Ala Pro Cys Gly Asp Gly Ala Leu Phe Asp  
165 170 175

Lys Ser Cys Ser Asp Arg Ala Met Glu Ser Thr Glu Ser Arg His Tyr  
180 185 190

Pro Val Phe Glu Asn Pro Lys Gln Gly Lys Leu Arg Thr Lys Val Glu  
195 200 205

Asn Gly Glu Gly Thr Ile Pro Val Glu Ser Ser Asp Ile Val Pro Thr  
210 215 220

Trp Asp Gly Ile Arg Leu Gly Glu Arg Leu Arg Thr Met Ser Cys Ser  
225 230 235 240

Asp Lys Ile Leu Arg Trp Asn Val Leu Gly Leu Gln Gly Ala Leu Leu  
245 250 255

Thr His Phe Leu Gln Pro Ile Tyr Leu Lys Ser Val Thr Leu Gly Tyr  
260 265 270

Leu Phe Ser Gln Gly His Leu Thr Arg Ala Ile Cys Cys Arg Val Thr  
275 280 285



Arg Asp Gly Ser Ala Phe Glu Asp Gly Leu Arg His Pro Phe Ile Val  
290 295 300

Asn His Pro Lys Val Gly Arg Val Ser Ile Tyr Asp Ser Lys Arg Gln  
305 310 315 320

Ser Gly Lys Thr Lys Glu Thr Ser Val Asn Trp Cys Leu Ala Asp Gly  
325 330 335

Tyr Asp Leu Glu Ile Leu Asp Gly Thr Arg Gly Thr Val Asp Gly Pro  
340 345 350

Arg Asn Glu Leu Ser Arg Val Ser Lys Lys Asn Ile Phe Leu Leu Phe  
355 360 365

Lys Lys Leu Cys Ser Phe Arg Tyr Arg Arg Asp Leu Leu Arg Leu Ser  
370 375 380

Tyr Gly Glu Ala Lys Lys Ala Ala Arg Asp Tyr Glu Thr Ala Lys Asn  
385 390 395 400

Tyr Phe Lys Lys Gly Leu Lys Asp Met Gly Tyr Gly Asn Trp Ile Ser  
405 410 415

Lys Pro Gln Glu Glu Lys Asn Phe Tyr Leu Cys Pro Val  
420 425

<210> 141

<211> 385

<212> PRT

<213> Unknown

<220>

<223> Description of Unknown:  
ADAR2 catalytic domain sequence

<400> 141

Gln Leu His Leu Pro Gln Val Leu Ala Asp Ala Val Ser Arg Leu Val  
1 5 10 15

Leu Gly Lys Phe Gly Asp Leu Thr Asp Asn Phe Ser Ser Pro His Ala  
20 25 30

Arg Arg Lys Val Leu Ala Gly Val Val Met Thr Thr Gly Thr Asp Val  
35 40 45

Lys Asp Ala Lys Val Ile Ser Val Ser Thr Gly Thr Lys Cys Ile Asn  
50 55 60

Gly Glu Tyr Met Ser Asp Arg Gly Leu Ala Leu Asn Asp Cys His Ala  
65 70 75 80

Glu Ile Ile Ser Arg Arg Ser Leu Leu Arg Phe Leu Tyr Thr Gln Leu  
85 90 95

Glu Leu Tyr Leu Asn Asn Lys Asp Asp Gln Lys Arg Ser Ile Phe Gln  
100 105 110

Lys Ser Glu Arg Gly Gly Phe Arg Leu Lys Glu Asn Val Gln Phe His  
115 120 125

Leu Tyr Ile Ser Thr Ser Pro Cys Gly Asp Ala Arg Ile Phe Ser Pro  
130 135 140

His Glu Pro Ile Leu Glu Glu Pro Ala Asp Arg His Pro Asn Arg Lys  
145 150 155 160

Ala Arg Gly Gln Leu Arg Thr Lys Ile Glu Ser Gly Glu Gly Thr Ile  
165 170 175

Pro Val Arg Ser Asn Ala Ser Ile Gln Thr Trp Asp Gly Val Leu Gln  
180 185 190

Gly Glu Arg Leu Leu Thr Met Ser Cys Ser Asp Lys Ile Ala Arg Trp  
195 200 205

Asn Val Val Gly Ile Gln Gly Ser Leu Leu Ser Ile Phe Val Glu Pro  
210 215 220

Ile Tyr Phe Ser Ser Ile Ile Leu Gly Ser Leu Tyr His Gly Asp His  
225 230 235 240

Leu Ser Arg Ala Met Tyr Gln Arg Ile Ser Asn Ile Glu Asp Leu Pro  
245 250 255

Pro Leu Tyr Thr Leu Asn Lys Pro Leu Leu Ser Gly Ile Ser Asn Ala  
260 265 270

Glu Ala Arg Gln Pro Gly Lys Ala Pro Asn Phe Ser Val Asn Trp Thr  
275 280 285

Val Gly Asp Ser Ala Ile Glu Val Ile Asn Ala Thr Thr Gly Lys Asp  
290 295 300

Glu Leu Gly Arg Ala Ser Arg Leu Cys Lys His Ala Leu Tyr Cys Arg  
305 310 315 320

Trp Met Arg Val His Gly Lys Val Pro Ser His Leu Leu Arg Ser Lys  
325 330 335

Ile Thr Lys Pro Asn Val Tyr His Glu Ser Lys Leu Ala Ala Lys Glu  
340 345 350

Tyr Gln Ala Ala Lys Ala Arg Leu Phe Thr Ala Phe Ile Lys Ala Gly

355

360

365

Leu Gly Ala Trp Val Glu Lys Pro Thr Glu Gln Asp Gln Phe Ser Leu  
 370 375 380

Thr  
 385

<210> 142  
 <211> 298  
 <212> PRT  
 <213> Unknown

<220>  
 <223> Description of Unknown:  
 ADAR dsRBD sequence

<400> 142  
 Met Asp Ile Glu Asp Glu Glu Asn Met Ser Ser Ser Ser Thr Asp Val  
 1 5 10 15

Lys Glu Asn Arg Asn Leu Asp Asn Val Ser Pro Lys Asp Gly Ser Thr  
 20 25 30

Pro Gly Pro Gly Glu Gly Ser Gln Leu Ser Asn Gly Gly Gly Gly Gly  
 35 40 45

Pro Gly Arg Lys Arg Pro Leu Glu Glu Gly Ser Asn Gly His Ser Lys  
 50 55 60

Tyr Arg Leu Lys Lys Arg Arg Lys Thr Pro Gly Pro Val Leu Pro Lys  
 65 70 75 80

Asn Ala Leu Met Gln Leu Asn Glu Ile Lys Pro Gly Leu Gln Tyr Thr  
 85 90 95

Leu Leu Ser Gln Thr Gly Pro Val His Ala Pro Leu Phe Val Met Ser  
 100 105 110

Val Glu Val Asn Gly Gln Val Phe Glu Gly Ser Gly Pro Thr Lys Lys  
 115 120 125

Lys Ala Lys Leu His Ala Ala Glu Lys Ala Leu Arg Ser Phe Val Gln  
 130 135 140

Phe Pro Asn Ala Ser Glu Ala His Leu Ala Met Gly Arg Thr Leu Ser  
 145 150 155 160

Val Asn Thr Asp Phe Thr Ser Asp Gln Ala Asp Phe Pro Asp Thr Leu  
 165 170 175

Phe Asn Gly Phe Glu Thr Pro Asp Lys Ala Glu Pro Pro Phe Tyr Val  
 180 185 190

Gly Ser Asn Gly Asp Asp Ser Phe Ser Ser Ser Gly Asp Leu Ser Leu  
 195 200 205

Ser Ala Ser Pro Val Pro Ala Ser Leu Ala Gln Pro Pro Leu Pro Val  
 210 215 220

Leu Pro Pro Phe Pro Pro Pro Ser Gly Lys Asn Pro Val Met Ile Leu  
 225 230 235 240

Asn Glu Leu Arg Pro Gly Leu Lys Tyr Asp Phe Leu Ser Glu Ser Gly  
 245 250 255

Glu Ser His Ala Lys Ser Phe Val Met Ser Val Val Val Asp Gly Gln  
 260 265 270

Phe Phe Glu Gly Ser Gly Arg Asn Lys Lys Leu Ala Lys Ala Arg Ala

275

280

285

Ala Gln Ser Ala Leu Ala Ala Ile Phe Asn  
290 295

&lt;210&gt; 143

&lt;211&gt; 564

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: Synthetic  
polypeptide

&lt;400&gt; 143

Met Leu Arg Ser Phe Val Gln Phe Pro Asn Ala Ser Glu Ala His Leu  
1 5 10 15

Ala Met Gly Arg Thr Leu Ser Val Asn Thr Asp Phe Thr Ser Asp Gln  
20 25 30

Ala Asp Phe Pro Asp Thr Leu Phe Asn Gly Phe Glu Thr Pro Asp Lys  
35 40 45

Ala Glu Pro Pro Phe Tyr Val Gly Ser Asn Gly Asp Asp Ser Phe Ser  
50 55 60

Ser Ser Gly Asp Leu Ser Leu Ser Ala Ser Pro Val Pro Ala Ser Leu  
65 70 75 80

Ala Gln Pro Pro Leu Pro Val Leu Pro Pro Phe Pro Pro Pro Ser Gly  
85 90 95

Lys Asn Pro Val Met Ile Leu Asn Glu Leu Arg Pro Gly Leu Lys Tyr  
100 105 110

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asp | Phe | Leu | Ser | Glu | Ser | Gly | Glu | Ser | His | Ala | Lys | Ser | Phe | Val | Met |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ser | Val | Val | Val | Asp | Gly | Gln | Phe | Phe | Glu | Gly | Ser | Gly | Arg | Asn | Lys |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Lys | Leu | Ala | Lys | Ala | Arg | Ala | Ala | Gln | Ser | Ala | Leu | Ala | Ala | Ile | Phe |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Asn | Leu | His | Leu | Asp | Gln | Thr | Pro | Ser | Arg | Gln | Pro | Ile | Pro | Ser | Glu |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Gly | Leu | Gln | Leu | His | Leu | Pro | Gln | Val | Leu | Ala | Asp | Ala | Val | Ser | Arg |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Leu | Val | Leu | Gly | Lys | Phe | Gly | Asp | Leu | Thr | Asp | Asn | Phe | Ser | Ser | Pro |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| His | Ala | Arg | Arg | Lys | Val | Leu | Ala | Gly | Val | Val | Met | Thr | Thr | Gly | Thr |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Asp | Val | Lys | Asp | Ala | Lys | Val | Ile | Ser | Val | Ser | Thr | Gly | Thr | Lys | Cys |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Ile | Asn | Gly | Glu | Tyr | Met | Ser | Asp | Arg | Gly | Leu | Ala | Leu | Asn | Asp | Cys |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| His | Ala | Glu | Ile | Ile | Ser | Arg | Arg | Ser | Leu | Leu | Arg | Phe | Leu | Tyr | Thr |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Gln | Leu | Glu | Leu | Tyr | Leu | Asn | Asn | Lys | Asp | Asp | Gln | Lys | Arg | Ser | Ile |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| Phe | Gln | Lys | Ser | Glu | Arg | Gly | Gly | Phe | Arg | Leu | Lys | Glu | Asn | Val | Gln |

290

295

300

Phe His Leu Tyr Ile Ser Thr Ser Pro Cys Gly Asp Ala Arg Ile Phe  
305 310 315 320

Ser Pro His Glu Pro Ile Leu Glu Glu Pro Ala Asp Arg His Pro Asn  
325 330 335

Arg Lys Ala Arg Gly Gln Leu Arg Thr Lys Ile Glu Ser Gly Glu Gly  
340 345 350

Thr Ile Pro Val Arg Ser Asn Ala Ser Ile Gln Thr Trp Asp Gly Val  
355 360 365

Leu Gln Gly Glu Arg Leu Leu Thr Met Ser Cys Ser Asp Lys Ile Ala  
370 375 380

Arg Trp Asn Val Val Gly Ile Gln Gly Ser Leu Leu Ser Ile Phe Val  
385 390 395 400

Glu Pro Ile Tyr Phe Ser Ser Ile Ile Leu Gly Ser Leu Tyr His Gly  
405 410 415

Asp His Leu Ser Arg Ala Met Tyr Gln Arg Ile Ser Asn Ile Glu Asp  
420 425 430

Leu Pro Pro Leu Tyr Thr Leu Asn Lys Pro Leu Leu Ser Gly Ile Ser  
435 440 445

Asn Ala Glu Ala Arg Gln Pro Gly Lys Ala Pro Asn Phe Ser Val Asn  
450 455 460

Trp Thr Val Gly Asp Ser Ala Ile Glu Val Ile Asn Ala Thr Thr Gly  
465 470 475 480



Lys Asp Glu Leu Gly Arg Ala Ser Arg Leu Cys Lys His Ala Leu Tyr  
485 490 495

Cys Arg Trp Met Arg Val His Gly Lys Val Pro Ser His Leu Leu Arg  
500 505 510

Ser Lys Ile Thr Lys Pro Asn Val Tyr His Glu Ser Lys Leu Ala Ala  
515 520 525

Lys Glu Tyr Gln Ala Ala Lys Ala Arg Leu Phe Thr Ala Phe Ile Lys  
530 535 540

Ala Gly Leu Gly Ala Trp Val Glu Lys Pro Thr Glu Gln Asp Gln Phe  
545 550 555 560

Ser Leu Thr Pro

<210> 144

<211> 36

<212> DNA

<213> Mus sp.

<400> 144

ctcacagaca cgcctcggtt tgtaaaactt ttcttc  
36

<210> 145

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 145

ctcacagaca            ccgctcagtt            tgtaaaactt            ttcttc  
36

<210> 146

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 146

ctcacagaca            ccgctcagtt            tgtaaaactt            ttcttc  
36

<210> 147

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 147

ctcacagaca            ccgctcatgt            cttatctagc            atgaca  
36

<210> 148

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 148

ctcacagaca            ccgctcgtgt            cttatctagc            atgaca  
36

<210> 149  
<211> 16  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 149  
c g g c c t c a g t g a g c g a  
16

<210> 150  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 150  
g g a a c c c c t a g t g a t g g a g t t  
21

<210> 151  
<211> 51  
<212> DNA  
<213> Homo sapiens

<400> 151  
agcggcagta ttctgtacag tagacacaag aattatgtac gccttttattc a  
51

<210> 152  
<211> 66  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>

<221> modified\_base

<222> (46)..(51)

<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (53)..(66)

<223> a, c, t, g, unknown or other

<400> 152

gtggaagagg agaacaatat gctaaatggt gttctcgtct cccacnnnnn ncnnnnnnnn  
60

|    |   |   |   |   |   |
|----|---|---|---|---|---|
| n  | n | n | n | n | n |
| 66 |   |   |   |   |   |

<210> 153

<211> 76

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>

<221> modified\_base

<222> (56)..(61)

<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (63)..(76)

<223> a, c, t, g, unknown or other

<400> 153

ggtgtcgaga agaggagaac aatatgctaa atgttggtct cgtctcctcg acaccnnnnn  
60

n c n n n n n n n n  
76

n n n n n n

<210> 154  
<211> 65  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 154  
gtggaagagg agaacaatat gctaaatggt gttctcgtct cccactgccg ccagctggat  
60

t t c c c  
65

<210> 155  
<211> 75  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 155  
ggtgtcgaga agaggagaac aatatgctaa atgttggtct cgtctcctcg acacctgccg  
60

c c a g c t g g a t t t c c c  
75

<210> 156  
<211> 65  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 156

gtggaagagg agaacaatat gctaaatggt gttctcgtct cccacctcct ccacccgacc  
60

c c g g g  
65

<210> 157

<211> 75

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 157

ggtgtcgaga agaggagaac aatatgctaa atgttggtct cgtctcctcg acaccctcct  
60

c c a c c c g a c c c c g g g  
75

<210> 158

<211> 65

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 158

gtggaatagt ataacaatat gctaaatggt gttatagtat cccacgagcc ccaaaattaa  
60

a t a g a  
65

<210> 159  
<211> 65  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 159  
gtggaatagt ataacaatat gctaaatggt gttatagtat cccacttacc cgaaattttc  
60

g a a g t  
65

<210> 160  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 160  
aacatgagga tcacccatgt cgagcccca aattaaatag aacatgagg atcacccatg  
60

t c  
62

<210> 161  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 161  
aacatgagga tcacccatgt cttacccgaa attttcgaag taacatgagg atcacccatg  
60

t c  
62

<210> 162  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 162  
aacatgagga tcacccatgt cccattccat tgctctttca aaacatgagg atcacccatg  
60

t c  
62

<210> 163  
<211> 65  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 163  
gtggaatagt ataacaatat gctaaatggt gttatagtat cccacactgc ccacagatga  
60

a c a a g  
65

<210> 164  
<211> 65  
<212> DNA  
<213> Artificial Sequence

<220>



<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 164

gtggaatagt ataacaatat gctaaatggt gttatagtat cccactgtac ccgaaggaga  
60

g                    a                    a                    t                    a  
65

<210> 165

<211> 65

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 165

gtggaatagt ataacaatat gctaaatggt gttatagtat cccacaaagt ccgaggagga  
60

a                    a                    a                    a                    g  
65

<210> 166

<211> 65

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 166

gtggaatagt ataacaatat gctaaatggt gttatagtat cccacaatac ccgtaggatt  
60

a                    a                    a                    t                    c  
65



t  
62

c

<210> 170  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 170  
aacatgagga tcacccatgt caaagtccga ggaggaaaaa gaacatgagg atcacccatg  
60

t  
62

c

<210> 171  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 171  
aacatgagga tcacccatgt caataccgt aggattaaat caacatgagg atcacccatg  
60

t  
62

c

<210> 172  
<211> 62  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 172

aacatgagga tcacccatgt caagttccgg aaaacaaaca aaacatgagg atcacccatg  
60

t c  
62

<210> 173

<211> 63

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>

<221> modified\_base

<222> (13)..(22)

<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (44)..(63)

<223> a, c, t, g, unknown or other

<400> 173

acatatatga tannnnnnnn nnaacatgag gatcacccat gtcnnnnnnnn nnnnnnnnnn  
60

n n n  
63

<210> 174

<211> 66

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<220>

<221> modified\_base

<222> (1)..(20)

<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (42)..(51)

<223> a, c, t, g, unknown or other

<400> 174

nnnnnnnnnn nnnnnnnnnn aacatgagga tcacccatgt cnnnnnnnnn nttgatcagt  
60

a                    t                    a                    t                    t                    a  
66

<210> 175

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 175

a c a t a t a t g a                    t a c a a t t t g a                    t c a g t a t a t t  
30

<210> 176

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 176

a c a t a t a t g a

30

t a c a a t t t g a

t c a g t a t a t t

<210> 177

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 177

a a c a t g a g g a

21

t c a c c c a t g t

c

<210> 178

<211> 120

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 178

gtggaatagt ataacaatat gctaaatggt gttatagtat cccactgccg ccagctggat  
60

ttcccaattc tgagtgtgga atagtataac aatatgctaa atgttggtat agtatccac  
120

<210> 179

<211> 130

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 179

gtggaatagt ataacaatat gctaaatggt gttatagtat ccactgccg ccagctggat  
60

ttccaattc tgagtaacac tctgctgga atagtataac aatatgctaa atgttggtat  
120

a g t a t c c a c  
130

<210> 180

<211> 150

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 180

gtggaatagt ataacaatat gctaaatggt gttatagtat ccactgccg ccagctggat  
60

ttccaattc tgagtaacac tctgcaatcc aaacagggtt caaccgtgga atagtataac  
120

a a t a t g c t a a a t g t t g t t a t a g t a t c c c a c  
150

<210> 181

<211> 170

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 181

gtggaatagt ataacaatat gctaaatggt gttatagtat ccactgccg ccagctggat  
60

ttccaattc tgagtaacac tctgcaatcc aaacagggtt caaccctcca ccttacaggc  
120

ctgcagtgga atagtataac aatatgctaa atgttggttat agtatccac  
170

<210> 182

<211> 190

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 182

gtggaatagt ataacaatat gctaaatggt gttatagtat cccactgccg ccagctggat  
60

ttcccaattc tgagtaacac tctgcaatcc aaacagggtt caaccctcca ccttacaggc  
120

ctgcattaca ggacttaaac acatagtgga atagtataac aatatgctaa atgttggttat  
180

a g t a t c c c a c  
190

<210> 183

<211> 210

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 183

gtggaatagt ataacaatat gctaaatggt gttatagtat cccactgccg ccagctggat  
60

ttcccaattc tgagtaacac tctgcaatcc aaacagggtt caaccctcca ccttacaggc  
120

ctgcattaca ggacttaaac acataatcca agaatttctt acactgtgga atagtataac  
180



a a t a t g c t a a  
210

a t g t t g t t a t

a g t a t c c c a c

<210> 184

<211> 110

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 184

gtggaatagt ataacaatat gctaaatggt gttatagtat cccacatact gccgccagct  
60

ggattgtgga atagtataac aatatgctaa atgttggttat agtatccac  
110

<210> 185

<211> 130

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 185

gtggaatagt ataacaatat gctaaatggt gttatagtat cccacactgt acagaatact  
60

gccgccagct ggatttccca attctgtgga atagtataac aatatgctaa atgttggttat  
120

a g t a t c c c a c  
130

<210> 186

<211> 150

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 186

gtggaatagt ataacaatat gctaaatggt gttatagtat cccactcttg tgtctactgt  
60

acagaatact gccgccagct ggatttccca attctgagta acactgtgga atagtataac  
120

a a t a t g c t a a                      a t g t t g t t a t                      a g t a t c c c a c  
150

<210> 187

<211> 170

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 187

gtggaatagt ataacaatat gctaaatggt gttatagtat cccaccgtac ataattcttg  
60

tgtctactgt acagaatact gccgccagct ggatttccca attctgagta acactctgca  
120

atccagtgga atagtataac aatatgctaa atgttggttat agtatccac  
170

<210> 188

<211> 190

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 188

gtggaatagt ataacaatat gctaaatggt gttatagtat cccactgata aaaggcgtag  
60

ataattcttg tgtctactgt acagaatact gccgccagct ggatttccca attctgagta  
120

acactctgca atccaaacag gggtcgtgga atagtataac aatatgctaa atgttggtat  
180

a g t a t c c c a c  
190

<210> 189

<211> 210

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 189

gtggaatagt ataacaatat gctaaatggt gttatagtat cccaccttaa gtctttgata  
60

aaaggcgtag ataattcttg tgtctactgt acagaatact gccgccagct ggatttccca  
120

attctgagta acactctgca atccaaacag gggtcaacc tccacgtgga atagtataac  
180

a a t a t g c t a a a t g t t g t t a t a g t a t c c c a c  
210

<210> 190

<211> 120

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

## polynucleotide

<400> 190

gtggaagagg agaacaatag gctaaacgtt gttctcgtct cccactgccg ccagctggat  
60

ttcccaattc tgagtgtgga agaggagaac aataggctaa acgttggttct cgtctccac  
120

<210> 191

<211> 130

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 191

gtggaagagg agaacaatag gctaaacgtt gttctcgtct cccactgccg ccagctggat  
60

ttcccaattc tgagtaacac tctgcgtgga agaggagaac aataggctaa acgttggttct  
120

c g t c t c c c a c  
130

<210> 192

<211> 150

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 192

gtggaagagg agaacaatag gctaaacgtt gttctcgtct cccactgccg ccagctggat  
60

ttcccaattc tgagtaacac tctgcaatcc aaacagggtt caaccgtgga agaggagaac  
120

a a t a g g c t a a

150

a c g t t g t t c t

c g t c t c c c a c

<210> 193

<211> 170

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 193

gtggaagagg agaacaatag gctaaacgtt gttctcgtct cccactgccg ccagctggat  
60

ttcccaattc tgagtaacac tctgcaatcc aaacagggtt caaccctcca ccttacaggc  
120

ctgcagtgga agaggagaac aataggctaa acgttggttct cgtctccac  
170

<210> 194

<211> 190

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 194

gtggaagagg agaacaatag gctaaacgtt gttctcgtct cccactgccg ccagctggat  
60

ttcccaattc tgagtaacac tctgcaatcc aaacagggtt caaccctcca ccttacaggc  
120

ctgcattaca ggacttaaac acatagtgga agaggagaac aataggctaa acgttggttct  
180

c g t c t c c c a c  
190

<210> 195  
<211> 210  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 195  
gtggaagagg agaacaatag gctaaacggtt gttctcgtct cccactgccg ccagctggat  
60

ttcccaattc tgagtaacac tctgcaatcc aaacaggggtt caaccctcca ccttacaggc  
120

ctgcattaca ggacttaaac acataatcca agaatttctt acactgtgga agaggagaac  
180

a a t a g g c t a a                      a c g t t g t t c t                      c g t c t c c c a c  
210

<210> 196  
<211> 110  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 196  
gtggaagagg agaacaatag gctaaacggtt gttctcgtct cccacatact gccgccagct  
60

ggattgtgga agaggagaac aataggctaa acgttggttct cgtctcccac  
110

<210> 197  
<211> 130  
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 197

gtggaagagg agaacaatag gctaaacggt gttctcgtct cccacactgt acagaatact  
60

gccgccagct ggatttccca attctgtgga agaggagaac aataggctaa acgttggtct  
120

c g t c t c c c a c  
130

<210> 198

<211> 150

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 198

gtggaagagg agaacaatag gctaaacggt gttctcgtct cccactcttg tgtctactgt  
60

acagaatact gccgccagct ggatttccca attctgagta acactgtgga agaggagaac  
120

a a t a g g c t a a a c g t t g t t c t c g t c t c c c a c  
150

<210> 199

<211> 170

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 199

gtggaagagg agaacaatag gctaaacggt gttctcgtct cccaccgtac ataattcttg  
60

tgtctactgt acagaatact gccgccagct ggatttccca attctgagta acactctgca  
120

atccagtgga agaggagAAC aataggctaa acgttggttct cgtctccac  
170

<210> 200

<211> 190

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 200

gtggaagagg agaacaatag gctaaacggt gttctcgtct cccactgata aaaggcgtac  
60

ataattcttg tgtctactgt acagaatact gccgccagct ggatttccca attctgagta  
120

acactctgca atccaaacag ggttcgtgga agaggagAAC aataggctaa acgttggttct  
180

c g t c t c c c a c  
190

<210> 201

<211> 210

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<400> 201



gtggaagagg agaacaatag gctaaacgtt gttctcgtct cccaccttaa gtctttgata  
60

aaaggcgtag ataattcttg tgtctactgt acagaatact gccgccagct ggatttccca  
120

attctgagta acactctgca atccaaacag ggttcaaccc tccacgtgga agaggagaac  
180

a a t a g g c t a a                      a c g t t g t t c t                      c g t c t c c c a c  
210

<210> 202

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 202

a c a a a c c g a g                                      c g g t g t c t g t  
20

<210> 203

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 203

g c c a t t c c a t                                      t g c t c t t t c a  
20

<210> 204

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 204

t g c c g c c a g c  
20

t g g a t t t c c c

<210> 205

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 205

c t g t a c c a g c  
20

c a g t c a a t t a

<210> 206

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 206

c t t c t c c a c a  
20

g c c c g a a g c a

<210> 207

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 207

c t c c t c c a c c  
20

c g a c c c c g g g

<210> 208

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 208

g g g t g c c a a g  
20

c a g t t g g t g g

<210> 209

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 209

c t t g t c c a c c  
20

t t g a t g c c c a

<210> 210

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 210

t t c a t c c a a t  
20

g g c t g g t t a t

<210> 211

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 211

c a a g g c c a a g  
20

g g c t c g c c a g

<210> 212

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 212

t c c a a c c a c c  
20

a c a a g t t t a t

<210> 213

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 213

tacagaatac    tgccgccagc    tggatttccc    aattctgagt    aacactctgc  
50

<210> 214  
<211> 50  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 214  
gaaaagtttt acaaaccgag cggtgtctgt gagactttca ttcacaccca  
50

<210> 215  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 215  
a t a a t t t c t a t t a t a t t a c a  
20

<210> 216  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 216  
a t t t c a g g t a a g c c g a g g t t  
20

<210> 217  
<211> 20

<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 217

a t a c t g c c g c  
20

c a g c t g g a t t

<210> 218

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 218

tgccgccagc tggatttccc aattctgagt aacactctgc aatccaaaca gggttcaacc  
60

<210> 219

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 219

tcttgtgtct actgtacaga atactgccgc cagctggatt tcccaattct gagtaacact  
60

<210> 220

<211> 100

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 220

tgccgccagc tggatttccc aattctgagt aacactctgc aatccaaaca gggttcaacc  
60

ctccacctta caggcctgca ttacaggact taaacacata  
100

<210> 221

<211> 100

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 221

tgataaaagg cgtacataat tcttgtgtct actgtacaga atactgccgc cagctggatt  
60

tcccaattct gagtaacact ctgcaatcca aacaggggttc  
100

<210> 222

<211> 100

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 222

tgaattagct gtatcgtcaa ggcactcttg cctacgccac cagctccaac caccacaagt  
60

ttatatattcag tcatttttcag caggcctctc tcccgcacct  
100

<210> 223  
<211> 100  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 223  
atcaaaaaaa taaactctac caaggggtgac ggaagtctct acagcaaggc caagggctcg  
60

ccagacggcg aacatcaggg gtgcatgggtg ggcactgccc  
100

<210> 224  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic primer

<400> 224  
c t c t c t g t a c c t t a t c t t a g t g t t a c t g a  
29

<210> 225  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic primer

<400> 225  
a t t t c t g g c a t a t t t c t g a a g g t g  
24



<210> 226  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 226  
a c c c t t c c t t t c t t a c c a c a c a  
22

<210> 227  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 227  
c a g g g t g t c c a g a t c t g a t t g t t  
23

<210> 228  
<211> 29  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 228  
c t t c t c t t t t a a a c t a a c c c a t c a g a g t t  
29

<210> 229  
<211> 27

<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 229

c a a g c a g t c a                      g a c c a a a a t a                      c c t a c t g  
27

<210> 230

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 230

t c t t a g c a t g                      c t t c g a t g t g                      g c a t a  
25

<210> 231

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 231

c a t c a a c a a g                                              c c a g g g c c t g  
20

<210> 232

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 232

g a a g a g g a a a                      t g t c c g t c t c                      c a c  
23

<210> 233

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 233

a g g c c t g t a a                                      g g t g g a g g g  
19

<210> 234

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 234

t g a a a t a a c g                      g c a a t t t a t c                      c a t t g c a c a t                      a c  
32

<210> 235

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 235

g g g a g c a g c a  
19

t g g a g c c t t

<210> 236

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
primer

<400> 236

t c c g a c c g t a  
22

a c t a t t c g g t

g c

<210> 237

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
primer

<400> 237

t g g g t g t g a a  
23

c c a t g a g a a g

t a t

<210> 238

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
primer

<400> 238

t g g c a t g g a c  
20

t g t g g t c a t g

<210> 239  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 239  
c c t a a c t t a t                      t g c c t g g g c a                      g t g  
23

<210> 240  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 240  
g c a t c a g c a g                      t a t c t t a g c c                      a t c a a  
25

<210> 241  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic  
primer

<400> 241  
c a g a g g c t c a                                      g c g g c t c c  
18

<210> 242

<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
primer

<400> 242  
t a g c t g t a t c                      g t c a a g g c a c                      t c  
22

<210> 243  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
primer

<400> 243  
c a c a c c t g t c                      t g t g c a c t t g                      t a  
22

<210> 244  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
primer

<400> 244  
c g g t c c a c a g                                      c t c a g g a a c c  
20

<210> 245  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 245

a c c a g a a g g c  
20

g g a t g a t g g g

<210> 246

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 246

c t c a g a c a g c  
20

c c a t c c a a c c

<210> 247

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 247

t a c t a c t t g c  
27

t t c c t g t a g g

a a t c c t c

<210> 248

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

primer

<400> 248

a g c c c t g c t g  
21

c t t c c t a a c t

t

<210> 249

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
primer

<400> 249

a c c c t a g t t t  
26

a t t t c a g c a t

c a g c a g

<210> 250

<211> 100

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
polynucleotide

<220>

<221> modified\_base

<222> (1)..(50)

<223> a, c, t, g, unknown or other

<220>

<221> modified\_base

<222> (52)..(100)

<223> a, c, t, g, unknown or other

<400> 250

nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn nnnnnnnnnnnn cnnnnnnnnnn  
60



nnnnnnnnnnnn      nnnnnnnnnnnn      nnnnnnnnnnnn      nnnnnnnnnnnn  
100

<210> 251  
<211> 190  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Synthetic polynucleotide

<220>  
<221> modified\_base  
<222> (46)..(95)  
<223> a, c, t, g, unknown or other

<220>  
<221> modified\_base  
<222> (97)..(145)  
<223> a, c, t, g, unknown or other

<400> 251  
gtggaatagt ataacaatat gctaaatgtt gttatagtat cccacnnnnn nnnnnnnnnn  
60

nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnncnnnn nnnnnnnnnnn nnnnnnnnnnn  
120

nnnnnnnnnnn nnnnnnnnnnn nnnnngtgga atagtataac aatatgctaa atgttggtat  
180

a      g      t      a      t      c      c      c      a      c  
190

<210> 252  
<211> 46  
<212> DNA  
<213> Unknown

<220>  
<223> Description of Unknown:

target sequence

<400> 252

tgaatgaaag      tctcacagac      accgctcagt      ttgtaaaact      tttctt  
46

<210> 253

<211> 24

<212> DNA

<213> Unknown

<220>

<223> Description of Unknown:  
target sequence

<400> 253

t c a c a g a c a c                      c g c t c a g t t t                      g t a a  
24

<210> 254

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 254

c a g a c a c c g c                      t c a g t t t g t a                      a a a c t  
25

<210> 255

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic  
oligonucleotide

<400> 255

g g a a a t c c a g  
25

c t a g c g g c a g

t a t t c