

CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING USHER SYNDROME AND RETINITIS PIGMENTOSA

REFERENCE TO RELATED APPLICATIONS

5 The present application claims the benefit of U.S. Provisional Application No. 61/948,520, filed March 5, 2014, the contents of which are hereby incorporated by reference in their entirety.

FIELD OF THE INVENTION

10 The invention relates to CRISPR/Cas-related methods and components for editing of a target nucleic acid sequence, and applications thereof in connection with Usher syndrome and retinitis pigmentosa.

BACKGROUND

15 Usher Syndrome is a common form of inherited combined hearing and vision loss. It affects 1 in 6,000 individuals (Kimberling *et al.*, Genetics in Medicine 2010; 12(8): 512-516). Usher Syndrome is known to be caused by mutations in at least 9 different genes. Usher syndrome type IIA is caused by mutations in the *USH2A* gene (also known as the RP39 gene). Usher syndrome type II accounts for approximately 50% of all Usher cases (Eudy *et al.*, Science
20 1998; 280(5370):1753-1757). Usher syndrome type IIA accounts for approximately 80% of all Usher type II cases (Le Quesne Stabel *et al.*, Journal of Molecular Genetics 2012; 49(1):27-36), or 40% of all Usher cases.

 The *USH2A* gene is 800,503 base pairs and codes for the usherin protein (1,551 amino acids in length). A common mutation in subjects with Usher syndrome type II or non-syndromic
25 retinitis pigmentosa (RP39) is a single nucleotide deletion, e.g., a guanine deletion, at nucleotide position c.2299 (2299delG) in the *USH2A* gene, which is responsible for between 15% and 78% of *USH2A* mutations, depending on the population (Baux *et al.* European Journal of Human Genetics 2010; 18:788-793. Yan *et al.*, Journal of Human Genetics 2009; 54:732-738. Weston *et al.*, American Journal of Human Genetics 2000; 66(4):1199-1210). The deletion of guanine at
30 position 2299 results in a premature stop codon, which leads to a truncated usherin protein. The truncated usherin protein disrupts vision and hearing, leading to visual and hearing loss.

Visual loss in Usher syndrome usually begins between the ages of 10 and 20. The vision loss is described as retinitis pigmentosa (RP), a retinal dystrophy that tends to affect peripheral visual fields initially. The visual field defect generally progresses inwards, constricting the subject's visual field and over time leading to blindness. Subjects commonly experience loss of night vision early in the disease, followed by loss of peripheral vision, followed by loss of visual acuity (a measure of the central visual field).

The visual loss associated with Usher syndrome type II is called 'syndromic' retinitis pigmentosa, because it is frequently associated with hearing loss. Alternatively, patients can have mutations in *USH2A* that are not associated with hearing loss. In this case, the patients are defined as having 'non-syndromic' retinitis pigmentosa. Non-syndromic retinitis pigmentosa caused by mutations in the *USH2A* gene may be called retinitis pigmentosa 39, or RP39.

Usher syndrome also causes deafness. In Usher syndrome type IIA, the age of onset of deafness is most often at birth and consists of moderate to severe hearing impairment which is generally non-progressive. However, in subjects with Usher type IIA, hearing loss may present after birth into teenage years and may be progressive. Usher syndrome type IIA subjects have normal vestibular function. Usher type I subjects are generally born profoundly deaf with absent vestibular function.

Treatment for the visual loss associated with Usher syndrome type IIA and/or RP-39 is limited. There is currently no approved treatment that substantially reverses or halts the progression of disease in Usher syndrome type 2 or in RP-39. Vitamin A supplementation may delay onset of disease and slow progression. An electrical implant known as the Argus II retinal implant was recently approved for use, but it only offers minimal improvement in vision in patients with RP. The best visual acuity achieved in trials by the device was 20/1260 (legal blindness is defined as 20/200 vision). In addition, current gene therapy delivery techniques are not able to deliver genes encoding large proteins, e.g., the *USH2A* gene.

There is also no curative treatment for hearing loss in Usher syndrome type IIA. Subjects with Usher syndrome commonly use hearing aids and cochlear implants. Both are helpful in providing some degree of auditory function but do not restore hearing. Subjects would benefit greatly from a therapeutic which restored hearing and/or prevented further hearing loss.

Despite advances that have been made in gene therapy and by using cochlear implants, there remains a need for therapeutics to treat the visual loss and deafness associated with Usher syndrome, including Usher syndrome type IIA, and retinitis pigmentosa.

5

SUMMARY OF THE INVENTION

Methods and compositions discussed herein, allow the correction of genetic disorders of the eye and the inner ear, e.g., disorders that affect retinal cells (e.g., photoreceptor cells), cells of the inner ear (e.g., inner hair cells or outer hair cells), or both.

Methods and compositions discussed herein, provide for treating or delaying the onset or
10 progression of Usher syndrome and retinitis pigmentosa, e.g., Usher Syndrome type IIA (USH2A,USHIIA) and retinitis pigmentosa 39 (RP39). Symptoms associated with Usher syndrome and retinitis pigmentosa, such as vision loss and hearing loss, can also be treated by the methods and compositions disclosed herein.

Methods and compositions discussed herein, provide for treating or delaying the onset or
15 progression of a disorder caused by mutations in the *USH2A* gene, including the mutation 2299delG (which causes a premature termination codon).

Methods and compositions discussed herein, provide for treating or delaying the onset or progression of usher syndrome and retinitis pigmentosa, e.g., Usher Syndrome type IIA (USH2A,USHIIA) and retinitis pigmentosa 39 (RP39) by gene editing, e.g., using CRISPR-
20 Cas9 mediated methods to correct the guanine deletion at position 2299 in the *USH2A* gene (e.g., replace the deleted guanine residue at position 2299 in the *USH2A* gene).

In one aspect, disclosed herein is a gRNA molecule, e.g., an isolated or non-naturally occurring gRNA molecule, comprising a targeting domain which is complementary with a target domain from the *USH2A* gene. *USH2A* is also known as *US2*, *RP39*, *USH2*, and *dJ1111A8.1*.

25 In an embodiment, the targeting domain is configured to provide a cleavage event, e.g., a double strand break or a single strand break, within 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 300 nucleotides of a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene.

In an embodiment, the targeting domain is configured to provide a cleavage event, e.g., a
30 double strand break or a single strand break, within 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 300 nucleotides of a target position in the *USH2A* gene, e.g., a deletion of guanine at

nucleotide position 2299 (2299delG). In an embodiment, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 1**. In some embodiments, the targeting domain is selected from those in **Table 1**. For example, in certain embodiments, the targeting domain is

5 GAGUGCAAAAAAGAAGCCAA;
 GUUAGAUGUCACCAAUUGUA;
 GGUGUCACACUGAAGUCCUU;
 GCCAUGGAGGUUACACUGGC;
 GUCACAGGCCUUACAAU;
 10 GUCACACUGAAGUCCUU;
 UGCAAAAAAGAAGCCAA;
 UGCAGAGAAAACUUUUA;
 UGUUCACUGAGCCAUGG; or
 AUGGAGGUUACACUGGC.

15 In other embodiments, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 2**. In an embodiment, the targeting domain is selected from **Table 2**.

In other embodiments, the targeting domain comprises a sequence that is the same as, or
 20 differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 3**. In an embodiment, the targeting domain is selected from **Table 3**.

In other embodiments, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 4A-4E**. In an embodiment, the targeting domain is selected from **Tables 4A-4E**.

25 In certain embodiments, the targeting domain is

GCAAGCCCAAUGUUGAA;
 GCAUUACAGACAGUCCC;
 GUCACACUGAAGUCCUU;
 GUCACAGGCCUUACAAU;
 GUCUGUAAUGCUAAGAC;
 GACACAGCUGGAUCCCUCCC;
 GAGACAGUGCAAUAAAUGUU;
 GCACUACACUGCCCAGAGUG;
 GCACUGUCUCCCUUCAACAU;
 GCCAUGGAGGUUACACUGGC;
 GCCUGUGACUGUGACACAGC;
 GGUGUCACACUGAAGUCCUU; or
 GUUAGAUGUCACCAAUUGUA.

In other embodiments, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 5A-5F**. In an embodiment, the targeting domain is selected from **Tables 5A-5F**.

In certain embodiments, the targeting domain is

GCACUACACUGCCCAGAGU;
GCCUGUGACUGUGACACAG;
GGCCUGUGACUGUGACACAG;
GGUGUGAUCAUUGCAAUU;
GACACCUGCAGAGAAAACUUUU;
GCAUUACAGACAGUCCCAGGG;
GCUUAGGUGUGAUCAUUGCAAUU;
GCUUCUUUUUUGCACUACACUGCC;
GGCUUAGGUGUGAUCAUUGCAAUU;
GUAAGGCCUGUGACUGUGACACAG; or
GUGACACCUGCAGAGAAAACUUUU.

5

In other embodiments, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 6A-6D**. In an embodiment, the targeting domain is selected from **Tables 6A-6D**.

In certain embodiments, the targeting domain is

GUGUCACACUGAAGUCC;
GGUGUGAUCAUUGCAAU; or
GGGCUCACAUCCAACAUCAU.

10 In an embodiment, the gRNA, e.g., a gRNA comprising a targeting domain which is complementary with a target domain from the *USH2A* gene, is a modular gRNA. In other embodiments, the gRNA is a chimeric gRNA.

In an embodiment, when two gRNAs are used to position two breaks, e.g., two single strand breaks, in the target nucleic acid sequence, each guide RNA is independently selected
15 from one or more of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**.

In an embodiment, the targeting domain which is complementary with a target domain from the *USH2A* gene target position in the *USH2A* gene is 16 nucleotides or more in length. In an embodiment, the targeting domain is 16 nucleotides in length. In an embodiment, the targeting domain is 17 nucleotides in length. In other embodiments, the targeting domain is 18

nucleotides in length. In still other embodiments, the targeting domain is 19 nucleotides in length. In still other embodiments, the targeting domain is 20 nucleotides in length. In an embodiment, the targeting domain is 21 nucleotides in length. In an embodiment, the targeting domain is 22 nucleotides in length. In an embodiment, the targeting domain is 23 nucleotides in length. In an embodiment, the targeting domain is 24 nucleotides in length. In an embodiment, the targeting domain is 25 nucleotides in length. In an embodiment, the targeting domain is 26 nucleotides in length.

In an embodiment, the targeting domain comprises 16 nucleotides.

In an embodiment, the targeting domain comprises 17 nucleotides.

In an embodiment, the targeting domain comprises 18 nucleotides.

In an embodiment, the targeting domain comprises 19 nucleotides.

In an embodiment, the targeting domain comprises 20 nucleotides.

In an embodiment, the targeting domain comprises 21 nucleotides.

In an embodiment, the targeting domain comprises 22 nucleotides.

In an embodiment, the targeting domain comprises 23 nucleotides.

In an embodiment, the targeting domain comprises 24 nucleotides.

In an embodiment, the targeting domain comprises 25 nucleotides.

In an embodiment, the targeting domain comprises 26 nucleotides.

A gRNA as described herein may comprise from 5' to 3': a targeting domain (comprising a "core domain", and optionally a "secondary domain"); a first complementarity domain; a linking domain; a second complementarity domain; a proximal domain; and a tail domain. In some embodiments, the proximal domain and tail domain are taken together as a single domain.

In an embodiment, a gRNA comprises a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 20 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In another embodiment, a gRNA comprises a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 30 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In another embodiment, a gRNA comprises a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 30 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

5 In another embodiment, a gRNA comprises a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 40 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

10 A cleavage event, e.g., a double strand or single strand break, is generated by a Cas9 molecule. The Cas9 molecule may be an enzymatically active Cas9 (eaCas9) molecule, e.g., an eaCas9 molecule that forms a double strand break in a target nucleic acid or an eaCas9 molecule forms a single strand break in a target nucleic acid (e.g., a nickase molecule).

In an embodiment, the eaCas9 molecule catalyzes a double strand break.

15 In some embodiments, the eaCas9 molecule comprises HNH-like domain cleavage activity but has no, or no significant, N-terminal RuvC-like domain cleavage activity. In this case, the eaCas9 molecule is an HNH-like domain nickase, e.g., the eaCas9 molecule comprises a mutation at D10, e.g., D10A. In other embodiments, the eaCas9 molecule comprises N-terminal RuvC-like domain cleavage activity but has no, or no significant, HNH-like domain cleavage activity. In an embodiment, the eaCas9 molecule is an N-terminal RuvC-like domain
20 nickase, e.g., the eaCas9 molecule comprises a mutation at H840, e.g., H840A. In an embodiment, the eaCas9 molecule is an N-terminal RuvC-like domain nickase, e.g., the eaCas9 molecule comprises a mutation at H863, e.g., H863A.

25 In an embodiment, a single strand break is formed in the strand of the target nucleic acid to which the targeting domain of said gRNA is complementary. In another embodiment, a single strand break is formed in the strand of the target nucleic acid other than the strand to which the targeting domain of said gRNA is complementary.

30 In another aspect, disclosed herein is a nucleic acid, e.g., an isolated or non-naturally occurring nucleic acid, e.g., DNA, that comprises (a) a sequence that encodes a gRNA molecule comprising a targeting domain that is complementary with a target domain in *USH2A* gene as disclosed herein.

In an embodiment, the nucleic acid encodes a gRNA molecule, e.g., the first gRNA molecule, comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any one of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. In an embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain that is selected from those in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**.

In an embodiment, a nucleic acid encodes a gRNA comprising from 5' to 3': a targeting domain (comprising a "core domain", and optionally a "secondary domain"); a first complementarity domain; a linking domain; a second complementarity domain; a proximal domain; and a tail domain. In some embodiments, the proximal domain and tail domain are taken together as a single domain.

In an embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain is configured to provide a cleavage event, e.g., a double strand break or a single strand break, within 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 300 nucleotides of a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene.

In an embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain is configured to provide a cleavage event, e.g., a double strand break or a single strand break, within 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 300 nucleotides of a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG). In an embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 1**. In an embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain is selected from those in **Table 1**. For example, in certain embodiments, the targeting domain is

GAGUGCAAAAAGAAGCCAA;
 GUUAGAUGUCACCAAUUGUA;
 GGUGUCACACUGAAGUCCUU;
 GCCAUGGAGGUUACACUGGC;
 GUCACAGGCCUUACAAU;
 GUCACACUGAAGUCCUU;
 UGCAAAAAGAAGCCAA;
 UGCAGAGAAAACUUUUA;
 UGUUCACUGAGCCAUGG; or

AUGGAGGUUACACUGGC.

In another embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 2**. In an embodiment, the targeting domain is selected from **Table 2**.

In another embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 3**. In an embodiment, the targeting domain is selected from **Table 3**.

In an embodiment, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 4A-4E**. In an embodiment, the targeting domain is selected from **Tables 4A-4E**.

In certain embodiments, the targeting domain is

GCAAGCCCAAUGUUGAA;
 GCAUUACAGACAGUCCC;
 GUCACACUGAAGUCCUU;
 GUCACAGGCCUUACAAU;
 GUCUGUAAUGCUAAGAC;
 GACACAGCUGGAUCCCUCCC;
 GAGACAGUGCAAUAAAUGUU;
 GCACUACACUGCCCAGAGUG;
 GCACUGUCUCCCUUCAACAU;
 GCCAUGGAGGUUACACUGGC;
 GCCUGUGACUGUGACACAGC;
 GGUGUCACACUGAAGUCCUU; or
 GUUAGAUGUCACCAAUUGUA.

In another embodiment, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 5A-5F**. In an embodiment, the targeting domain is selected from **Tables 5A-5F**.

In certain embodiments, the targeting domain is

GCACUACACUGCCCAGAGU;
 GCCUGUGACUGUGACACAG;
 GGCCUGUGACUGUGACACAG;
 GGUGUGAUCAUUGCAAUU;

GACACCUGCAGAGAAAACUUUU;
 GCAUUACAGACAGUCCCAGGG;
 GCUUAGGUGUGAUCAUUGCAAUU;
 GCUUCUUUUUUGCACUACACUGCC;
 GGCUUAGGUGUGAUCAUUGCAAUU;
 GUAAGGCCUGUGACUGUGACACAG; or
 GUGACACCUGCAGAGAAAACUUUU.

In yet another embodiment, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 6A-6D**. In an embodiment, the targeting domain is selected from **Tables 6A-6D**.

5 In certain embodiments, the targeting domain is

GUGUCACACUGAAGUCC;
 GGUGUGAUCAUUGCAAU; or
 GGGCUCACAUCCAACAUCAU.

In an embodiment, the nucleic acid encodes a modular gRNA, e.g., one or more nucleic acids encode a modular gRNA. In other embodiments, the nucleic acid encodes a chimeric gRNA. The nucleic acid may encode a gRNA, e.g., the first gRNA molecule, comprising a
 10 targeting domain comprising 16 nucleotides or more in length. In one embodiment, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 16 nucleotides in length. In other embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 17 nucleotides in length. In still other
 15 embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 18 nucleotides in length. In still other embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 19 nucleotides in length. In still other embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 20 nucleotides in length. In still other
 20 embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 21 nucleotides in length. In still other embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 22 nucleotides in length. In still other embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 23 nucleotides in length. In still other
 embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a

targeting domain that is 24 nucleotides in length. In still other embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 25 nucleotides in length. In still other embodiments, the nucleic acid encodes a gRNA, e.g., the first gRNA molecule, comprising a targeting domain that is 26 nucleotides in length.

5 In an embodiment, a nucleic acid encodes a gRNA comprising from 5' to 3': a targeting domain (comprising a "core domain", and optionally a "secondary domain"); a first complementarity domain; a linking domain; a second complementarity domain; a proximal domain; and a tail domain. In some embodiments, the proximal domain and tail domain are taken together as a single domain.

10 In an embodiment, a nucleic acid encodes a gRNA comprising a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 20 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

15 In an embodiment, a nucleic acid encodes a gRNA comprising a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 30 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

20 In an embodiment, a nucleic acid encodes a gRNA comprising a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 30 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

25 In an embodiment, a nucleic acid encodes a gRNA comprising a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 40 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In an embodiment, a nucleic acid comprises (a) a sequence that encodes a gRNA molecule comprising a targeting domain that is complementary with a target domain in the *USH2A* gene as disclosed herein, and further comprising (b) a sequence that encodes a Cas9 molecule.

30 The Cas9 molecule may be a nickase molecule, a enzymatically activating Cas9 (eaCas9) molecule, e.g., an eaCas9 molecule that forms a double strand break in a target nucleic acid and

an eaCas9 molecule forms a single strand break in a target nucleic acid. In an embodiment, a single strand break is formed in the strand of the target nucleic acid to which the targeting domain of said gRNA is complementary. In another embodiment, a single strand break is formed in the strand of the target nucleic acid other than the strand to which the targeting domain of said gRNA is complementary.

In an embodiment, the eaCas9 molecule catalyzes a double strand break.

In some embodiments, the eaCas9 molecule comprises HNH-like domain cleavage activity but has no, or no significant, N-terminal RuvC-like domain cleavage activity. In other embodiments, the said eaCas9 molecule is an HNH-like domain nickase, e.g., the eaCas9 molecule comprises a mutation at D10, e.g., D10A. In other embodiments, the eaCas9 molecule comprises N-terminal RuvC-like domain cleavage activity but has no, or no significant, HNH-like domain cleavage activity. In another embodiment, the eaCas9 molecule is an N-terminal RuvC-like domain nickase, e.g., the eaCas9 molecule comprises a mutation at H840, e.g., H840A. In another embodiment, the eaCas9 molecule is an N-terminal RuvC-like domain nickase, e.g., the eaCas9 molecule comprises a mutation at H863, e.g., H863A.

A nucleic acid disclosed herein may comprise (a) a sequence that encodes a gRNA molecule comprising a targeting domain that is complementary with a target domain in the *USH2A* gene as disclosed herein; and (b) a sequence that encodes a Cas9 molecule.

A nucleic acid disclosed herein may comprise (a) a sequence that encodes a gRNA molecule comprising a targeting domain that is complementary with a target domain in the *USH2A* gene as disclosed herein; (b) a sequence that encodes a Cas9 molecule; and further may comprises (c)(i) a sequence that encodes a second gRNA molecule described herein having a targeting domain that is complementary to a second target domain of the *USH2A* gene, and optionally, (c)(ii) a sequence that encodes a third gRNA molecule described herein having a targeting domain that is complementary to a third target domain of the *USH2A* gene; and optionally, (c)(iii) a sequence that encodes a fourth gRNA molecule described herein having a targeting domain that is complementary to a fourth target domain of the *USH2A* gene. In an embodiment, a nucleic acid encoding a second gRNA molecule comprising a targeting domain is configured to provide a cleavage event, e.g., a double strand break or a single strand break, within 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 300 nucleotides of a target position in

the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene.

In an embodiment, a nucleic acid encodes a second gRNA molecule comprising a targeting domain configured to provide a cleavage event, e.g., a double strand break or a single strand break, sufficiently close to the target position in the *USH2A* gene to allow alteration, either alone or in combination with the break positioned by the first gRNA molecule.

In an embodiment, a nucleic acid encodes a third gRNA molecule comprising a targeting domain configured to provide a cleavage event, e.g., a double strand break or a single strand break, sufficiently close to the target position in the *USH2A* gene to allow alteration, either alone or in combination with the break positioned by the first and/or second gRNA molecule.

In an embodiment, a nucleic acid encodes a fourth gRNA molecule comprising a targeting domain configured to provide a cleavage event, e.g., a double strand break or a single strand break, sufficiently close to the target position in the *USH2A* gene to allow alteration, either alone or in combination with the break positioned by the first gRNA molecule, the second gRNA molecule and/or the third gRNA molecule.

In an embodiment, a nucleic acid encodes a second gRNA molecule comprising a targeting domain configured to provide a cleavage event, e.g., a double strand break or a single strand break, in combination with the break position by said first gRNA molecule, sufficiently close to the target position in the *USH2A* gene to allow alteration of the target position, either alone or in combination with the break positioned by said first gRNA molecule.

In an embodiment, a nucleic acid encodes a third gRNA molecule comprising a targeting domain configured to provide a cleavage event, e.g., a double strand break or a single strand break, in combination with the break position by said first and/or second gRNA molecule, sufficiently close to the target position in the *USH2A* gene to allow alteration, either alone or in combination with the break positioned by the first and/or second gRNA molecule.

In an embodiment, a nucleic acid encodes a fourth gRNA molecule comprising a targeting domain configured to provide a cleavage event, e.g., a double strand break or a single strand break, in combination with the break positioned by the first gRNA molecule, the second gRNA molecule and/or the third gRNA molecule, sufficiently close to the target position in the *USH2A* gene to allow alteration, either alone or in combination with the break positioned by the first gRNA molecule, the second gRNA molecule and/or the third gRNA molecule.

In an embodiment, a nucleic acid encoding a second gRNA molecule comprising a targeting domain is configured to provide a cleavage event, e.g., a double strand break or a single strand break, within 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 300 nucleotides of a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG). In an embodiment, the nucleic acid encodes a second gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 amino acids from, a targeting domain sequence from **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. In an embodiment, the nucleic acid encodes a second gRNA molecule comprising a targeting domain is selected from those in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. For example, in certain embodiments, the targeting domain is

GAGUGCAAAAAAGAAGCCAA;
 GUUAGAUGUCACCAAUUGUA;
 GGUGUCACACUGAAGUCCUU;
 GCCAUGGAGGUUACACUGGC;
 GUCACAGGCCUUACAAU;
 GUCACACUGAAGUCCUU;
 UGCAAAAAAGAAGCCAA;
 UGCAGAGAAAACUUUUA;
 UGUUCACUGAGCCAUGG; or
 AUGGAGGUUACACUGGC.

In certain embodiments, the targeting domain is

GCAAGCCCCAAUGUUGAA;
 GCAUUACAGACAGUCCC;
 GUCACACUGAAGUCCUU;
 GUCACAGGCCUUACAAU;
 GUCUGUAAUGCUAAGAC;
 GACACAGCUGGAUCCCUC;
 GAGACAGUGCAAUAAAUGUU;
 GCACUACACUGCCCAGAGUG;
 GCACUGUCUCCCUUCAACAU;
 GCCAUGGAGGUUACACUGGC;
 GCCUGUGACUGUGACACAGC;
 GGUGUCACACUGAAGUCCUU; or
 GUUAGAUGUCACCAAUUGUA.

In certain embodiments, the targeting domain is

GCACUACACUGCCCAGAGU;
 GCCUGUGACUGUGACACAG;

GGCCUGUGACUGUGACACAG;
 GGUGUGAUCAUUGCAAUU;
 GACACCUGCAGAGAAAACUUUU;
 GCAUUACAGACAGUCCCAGGG;
 GCUUAGGUGUGAUCAUUGCAAUU;
 GCUUCUUUUUUGCACUACACUGCC;
 GGCUUAGGUGUGAUCAUUGCAAUU;
 GUAAGGCCUGUGACUGUGACACAG; or
 GUGACACCUGCAGAGAAAACUUUU.

In certain embodiments, the targeting domain is

GUGUCACACUGAAGUCC;
 GGUGUGAUCAUUGCAAU; or
 GGGCUCACAUCCAACAUCAU.

In an embodiment, a nucleic acid encoding a third gRNA molecule comprising a
 5 targeting domain is configured to provide a cleavage event, e.g., a double strand break
 or a single strand break, within 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 300
 nucleotides of a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide
 position 2299 (2299delG). In an embodiment, the nucleic acid encodes a second gRNA molecule
 comprising a targeting domain comprising a sequence that is the same as, or differs by no more
 10 than 1, 2, 3, 4, or 5 amino acids from, a targeting domain sequence from **Tables 1-3, 4A-4E, 5A-
 5F, or 6A-6D**. In an embodiment, the nucleic acid encodes a third gRNA molecule comprising a
 targeting domain is selected from those in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. For example,
 in certain embodiments, the targeting domain is

15 GAGUGCAAAAAGAAGCCAA;
 GUUAGAUGUCACCAAUUGUA;
 GGUGUCACACUGAAGUCCUU;
 GCCAUGGAGGUUACACUGGC;
 GUCACAGGCCUUAACAAU;
 GUCACACUGAAGUCCUU;
 20 UGCAAAAAGAAGCCAA;
 UGCAGAGAAAACUUUUA;
 UGUUCACUGAGCCAUGG; or
 AUGGAGGUUACACUGGC.

25 In certain embodiments, the targeting domain is

GCAAGCCCAAUGUUGAA;
 GCAUUACAGACAGUCCC;

GUCACACUGAAGUCCUU;
 GUCACAGGCCUUACAAU;
 GUCUGUAAUGCUAAGAC;
 GACACAGCUGGAUCCCUCCC;
 GAGACAGUGCAAUAAAUGUU;
 GCACUACACUGCCCAGAGUG;
 GCACUGUCUCCCUUCAACAU;
 GCCAUGGAGGUUACACUGGC;
 GCCUGUGACUGUGACACAGC;
 GGUGUCACACUGAAGUCCUU; or
 GUUAGAUGUCACCAAUUGUA.

In certain embodiments, the targeting domain is

GCACUACACUGCCCAGAGU;
 GCCUGUGACUGUGACACAG;
 GGCCUGUGACUGUGACACAG;
 GGUGUGAUCAUUGCAAUU;
 GACACCUGCAGAGAAAACUUUU;
 GCAUUACAGACAGUCCCAGGG;
 GCUUAGGUGUGAUCAUUGCAAUU;
 GCUUCUUUUUUGCACUACACUGCC;
 GGCUUAGGUGUGAUCAUUGCAAUU;
 GUAAGGCCUGUGACUGUGACACAG; or
 GUGACACCUGCAGAGAAAACUUUU.

In certain embodiments, the targeting domain is

GUGUCACACUGAAGUCC;
 GGUGUGAUCAUUGCAAU; or
 GGGCUCACAUCCAACAUCAU.

5

In an embodiment, a nucleic acid encoding a fourth gRNA molecule comprising a targeting domain is configured to provide a cleavage event, e.g., a double strand break or a single strand break, within 10, 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, or 300 nucleotides of a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG). In an embodiment, the nucleic acid encodes a fourth gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 amino acids from, a targeting domain sequence from **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. In an embodiment, the nucleic acid encodes a second gRNA molecule comprising

10

a targeting domain is selected from those in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. For example, in certain embodiments, the targeting domain is

5 GAGUGCAAAAAAGAAGCCAA;
 GUUAGAUGUCACCAAUUGUA;
 GGUGUCACACUGAAGUCCUU;
 GCCAUGGAGGUUACACUGGC;
 GUCACAGGCCUUACAAU;
 GUCACACUGAAGUCCUU;
 10 UGCAAAAAAGAAGCCAA;
 UGCAGAGAAAACUUUUA;
 UGUUCACUGAGCCAUGG; or
 AUGGAGGUUACACUGGC.

In certain embodiments, the targeting domain is

GCAAGCCCCAAUGUUGAA;
 GCAUUACAGACAGUCCC;
 GUCACACUGAAGUCCUU;
 GUCACAGGCCUUACAAU;
 GUCUGUAAUGCUAAGAC;
 GACACAGCUGGAUCCCUCCC;
 GAGACAGUGCAAUAAAUGUU;
 GCACUACACUGCCCAGAGUG;
 GCACUGUCUCCCUUCAACAU;
 GCCAUGGAGGUUACACUGGC;
 GCCUGUGACUGUGACACAGC;
 GGUGUCACACUGAAGUCCUU; or
 15 GUUAGAUGUCACCAAUUGUA.

In certain embodiments, the targeting domain is

GCACUACACUGCCCAGAGU;
 GCCUGUGACUGUGACACAG;
 GGCCUGUGACUGUGACACAG;
 GGUGUGAUCAUUGCAAUU;
 GACACCUGCAGAGAAAACUUUU;
 GCAUUACAGACAGUCCCAGGG;
 GCUUAGGUGUGAUCAUUGCAAUU;
 GCUUCUUUUUUGCACUACACUGCC;
 GGCUUAGGUGUGAUCAUUGCAAUU;
 GUAAGGCCUGUGACUGUGACACAG; or
 GUGACACCUGCAGAGAAAACUUUU.

In certain embodiments, the targeting domain is

GUGUCACACUGAAGUCC;
GGUGUGAUCAUUGCAAU; or
GGGCUCAUCAUCCAACAUCAU.

In another embodiment, the nucleic acid encodes a second gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 1**. In an embodiment, the
5 targeting domain is selected from **Table 1**. In another embodiment, the nucleic acid encodes a third gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 1**. In an embodiment, the targeting domain is selected from **Table 1**. In another
10 embodiment, the nucleic acid encodes a fourth gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 1**. In an embodiment, the targeting domain is selected from **Table 1**.

In another embodiment, the nucleic acid encodes a second gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4,
15 or 5 nucleotides from, a targeting domain sequence from **Table 2**. In an embodiment, the targeting domain is selected from **Table 2**. In another embodiment, the nucleic acid encodes a third gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 2**. In an embodiment, the targeting domain is selected from **Table 2**. In another
20 embodiment, the nucleic acid encodes a fourth gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 2**. In an embodiment, the targeting domain is selected from **Table 2**.

In another embodiment, the nucleic acid encodes a second gRNA molecule comprising a
25 targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 3**. In an embodiment, the targeting domain is selected from **Table 3**. In another embodiment, the nucleic acid encodes a third gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from

Table 3. In an embodiment, the targeting domain is selected from **Table 3**. In another embodiment, the nucleic acid encodes a fourth gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Table 3**. In an embodiment, the targeting domain is selected from **Table 3**.

In an embodiment, the nucleic acid encodes a second gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 4A-4E**. In an embodiment, the targeting domain is selected from **Tables 4A-4E**. In another embodiment, the nucleic acid encodes a third gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 4A-4E**. In an embodiment, the targeting domain is selected from **Tables 4A-4E**. In yet another embodiment, the nucleic acid encodes a fourth gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 4A-4E**. In an embodiment, the targeting domain is selected from **Tables 4A-4E**.

In an embodiment, the nucleic acid encodes a second gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 5A-5F**. In an embodiment, the targeting domain is selected from **Tables 5A-5F**. In another embodiment, the nucleic acid encodes a third gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 5A-5F**. In an embodiment, the targeting domain is selected from **Tables 5A-5F**. In yet another embodiment, the nucleic acid encodes a third gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 5A-5F**. In an embodiment, the targeting domain is selected from **Tables 5A-5F**.

In an embodiment, the nucleic acid encodes a second gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 6A-6D**. In an embodiment, the targeting domain is selected from **Tables 6A-6D**. In another embodiment, the nucleic acid

encodes a second gRNA molecule comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 6A-6D**. In an embodiment, the targeting domain is selected from **Tables 6A-6D**. In yet another embodiment, the nucleic acid encodes a second gRNA molecule

5 comprising a targeting domain comprising a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from **Tables 6A-6D**. In an embodiment, the targeting domain is selected from **Tables 6A-6D**.

In an embodiment, the nucleic acid encodes a second gRNA which is a modular gRNA, e.g., wherein one or more nucleic acid molecules encode a modular gRNA. In another

10 embodiment, the nucleic acid encoding a second gRNA is a chimeric gRNA. In yet another embodiment, when a nucleic acid encodes a third or fourth gRNA, the third and fourth gRNA may be a modular gRNA or a chimeric gRNA. When multiple gRNAs are used, any combination of modular or chimeric gRNAs may be used.

A nucleic acid may encode a second, a third, and/or a fourth gRNA, each independently,

15 comprising a targeting domain comprising 16 nucleotides or more in length.. In an embodiment, the nucleic acid encodes a second gRNA comprising a targeting domain that is 16 nucleotides in length. In an embodiment, the nucleic acid encodes a second gRNA comprising a targeting domain that is 17 nucleotides in length. In other embodiments, the nucleic acid encodes a second gRNA comprising a targeting domain that is 18 nucleotides in length. In still other

20 embodiments, the nucleic acid encodes a second gRNA comprising a targeting domain that is 19 nucleotides in length. In still other embodiments, the nucleic acid encodes a second gRNA comprising a targeting domain that is 20 nucleotides in length. In still other embodiments, the nucleic acid encodes a second gRNA comprising a targeting domain that is 21 nucleotides in length. In still other embodiments, the nucleic acid encodes a second gRNA comprising a targeting domain that is 22 nucleotides in length. In still other embodiments, the nucleic acid

25 encodes a second gRNA comprising a targeting domain that is 23 nucleotides in length. In still other embodiments, the nucleic acid encodes a second gRNA comprising a targeting domain that is 24 nucleotides in length. In still other embodiments, the nucleic acid encodes a second gRNA comprising a targeting domain that is 25 nucleotides in length. In still other embodiments, the

30 nucleic acid encodes a second gRNA comprising a targeting domain that is 26 nucleotides in length.

In an embodiment, the targeting domain comprises 16 nucleotides.

In an embodiment, the targeting domain comprises 17 nucleotides.

In an embodiment, the targeting domain comprises 18 nucleotides.

In an embodiment, the targeting domain comprises 19 nucleotides.

5 In an embodiment, the targeting domain comprises 20 nucleotides.

In an embodiment, the targeting domain comprises 21 nucleotides.

In an embodiment, the targeting domain comprises 22 nucleotides.

In an embodiment, the targeting domain comprises 23 nucleotides.

In an embodiment, the targeting domain comprises 24 nucleotides.

10 In an embodiment, the targeting domain comprises 25 nucleotides.

In an embodiment, the targeting domain comprises 26 nucleotides.

In an embodiment, a nucleic acid encodes a second, a third, and/or a fourth gRNA, each independently, comprising from 5' to 3': a targeting domain (comprising a "core domain", and optionally a "secondary domain"); a first complementarity domain; a linking domain; a second
15 complementarity domain; a proximal domain; and a tail domain. In some embodiments, the proximal domain and tail domain are taken together as a single domain.

In an embodiment, a nucleic acid encodes a second gRNA comprising a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 20 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20,
20 21, 22, 23, 24, 25 or 26 nucleotides in length.

In an embodiment, a nucleic acid encodes a second gRNA comprising a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 30 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20,
25 21, 22, 23, 24, 25 or 26 nucleotides in length.

In an embodiment, a nucleic acid encodes a second gRNA comprising a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 30 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20,
30 21, 22, 23, 24, 25 or 26 nucleotides in length.

In an embodiment, a nucleic acid encodes a second gRNA comprising a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at

least 40 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In some embodiments, the nucleic acid encodes (a) a sequence that encodes a gRNA molecule comprising a targeting domain that is complementary with a target domain in the
5 *USH2A* gene as disclosed herein; (b) a sequence that encodes a Cas9 molecule; and further comprises (c)(i) a sequence that encodes a second gRNA molecule described herein having a targeting domain that is complementary to a second target domain of the *USH2A* gene, and optionally, (c)(ii) a sequence that encodes a third gRNA molecule described herein having a targeting domain that is complementary to a third target domain of the *USH2A* gene; and
10 optionally, (c)(iii) a sequence that encodes a fourth gRNA molecule described herein having a targeting domain that is complementary to a fourth target domain of the *USH2A* gene. . In some embodiments, the targeting domain of the gRNA molecule and the targeting domain of the second gRNA molecules are complementary to opposite strands of the target nucleic acid molecule. In some embodiments, the gRNA molecule and the second gRNA molecule are
15 configured such that the PAMs are oriented outward.

In some embodiments, the gRNA molecule and said second gRNA molecule are configured such that they do not overlap and are separated by as much as 50, 100, or 200 nucleotides. The gRNA and second gRNA may be configured such that single strand breaks are formed on each strand of the target nucleic acid. In an embodiment, the gRNA and the second
20 gRNA are configured such that single strand breaks are formed on each strand of the target nucleic acid and the single strand breaks are within 50-100 nucleotides of one another.

In an embodiment, the gRNA molecule and the second gRNA molecule are configured such that the first and second breaks are 5' to a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG). In another embodiment, the gRNA
25 molecule and the second gRNA molecule are configured such that the first and second breaks are 3' to a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG). In another embodiment, the gRNA molecule and said second gRNA molecule are configured such that the first and second breaks flank a target position in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG).

30 In some embodiments, the nucleic acid encodes (a) a sequence that encodes a gRNA molecule comprising a targeting domain that is complementary with a target domain in the

USH2A gene as disclosed herein; (b) a sequence that encodes a Cas9 molecule; (c) a sequence that encodes a second, third and/or fourth gRNA molecule described herein having a targeting domain that is complementary to a second target domain of the *USH2A* gene; and further comprising (d) a template nucleic acid. In an embodiment, the template nucleic acid is a single stranded nucleic acid. In another embodiment, the template nucleic acid is a double stranded nucleic acid. In some embodiments, the template nucleic acid comprises a nucleotide sequence, e.g., of one or more nucleotides, that will be added to or will template a change in the target nucleic acid. In other embodiments, the template nucleic acid comprises a nucleotide sequence that may be used to modify the target position. In other embodiments, the template nucleic acid comprises a nucleotide sequence, e.g., of one or more nucleotides, that corresponds to wildtype sequence of the target nucleic acid, e.g., of the target position.

The template nucleic acid may comprise a replacement sequence, e.g., a replacement sequence from the **Table 13**. In some embodiments, the template nucleic acid comprises a 5' homology arm, e.g., a 5' homology arm from **Table 13**. In other embodiments, the template nucleic acid comprises a 3' homology arm, e.g., a 3' homology arm from **Table 13**.

As described above, a nucleic acid may comprise (a) a sequence encoding a gRNA molecule comprising a targeting domain that is complementary with a target domain in *USH2A* gene, and (b) a sequence encoding a Cas9 molecule. In some embodiments, (a) and (b) are present on the same nucleic acid molecule, e.g., the same vector, e.g., the same viral vector, e.g., the same adeno-associated virus (AAV) vector. In an embodiment, the nucleic acid molecule is an AAV vector.

In other embodiments, (a) is present on a first nucleic acid molecule, e.g. a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (b) is present on a second nucleic acid molecule, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecules may be AAV vectors.

In other embodiments, the nucleic acid may further comprise (c) a sequence that encodes a second, third and/or fourth gRNA molecule as described herein. In some embodiments, the nucleic acid comprises (a), (b) and (c), but not (d), a template nucleic acid. Each of (a) and (c) may be present on the same nucleic acid molecule, e.g., the same vector, e.g., the same viral vector, e.g., the same adeno-associated virus (AAV) vector. In an embodiment, the nucleic acid molecule is an AAV vector.

In other embodiment, (a) and (c) are on different vectors. For example, (a) may be present on a first nucleic acid molecule, e.g. a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (c) may be present on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. In an embodiment, the first and second nucleic acid molecules are AAV vectors.

In another embodiment, each of (a), (b), and (c) are present on the same nucleic acid molecule, e.g., the same vector, e.g., the same viral vector, e.g., an AAV vector. In an embodiment, the nucleic acid molecule is an AAV vector. In an alternate embodiment, one of (a), (b), and (c) is encoded on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and a second and third of (a), (b), and (c) is encoded on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In an embodiment, (a) is present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, a first AAV vector; and (b) and (c) are present on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (b) is present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (a) and (c) are present on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (c) is present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (a) and (b) are present on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In another embodiment, each of (a), (b), (c) and (d) are present on the same nucleic acid molecule, e.g., the same vector, e.g., the same viral vector, e.g., an AAV vector. In an embodiment, the nucleic acid molecule may be an AAV vector.

In other embodiments, one of (a), (b), (c) and (d) is encoded on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and a second, third, and fourth of (a), (b), (c) and (d) is encoded on a second nucleic acid molecule, e.g., a second

vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (a) is present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (b), (c), and (d) are present on a second
 5 nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (b) is present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (a), (c), and (d) are present on a second
 10 nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (c) is present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (a), (b), and (d) are present on a second
 nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second vector, e.g., a
 second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (d) is present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (a), (b), and (c) are present on a second
 15 nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, a first and second of (a), (b), (c) and (d) is encoded on a first
 20 nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first viral vector, e.g., a first AAV vector; and a third and fourth of (a), (b), (c) and (d) is encoded on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (a) and (b) are present on a first nucleic acid molecule, e.g., a first
 25 vector, e.g., a first viral vector, e.g., a first AAV vector; and (c) and (d) are present on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (a) and (c) are present on a first nucleic acid molecule, e.g., a first
 vector, e.g., a first viral vector, e.g., a first AAV vector; and (b) and (d) are present on a second
 30 nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

In other embodiments, (a) and (d) are present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (b) and (c) are present on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

5 In other embodiments, (b) and (d) are present on a first nucleic acid molecule, e.g., a first vector, e.g., a first viral vector, e.g., a first AAV vector; and (a) and (c) are present on a second nucleic acid molecule, e.g., a second vector, e.g., a second vector, e.g., a second vector, e.g., a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

10 In other embodiments, the first nucleic acid molecule is other than an AAV vector and the second nucleic acid molecule is an AAV vector. In still other embodiments, the first nucleic acid molecule is an AAV vector and the second nucleic acid molecule is other than an AAV vector.

The nucleic acids described herein may comprise a promoter operably linked to the sequence that encodes said gRNA molecule of (a), e.g., a promoter described herein. The nucleic acid may further comprise a second promoter operably linked to the sequence that encodes the second gRNA molecule of (c), e.g., a promoter described herein. The promoter and second promoter differ from one another. In some embodiments, the promoter and second promoter are the same.

20 The nucleic acids described herein may further comprise a promoter operably linked to the sequence that encodes the Cas9 molecule of (b), e.g., a promoter described herein.

In another aspect, disclosed herein is a composition comprising (a) a gRNA molecule comprising a targeting domain that is complementary with a target domain in *USH2A* gene, as described herein. The composition of (a) may further comprise (b) a Cas9 molecule, e.g., a Cas9 molecule as described herein. A composition of (a) and (b) may further comprise (c) a second gRNA molecule, e.g., a second, third and/or fourth gRNA molecule, e.g., a second, third and/or fourth gRNA molecule, as described herein. A composition of (a), (b) and (c) may further comprise (d) a template nucleic acid, e.g., a template nucleic acid described herein, e.g., a template nucleic acid, as described herein. In an embodiment, the composition is a pharmaceutical composition. The Compositions described herein, e.g., pharmaceutical compositions described herein, can be used in treating Usher Syndrome or retinitis pigmentosa 39 in a subject, e.g., in accordance with a method disclosed herein.

In another aspect, disclosed herein is a method of altering a cell, e.g., altering the structure, e.g., altering the sequence, of a target nucleic acid of a cell, comprising contacting said cell with: (a) a gRNA that targets the *USH2A* gene, e.g., a gRNA as described herein; (b) a Cas9 molecule, e.g., a Cas9 molecule as described herein; and optionally, (c) a second, third and/or fourth gRNA that targets *USH2A* gene, e.g., a second, third and/or fourth gRNA as described herein; and (d) a template nucleic acid, e.g., a template nucleic acid as described herein.

In some embodiments, the method comprises contacting said cell with (a), (b), (c), and (d). The gRNA of (a) may be selected from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, or a gRNA that differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. The gRNA of (c) may be selected from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, or a gRNA that differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**.

In some embodiments, the method comprises contacting a cell from a subject. The cell may be from a subject having a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene. In an embodiment, the cell is from a subject suffering from Usher syndrome, e.g., Usher syndrome type 2A. In another embodiment, the cell is from a subject suffering from retinitis pigmentosa, e.g., retinitis pigmentosa 39.

In some embodiments, the cell being contacted in the disclosed method is a photoreceptor cell. The contacting may be performed *ex vivo* and the contacted cell may be returned to the subject's body after the contacting step. In other embodiments, the contacting step may be performed *in vivo*.

In some embodiments, the cell being contacted in the disclosed method is an inner hair cell or an outer hair cell. The contacting may be performed *ex vivo* and the contacted cell may be returned to the subject's body after the contacting step. In other embodiments, the contacting step may be performed *in vivo*.

In some embodiments, the method of altering a cell as described herein comprises acquiring knowledge of a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in said cell, prior to the contacting step. Acquiring knowledge of the presence of a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the cell may be by sequencing a portion of the *USH2A* (or *RP39*) gene. In some

embodiments, acquiring knowledge of a mutation in the *USH2A* (or *RP39*) gene is used to treat a subject (or a cell from the subject) likely to develop Usher syndrome or retinitis pigmentosa (e.g., correct the guanine deletion at nucleotide position 2299).

5 Based on the presence of a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG), the method may further comprise selecting a template nucleic acid, e.g., to correct the mutation in the cell. For example, the method may comprise correcting a guanine deletion at nucleotide position 2299 in the *USH2A* gene.

10 In some embodiments, the contacting step of the method comprises contacting the cell with a nucleic acid, e.g., a vector, e.g., an AAV vector, that expresses at least one of (a), (b), and (c). In some embodiments, the contacting step of the method comprises contacting the cell with a nucleic acid, e.g., a vector, e.g., an AAV vector, that expresses each of (a), (b), and (c). In another embodiment, the contacting step of the method comprises delivering to the cell the Cas9 molecule of (b) and a nucleic acid which encodes a gRNA of (a) and optionally, a second, third and/or fourth gRNA of (c).

15 In an embodiment, the contacting step comprises contacting the cell with a nucleic acid, e.g., a vector, e.g., an AAV vector, described herein.

In an embodiment, the contacting step comprises delivering to the cell the Cas9 molecule of (b), as a protein or an mRNA, and a nucleic acid which encodes a gRNA of (a) and optionally a second, third and/or fourth gRNA of (c).

20 In an embodiment, the contacting step comprises delivering to the cell the Cas9 molecule of (b), as a protein or an mRNA, said gRNA of (a), as an RNA, and optionally said second, third and/or fourth gRNA of (c), as an RNA.

25 In an embodiment, the contacting step comprises delivering to the cell the gRNA of (a) as an RNA, optionally the second, third and/or fourth gRNA of (c) as an RNA, and a nucleic acid that encodes the Cas9 molecule of (b).

In another aspect, disclosed herein is a method of treating a subject having or likely to develop Usher Syndrome, e.g., by altering the structure, e.g., the sequence, of a target nucleic acid of the subject, comprising contacting said subject (or a cell from said subject) with:

- 30 (a) a gRNA that targets the *USH2A* gene, e.g., a gRNA disclosed herein;
(b) a Cas9 molecule, e.g., a Cas9 molecule disclosed herein;

optionally, (c)(i) a second gRNA that targets *USH2A* gene, e.g., a second gRNA disclosed herein; and further optionally, (c)(ii) a third gRNA, and still further optionally, (c)(iii) a fourth gRNA that target the *CEP290*, e.g., a fourth gRNA disclosed herein, and
(d) a template nucleic acid, e.g., a template nucleic acid disclosed herein.

5 In an embodiment, contacting comprises contacting with (a), (b), and (d).

In an embodiment, contacting comprises contacting with (a), (b), (c)(i), and (d).

In an embodiment, contacting comprises contacting with (a), (b), (c)(i), (c)(ii), and (d).

In an embodiment, contacting comprises contacting with (a), (b), (c)(i), (c)(ii), (c)(iii), and (d).

10 The gRNA of (a) or (c) (e.g., (c)(i), (c)(ii), or (c)(iii)) may be independently selected from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, or a gRNA that differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**.

In an embodiment, said subject is suffering from Usher syndrome. In an embodiment,
15 said subject has a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene.

In an embodiment, the method comprises acquiring knowledge of the presence of a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene, in said subject.

20 In an embodiment, the method comprises acquiring knowledge of the presence of a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene, in said subject by sequencing a portion of the *USH2A* gene.

In an embodiment, a cell of said subject is contacted *ex vivo* with (a), (b), (d), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In an embodiment,
25 said cell is returned to the subject's body.

In an embodiment, the method comprises a treatment comprising introducing a cell into said subject's body, wherein said cell subject was contacted *ex vivo* with (a), (b), (d), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In an embodiment, the method comprises said contacting, e.g., contacting a cell of the
30 subject, is performed *in vivo*. In an embodiment, contacting the cell of a subject *in vivo* is by

subretinal delivery. In an embodiment, contacting the cell of a subject *in vivo* is by subretinal injection.

In an embodiment, the contacting step comprises contacting said subject with a nucleic acid, e.g., a vector, e.g., an AAV vector, described herein, e.g., a nucleic acid that expresses at least one of (a), (b), (c)(i), (c)(ii), or (c)(iii).

In an embodiment, the contacting step comprises delivering to said subject said Cas9 molecule of (b), as a protein or mRNA, and a nucleic acid which encodes (a), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In an embodiment, the contacting step comprises delivering to said subject said Cas9 molecule of (b), as a protein or mRNA, said gRNA of (a), as an RNA, and optionally said second gRNA of (c)(i), further optionally said third gRNA of (c)(ii), and still further optionally said fourth gRNA of (c)(iii), as an RNA.

In an embodiment, the contacting step comprises delivering to said subject said gRNA of (a), as an RNA, optionally said second gRNA of (c)(i), further optionally said third gRNA of (c)(ii), and still further optionally said fourth gRNA of (c)(iii), as an RNA, and a nucleic acid that encodes the Cas9 molecule of (b).

In another aspect, disclosed herein is a method of treating a subject having or likely to develop retinitis pigmentosa, e.g., by altering the structure, e.g., the sequence, of a target nucleic acid of the subject, comprising contacting said subject (or a cell from said subject) with:

(a) a gRNA that targets the *RP39* (also known as *USH2A*) gene, e.g., a gRNA disclosed herein;

(b) a Cas9 molecule, e.g., a Cas9 molecule disclosed herein;
optionally, (c)(i) a second gRNA that targets *USH2A* gene, e.g., a second gRNA disclosed herein; and further optionally, (c)(ii) a third gRNA, and still further optionally, (c)(iii) a fourth gRNA that target the *CEP290*, e.g., a third and fourth gRNA disclosed herein, and
(d) a template nucleic acid, e.g., a template nucleic acid disclosed herein.

In an embodiment, contacting comprises contacting with (a), (b), and (d).

In an embodiment, contacting comprises contacting with (a), (b), (c)(i), and (d).

In an embodiment, contacting comprises contacting with (a), (b), (c)(i), (c)(ii), and (d).

In an embodiment, contacting comprises contacting with (a), (b), (c)(i), (c)(ii), (c)(iii), and (d).

The gRNA of (a) may be selected from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, or a gRNA that differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**.

5 The gRNA of (c) may be selected from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, or a gRNA that differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**.

In an embodiment, said subject is suffering from Usher syndrome or retinitis pigmentosa. In an embodiment, said subject has a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene.

10 In an embodiment, the method comprises acquiring knowledge of the presence of a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene, in said subject.

In an embodiment, the method comprises acquiring knowledge of the presence of a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene, in said subject by sequencing a portion of the *USH2A* gene.

15 In an embodiment, said subject is suffering from retinitis pigmentosa, e.g., retinitis pigmentosa 39. In an embodiment, said subject has a mutation in the *RP39* (also known as *USH2A*) gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *RP39* gene.

20 In an embodiment, the method comprises acquiring knowledge of the presence of a mutation in the *RP39* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *RP39* gene, in said subject.

In an embodiment, the method comprises acquiring knowledge of the presence of a mutation in the *RP39* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the *RP39* gene, in said subject by sequencing a portion of the *USH2A* gene.

25 In an embodiment, the method comprises, based on the presence of a mutation in the *USH2A* (or *RP39*) gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG), selecting a template nucleic acid.

In an embodiment, the method comprises correcting a deletion of a guanine at nucleotide position 2299 (2299delG) in the *USH2A* (or *RP39*) gene.

30

In an embodiment, a cell of said subject is contacted *ex vivo* with (a), (b), (d), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In an embodiment, said cell is returned to the subject's body.

5 In an embodiment, the method comprises a treatment comprising introducing a cell into said subject's body, wherein said cell subject was contacted *ex vivo* with (a), (b), (d), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

10 In an embodiment, the method comprises said contacting, e.g., contacting a cell of the subject, is performed *in vivo*. In an embodiment, contacting the cell of a subject *in vivo* is by subretinal delivery. In an embodiment, contacting the cell of a subject *in vivo* is by subretinal injection.

In an embodiment, the contacting step comprises contacting said subject with a nucleic acid, e.g., a vector, e.g., an AAV vector, described herein, e.g., a nucleic acid that expresses at least one of (a), (b), (c)(i), c(ii), or c(iii).

15 In an embodiment, the contacting step comprises delivering to said subject said Cas9 molecule of (b), as a protein or mRNA, and a nucleic acid which encodes (a), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

20 In an embodiment, the contacting step comprises delivering to said subject said Cas9 molecule of (b), as a protein or mRNA, said gRNA of (a), as an RNA, and optionally said second gRNA of (c)(i), further optionally said third gRNA of (c)(ii), and still further optionally said third gRNA of (c)(iii), as an RNA.

In an embodiment, the contacting step comprises delivering to said subject said gRNA of (a), as an RNA, optionally said second gRNA of (c)(i), further optionally said third gRNA of (c)(ii), and still further optionally said third gRNA of (c)(iii), as an RNA, and a nucleic acid that encodes the Cas9 molecule of (b).

25 In another aspect, disclosed herein is a reaction mixture comprising a gRNA, a nucleic acid, or a composition described herein, and a cell, e.g., a cell from a subject having Usher syndrome or retinitis pigmentosa 39, or a subject having a mutation in the *USH2A* (or *RP39*) gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG).

30 In another aspect, disclosed herein is a kit comprising (a) gRNA molecule described herein, or nucleic acid that encodes said gRNA, and one or more of the following:

(b) a Cas9 molecule, e.g., a Cas9 molecule described herein;

- (c)(i) a second gRNA molecule, e.g., a second gRNA molecule described herein;
- (c)(ii) a third gRNA molecule, e.g., a second gRNA molecule described herein; or
- (c)(iii) a fourth gRNA molecule, e.g., a second gRNA molecule described herein;
- (d) a template nucleic acid e.g, a template nucleic acid described herein;
- 5 (e) nucleic acid that encodes one or more of (b), (c)(i), (c)(ii), (c)(iii), or (d).

In an embodiment, the kit comprises a nucleic acid, e.g., an AAV vector, that encodes one or more of (a), (b), (c)(i), (c)(ii), or c(iii).

In an embodiment, the kit further comprises a template nucleic acid, e.g., a single strand DNA that comprises said template nucleic acid.

10 In another aspect, disclosed herein is non-naturally occurring template nucleic acid described herein.

In yet another aspect, disclosed herein is a gRNA molecule, e.g., a gRNA molecule described herein, for use in treating Usher Syndrome or retinitis pigmentosa 39 in a subject, e.g., in accordance with a method of treating Usher Syndrome or retinitis pigmentosa 39 as described
15 herein.

In an embodiment, the gRNA molecule is used in combination with a Cas9 molecule, e.g., a Cas9 molecule described herein. Additionally or alternatively, in an embodiment, the gRNA molecule is used in combination with a second, third and/or fourth gRNA molecule, e.g., a second, third and/or fourth gRNA molecule described herein. Additionally or alternatively, in an
20 embodiment, the gRNA molecule is used in combination with a template nucleic acid, e.g., a template nucleic acid described herein.

In still another aspect, disclosed herein is use of a gRNA molecule, e.g., a gRNA molecule described herein, in the manufacture of a medicament for treating Usher Syndrome or retinitis pigmentosa 39 in a subject, e.g., in accordance with a method of treating Usher
25 Syndrome or retinitis pigmentosa 39 as described herein.

In an embodiment, the medicament comprises a Cas9 molecule, e.g., a Cas9 molecule described herein. Additionally or alternatively, in an embodiment, the medicament comprises a second, third and/or fourth gRNA molecule, e.g., a second, third and/or fourth gRNA molecule described herein. Additionally or alternatively, in an embodiment, the medicament comprises a
30 template nucleic acid, e.g., a template nucleic acid described herein.

The gRNA molecules and methods, as disclosed herein, can be used in combination with a governing gRNA molecule. As used herein, a governing gRNA molecule refers to a gRNA molecule comprising a targeting domain which is complementary to a target domain on a nucleic acid that encodes a component of the CRISPR/Cas system introduced into a cell or subject. For example, the methods described herein can further include contacting a cell or subject with a governing gRNA molecule or a nucleic acid encoding a governing molecule. In an embodiment, the governing gRNA molecule targets a nucleic acid that encodes a Cas9 molecule or a nucleic acid that encodes a target gene gRNA molecule. In an embodiment, the governing gRNA comprises a targeting domain that is complementary to a target domain in a sequence that encodes a Cas9 component, e.g., a Cas9 molecule or target gene gRNA molecule. In an embodiment, the target domain is designed with, or has, minimal homology to other nucleic acid sequences in the cell, e.g., to minimize off-target cleavage. For example, the targeting domain on the governing gRNA can be selected to reduce or minimize off-target effects. In an embodiment, a target domain for a governing gRNA can be disposed in the control or coding region of a Cas9 molecule or disposed between a control region and a transcribed region. In an embodiment, a target domain for a governing gRNA can be disposed in the control or coding region of a target gene gRNA molecule or disposed between a control region and a transcribed region for a target gene gRNA. While not wishing to be bound by theory, in an embodiment, it is believed that altering, e.g., inactivating, a nucleic acid that encodes a Cas9 molecule or a nucleic acid that encodes a target gene gRNA molecule can be effected by cleavage of the targeted nucleic acid sequence or by binding of a Cas9 molecule/governing gRNA molecule complex to the targeted nucleic acid sequence.

The compositions, reaction mixtures and kits, as disclosed herein, can also include a governing gRNA molecule, e.g., a governing gRNA molecule disclosed herein,

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 invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Headings, including numeric and alphabetical headings and subheadings, are for organization and presentation and are not intended to be limiting.

Other features and advantages of the invention will be apparent from the detailed description, drawings, and from the claims.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-1G are representations of several exemplary gRNAs.

Fig. 1A depicts a modular gRNA molecule derived in part (or modeled on a sequence in part) from *Streptococcus pyogenes* (*S. pyogenes*) as a duplexed structure (SEQ ID NOS: 42 and 10 43, respectively, in order of appearance);

Fig. 1B depicts a unimolecular (or chimeric) gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO: 44);

Fig. 1C depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO: 45);

15 **Fig. 1D** depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO: 46);

Fig. 1E depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO: 47);

Fig. 1F depicts a modular gRNA molecule derived in part from *Streptococcus thermophilus* (*S. thermophilus*) as a duplexed structure (SEQ ID NOS: 48 and 49, respectively, 20 in order of appearance);

Fig. 1G depicts an alignment of modular gRNA molecules of *S. pyogenes* and *S. thermophilus* (SEQ ID NOS: 50-53, respectively, in order of appearance).

Figs. 2A-2G depict an alignment of Cas9 sequences from Chylinski *et al.* (RNA Biol. 25 2013; 10(5): 726–737). The N-terminal RuvC-like domain is boxed and indicated with a “Y”. The other two RuvC-like domains are boxed and indicated with a “B”. The HNH-like domain is boxed and indicated by a “G”. Sm: *S. mutans* (SEQ ID NO: 1); Sp: *S. pyogenes* (SEQ ID NO: 2); St: *S. thermophilus* (SEQ ID NO: 3); Li: *L. innocua* (SEQ ID NO: 4). Motif: this is a motif based on the four sequences: residues conserved in all four sequences are indicated by single 30 letter amino acid abbreviation; “*” indicates any amino acid found in the corresponding position

of any of the four sequences; and “-” indicates any amino acid, e.g., any of the 20 naturally occurring amino acids, or absent.

Figs. 3A-3B show an alignment of the N-terminal RuvC-like domain from the Cas9 molecules disclosed in Chylinski *et al* (SEQ ID NOS: 54-103, respectively, in order of appearance). The last line of **Fig. 3B** identifies 4 highly conserved residues.

Figs. 4A-4B show an alignment of the N-terminal RuvC-like domain from the Cas9 molecules disclosed in Chylinski *et al.* with sequence outliers removed (SEQ ID NOS: 104-177, respectively, in order of appearance). The last line of **Fig. 4B** identifies 3 highly conserved residues.

Figs. 5A-5C show an alignment of the HNH-like domain from the Cas9 molecules disclosed in Chylinski *et al* (SEQ ID NOS: 178-252, respectively, in order of appearance). The last line of **Fig. 5C** identifies conserved residues.

Figs. 6A-6B show an alignment of the HNH-like domain from the Cas9 molecules disclosed in Chylinski *et al.* with sequence outliers removed (SEQ ID NOS: 253-302, respectively, in order of appearance). The last line of **Fig. 6B** identifies 3 highly conserved residues.

Figs. 7A-7B depict an alignment of Cas9 sequences from *S. pyogenes* and *Neisseria meningitidis* (*N. meningitidis*). The N-terminal RuvC-like domain is boxed and indicated with a “Y”. The other two RuvC-like domains are boxed and indicated with a “B”. The HNH-like domain is boxed and indicated with a “G”. Sp: *S. pyogenes*; Nm: *N. meningitidis*. Motif: this is a motif based on the two sequences: residues conserved in both sequences are indicated by a single amino acid designation; “*” indicates any amino acid found in the corresponding position of any of the two sequences; “-” indicates any amino acid, e.g., any of the 20 naturally occurring amino acids, and “-” indicates any amino acid, e.g., any of the 20 naturally occurring amino acids, or absent.

Fig. 8 shows a nucleic acid sequence encoding Cas9 of *N. meningitidis* (SEQ ID NO: 303). Sequence indicated by an “R” is an SV40 NLS; sequence indicated as “G” is an HA tag; and sequence indicated by an “O” is a synthetic NLS sequence; the remaining (unmarked) sequence is the open reading frame (ORF).

Figs. 9A and 9B are schematic representations of the domain organization of *S. pyogenes* Cas 9. **Fig. 9A** shows the organization of the Cas9 domains, including amino acid positions, in

reference to the two lobes of Cas9 (recognition (REC) and nuclease (NUC) lobes). **Fig. 9B** shows the percent homology of each domain across 83 Cas9 orthologs.

DETAILED DESCRIPTION

5 Definitions

Domain, as used herein, is used to describe segments of a protein or nucleic acid. Unless otherwise indicated, a domain is not required to have any specific functional property.

Calculations of homology or sequence identity between two sequences (the terms are used interchangeably herein) are performed as follows. The sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The optimal alignment is determined as the best score using the GAP program in the GCG software package with a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences.

20 “Governing gRNA molecule”, as used herein, refers to a gRNA molecule that comprises a targeting domain that is complementary to a target domain on a nucleic acid that comprises a sequence that encodes a component of the CRISPR/Cas system that is introduced into a cell or subject. A governing gRNA does not target an endogenous cell or subject sequence. In an embodiment, a governing gRNA molecule comprises a targeting domain that is complementary with a target sequence on: (a) a nucleic acid that encodes a Cas9 molecule; (b) a nucleic acid that encodes a gRNA which comprises a targeting domain that targets the *USH2A* gene (a target gene gRNA); or on more than one nucleic acid that encodes a CRISPR/Cas component, e.g., both (a) and (b). In an embodiment, a nucleic acid molecule that encodes a CRISPR/Cas component, e.g., that encodes a Cas9 molecule or a target gene gRNA, comprises more than one target domain that is complementary with a governing gRNA targeting domain. While not wishing to be bound by theory, in an embodiment, it is believed that a governing gRNA molecule

complexes with a Cas9 molecule and results in Cas9 mediated inactivation of the targeted nucleic acid, e.g., by cleavage or by binding to the nucleic acid, and results in cessation or reduction of the production of a CRISPR/Cas system component. In an embodiment, the Cas9 molecule forms two complexes: a complex comprising a Cas9 molecule with a target gene gRNA, which complex will alter the *USH2A* gene; and a complex comprising a Cas9 molecule with a governing gRNA molecule, which complex will act to prevent further production of a CRISPR/Cas system component, e.g., a Cas9 molecule or a target gene gRNA molecule. In an embodiment, a governing gRNA molecule/Cas9 molecule complex binds to or promotes cleavage of a control region sequence, e.g., a promoter, operably linked to a sequence that encodes a Cas9 molecule, a sequence that encodes a transcribed region, an exon, or an intron, for the Cas9 molecule. In an embodiment, a governing gRNA molecule/Cas9 molecule complex binds to or promotes cleavage of a control region sequence, e.g., a promoter, operably linked to a gRNA molecule, or a sequence that encodes the gRNA molecule. In an embodiment, the governing gRNA, e.g., a Cas9-targeting governing gRNA molecule, or a target gene gRNA-targeting governing gRNA molecule, limits the effect of the Cas9 molecule/target gene gRNA molecule complex-mediated gene targeting. In an embodiment, a governing gRNA places temporal, level of expression, or other limits, on activity of the Cas9 molecule/target gene gRNA molecule complex. In an embodiment, a governing gRNA reduces off-target or other unwanted activity. In an embodiment, a governing gRNA molecule inhibits, e.g., entirely or substantially entirely inhibits, the production of a component of the Cas9 system and thereby limits, or governs, its activity.

“Modulator”, as used herein, refers to an entity, e.g., a drug, that can alter the activity (e.g., enzymatic activity, transcriptional activity, or translational activity), amount, distribution, or structure of a subject molecule or genetic sequence. In an embodiment, modulation comprises cleavage, e.g., breaking of a covalent or non-covalent bond, or the forming of a covalent or non-covalent bond, e.g., the attachment of a moiety, to the subject molecule. In an embodiment, a modulator alters the, three dimensional, secondary, tertiary, or quaternary structure, of a subject molecule. A modulator can increase, decrease, initiate, or eliminate a subject activity.

“Large molecule”, as used herein, refers to a molecule having a molecular weight of at least 2, 3, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 kD. Large molecules include proteins, polypeptides, nucleic acids, biologics and carbohydrates.

“Polypeptide”, as used herein, refers to a polymer of amino acids having less than 100 amino acid residues. In an embodiment it has less than 50, 20, or 10 amino acid residues.

“Reference molecule”, e.g., a reference Cas9 molecule or reference gRNA, as used herein, refers to a molecule to which a subject molecule, e.g., a subject Cas9 molecule of subject gRNA molecule, e.g., a modified or candidate Cas9 molecule is compared. For example, a Cas9 molecule may be characterized as having no more than 10% of the nuclease activity of a reference Cas9 molecule. Examples of reference Cas9 molecules include naturally occurring unmodified Cas9 molecules, e.g., a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, *S. aureus* or *S. thermophilus*. In an embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology with the Cas9 molecule to which it is being compared. In an embodiment, the reference Cas9 molecule is a sequence, e.g., a naturally occurring or known sequence, which is the parental form on which a change, e.g., a mutation has been made.

“Replacement”, or “replaced”, as used herein with reference to a modification of a molecule does not require a process limitation but merely indicates that the replacement entity is present.

“Small molecule”, as used herein, refers to a compound having a molecular weight less than about 2 kD, e.g., less than about 2 kD, less than about 1.5 kD, less than about 1 kD, or less than about 0.75 kD.

“Subject”, as used herein, may mean either a human or non-human animal. The term includes, but is not limited to, mammals (e.g., humans, other primates, pigs, rodents (e.g., mice and rats or hamsters), rabbits, guinea pigs, cows, horses, cats, dogs, sheep, and goats). In an embodiment the subject is a human. In other embodiments the subject is poultry.

“Treat”, “treating” and “treatment”, as used herein, mean the treatment of a disease in a mammal, e.g., in a human, including (a) inhibiting the disease, i.e., arresting or preventing its development; (b) relieving the disease, i.e., causing regression of the disease state; and (c) curing the disease.

“X”, as used herein, in the context of an amino acid sequence, refers to any amino acid (e.g., any of the twenty natural amino acids) unless otherwise specified.

Usher Syndrome

Usher syndrome is a disease characterized by progressive loss of vision beginning between the ages of 10 and 20. Usher syndrome type 1 symptoms are generally more severe and have an earlier onset than those of Usher syndrome type 2 (e.g., Usher syndrome type 2A). The vision loss in Usher syndrome is described as retinitis pigmentosa (RP), a group of inherited retinal dystrophies that affect photoreceptors and retinal pigment epithelium cells.

Subjects suffering from Usher syndrome type II have mutations in the *USH2A* gene (also known as the *RP39* gene) and develop vision loss that is accompanied by hearing loss (and/or balance problems). The visual loss associated with Usher syndrome type II is called ‘syndromic’ retinitis pigmentosa, because it is associated with hearing loss. Alternatively, patients can have mutations in *USH2A* that are not associated with hearing loss. In this case, the patients are defined as having ‘non-syndromic’ retinitis pigmentosa. Non-syndromic retinitis pigmentosa caused by mutations in the *USH2A* gene is also called retinitis pigmentosa 39, or RP39. In both syndromic and non-syndromic RP, repair of the *USH2A* mutations within the eye may ameliorate or slow the progression of retinitis pigmentosa. In syndromic RP, repair of *USH2A* mutations may ameliorate vision loss but not address hearing loss. In non-syndromic RP, repair of *USH2A* may ameliorate vision loss (but not hearing loss as there is no hearing loss in non-syndromic RP).

The *USH2A* gene is 85,000 base pairs and codes for the usherin protein. Usherin is expressed in photoreceptors of the retina and in inner hair cells and outer hair cells in the inner ear. The most common mutation in subjects with Usher syndrome type II or non-syndromic retinitis pigmentosa (RP39) is a single nucleotide deletion, e.g., a guanine deletion, at nucleotide position 2299 (2299delG) in the *USH2A* gene, which is responsible for somewhere between 15% and 40% of *USH2A* mutations. The deletion of guanine at position 2299 results in a premature stop codon.

The *USH2A* gene is expressed in retinal photoreceptor (PR) rods and cones. Photoreceptors cells have an outer segment made of a cilium that plays an important role in the retinoid cycle and the phototransduction cascade. The *USH2A* gene encodes the usherin protein which is responsible for protein trafficking in the PR outer segment. Mutations in the *USH2A* gene leads to interrupted protein transport between the ciliary inner segment and outer segment. This causes PR dysfunction and loss of vision in retinitis pigmentosa.

As RP progresses, PR rods generally degenerate first. In most cases of RP, rod photoreceptor cells function poorly and begin to die at the earliest stages of disease, resulting in poor night vision and declining peripheral vision. PR cones generally degenerate late in the course of disease. This causes the typical phenotypic progression experienced by RP patients.

5 They experience loss of peripheral visual fields followed by loss of central visual fields (the latter measured by decreases in visual acuity).

Methods to Treat or Prevent Usher Syndrome type 2A and/or Retinitis Pigmentosa 39

Treatment for RP is limited and there is currently no approved treatment that
 10 substantially reverses or halts the progression of disease in Usher Syndrome type 2 or in RP-39. Vitamin A supplementation may delay onset of disease and slow progression. An electrical implant known as the Argus II retinal implant was recently approved for use, but it only offers minimal improvement in vision in patients with RP. The best visual acuity achieved in trials by the device was 20/1260 (legal blindness is defined as 20/200 vision). In addition, current gene
 15 therapy delivery techniques are not able to deliver genes encoding large proteins, e.g., the *USH2A* gene.

In the retina, the *USH2A* gene is expressed in retinal photoreceptor (PR) rods and cones. Photoreceptors cells have an outer segment made of a cilium that plays an important role in the retinoid cycle and the phototransduction cascade. The *USH2A* gene encodes the usherin protein
 20 that is responsible for protein trafficking in the PR outer segment. Mutations in the *USH2A* gene leads to interrupted protein transport between the ciliary inner segment and outer segment. This causes PR dysfunction and eventual loss of vision in retinitis pigmentosa.

As RP progresses, PR rods generally degenerate first. In most cases of RP, rod photoreceptor cells function poorly and begin to die at the earliest stages of disease, resulting in
 25 poor night vision and declining peripheral vision. PR cones generally degenerate late in the course of disease. This causes the typical phenotypic progression experienced by RP patients. They experience loss of peripheral visual fields followed by loss of central visual fields (the latter measured by decreases in visual acuity).

Correction of the *USH2A* gene (e.g., insertion of the deleted guanine residue at
 30 nucleotide position 2299) in the eye may delay disease progression or improve in vision, or both. Restoring functional usherin to PR rods and cones is predicted to preserve communication and

functioning within PR cells. This may delay or prevent PR cell death in subjects with Usher syndrome type 2 and RP39. Following correction of the *USH2A* gene, subjects can experience delayed disease progression and/or improvements in vision.

5 In the inner ear, the *USH2A* gene is expressed in inner and outer hair cells. Hair cells are responsible for mechanotransduction within the inner ear, a process in which sound waves are converted to electrical signals that are picked up by neurons in the inner ear and converted into sounds. Stereocilia within hair cells rely on functional usherin to interact with myosin 7A, whirlin and harmonin proteins for effective mechanotransduction (see Adato et al., Human Molecular Genetics 2005; 14(24):3921-3932, in particular, Figure 6). Truncated or errant
10 splicing of harmonin leads to dysfunction of the interconnections of harmonin and other stereociliary proteins, which leads to disruption in hearing.

Correction of the *USH2A* gene in the inner ear can delay progression of hearing loss or improve hearing or both. Following correction of the *USH2A* gene, subjects can experience delayed disease progression and/or improvements in hearing.

15 As disclosed herein, *USH2A* mutations may be corrected by gene editing, e.g., using CRISPR-Cas9 mediated methods to correct the guanine deletion at position 2299 in the *USH2A* gene (i.e., replace the deleted guanine residue at position 2299 in the *USH2A* gene).

Described herein are methods for treating or delaying the onset or progression of Usher syndrome type 2A and/or retinitis pigmentosa 39 (RP39), e.g., caused by mutations in the
20 *USH2A* gene, including but not limited to the mutations: c.2299delG. The disclosed methods for treating or delaying the onset or progression of Usher type 2A and/or RP39 alter the *USH2A* gene by genome editing using a gRNA targeting the Usher type 2A and/or RP39 target position and a Cas9 enzyme. Details on gRNAs targeting the Usher type 2A and/or RP39 target position and Cas9 enzymes are provided below.

25 In a method disclosed herein, a mutation is targeted by cleaving with either a single nuclease or dual nickase, e.g., to induce HDR with a donor template, that corrects the point mutation (e.g., the single nucleotide, e.g., guanine, deletion). The method can include acquiring knowledge of the mutation carried by the subject, e.g., by sequencing the appropriate portion of the *USH2A* gene.

30 Usher syndrome involves, e.g., hearing loss and a progressive decline in visual acuity and treatment during the earlier stages of the disease may prevent further decline in visual acuity.

Some subjects with Usher syndrome may benefit from treatment at later stages of the disease. Physicians detecting hearing loss or loss of visual acuity in a young subject may consider determining or acquiring the relevant *USH2A* sequence in the subject to determine whether the hearing loss or loss of visual acuity is due to a mutation in the *USH2A* gene. If so, the subject may be a candidate for treatment.

In an embodiment, treatment is initiated prior to onset of the disease.

In an embodiment, treatment is initiated after onset of the disease.

In an embodiment, treatment is initiated prior to loss of visual acuity.

In an embodiment, treatment is initiated at onset of loss of visual acuity.

In an embodiment, treatment is initiated after onset of loss of visual acuity.

In an embodiment, treatment is initiated prior to loss of hearing.

In an embodiment, treatment is initiated at onset of loss of hearing.

In an embodiment, treatment is initiated after onset of loss of hearing.

In an embodiment, the subject undergoes genetic testing and is found to have a mutation in the *USH2A* gene.

In an embodiment, treatment is initiated at the appearance of any of the following symptoms: declining peripheral vision, poor night vision or night blindness, progressive visual loss, and/or progression constriction of the visual field.

In an embodiment, treatment is initiated before the appearance of any of the following symptoms: declining peripheral vision, poor night vision or night blindness, progressive visual loss, and/or progression constriction of the visual field.

In an embodiment, treatment is initiated after the appearance of any of the following symptoms: declining peripheral vision, poor night vision or night blindness, progressive visual loss, and/or progression constriction of the visual field.

In an embodiment, treatment is initiated at the appearance of any of the following findings consistent with Usher syndrome or RP on exam, including but not limited to, bone spicule pigmentation, narrowing of the visual fields, retinal atrophy, attenuated retinal vasculature, loss of retinal pigment epithelium, and/or pallor of the optic nerve.

In an embodiment, treatment is initiated before the appearance of any of the following findings consistent with Usher syndrome or RP on exam, including but not limited to, bone

spicule pigmentation, narrowing of the visual fields, retinal atrophy, attenuated retinal vasculature, loss of retinal pigment epithelium, and/or pallor of the optic nerve.

In an embodiment, treatment is initiated after the appearance of any of the following findings consistent with Usher syndrome or RP on exam, including but not limited to, bone
5 spicule pigmentation, narrowing of the visual fields, retinal atrophy, attenuated retinal vasculature, loss of retinal pigment epithelium, and/or pallor of the optic nerve.

In an embodiment, treatment is initiated at the appearance of any of the following symptoms: hearing loss, hearing impairment, reduced hearing, and/or profound deafness.

10 In an embodiment, treatment is initiated before the appearance of any of the following symptoms: hearing loss, hearing impairment, reduced hearing, and/or profound deafness.

In an embodiment, treatment is initiated after the appearance of any of the following symptoms: hearing loss, hearing impairment, reduced hearing, and/or profound deafness.

15 In an embodiment, treatment is initiated at the appearance of any of the following findings consistent with hearing loss on exam, including but not limited to, down-sloping configuration on audiogram, hearing loss on otoacoustic emissions (OAE) test, and/or hearing loss on Electrocochleography.

20 In an embodiment, treatment is initiated before the appearance of any of the following findings consistent with hearing loss on exam, including but not limited to, down-sloping configuration on audiogram, hearing loss on otoacoustic emissions (OAE) test, and/or hearing loss on Electrocochleography.

In an embodiment, treatment is initiated after the appearance of any of the following findings consistent with hearing loss on exam, including but not limited to, down-sloping configuration on audiogram, hearing loss on otoacoustic emissions (OAE) test, and/or hearing loss on Electrocochleography.

25 In an embodiment, treatment is initiated between the ages of 10 and 20.

In an embodiment, treatment is initiated prior to the age of 10.

In an embodiment, treatment is initiated prior to the age of 20.

In an embodiment, treatment is initiated after the age of 20.

In an embodiment, treatment is initiated after the age of 30.

30 In an embodiment, treatment is initiated after the age of 40.

In an embodiment, treatment is initiated after the age of 50.

In an embodiment, treatment is initiated after the age of 60.

In an embodiment, treatment is initiated at the appearance of loss of visual acuity in a subject's first two decades of life.

5 In an embodiment, treatment is initiated at the appearance of loss of hearing in a subject's first two decades of life.

In an embodiment, treatment is initiated after a subject is determined to have a mutation, e.g., a guanine deletion at position 2299 in *USH2A* by genetic screening, e.g., genotyping, wherein the genetic testing was performed prior to or after disease onset.

10 A subject's vision can be evaluated, e.g., prior to treatment, or after treatment, e.g., to monitor the progress of the treatment. In an embodiment, a subject's vision is evaluated prior to treatment, e.g., to determine the need for treatment. In an embodiment, a subject's vision is evaluated after treatment has been initiated, e.g., to assess the effectiveness of the treatment. Vision can be evaluated by one or more of: evaluating changes in function relative to the contralateral eye, e.g., by utilizing retinal analytical techniques; by evaluating mean, median and
15 distribution of change in best corrected visual acuity (BCVA); evaluation by Optical Coherence Tomography; evaluation of changes in visual field using perimetry; evaluation by full-field electroretinography (ERG); evaluation by slit lamp examination; evaluation of intraocular pressure; evaluation of autofluorescence, evaluation with fundoscopy; evaluation with fundus photography; evaluation with fluorescein angiography (FA); or evaluation of visual field
20 sensitivity (FFST).

A subject's hearing can be evaluated, e.g., prior to treatment, or after treatment, e.g., to monitor the progress of the treatment. In an embodiment, a subject's hearing is evaluated prior to treatment, e.g., to determine the need for treatment. In an embodiment, a subject's hearing is evaluated after treatment has been initiated, e.g., to assess the effectiveness of the treatment.
25 Hearing can be evaluated by one or more of: evaluating changes in function relative to the contralateral ear, e.g., by evaluating by physical exam, e.g., by evaluating by audiogram, e.g., by evaluating by otoacoustic emissions (OAE) test, e.g., by evaluating by electrocochleography.

Methods of Altering *USH2A*

30 As disclosed herein, *USH2A* mutations can be corrected by gene editing, e.g., using CRISPR-Cas9 mediated methods to correct a mutation in the *USH2A* gene, e.g., the guanine

deletion at position 2299 in the *USH2A* gene (e.g., replace the deleted guanine residue at position 2299 in the *USH2A* gene).

In a method disclosed herein, a mutation is targeted by cleaving with either one or more nuclease, one or more nickase, or a combination thereof, e.g., to induce HDR with a donor
5 template that corrects the point mutation (e.g., the single nucleotide, e.g., guanine, deletion). The method can include acquiring knowledge of the mutation carried by the subject, e.g., by sequencing the appropriate portion of the *USH2A* gene.

Methods and compositions discussed herein, provide for altering the *USH2A* target position in the *USH2A* gene. *USH2A* target position can be altered (e.g., corrected) by gene
10 editing, e.g., using CRISPR-Cas9 mediated methods to correct a mutation in the *USH2A* gene, e.g., the guanine deletion at position 2299 in the *USH2A* gene (e.g., replace the deleted guanine residue at position 2299, e.g., 2299delG in the *USH2A* gene).

The alteration (e.g., correction) of the mutant *USH2A* gene can be mediated by any mechanism. Exemplary mechanisms that can be associated with the alteration (e.g., correction)
15 of the mutant *USH2A* gene include, but are not limited to, non-homologous end joining (e.g., classical or alternative), microhomology-mediated end joining (MMEJ), homology-directed repair (e.g., endogenous donor template mediated), SDSA (synthesis dependent strand annealing), single strand annealing or single strand invasion.

The methods and compositions described herein introduce one or more breaks near the target position (e.g., 2299delG) in the *USH2A* gene. In an embodiment, a mutation (e.g.,
20 2299delG) is targeted by cleaving with either one or more nucleases, one or more nickases or any combination thereof to induce HDR with a donor template that corrects the point mutation (e.g., the single nucleotide, e.g., guanine, deletion, e.g., 2299delG). The method can include acquiring knowledge of the mutation carried by the subject, e.g., by sequencing the appropriate
25 portion of the *USH2A* gene.

In an embodiment, guide RNAs were designed to target a mutation (e.g., 2299delG) in the *USH2A* gene. A single gRNA with a Cas9 nuclease or a Cas9 nickase could be used to generate a break (e.g., a single strand break or a double strand break) in close proximity to a mutation (e.g., 2299delG). While not bound by theory, in an embodiment, it is believed that
30 HDR-mediated repair (e.g., with a donor template) of the break (e.g., a single strand break or a

double strand break) allows for the correction of the mutation (e.g., 2299delG) which results in restoration of a functional usherin protein.

In another embodiment, two gRNAs with two Cas9 nickases could be used to generate two single strand breaks in close proximity to a mutation (e.g., 2299delG). While not bound by theory, in an embodiment, it is believed that HDR-mediated repair (e.g., with a donor template) of the breaks (e.g., the two single strand breaks) allow for the correction of the mutation (e.g., 2299delG) which results in restoration of a functional usherin protein.

In another embodiment, more than two gRNAs may be used in a dual-targeting approach to generate two sets of breaks (e.g., two double strand breaks, one double strand break and a pair of single strand breaks or two pairs of single strand breaks) in close proximity to a mutation (e.g., 2299delG) or delete a genomic sequence containing a mutation (e.g., 2299delG) in the *USH2A* gene. While not bound by theory, in an embodiment, it is believed that HDR-mediated repair (e.g., with a donor template) of the breaks (e.g., two double strand breaks, one double strand break and a pair of single strand breaks or two pairs of single strand breaks) allow for the correction of the mutation (e.g., 2299delG) which results in restoration of a functional usherin protein.

In an embodiment, a single strand break is introduced (e.g., positioned by one gRNA molecules) in close proximity to a mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, when a single gRNA molecule is used to target a Cas9 nickase to create a single strand break in close proximity to the mutation, eg., the gRNA is used to target either upstream of (e.g., within 200 bp upstream of the mutation), or downstream of (e.g., within 200 bp downstream of the mutation) in the *USH2A* gene. In an embodiment, the break is positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat.

In an embodiment, a double strand break is introduced (e.g., positioned by one gRNA molecules) in close proximity to a mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, when a single gRNA molecule is used to target a Cas9 nuclease to create a double strand break in close proximity to the mutation, eg., the gRNA is used to target either upstream of (e.g., within 200 bp upstream of the mutation), or downstream of (e.g., within 200 bp downstream of the mutation) in the *USH2A* gene. In an embodiment, the break is positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat.

In an embodiment, two single strand breaks are introduced (e.g., positioned by two gRNA molecules) in close proximity to a mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, when two gRNA molecules are used to target two Cas9 nickcases to create two single strand breaks in close proximity to the mutation, e.g., both gRNAs are used to target upstream of (e.g., within 200 bp upstream of the mutation), both gRNAs are used to target downstream of (e.g., within 200 bp downstream of the mutation), or one is upstream (e.g., within 200 bp upstream of the mutation) and the second one is downstream (e.g., within 200 bp downstream of the mutation) of the mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, the break is positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat.

In an embodiment, two sets of breaks (e.g., two double strand breaks) are introduced (e.g., positioned by two gRNA molecules) in close proximity to a mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, two gRNA molecule are used to target two Cas9 nucleases to create two double strand breaks to flank a mutation (e.g., 2299delG), eg., one gRNA is used to target upstream of (e.g., within 200 bp upstream of the mutation) while a second gRNA is used to target downstream of (e.g., within 200 bp downstream of the mutation) of a mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, the breaks are positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat.

In an embodiment, two sets of breaks (e.g., one double strand break and a pair of nickases) are introduced (e.g., positioned by three gRNA molecules) in close proximity to a mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, three gRNA molecules are used to target three Cas9 molecules to create two sets of breaks (e.g., one double strand break and a pair of nickases)) to flank a mutation (e.g., 2299delG), eg., one gRNA molecule is used to target upstream or downstream of (e.g., within 200 bp upstream or downstream of the mutation) while a second and a third gRNA molecules are used to target the opposite site (e.g., within 200 bp downstream or upstream) of of a mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, the breaks are positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat.

In an embodiment, two sets of breaks (e.g., two pairs of strand breaks) are introduced (e.g., positioned by four gRNA molecules) in close proximity to a mutation (e.g., 2299delG) in the *USH2A* gene. In an embodiment, four gRNA molecule are used to target four Cas9 nickases

to create two pairs of single strand breaks to flank a mutation (e.g., 2299delG), eg., one and a second gRNA molecules are used to target upstream of (e.g., within 200 bp upstream of the mutation) while a third and a fourth gRNA molecules are used to target downstream of (e.g., within 200 bp downstream of the mutation) of a mutation (e.g., 2299delG) in the *USH2A* gene.

5 In an embodiment, the breaks are positioned to avoid unwanted target chromosome elements, such as repeat elements, e.g., an *Alu* repeat.

When two gRNAs designed for use to target two Cas9 enzymes, one Cas9 can be one species, the second Cas9 can be from a different species. Both Cas9 species are used to generate a single or double-strand break, as desired.

10

I. gRNA Molecules

A gRNA molecule, as that term is used herein, refers to a nucleic acid that promotes the specific targeting or homing of a gRNA molecule/Cas9 molecule complex to a target nucleic acid. gRNA molecules can be unimolecular (having a single RNA molecule), sometimes
 15 referred to herein as “chimeric” gRNAs, or modular (comprising more than one, and typically two, separate RNA molecules). A gRNA molecule comprises a number of domains. The gRNA molecule domains are described in more detail below.

Several exemplary gRNA structures, with domains indicated thereon, are provided in **Fig. 1A-1G**. While not wishing to be bound by theory, in an embodiment, with regard to the three
 20 dimensional form, or intra- or inter-strand interactions of an active form of a gRNA, regions of high complementarity are sometimes shown as duplexes in **Figs. 1A-1G** and other depictions provided herein.

In an embodiment, a unimolecular, or chimeric, gRNA comprises, preferably from 5' to 3':

- 25 a targeting domain (which is complementary to a target nucleic acid in the *USH2A* gene, e.g., a targeting domain from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**;
- a first complementarity domain;
- a linking domain;
- 30 a second complementarity domain (which is complementary to the first complementarity domain);

a proximal domain; and
optionally, a tail domain.

In an embodiment, a modular gRNA comprises:

- 5 a first strand comprising, preferably from 5' to 3';
 - a targeting domain (which is complementary to a target nucleic acid in the *USH2A* gene, e.g., a targeting domain from any of **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**; and
 - a first complementarity domain; and
- a second strand, comprising, preferably from 5' to 3':
 - 10 optionally, a 5' extension domain;
 - a second complementarity domain;
 - a proximal domain; and
 - optionally, a tail domain.

15 The domains are discussed briefly below:

The Targeting Domain

Figs. 1A-1G provide examples of the placement of targeting domains.

The targeting domain comprises a nucleotide sequence that is complementary, e.g., at least 80, 85, 90, or 95% complementary, e.g., fully complementary, to the target sequence on the target nucleic acid. The targeting domain is part of an RNA molecule and therefore comprises the base uracil (U), while any DNA encoding the gRNA molecule comprises the base thymine (T). While not wishing to be bound by theory, in an embodiment, it is believed that the complementarity of the targeting domain with the target sequence contributes to specificity of the interaction of the gRNA molecule/Cas9 molecule complex with a target nucleic acid. It is understood that in a targeting domain and target sequence pair, the uracil bases in the targeting domain will pair with the adenine bases in the target sequence. In an embodiment, the target domain itself comprises two domains, which are, in the 5' to 3' direction, an optional secondary domain, and a core domain. In an embodiment, the core domain is fully complementary with the target sequence. In an embodiment, the targeting domain is 5 to 50 nucleotides in length. The strand of the target nucleic acid with which the targeting domain is complementary is referred to

herein as the complementary strand. Some or all of the nucleotides of the domain can have a modification, e.g., a modification found in Section VIII herein.

In an embodiment, the targeting domain is 16 nucleotides in length.

In an embodiment, the targeting domain is 17 nucleotides in length.

5 In an embodiment, the targeting domain is 18 nucleotides in length.

In an embodiment, the targeting domain is 19 nucleotides in length.

In an embodiment, the targeting domain is 20 nucleotides in length.

In an embodiment, the targeting domain is 21 nucleotides in length.

In an embodiment, the targeting domain is 22 nucleotides in length.

10 In an embodiment, the targeting domain is 23 nucleotides in length.

In an embodiment, the targeting domain is 24 nucleotides in length.

In an embodiment, the targeting domain is 25 nucleotides in length.

In an embodiment, the targeting domain is 26 nucleotides in length.

In an embodiment, the targeting domain comprises 16 nucleotides.

15 In an embodiment, the targeting domain comprises 17 nucleotides.

In an embodiment, the targeting domain comprises 18 nucleotides.

In an embodiment, the targeting domain comprises 19 nucleotides.

In an embodiment, the targeting domain comprises 20 nucleotides.

In an embodiment, the targeting domain comprises 21 nucleotides.

20 In an embodiment, the targeting domain comprises 22 nucleotides.

In an embodiment, the targeting domain comprises 23 nucleotides.

In an embodiment, the targeting domain comprises 24 nucleotides.

In an embodiment, the targeting domain comprises 25 nucleotides.

In an embodiment, the targeting domain comprises 26 nucleotides.

25 Targeting domains are discussed in more detail below.

The First Complementarity Domain

Figs. 1A-1G provide examples of first complementarity domains.

The first complementarity domain is complementary with the second complementarity domain, and in an embodiment, has sufficient complementarity to the second complementarity domain to form a duplexed region under at least some physiological conditions. In an

30

embodiment, the first complementarity domain is 5 to 30 nucleotides in length. In an embodiment, the first complementarity domain is 5 to 25 nucleotides in length. In an embodiment, the first complementarity domain is 7 to 25 nucleotides in length. In an embodiment, the first complementarity domain is 7 to 22 nucleotides in length. In an embodiment, the first
 5 complementary domain is 7 to 18 nucleotides in length. In an embodiment, the first complementary domain is 7 to 15 nucleotides in length. In an embodiment, the first complementary domain is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length.

In an embodiment, the first complementarity domain comprises 3 subdomains, which, in the
 10 5' to 3' direction are: a 5' subdomain, a central subdomain, and a 3' subdomain. In an embodiment, the 5' subdomain is 4-9, e.g., 4, 5, 6, 7, 8 or 9 nucleotides in length. In an embodiment, the central subdomain is 1, 2, or 3, e.g., 1, nucleotide in length. In an embodiment, the 3' subdomain is 3 to 25, e.g., 4-22, 4-18, or 4 to 10, or 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25, nucleotides in length.

15 The first complementarity domain can share homology with, or be derived from, a naturally occurring first complementarity domain. In an embodiment, it has at least 50% homology with a first complementarity domain disclosed herein, e.g., an *S. pyogenes*, *S. aureus* or *S. thermophilus*, first complementarity domain.

Some or all of the nucleotides of the domain can have a modification, e.g., modification
 20 found in Section VIII herein.

First complementarity domains are discussed in more detail below.

The Linking Domain

Figs. 1A-1G provide examples of linking domains.

25 A linking domain serves to link the first complementarity domain with the second complementarity domain of a unimolecular gRNA. The linking domain can link the first and second complementarity domains covalently or non-covalently. In an embodiment, the linkage is covalent. In an embodiment, the linking domain covalently couples the first and second complementarity domains, see, e.g., **Figs. 1B-1E**. In an embodiment, the linking domain is, or
 30 comprises, a covalent bond interposed between the first complementarity domain and the second

complementarity domain. Typically the linking domain comprises one or more, e.g., 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides.

In modular gRNA molecules the two molecules are associated by virtue of the hybridization of the complementarity domains see e.g., **Fig. 1A**.

5 A wide variety of linking domains are suitable for use in unimolecular gRNA molecules. Linking domains can consist of a covalent bond, or be as short as one or a few nucleotides, e.g., 1, 2, 3, 4, or 5 nucleotides in length. In an embodiment, a linking domain is 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or 25 or more nucleotides in length. In an embodiment, a linking domain is 2 to 50, 2 to 40, 2 to 30, 2 to 20, 2 to 10, or 2 to 5 nucleotides in length. In an embodiment, a linking
10 domain shares homology with, or is derived from, a naturally occurring sequence, e.g., the sequence of a tracrRNA that is 5' to the second complementarity domain. In an embodiment, the linking domain has at least 50% homology with a linking domain disclosed herein.

Some or all of the nucleotides of the domain can have a modification, e.g., a modification found in Section VIII herein.

15 Linking domains are discussed in more detail below.

The 5' Extension Domain

In an embodiment, a modular gRNA can comprise additional sequence, 5' to the second complementarity domain, referred to herein as the 5' extension domain, see, e.g., **Fig. 1A**. In an
20 embodiment, the 5' extension domain is 2-10, 2-9, 2-8, 2-7, 2-6, 2-5, 2-4 nucleotides in length. In an embodiment, the 5' extension domain is 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more nucleotides in length.

The Second Complementarity Domain

25 **Figs. 1A-1G** provide examples of second complementarity domains.

The second complementarity domain is complementary with the first complementarity domain, and in an embodiment, has sufficient complementarity to the second complementarity domain to form a duplexed region under at least some physiological conditions. In an embodiment, e.g., as shown in **Figs. 1A-1B**, the second complementarity domain can include
30 sequence that lacks complementarity with the first complementarity domain, e.g., sequence that loops out from the duplexed region.

In an embodiment, the second complementarity domain is 5 to 27 nucleotides in length. In an embodiment, it is longer than the first complementarity region. In an embodiment, the second complementary domain is 7 to 27 nucleotides in length. In an embodiment, the second complementary domain is 7 to 25 nucleotides in length. In an embodiment, the second complementary domain is 7 to 20 nucleotides in length. In an embodiment, the second complementary domain is 7 to 17 nucleotides in length. In an embodiment, the complementary domain is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides in length.

In an embodiment, the second complementarity domain comprises three subdomains, which, in the 5' to 3' direction are: a 5' subdomain, a central subdomain, and a 3' subdomain. In an embodiment, the 5' subdomain is 3 to 25, e.g., 4 to 22, 4 to 18, or 4 to 10, or 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides in length. In an embodiment, the central subdomain is 1, 2, 3, 4 or 5, e.g., 3, nucleotides in length. In an embodiment, the 3' subdomain is 4 to 9, e.g., 4, 5, 6, 7, 8 or 9 nucleotides in length.

In an embodiment, the 5' subdomain and the 3' subdomain of the first complementarity domain, are respectively, complementary, e.g., fully complementary, with the 3' subdomain and the 5' subdomain of the second complementarity domain.

The second complementarity domain can share homology with or be derived from a naturally occurring second complementarity domain. In an embodiment it has at least 50% homology with a second complementarity domain disclosed herein, e.g., an *S. pyogenes*, *S. aureus* or *S. thermophilus*, second complementarity domain.

Some or all of the nucleotides of the domain can have a modification, e.g., modification found in Section VIII herein.

The Proximal Domain

Figs. 1A-1G provide examples of proximal domains.

In an embodiment, the proximal domain is 5 to 20 nucleotides in length. In an embodiment, the proximal domain can share homology with or be derived from a naturally occurring proximal domain. In an embodiment, it has at least 50% homology with a proximal domain disclosed herein, e.g., an *S. pyogenes*, *S. aureus* or *S. thermophilus*, proximal domain.

Some or all of the nucleotides of the domain can have a modification, e.g., modification found in Section VIII herein.

The Tail Domain

Figs. 1A-1G provide examples of tail domains.

As can be seen by inspection of the tail domains in **Figs. 1A-1G**, a broad spectrum of tail domains are suitable for use in gRNA molecules. In an embodiment, the tail domain is 0 (absent), 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In embodiment, the tail domain nucleotides are from or share homology with sequence from the 5' end of a naturally occurring tail domain, see e.g., **Fig. 1D** or **1E**. In an embodiment, the tail domain includes sequences that are complementary to each other and which, under at least some physiological conditions, form a duplexed region.

In an embodiment, the tail domain is absent or is 1 to 50 nucleotides in length. In an embodiment, the tail domain can share homology with or be derived from a naturally occurring proximal tail domain. In an embodiment, it has at least 50% homology with a tail domain disclosed herein, e.g., an *S. pyogenes*, *S. aureus* or *S. thermophilus*, tail domain.

In an embodiment, the tail domain includes nucleotides at the 3' end that are related to the method of in vitro or in vivo transcription. When a T7 promoter is used for in vitro transcription of the gRNA, these nucleotides may be any nucleotides present before the 3' end of the DNA template. When a U6 promoter is used for in vivo transcription, these nucleotides may be the sequence UUUUUU. When alternate pol-III promoters are used, these nucleotides may be various numbers or uracil bases or may include alternate bases.

The domains of gRNA molecules are described in more detail below.

The Targeting Domain

The "targeting domain" of the gRNA is complementary to the "target domain" on the target nucleic acid. The strand of the target nucleic acid comprising the core domain target is referred to herein as the "complementary strand" of the target nucleic acid. Guidance on the selection of targeting domains can be found, e.g., in Fu Y *et al.*, NAT BIOTECHNOL 2014 (doi: 10.1038/nbt.2808) and Sternberg SH *et al.*, NATURE 2014 (doi: 10.1038/nature13011).

In an embodiment, the targeting domain is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In an embodiment, the targeting domain is 16 nucleotides in length.

In an embodiment, the targeting domain is 17 nucleotides in length.

5 In an embodiment, the targeting domain is 18 nucleotides in length.

In an embodiment, the targeting domain is 19 nucleotides in length.

In an embodiment, the targeting domain is 20 nucleotides in length.

In an embodiment, the targeting domain is 21 nucleotides in length.

In an embodiment, the targeting domain is 22 nucleotides in length.

10 In an embodiment, the targeting domain is 23 nucleotides in length.

In an embodiment, the targeting domain is 24 nucleotides in length.

In an embodiment, the targeting domain is 25 nucleotides in length.

In an embodiment, the targeting domain is 26 nucleotides in length.

In an embodiment, the targeting domain comprises 16 nucleotides.

15 In an embodiment, the targeting domain comprises 17 nucleotides.

In an embodiment, the targeting domain comprises 18 nucleotides.

In an embodiment, the targeting domain comprises 19 nucleotides.

In an embodiment, the targeting domain comprises 20 nucleotides.

In an embodiment, the targeting domain comprises 21 nucleotides.

20 In an embodiment, the targeting domain comprises 22 nucleotides.

In an embodiment, the targeting domain comprises 23 nucleotides.

In an embodiment, the targeting domain comprises 24 nucleotides.

In an embodiment, the targeting domain comprises 25 nucleotides.

In an embodiment, the targeting domain comprises 26 nucleotides.

25 In an embodiment, the targeting domain is 10 +/-5, 20 +/-5, 30 +/-5, 40 +/-5, 50 +/-5, 60 +/-5, 70 +/-5, 80 +/-5, 90 +/-5, or 100 +/-5 nucleotides, in length.

In an embodiment, the targeting domain is 20 +/-5 nucleotides in length.

In an embodiment, the targeting domain is 20 +/-10, 30 +/-10, 40 +/-10, 50 +/-10, 60 +/-10, 70 +/-10, 80 +/-10, 90 +/-10, or 100 +/-10 nucleotides, in length.

30 In an embodiment, the targeting domain is 30 +/-10 nucleotides in length.

In an embodiment, the targeting domain is 10 to 100, 10 to 90, 10 to 80, 10 to 70, 10 to 60, 10 to 50, 10 to 40, 10 to 30, 10 to 20 or 10 to 15 nucleotides in length. In other embodiments, the targeting domain is 20 to 100, 20 to 90, 20 to 80, 20 to 70, 20 to 60, 20 to 50, 20 to 40, 20 to 30, or 20 to 25 nucleotides in length.

5 Typically the targeting domain has full complementarity with the target sequence. In some embodiments, the targeting domain has or includes 1, 2, 3, 4, 5, 6, 7 or 8 nucleotides that are not complementary with the corresponding nucleotide of the targeting domain.

In an embodiment, the target domain includes 1, 2, 3, 4 or 5 nucleotides that are complementary with the corresponding nucleotide of the targeting domain within 5 nucleotides
10 of its 5' end. In an embodiment, the target domain includes 1, 2, 3, 4 or 5 nucleotides that are complementary with the corresponding nucleotide of the targeting domain within 5 nucleotides of its 3' end.

In an embodiment, the target domain includes 1, 2, 3, or 4 nucleotides that are not complementary with the corresponding nucleotide of the targeting domain within 5 nucleotides
15 of its 5' end. In an embodiment, the target domain includes 1, 2, 3, or 4 nucleotides that are not complementary with the corresponding nucleotide of the targeting domain within 5 nucleotides of its 3' end.

In an embodiment, the degree of complementarity, together with other properties of the gRNA, is sufficient to allow targeting of a Cas9 molecule to the target nucleic acid.

20 In some embodiments, the targeting domain comprises two consecutive nucleotides that are not complementary to the target domain ("non-complementary nucleotides"), e.g., two consecutive noncomplementary nucleotides that are within 5 nucleotides of the 5' end of the targeting domain, within 5 nucleotides of the 3' end of the targeting domain, or more than 5 nucleotides away from one or both ends of the targeting domain.

25 In an embodiment, no two consecutive nucleotides within 5 nucleotides of the 5' end of the targeting domain, within 5 nucleotides of the 3' end of the targeting domain, or within a region that is more than 5 nucleotides away from one or both ends of the targeting domain, are not complementary to the targeting domain.

In an embodiment, there are no non-complementary nucleotides within 5 nucleotides of
30 the 5' end of the targeting domain, within 5 nucleotides of the 3' end of the targeting domain, or

within a region that is more than 5 nucleotides away from one or both ends of the targeting domain.

In an embodiment, the targeting domain nucleotides do not comprise modifications, e.g., modifications of the type provided in Section VIII. However, in an embodiment, the targeting domain comprises one or more modifications, e.g., modifications that it render it less susceptible to degradation or more bio-compatible, e.g., less immunogenic. By way of example, the backbone of the targeting domain can be modified with a phosphorothioate, or other modification(s) from Section VIII. In an embodiment, a nucleotide of the targeting domain can comprise a 2' modification, e.g., a 2'-acetylation, e.g., a 2' methylation, or other modification(s) from Section VIII.

In some embodiments, the targeting domain includes 1, 2, 3, 4, 5, 6, 7 or 8 or more modifications. In an embodiment, the targeting domain includes 1, 2, 3, or 4 modifications within 5 nucleotides of its 5' end. In an embodiment, the targeting domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 3' end.

In some embodiments, the targeting domain comprises modifications at two consecutive nucleotides, e.g., two consecutive nucleotides that are within 5 nucleotides of the 5' end of the targeting domain, within 5 nucleotides of the 3' end of the targeting domain, or more than 5 nucleotides away from one or both ends of the targeting domain.

In an embodiment, no two consecutive nucleotides are modified within 5 nucleotides of the 5' end of the targeting domain, within 5 nucleotides of the 3' end of the targeting domain, or within a region that is more than 5 nucleotides away from one or both ends of the targeting domain. In an embodiment, no nucleotide is modified within 5 nucleotides of the 5' end of the targeting domain, within 5 nucleotides of the 3' end of the targeting domain, or within a region that is more than 5 nucleotides away from one or both ends of the targeting domain.

Modifications in the targeting domain can be selected so as to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in the system described in Section IV. gRNAs having a candidate targeting domain having a selected length, sequence, degree of complementarity, or degree of modification, can be evaluated in a system in Section IV. The candidate targeting domain can be placed, either alone, or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

In some embodiments, all of the modified nucleotides are complementary to and capable of hybridizing to corresponding nucleotides present in the target domain. In other embodiments, 1, 2, 3, 4, 5, 6, 7 or 8 or more modified nucleotides are not complementary to or capable of hybridizing to corresponding nucleotides present in the target domain.

5 In an embodiment, the targeting domain comprises, preferably in the 5'→3' direction: a secondary domain and a core domain. These domains are discussed in more detail below.

The Core Domain and Secondary Domain of the Targeting Domain

10 The “core domain” of the targeting domain is complementary to the “core domain target” on the target nucleic acid. In an embodiment, the core domain comprises about 8 to about 13 nucleotides from the 3' end of the targeting domain (e.g., the most 3' 8 to 13 nucleotides of the targeting domain). In an embodiment, the secondary domain is absent or optional.

In an embodiment, the secondary domain is absent or optional.

15 In an embodiment, the core domain and targeting domain, are independently, 6 +/-2, 7 +/-2, 8 +/-2, 9 +/-2, 10 +/-2, 11 +/-2, 12 +/-2, 13 +/-2, 14 +/-2, 15 +/-2, 16 +/-2, 17 +/-2, or 18 +/-2, nucleotides in length.

In an embodiment, the core domain and targeting domain, are independently, 10 +/-2 nucleotides in length.

20 In an embodiment, the core domain and targeting domain are independently 10 +/-4 nucleotides in length.

In an embodiment, the core domain and targeting domain are independently 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18, nucleotides in length.

In an embodiment, the core domain and targeting domain are independently 3 to 20, 4 to 20, 5 to 20, 6 to 20, 7 to 20, 8 to 20, 9 to 20, 10 to 20 or 15 to 20 nucleotides in length.

25 In an embodiment, the core domain and targeting domain are independently 3 to 15, e.g., 6 to 15, 7 to 14, 7 to 13, 6 to 12, 7 to 12, 7 to 11, 7 to 10, 8 to 14, 8 to 13, 8 to 12, 8 to 11, 8 to 10 or 8 to 9 nucleotides in length.

30 The core domain is complementary with the core domain target. Typically the core domain has exact complementarity with the core domain target. In some embodiments, the core domain can have 1, 2, 3, 4 or 5 nucleotides that are not complementary with the corresponding nucleotide of the core domain. In an embodiment, the degree of complementarity, together with

other properties of the gRNA, is sufficient to allow targeting of a Cas9 molecule to the target nucleic acid.

The “secondary domain” of the targeting domain of the gRNA is complementary to the “secondary domain target” of the target nucleic acid.

5 In an embodiment, the secondary domain is positioned 5' to the core domain.

In an embodiment, the secondary domain is absent or optional.

In an embodiment, if the targeting domain is 26 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 12 to 17 nucleotides in length.

10 In an embodiment, if the targeting domain is 25 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 12 to 17 nucleotides in length.

In an embodiment, if the targeting domain is 24 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the
15 secondary domain is 11 to 16 nucleotides in length.

In an embodiment, if the targeting domain is 23 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 10 to 15 nucleotides in length.

In an embodiment, if the targeting domain is 22 nucleotides in length and the core
20 domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 9 to 14 nucleotides in length.

In an embodiment, if the targeting domain is 21 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 8 to 13 nucleotides in length.

25 In an embodiment, if the targeting domain is 20 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 7 to 12 nucleotides in length.

In an embodiment, if the targeting domain is 19 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the
30 secondary domain is 6 to 11 nucleotides in length.

In an embodiment, if the targeting domain is 18 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 5 to 10 nucleotides in length.

5 In an embodiment, if the targeting domain is 17 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 4 to 9 nucleotides in length.

In an embodiment, if the targeting domain is 16 nucleotides in length and the core domain (counted from the 3' end of the targeting domain) is 8 to 13 nucleotides in length, the secondary domain is 3 to 8 nucleotides in length.

10 In an embodiment, the secondary domain is 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 nucleotides in length.

The secondary domain is complementary with the secondary domain target. Typically the secondary domain has exact complementarity with the secondary domain target. In some embodiments the secondary domain can have 1, 2, 3, 4 or 5 nucleotides that are not
15 complementary with the corresponding nucleotide of the secondary domain. In an embodiment, the degree of complementarity, together with other properties of the gRNA, is sufficient to allow targeting of a Cas9 molecule to the target nucleic acid.

In an embodiment, the core domain nucleotides do not comprise modifications, e.g.,
20 modifications of the type provided in Section VIII. However, in an embodiment, the core domain comprise one or more modifications, e.g., modifications that it render it less susceptible to degradation or more bio-compatible, e.g., less immunogenic. By way of example, the backbone of the core domain can be modified with a phosphorothioate, or other modification(s) from Section VIII. In an embodiment a nucleotide of the core domain can comprise a 2'
25 modification, e.g., a 2'-acetylation, e.g., a 2' methylation, or other modification(s) from Section VIII. Typically, a core domain will contain no more than 1, 2, or 3 modifications.

Modifications in the core domain can be selected so as to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in the system described in Section IV. gRNAs having a candidate core domain having a selected length, sequence, degree
30 of complementarity, or degree of modification, can be evaluated in the system described at Section IV. The candidate core domain can be placed, either alone, or with one or more other

candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

In an embodiment, the secondary domain nucleotides do not comprise modifications, e.g., modifications of the type provided in Section VIII. However, in an embodiment, the secondary domain comprises one or more modifications, e.g., modifications that it render it less susceptible to degradation or more bio-compatible, e.g., less immunogenic. By way of example, the backbone of the secondary domain can be modified with a phosphorothioate, or other modification(s) from Section VIII. In an embodiment a nucleotide of the secondary domain can comprise a 2' modification, e.g., a 2-acetylation, e.g., a 2' methylation, or other modification(s) from Section VIII. Typically, a secondary domain will contain no more than 1, 2, or 3 modifications.

Modifications in the secondary domain can be selected so as to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in the system described in Section IV. gRNAs having a candidate secondary domain having a selected length, sequence, degree of complementarity, or degree of modification, can be evaluated in the system described at Section IV. The candidate secondary domain can be placed, either alone, or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

In an embodiment, (1) the degree of complementarity between the core domain and its target, and (2) the degree of complementarity between the secondary domain and its target, may differ. In an embodiment, (1) may be greater (2). In an embodiment, (1) may be less than (2). In an embodiment, (1) and (2) may be the same, e.g., each may be completely complementary with its target.

In an embodiment, (1) the number of modification (e.g., modifications from Section VIII) of the nucleotides of the core domain and (2) the number of modification (e.g., modifications from Section VIII) of the nucleotides of the secondary domain, may differ. In an embodiment, (1) may be less than (2). In an embodiment, (1) may be greater than (2). In an embodiment, (1) and (2) may be the same, e.g., each may be free of modifications.

The First and Second Complementarity Domains

The first complementarity domain is complementary with the second complementarity domain.

Typically the first domain does not have exact complementarity with the second complementarity domain target. In some embodiments, the first complementarity domain can have 1, 2, 3, 4 or 5 nucleotides that are not complementary with the corresponding nucleotide of the second complementarity domain. In an embodiment, 1, 2, 3, 4, 5 or 6, e.g., 3 nucleotides, do not pair in the duplex, and, e.g., form a non-duplexed or looped-out region. In an embodiment an unpaired, or loop-out, region, e.g., a loop-out of 3 nucleotides, is present on the second complementarity domain. In an embodiment, the unpaired region begins 1, 2, 3, 4, 5, or 6, e.g., 4, nucleotides from the 5' end of the second complementarity domain.

In an embodiment, the degree of complementarity, together with other properties of the gRNA, is sufficient to allow targeting of a Cas9 molecule to the target nucleic acid.

In an embodiment, the first and second complementarity domains are:

independently, 6 +/-2, 7 +/-2, 8 +/-2, 9 +/-2, 10 +/-2, 11 +/-2, 12 +/-2, 13 +/-2, 14 +/-2, 15 +/-2, 16 +/-2, 17 +/-2, 18 +/-2, 19 +/-2, or 20 +/-2, 21 +/-2, 22 +/-2, 23 +/-2, or 24 +/-2 nucleotides in length;

independently, 6, 7, 8, 9, 10, 11, 12, 13, 14, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26, nucleotides in length;

independently, 5 to 24, 5 to 23, 5 to 22, 5 to 21, 5 to 20, 7 to 18, 9 to 16, or 10 to 14 nucleotides in length.

In an embodiment, the second complementarity domain is longer than the first complementarity domain, e.g., 2, 3, 4, 5, or 6, e.g., 6, nucleotides longer.

In an embodiment, the first and second complementary domains, independently, do not comprise modifications, e.g., modifications of the type provided in Section VIII.

In an embodiment, the first and second complementary domains, independently, comprise one or more modifications, e.g., modifications that render the domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. By way of example, the backbone of the domain can be modified with a phosphorothioate, or other modification(s) from Section VIII. In an embodiment a nucleotide of the domain can comprise a 2' modification, e.g., a 2-acetylation, e.g., a 2' methylation, or other modification(s) from Section VIII.

In an embodiment, the first and second complementary domains, independently, include 1, 2, 3, 4, 5, 6, 7 or 8 or more modifications. In an embodiment, the first and second complementary domains, independently, include 1, 2, 3, or 4 modifications within 5 nucleotides of its 5' end. In an embodiment, the first and second complementary domains, independently, include as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 3' end.

In an embodiment, the first and second complementary domains, independently, include modifications at two consecutive nucleotides, e.g., two consecutive nucleotides that are within 5 nucleotides of the 5' end of the domain, within 5 nucleotides of the 3' end of the domain, or more than 5 nucleotides away from one or both ends of the domain. In an embodiment, the first and second complementary domains, independently, include no two consecutive nucleotides that are modified, within 5 nucleotides of the 5' end of the domain, within 5 nucleotides of the 3' end of the domain, or within a region that is more than 5 nucleotides away from one or both ends of the domain. In an embodiment, the first and second complementary domains, independently, include no nucleotide that is modified within 5 nucleotides of the 5' end of the domain, within 5 nucleotides of the 3' end of the domain, or within a region that is more than 5 nucleotides away from one or both ends of the domain.

Modifications in a complementarity domain can be selected so as to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in the system described in Section IV. gRNAs having a candidate complementarity domain having a selected length, sequence, degree of complementarity, or degree of modification, can be evaluated in the system described in Section IV. The candidate complementarity domain can be placed, either alone, or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

In an embodiment, the first complementarity domain has at least 60, 70, 80, 85%, 90% or 95% homology with, or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from, a reference first complementarity domain, e.g., a naturally occurring, e.g., an *S. pyogenes*, *S. aureus* or *S. thermophilus*, first complementarity domain, or a first complementarity domain described herein, e.g., from **Figs. 1A-1G**.

In an embodiment, the second complementarity domain has at least 60, 70, 80, 85%, 90%, or 95% homology with, or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from, a reference second complementarity domain, e.g., a naturally occurring, e.g., an *S. pyogenes*, *S.*

aureus or *S. thermophilus*, second complementarity domain, or a second complementarity domain described herein, e.g., from **Figs. 1A-1G**.

The duplexed region formed by first and second complementarity domains is typically 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 base pairs in length (excluding any looped out or unpaired nucleotides).

In some embodiments, the first and second complementarity domains, when duplexed, comprise 11 paired nucleotides, for example in the gRNA sequence (one paired strand underlined, one bolded):

NNNNNNNNNNNNNNNNNNNNNNNGUUUUAGAGGCUAGAAA**UAGCAAGU**AAAAUAAGG
CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 5).

In some embodiments the first and second complementarity domains, when duplexed, comprise 15 paired nucleotides, for example in the gRNA sequence (one paired strand underlined, one bolded):

NNNNNNNNNNNNNNNNNNNNNNNGUUUUAGAGGCUAUGCUGAAA**AGCAUAGCAAGU**UA
AAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ
ID NO: 27).

In some embodiments the first and second complementarity domains, when duplexed, comprise 16 paired nucleotides, for example in the gRNA sequence (one paired strand underlined, one bolded):

NNNNNNNNNNNNNNNNNNNNNNNGUUUUAGAGGCUAUGCUGGAA**ACAGCAUAGCAAGU**
UAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC
(SEQ ID NO: 28).

In some embodiments the first and second complementarity domains, when duplexed, comprise 21 paired nucleotides, for example in the gRNA sequence (one paired strand underlined, one bolded):

NNNNNNNNNNNNNNNNNNNNNNNGUUUUAGAGGCUAUGCUGUUUUGGAA**ACAAAACAG**
CAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA
GUCGGUGC (SEQ ID NO: 29).

In some embodiments, nucleotides are exchanged to remove poly-U tracts, for example in the gRNA sequences (exchanged nucleotides underlined):

NNNNNNNNNNNNNNNNNNNNNNNNNGUAUUAGAGCUAGAAAUAGCAAGUAAUAUAAGG
CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 30).;

NNNNNNNNNNNNNNNNNNNNNNNNNGUUUAAGAGCUAGAAAUAGCAAGUUUAAAUAAGG
CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO: 31).; or

NNNNNNNNNNNNNNNNNNNNNNNNNGUAUUAGAGCUAUGCUGUAUUGGAAACAAAACAG
CAUAGCAAGUAAUAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA
GUCGGUGC (SEQ ID NO: 32)..

The 5' Extension Domain

In an embodiment, a modular gRNA can comprise additional sequence, 5' to the second complementarity domain. In an embodiment, the 5' extension domain is 2 to 10, 2 to 9, 2 to 8, 2 to 7, 2 to 6, 2 to 5, or 2 to 4 nucleotides in length. In an embodiment, the 5' extension domain is 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more nucleotides in length.

In an embodiment, the 5' extension domain nucleotides do not comprise modifications, e.g., modifications of the type provided in Section VIII. However, in an embodiment, the 5' extension domain comprises one or more modifications, e.g., modifications that it render it less susceptible to degradation or more bio-compatible, e.g., less immunogenic. By way of example, the backbone of the 5' extension domain can be modified with a phosphorothioate, or other modification(s) from Section VIII. In an embodiment, a nucleotide of the 5' extension domain can comprise a 2' modification, e.g., a 2-acetylation, e.g., a 2' methylation, or other modification(s) from Section VIII.

In some embodiments, the 5' extension domain can comprise as many as 1, 2, 3, 4, 5, 6, 7 or 8 modifications. In an embodiment the 5' extension domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 5' end, e.g., in a modular gRNA molecule. In an embodiment the 5' extension domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 3' end, e.g., in a modular gRNA molecule.

In some embodiments, the 5' extension domain comprises modifications at two consecutive nucleotides, e.g., two consecutive nucleotides that are within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension domain, or

more than 5 nucleotides away from one or both ends of the 5' extension domain. In an embodiment, no two consecutive nucleotides are modified within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension domain, or within a region that is more than 5 nucleotides away from one or both ends of the 5' extension domain.

- 5 In an embodiment, no nucleotide is modified within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension domain, or within a region that is more than 5 nucleotides away from one or both ends of the 5' extension domain.

Modifications in the 5' extension domain can be selected so as to not interfere with gRNA molecule efficacy, which can be evaluated by testing a candidate modification in the system described in Section IV. gRNAs having a candidate 5' extension domain having a selected length, sequence, degree of complementarity, or degree of modification, can be evaluated in the system described at Section IV. The candidate 5' extension domain can be placed, either alone, or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

- 15 In an embodiment, the 5' extension domain has at least 60, 70, 80, 85, 90 or 95% homology with, or differs by no more than 1, 2, 3, 4, 5 or 6 nucleotides from, a reference 5' extension domain, e.g., a naturally occurring, e.g., an *S. pyogenes*, *S. aureus* or *S. thermophilus*, 5' extension domain, or a 5' extension domain described herein, e.g., from **Figs. 1A-1G**.

20 The Linking Domain

In a unimolecular gRNA molecule, the linking domain is disposed between the first and second complementarity domains. In a modular gRNA molecule, the two molecules are associated with one another by the complementarity domains.

- In an embodiment, the linking domain is 10 +/-5, 20+/-5, 30+/-5, 40+/-5, 50+/-5, 60+/-5, 25 70+/-5, 80+/-5, 90+/-5, or 100+/-5 nucleotides, in length.

In an embodiment, the linking domain is 20+/-10, 30+/-10, 40+/-10, 50+/-10, 60+/-10, 70+/-10, 80+/-10, 90+/-10, or 100+/-10 nucleotides in length.

- In an embodiment, the linking domain is 10 to 100, 10 to 90, 10 to 80, 10 to 70, 10 to 60, 10 to 50, 10 to 40, 10 to 30, 10 to 20 or 10 to 15 nucleotides in length. In other embodiments, 30 the linking domain is 20 to 100, 20 to 90, 20 to 80, 20 to 70, 20 to 60, 20 to 50, 20 to 40, 20 to 30, or 20 to 25 nucleotides in length.

In an embodiment, the linking domain is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length.

In an embodiment, the linking domain is a covalent bond.

In an embodiment, the linking domain comprises a duplexed region, typically adjacent to
5 or within 1, 2, or 3 nucleotides of the 3' end of the first complementarity domain and/or the 5-
end of the second complementarity domain. In an embodiment, the duplexed region can be
20+/-10 base pairs in length. In an embodiment, the duplexed region can be 10+/-5, 15+/-5,
20+/-5, or 30+/-5 base pairs in length. In an embodiment, the duplexed region can be 1, 2, 3, 4,
5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 base pairs in length.

10 Typically the sequences forming the duplexed region have exact complementarity with
one another, though in some embodiments as many as 1, 2, 3, 4, 5, 6, 7 or 8 nucleotides are not
complementary with the corresponding nucleotides.

In an embodiment, the linking domain nucleotides do not comprise modifications, e.g.,
modifications of the type provided in Section VIII. However, in an embodiment, the linking
15 domain comprises one or more modifications, e.g., modifications that it render it less susceptible
to degradation or more bio-compatible, e.g., less immunogenic. By way of example, the
backbone of the linking domain can be modified with a phosphorothioate, or other
modification(s) from Section VIII. In an embodiment, a nucleotide of the linking domain can
comprise a 2' modification, e.g., a 2'-acetylation, e.g., a 2' methylation, or other modification(s)
20 from Section VIII.

In some embodiments, the linking domain can comprise as many as 1, 2, 3, 4, 5, 6, 7 or 8
modifications.

Modifications in a linking domain can be selected so as to not interfere with targeting
efficacy, which can be evaluated by testing a candidate modification in the system described in
25 Section IV. gRNAs having a candidate linking domain having a selected length, sequence,
degree of complementarity, or degree of modification, can be evaluated a system described in
Section IV. A candidate linking domain can be placed, either alone, or with one or more other
candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a
selected target and evaluated.

In an embodiment, the linking domain has at least 60, 70, 80, 85, 90 or 95% homology with, or differs by no more than 1, 2, 3, 4, 5 or 6 nucleotides from, a reference linking domain, e.g., a linking domain described herein, e.g., from **Figs. 1A-1G**.

The Proximal Domain

In an embodiment, the proximal domain is 6 +/-2, 7 +/-2, 8 +/-2, 9 +/-2, 10 +/-2, 11 +/-2, 12 +/-2, 13 +/-2, 14 +/-2, 14 +/-2, 16 +/-2, 17 +/-2, 18 +/-2, 19 +/-2, or 20 +/-2 nucleotides in length.

In an embodiment, the proximal domain is 6, 7, 8, 9, 10, 11, 12, 13, 14, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In an embodiment, the proximal domain is 5 to 20, 7, to 18, 9 to 16, or 10 to 14 nucleotides in length.

In an embodiment, the proximal domain nucleotides do not comprise modifications, e.g., modifications of the type provided in Section VIII. However, in an embodiment, the proximal domain comprises one or more modifications, e.g., modifications that it render it less susceptible to degradation or more bio-compatible, e.g., less immunogenic. By way of example, the backbone of the proximal domain can be modified with a phosphorothioate, or other modification(s) from Section VIII. In an embodiment, a nucleotide of the proximal domain can comprise a 2' modification, e.g., a 2-acetylation, e.g., a 2' methylation, or other modification(s) from Section VIII.

In some embodiments, the proximal domain can comprise as many as 1, 2, 3, 4, 5, 6, 7 or 8 modifications. In an embodiment, the proximal domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 5' end, e.g., in a modular gRNA molecule. In an embodiment, the target domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 3' end, e.g., in a modular gRNA molecule.

In some embodiments, the proximal domain comprises modifications at two consecutive nucleotides, e.g., two consecutive nucleotides that are within 5 nucleotides of the 5' end of the proximal domain, within 5 nucleotides of the 3' end of the proximal domain, or more than 5 nucleotides away from one or both ends of the proximal domain. In an embodiment, no two consecutive nucleotides are modified within 5 nucleotides of the 5' end of the proximal domain, within 5 nucleotides of the 3' end of the proximal domain, or within a region that is more than 5 nucleotides away from one or both ends of the proximal domain. In an embodiment, no

nucleotide is modified within 5 nucleotides of the 5' end of the proximal domain, within 5 nucleotides of the 3' end of the proximal domain, or within a region that is more than 5 nucleotides away from one or both ends of the proximal domain.

Modifications in the proximal domain can be selected so as to not interfere with gRNA molecule efficacy, which can be evaluated by testing a candidate modification in the system described in Section IV. gRNAs having a candidate proximal domain having a selected length, sequence, degree of complementarity, or degree of modification, can be evaluated in the system described at Section IV. The candidate proximal domain can be placed, either alone, or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

In an embodiment, the proximal domain has at least 60, 70, 80, 85 90 or 95% homology with, or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from, a reference proximal domain, e.g., a naturally occurring, e.g., an *S. pyogenes*, *S. aureus* or *S. thermophilus*, proximal domain, or a proximal domain described herein, e.g., from **Figs. 1A-1G**.

The Tail Domain

In an embodiment, the tail domain is 10 +/-5, 20+/-5, 30+/-5, 40+/-5, 50+/-5, 60+/-5, 70+/-5, 80+/-5, 90+/-5, or 100+/-5 nucleotides in length.

In an embodiment, the tail domain is 20+/-5 nucleotides in length.

In an embodiment, the tail domain is 20+/-10, 30+/-10, 40+/-10, 50+/-10, 60+/-10, 70+/-10, 80+/-10, 90+/-10, or 100+/-10 nucleotides, in length.

In an embodiment, the tail domain is 25+/-10 nucleotides in length.

In an embodiment, the tail domain is 10 to 100, 10 to 90, 10 to 80, 10 to 70, 10 to 60, 10 to 50, 10 to 40, 10 to 30, 10 to 20 or 10 to 15 nucleotides in length.

In other embodiments, the tail domain is 20 to 100, 20 to 90, 20 to 80, 20 to 70, 20 to 60, 20 to 50, 20 to 40, 20 to 30, or 20 to 25 nucleotides in length.

In an embodiment, the tail domain is 1 to 20, 1 to 1, 1 to 10, or 1 to 5 nucleotides in length.

In an embodiment, the tail domain nucleotides do not comprise modifications, e.g., modifications of the type provided in Section VIII. However, in an embodiment, the tail domain comprises one or more modifications, e.g., modifications that it render it less susceptible to

degradation or more bio-compatible, e.g., less immunogenic. By way of example, the backbone of the tail domain can be modified with a phosphorothioate, or other modification(s) from Section VIII. In an embodiment, a nucleotide of the tail domain can comprise a 2' modification, e.g., a 2-acetylation, e.g., a 2' methylation, or other modification(s) from Section VIII.

5 In some embodiments, the tail domain can have as many as 1, 2, 3, 4, 5, 6, 7 or 8 modifications. In an embodiment, the target domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 5' end. In an embodiment, the target domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 3' end.

10 In an embodiment, the tail domain comprises a tail duplex domain, which can form a tail duplexed region. In an embodiment, the tail duplexed region can be 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 base pairs in length. In an embodiment, a further single stranded domain, exists 3' to the tail duplexed domain. In an embodiment, this domain is 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In an embodiment, it is 4 to 6 nucleotides in length.

15 In an embodiment, the tail domain has at least 60, 70, 80, 90 or 95% homology with, or differs by no more than 1, 2, 3, 4, 5 or 6 nucleotides from, a reference tail domain, e.g., a naturally occurring, e.g., an *S. pyogenes*, *S. aureus* or *S. thermophilus*, tail domain, or a tail domain described herein, e.g., from **Figs. 1A-1G**.

In an embodiment, the proximal and tail domain, taken together comprise the following sequences:

20 AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCU (SEQ ID NO: 33), or
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC (SEQ ID NO: 34), or
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCGGAUC (SEQ
25 ID NO: 35), or
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUG (SEQ ID NO: 36), or
AAGGCUAGUCCGUUAUCA (SEQ ID NO: 37), or AAGGCUAGUCCG (SEQ ID NO: 38).

30 In an embodiment, the tail domain comprises the 3' sequence UUUUUU, e.g., if a U6 promoter is used for transcription.

In an embodiment, the tail domain comprises the 3' sequence UUUU, e.g., if an H1 promoter is used for transcription.

In an embodiment, tail domain comprises variable numbers of 3' Us depending, e.g., on the termination signal of the pol-III promoter used.

In an embodiment, the tail domain comprises variable 3' sequence derived from the DNA template if a T7 promoter is used.

In an embodiment, the tail domain comprises variable 3' sequence derived from the DNA template, e.g., if in vitro transcription is used to generate the RNA molecule.

5 In an embodiment, the tail domain comprises variable 3' sequence derived from the DNA template, e., if a pol-II promoter is used to drive transcription.

Modifications in the tail domain can be selected so as to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in the system described in Section IV. gRNAs having a candidate tail domain having a selected length, sequence, degree of
10 complementarity, or degree of modification, can be evaluated in the system described in Section IV. The candidate tail domain can be placed, either alone, or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

In some embodiments, the tail domain comprises modifications at two consecutive
15 nucleotides, e.g., two consecutive nucleotides that are within 5 nucleotides of the 5' end of the tail domain, within 5 nucleotides of the 3' end of the tail domain, or more than 5 nucleotides away from one or both ends of the tail domain. In an embodiment, no two consecutive nucleotides are modified within 5 nucleotides of the 5' end of the tail domain, within 5
20 nucleotides of the 3' end of the tail domain, or within a region that is more than 5 nucleotides away from one or both ends of the tail domain. In an embodiment, no nucleotide is modified within 5 nucleotides of the 5' end of the tail domain, within 5 nucleotides of the 3' end of the tail domain, or within a region that is more than 5 nucleotides away from one or both ends of the tail domain.

In an embodiment, a gRNA has the following structure:

25 5' [targeting domain]-[first complementarity domain]-[linking domain]-[second complementarity domain]-[proximal domain]-[tail domain]-3',

wherein the targeting domain comprises a core domain and, optionally, a secondary domain, and is 10 to 50 nucleotides in length;

the first complementarity domain is 5 to 25 nucleotides in length and, in an embodiment,
30 has at least 50, 60, 70, 80, 85, 90 or 95% homology with a reference first complementarity domain disclosed herein;

the linking domain is 1 to 5 nucleotides in length;

the second complementarity domain is 5 to 27 nucleotides in length and, in an embodiment has at least 50, 60, 70, 80, 85, 90 or 95% homology with a reference second complementarity domain disclosed herein;

5 the proximal domain is 5 to 20 nucleotides in length and, in an embodiment has at least 50, 60, 70, 80, 85, 90 or 95% homology with a reference proximal domain disclosed herein;

and the tail domain is absent or a nucleotide sequence is 1 to 50 nucleotides in length and, in an embodiment, has at least 50, 60, 70, 80, 85, 90 or 95% homology with a reference tail domain disclosed herein.

10

Exemplary Chimeric gRNAs

In an embodiment, a unimolecular, or chimeric, gRNA comprises, preferably from 5' to 3':

a targeting domain (which is complementary to a target nucleic acid);

15 a first complementarity domain, e.g., comprising 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides;

a linking domain;

a second complementarity domain (which is complementary to the first complementarity domain);

20 a proximal domain; and

a tail domain,

wherein,

(a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides;

25 (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; or

(c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

30

In an embodiment, the sequence from (a), (b), or (c), has at least 60, 75, 80, 85, 90, 95, or 99% homology with the corresponding sequence of a naturally occurring gRNA, or with a gRNA described herein.

In an embodiment, the proximal and tail domain, when taken together, comprise at least
5 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is
10 complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

15 In an embodiment, the targeting domain comprises, has, or consists of, 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the
20 targeting domain is 17 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 18 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 19 nucleotides
25 (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 19 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 20 nucleotides (e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 20 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 21 nucleotides (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 21 nucleotides in length.

5 In an embodiment, the targeting domain comprises, has, or consists of, 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 22 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 23 nucleotides in length.

10 In an embodiment, the targeting domain comprises, has, or consists of, 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 24 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 25 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length.

20 In an embodiment, the targeting domain comprises, has, or consists of, 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

25 In an embodiment, the targeting domain comprises, has, or consists of, 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 17 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

5 In an embodiment, the targeting domain comprises, has, or consists of, 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 17 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

10 In an embodiment, the targeting domain comprises, has, or consists of, 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 17 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

15 In an embodiment, the targeting domain comprises, has, or consists of, 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 18 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

20 In an embodiment, the targeting domain comprises, has, or consists of, 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 18 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

25 In an embodiment, the targeting domain comprises, has, or consists of, 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 18 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

30 In an embodiment, the targeting domain comprises, has, or consists of, 19 nucleotides (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 19 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 19 nucleotides (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 19 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

5 In an embodiment, the targeting domain comprises, has, or consists of, 19 nucleotides (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 19 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

10 In an embodiment, the targeting domain comprises, has, or consists of, 20 nucleotides (e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 20 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 20 nucleotides
15 (e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 20 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 20 nucleotides
20 (e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 20 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 21 nucleotides
25 (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 21 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 21 nucleotides
30 (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 21 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 21 nucleotides (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 21 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 22 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 22 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 22 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 23 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 23 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 23 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 24 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

5 In an embodiment, the targeting domain comprises, has, or consists of, 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 24 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

10 In an embodiment, the targeting domain comprises, has, or consists of, 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 24 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

15 In an embodiment, the targeting domain comprises, has, or consists of, 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 25 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

20 In an embodiment, the targeting domain comprises, has, or consists of, 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 25 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

25 In an embodiment, the targeting domain comprises, has, or consists of, 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 25 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

30 In an embodiment, the targeting domain comprises, has, or consists of, 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

5 In an embodiment, the targeting domain comprises, has, or consists of, 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

10 In an embodiment, the unimolecular, or chimeric, gRNA molecule (comprising a targeting domain, a first complementary domain, a linking domain, a second complementary domain, a proximal domain and, optionally, a tail domain) comprises the following sequence in which the targeting domain is depicted as 20 Ns but could be any sequence and range in length from 16 to 26 nucleotides and in which the gRNA sequence is followed by 6 Us, which serve as
 15 a termination signal for the U6 promoter, but which could be either absent or fewer in number:
 NNNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGG
 CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU. In an
 embodiment, the unimolecular, or chimeric, gRNA molecule is a *S. pyogenes* gRNA molecule.

20 In some embodiments, the unimolecular, or chimeric, gRNA molecule (comprising a targeting domain, a first complementary domain, a linking domain, a second complementary domain, a proximal domain and, optionally, a tail domain) comprises the following sequence in which the targeting domain is depicted as 20 Ns but could be any sequence and range in length from 16 to 26 nucleotides and in which the gRNA sequence is followed by 6 Us, which serve as
 a termination signal for the U6 promoter, but which could be either absent or fewer in number:
 25 NNNNNNNNNNNNNNNNNNNNNNGUUUUAGUACUCUGGAAACAGAAUCUACUAAAAC
 AAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUUUU. In an
 embodiment, the unimolecular, or chimeric, gRNA molecule is a *S. aureus* gRNA molecule.

Exemplary Modular gRNAs

30 In an embodiment, a modular gRNA comprises:

a first strand comprising, preferably from 5' to 3';

a targeting domain, e.g., comprising 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides;

a first complementarity domain; and

a second strand, comprising, preferably from 5' to 3';

5 optionally a 5' extension domain;

a second complementarity domain;

a proximal domain; and

a tail domain,

wherein:

10 (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides;

(b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; or

15 (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

20 In an embodiment, the sequence from (a), (b), or (c), has at least 60, 75, 80, 85, 90, 95, or 99% homology with the corresponding sequence of a naturally occurring gRNA, or with a gRNA described herein.

In an embodiment, the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

25 In an embodiment, there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

30 In an embodiment, the targeting domain comprises, has, or consists of, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length.

5 In an embodiment, the targeting domain comprises, has, or consists of, 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 17 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 18 nucleotides in length.

10 In an embodiment, the targeting domain comprises, has, or consists of, 19 nucleotides (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 19 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 20 nucleotides (e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 20 nucleotides in length.

15 In an embodiment, the targeting domain comprises, has, or consists of, 21 nucleotides (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 21 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 22 nucleotides in length.

20 In an embodiment, the targeting domain comprises, has, or consists of, 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 23 nucleotides in length.

25 In an embodiment, the targeting domain comprises, has, or consists of, 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 24 nucleotides in length.

In an embodiment, the targeting domain comprises, has, or consists of, 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 25 nucleotides in length.

30

In an embodiment, the targeting domain comprises, has, or consists of, 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length.

5 In an embodiment, the targeting domain comprises, has, or consists of, 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

10 In an embodiment, the targeting domain comprises, has, or consists of, 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

15 In an embodiment, the targeting domain comprises, has, or consists of, 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

20 In an embodiment, the targeting domain comprises, has, or consists of, 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 17 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 17 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

25 In an embodiment, the targeting domain comprises, has, or consists of, 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 17 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

30 In an embodiment, the targeting domain comprises, has, or consists of, 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the

targeting domain is 18 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the
5 targeting domain is 18 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the
10 targeting domain is 18 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 19 nucleotides (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the
15 targeting domain is 19 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 19 nucleotides (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the
20 targeting domain is 19 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 19 nucleotides (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the
25 targeting domain is 19 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 20 nucleotides (e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the
30 targeting domain is 20 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 20 nucleotides
(e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the

targeting domain is 20 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 20 nucleotides (e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the
5 targeting domain is 20 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 21 nucleotides (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the
10 targeting domain is 21 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 21 nucleotides (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the
15 targeting domain is 21 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 21 nucleotides (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the
20 targeting domain is 21 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the
25 targeting domain is 22 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the
30 targeting domain is 22 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the
30 targeting domain is 22 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41,

46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the
5 targeting domain is 23 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the
10 targeting domain is 23 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the
15 targeting domain is 23 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the
20 targeting domain is 24 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the
25 targeting domain is 24 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the
30 targeting domain is 24 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the

targeting domain is 25 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 25 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 25 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In an embodiment, the targeting domain comprises, has, or consists of, 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In an embodiment, the targeting domain comprises, has, or consists of, 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

II. Methods for Designing gRNAs

Methods for designing gRNAs are described herein, including methods for selecting, designing and validating target domains. Exemplary targeting domains are also provided herein. Targeting Domains discussed herein can be incorporated into the gRNAs described herein.

Methods for selection and validation of target sequences as well as off-target analyses are described, e.g., in Mali *et al.*, SCIENCE 2013, 339(6121): 823-826; Hsu *et al.*, NAT BIOTECHNOL,

published on July 21, 2013; Fu *et al.*, NAT BIOTECHNOL 2014 Jan 26 (doi: 10.1038/nbt.2808. PubMed PMID: 24463574); Heigwer *et al.*, NAT METHODS 2014, 11(2):122-3 (doi: 10.1038/nmeth.2812. PubMed PMID: 24481216); Bae *et al.*, BIOINFORMATICS, 2014 Jan 24 (PubMed PMID: 24463181); Xiao A *et al.*, BIOINFORMATICS, 2014 Jan 21 (PubMed PMID: 24389662).

For example, a software tool can be used to optimize the choice of gRNA within a user's target sequence, e.g., to minimize total off-target activity across the genome. Off target activity may be other than cleavage. For each possible gRNA choice, the tool can identify all off-target sequences (preceding either NAG or NGG PAMs) across the genome that contain up to certain number (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of mismatched base-pairs. The cleavage efficiency at each off-target sequence can be predicted, e.g., using an experimentally-derived weighting scheme. Each possible gRNA is then ranked according to its total predicted off-target cleavage; the top-ranked gRNAs represent those that are likely to have the greatest on-target and the least off-target cleavage. Other functions, e.g., automated reagent design for CRISPR construction, primer design for the on-target Surveyor assay, and primer design for high-throughput detection and quantification of off-target cleavage via next-gen sequencing, can also be included in the tool. Candidate gRNA molecules can be evaluated by art-known methods or as described in Section IV herein.

Guide RNAs (gRNAs) for use with *S. pyogenes*, *S. aureus* and *N. meningitidis* Cas9s were identified using a DNA sequence searching algorithm. Guide RNA design was carried out using a custom guide RNA design software based on the public tool cas-offinder (reference: Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases., Bioinformatics. 2014 Feb 17. Bae S1, Park J, Kim JS. PMID:24463181). Said custom guide RNA design software scores guides after calculating their genomewide off-target propensity. Typically matches ranging from perfect matches to 7 mismatches are considered for guides ranging in length from 17 to 24 . Once the off-target sites are computationally determined , an aggregate score is calculated for each guide and summarized in a tabular output using a web-interface. In addition to identifying potential gRNA sites adjacent to PAM sequences, the software also identifies all PAM adjacent sequences that differ by 1, 2, 3 or more nucleotides from the selected gRNA sites. Genomic DNA sequence for each gene was obtained from the UCSC Genome browser and sequences were screened for repeat elements

using the publically available RepeatMasker program. RepeatMasker searches input DNA sequences for repeated elements and regions of low complexity. The output is a detailed annotation of the repeats present in a given query sequence.

Following identification, gRNAs were ranked into tiers based on their distance to the target site, their orthogonality and presence of a 5' G (based on identification of close matches in the human genome containing a relevant PAM, e.g., in the case of *S. pyogenes*, a NGG PAM, in the case of *S. aureus*, NNGRR (e.g., a NNGRRT or NNGRRV) PAM, and in the case of *N. meningitidis*, a NNNNGATT or NNNNGCTT PAM. Orthogonality refers to the number of sequences in the human genome that contain a minimum number of mismatches to the target sequence. A "high level of orthogonality" or "good orthogonality" may, for example, refer to 20-mer gRNAs that have no identical sequences in the human genome besides the intended target, nor any sequences that contain one or two mismatches in the target sequence. Targeting domains with good orthogonality are selected to minimize off-target DNA cleavage.

As an example, for *S. pyogenes* and *N. meningitidis* targets, 17-mer, or 20-mer gRNAs were designed. As another example, for *S. aureus* targets, 18-mer, 19-mer, 20-mer, 21-mer, 22-mer, 23-mer and 24-mer gRNAs were designed. Targeting domains, disclosed herein, may comprise the 17-mer described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, e.g., the targeting domains of 18 or more nucleotides may comprise the 17-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. Targeting domains, disclosed herein, may comprises the 18-mer described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, e.g., the targeting domains of 19 or more nucleotides may comprise the 18-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. Targeting domains, disclosed herein, may comprises the 19-mer described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, e.g., the targeting domains of 20 or more nucleotides may comprise the 19-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. Targeting domains, disclosed herein, may comprises the 20-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, e.g., the targeting domains of 21 or more nucleotides may comprise the 20-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. Targeting domains, disclosed herein, may comprises the 21-mer described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, e.g., the targeting domains of 22 or more nucleotides may comprise the 21-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. Targeting domains, disclosed herein, may comprises the 22-mer described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, e.g., the targeting domains of 23 or more nucleotides may comprise

the 22-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. Targeting domains, disclosed herein, may comprise the 23-mer described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, e.g., the targeting domains of 24 or more nucleotides may comprise the 23-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**. Targeting domains, disclosed herein, may comprise the 24-mer described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**, e.g., the targeting domains of 25 or more nucleotides may comprise the 24-mer gRNAs described in **Tables 1-3, 4A-4E, 5A-5F, or 6A-6D**.

gRNAs were identified for both single-gRNA nuclease cleavage and for a dual-gRNA paired “nickase” strategy. Criteria for selecting gRNAs and for determining which gRNAs are used in a selected strategy is based on several considerations:

1. gRNA pairs should be oriented on the DNA such that PAMs are facing out and cutting with the D10A Cas9 nickase will result in 5' overhangs.
 2. An assumption that cleaving with dual nickase pairs results in deletion of the entire intervening sequence at a reasonable frequency. However, use of dual nickase pairs also typically results in indel mutations at the site of only one of the gRNAs.
- Candidate pair members can be tested to determine how efficiently they remove the entire sequence versus producing indel mutations at the site of one gRNA.

The targeting domains discussed herein can be incorporated into the gRNAs described herein.

As an example, two strategies were utilized to identify gRNAs for use with *S. pyogenes*, *S. aureus* and *N. meningitidis* Cas9 enzymes.

In one strategy, gRNAs were designed for use with *S. pyogenes* Cas9 enzymes (**Tables 1-3**). While it can be desirable to have gRNAs start with a 5'G, this requirement was relaxed for some gRNAs in tier 1 to identify guides in the correct orientation, within a reasonable distance to the mutation and with a high level of orthogonality. To find a pair of gRNAs for the dual-nickase strategy, the distance from the mutation was extended or the requirement for the 5'G was removed. For selection of tier 2 gRNAs, the distance restriction was relaxed in some cases such that a longer sequence was scanned, but the 5'G was required for all gRNAs. Whether or not the distance requirement was relaxed depended on how many sites were found within the original search window. Tier 3 uses the same distance restriction as tier 2, but removes the requirement for a 5'G. Note that tiers are non-inclusive (each gRNA is listed only once).

As discussed above, gRNAs were identified for single-gRNA nuclease cleavage as well as for a dual-gRNA paired “nickase” strategy, as indicated.

In a second strategy, gRNAs were designed for use with *S. pyogenes*, *S. aureus* and *N. meningitidis* Cas9 enzymes. The gRNAs were identified and ranked into 4 tiers for *S. pyogenes* (Tables 4A-4E). The targeting domain to be used with *S. pyogenes* Cas9 enzymes for tier 1 gRNA molecules were selected based on (1) proximity to the mutation, e.g., within 200bp (e.g., upstream or downstream) of mutation, (2) a high level of orthogonality, and (3) the presence of a 5' G. For selection of tier 2 gRNAs, a reasonable distance and high orthogonality were required but the presence of a 5'G was not required. Tier 3 uses the same distance restriction and the requirement for a 5'G, but removes the requirement of good orthogonality. Tier 4 uses the same distance restriction but removes the requirement of good orthogonality and the 5'G. The gRNAs were identified and ranked into 5 tiers for *S. aureus*, when the relevant PAM was NNGRRT or NNGRRV (Tables 5A-5F). The targeting domain to be used with *S. aureus* Cas9 enzymes for tier 1 gRNA molecules were selected based on (1) proximity to the mutation, e.g., within 200bp (e.g., upstream or downstream) of mutation, (2) a high level of orthogonality, (3) the presence of a 5' G and (4) PAM was NNGRRT. For selection of tier 2 gRNAs, a reasonable distance and high orthogonality were required but the presence of a 5'G was not required, and PAM was NNGRRT. Tier 3 uses the same distance restriction and the requirement for a 5'G, but removes the requirement of good orthogonality, and PAM was NNGRRT. Tier 4 uses the same distance restriction but removes the requirement of good orthogonality and the 5'G, and PAM was NNGRRT. Tier 5 uses the same distance restriction but removes the requirement of good orthogonality and the 5'G, and PAM was NNGRRV. The gRNAs were identified and ranked into 4 tiers for *N. meningitidis* (Tables 6A-6D). The targeting domain to be used with *N. meningitidis* Cas9 enzymes for tier 1 gRNA molecules were selected based on (1) proximity to the mutation, e.g., within 200bp (e.g., upstream or downstream) of mutation, (2) a high level of orthogonality, and (3) the presence of a 5' G. For selection of tier 2 gRNAs, a reasonable distance and high orthogonality were required but the presence of a 5'G was not required. Tier 3 uses the same distance restriction and the requirement for a 5'G, but removes the requirement of good orthogonality. Tier 4 uses the same distance restriction but removes the requirement of good orthogonality and the 5'G. Note that tiers are non-inclusive (each gRNA is listed only once

for the strategy). In certain instances, no gRNA was identified based on the criteria of the particular tier.

In an embodiment, when a single gRNA molecule is used to target a Cas9 nickase to create a single strand break in close proximity to the mutation, eg., the gRNA is used to target
5 either upstream of (e.g., within 200 bp upstream of the mutation), or downstream of (e.g., within 200 bp downstream of the mutation) in the *USH2A* gene.

In an embodiment, when a single gRNA molecule is used to target a Cas9 nuclease to create a double strand break to in closeproximity to the mutation, e.g., the gRNA is used to target
10 either upstream of (e.g., within 200 bp upstream of the mutation), or downstream of (e.g., within 200 bp downstream of the mutation) in the *USH2A* gene.

In an embodiment, dual targeting is used to create two double strand breaks to in closeproximity to the mutation, e.g., the gRNA is used to target either upstream of (e.g., within 200 bp upstream of the mutation), or downstream of (e.g., within 200 bp downstream of the mutation) in the *USH2A* gene. In an embodiment, the first and second gRNAs are used target
15 two Cas9 nucleases to flank, e.g., the first of gRNA is used to target upstream of (e.g., within 200 bp upstream of the mutation), and the second gRNA is used to target downstream of (e.g., within 200 bp downstream of the mutation) in the *USH2A* gene.

In an embodiment, dual targeting is used to create a double strand break and a pair of single strand breaks to delete a genomic sequence including the mutation. In an embodiment, the
20 first, second and third gRNAs are used to target one Cas9 nuclease and two Cas9 nickases to flank, e.g., the first gRNA that will be used with the Cas9 nuclease is used to target upstream of (e.g., within 200 bp upstream of the mutation) or downstream of (e.g., within 200 bp downstream of the mutation), and the second and third gRNAs that will be used with the Cas9 nickase pair are used to target the opposite side of the mutation (e.g., within 200 bp upstream or downstream
25 of the mutation) in the *USH2A* gene.

In an embodiment, when four gRNAs (e.g., two pairs) are used to target four Cas9 nickases to create four single strand breaks to delete genomic sequence including the mutation, the first pair and second pair of gRNAs are used to target four Cas9 nickases to flank, e.g., the first pair of gRNAs are used to target upstream of (e.g., within 200 bp upstream of the mutation),
30 and the second pair of gRNAs are used to target downstream of (e.g., within 200 bp downstream of the mutation) in the *USH2A* gene.

Any of the targeting domains in the tables described herein can be used with a Cas9 nickase molecule to generate a single strand break.

Any of the targeting domains in the tables described herein can be used with a Cas9 nuclease molecule to generate a double strand break.

5 In an embodiment, dual targeting (e.g., dual nicking) is used to create two nicks on opposite DNA strands by using *S. pyogenes*, *S. aureus* and *N. meningitidis* Cas9 nickases with two targeting domains that are complementary to opposite DNA strands, e.g., a gRNA comprising any minus strand targeting domain may be paired any gRNA comprising a plus strand targeting domain provided that the two gRNAs are oriented on the DNA such that PAMs
10 face outward and the distance between the 5' ends of the gRNAs is 0-50 bp. Exemplary nickase pairs including selecting a targeting domain from Group A and a second targeting domain from Group B, or selecting a targeting domain from Group C and a second targeting domain from Group D, in **Table 4E** (for *S. pyogenes*), selecting a targeting domain from Group A and a second targeting domain from Group B in **Table 5F** (for *S. aureus*) or selecting a targeting
15 domain from Group A and a second targeting domain from Group B in **Table 6D** (for *N. meningitidis*). It is contemplated herein that in an embodiment a targeting domain of Group A can be combined with any of the targeting domains of Group B, or a targeting domain of Group C can be combined with any of the targeting domains of Group D in **Table 4E** (for *S. pyogenes*). For example, USH2A-182 can be combined with USH2A-179, USH2A-177 can be combined
20 with USH2A-176, or USH2A-187 can be combined with USH2A-176. It is contemplated herein that in an embodiment a targeting domain of Group A can be combined with any of the targeting domains of Group B in **Table 5F** (for *S. aureus*). For example, USH2A-288 can be combined with USH2A-448. It is contemplated herein that in an embodiment a targeting domain of Group A can be combined with any of the targeting domains of Group B in **Table 6D** (for *N. meningitidis*). For example, USH2A-266 can be combined with USH2A-261 or USH2A-268 can
25 be combined with USH2A-261.

When two gRNAs designed for use to target two Cas9 molecules, one Cas9 can be one species, the second Cas9 can be from a different species. Both Cas9 species are used to generate a single or double-strand break, as desired.

30

Exemplary Targeting Domains

Table 1 provides targeting domains for the 2299delG site selected according to first tier parameters, and are selected based on the presence of a 5' G, close proximity and orientation to mutation and orthogonality in the human genome. In an embodiment, the targeting domain is the exact complement of the target domain. Any of the targeting domains can be used with a Cas9 molecule that gives double stranded cleavage. Any of the targeting domains in the table can be used with single-stranded break nucleases (nickases). In an embodiment, dual targeting is used to create two nicks. In an embodiment, 20-mer dual nickase pairs are used, e.g., USH2A-1 and USH2A-6, or USH2A-2 and USH2A-6 are used. In an embodiment, 17-mer dual nickase pairs are used, e.g., USH2A-15 and USH2A-20, USH2A-15 and USH2A-22, USH2A-16 and USH2A-20, or USH2A-16 and USH2A-22 are used.

Table 1

1st Tier	selected based on the presence of a 5' G (only for USH2A-1, 2, 5, 6, 10, 11), close proximity and orientation to mutation and orthogonality in the human genome			
gRNA Name	DNA Strand	Target Site Sequence (does not include PAM)	Target Site Length	Distance to mutation
USH2A-1	-	GAGUGCAAAAAAGAAGCCAA	20	16bp downstream
USH2A-2	-	GUUAGAUGUCACCAAUUGUA	20	75bp downstream
USH2A-5	+	GGUGUCACACUGAAGUCCUU	20	21bp downstream
USH2A-6	+	GCCAUGGAGGUACACUGGC	20	56bp upstream
USH2A-10	+	GUCACAGGCCUUACAAU	17	75bp downstream
USH2A-11	+	GUCACACUGAAGUCCUU	17	21bp downstream
USH2A-15	-	UGCAAAAAAGAAGCCAA	17	16bp downstream
USH2A-16	-	UGCAGAGAAAACUUUUA	17	52bp downstream
USH2A-20	+	UGUUCACUGAGCCAUGG	17	43bp upstream
USH2A-22	+	AUGGAGGUACACUGGC	17	56bp upstream

Table 2 provides targeting domains for the 2299delG site selected according to Second Tier parameters, as described above, and are selected based on the presence of a 5' G and reasonable proximity to mutation.

Table 2

2nd Tier	Selected based on the presence of a 5' G and reasonable proximity to mutation		
gRNA Name	DNA Strand	Target Site Sequence (does not include PAM)	Target Site Length
USH2A-3	-	GCCUGUGACUGUGACACAGC	20

USH2A-4	-	GACACAGCUGGAUCCCUCCC	20
USH2A-7	-	GCAGAGAAAACUUUUUAU	17
USH2A-8	-	GUCUGUAAUGCUAAGAC	17
USH2A-9	+	GCAUUACAGACAGUCCC	17

Table 3 provides targeting domains for the 2299delG site selected according to Third Tier parameters, as described above, and are selected based on reasonable proximity to mutation.

Table 3

3rd Tier	Selected based on reasonable proximity to mutation		
gRNA Name	DNA Strand	Target Site Sequence (does not include PAM)	Target Site Length
USH2A-12	-	UGCCAGUGUAACCUCCA	17
USH2A-13	-	UUCUGCAAUCCUCACUC	17
USH2A-14	-	UCUGCAAUCCUCACUCU	17
USH2A-17	+	AUAAAAGUUUUCUCUGC	17
USH2A-18	+	UCACACUGCCCAGAGUG	17
USH2A-19	+	AUUUGUUCACUGAGCCA	17
USH2A-21	+	AGCCAUGGAGGUUACAC	17
USH2A-23	+	CUACACUGCCCAGAGUG	17
USH2A-24	-	AAAUUCUGCAAUCCUCACUC	20
USH2A-25	-	AAUUCUGCAAUCCUCACUCU	20
USH2A-26	-	ACACAGCUGGAUCCCUCCC	20
USH2A-27	-	ACCUGCAGAGAAAACUUUUA	20
USH2A-28	-	ACUGUCUGUAAUGCUAAGAC	20
USH2A-29	-	AGGUGUGAUCAUUGCAAUUU	20
USH2A-30	-	AUAUUUUUAUCUUUAGGGCUU	20
USH2A-31	-	CCCUGCCAGUGUAACCUCCA	20
USH2A-32	-	CCUGCAGAGAAAACUUUUUAU	20
USH2A-33	-	CUCCGAAGCUUUAUGAUGU	20
USH2A-34	-	CUGUCUGUAAUGCUAAGACA	20
USH2A-35	+	ACAGUCACAGGCCUUACAAU	20
USH2A-36	+	AGAAUUUGUUCACUGAGCCA	20
USH2A-37	+	AUCCAACAUCAUUAAAGCUU	20
USH2A-38	+	AUUACAGACAGUCCCAGGGA	20
USH2A-39	+	AUUUGUUCACUGAGCCAUGG	20
USH2A-40	+	CACUCACACUGCCCAGAGUG	20
USH2A-41	+	CAUUACAGACAGUCCCAGGG	20
USH2A-42	+	CCAUGGAGGUUACACUGGCA	20
USH2A-43	+	CCCAUAAAAGUUUUCUCUGC	20
USH2A-44	+	CUGAGCCAUGGAGGUUACAC	20
USH2A-45	+	UAGCAUUACAGACAGUCCCA	20

USH2A-46	+	UCCAGCUGUGUCACAGUCAC	20
USH2A-47	+	UUAGCAUUACAGACAGUCCC	20
USH2A-48	-	AAUAUAUUUUUAUCUUUA	17
USH2A-49	-	UUUUUAUCUUUAGGGCUU	17
USH2A-50	-	UGUGAUCAUUGCAAUUU	17
USH2A-51	-	CGAAGCUUUAUGAUGU	17
USH2A-52	-	AGAUGUCACCAAUUGUA	17
USH2A-53	-	UGUGACUGUGACACAGC	17
USH2A-54	-	ACAGCUGGAUCCCUCCC	17
USH2A-55	-	CAGCUGGAUCCCUCCCU	17
USH2A-56	-	UCUGUAAUGCUAAGACA	17
USH2A-57	+	CAUUACAGACAGUCCCA	17
USH2A-58	+	UACAGACAGUCCCAGGG	17
USH2A-59	+	ACAGACAGUCCCAGGGA	17
USH2A-60	+	AGCUGUGUCACAGUCAC	17
USH2A-61	+	UGGAGGUUACACUGGCA	17
USH2A-62	+	CAACAUCAUUAAGCUU	17

Table 4A provides targeting domains for the 2299delG site in the USH2A gene selected according to the first tier parameters. The targeting domains are within 200 bases of the 2299deG site, have good orthogonality, and start with G. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. pyogenes* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 4A

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-230	-	GCAAGCCCAAUGUUGAA	17
USH2A-225	+	GCAUUACAGACAGUCCC	17
USH2A-221	+	GUCACACUGAAGUCCUU	17
USH2A-217	+	GUCACAGGCCUUAACAAU	17
USH2A-226	-	GUCUGUAAUGCUAAGAC	17
USH2A-198	-	GACACAGCUGGAUCCCUCCC	20
USH2A-204	-	GAGACAGUGCAAUAAAUGUU	20
USH2A-184	+	GCACUACACUGCCCAGAGUG	20
USH2A-197	+	GCACUGUCUCCCUUCAACAU	20
USH2A-194	+	GCCAUGGAGGUUACACUGGC	20
USH2A-192	-	GCCUGUGACUGUGACACAGC	20
USH2A-188	+	GGUGUCACACUGAAGUCCUU	20

USH2A-179	-	GUUAGAUGUCACCAAUUGUA	20
-----------	---	----------------------	----

Table 4B provides targeting domains for the 2299delG site in the USH2A gene selected according to the second tier parameters. The targeting domains are within 200 bases of the 2299deG site, have good orthogonality, and do not start with G. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. pyogenes* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 4B

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-244	-	ACAGCUGGAUCCCUC	17
USH2A-237	-	ACAGUGCAAUAAAUGUU	17
USH2A-220	-	AGAUGUCACCAAUUGUA	17
USH2A-231	+	AGCCAUGGAGGUUACAC	17
USH2A-241	+	AUAAAAGUUUUCUCUGC	17
USH2A-219	+	AUGGAGGUUACACUGGC	17
USH2A-247	+	AUUUAAAAGGUGAGGAU	17
USH2A-245	+	AUUUGUUCACUGAGCCA	17
USH2A-242	+	CAACAUCAUUAAGCUU	17
USH2A-228	-	CAGCUGGAUCCCUC	17
USH2A-222	+	CAUUACAGACAGUCCCA	17
USH2A-218	-	CGAAGCUUUAUGAUGU	17
USH2A-235	+	CUACACUGCCCAGAGUG	17
USH2A-234	+	CUGUCUCCCUCAACAU	17
USH2A-232	+	UACAGACAGUCCCAGGG	17
USH2A-229	-	UCUGCAAUCCUCACUCU	17
USH2A-224	-	UCUGUAAUGCUAAGACA	17
USH2A-240	-	UGCAAGCCCAAUGUUGA	17
USH2A-246	-	UGCAGAGAAAACUUUUA	17
USH2A-233	-	UGCCAGUGUAACCUCCA	17
USH2A-227	+	UGGAGGUUACACUGGCA	17
USH2A-223	+	UGUCUCCCUCAACAUU	17
USH2A-238	-	UGUGAUCAUUGCAAUUU	17
USH2A-239	+	UGUUCACUGAGCCAUGG	17
USH2A-236	-	UUCUGCAAUCCUCACUC	17

USH2A-243	-	UUUUAUCUUUAGGGCUU	17
USH2A-178	-	AAAUUCUGCAAUCCUCACUC	20
USH2A-186	-	AAUUCUGCAAUCCUCACUCU	20
USH2A-191	-	ACACAGCUGGAUCCCUCCCU	20
USH2A-175	+	ACAGUCACAGGCCUUACAAU	20
USH2A-206	-	ACAGUGCAAUAAAUGUUUGG	20
USH2A-201	-	ACCUGCAGAGAAAACUUUUA	20
USH2A-196	-	ACUGUCUGUAAUGCUAAGAC	20
USH2A-199	+	AGAAUUUGUUCACUGAGCCA	20
USH2A-185	-	AGGUGUGAUCAUUGCAAUUU	20
USH2A-193	-	AUAUUUUUAUCUUUAGGGCUU	20
USH2A-202	+	AUCCAACAUCAUUAAAGCUU	20
USH2A-176	-	AUCUGCAAGCCCAUGUUGA	20
USH2A-205	+	AUUACAGACAGUCCCAGGGA	20
USH2A-200	+	CACUGUCUCCCUUCAACAUU	20
USH2A-203	-	CAGUGCAAUAAAUGUUUGGA	20
USH2A-177	+	CAUUACAGACAGUCCCAGGG	20
USH2A-180	+	CCAUGGAGGUUACACUGGCA	20
USH2A-182	+	CCCAUAAAAGUUUUCUCUGC	20
USH2A-183	-	CCCUGCCAGUGUAACCUCCA	20
USH2A-174	-	CUCCGAAGCUUUAAUGAUGU	20
USH2A-189	+	CUGAGCCAUGGAGGUUACAC	20
USH2A-181	-	CUGUCUGUAAUGCUAAGACA	20
USH2A-187	+	UAGCAUUACAGACAGUCCCA	20
USH2A-190	-	UCUGCAAGCCCAUGUUGAA	20
USH2A-195	+	UUAGCAUUACAGACAGUCCC	20

Table 4C provides targeting domains for the 2299delG site in the USH2A gene selected according to the third tier parameters. The targeting domains are within 200 bases of the 2299deG site, and start with G. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. pyogenes* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 4C

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-259	+	GAUAAAAUAUUAUUAAA	17
USH2A-249	-	GCAGAGAAAACUUUUUAU	17

USH2A-255	-	GUGCAAUAAAUGUUUGG	17
-----------	---	-------------------	----

Table 4D provides targeting domains for the 2299delG site in the USH2A gene selected according to the fourth tier parameters. The targeting domains are within 200 bases of the 2299deG site. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. pyogenes* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 4D

4th Tier			
gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-258	-	AAAUAUAUUUAUCUUU	17
USH2A-257	+	AAUAUAUUUAAAAGGUG	17
USH2A-253	-	AAUAUAUUUAUCUUUA	17
USH2A-251	+	ACAGACAGUCCCAGGGA	17
USH2A-254	+	AGCUGUGUCACAGUCAC	17
USH2A-252	+	UAUUUAAAAGGUGAGGA	17
USH2A-256	-	UGCAAAAAGAAGCCAA	17
USH2A-248	-	UGCAAUAAAUGUUUGGA	17
USH2A-250	-	UGUGACUGUGACACAGC	17
USH2A-216	+	AAAGAUAAAUAUAUUUAAA	20
USH2A-208	+	AUAUAUUUAAAAGGUGAGGA	20
USH2A-210	+	AUUUGUUCACUGAGCCAUGG	20
USH2A-211	-	CCUGCAGAGAAAACUUUUUAU	20
USH2A-209	+	UAAAUAUAUUUAAAAGGUG	20
USH2A-212	-	UAGUGCAAAAAGAAGCCAA	20
USH2A-207	+	UAUAUUUAAAAGGUGAGGAU	20
USH2A-215	+	UCCAGCUGUGUCACAGUCAC	20
USH2A-213	-	UUAAAUAUAUUUAUCUUUA	20
USH2A-214	-	UUUAAAUAUAUUUAUCUUU	20

Table 4E provides targeting domains for the 2299delG site in the USH2A gene that can be used for dual targeting. Any of the targeting domains in the table can be used with a *S. pyogenes* Cas9 (nickase) molecule to generate a single stranded break.

Exemplary nickase pairs including selecting a targeting domain from Group A and a second targeting domain from Group B, or a targeting domain from Group C and a second

targeting domain from Group D. It is contemplated herein that in an embodiment a targeting domain of Group A can be combined with any of the targeting domains of Group B or a targeting domain of Group C can be combined with any of the targeting domains of Group D. For example, USH2A-182 can be combined with USH2A-179, USH2A-177 can be combined with USH2A-176, or USH2A-187 can be combined with USH2A-176.

Table 4E

Group A	Group B	Group C	Group D
USH2A-182	USH2A-179	USH2A-177	USH2A-176
		USH2A-187	

Table 5A provides targeting domains for the 2299delG site in the USH2A selected according to the first tier parameters. The targeting domains are within 200 bases of the 2299deG site, have good orthogonality, start with G and PAM is NNGRRT. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. aureus* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 5A

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-292	+	GCACUACACUGCCCAGAGU	19
USH2A-298	-	GCCUGUGACUGUGACACAG	19
USH2A-297	-	GGCCUGUGACUGUGACACAG	20
USH2A-284	-	GGUGUGAUCAUUGCAAUU	18
USH2A-448	-	GACACCUGCAGAGAAAACUUUU	22
USH2A-445	+	GCAUUACAGACAGUCCCAGGG	21
USH2A-427	-	GCUUAGGUGUGAUCAUUGCAAUU	23
USH2A-430	+	GCUUCUUUUUUGCACUACACUGCC	24
USH2A-426	-	GGCUUAGGUGUGAUCAUUGCAAUU	24
USH2A-438	-	GUAAGGCCUGUGACUGUGACACAG	24
USH2A-446	-	GUGACACCUGCAGAGAAAACUUUU	24

Table 5B provides targeting domains for the 2299delG site in the USH2A selected according to the second tier parameters. The targeting domains are within 200 bases of the 2299deG site, have good orthogonality and PAM is NNGRRT. It is contemplated herein that in

an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. aureus* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

5 **Table 5B**

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-295	-	ACCUGCAGAGAAAACUUUU	19
USH2A-288	+	ACUGCCCAGAGUGAGGAUUG	20
USH2A-283	-	AGGUGUGAUCAUUGCAAUU	19
USH2A-280	+	AUUACAGACAGUCCCAGGG	19
USH2A-294	-	CACCUGCAGAGAAAACUUUU	20
USH2A-293	+	CACUACACUGCCCAGAGU	18
USH2A-279	+	CAUUACAGACAGUCCCAGGG	20
USH2A-296	-	CCUGCAGAGAAAACUUUU	18
USH2A-299	-	CCUGUGACUGUGACACAG	18
USH2A-277	-	CUCCGAAGCUUUAUGAUG	19
USH2A-289	+	CUGCCCAGAGUGAGGAUUG	19
USH2A-285	+	CUUUUUUGCACUACACUGCC	20
USH2A-282	-	UAGGUGUGAUCAUUGCAAUU	20
USH2A-278	-	UCCGAAGCUUUAUGAUG	18
USH2A-276	-	UCUCCGAAGCUUUAUGAUG	20
USH2A-291	+	UGCACUACACUGCCCAGAGU	20
USH2A-290	+	UGCCCAGAGUGAGGAUUG	18
USH2A-281	+	UUACAGACAGUCCCAGGG	18
USH2A-287	+	UUUUUUGCACUACACUGCC	18
USH2A-286	+	UUUUUUGCACUACACUGCC	19
USH2A-440	-	AAGGCCUGUGACUGUGACACAG	22
USH2A-450	-	AAUUUCUCCGAAGCUUUAUGAUG	24
USH2A-449	-	ACACCUGCAGAGAAAACUUUU	21
USH2A-456	+	ACACUGCCCAGAGUGAGGAUUG	22
USH2A-444	+	AGCAUUACAGACAGUCCCAGGG	22
USH2A-441	-	AGGCCUGUGACUGUGACACAG	21
USH2A-451	-	AUUUCUCCGAAGCUUUAUGAUG	23
USH2A-457	+	CACUGCCCAGAGUGAGGAUUG	21
USH2A-454	+	CUACACUGCCCAGAGUGAGGAUUG	24
USH2A-428	-	CUUAGGUGUGAUCAUUGCAAUU	22
USH2A-431	+	CUUCUUUUUUGCACUACACUGCC	23
USH2A-439	-	UAAGGCCUGUGACUGUGACACAG	23

USH2A-455	+	UACACUGCCCAGAGUGAGGAUUG	23
USH2A-443	+	UAGCAUUACAGACAGUCCCAGGG	23
USH2A-433	+	UCUUUUUUUGCACUACACUGCC	21
USH2A-447	-	UGACACCGUCAGAGAAAACUUUU	23
USH2A-442	+	UUAGCAUUACAGACAGUCCCAGGG	24
USH2A-429	-	UUAGGUGUGAUCAUUGCAAUU	21
USH2A-453	-	UUCUCCGAAGCUUUAUGAUG	21
USH2A-432	+	UUCUUUUUUUGCACUACACUGCC	22
USH2A-437	+	UUGCACUACACUGCCCAGAGU	21
USH2A-452	-	UUUCUCCGAAGCUUUAUGAUG	22
USH2A-436	+	UUUGCACUACACUGCCCAGAGU	22
USH2A-435	+	UUUUGCACUACACUGCCCAGAGU	23
USH2A-434	+	UUUUUGCACUACACUGCCCAGAGU	24

Table 5C provides targeting domains for the 2299delG site in the USH2A selected according to the third tier parameters. The targeting domains are within 200 bases of the 2299deG site, start with 5' G and PAM is NNGRRT. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. aureus* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 5C

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-461	+	GAUAAAAUAUAUUUAAAAGGU	21

Table 5D provides targeting domains for the 2299delG site in the USH2A selected according to the fourth tier parameters. The targeting domains are within 200 bases of the 2299deG site and PAM is NNGRRT. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. aureus* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 5D

4th Tier			
gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-300	+	AUAAAAUAUAUUUAAAAGGU	20

USH2A-301	+	UAAAAUAUAUUUAAAAGGU	19
USH2A-302	+	AAAAUAUAUUUAAAAGGU	18
USH2A-458	+	AAAGAUAAAAUAUAUUUAAAAGGU	24
USH2A-459	+	AAGAUAAAAUAUAUUUAAAAGGU	23
USH2A-460	+	AGAUAAAAUAUAUUUAAAAGGU	22

Table 5E provides targeting domains for the 2299delG site in the USH2A selected according to the fifth tier parameters. The targeting domains are within 200 bases of the 2299delG site and PAM is NNGRRV. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *S. aureus* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 5E

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-303	+	AUAUAUUUAAAAGGUGAGGA	20
USH2A-304	+	UAUAUUUAAAAGGUGAGGA	19
USH2A-305	+	AUAUUUAAAAGGUGAGGA	18
USH2A-306	+	UUUUCUCUGCAGGUGUCACA	20
USH2A-307	+	UUUCUCUGCAGGUGUCACA	19
USH2A-308	+	UUCUCUGCAGGUGUCACA	18
USH2A-309	+	UGCACUGUCUCCCUUCAACA	20
USH2A-310	+	GCACUGUCUCCCUUCAACA	19
USH2A-311	+	CACUGUCUCCCUUCAACA	18
USH2A-312	-	GUGCAUCUGCAAGCCCAAUG	20
USH2A-313	-	UGCAUCUGCAAGCCCAAUG	19
USH2A-314	-	GCAUCUGCAAGCCCAAUG	18
USH2A-315	-	CAAAUUCUGCAAUCCUCACU	20
USH2A-316	-	AAAUUCUGCAAUCCUCACU	19
USH2A-317	-	AAUUCUGCAAUCCUCACU	18
USH2A-318	-	UCUGCAAGCCCAAUGUUGAA	20
USH2A-319	-	CUGCAAGCCCAAUGUUGAA	19
USH2A-320	-	UGCAAGCCCAAUGUUGAA	18
USH2A-321	+	UUAGCAUUACAGACAGUCCC	20
USH2A-322	+	UAGCAUUACAGACAGUCCC	19
USH2A-323	+	AGCAUUACAGACAGUCCC	18
USH2A-324	-	GACAGUGCAAUAAAUGUUUG	20
USH2A-325	-	ACAGUGCAAUAAAUGUUUG	19

USH2A-326	-	CAGUGCAAUAAAUGUUUG	18
USH2A-327	-	UGACACAGCUGGAUCCCUCC	20
USH2A-328	-	GACACAGCUGGAUCCCUCC	19
USH2A-329	-	ACACAGCUGGAUCCCUCC	18
USH2A-330	+	CUUAGCAUUACAGACAGUCC	20
USH2A-331	+	UUAGCAUUACAGACAGUCC	19
USH2A-332	+	UAGCAUUACAGACAGUCC	18
USH2A-333	-	AAUUUUGGAUUUAAAUUUCU	20
USH2A-334	-	AUUUUGGAUUUAAAUUUCU	19
USH2A-335	-	UUUUGGAUUUAAAUUUCU	18
USH2A-336	+	UUUGCACUACACUGCCCAGA	20
USH2A-337	+	UUGCACUACACUGCCCAGA	19
USH2A-338	+	UGCACUACACUGCCCAGA	18
USH2A-339	-	GGAGACAGUGCAAUAAAUGU	20
USH2A-340	-	GAGACAGUGCAAUAAAUGU	19
USH2A-341	-	AGACAGUGCAAUAAAUGU	18
USH2A-342	+	AAUGAUUUCAUUCAAGAUAG	20
USH2A-343	+	AUGAUUUCAUUCAAGAUAG	19
USH2A-344	+	UGAUUUCAUUCAAGAUAG	18
USH2A-345	-	ACAGUGCAAUAAAUGUUUGG	20
USH2A-346	-	CAGUGCAAUAAAUGUUUGG	19
USH2A-347	-	AGUGCAAUAAAUGUUUGG	18
USH2A-348	+	AGAUAAAUAUAUUUAAAAG	20
USH2A-349	+	GAUAAAUAUAUUUAAAAG	19
USH2A-350	+	AUAAAUAUAUUUAAAAG	18
USH2A-351	+	UAUAUUUAAAAGGUGAGGAU	20
USH2A-352	+	AUAUUUAAAAGGUGAGGAU	19
USH2A-353	+	UAUUUAAAAGGUGAGGAU	18
USH2A-354	-	CUGGGCAGUGUAGUGCAAAA	20
USH2A-355	-	UGGGCAGUGUAGUGCAAAA	19
USH2A-356	-	GGGCAGUGUAGUGCAAAA	18
USH2A-357	+	CCAACAUCAUUAAGCUUCG	20
USH2A-358	+	CAACAUCAUUAAGCUUCG	19
USH2A-359	+	AACAUCAUUAAGCUUCG	18
USH2A-360	-	UUGUGUCUCGUCUAUCUUGA	20
USH2A-361	-	UGUGUCUCGUCUAUCUUGA	19
USH2A-362	-	GUGUCUCGUCUAUCUUGA	18
USH2A-363	+	UAGCAUUACAGACAGUCCCA	20
USH2A-364	+	AGCAUUACAGACAGUCCCA	19
USH2A-365	+	GCAUUACAGACAGUCCCA	18

USH2A-366	-	AUCUGCAAGCCCAAUGUUGA	20
USH2A-367	-	UCUGCAAGCCCAAUGUUGA	19
USH2A-368	-	CUGCAAGCCCAAUGUUGA	18
USH2A-369	-	UUUUAAAUAUAUUUUAUCUU	20
USH2A-370	-	UUUAAAUAUAUUUUAUCUU	19
USH2A-371	-	UUAAAUAUAUUUUAUCUU	18
USH2A-372	+	CAUCCAACAUCAUUAAGCU	20
USH2A-373	+	AUCCAACAUCAUUAAGCU	19
USH2A-374	+	UCCAACAUCAUUAAGCU	18
USH2A-375	+	GCAUUACAGACAGUCCCAGG	20
USH2A-376	+	CAUUACAGACAGUCCCAGG	19
USH2A-377	+	AUUACAGACAGUCCCAGG	18
USH2A-378	+	CAGAAUUUGUUCACUGAGCC	20
USH2A-379	+	AGAAUUUGUUCACUGAGCC	19
USH2A-380	+	GAAUUUGUUCACUGAGCC	18
USH2A-381	-	ACUUCAGUGUGACACCUGCA	20
USH2A-382	-	CUUCAGUGUGACACCUGCA	19
USH2A-383	-	UUCAGUGUGACACCUGCA	18
USH2A-384	-	CAGUGCAAUAAAUGUUUGGA	20
USH2A-385	-	AGUGCAAUAAAUGUUUGGA	19
USH2A-386	-	GUGCAAUAAAUGUUUGGA	18
USH2A-387	-	GAGACAGUGCAAUAAAUGUU	20
USH2A-388	-	AGACAGUGCAAUAAAUGUU	19
USH2A-389	-	GACAGUGCAAUAAAUGUU	18
USH2A-390	-	AAGCUUUAAUGAUGUUGGAU	20
USH2A-391	-	AGCUUUAAUGAUGUUGGAU	19
USH2A-392	-	GCUUUAAUGAUGUUGGAU	18
USH2A-393	+	AAUAUAUUUAAAAGGUGAGG	20
USH2A-394	+	AUAUAUUUAAAAGGUGAGG	19
USH2A-395	+	UAUAUUUAAAAGGUGAGG	18
USH2A-396	-	GGACUUCAGUGUGACACCUG	20
USH2A-397	-	GACUUCAGUGUGACACCUG	19
USH2A-398	-	ACUUCAGUGUGACACCUG	18
USH2A-399	+	AUCCAACAUCAUUAAGCUU	20
USH2A-400	+	UCCAACAUCAUUAAGCUU	19
USH2A-401	+	CCAACAUCAUUAAGCUU	18
USH2A-402	-	AGUGUAACCUCCAUGGCUCA	20
USH2A-403	-	GUGUAACCUCCAUGGCUCA	19
USH2A-404	-	UGUAACCUCCAUGGCUCA	18
USH2A-405	+	AGAAUUUGUUCACUGAGCCA	20

USH2A-406	+	GAAUUUGUUCACUGAGCCA	19
USH2A-407	+	AAUUUGUUCACUGAGCCA	18
USH2A-408	-	GUAGUGCAAAAAAGAAGCCA	20
USH2A-409	-	UAGUGCAAAAAAGAAGCCA	19
USH2A-410	-	AGUGCAAAAAAGAAGCCA	18
USH2A-411	-	CAUCUGCAAGCCCAUGUUG	20
USH2A-412	-	AUCUGCAAGCCCAUGUUG	19
USH2A-413	-	UCUGCAAGCCCAUGUUG	18
USH2A-414	-	GACUGUCUGUAAUGCUAAGA	20
USH2A-415	-	ACUGUCUGUAAUGCUAAGA	19
USH2A-416	-	CUGUCUGUAAUGCUAAGA	18
USH2A-417	-	GACACAGCUGGAUCCCUCCC	20
USH2A-418	-	ACACAGCUGGAUCCCUCCC	19
USH2A-419	-	CACAGCUGGAUCCCUCCC	18
USH2A-420	+	AGGAUUGCAGAAUUUGUUCA	20
USH2A-421	+	GGAUUGCAGAAUUUGUUCA	19
USH2A-422	+	GAUUGCAGAAUUUGUUCA	18
USH2A-423	+	AGCCAUGGAGGUUACACUGG	20
USH2A-424	+	GCCAUGGAGGUUACACUGG	19
USH2A-425	+	CCAUGGAGGUUACACUGG	18
USH2A-462	+	CUCACAUCCAACAUCAUUAAGCU	24
USH2A-463	+	UCACAUCCAACAUCAUUAAGCU	23
USH2A-464	+	CACAUCCAACAUCAUUAAGCU	22
USH2A-465	+	ACAUCCAACAUCAUUAAGCU	21
USH2A-466	+	AUUGCAGAAUUUGUUCACUGAGCC	24
USH2A-467	+	UUGCAGAAUUUGUUCACUGAGCC	23
USH2A-468	+	UGCAGAAUUUGUUCACUGAGCC	22
USH2A-469	+	GCAGAAUUUGUUCACUGAGCC	21
USH2A-470	-	CUGGGACUGUCUGUAAUGCUAAGA	24
USH2A-471	-	UGGGACUGUCUGUAAUGCUAAGA	23
USH2A-472	-	GGGACUGUCUGUAAUGCUAAGA	22
USH2A-473	-	GGACUGUCUGUAAUGCUAAGA	21
USH2A-474	+	UUGCAGAAUUUGUUCACUGAGCCA	24
USH2A-475	+	UGCAGAAUUUGUUCACUGAGCCA	23
USH2A-476	+	GCAGAAUUUGUUCACUGAGCCA	22
USH2A-477	+	CAGAAUUUGUUCACUGAGCCA	21
USH2A-478	-	GAAGGGAGACAGUGCAAUAAAUGU	24
USH2A-479	-	AAGGGAGACAGUGCAAUAAAUGU	23
USH2A-480	-	AGGGAGACAGUGCAAUAAAUGU	22
USH2A-481	-	GGGAGACAGUGCAAUAAAUGU	21

USH2A-482	+	AAAGUUUUCUCUGCAGGUGUCACA	24
USH2A-483	+	AAGUUUUCUCUGCAGGUGUCACA	23
USH2A-484	+	AGUUUUCUCUGCAGGUGUCACA	22
USH2A-485	+	GUUUUUCUCUGCAGGUGUCACA	21
USH2A-486	+	CUUAGCAUACAGACAGUCCCAGG	24
USH2A-487	+	UUAGCAUACAGACAGUCCCAGG	23
USH2A-488	+	UAGCAUACAGACAGUCCCAGG	22
USH2A-489	+	AGCAUACAGACAGUCCCAGG	21
USH2A-490	+	CUAAAGAUAAAAUAUUAUUAAG	24
USH2A-491	+	UAAAGAUAAAAUAUUAUUAAG	23
USH2A-492	+	AAAGAUAAAAUAUUAUUAAG	22
USH2A-493	+	AAGAUAAAAUAUUAUUAAG	21
USH2A-494	-	UGCAUCUGCAAGCCCAUGUUGAA	24
USH2A-495	-	GCAUCUGCAAGCCCAUGUUGAA	23
USH2A-496	-	CAUCUGCAAGCCCAUGUUGAA	22
USH2A-497	-	AUCUGCAAGCCCAUGUUGAA	21
USH2A-498	+	ACUGAGCCAUGGAGGUACACUGG	24
USH2A-499	+	CUGAGCCAUGGAGGUACACUGG	23
USH2A-500	+	UGAGCCAUGGAGGUACACUGG	22
USH2A-501	+	GAGCCAUGGAGGUACACUGG	21
USH2A-502	-	AAGGACUUCAGUGUGACACCUGCA	24
USH2A-503	-	AGGACUUCAGUGUGACACCUGCA	23
USH2A-504	-	GGACUUCAGUGUGACACCUGCA	22
USH2A-505	-	GACUUCAGUGUGACACCUGCA	21
USH2A-506	+	AGUGAGGAUUGCAGAAUUUGUUCA	24
USH2A-507	+	GUGAGGAUUGCAGAAUUUGUUCA	23
USH2A-508	+	UGAGGAUUGCAGAAUUUGUUCA	22
USH2A-509	+	GAGGAUUGCAGAAUUUGUUCA	21
USH2A-510	+	UAAAAUAUUAUUAAGGUGAGGA	24
USH2A-511	+	AAAAUAUUAUUAAGGUGAGGA	23
USH2A-512	+	AAAUUAUUAUUAAGGUGAGGA	22
USH2A-513	+	AAUAUAUUAUUAAGGUGAGGA	21
USH2A-514	-	CUGUGACACAGCUGGAUCCCUCCC	24
USH2A-515	-	UGUGACACAGCUGGAUCCCUCCC	23
USH2A-516	-	GUGACACAGCUGGAUCCCUCCC	22
USH2A-517	-	UGACACAGCUGGAUCCCUCCC	21
USH2A-518	+	CUGUCUAGCAUACAGACAGUCC	24
USH2A-519	+	UGUCUAGCAUACAGACAGUCC	23
USH2A-520	+	GUCUAGCAUACAGACAGUCC	22
USH2A-521	+	UCUAGCAUACAGACAGUCC	21

USH2A-522	-	UGAACAAAUUCUGCAAUCCUCACU	24
USH2A-523	-	GAACAAAUUCUGCAAUCCUCACU	23
USH2A-524	-	AACAAAUUCUGCAAUCCUCACU	22
USH2A-525	-	ACAAAUUCUGCAAUCCUCACU	21
USH2A-526	-	CAAAGGACUUCAGUGUGACACCUG	24
USH2A-527	-	AAAGGACUUCAGUGUGACACCUG	23
USH2A-528	-	AAGGACUUCAGUGUGACACCUG	22
USH2A-529	-	AGGACUUCAGUGUGACACCUG	21
USH2A-530	-	CACCUUUUAAAUUAUUAUUUAUCUU	24
USH2A-531	-	ACCUUUUAAAUUAUUAUUUAUCUU	23
USH2A-532	-	CCUUUUAAAUUAUUAUUUAUCUU	22
USH2A-533	-	CUUUUAAAUUAUUAUUUAUCUU	21
USH2A-534	-	GUGCAUCUGCAAGCCCAAUGUUGA	24
USH2A-535	-	UGCAUCUGCAAGCCCAAUGUUGA	23
USH2A-536	-	GCAUCUGCAAGCCCAAUGUUGA	22
USH2A-537	-	CAUCUGCAAGCCCAAUGUUGA	21
USH2A-538	+	GUCUUAGCAUUACAGACAGUCCCA	24
USH2A-539	+	UCUUAGCAUUACAGACAGUCCCA	23
USH2A-540	+	CUUAGCAUUACAGACAGUCCCA	22
USH2A-541	+	UUAGCAUUACAGACAGUCCCA	21
USH2A-542	-	AGUGCAUCUGCAAGCCCAAUGUUG	24
USH2A-543	-	GUGCAUCUGCAAGCCCAAUGUUG	23
USH2A-544	-	UGCAUCUGCAAGCCCAAUGUUG	22
USH2A-545	-	GCAUCUGCAAGCCCAAUGUUG	21
USH2A-546	-	CACUCUGGGCAGUGUAGUGCAAAA	24
USH2A-547	-	ACUCUGGGCAGUGUAGUGCAAAA	23
USH2A-548	-	CUCUGGGCAGUGUAGUGCAAAA	22
USH2A-549	-	UCUGGGCAGUGUAGUGCAAAA	21
USH2A-550	+	UGUCUUAGCAUUACAGACAGUCCC	24
USH2A-551	+	GUCUUAGCAUUACAGACAGUCCC	23
USH2A-552	+	UCUUAGCAUUACAGACAGUCCC	22
USH2A-553	+	CUUAGCAUUACAGACAGUCCC	21
USH2A-554	+	CUUUUUUGCACUACACUGCCCAGA	24
USH2A-555	+	UUUUUUUGCACUACACUGCCCAGA	23
USH2A-556	+	UUUUUGCACUACACUGCCCAGA	22
USH2A-557	+	UUUUGCACUACACUGCCCAGA	21
USH2A-558	-	CAGUGUAGUGCAAAAAAGAAGCCA	24
USH2A-559	-	AGUGUAGUGCAAAAAAGAAGCCA	23
USH2A-560	-	GUGUAGUGCAAAAAAGAAGCCA	22
USH2A-561	-	UGUAGUGCAAAAAAGAAGCCA	21

USH2A-562	+	AAAAUAUAUUUAAAAGGUGAGGAU	24
USH2A-563	+	AAAUUAUAUUUAAAAGGUGAGGAU	23
USH2A-564	+	AAUAUAUAUUUAAAAGGUGAGGAU	22
USH2A-565	+	AUAUAUAUUUAAAAGGUGAGGAU	21
USH2A-566	-	ACUGUGACACAGCUGGAUCCCUCC	24
USH2A-567	-	CUGUGACACAGCUGGAUCCCUCC	23
USH2A-568	-	UGUGACACAGCUGGAUCCCUCC	22
USH2A-569	-	GUGACACAGCUGGAUCCCUCC	21
USH2A-570	-	UGCCAGUGUAACCUCCAUGGCUCA	24
USH2A-571	-	GCCAGUGUAACCUCCAUGGCUCA	23
USH2A-572	-	CCAGUGUAACCUCCAUGGCUCA	22
USH2A-573	-	CAGUGUAACCUCCAUGGCUCA	21
USH2A-574	-	UUGCAAUUUUGGAUUUAAAUUUCU	24
USH2A-575	-	UGCAAUUUUGGAUUUAAAUUUCU	23
USH2A-576	-	GCAAUUUUGGAUUUAAAUUUCU	22
USH2A-577	-	CAAUUUUGGAUUUAAAUUUCU	21
USH2A-578	-	GAGACAGUGCAAUAAAUGUUUGGA	24
USH2A-579	-	AGACAGUGCAAUAAAUGUUUGGA	23
USH2A-580	-	GACAGUGCAAUAAAUGUUUGGA	22
USH2A-581	-	ACAGUGCAAUAAAUGUUUGGA	21
USH2A-582	+	GGAAAAUGAUUUCAUUAAGAUAG	24
USH2A-583	+	GAAAAUGAUUUCAUUAAGAUAG	23
USH2A-584	+	AAAAUGAUUUCAUUAAGAUAG	22
USH2A-585	+	AAAUGAUUUCAUUAAGAUAG	21
USH2A-586	+	AUAAAAUAUAUUUAAAAGGUGAGG	24
USH2A-587	+	UAAAAUAUAUUUAAAAGGUGAGG	23
USH2A-588	+	AAAAUAUAUUUAAAAGGUGAGG	22
USH2A-589	+	AAAUUAUAUUUAAAAGGUGAGG	21
USH2A-590	-	AAGGGAGACAGUGCAAUAAAUGUU	24
USH2A-591	-	AGGGAGACAGUGCAAUAAAUGUU	23
USH2A-592	-	GGGAGACAGUGCAAUAAAUGUU	22
USH2A-593	-	GGAGACAGUGCAAUAAAUGUU	21
USH2A-594	+	UCACAUCCAACAUCAUUAAAGCUU	24
USH2A-595	+	CACAUCCAACAUCAUUAAAGCUU	23
USH2A-596	+	ACAUCCAACAUCAUUAAAGCUU	22
USH2A-597	+	CAUCCAACAUCAUUAAAGCUU	21
USH2A-598	-	GGAGACAGUGCAAUAAAUGUUUGG	24
USH2A-599	-	GAGACAGUGCAAUAAAUGUUUGG	23
USH2A-600	-	AGACAGUGCAAUAAAUGUUUGG	22
USH2A-601	-	GACAGUGCAAUAAAUGUUUGG	21

USH2A-602	+	UUAUUGCACUGUCUCCCUUCAACA	24
USH2A-603	+	UAUUGCACUGUCUCCCUUCAACA	23
USH2A-604	+	AUUGCACUGUCUCCCUUCAACA	22
USH2A-605	+	UUGCACUGUCUCCCUUCAACA	21
USH2A-606	+	ACAUCCAACAUCAUUAAGCUUCG	24
USH2A-607	+	CAUCCAACAUCAUUAAGCUUCG	23
USH2A-608	+	AUCCAACAUCAUUAAGCUUCG	22
USH2A-609	+	UCCAACAUCAUUAAGCUUCG	21
USH2A-610	-	UCCGAAGCUUUAUGAUGUUGGAU	24
USH2A-611	-	CCGAAGCUUUAUGAUGUUGGAU	23
USH2A-612	-	CGAAGCUUUAUGAUGUUGGAU	22
USH2A-613	-	GAAGCUUUAUGAUGUUGGAU	21
USH2A-614	-	GGCAGUGCAUCUGCAAGCCCAAUG	24
USH2A-615	-	GCAGUGCAUCUGCAAGCCCAAUG	23
USH2A-616	-	CAGUGCAUCUGCAAGCCCAAUG	22
USH2A-617	-	AGUGCAUCUGCAAGCCCAAUG	21
USH2A-618	-	GGGAGACAGUGCAAUAAAUGUUUG	24
USH2A-619	-	GGAGACAGUGCAAUAAAUGUUUG	23
USH2A-620	-	GAGACAGUGCAAUAAAUGUUUG	22
USH2A-621	-	AGACAGUGCAAUAAAUGUUUG	21

Table 5F provides targeting domains for the 2299delG site in the USH2A gene that can be used for dual targeting. Any of the targeting domains in the table can be used with a *S. aureus* Cas9 (nickase) molecule to generate a single stranded break.

- 5 Exemplary nickase pairs including selecting a targeting domain from Group A and a second targeting domain from Group B. It is contemplated herein that in an embodiment a targeting domain of Group A can be combined with any of the targeting domains of Group B. For example, USH2A-288 can be combined with USH2A-448.

Table 5F

Group A	Group B
USH2A-288	USH2A-448

10

Table 6A provides targeting domains for the 2299delG site in the USH2A selected according to the first tier parameters. The targeting domains are within 200 bases of the 2299deG site, have good orthogonality, and start with G. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any

of the targeting domains in the table can be used with a *N. meningitidis* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 6A

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-264	+	GUGUCACACUGAAGUCC	17
USH2A-261	-	GGUGUGAUCAUUGCAAU	17
USH2A-270	+	GGGCUCACAUCCAACAUCAU	20

5 **Table 6B** provides targeting domains for the 2299delG site in the USH2A selected according to the second tier parameters. The targeting domains are within 200 bases of the 2299deG site and have good orthogonality. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *N. meningitidis* Cas9 molecule that
10 generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 6B

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-263	+	CACUACACUGCCCAGAG	17
USH2A-266	+	AAAAGGUGAGGAUGGGA	17
USH2A-260	+	CUCACAUCCAACAUCAU	17
USH2A-262	+	ACUGUCUCCCUUCAACA	17
USH2A-273	+	CAGGUGUCACACUGAAGUCC	20
USH2A-268	-	UUAGGUGUGAUCAUUGCAAU	20
USH2A-269	+	UUGCACUACACUGCCCAGAG	20
USH2A-271	+	UGCACUGUCUCCCUUCAACA	20
USH2A-274	+	UUUAAAAGGUGAGGAUGGGA	20

15 **Table 6C** provides targeting domains for the 2299delG site in the USH2A selected according to the fourth tier parameters. The targeting domains are within 200 bases of the 2299deG site. It is contemplated herein that in an embodiment the targeting domain hybridizes to the target domain through complementary base pairing. Any of the targeting domains in the table can be used with a *N. meningitidis* Cas9 molecule that generates a double stranded break (Cas9 nuclease) or a single-stranded break (Cas9 nickase).

Table 6C

gRNA Name	DNA Strand	Targeting Domain	Target Site Length
USH2A-267	-	UAAAUUAUUUUUAUCUU	17
USH2A-265	-	UUGGAUUUAAAUUUCUC	17
USH2A-272	-	UUUUAAAUAUAUUUUAUCUU	20
USH2A-275	-	AUUUUGGAUUUAAAUUUCUC	20

Table 6D provides targeting domains for the 2299delG site in the USH2A gene that can be used for dual targeting. Any of the targeting domains in the table can be used with a *N. meningitidis* Cas9 (nickase) molecule to generate a single stranded break.

- 5 Exemplary nickase pairs including selecting a targeting domain from Group A and a second targeting domain from Group B. It is contemplated herein that in an embodiment a targeting domain of Group A can be combined with any of the targeting domains of Group B. For example, USH2A-266 can be combined with USH2A-261 or USH2A-268 can be combined with USH2A-261.

10 **Table 6D**

Group A	Group B
USH2A-266	USH2A-261
USH2A-268	

III. Cas9 Molecules

- Cas9 molecules of a variety of species can be used in the methods and compositions described herein. While the *S. pyogenes*, *S. aureus*, and *S. thermophilus* Cas9 molecules are the subject of much of the disclosure herein, Cas9 molecules of, derived from, or based on the Cas9 proteins of other species listed herein can be used as well. In other words, while much of the description herein uses *S. pyogenes* and *S. thermophilus* Cas9 molecules Cas9 molecules from the other species can replace them. Such species include: *Acidovorax avenae*, *Actinobacillus pleuropneumoniae*, *Actinobacillus succinogenes*, *Actinobacillus suis*, *Actinomyces* sp.,
- 15 *Cycliphilus denitrificans*, *Aminomonas paucivorans*, *Bacillus cereus*, *Bacillus smithii*, *Bacillus thuringiensis*, *Bacteroides* sp., *Blastopirellula marina*, *Bradyrhizobium* sp., *Brevibacillus laterosporus*, *Campylobacter coli*, *Campylobacter jejuni*, *Campylobacter lari*, *Candidatus puniceispirillum*, *Clostridium cellulolyticum*, *Clostridium perfringens*, *Corynebacterium accolens*, *Corynebacterium diphtheria*, *Corynebacterium matruchotii*, *Dinoroseobacter shibae*,
- 20

Eubacterium dolichum, *gamma proteobacterium*, *Gluconacetobacter diazotrophicus*,
Haemophilus parainfluenzae, *Haemophilus sputorum*, *Helicobacter canadensis*, *Helicobacter*
cinaedi, *Helicobacter mustelae*, *Ilyobacter polytropus*, *Kingella kingae*, *Lactobacillus crispatus*,
Listeria ivanovii, *Listeria monocytogenes*, *Listeriaceae bacterium*, *Methylocystis* sp.,
5 *Methylosinus trichosporium*, *Mobiluncus mulieris*, *Neisseria bacilliformis*, *Neisseria cinerea*,
Neisseria flavescens, *Neisseria lactamica*, *Neisseria meningitidis*, *Neisseria* sp., *Neisseria*
wadsworthii, *Nitrosomonas* sp., *Parvibaculum lavamentivorans*, *Pasteurella multocida*,
Phascolarctobacterium succinatutens, *Ralstonia syzygii*, *Rhodopseudomonas palustris*,
Rhodovulum sp., *Simonsiella muelleri*, *Sphingomonas* sp., *Sporolactobacillus vineae*,
10 *Staphylococcus aureus*, *Staphylococcus lugdunensis*, *Streptococcus* sp., *Subdoligranulum* sp.,
Tistrella mobilis, *Treponema* sp., or *Verminephrobacter eiseniae*.

A Cas9 molecule, or Cas9 polypeptide, as that term is used herein, refers to a molecule or polypeptide that can interact with a guide RNA (gRNA) molecule and, in concert with the gRNA molecule, home or localizes to a site which comprises a target domain and PAM sequence. Cas9
15 molecule and Cas9 polypeptide, as those terms are used herein, refer to naturally occurring Cas9 molecules and to engineered, altered, or modified Cas9 molecules or Cas9 polypeptides that differ, e.g., by at least one amino acid residue, from a reference sequence, e.g., the most similar naturally occurring Cas9 molecule or a sequence of **Table 7**.

20 Cas9 Domains

Crystal structures have been determined for two different naturally occurring bacterial Cas9 molecules (Jinek et al., Science, 343(6176):1247997, 2014) and for *S. pyogenes* Cas9 with a guide RNA (e.g., a synthetic fusion of crRNA and tracrRNA) (Nishimasu et al., Cell, 156:935-949, 2014; and Anders et al., Nature, 2014, doi: 10.1038/nature13579).

25 A naturally occurring Cas9 molecule comprises two lobes: a recognition (REC) lobe and a nuclease (NUC) lobe; each of which further comprises domains described herein. **Figs. 9A-9B** provide a schematic of the organization of important Cas9 domains in the primary structure. The domain nomenclature and the numbering of the amino acid residues encompassed by each domain used throughout this disclosure is as described in Nishimasu et al. The numbering of the
30 amino acid residues is with reference to Cas9 from *S. pyogenes*.

The REC lobe comprises the arginine-rich bridge helix (BH), the REC1 domain, and the REC2 domain. The REC lobe does not share structural similarity with other known proteins, indicating that it is a Cas9-specific functional domain. The BH domain is a long α helix and arginine rich region and comprises amino acids 60-93 of the sequence of *S. pyogenes* Cas9. The REC1 domain is important for recognition of the repeat:anti-repeat duplex, e.g., of a gRNA or a tracrRNA, and is therefore critical for Cas9 activity by recognizing the target sequence. The REC1 domain comprises two REC1 motifs at amino acids 94 to 179 and 308 to 717 of the sequence of *S. pyogenes* Cas9. These two REC1 domains, though separated by the REC2 domain in the linear primary structure, assemble in the tertiary structure to form the REC1 domain. The REC2 domain, or parts thereof, may also play a role in the recognition of the repeat:anti-repeat duplex. The REC2 domain comprises amino acids 180-307 of the sequence of *S. pyogenes* Cas9.

The NUC lobe comprises the RuvC domain (also referred to herein as RuvC-like domain), the HNH domain (also referred to herein as HNH-like domain), and the PAM-interacting (PI) domain. The RuvC domain shares structural similarity to retroviral integrase superfamily members and cleaves a single strand, e.g., the non-complementary strand of the target nucleic acid molecule. The RuvC domain is assembled from the three split RuvC motifs (RuvC I, RuvCII, and RuvCIII, which are often commonly referred to in the art as RuvCI domain, or N-terminal RuvC domain, RuvCII domain, and RuvCIII domain) at amino acids 1-59, 718-769, and 909-1098, respectively, of the sequence of *S. pyogenes* Cas9. Similar to the REC1 domain, the three RuvC motifs are linearly separated by other domains in the primary structure, however in the tertiary structure, the three RuvC motifs assemble and form the RuvC domain. The HNH domain shares structural similarity with HNH endonucleases, and cleaves a single strand, e.g., the complementary strand of the target nucleic acid molecule. The HNH domain lies between the RuvC II-III motifs and comprises amino acids 775-908 of the sequence of *S. pyogenes* Cas9. The PI domain interacts with the PAM of the target nucleic acid molecule, and comprises amino acids 1099-1368 of the sequence of *S. pyogenes* Cas9.

A RuvC-like domain and an HNH-like domain

In an embodiment, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain and a RuvC-like domain. In an embodiment, cleavage activity is dependent on a RuvC-like domain and an HNH-like domain. A Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9

molecule or eaCas9 polypeptide, can comprise one or more of the following domains: a RuvC-like domain and an HNH-like domain. In an embodiment, a Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide and the eaCas9 molecule or eaCas9 polypeptide comprises a RuvC-like domain, e.g., a RuvC-like domain described below, and/or an HNH-like domain, e.g., an HNH-like domain described below.

RuvC-like domains

In an embodiment, a RuvC-like domain cleaves, a single strand, e.g., the non-complementary strand of the target nucleic acid molecule. The Cas9 molecule or Cas9 polypeptide can include more than one RuvC-like domain (e.g., one, two, three or more RuvC-like domains). In an embodiment, a RuvC-like domain is at least 5, 6, 7, 8 amino acids in length but not more than 20, 19, 18, 17, 16 or 15 amino acids in length. In an embodiment, the Cas9 molecule or Cas9 polypeptide comprises an N-terminal RuvC-like domain of about 10 to 20 amino acids, e.g., about 15 amino acids in length.

N-terminal RuvC-like domains

Some naturally occurring Cas9 molecules comprise more than one RuvC-like domain with cleavage being dependent on the N-terminal RuvC-like domain. Accordingly, Cas9 molecules or Cas9 polypeptide can comprise an N-terminal RuvC-like domain. Exemplary N-terminal RuvC-like domains are described below.

In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an N-terminal RuvC-like domain comprising an amino acid sequence of formula I:

D-X1-G-X2-X3-X4-X5-G-X6-X7-X8-X9 (SEQ ID NO: 8),

wherein,

X1 is selected from I, V, M, L and T (e.g., selected from I, V, and L);

X2 is selected from T, I, V, S, N, Y, E and L (e.g., selected from T, V, and I);

X3 is selected from N, S, G, A, D, T, R, M and F (e.g., A or N);

X4 is selected from S, Y, N and F (e.g., S);

X5 is selected from V, I, L, C, T and F (e.g., selected from V, I and L);

X6 is selected from W, F, V, Y, S and L (e.g., W);

X7 is selected from A, S, C, V and G (e.g., selected from A and S);

X8 is selected from V, I, L, A, M and H (e.g., selected from V, I, M and L); and

X9 is selected from any amino acid or is absent, designated by Δ (e.g., selected from T,

V, I, L, Δ, F, S, A, Y, M and R, or, e.g., selected from T, V, I, L and Δ).

In an embodiment, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:8, by as many as 1 but no more than 2, 3, 4, or 5 residues.

In embodiment, the N-terminal RuvC-like domain is cleavage competent.

5 In embodiment, the N-terminal RuvC-like domain is cleavage incompetent.

In an embodiment, a eaCas9 molecule or eaCas9 polypeptide comprises an N-terminal RuvC-like domain comprising an amino acid sequence of formula II:

D-X1-G-X2-X3-S-X5-G-X6-X7-X8-X9, (SEQ ID NO: 9),

wherein

10 X1 is selected from I, V, M, L and T (e.g., selected from I, V, and L);

X2 is selected from T, I, V, S, N, Y, E and L (e.g., selected from T, V, and I);

X3 is selected from N, S, G, A, D, T, R, M and F (e.g., A or N);

X5 is selected from V, I, L, C, T and F (e.g., selected from V, I and L);

X6 is selected from W, F, V, Y, S and L (e.g., W);

15 X7 is selected from A, S, C, V and G (e.g., selected from A and S);

X8 is selected from V, I, L, A, M and H (e.g., selected from V, I, M and L); and

X9 is selected from any amino acid or is absent (e.g., selected from T, V, I, L, Δ, F, S, A, Y, M and R or selected from e.g., T, V, I, L and Δ).

20 In an embodiment, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:9 by as many as 1 but no more than 2, 3, 4, or 5 residues.

In an embodiment, the N-terminal RuvC-like domain comprises an amino acid sequence of formula III:

D-I-G-X2-X3-S-V-G-W-A-X8-X9 (SEQ ID NO: 10),

wherein

25 X2 is selected from T, I, V, S, N, Y, E and L (e.g., selected from T, V, and I);

X3 is selected from N, S, G, A, D, T, R, M and F (e.g., A or N);

X8 is selected from V, I, L, A, M and H (e.g., selected from V, I, M and L); and

X9 is selected from any amino acid or is absent (e.g., selected from T, V, I, L, Δ, F, S, A, Y, M and R or selected from e.g., T, V, I, L and Δ).

30 In an embodiment, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:10 by as many as 1 but no more than, 2, 3, 4, or 5 residues.

In an embodiment, the N-terminal RuvC-like domain comprises an amino acid sequence of formula III:

D-I-G-T-N-S-V-G-W-A-V-X (SEQ ID NO: 11),

wherein

5 X is a non-polar alkyl amino acid or a hydroxyl amino acid, e.g., X is selected from V, I, L and T (e.g., the eaCas9 molecule can comprise an N-terminal RuvC-like domain shown in **Figs. 2A-2G** (is depicted as Y)).

In an embodiment, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:11 by as many as 1 but no more than, 2, 3, 4, or 5 residues.

10 In an embodiment, the N-terminal RuvC-like domain differs from a sequence of an N-terminal RuvC like domain disclosed herein, e.g., in **Figs. 3A-3B** or **Figs. 7A-7B**, as many as 1 but no more than 2, 3, 4, or 5 residues. In an embodiment, 1, 2, 3 or all of the highly conserved residues identified in **Figs. 3A-3B** or **Figs. 7A-7B** are present.

15 In an embodiment, the N-terminal RuvC-like domain differs from a sequence of an N-terminal RuvC-like domain disclosed herein, e.g., in **Figs. 4A-4B** or **Figs. 7A-7B**, as many as 1 but no more than 2, 3, 4, or 5 residues. In an embodiment, 1, 2, or all of the highly conserved residues identified in **Figs. 4A-4B** or **Figs. 7A-7B** are present.

Additional RuvC-like domains

20 In addition to the N-terminal RuvC-like domain, the Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, can comprise one or more additional RuvC-like domains. In an embodiment, the Cas9 molecule or Cas9 polypeptide can comprise two additional RuvC-like domains. Preferably, the additional RuvC-like domain is at least 5 amino acids in length and, e.g., less than 15 amino acids in length, e.g., 5 to 10 amino acids in length, e.g., 8 amino acids in length.

25 An additional RuvC-like domain can comprise an amino acid sequence:

I-X1-X2-E-X3-A-R-E (SEQ ID NO:12), wherein

X1 is V or H,

X2 is I, L or V (e.g., I or V); and

X3 is M or T.

30 In an embodiment, the additional RuvC-like domain comprises the amino acid sequence:

I-V-X2-E-M-A-R-E (SEQ ID NO:13), wherein

X2 is I, L or V (e.g., I or V) (e.g., the eaCas9 molecule or eaCas9 polypeptide can comprise an additional RuvC-like domain shown in **Fig. 2A-2G** or **Figs. 7A-7B** (depicted as B)).

An additional RuvC-like domain can comprise an amino acid sequence:

5 H-H-A-X1-D-A-X2-X3 (SEQ ID NO:14), wherein

X1 is H or L;

X2 is R or V; and

X3 is E or V.

In an embodiment, the additional RuvC-like domain comprises the amino acid sequence:

10 H-H-A-H-D-A-Y-L (SEQ ID NO:15).

In an embodiment, the additional RuvC-like domain differs from a sequence of SEQ ID NO:13, 15, 12 or 14 by as many as 1 but no more than 2, 3, 4, or 5 residues.

In some embodiments, the sequence flanking the N-terminal RuvC-like domain is a sequences of formula V:

15 K-X1'-Y-X2'-X3'-X4'-Z-T-D-X9'-Y, (SEQ ID NO:16).

wherein

X1' is selected from K and P,

X2' is selected from V, L, I, and F (e.g., V, I and L);

X3' is selected from G, A and S (e.g., G),

20 X4' is selected from L, I, V and F (e.g., L);

X9' is selected from D, E, N and Q; and

Z is an N-terminal RuvC-like domain, e.g., as described above.

HNH-like domains

In an embodiment, an HNH-like domain cleaves a single stranded complementary
25 domain, e.g., a complementary strand of a double stranded nucleic acid molecule. In an embodiment, an HNH-like domain is at least 15, 20, 25 amino acids in length but not more than 40, 35 or 30 amino acids in length, e.g., 20 to 35 amino acids in length, e.g., 25 to 30 amino acids in length. Exemplary HNH-like domains are described below.

In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an HNH-like
30 domain having an amino acid sequence of formula VI:

X1-X2-X3-H-X4-X5-P-X6-X7-X8-X9-X10-X11-X12-X13-X14-X15-N-X16-X17-X18-X19-X20-X21-X22-X23-N (SEQ ID NO:17), wherein

X1 is selected from D, E, Q and N (e.g., D and E);

X2 is selected from L, I, R, Q, V, M and K;

5 X3 is selected from D and E;

X4 is selected from I, V, T, A and L (e.g., A, I and V);

X5 is selected from V, Y, I, L, F and W (e.g., V, I and L);

X6 is selected from Q, H, R, K, Y, I, L, F and W;

X7 is selected from S, A, D, T and K (e.g., S and A);

10 X8 is selected from F, L, V, K, Y, M, I, R, A, E, D and Q (e.g., F);

X9 is selected from L, R, T, I, V, S, C, Y, K, F and G;

X10 is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

X11 is selected from D, S, N, R, L and T (e.g., D);

X12 is selected from D, N and S;

15 X13 is selected from S, A, T, G and R (e.g., S);

X14 is selected from I, L, F, S, R, Y, Q, W, D, K and H (e.g., I, L and F);

X15 is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y and V;

X16 is selected from K, L, R, M, T and F (e.g., L, R and K);

X17 is selected from V, L, I, A and T;

20 X18 is selected from L, I, V and A (e.g., L and I);

X19 is selected from T, V, C, E, S and A (e.g., T and V);

X20 is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H and A;

X21 is selected from S, P, R, K, N, A, H, Q, G and L;

X22 is selected from D, G, T, N, S, K, A, I, E, L, Q, R and Y; and

25 X23 is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D and F.

In an embodiment, a HNH-like domain differs from a sequence of SEQ ID NO:17 by at least one but no more than, 2, 3, 4, or 5 residues.

In an embodiment, the HNH-like domain is cleavage competent.

In an embodiment, the HNH-like domain is cleavage incompetent.

30 In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an HNH-like domain comprising an amino acid sequence of formula VII:

X1-X2-X3-H-X4-X5-P-X6-S-X8-X9-X10-D-D-S-X14-X15-N-K-V-L-X19-X20-X21-X22-X23-N (SEQ ID NO:18),

wherein

X1 is selected from D and E;

5 X2 is selected from L, I, R, Q, V, M and K;

X3 is selected from D and E;

X4 is selected from I, V, T, A and L (e.g., A, I and V);

X5 is selected from V, Y, I, L, F and W (e.g., V, I and L);

X6 is selected from Q, H, R, K, Y, I, L, F and W;

10 X8 is selected from F, L, V, K, Y, M, I, R, A, E, D and Q (e.g., F);

X9 is selected from L, R, T, I, V, S, C, Y, K, F and G;

X10 is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

X14 is selected from I, L, F, S, R, Y, Q, W, D, K and H (e.g., I, L and F);

X15 is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y and V;

15 X19 is selected from T, V, C, E, S and A (e.g., T and V);

X20 is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H and A;

X21 is selected from S, P, R, K, N, A, H, Q, G and L;

X22 is selected from D, G, T, N, S, K, A, I, E, L, Q, R and Y; and

X23 is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D and F.

20 In an embodiment, the HNH-like domain differs from a sequence of SEQ ID NO:18 by 1, 2, 3, 4, or 5 residues.

In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an HNH-like domain comprising an amino acid sequence of formula VII:

25 X1-V-X3-H-I-V-P-X6-S-X8-X9-X10-D-D-S-X14-X15-N-K-V-L-T-X20-X21-X22-X23-N (SEQ ID NO:19),

wherein

X1 is selected from D and E;

X3 is selected from D and E;

X6 is selected from Q, H, R, K, Y, I, L and W;

30 X8 is selected from F, L, V, K, Y, M, I, R, A, E, D and Q (e.g., F);

X9 is selected from L, R, T, I, V, S, C, Y, K, F and G;

X10 is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

X14 is selected from I, L, F, S, R, Y, Q, W, D, K and H (e.g., I, L and F);

X15 is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y and V;

X20 is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H and A;

5 X21 is selected from S, P, R, K, N, A, H, Q, G and L;

X22 is selected from D, G, T, N, S, K, A, I, E, L, Q, R and Y; and

X23 is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D and F.

In an embodiment, the HNH-like domain differs from a sequence of SEQ ID NO:19 by 1, 2, 3, 4, or 5 residues.

10 In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an HNH-like domain having an amino acid sequence of formula VIII:

D-X2-D-H-I-X5-P-Q-X7-F-X9-X10-D-X12-S-I-D-N-X16-V-L-X19-X20-S-X22-X23-N
(SEQ ID NO:20),

wherein

15 X2 is selected from I and V;

X5 is selected from I and V;

X7 is selected from A and S;

X9 is selected from I and L;

X10 is selected from K and T;

20 X12 is selected from D and N;

X16 is selected from R, K and L; X19 is selected from T and V;

X20 is selected from S and R;

X22 is selected from K, D and A; and

25 X23 is selected from E, K, G and N (e.g., the eaCas9 molecule or eaCas9 polypeptide can comprise an HNH-like domain as described herein).

In an embodiment, the HNH-like domain differs from a sequence of SEQ ID NO:20 by as many as 1 but no more than 2, 3, 4, or 5 residues.

In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises the amino acid sequence of formula IX:

30 L-Y-Y-L-Q-N-G-X1'-D-M-Y-X2'-X3'-X4'-X5'-L-D-I—X6'-X7'-L-S-X8'-Y-Z-N-R-X9'-K-X10'-D-X11'-V-P (SEQ ID NO:21),

wherein

X1' is selected from K and R;

X2' is selected from V and T;

X3' is selected from G and D;

5 X4' is selected from E, Q and D;

X5' is selected from E and D;

X6' is selected from D, N and H;

X7' is selected from Y, R and N;

X8' is selected from Q, D and N; X9' is selected from G and E;

10 X10' is selected from S and G;

X11' is selected from D and N; and

Z is an HNH-like domain, e.g., as described above.

In an embodiment, the eaCas9 molecule or eaCas9 polypeptide comprises an amino acid sequence that differs from a sequence of SEQ ID NO:21 by as many as 1 but no more than 2, 3,
15 4, or 5 residues.

In an embodiment, the HNH-like domain differs from a sequence of an HNH-like domain disclosed herein, e.g., in **Figs. 5A-5C** or **Figs. 7A-7B**, as many as 1 but no more than 2, 3, 4, or 5 residues. In an embodiment, 1 or both of the highly conserved residues identified in **Figs. 5A-5C** or **Figs. 7A-7B** are present.

20 In an embodiment, the HNH-like domain differs from a sequence of an HNH-like domain disclosed herein, e.g., in **Figs. 6A-6B** or **Figs. 7A-7B**, as many as 1 but no more than 2, 3, 4, or 5 residues. In an embodiment, 1, 2, all 3 of the highly conserved residues identified in **Figs. 6A-6B** or **Figs. 7A-7B** are present.

25 Cas9 Activities

Nuclease and Helicase Activities

In an embodiment, the Cas9 molecule or Cas9 polypeptide is capable of cleaving a target nucleic acid molecule. Typically wild type Cas9 molecules cleave both strands of a target nucleic acid molecule. Cas9 molecules and Cas9 polypeptides can be engineered to alter
30 nuclease cleavage (or other properties), e.g., to provide a Cas9 molecule or Cas9 polypeptide which is a nickase, or which lacks the ability to cleave target nucleic acid. A Cas9 molecule or

Cas9 polypeptide that is capable of cleaving a target nucleic acid molecule is referred to herein as an eaCas9 molecule or eaCas9 polypeptide. In an embodiment, an eaCas9 molecule or Cas9 polypeptide comprises one or more of the following activities:

a nickase activity, i.e., the ability to cleave a single strand, e.g., the non-complementary strand or the complementary strand, of a nucleic acid molecule;

a double stranded nuclease activity, i.e., the ability to cleave both strands of a double stranded nucleic acid and create a double stranded break, which in a embodiment is the presence of two nickase activities;

an endonuclease activity;

an exonuclease activity; and

a helicase activity, i.e., the ability to unwind the helical structure of a double stranded nucleic acid.

In an embodiment, an enzymatically active Cas9 or eaCas9 molecule or eaCas9 polypeptide cleaves both strands and results in a double stranded break. In an embodiment, an eaCas9 molecule cleaves only one strand, e.g., the strand to which the gRNA hybridizes to, or the strand complementary to the strand the gRNA hybridizes with. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH-like domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an N-terminal RuvC-like domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH-like domain and cleavage activity associated with an N-terminal RuvC-like domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an active, or cleavage competent, HNH-like domain and an inactive, or cleavage incompetent, N-terminal RuvC-like domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an inactive, or cleavage incompetent, HNH-like domain and an active, or cleavage competent, N-terminal RuvC-like domain.

Some Cas9 molecules or Cas9 polypeptides have the ability to interact with a gRNA molecule, and in conjunction with the gRNA molecule localize to a core target domain, but are incapable of cleaving the target nucleic acid, or incapable of cleaving at efficient rates. Cas9 molecules having no, or no substantial, cleavage activity are referred to herein as an eiCas9 molecule or eiCas9 polypeptide. For example, an eiCas9 molecule or eiCas9 polypeptide can

lack cleavage activity or have substantially less, e.g., less than 20, 10, 5, 1 or 0.1 % of the cleavage activity of a reference Cas9 molecule or eiCas9 polypeptide, as measured by an assay described herein.

Targeting and PAMs

5 A Cas9 molecule or Cas9 polypeptide, is a polypeptide that can interact with a guide RNA (gRNA) molecule and, in concert with the gRNA molecule, localizes to a site which comprises a target domain and PAM sequence.

In an embodiment, the ability of an eaCas9 molecule or eaCas9 polypeptide to interact with and cleave a target nucleic acid is PAM sequence dependent. A PAM sequence is a
 10 sequence in the target nucleic acid. In an embodiment, cleavage of the target nucleic acid occurs upstream from the PAM sequence. eaCas9 molecules from different bacterial species can recognize different sequence motifs (e.g., PAM sequences). In an embodiment, an eaCas9 molecule of *S. pyogenes* recognizes the sequence motif NGG and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, base pairs upstream from that sequence. See, e.g.,
 15 Mali *et al.*, SCIENCE 2013; 339(6121): 823-826. In an embodiment, an eaCas9 molecule of *S. thermophilus* recognizes the sequence motif NGGNG and NNAGAAW (W = A or T) and directs cleavage of a core target nucleic acid sequence 1 to 10, e.g., 3 to 5, base pairs upstream from these sequences. See, e.g., Horvath *et al.*, SCIENCE 2010; 327(5962):167-170, and Deveau *et al.*, J BACTERIOL 2008; 190(4): 1390-1400. In an embodiment, an eaCas9 molecule of *S. mutans*
 20 recognizes the sequence motif NGG and/or NAAR (R = A or G) and directs cleavage of a core target nucleic acid sequence 1 to 10, e.g., 3 to 5 base pairs, upstream from this sequence. See, e.g., Deveau *et al.*, J BACTERIOL 2008; 190(4): 1390-1400. In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R = A or G) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, base pairs upstream from that sequence. In
 25 an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R = A or G) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, base pairs upstream from that sequence. In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRT (R = A or G) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, base pairs upstream from that sequence. In an embodiment, an eaCas9
 30 molecule of *S. aureus* recognizes the sequence motif NNGRRV (R = A or G, V = A, G or C) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, base pairs upstream from

that sequence. In an embodiment, an eaCas9 molecule of *Neisseria meningitidis* recognizes the sequence motif NNNNGATT or NNNGCTT and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, base pairs upstream from that sequence. See, e.g., Hou *et al.*, PNAS Early Edition 2013, 1-6. The ability of a Cas9 molecule to recognize a PAM sequence
5 can be determined, e.g., using a transformation assay described in Jinek *et al.*, SCIENCE 2012 337:816. In the aforementioned embodiments, N can be any nucleotide residue, e.g., any of A, G, C or T.

As is discussed herein, Cas9 molecules can be engineered to alter the PAM specificity of the Cas9 molecule.

10 Exemplary naturally occurring Cas9 molecules are described in Chylinski *et al.*, RNA BIOLOGY 2013 10:5, 727-737. Such Cas9 molecules include Cas9 molecules of a cluster 1 bacterial family, cluster 2 bacterial family, cluster 3 bacterial family, cluster 4 bacterial family, cluster 5 bacterial family, cluster 6 bacterial family, a cluster 7 bacterial family, a cluster 8 bacterial family, a cluster 9 bacterial family, a cluster 10 bacterial family, a cluster 11 bacterial
15 family, a cluster 12 bacterial family, a cluster 13 bacterial family, a cluster 14 bacterial family, a cluster 15 bacterial family, a cluster 16 bacterial family, a cluster 17 bacterial family, a cluster 18 bacterial family, a cluster 19 bacterial family, a cluster 20 bacterial family, a cluster 21 bacterial family, a cluster 22 bacterial family, a cluster 23 bacterial family, a cluster 24 bacterial family, a cluster 25 bacterial family, a cluster 26 bacterial family, a cluster 27 bacterial family, a cluster 28
20 bacterial family, a cluster 29 bacterial family, a cluster 30 bacterial family, a cluster 31 bacterial family, a cluster 32 bacterial family, a cluster 33 bacterial family, a cluster 34 bacterial family, a cluster 35 bacterial family, a cluster 36 bacterial family, a cluster 37 bacterial family, a cluster 38 bacterial family, a cluster 39 bacterial family, a cluster 40 bacterial family, a cluster 41 bacterial family, a cluster 42 bacterial family, a cluster 43 bacterial family, a cluster 44 bacterial family, a
25 cluster 45 bacterial family, a cluster 46 bacterial family, a cluster 47 bacterial family, a cluster 48 bacterial family, a cluster 49 bacterial family, a cluster 50 bacterial family, a cluster 51 bacterial family, a cluster 52 bacterial family, a cluster 53 bacterial family, a cluster 54 bacterial family, a cluster 55 bacterial family, a cluster 56 bacterial family, a cluster 57 bacterial family, a cluster 58 bacterial family, a cluster 59 bacterial family, a cluster 60 bacterial family, a cluster 61 bacterial
30 family, a cluster 62 bacterial family, a cluster 63 bacterial family, a cluster 64 bacterial family, a cluster 65 bacterial family, a cluster 66 bacterial family, a cluster 67 bacterial family, a cluster 68

bacterial family, a cluster 69 bacterial family, a cluster 70 bacterial family, a cluster 71 bacterial family, a cluster 72 bacterial family, a cluster 73 bacterial family, a cluster 74 bacterial family, a cluster 75 bacterial family, a cluster 76 bacterial family, a cluster 77 bacterial family, or a cluster 78 bacterial family.

5 Exemplary naturally occurring Cas9 molecules include a Cas9 molecule of a cluster 1 bacterial family. Examples include a Cas9 molecule of: *S. pyogenes* (e.g., strain SF370, MGAS10270, MGAS10750, MGAS2096, MGAS315, MGAS5005, MGAS6180, MGAS9429, NZ131 and SSI-1), *S. thermophilus* (e.g., strain LMD-9), *S. pseudoporcinus* (e.g., strain SPIN 20026), *S. mutans* (e.g., strain UA159, NN2025), *S. macacae* (e.g., strain NCTC11558), *S.*
 10 *gallolyticus* (e.g., strain UCN34, ATCC BAA-2069), *S. equines* (e.g., strain ATCC 9812, MGCS 124), *S. dysdalactiae* (e.g., strain GGS 124), *S. bovis* (e.g., strain ATCC 700338), *S. anginosus* (e.g., strain F0211), *S. agalactiae* (e.g., strain NEM316, A909), *Listeria monocytogenes* (e.g., strain F6854), *Listeria innocua* (*L. innocua*, e.g., strain Clip11262), *Enterococcus italicus* (e.g., strain DSM 15952), or *Enterococcus faecium* (e.g., strain 1,231,408). Another exemplary Cas9
 15 molecule is a Cas9 molecule of *Neisseria meningitides* (Hou *et al.*, PNAS Early Edition 2013, 1-6.

In an embodiment, a Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, comprises an amino acid sequence:

20 having 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with;
 differs at no more than, 2, 5, 10, 15, 20, 30, or 40% of the amino acid residues when compared with;
 differs by at least 1, 2, 5, 10 or 20 amino acids but by no more than 100, 80, 70, 60, 50, 40 or 30 amino acids from; or

25 is identical to any Cas9 molecule sequence described herein, or a naturally occurring Cas9 molecule sequence, e.g., a Cas9 molecule from a species listed herein or described in Chylinski *et al.*, RNA BIOLOGY 2013 10:5, 727-737; Hou *et al.*, PNAS Early Edition 2013, 1-6; e.g., SEQ ID NOs:1-4. In an embodiment, the Cas9 molecule or Cas9 polypeptide comprises one or more of the following activities: a nickase activity; a double stranded cleavage activity
 30 (e.g., an endonuclease and/or exonuclease activity); a helicase activity; or the ability, together with a gRNA molecule, to home to a target nucleic acid.

In an embodiment, a Cas9 molecule or Cas9 polypeptide comprises any of the amino acid sequence of the consensus sequence of **Figs. 2A-2G**, wherein “*” indicates any amino acid found in the corresponding position in the amino acid sequence of a Cas9 molecule of *S.*

pyogenes, *S. thermophilus*, *S. mutans* and *L. innocua*, and “-” indicates any amino acid. In an

embodiment a Cas9 molecule or Cas9 polypeptide differs from the sequence of the consensus sequence disclosed in **Figs. 2A-2G** by at least 1, but no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10

amino acid residues. In an embodiment, a Cas9 molecule or Cas9 polypeptide comprises the amino acid sequence of SEQ ID NO:7 of **Figs. 7A-7B**, wherein “*” indicates any amino acid found in the corresponding position in the amino acid sequence of a Cas9 molecule of *S.*

pyogenes, or *N. meningitides*, “-” indicates any amino acid, and “-” indicates any amino acid or absent. In an embodiment, a Cas9 molecule or Cas9 polypeptide differs from the sequence of SEQ ID NO:6 or 7 disclosed in **Figs. 7A-7B** by at least 1, but no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues.

A comparison of the sequence of a number of Cas9 molecules indicate that certain regions are conserved. These are identified below as:

region 1 (residues1 to 180, or in the case of region 1' residues 120 to 180)

region 2 (residues360 to 480);

region 3 (residues 660 to 720);

region 4 (residues 817 to 900); and

region 5 (residues 900 to 960);

In an embodiment, a Cas9 molecule or Cas9 polypeptide comprises regions 1-5, together with sufficient additional Cas9 molecule sequence to provide a biologically active molecule, e.g., a Cas9 molecule having at least one activity described herein. In an embodiment, each of regions 1-6, independently, have, 50%, 60%, 70%, or 80% homology with the corresponding residues of a Cas9 molecule or Cas9 polypeptide described herein, e.g., a sequence from **Fig. 2A-2G** or from **Figs. 7A-7B**.

In an embodiment, a Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, comprises an amino acid sequence referred to as region 1:

having 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with amino acids 1-180 (the numbering is according to the motif sequence in **Figs. 2A-2G**; 52% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of

Cas9 of *S. pyogenes*;

differs by at least 1, 2, 5, 10 or 20 amino acids but by no more than 90, 80, 70, 60, 50, 40 or 30 amino acids from amino acids 1-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *Listeria innocua*; or

5 is identical to 1-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*.

In an embodiment, a Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, comprises an amino acid sequence referred to as region 1':

10 having 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with amino acids 120-180 (55% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 120-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua* ; or

15 is identical to 120-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*.

In an embodiment, a Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, comprises an amino acid sequence referred to as region 2:

20 having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with amino acids 360-480 (52% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 360-480 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*; or

25 is identical to 360-480 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*.

In an embodiment, a Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, comprises an amino acid sequence referred to as region 3:

30 having 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%

homology with amino acids 660-720 (56% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 660-720 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*; or

is identical to 660-720 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*.

In an embodiment, a Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, comprises an amino acid sequence referred to as region 4:

having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 817-900 (55% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 817-900 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*; or

is identical to 817-900 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*.

In an embodiment, a Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, comprises an amino acid sequence referred to as region 5:

having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 900-960 (60% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 900-960 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*; or

is identical to 900-960 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*.

Engineered or Altered Cas9 Molecules and Cas9 Polypeptides

Cas9 molecules and Cas9 polypeptides described herein, e.g., naturally occurring Cas9 molecules, can possess any of a number of properties, including: nickase activity, nuclease activity (e.g., endonuclease and/or exonuclease activity); helicase activity; the ability to associate functionally with a gRNA molecule; and the ability to target (or localize to) a site on a nucleic acid (e.g., PAM recognition and specificity). In an embodiment, a Cas9 molecule or Cas9 polypeptide can include all or a subset of these properties. In typical embodiments, a Cas9 molecule or Cas9 polypeptide has the ability to interact with a gRNA molecule and, in concert with the gRNA molecule, localize to a site in a nucleic acid. Other activities, e.g., PAM specificity, cleavage activity, or helicase activity can vary more widely in Cas9 molecules and Cas9 polypeptides.

Cas9 molecules include engineered Cas9 molecules and engineered Cas9 polypeptides (engineered, as used in this context, means merely that the Cas9 molecule or Cas9 polypeptide differs from a reference sequences, and implies no process or origin limitation). An engineered Cas9 molecule or Cas9 polypeptide can comprise altered enzymatic properties, e.g., altered nuclease activity, (as compared with a naturally occurring or other reference Cas9 molecule) or altered helicase activity. As discussed herein, an engineered Cas9 molecule or Cas9 polypeptide can have nickase activity (as opposed to double strand nuclease activity). In an embodiment an engineered Cas9 molecule or Cas9 polypeptide can have an alteration that alters its size, e.g., a deletion of amino acid sequence that reduces its size, e.g., without significant effect on one or more, or any Cas9 activity. In an embodiment, an engineered Cas9 molecule or Cas9 polypeptide can comprise an alteration that affects PAM recognition. E.g., an engineered Cas9 molecule can be altered to recognize a PAM sequence other than that recognized by the endogenous wild-type PI domain. In an embodiment, a Cas9 molecule or Cas9 polypeptide can differ in sequence from a naturally occurring Cas9 molecule but not have significant alteration in one or more Cas9 activities.

Cas9 molecules or Cas9 polypeptides with desired properties can be made in a number of ways, e.g., by alteration of a parental, e.g., naturally occurring Cas9 molecules or Cas9 polypeptides to provide an altered Cas9 molecule or Cas9 polypeptide having a desired property. For example, one or more mutations or differences relative to a parental Cas9 molecule, e.g., a naturally occurring or engineered Cas9 molecule, can be introduced. Such mutations and

differences comprise: substitutions (e.g., conservative substitutions or substitutions of non-essential amino acids); insertions; or deletions. In an embodiment, a Cas9 molecule or Cas9 polypeptide can comprises one or more mutations or differences, e.g., at least 1, 2, 3, 4, 5, 10, 15, 20, 30, 40 or 50 mutations, but less than 200, 100, or 80 mutations relative to a reference, e.g., a parental, Cas9 molecule.

In an embodiment, a mutation or mutations do not have a substantial effect on a Cas9 activity, e.g. a Cas9 activity described herein. In an embodiment, a mutation or mutations have a substantial effect on a Cas9 activity, e.g. a Cas9 activity described herein.

Non-Cleaving and Modified-Cleavage Cas9 Molecules and Cas9 Polypeptides

In an embodiment, a Cas9 molecule or Cas9 polypeptide comprises a cleavage property that differs from naturally occurring Cas9 molecules, e.g., that differs from the naturally occurring Cas9 molecule having the closest homology. For example, a Cas9 molecule or Cas9 polypeptide can differ from naturally occurring Cas9 molecules, e.g., a Cas9 molecule of *S. pyogenes*, as follows: its ability to modulate, e.g., decreased or increased, cleavage of a double stranded nucleic acid (endonuclease and/or exonuclease activity), e.g., as compared to a naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S. pyogenes*); its ability to modulate, e.g., decreased or increased, cleavage of a single strand of a nucleic acid, e.g., a non-complementary strand of a nucleic acid molecule or a complementary strand of a nucleic acid molecule (nickase activity) , e.g., as compared to a naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S. pyogenes*); or the ability to cleave a nucleic acid molecule, e.g., a double stranded or single stranded nucleic acid molecule, can be eliminated.

Modified Cleavage eaCas9 Molecules and eaCas9 Polypeptides

In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises one or more of the following activities: cleavage activity associated with an N-terminal RuvC-like domain; cleavage activity associated with an HNH-like domain; cleavage activity associated with an HNH-like domain and cleavage activity associated with an N-terminal RuvC-like domain.

In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an active, or cleavage competent, HNH-like domain (e.g., an HNH-like domain described herein, e.g., SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20 or SEQ ID NO: 21) and an inactive,

or cleavage incompetent, N-terminal RuvC-like domain. An exemplary inactive, or cleavage incompetent N-terminal RuvC-like domain can have a mutation of an aspartic acid in an N-terminal RuvC-like domain, e.g., an aspartic acid at position 9 of the consensus sequence disclosed in **Figs. 2A-2G** or an aspartic acid at position 10 of SEQ ID NO:7, e.g., can be substituted with an alanine. In an embodiment, the eaCas9 molecule or eaCas9 polypeptide differs from wild type in the N-terminal RuvC-like domain and does not cleave the target nucleic acid, or cleaves with significantly less efficiency, e.g., less than 20, 10, 5, 1 or .1 % of the cleavage activity of a reference Cas9 molecule, e.g., as measured by an assay described herein. The reference Cas9 molecule can be a naturally occurring unmodified Cas9 molecule, e.g., a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, or *S. thermophilus*. In an embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology.

In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an inactive, or cleavage incompetent, HNH domain and an active, or cleavage competent, N-terminal RuvC-like domain (e.g., an HNH-like domain described herein, e.g., SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14 or SEQ ID NO:15). Exemplary inactive, or cleavage incompetent HNH-like domains can have a mutation at one or more of: a histidine in an HNH-like domain, e.g., a histidine shown at position 856 of the consensus sequence disclosed in **Figs. 2A-2G**, e.g., can be substituted with an alanine; and one or more asparagines in an HNH-like domain, e.g., an asparagine shown at position 870 of the consensus sequence disclosed in **Figs. 2A-2G** and/or at position 879 of the consensus sequence disclosed in **Figs. 2A-2G**, e.g., can be substituted with an alanine. In an embodiment, the eaCas9 differs from wild type in the HNH-like domain and does not cleave the target nucleic acid, or cleaves with significantly less efficiency, e.g., less than 20, 10, 5, 1 or 0.1% of the cleavage activity of a reference Cas9 molecule, e.g., as measured by an assay described herein. The reference Cas9 molecule can be a naturally occurring unmodified Cas9 molecule, e.g., a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, or *S. thermophilus*. In an embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology.

Alterations in the Ability to Cleave One or Both Strands of a Target Nucleic Acid

In an embodiment, exemplary Cas9 activities comprise one or more of PAM specificity, cleavage activity, and helicase activity. A mutation(s) can be present, e.g., in one or more RuvC-like domain, e.g., an N-terminal RuvC-like domain; an HNH-like domain; a region outside the RuvC-like domains and the HNH-like domain. In some embodiments, a mutation(s) is present in a RuvC-like domain, e.g., an N-terminal RuvC-like domain. In some embodiments, a mutation(s) is present in an HNH-like domain. In some embodiments, mutations are present in both a RuvC-like domain, e.g., an N-terminal RuvC-like domain and an HNH-like domain.

Exemplary mutations that may be made in the RuvC domain or HNH domain with reference to the *S. pyogenes* sequence include: D10A, E762A, H840A, N854A, N863A and/or D986A.

In an embodiment, a Cas9 molecule or Cas9 polypeptide is an eiCas9 molecule or eiCas9 polypeptide comprising one or more differences in a RuvC domain and/or in an HNH domain as compared to a reference Cas9 molecule, and the eiCas9 molecule or eiCas9 polypeptide does not cleave a nucleic acid, or cleaves with significantly less efficiency than does wildtype, e.g., when compared with wild type in a cleavage assay, e.g., as described herein, cuts with less than 50, 25, 10, or 1% of a reference Cas9 molecule, as measured by an assay described herein.

Whether or not a particular sequence, e.g., a substitution, may affect one or more activity, such as targeting activity, cleavage activity, etc, can be evaluated or predicted, e.g., by evaluating whether the mutation is conservative or by the method described in Section IV. In an embodiment, a “non-essential” amino acid residue, as used in the context of a Cas9 molecule, is a residue that can be altered from the wild-type sequence of a Cas9 molecule, e.g., a naturally occurring Cas9 molecule, e.g., an eaCas9 molecule, without abolishing or more preferably, without substantially altering a Cas9 activity (e.g., cleavage activity), whereas changing an “essential” amino acid residue results in a substantial loss of activity (e.g., cleavage activity).

In an embodiment, a Cas9 molecule or Cas9 polypeptide comprises a cleavage property that differs from naturally occurring Cas9 molecules, e.g., that differs from the naturally occurring Cas9 molecule having the closest homology. For example, a Cas9 molecule or Cas9 polypeptide can differ from naturally occurring Cas9 molecules, e.g., a Cas9 molecule of *S. aureus*, *S. pyogenes*, or *C. jejuni* as follows: its ability to modulate, e.g., decreased or increased, cleavage of a double stranded break (endonuclease and/or exonuclease activity), e.g., as

compared to a naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S. aureus*, *S. pyogenes*, or *C. jejuni*); its ability to modulate, e.g., decreased or increased, cleavage of a single strand of a nucleic acid, e.g., a non-complimentary strand of a nucleic acid molecule or a complementary strand of a nucleic acid molecule (nickase activity), e.g., as compared to a
 5 naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S. aureus*, *S. pyogenes*, or *C. jejuni*); or the ability to cleave a nucleic acid molecule, e.g., a double stranded or single stranded nucleic acid molecule, can be eliminated.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising one or more of the following activities: cleavage activity
 10 associated with a RuvC domain; cleavage activity associated with an HNH domain; cleavage activity associated with an HNH domain and cleavage activity associated with a RuvC domain.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eiCas9 molecule or eiCas9 polypeptide which does not cleave a nucleic acid molecule (either double stranded or single stranded nucleic acid molecules) or cleaves a nucleic acid molecule with significantly less
 15 efficiency, e.g., less than 20, 10, 5, 1 or 0.1% of the cleavage activity of a reference Cas9 molecule, e.g., as measured by an assay described herein. The reference Cas9 molecule can be a naturally occurring unmodified Cas9 molecule, e.g., a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, *S. thermophilus*, *S. aureus*, *C. jejuni* or *N. meningitidis*. In an embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the
 20 closest sequence identity or homology. In an embodiment, the eiCas9 molecule or eiCas9 polypeptide lacks substantial cleavage activity associated with a RuvC domain and cleavage activity associated with an HNH domain.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the fixed amino acid residues of *S. pyogenes* shown in the
 25 consensus sequence disclosed in **Figs. 2A-2G**, and has one or more amino acids that differ from the amino acid sequence of *S. pyogenes* (e.g., has a substitution) at one or more residue (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, 200 amino acid residues) represented by an “-” in the consensus sequence disclosed in **Figs. 2A-2G** or SEQ ID NO:7.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide comprises a sequence
 30 in which:

the sequence corresponding to the fixed sequence of the consensus sequence disclosed in **Figs. 2A-2G** differs at no more than 1, 2, 3, 4, 5, 10, 15, or 20% of the fixed residues in the consensus sequence disclosed in **Figs. 2A-2G**;

the sequence corresponding to the residues identified by “*” in the consensus sequence disclosed in **Figs. 2A-2G** differ at no more than 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, or 40% of the “*” residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an *S. pyogenes* Cas9 molecule; and,

the sequence corresponding to the residues identified by “-” in the consensus sequence disclosed in **Figs. 2A-2G** differ at no more than 5, 10, 15, 20, 25, 30, 35, 40, 45, 55, or 60% of the “-” residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an *S. pyogenes* Cas9 molecule.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the fixed amino acid residues of *S. thermophilus* shown in the consensus sequence disclosed in **Figs. 2A-2G**, and has one or more amino acids that differ from the amino acid sequence of *S. thermophilus* (e.g., has a substitution) at one or more residue (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, 200 amino acid residues) represented by an “-” in the consensus sequence disclosed in **Figs. 2A-2G**.

In an embodiment the altered Cas9 molecule or Cas9 polypeptide comprises a sequence in which:

the sequence corresponding to the fixed sequence of the consensus sequence disclosed in **Figs. 2A-2G** differs at no more than 1, 2, 3, 4, 5, 10, 15, or 20% of the fixed residues in the consensus sequence disclosed in **Figs. 2A-2G**;

the sequence corresponding to the residues identified by “*” in the consensus sequence disclosed in **Figs. 2A-2G** differ at no more than 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, or 40% of the “*” residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an *S. thermophilus* Cas9 molecule; and,

the sequence corresponding to the residues identified by “-” in the consensus sequence disclosed in **Figs. 2A-2G** differ at no more than 5, 10, 15, 20, 25, 30, 35, 40, 45, 55, or 60% of the “-” residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an *S. thermophilus* Cas9 molecule.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the fixed amino acid residues of *S. mutans* shown in the consensus sequence disclosed in **Figs. 2A-2G**, and has one or more amino acids that differ from the amino acid sequence of *S. mutans* (e.g., has a substitution) at one or more residue (e.g., 2, 3,
 5 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, 200 amino acid residues) represented by an “-” in the consensus sequence disclosed in **Figs. 2A-2G**.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide comprises a sequence in which:

the sequence corresponding to the fixed sequence of the consensus sequence disclosed in
 10 **Figs. 2A-2G** differs at no more than 1, 2, 3, 4, 5, 10, 15, or 20% of the fixed residues in the consensus sequence disclosed in **Figs. 2A-2G**;

the sequence corresponding to the residues identified by “*” in the consensus sequence disclosed in **Figs. 2A-2G** differ at no more than 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, or 40% of the “*” residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an
 15 *S. mutans* Cas9 molecule; and,

the sequence corresponding to the residues identified by “-” in the consensus sequence disclosed in **Figs. 2A-2G** differ at no more than 5, 10, 15, 20, 25, 30, 35, 40, 45, 55, or 60% of the “-” residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an
 20 *S. mutans* Cas9 molecule.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the fixed amino acid residues of *L. innocua* shown in the consensus sequence disclosed in **Figs. 2A-2G**, and has one or more amino acids that differ from the amino acid sequence of *L. innocua* (e.g., has a substitution) at one or more residue (e.g., 2, 3,
 5 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, 200 amino acid residues) represented by an “-” in the
 25 consensus sequence disclosed in **Figs. 2A-2G**.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide comprises a sequence in which:

the sequence corresponding to the fixed sequence of the consensus sequence disclosed in
Figs. 2A-2G differs at no more than 1, 2, 3, 4, 5, 10, 15, or 20% of the fixed residues in the
 30 consensus sequence disclosed in **Figs. 2A-2G**;

the sequence corresponding to the residues identified by “*” in the consensus sequence disclosed in **Figs. 2A-2G** differ at no more than 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, or 40% of the “*” residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an *L. innocula* Cas9 molecule; and,

the sequence corresponding to the residues identified by “-” in the consensus sequence disclosed in **Figs. 2A-2G** differ at no more than 5, 10, 15, 20, 25, 30, 35, 40, 45, 55, or 60% of the “-” residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an *L. innocula* Cas9 molecule.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, can be a fusion, e.g., of two or more different Cas9 molecules, e.g., of two or more naturally occurring Cas9 molecules of different species. For example, a fragment of a naturally occurring Cas9 molecule of one species can be fused to a fragment of a Cas9 molecule of a second species. As an example, a fragment of Cas9 molecule of *S. pyogenes* comprising an N-terminal RuvC-like domain can be fused to a fragment of Cas9 molecule of a species other than *S. pyogenes* (e.g., *S. thermophilus*) comprising an HNH-like domain.

Cas9 Molecules and Cas9 Polypeptides with Altered PAM Recognition or No PAM Recognition

Naturally occurring Cas9 molecules can recognize specific PAM sequences, for example the PAM recognition sequences described above for *S. pyogenes*, *S. thermophilus* and *S. mutans*.

In an embodiment, a Cas9 molecule or Cas9 polypeptide has the same PAM specificities as a naturally occurring Cas9 molecule. In other embodiments, a Cas9 molecule or Cas9 polypeptide has a PAM specificity not associated with a naturally occurring Cas9 molecule, or a PAM specificity not associated with the naturally occurring Cas9 molecule to which it has the closest sequence homology. For example, a naturally occurring Cas9 molecule can be altered, e.g., to alter PAM recognition, e.g., to alter the PAM sequence that the Cas9 molecule recognizes to decrease off target sites and/or improve specificity; or eliminate a PAM recognition requirement. In an embodiment, a Cas9 molecule or Cas9 polypeptide can be altered, e.g., to increase length of PAM recognition sequence and/or improve Cas9 specificity to high level of identity (e.g., 98%, 99% or 100% match between gRNA and a PAM sequence), e.g., to decrease off target sites and increase specificity. In an embodiment, the length of the PAM recognition

sequence is at least 4, 5, 6, 7, 8, 9, 10 or 15 amino acids in length. In an embodiment, the Cas9 specificity requires at least 90%, 95%, 96%, 97%, 98%, 99% or more homology between the gRNA and the PAM sequence. Cas9 molecules or Cas9 polypeptides that recognize different PAM sequences and/or have reduced off-target activity can be generated using directed evolution. Exemplary methods and systems that can be used for directed evolution of Cas9 molecules are described, e.g., in Esvelt *et al.* NATURE 2011, 472(7344): 499-503. Candidate Cas9 molecules can be evaluated, e.g., by methods described in Section IV.

Alterations of the PI domain, which mediates PAM recognition, are discussed below.

Synthetic Cas9 Molecules and Cas9 Polypeptides with Altered PI Domains

Current genome-editing methods are limited in the diversity of target sequences that can be targeted by the PAM sequence that is recognized by the Cas9 molecule utilized. A synthetic Cas9 molecule (or Syn-Cas9 molecule), or synthetic Cas9 polypeptide (or Syn-Cas9 polypeptide), as that term is used herein, refers to a Cas9 molecule or Cas9 polypeptide that comprises a Cas9 core domain from one bacterial species and a functional altered PI domain, i.e., a PI domain other than that naturally associated with the Cas9 core domain, e.g., from a different bacterial species.

In an embodiment, the altered PI domain recognizes a PAM sequence that is different from the PAM sequence recognized by the naturally-occurring Cas9 from which the Cas9 core domain is derived. In an embodiment, the altered PI domain recognizes the same PAM sequence recognized by the naturally-occurring Cas9 from which the Cas9 core domain is derived, but with different affinity or specificity. A Syn-Cas9 molecule or Syn-Cas9 polypeptide can be, respectively, a Syn-eaCas9 molecule or Syn-eaCas9 polypeptide or a Syn-eiCas9 molecule Syn-eiCas9 polypeptide.

An exemplary Syn-Cas9 molecule or Syn-Cas9 polypeptide comprises:

- a) a Cas9 core domain, e.g., a Cas9 core domain from **Table 7** or **8**, e.g., a *S. aureus*, *S. pyogenes*, or *C. jejuni* Cas9 core domain; and
- b) an altered PI domain from a species X Cas9 sequence selected from **Tables 10** and **11**.

In an embodiment, the RKR motif (the PAM binding motif) of said altered PI domain comprises: differences at 1, 2, or 3 amino acid residues; a difference in amino acid sequence at the first, second, or third position; differences in amino acid sequence at the first and second

positions, the first and third positions, or the second and third positions; as compared with the sequence of the RKR motif of the native or endogenous PI domain associated with the Cas9 core domain.

5 In an embodiment, the Cas9 core domain comprises the Cas9 core domain from a species X Cas9 from **Table 7** and said altered PI domain comprises a PI domain from a species Y Cas9 from **Table 7**.

In an embodiment, the RKR motif of the species X Cas9 is other than the RKR motif of the species Y Cas9.

10 In an embodiment, the RKR motif of the altered PI domain is selected from XXY, XNG, and XNQ.

In an embodiment, the altered PI domain has at least 60, 70, 80, 90, 95, or 100% homology with the amino acid sequence of a naturally occurring PI domain of said species Y from **Table 7**.

15 In an embodiment, the altered PI domain differs by no more than 50, 40, 30, 25, 20, 15, 10, 5, 4, 3, 2, or 1 amino acid residue from the amino acid sequence of a naturally occurring PI domain of said second species from **Table 7**.

20 In an embodiment, the Cas9 core domain comprises a *S. aureus* core domain and altered PI domain comprises: an *A. denitrificans* PI domain; a *C. jejuni* PI domain; a *H. mustelae* PI domain; or an altered PI domain of species X PI domain, wherein species X is selected from **Table 11**.

In an embodiment, the Cas9 core domain comprises a *S. pyogenes* core domain and the altered PI domain comprises: an *A. denitrificans* PI domain; a *C. jejuni* PI domain; a *H. mustelae* PI domain; or an altered PI domain of species X PI domain, wherein species X is selected from **Table 11**.

25 In an embodiment, the Cas9 core domain comprises a *C. jejuni* core domain and the altered PI domain comprises: an *A. denitrificans* PI domain; a *H. mustelae* PI domain; or an altered PI domain of species X PI domain, wherein species X is selected from **Table 11**.

In an embodiment, the Cas9 molecule or Cas9 polypeptide further comprises a linker disposed between said Cas9 core domain and said altered PI domain.

In an embodiment, the linker comprises: a linker described elsewhere herein disposed between the Cas9 core domain and the heterologous PI domain. Suitable linkers are further described in Section V.

Exemplary altered PI domains for use in Syn-Cas9 molecules are described in **Tables 10** and **11**. The sequences for the 83 Cas9 orthologs referenced in **Tables 10** and **11** are provided in **Table 7**. **Table 9** provides the Cas9 orthologs with known PAM sequences and the corresponding RKR motif.

In an embodiment, a Syn-Cas9 molecule or Syn-Cas9 polypeptide may also be size-optimized, e.g., the Syn-Cas9 molecule or Syn-Cas9 polypeptide comprises one or more deletions, and optionally one or more linkers disposed between the amino acid residues flanking the deletions. In an embodiment, a Syn-Cas9 molecule or Syn-Cas9 polypeptide comprises a REC deletion.

Size-Optimized Cas9 Molecules and Cas9 Polypeptides

Engineered Cas9 molecules and engineered Cas9 polypeptides described herein include a Cas9 molecule or Cas9 polypeptide comprising a deletion that reduces the size of the molecule while still retaining desired Cas9 properties, e.g., essentially native conformation, Cas9 nuclease activity, and/or target nucleic acid molecule recognition. Provided herein are Cas9 molecules or Cas9 polypeptides comprising one or more deletions and optionally one or more linkers, wherein a linker is disposed between the amino acid residues that flank the deletion. Methods for identifying suitable deletions in a reference Cas9 molecule, methods for generating Cas9 molecules with a deletion and a linker, and methods for using such Cas9 molecules will be apparent to one of ordinary skill in the art upon review of this document.

A Cas9 molecule, e.g., a *S. aureus*, *S. pyogenes*, or *C. jejuni*, Cas9 molecule, having a deletion is smaller, e.g., has reduced number of amino acids, than the corresponding naturally-occurring Cas9 molecule. The smaller size of the Cas9 molecules allows increased flexibility for delivery methods, and thereby increases utility for genome-editing. A Cas9 molecule or Cas9 polypeptide can comprise one or more deletions that do not substantially affect or decrease the activity of the resultant Cas9 molecules or Cas9 polypeptides described herein. Activities that are retained in the Cas9 molecules or Cas9 polypeptides comprising a deletion as described herein include one or more of the following:

a nickase activity, i.e., the ability to cleave a single strand, e.g., the non-complementary strand or the complementary strand, of a nucleic acid molecule; a double stranded nuclease activity, i.e., the ability to cleave both strands of a double stranded nucleic acid and create a double stranded break, which in an embodiment is the presence of two nickase activities;

5 an endonuclease activity;

an exonuclease activity;

a helicase activity, i.e., the ability to unwind the helical structure of a double stranded nucleic acid;

and recognition activity of a nucleic acid molecule, e.g., a target nucleic acid or a gRNA.

10 Activity of the Cas9 molecules or Cas9 polypeptides described herein can be assessed using the activity assays described herein or in the art.

Identifying regions suitable for deletion

Suitable regions of Cas9 molecules for deletion can be identified by a variety of methods.

15 Naturally-occurring orthologous Cas9 molecules from various bacterial species, e.g., any one of those listed in **Table 7**, can be modeled onto the crystal structure of *S. pyogenes* Cas9 (Nishimasu et al., Cell, 156:935-949, 2014) to examine the level of conservation across the selected Cas9 orthologs with respect to the three-dimensional conformation of the protein. Less conserved or unconserved regions that are spatially located distant from regions involved in Cas9 activity, e.g., interface with the target nucleic acid molecule and/or gRNA, represent regions or domains are candidates for deletion without substantially affecting or decreasing Cas9 activity.

REC-Optimized Cas9 Molecules and Cas9 Polypeptides

25 A REC-optimized Cas9 molecule, or a REC-optimized Cas9 polypeptide, as that term is used herein, refers to a Cas9 molecule or Cas9 polypeptide that comprises a deletion in one or both of the REC2 domain and the RE1_{CT} domain (collectively a REC deletion), wherein the deletion comprises at least 10% of the amino acid residues in the cognate domain. A REC-optimized Cas9 molecule or Cas9 polypeptide can be an eaCas9 molecule or eaCas9 polypeptide, or an eiCas9 molecule or eiCas9 polypeptide. An exemplary REC-optimized Cas9 molecule or

30 REC-optimized Cas9 polypeptide comprises:

a) a deletion selected from:

- i) a REC2 deletion;
- ii) a REC1_{CT} deletion; or
- iii) a REC1_{SUB} deletion.

Optionally, a linker is disposed between the amino acid residues that flank the deletion.

5 In an embodiment, a Cas9 molecule or Cas9 polypeptide includes only one deletion, or only two deletions. A Cas9 molecule or Cas9 polypeptide can comprise a REC2 deletion and a REC1_{CT} deletion. A Cas9 molecule or Cas9 polypeptide can comprise a REC2 deletion and a REC1_{SUB} deletion.

Generally, the deletion will contain at least 10% of the amino acids in the cognate
10 domain, e.g., a REC2 deletion will include at least 10% of the amino acids in the REC2 domain.

A deletion can comprise: at least 10, 20, 30, 40, 50, 60, 70, 80, or 90% of the amino acid residues of its cognate domain; all of the amino acid residues of its cognate domain; an amino acid residue outside its cognate domain; a plurality of amino acid residues outside its cognate domain; the amino acid residue immediately N terminal to its cognate domain; the amino acid
15 residue immediately C terminal to its cognate domain; the amino acid residue immediately N terminal to its cognate and the amino acid residue immediately C terminal to its cognate domain; a plurality of, e.g., up to 5, 10, 15, or 20, amino acid residues N terminal to its cognate domain; a plurality of, e.g., up to 5, 10, 15, or 20, amino acid residues C terminal to its cognate domain; a plurality of, e.g., up to 5, 10, 15, or 20, amino acid residues N terminal to to its cognate domain
20 and a plurality of e.g., up to 5, 10, 15, or 20, amino acid residues C terminal to its cognate domain.

In an embodiment, a deletion does not extend beyond: its cognate domain; the N terminal amino acid residue of its cognate domain; the C terminal amino acid residue of its cognate domain.

25 A REC-optimized Cas9 molecule or REC-optimized Cas9 polypeptide can include a linker disposed between the amino acid residues that flank the deletion. Any linkers known in the art that maintain the conformation or native fold of the Cas9 molecule (thereby retaining Cas9 activity) can be used between the amino acid residues that flank a REC deletion in a REC-optimized Cas9 molecule or REC-optimized Cas9 polypeptide. Linkers for use in generating
30 recombinant proteins, e.g., multi-domain proteins, are known in the art (Chen et al., *Adv Drug Delivery Rev*, 65:1357-69, 2013).

In an embodiment, a REC-optimized Cas9 molecule or REC-optimized Cas9 polypeptide comprises an amino acid sequence that, other than any REC deletion and associated linker, has at least 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 99, or 100% homology with the amino acid sequence of a naturally occurring Cas 9, e.g., a Cas9 molecule described in **Table 7**, e.g., a *S. aureus* Cas9 molecule, a *S. pyogenes* Cas9 molecule, or a *C. jejuni* Cas9 molecule.

In an embodiment, a a REC-optimized Cas9 molecule or REC-optimized Cas9 polypeptide comprises an amino acid sequence that, other than any REC deletion and associated linker, differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or 25, amino acid residues from the amino acid sequence of a naturally occurring Cas 9, e.g., a Cas9 molecule described in **Table 7**, e.g., a *S. aureus* Cas9 molecule, a *S. pyogenes* Cas9 molecule, or a *C. jejuni* Cas9 molecule.

In an embodiment, a REC-optimized Cas9 molecule or REC-optimized Cas9 polypeptide comprises an amino acid sequence that, other than any REC deletion and associate linker, differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or 25% of the, amino acid residues from the amino acid sequence of a naturally occurring Cas 9, e.g., a Cas9 molecule described in **Table 7**, e.g., a *S. aureus* Cas9 molecule, a *S. pyogenes* Cas9 molecule, or a *C. jejuni* Cas9 molecule.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman, (1970) Adv. Appl. Math. 2:482c, by the homology alignment algorithm of Needleman and Wunsch, (1970) J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman, (1988) Proc. Nat'l. Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (see, e.g., Brent et al., (2003) Current Protocols in Molecular Biology).

Two examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., (1977) Nuc. Acids Res. 25:3389-3402; and Altschul et al., (1990) J. Mol. Biol. 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information.

The percent identity between two amino acid sequences can also be determined using the algorithm of E. Meyers and W. Miller, (1988) Comput. Appl. Biosci. 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. In addition, the percent identity between two amino acid sequences can be determined using the Needleman and Wunsch (1970) J. Mol. Biol. 48:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available at www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6.

Sequence information for exemplary REC deletions are provided for 83 naturally-occurring Cas9 orthologs in **Table 7**.

The amino acid sequences of exemplary Cas9 molecules from different bacterial species are shown below.

Table 7. Amino Acid Sequence of Cas9 Orthologs

<u>Species / Composite ID</u>	<u>Amino acid sequence</u>	REC2			REC1 _{CT}			Rec _{sub}		
		start (AA pos)	stop (AA pos)	# AA delete d (n)	start (AA pos)	stop (AA pos)	# AA delete d (n)	start (AA pos)	stop (AA pos)	# AA delete d (n)
Staphylococcus Aureus trIJ7RUA5IJ7RUA5_STAAU	SEQ ID NO: 304	126	166	41	296	352	57	296	352	57
Streptococcus Pyogenes splQ99ZW2ICAS9_STRP1	SEQ ID NO: 305	176	314	139	511	592	82	511	592	82
Campylobacter jejuni NCTC 11168 gil218563121reflYP_002344900.1	SEQ ID NO: 306	137	181	45	316	360	45	316	360	45
Bacteroides fragilis NCTC 9343 gil60683389reflYP_213533.11	SEQ ID NO: 307	148	339	192	524	617	84	524	617	84
Bifidobacterium bifidum S17 gil310286728reflYP_003937986	SEQ ID NO: 308	173	335	163	516	607	87	516	607	87
Veillonella atypica ACS-134-V-Col7a gil303229466reflZP_07316256.1	SEQ ID NO: 309	185	339	155	574	663	79	574	663	79

Lactobacillus rhamnosus GG gil258509199 reflYP_003171950.1	SEQ ID NO: 310	169	320	152	559	645	78	559	645	78
Filifactor alocis ATCC 35896 gil374307738 reflYP_005054169.1	SEQ ID NO: 311	166	314	149	508	592	76	508	592	76
Oenococcus kitaharae DSM 17330 gil366983953 gb EHN59352.1	SEQ ID NO: 312	169	317	149	555	639	80	555	639	80
Fructobacillus fructosus KCTC 3544 gil339625081 reflZP_08660870.1	SEQ ID NO: 313	168	314	147	488	571	76	488	571	76
Catenibacterium mitsuokai DSM 15897 gil224543312 reflZP_03683851.1	SEQ ID NO: 314	173	318	146	511	594	78	511	594	78
Finegoldia magna ATCC 29328 gil169823755 reflYP_001691366.1	SEQ ID NO: 315	168	313	146	452	534	77	452	534	77
Coriobacterium glomerans PW2 gil328956315 reflYP_004373648.1	SEQ ID NO: 316	175	318	144	511	592	82	511	592	82
Eubacterium yurii ATCC 43715 gil306821691 reflZP_07455288.1	SEQ ID NO: 317	169	310	142	552	633	76	552	633	76
Peptoniphilus duerdenii ATCC BAA-1640 gil304438954 reflZP_07398877.1	SEQ ID NO: 318	171	311	141	535	615	76	535	615	76
Acidaminococcus sp. D21 gil227824983 reflZP_03989815.1	SEQ ID NO: 319	167	306	140	511	591	75	511	591	75
Lactobacillus farciminis KCTC 3681 gil336394882 reflZP_08576281.1	SEQ ID NO: 320	171	310	140	542	621	85	542	621	85
Streptococcus sanguinis SK49 gil422884106 reflZP_16930555.1	SEQ ID NO: 321	185	324	140	411	490	85	411	490	85
Coprococcus catus GD-7 gil291520705 embl CBK78998.1	SEQ ID NO: 322	172	310	139	556	634	76	556	634	76
Streptococcus mutans UA159 gil24379809 reflNP_721764.1	SEQ ID NO: 323	176	314	139	392	470	84	392	470	84
Streptococcus pyogenes M1 GAS gil13622193 gb AAK33936.1	SEQ ID NO: 324	176	314	139	523	600	82	523	600	82
Streptococcus thermophilus LMD-9 gil116628213 reflYP_820832.1	SEQ ID NO: 325	176	314	139	481	558	81	481	558	81
Fusobacterium nucleatum ATCC49256 gil34762592 reflZP_00143587.1	SEQ ID NO: 326	171	308	138	537	614	76	537	614	76
Planococcus antarcticus DSM 14505 gil389815359 reflZP_10206685.1	SEQ ID NO: 327	162	299	138	538	614	94	538	614	94

Treponema denticola ATCC 35405 gi42525843 reflNP_970941.1	SEQ ID NO: 328	169	305	137	524	600	81	524	600	81
Solobacterium moorei F0204 gi320528778 reflZP_08029929.1	SEQ ID NO: 329	179	314	136	544	619	77	544	619	77
Staphylococcus pseudintermedius ED99 gi323463801 gblADX75954.1	SEQ ID NO: 330	164	299	136	531	606	92	531	606	92
Flavobacterium branchiophilum FL-15 gi347536497 reflYP_004843922.1	SEQ ID NO: 331	162	286	125	538	613	63	538	613	63
Ignavibacterium album JCM 16511 gi385811609 reflYP_005848005.1	SEQ ID NO: 332	223	329	107	357	432	90	357	432	90
Bergeyella zoohelcum ATCC 43767 gi423317190 reflZP_17295095.1	SEQ ID NO: 333	165	261	97	529	604	56	529	604	56
Nitrobacter hamburgensis X14 gi92109262 reflYP_571550.1	SEQ ID NO: 334	169	253	85	536	611	48	536	611	48
Odoribacter laneus YIT 12061 gi374384763 reflZP_09642280.1	SEQ ID NO: 335	164	242	79	535	610	63	535	610	63
Legionella pneumophila str. Paris gi54296138 reflYP_122507.1	SEQ ID NO: 336	164	239	76	402	476	67	402	476	67
Bacteroides sp. 20 3 gi301311869 reflZP_07217791.1	SEQ ID NO: 337	198	269	72	530	604	83	530	604	83
Akkermansia muciniphila ATCC BAA-835 gi187736489 reflYP_001878601	SEQ ID NO: 338	136	202	67	348	418	62	348	418	62
Prevotella sp. C561 gi345885718 reflZP_08837074.1	SEQ ID NO: 339	184	250	67	357	425	78	357	425	78
Wolinella succinogenes DSM 1740 gi34557932 reflNP_907747.1	SEQ ID NO: 340	157	218	36	401	468	60	401	468	60
Alicyclobacillus hesperidum URH17-3-68 gi403744858 reflZP_10953934.1	SEQ ID NO: 341	142	196	55	416	482	61	416	482	61
Caenispirillum salinarum AK4 gi427429481 reflZP_18919511.1	SEQ ID NO: 342	161	214	54	330	393	68	330	393	68
Eubacterium rectale ATCC 33656 gi238924075 reflYP_002937591.1	SEQ ID NO: 343	133	185	53	322	384	60	322	384	60
Mycoplasma synoviae 53 gi71894592 reflYP_278700.1	SEQ ID NO: 344	187	239	53	319	381	80	319	381	80
Porphyromonas sp. oral taxon 279 str. F0450 gi402847315 reflZP_10895610.1	SEQ ID NO: 345	150	202	53	309	371	60	309	371	60

Streptococcus thermophilus LMD-9 gil116627542reflYP_820161.11	SEQ ID NO: 346	127	178	139	424	486	81	424	486	81
Roseburia inulinivorans DSM 16841 gil225377804reflZP_03755025.1	SEQ ID NO: 347	154	204	51	318	380	69	318	380	69
Methylosinus trichosporium OB3b gil296446027reflZP_06887976.1	SEQ ID NO: 348	144	193	50	426	488	64	426	488	64
Ruminococcus albus 8 gil325677756reflZP_08157403.1	SEQ ID NO: 349	139	187	49	351	412	55	351	412	55
Bifidobacterium longum DJO10A gil189440764reflYP_001955845	SEQ ID NO: 350	183	230	48	370	431	44	370	431	44
Enterococcus faecalis TX0012 gil315149830lgbIEFT93846.11	SEQ ID NO: 351	123	170	48	327	387	60	327	387	60
Mycoplasma mobile 163K gil47458868reflYP_015730.11	SEQ ID NO: 352	179	226	48	314	374	79	314	374	79
Actinomyces coleocanis DSM 15436 gil227494853reflZP_03925169.1	SEQ ID NO: 353	147	193	47	358	418	40	358	418	40
Dinoroseobacter shibae DFL 12 gil159042956reflYP_001531750 .1	SEQ ID NO: 354	138	184	47	338	398	48	338	398	48
Actinomyces sp. oral taxon 180 str. F0310 gil315605738reflZP_07880770.1	SEQ ID NO: 355	183	228	46	349	409	40	349	409	40
Alcanivorax sp. W11-5 gil407803669reflZP_11150502.1	SEQ ID NO: 356	139	183	45	344	404	61	344	404	61
Aminomonas paucivorans DSM 12260 gil312879015reflZP_07738815.1	SEQ ID NO: 357	134	178	45	341	401	63	341	401	63
Mycoplasma canis PG 14 gil384393286lgbIEIE39736.11	SEQ ID NO: 358	139	183	45	319	379	76	319	379	76
Lactobacillus coryniformis KCTC 3535 gil336393381reflZP_08574780.1	SEQ ID NO: 359	141	184	44	328	387	61	328	387	61
Elusimicrobium minutum Pei191 gil187250660reflYP_001875142 .1	SEQ ID NO: 360	177	219	43	322	381	47	322	381	47
Neisseria meningitidis Z2491 gil218767588reflYP_002342100 .1	SEQ ID NO: 361	147	189	43	360	419	61	360	419	61
Pasteurella multocida str. Pm70 gil15602992reflNP_246064.11	SEQ ID NO: 362	139	181	43	319	378	61	319	378	61
Rhodovulum sp. PH10 gil402849997reflZP_10898214.1	SEQ ID NO: 363	141	183	43	319	378	48	319	378	48
Eubacterium dolichum DSM 3991	SEQ ID NO: 364	131	172	42	303	361	59	303	361	59

gil160915782 ref ZP_02077990.1										
Nitratifactor salsuginis DSM 16511 gil319957206 ref YP_004168469.1	SEQ ID NO: 365	143	184	42	347	404	61	347	404	61
Rhodospirillum rubrum ATCC 11170 gil83591793 ref YP_425545.1	SEQ ID NO: 366	139	180	42	314	371	55	314	371	55
Clostridium cellulolyticum H10 gil220930482 ref YP_002507391.1	SEQ ID NO: 367	137	176	40	320	376	61	320	376	61
Helicobacter mustelae 12198 gil291276265 ref YP_003516037.1	SEQ ID NO: 368	148	187	40	298	354	48	298	354	48
Ilyobacter polytropus DSM 2926 gil310780384 ref YP_003968716.1	SEQ ID NO: 369	134	173	40	462	517	63	462	517	63
Sphaerochaeta globus str. Buddy gil325972003 ref YP_004248194.1	SEQ ID NO: 370	163	202	40	335	389	45	335	389	45
Staphylococcus lugdunensis M23590 gil315659848 ref ZP_07912707.1	SEQ ID NO: 371	128	167	40	337	391	57	337	391	57
Treponema sp. JC4 gil384109266 ref ZP_10010146.1	SEQ ID NO: 372	144	183	40	328	382	63	328	382	63
uncultured delta proteobacterium HF0070 07E19 gil297182908 gbl ADI19058.1	SEQ ID NO: 373	154	193	40	313	365	55	313	365	55
Alicyclophilus denitrificans K601 gil330822845 ref YP_004386148.1	SEQ ID NO: 374	140	178	39	317	366	48	317	366	48
Azospirillum sp. B510 gil288957741 ref YP_003448082.1	SEQ ID NO: 375	205	243	39	342	389	46	342	389	46
Bradyrhizobium sp. BTAi1 gil148255343 ref YP_001239928.1	SEQ ID NO: 376	143	181	39	323	370	48	323	370	48
Parvibaculum lavamentivorans DS-1 gil154250555 ref YP_001411379.1	SEQ ID NO: 377	138	176	39	327	374	58	327	374	58
Prevotella timonensis CRIS 5C-B1 gil282880052 ref ZP_06288774.1	SEQ ID NO: 378	170	208	39	328	375	61	328	375	61
Bacillus smithii 7 3 47FAA gil365156657 ref ZP_09352959.1	SEQ ID NO: 379	134	171	38	401	448	63	401	448	63
Cand. Puniceispirillum marinum IMCC1322 gil294086111 ref YP_003552871.1	SEQ ID NO: 380	135	172	38	344	391	53	344	391	53
Barnesiella intestinihominis YIT 11860 gil404487228 ref ZP_11022414.1	SEQ ID NO: 381	140	176	37	371	417	60	371	417	60
Ralstonia syzygii R24	SEQ ID NO:	140	176	37	395	440	50	395	440	50

gil344171927 embl CCA84553.1	382									
Wolinella succinogenes DSM 1740 gil34557790 ref NP_907605.1	SEQ ID NO: 383	145	180	36	348	392	60	348	392	60
Mycoplasma gallisepticum str. F gil284931710 gbl ADC31648.1	SEQ ID NO: 384	144	177	34	373	416	71	373	416	71
Acidothermus cellulolyticus 11B gil117929158 ref YP_873709.1	SEQ ID NO: 385	150	182	33	341	380	58	341	380	58
Mycoplasma ovipneumoniae SC01 gil363542550 ref ZP_09312133.1	SEQ ID NO: 386	156	184	29	381	420	62	381	420	62

Table 8. Amino Acid Sequence of Cas9 Core Domains

Strain Name	Cas9 Start (AA pos)	Cas9 Stop (AA pos)
	Start and Stop numbers refer to the sequence in Table 7	
Staphylococcus Aureus	1	772
Streptococcus Pyogenes	1	1099
Campulobacter Jejuni	1	741

Table 9. Identified PAM sequences and corresponding RKR motifs.

Strain Name	PAM sequence (NA)	RKR motif (AA)
Streptococcus pyogenes	NGG	RKR
Streptococcus mutans	NGG	RKR
Streptococcus thermophilus A	NGGNG	RYR
Treponema denticola	NAAAAAN	VAK
Streptococcus thermophilus B	NNAAAAW	IYK
Campylobacter jejuni	NNNNACA	NLK
Pasteurella multocida	GNNNCNNA	KDG
Neisseria meningitidis	NNNNGATT or	IGK
Staphylococcus aureus	NNGRRV (R = A or G; V = A, G or C) NNGRRT (R = A or G)	NDK

PI domains are provided in **Tables 10** and **11**.

Table 10. Altered PI Domains

Strain Name	PI Start (AA pos)	PI Stop (AA pos)	Length of PI (AA)	RKR motif (AA)
	Start and Stop numbers refer to the sequences in Table 100			
Alicyclophilus denitrificans K601	837	1029	193	--Y
Campylobacter jejuni NCTC 11168	741	984	244	-NG
Helicobacter mustelae 12198	771	1024	254	-NQ

5 Table 11. Other Altered PI Domains

Strain Name	PI Start (AA pos)	PI Stop (AA pos)	Length of PI (AA)	RKR motif (AA)
	Start and Stop numbers refer to the sequences in Table 7			
Akkermansia muciniphila ATCC BAA-835	871	1101	231	ALK
Ralstonia syzygii R24	821	1062	242	APY
Cand. Puniceispirillum marinum IMCC1322	815	1035	221	AYK
Fructobacillus fructosus KCTC 3544	1074	1323	250	DGN
Eubacterium yurii ATCC 43715	1107	1391	285	DGY
Eubacterium dolichum DSM 3991	779	1096	318	DKK
Dinoroseobacter shibae DFL 12	851	1079	229	DPI
Clostridium cellulolyticum H10	767	1021	255	EGK
Pasteurella multocida str. Pm70	815	1056	242	ENN
Mycoplasma canis PG 14	907	1233	327	EPK
Porphyromonas sp. oral taxon 279 str. F0450	935	1197	263	EPT
Filifactor alocis ATCC 35896	1094	1365	272	EVD
Aminomonas paucivorans DSM 12260	801	1052	252	EVY
Wolinella succinogenes DSM 1740	1034	1409	376	EYK
Oenococcus kitaharae DSM 17330	1119	1389	271	GAL
Coriobacterium glomerans PW2	1126	1384	259	GDR
Peptoniphilus duerdenii ATCC BAA-1640	1091	1364	274	GDS
Bifidobacterium bifidum S17	1138	1420	283	GGL
Alicyclobacillus hesperidum URH17-3-68	876	1146	271	GGR
Roseburia inulinivorans DSM 16841	895	1152	258	GGT
Actinomyces coleocanis DSM 15436	843	1105	263	GKK
Odoribacter laneus YIT 12061	1103	1498	396	GKV

Coprococcus catus GD-7	1063	1338	276	GNQ
Enterococcus faecalis TX0012	829	1150	322	GRK
Bacillus smithii 7 3 47FAA	809	1088	280	GSK
Legionella pneumophila str. Paris	1021	1372	352	GTM
Bacteroides fragilis NCTC 9343	1140	1436	297	IPV
Mycoplasma ovipneumoniae SC01	923	1265	343	IRI
Actinomyces sp. oral taxon 180 str. F0310	895	1181	287	KEK
Treponema sp. JC4	832	1062	231	KIS
Fusobacteriumnucleatum ATCC49256	1073	1374	302	KKV
Lactobacillus farciminis KCTC 3681	1101	1356	256	KKV
Nitratifractor salsuginis DSM 16511	840	1132	293	KMR
Lactobacillus coryniformis KCTC 3535	850	1119	270	KNK
Mycoplasma mobile 163K	916	1236	321	KNY
Flavobacterium branchiophilum FL-15	1182	1473	292	KQK
Prevotella timonensis CRIS 5C-B1	957	1218	262	KQQ
Methylosinus trichosporium OB3b	830	1082	253	KRP
Prevotella sp. C561	1099	1424	326	KRY
Mycoplasma gallisepticum str. F	911	1269	359	KTA
Lactobacillus rhamnosus GG	1077	1363	287	KYG
Wolinella succinogenes DSM 1740	811	1059	249	LPN
Streptococcus thermophilus LMD-9	1099	1388	290	MLA
Treponema denticola ATCC 35405	1092	1395	304	NDS
Bergeyella zoohelcum ATCC 43767	1098	1415	318	NEK
Veillonella atypica ACS-134-V-Col7a	1107	1398	292	NGF
Neisseria meningitidis Z2491	835	1082	248	NHN
Ignavibacterium album JCM 16511	1296	1688	393	NKK
Ruminococcus albus 8	853	1156	304	NNF
Streptococcus thermophilus LMD-9	811	1121	311	NNK
Barnesiella intestinihominis YIT 11860	871	1153	283	NPV
Azospirillum sp. B510	911	1168	258	PFH
Rhodospirillum rubrum ATCC 11170	863	1173	311	PRG
Planococcus antarcticus DSM 14505	1087	1333	247	PYY
Staphylococcus pseudintermedius ED99	1073	1334	262	QIV
Alcanivorax sp. W11-5	843	1113	271	RIE
Bradyrhizobium sp. BTAi1	811	1064	254	RIY
Streptococcus pyogenes M1 GAS	1099	1368	270	RKR
Streptococcus mutans UA159	1078	1345	268	RKR
Streptococcus Pyogenes	1099	1368	270	RKR
Bacteroides sp. 20 3	1147	1517	371	RNI

<i>S. aureus</i>	772	1053	282	RNK
<i>Solobacterium moorei</i> F0204	1062	1327	266	RSG
<i>Finegoldia magna</i> ATCC 29328	1081	1348	268	RTE
uncultured delta proteobacterium HF0070 07E19	770	1011	242	SGG
<i>Acidaminococcus</i> sp. D21	1064	1358	295	SIG
<i>Eubacterium rectale</i> ATCC 33656	824	1114	291	SKK
<i>Caenispirillum salinarum</i> AK4	1048	1442	395	SLV
<i>Acidothermus cellulolyticus</i> 11B	830	1138	309	SPS
<i>Catenibacterium mitsuokai</i> DSM 15897	1068	1329	262	SPT
<i>Parvibaculum lavamentivorans</i> DS-1	827	1037	211	TGN
<i>Staphylococcus lugdunensis</i> M23590	772	1054	283	TKK
<i>Streptococcus sanguinis</i> SK49	1123	1421	299	TRM
<i>Elusimicrobium minutum</i> Pei191	910	1195	286	TTG
<i>Nitrobacter hamburgensis</i> X14	914	1166	253	VAY
<i>Mycoplasma synoviae</i> 53	991	1314	324	VGf
<i>Sphaerochaeta globus</i> str. Buddy	877	1179	303	VKG
<i>Ilyobacter polytropus</i> DSM 2926	837	1092	256	VNG
<i>Rhodovulum</i> sp. PH10	821	1059	239	VPY
<i>Bifidobacterium longum</i> DJO10A	904	1187	284	VRK

Amino acid sequences described in Table 7 (in order of appearance):

SEQ ID NO: 304

5 MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRI
 QRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDT
 GNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSYVKEAKQLLKVKQKAYHQ
 LDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLY
 10 NALNDLNNLVI TRDENEKLEYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGK
 PEFTNLKVYHDIKDITARKEI IENAELLDQIAKILTIYQSSEDIQEELTNLNSLTQEEIEQIS
 NLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSP
 VVKRSFIQSIVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKRNRQTNERIEEII RTT
 GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII PRSVSFDNSFNKVLVK
 QEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRI SKTKKEYLLEERDINRF SVQKD
 FINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAD
 15 ALIIANADFIKFEWKKLDKAKKVMENQMFEEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKD
 YKYSHRVDKKNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHH
 DPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKNNGPVIKKIKYYGNKLNAHLDTDD
 YPNSRNKVVKLSLKPYPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA
 EFIA SFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKT
 20 QSIKKYSTDILGNLYEVKSKKHPQIIKKG

SEQ ID NO: 305

MDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFV LGNTDRHSIKKNLIGALLFDSGETAEATRL
 KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAY

HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTY
 NQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
 DLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
 MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF IKPILEKMD
 5 GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
 PYYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHS
 LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTF
 10 KEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQ
 TTQKGQKNSRERMKRIEEGIELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
 LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
 FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYKREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIK
 15 SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKG
 KSKKLKSVKELLGITIMERSSSFENPIDFLEAKGYKEVKKDLIIKLPKYSLEFELNGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLEVEQHKHYLDEIIIEQISEFSKRV
 ILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
 20 ATLIHQ SITGLYETRIDLSQLGGD

SEQ ID NO: 306

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKAR
 LNHLKHLIANEFKLNIEDYQSFDESLAKAYKGLISPYELRFRALNELLSKQDFARVILHIAKR
 25 RGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYE
 RCIAQSFLKDELKLIFKKQREFGFSFSKKFEEVLSVAFYKRALKDFSHLVGNCSSFTDEKRAP
 KNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDL NALLNEVLKNGTLTYKQTKKLLGLSDDYE
 FKGEKGTIFYIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDS
 LSKLEFKDHLNISFKALKLVTPMLLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVT
 30 NPVVLRAIKEYRKVLNALLKKYKGVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELEC
 EKLGLKINSKNILKLRLFKKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRFSFDDSYMNVK
 VFTKQKQEKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDT
 RYIARLVNLNYTKDYLDLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNH
 LHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLD
 35 KIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNKIVKNGDMFR
 VDFIKHKKTNKFYAVPIYTMDFAKVLPNKAVARSKKGEIKDWILMDENYEFCSLYKDSLIL
 QTKDMQEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVF
 EKYIVSALGEVTKAEFRQREDFKK

40 SEQ ID NO: 307

MKRILGLDLGTNSIGWALVNEAENKDERSSIVKLGVRVNPLTVDELTFEKGKSITTNADRTLK
 RGMRRNLQRYKLRRETLTEVLKEHKLITEDTILSENGNRTTFETYRLRAKAVTEEISLEEFARV
 LLMINKKRGYKSSRKAKGVEEGTLIDGMDIARELYNNNLTPGELCLQLLDAGKKFLPDFYRSDL
 QNELDRIWEKQKEYYPEILTDVLKEELRGKKRDAVWAICAKYFVWKENYTEWNKEKGKTEQQR
 45 EHKLEGIYSKRKRDEAKRENLQWRVNLKEKLSLEQLVIVFQEMNTQINNSSGYLGAI SDRSKE
 LYFNKQTVGQYQMEMLDKNPNASLRNMVFYRQDYLDENMLWEKQAVYHKELTEELKKEIRDII
 IFYQRRLKSQKGLIGFCEFESRQIEVDIDGKKKIKTVGNRVISRSSPLFQEFKIWQILNNIEVT

VVGKKRRRKLKENYSALFEELNDAEQLELNGSRRLCQEEKELLAQELFIRDKMTKSEVLKLLF
 DNPQELDLNFKTIDGNKTGYALFQAYSKMIEMSGHEPVDFKKPVEKVVEYIKAVFDLLNWNTDI
 LGFNSNEELDNQPYKLVHLLYSFEGDNTPTGNRLIQKMTELYGFEKEYATILANVSFQDDYG
 SLSAKAIHKILPHLKEGNRYDVACVYAGYRHSESSLTREEIANKVLKDRLMLLPKNSLHNPVVE
 5 KILNQMVNVINVIIDIYGPDEIRVELARELKKNAKEREEELTKSIAQTTKAHEEYKTLQTEFG
 LTNVSRDILRYKLYKELESCGYKTLYSNTYISREKLFKSKEFDIEHIIPQARLFDDSF SNKTLE
 ARSVNIEKGNKTAYDFVKEKFGESGADNSLEHYLNNIEDLFSKGKISKTKYNKLKMAEQDIPDG
 FIERDLRNTQYIAKKALSMLNEISHRVVATSGSVTDKLREDWQLIDVMKELNWEKYKALGLVEY
 FEDRDGRQIGRIKDWTKRNDHRHHAMDALTVAF TKDVF IQYFNNKNASLDPNANEHAIKNKYFQ
 10 NGRAIAPMPLREFRAEAKKHLENTLISIKAKNKVITGNINKTRKKGGVNKNMQQTPRGQLHLET
 IYSGSGKQYLTKEEKVNASFDMRKIGTVSKSAYRDALLKRLYENDNDPKKAFAGKNSLKDQPIWL
 DKEQMRKVPEKVKIVTLEAIYTIRKEISPDCLKVDKVIDVGVRKILIDRLNEYGNDAKKAFFSNLD
 KNPIWLNKEKGISIKRVTISGISNAQSLHVKKDKDGKPILDENGRNIPVDFVNTGNNHHVAVYY
 RPVIDKRGQLVVDEAGNPKEYELEEVVVSFFEAVTRANLGLPIIDKDYKTTEGWQFLFSMKQNEY
 15 FVFPNEKTGFNPKEIDLDDVENYGLISPNLFRVQKFSKKNYVFRHHLETTIKDTSSILRGITWI
 DFRSSKGLDTIVKVRVNHIGQIVSVGEY

SEQ ID NO: 308

MSRKNYVDDYAI SLDIGNASVGWSAFTPNYRLVRAKGHELIGVRLFPADTAESRRMARTTRRR
 20 YSRRRWRLRLLDALFDQALSEIDPSFLARRKYSWVHPDDENNADCWYGSVLFDSNEQDKRFYEK
 YPTIYHLRKALMEDDSQHDIREIYLAIHMMVKYRGNFLVEGTLESSNAFKEDELKLLGRITRY
 EMSEGEQNSDIEQDDENKL VAPANGQLADALCATRGSRSRMRVDNALEALS AVNDLSREQRAIVK
 AIFAGLEGNKLDLAKIFVSKEFSSENKKILGIYFNKSDYEEKCVQIVDSGLLDDEEREFLDRMQ
 GQYNAIALKQLLGRSTSVSDSKCASDYDAHRANWNLIKQLRRTKENEKDINENY GILVGWKIDSG
 25 QRKSVRGESAYENMRKKANVFFKKMIETSDLSETDKNRLIHDIEEDKLFPIQRDSNGVIPHQL
 HQNELKQIIKKQGKYYPFLDAFEKDGKQINKIEGLLTFRVPYFVGPLVVPEDLQKSDNSENHW
 MVRKKKG EITPWNFDEMVDKDASGRKFIERLVGTDSYLLGEPTLPKNSLLYQ EYEVNLNENVR
 LSVRTGNHWNDRMRRLGREEKTLLCQRLFMKGQTVTKRTAENLLRKEYGRTYELSGLSDESKF
 TSSLSTYGKMCRIFGKEYVNEHRDLMEKIVELQTVFEDKETLLHQLRQLEGISEADCALLVNTH
 30 YTGWGRLSRKLTTKAGECKISDDFAPRKHSIEIMRAEDRNLMEIITDKQLGFSWDIEWENLG
 AENGSSLMEVVDDL RVSPKVKRGIIQSIRLIDDISKAVGKRPSRIFLELADDIQPSGRTISRKS
 RLQDLYRNANLGKEFKGIADDELNACSDKDLQDDRLFLYYTQLGKDMYTGEELDLRLSSAYDID
 HIIPQAVTQNSIDNRVLVARAENARKTDSFTYMPQIADRMNFWQILLDNGLISRVKFERLTR
 QNEFSEREKERFVQ RSLVETRQIMKNVATLMRQRYGNSAAVIGLNAELTKEMHRYLGF SHKNRD
 35 INDYHHAQDALCVGIAGQFAANRGFFADGEVSDGAQNSYNQYLRDYLRGYREKLSAEDRKQGRA
 FGFIVGSMRSQDEQKRVNPRTEGEVWSEEDKDYL RKVMNYRKMLVTQKVGGDFGALYDETRYAA
 TDPKGIKIPFDGAKQDTSLYGGFSSAKPAYAVLIESKGKTRLVNVTMQEYSLLDGRPSDDEL R
 KVLAKKKSEYAKANILLRHVPKMQLIRYGGGLMVIKSAGELNNAQQLWLPYEEYCYFDDLSQ GK
 GSLEKDDLKLLDSILGSVQCLYPWHRFTEEELADLHVAFDKLPEDKKNVITGIVSALHADAK
 40 TANLSIVGMTG SWRRMNNKSGYTF SDEDEFIFQSPSGLFEKRVTVGELKRKAKKEVNSKYRTNE
 KRLPTLSGASQP

SEQ ID NO: 309

METQTSNQLITSHLKDYPKQDYFVGLDIGTNSVGWAVTNTSYELLKFHSHKMWGSRLFEEGES A
 45 VTRRGFRSMRRRLERRKLRLKLLLELFADAMAQVDSTFFIRLHESKYHYEDKTTGHSSKHILFI
 DEDYTDQDYFTEYPTIYHLRKDLMEGTDDIRKLF LAVHHILKYRGNFLYEGATFNSNAFTFED
 VLKQALVNITFNCFDTNSAIISSISNILMESGKTKSDKAKAIERLVDTYTVFDEVNTPDKPQKEQ

VKEDKKTLLKAFANLVLGLSANLIDLFSGVEDIDDDLKKLQIVGDTYDEKRDELAKVWGDEIHI
DDCKSVYDAIILMSIKEPGLTISQSKVKAFDKHKEDLVILKSLLKLDRNVYNEMFKSDKKGLHN
YVHYIKQGRTEETSCSREDFYKYTKKIVEGLADSKDKEYILNEIELQTLLPLQRIKDNGVIPYQ
LHLEELKVILDKCGPKFPFLHTVSDGFSVTEKLIKMLEFRIPYYVGPLNTHHNIDNGGFSWAVR
5 KQAGRVTWPWFEEKIDREKSAAAFIKNLTNKCTYLFGEDVLPKSSLLYSEFMLLNELNVRIDG
KALAQGVKQHLIDSIFKQDHHKMTKNRIELFLKDNNTYITKKHKPEITGLDGEIKNDLTSYRDMV
RILGNNFDVSMAEDIITDITIFGESKKMLRQTLRNKFGSQLNDETICKLSKLRYRDWGRLSKKL
LKGIDGCDKAGNGAPKTIIELMRNDSYNLMEILGDKFSFMECIEEENAKLAQGQVVNPHDIIDE
LALSPAVKRAVWQALRIVDEVAHIKKALPSRIFVEVARTNKSEKKKKDSRQKRLSDLYSAIKKD
10 DVLQSGLQDKEFGALKSGLANYDDAALRSKKLYLYYTQMGRCAYTGNIIDLNQLNTDNYDIDHI
YPRSLTKDDSFNDNLVLCERTANAKKSDIYPIDNRIQTKQKPFWAFLLKHQGLISERKYERLTRIA
PLTADDLSGFIARQLVETNQSVMKATTTLLRRLYPDIDVVFVKAENVSDFRHNNNF IKVRSNLHH
HHAKDAYLNIIVGVNVYHEKFTRNFRLLFFKKNGANRTYNLAKMFNYDVICTNAQDGKAWDVKTSM
NTVKKMMASNDVRVTRRLLEQSGALADATIYKASVAAKAKDGAYIGMKTYSVFADVTKYGGMT
15 KIKNAYSIIVQYTGKKGEEIKEIVPLPIYLINRNATDIELIDYVKSVIPKAKDISIKYRKL CIN
QLVKVNGFYYYLGGKTNDKIYIDNAIELVPHDIATYIKLLDKYDLLRKENKTLKASSITTSIY
NINTSTVVS LNKV GIDVFDYFMSKLRTPLYMKMKGNKVDELSSSTGRSKFIKMTLEEQSIYLLLEV
LNLLTNSKTTFDVKPLGITGSRSTIGVKIHNLD EFKIINESITGLYSNEVTIV

20 SEQ ID NO: 310
MTKLNQPYGIGLDIGSNSIGFAVVDANSHLLRLKGETAIGARLFREGQSAADRRGSRTTRRRLS
RTRWRLSFLRDF FAPHITKIDPDFL RQKYSEISPDKDRFKYEKRLFNDRD AEFYEDYPSMY
HLRLHLMTHTHKADPREIFLAIHHILKSRGHFLTPGA AKDFNTDKVDLEDIFPALTEAYA QVYP
DLELTFDLAKADDFKAKLLDEQATPSDTQKALVNLLSSDGEKEIVKKRKQVLTEFAKAITGLK
25 TKFNALGTEVDEADASNWQFSMGQLDDKWSNIETSM TDQGTEIFEQIQELYRARLLNGIVPAG
MSLSQAKVADYGQHKEDLELFKTYLKKLNDHELAKTIRGLYDRYINGDDAKPFLREDFVKALTK
EVT AHPNEVSEQLLNRMGQANFMLKQRTKANGAIP IQLQQRELDQIIANQSKYYDWLAAPNPVE
AHRWKMPYQLDEL LNFIPIYYVGPLITPKQQAESGENVF AWMVRKDPSGNI TPYNFDEKVDREA
SANTFIQRMKTTDTYLIGEDVLPKQSLLYQKYEV LNELNNVRINNECLGTDQKQRLIREVFERH
30 SSVTIKQVADNLVAHGDFARRPEIRGLADEKRF LSSLSYHQLKEILHEAIDDP TKLLDIENII
TWSTVFEDHTIFETKLAEIEWLDPKKINELSGIRYRGWGQFSRKLLDGLKLNGHTVIQELMLS
NHNLMQILADETLKETMTELNQDKLKTDDIEDVINDAYTSPSNKKALRQVLRVVEDIKHAANGQ
DPSWLF IETADGTGTAGKRTQSRQKQIQTVYANAAQELIDSAVRGELEDKIADKASFTDRLVLY
FMQGGRIYTGA PLNIDQLSHYDIDHILPQSLIKDDSLDNRVLVNATINREKNNVFASTLFAGK
35 MKATWRKWHEAGLISGRKLRNLMLRPDEIDKFAGGFVARQLVETRQIIKLTEQIAAAQYPNTKI
IAVKAGLSHQLREELDFPKNRDVNHYHHAFDAFLAARIGTYLLKRYPKLAPFFTYGEFAKVDVK
KREFNF IGALTHAKKNIIAKDTGEIVWDKERDIRELDRIYNFKRMLITHEVYFETADLFKQTI
YAAKDSKERGGSKQLIPKKQGYPTQVYGGYTQESGSYNALVRVAEADTTAYQVIKISAQNASKI
ASANLKSREK GKQLLNEIVVKQLAKRRKNWKPSANSFKIVIPRFGMGT LFNQAKYGLFMVNSDT
40 YYRNYQELWLSRENQKLLKKLFSIKYEKTQMNHDALQVYKAIIDQVEKFFKLYDINQFRAKLS
AIERFEKLPINTDGNKIGKTETLRQILIGLQANGTRSNVKNLGIKTDLGLLQVGS GIKLKDQTQ
IVYQSPSGLFKRRIP LADL

SEQ ID NO: 311
45 MTKEYYLGLDVGTNSVGWAVTDSQYNLCKFKKKDMWGIRLFESANTAKDRRLQRGNRRRLERKK
QRIDLLQEIFSPEICKIDPTFFIRLNESRLHLEDKSNDFKYPLFIEKDYSIDIEYYKEFPTIFHL
RKHLIESEEKQDIRLIYLALHNIKTRGHFLIDGDLQSAKQLRPILD TFLLSLQEEQNLSVSL S

ENQKDEYEEILKNRSIAKSEKVKKLKNLFEISDELEKEEKKAQSAVIENFCKFIVGNKGDVCKF
LRVSKEELEIDSFSFSEGGYEDDIVKNLEEKVPEKVYLFEQMKAMYDWNILVDILETEEYISFA
KVKQYEHKHTNLRLLRDIILKYCTKDEYNRMFNDEKEAGSYTAYVGKLKKNKKYWIEKKRNPE
EFYKSLGKLLDKIEPLKEDLEVLTMIEECKNHTLLPIQKNKDNGVIPHQVHEVELKKILENAK
5 KYYSFLTETDKDGYSVVQKIESIFRFRIPYYVGPLSTRHQEKGSNVWMVRKPGREDRIYPWNME
EIIDFEKSNENFITRMTNKTCTYLIGEDVLPKHSLLYSKYMVLNELNNVKVRGKKLPTSLKQKVF
EDLFENKSKVTGKNLLEYLQIQDKDIQIDDLSGFDKDFKTSLSYLDFFKKQIFGEEIEKESIQN
MIEDIIKWITIIYGNDKEMLRVIRANYSNQLTEEQMKKITGFQYSGWGNFSKMFLKGISGSDVS
TGETFDIITAMWETDNNLMQILSKKFTFMDNVEDFNSGKVGKIDKITDYDSTVKEMFLSPENKRA
10 VWQTIQVAEEIKKVMGCEPKKIFIEMARGGEKVKKRTKSRAQLLELYAACEEDCRELIKEIED
RDERDFNSMKLFLYYTQFGKCMYSGDDIDINELIRGNSKWDRDHIYPQSKIKDDSIDNLVLVNK
TYNAKKSNELLSEDIQKKMHSFWLSLLNKKLITKSKYDRLTRKGDFTDEELSGFIARQLVETRQ
STKAIADIFKQIYSSEVVYVKSSLSVDFRKKPLNYLKSRRVNDYHHAKDAYLNIIVGVNVYNKKF
TSNPIQWMKKNRDTNYSLNKVF EHDVVINGEVIWEKCTYHEDTNTYDGGTLDRIKIVERDNIL
15 YTEYAYCEKGELFNATIONKNGNSTVSLKKGLDVKKYGGYFSANTS YFSLIEFEDKKGDRARHI
IGVPIYIANMLEHSPSAFLEYCEQKGYQNVIRILVEKIKKNSLLIINGYPLRIRGENEVDTSFKR
AIQLKLDQKNYELVRNIEKFLEKYVEKKGNYPIDENRDHITHEKMNQLYEVLLSKMKKFNKKGM
ADPSDRIEKS KPFIKLEDLIDKINVINKMLNLLRCDNDTKADLSLIELPKNAGSFVVKKNTIG
KSKIILVNQSVTGLYENRREL

SEQ ID NO: 312

MARDYSVGLDIGTSSVGWAAIDNKYHLIRAKSKNLIGVRLFDSAVTAEKRRGYRTTRRRLSRRH
WRLRLLNDIFAGPLTDFGDENFLARLKYSWVHPQDQSNQAHFAAGLLFDSKEQDKDFYRKYPTI
YHLRLALMNDQKHDLREVYLAIHHLVKYRGHFLIEGDVKADSAFDVHTFADAIQRYAESNNSD
25 ENLLGKIDEKKLSAALTDKHGSKSQRAETAETAFDIIDLQSKKQIQAILKSVVGNQANLMAIFG
LDSSAISKDEQKNYKFSFDDADIDEKIIDSEALLSDTEFEFLCDLKAADFGLTLKMLLGDDKT
SAAMVRRFNEHQDWEYIKSHIRNAKNAGNGLYEKSKKFDGINAAYLALQSDNEDDRKKAKKIF
QDEISSADIPDDVKADFLKKIDDDQFLPIQRTKNNGTIPHQLHRNELEQIIIEKQGIYYPFLKDT
YQENSHELNKITALINFRVPYYVGPLVEEEQKIADDGKNIPDPTNHWMVRKSNDTITPWNLSQV
30 VDLKSGRRFIERLTGTDTYLIGEPTLPKNSLLYQKFDVLQELNNIRVSGRRLDIRAKQDAFEH
LFKVQKTVSATNLKDFLVQAGYISEDQIEGLADVNGKNFNALTTYNLVSVLGREFVENPSN
EELLEIEITELQTVFEDKKVLRRLDQLDGLSDHNREKLSRKHYTGWGRISKKLLTTKIVQNADK
IDNQTFDVPRMNQSIIDTLYNTKMNLMEIINNAEDDFGVRAWIDKQNTTDGDEQDVYSLIDELA
GPKEIKRGIVQSFRILDDITKAVGYAPKRVYLEFARKTQESHLTNSRKNQLSTLLKNAGLSELV
35 TQVSQYDAAALQNDRLYLFLQQGKDMYSGEKLNLNLSNYDIDHIIPQAYTKDNSLDNRVLVS
NITNRRKSDSSNYLPALIDKMRPFWSVLQGLLSKHKFANLTRTRDFDDMEKERFIARSLVET
RQIIKNVASLIDSHFGGETKAVAIRSSLTADMRRYVDIPKNRDINDYHHAFDALLFSTVGQYTE
NSGLMKGQLSDSAGNQYNRYIKEWIHAARLNAQSQRVNPFVGVGSMRNAAPGKLNPEIT
PEENADWSIADLDYLHKVMNFRKITVTRRLKDQKQLYDESRYPVLHDAKSKASINFDKHKPV
40 DLYGGFSSAKPAYAALIKFKNKFRLVNVLQWQTYSDKNSDYILEQIRGKYPKAEMVLSHIPYG
QLVKKDGALVTISSATELHNFEQLWLPLADYKLINTLLKTKEDNLVDILHNRLDLPMTIESAF
YKAFDSILSFAFNRYALHQNALVKLQAHRRDFNALNYEDKQQTLEIRILDALHASPASSDLKKIN
LSSGFGRFLFSPSHFTLADTDEFIFQSVTGLFSTQKTVAQLYQETK

SEQ ID NO: 313

MVYDVGLDIGTGSVGWVALDENGKLARAKGKNLVGVRLFDTAQTAADRRGFRTTRRRLSRRKWR
LRLLELFSAEINEIDSSFFQRLKYSYVHPKDEENKAHYGGYLFPTTEEETKKFHRSYPTIYHL

RQELMAQPNKRFDIRIYLAIHHLVKYRGHFLSSQEKITIGSTYNPEDLANAIEVYADEKGLSW
 ELNNPEQLTEIISGEAGYGLNKSMADEALKLFEFDNNQDKVAIKTLLAGLTGNQIDFAKLFGK
 DISDKDEAKLWKLKLDDEALEEKSQTILSQLTDEEIELFHAVVQAYDGFVLIGLLNGADSVSAA
 MVQLYDQHREDRKLLKSLAQKAGLKHKRFSEIYEQLALATDEATIKNGISTARELVEESNLSKE
 5 VKEDTLRRLDENEFKPKQRTKANSVIPHQLHLAELQKILQNQGQYYPFLLDTFEKEDGQDNKIE
 ELLRFRIPYVVGPLVTKKDVEHAGGDADNHWVERNEGFEKSRVTPWNFDKVFNRDKAARDFIER
 LTGNDTYLIGECTLPQNSLRYQLFTVLNELNNVRVNGKKFDSKTKADLINDLFKARKTVSLSAL
 KDYLKAQGKGDVTITGLADESKFNSSLSSYNLKKTFDAEYLENEDNQETLEKIIIEIQTVFEDS
 KIASRELSKLPLDDDQVKKLSQTHYTGWGRLEKLLDSKIIDERGQKVSILDKLKSTSQNFMISI
 10 INNDKYGVQAWITEQNTGSSKLTDEKVNELTTSPANKRGIKQSFVAVLNDIKKAMKEEPRRVYL
 EFAREDQTSVRSVPRYNQLKEYQSKSLSEEAKVLKKTLDGNKNKMSDDRYFLYFQQQGKDMYT
 GRPINFERLSQDYDIDHIIPQAFTKDDSLDNRVLVSRPENARKSDSFAYTDEVQKQDGSLWTSLS
 LKSGFINRKKYERLTAKAGKYLDGQKTGFARQLVETRQIIKNVASLIEGEYENSKAVAIRSEIT
 ADMRLLVGIIKKHREINSFHAFDALLITAAGQYMQNRYPDNRDSTNVYNEFDRYTNDYLNLRQL
 15 SSRDEVRLKSFSGFVVGTMKGNEDWSEENTSYLRKVMFMKNILTTKKTEKDRGPLNKETIFSP
 KSGKKLIPLNSKRSDTALYGGYSNVYSAYMTLVRANGKNLLIKIPISIANQIEVGNLKINDYIV
 NNPAIKKFEKILISKPLPLGQLVNEDGNLIYLASNEYRHNALQWLSTTDADKIASISENSSDEE
 LLEAYDILTSENVKNRFPFFKKDIDKLSQVRDEFLDSDKRIAVIQTILRGLQIDAAYQAPVKIIS
 SKKVSVDWHKLQQSGGIKLSDNSEMIYQSATGIFETRVKISDLL

SEQ ID NO: 314

IVDYCIGLDLGTGSGVWAVVDMNHRMLMKRNGKHLWGSRLFSNAETAANRRASRSIRRRYNKRRE
 RIRLLRAILQDMVLEKDPFFIRLEHTSFLDEEDKAKYLGTDYKDNYNLFIDEDFNDYTYHYKY
 PTIYHLRKALCESTEKADPRLIYLALHHIVKYRGNFLYEGQKFNMDASNIEDKLSDIFTQFTSF
 25 NNIPYEDDEKKNLEILEILKKPLSKKAKVDEVMTLIAPEKDYKSAFKELVTGIAGNKMNVTKMI
 LCEPIKQGDSEIKLKFSDSNYDDQFSEVEKDLGEYVEFVDALHNVYSWVELQTIMGATHDNAS
 ISEAMVSRYNKHHDLLKLLKDCIKNNVPNKYFDMFRNDSEKSKGYNYINRPSKAPVDEFYKYV
 KKCIKVDTPPEAKQILNDIELENFLKQNSRTNGSVPYQMQLDEMIKIIDNQAEYYPILKEKRE
 QLLSILTFRIPIYFGLPNETSEHAWIKRLEGKENQRILPWNQYQDIVDVDATAEGFIKRMRSYCT
 30 YFPDEEVLPKNSLIVSKYEVYNELNKIRVDDKLEVDVKNDIYNELFMKNKTVTEKKLKNWLVN
 NQCCSKDAEIKGFQKENQFSTSLTPWIDFTNIFGKIDQSNFDLIENIYDLTVFEDKKIMKRRL
 KKKYALPDDKVKQILKLKYKDW SRLSKLLDGIVADNRFGSSVTVLVDVLEMSRLNLMEIINDKD
 LGYAQMIEEATSCPEDGKFTYEEVERLAGSPALKRGIWQSLQIVEEITKVMKCRPKYIYIEFER
 SEEAKERTESKIKKLENVYKDLDEQTKKEYKSVLEELKGFNDTKKISSDSLFLYFTQLGKCMYS
 35 GKKLDIDSLDKYQIDHIVPQSLVKDDSFDRNLVVPSENQRKLDLVPFDIRDKMYRFWKLLF
 DHELISPKKFYSLIKTEYTERDEERFINRQLVETRQITKNVTQIIEDHYSTTKVAAIRANLSHE
 FRVKNHIYKNRDINDYHHAHDAYIVALIGGFMRDRYPNMHDSKAVYSEYMKMRKNKNDQKRWK
 DGFVINSNMNYPYEVGDGLIWNPDLINEIKKCFYKDCYCTTKLDQKSGQLFNLTVLSNDAHADK
 GVTKAVVPVNKNRSDVHKYGGFSGLQYTIVAIEGQKKKGKKTENVKKISGVPLHLKAASINEKI
 40 NYIEEKEGLSDVRIKDNIPVNQMIEMDGGEYLLTSPTEYVNARQLVLNEKQCALIADIYNAIY
 KQDYDNLDLILMIQLYIELTNKMKVLYPAYRGIAEKFESEMNENYVVISKEEKANIKQMLIVMH
 RGPQNGNIVYDDFKISDRIGRLKTKNHNLNINIVFISQSPTGIYTKKYKL

SEQ ID NO: 315

MKSEKKYYIGLDVGTNSVGWAVTDEFYNILRAK GKDLWGVRLF EKADTAANTRIFRSGRRRNDR
 KGMRLQILREIFEDEIKKVDKDFYDRLDESKFWAEDKKVSGKYSLFNDKNFSDKQYFEKFPTIF
 HLRKYLMEEHGKVDIRYYFLAINQMMKRRGHFLIDGQISHVTDDKPLKEQLILLINDLLKIELE

EELMDSIFEILADVNEKRTDKNNLKELIKQDFNKQEGNILNSIFESIVTGKAKIKNIISDED
 ILEKIKEDNKEDFVLTGDSYEENLQYFEEVLQENITLFTNTLKSTYDFLILQSILKGKSTLSDAQ
 VERYDEHKKDLEILKKVIKKYDEDGKLFKQVFKEDNGNGYVSYIGYYLNKNKKITAKKKISNIE
 FTKYVKGILEKQDCEDDEDVKYLLGKIEQENFLKQISSINSVIPHQIHLFELDKILENLAKNY
 5 PSFNNKKEEFTKIEKIRKTFTRIPYYVGPLNDYHKNNNGNAWIFRNKGEKIRPWNFEKIVDLH
 KSEEEFIKRLNLQCTYLPEETVLPKSSILYSEYMLNELNNLRINGKPLDSTDVKLKLIEELFKK
 KTKVTLKSIRDYMRNNFADKEDFDNSEKNLEIASNMKSYIDFNNILEDKFDVEMVEDLIEKIT
 IHTGNKKLLKKYIEETYPDLSSSQIQKIINLKYKDWGRLSRKLLDGIKGTKEKTEKTDTVINFL
 RNSSDNLMQIIGSQNYSFNEYIDKLRKKYIPQEISYEVENLYVSPSVKKMIWQVIRVTEEITK
 10 VMGYDPDKIFIEMAKSEEEKTTISRKNKLLDLYKAIKKDERDSQYEKLLTGLNKLDDSDLRSR
 KLYLYYTQMGDMYTGEKIDLDKLFDDSTHYDKDHIIPQSMKKDDSIINNVLVNKNANQTTKGN
 IYPVPSSIRNNPKIYNYWKYLMEKEFISKEKYNRLIRNTPLTNEELGGFINRQLVETRQSTKAI
 KELFEKFYQKSKIIPVKASLASDLRKDMNTLKSREVNDLHHAHDAFLNIVAGDVWNREFTSNPI
 NYVKENREGDKVKYSLSKDFTPRKSKGKVIWTPEKGRKLIIVDTLNKPSVLISNESHVKKGELF
 15 NATIAGKKDYKKGKIYLPKKDDRLQDVSKYGGYKAINGAFFFLVEHTKSKKRIRSIELFPLHL
 LSKFYEDKNTVLDYAINVLQLQDPKIIIDKINYRTEI IIDNFSYLISTKSNDGSITVKPNEQMY
 WRVDEISNLKKIENKYKKDAILTEEDRKIMESYIDKIYQQFKAGKYKNRRTTDTIEKYEIIDL
 DTLDNKQLYQLLVAFISLSYKTSNNAVDFTVIGLGTGCKPRITNLPDNTYLVYKSITGIYEKR
 IRIK

SEQ ID NO: 316

MKLRGIEDDYSIGLDMGTSSVGWAVTDERGTLAHFKRKPTWGSRLFREAQTAAVARMPRGQRRR
 YVRRRWRLDLLQKLFEQQMEQADPDDFIRLRQSRLLRDDRAEEHADYRWPLFNDCKFTERDYYQ
 RFPTIYHVRSWLMETDEQADIRLIYLALHNIVKHGRNFLREGQSLSAKSARPDEALNHLRETLR
 25 VWSSERGFECESIADNGSILAMLTHPDLSPSDRRKKIAPLFDVKSDDAADKKLGIALAGAVIGL
 KTEFKNIFGDFPCEDSSIYLSNDEAVDAVRSACPDDCAELFDRLCEVYSAYVLQGLLSYAPGQT
 ISANMVEKYRRYGEDLALLKKLVKIYAPDQYRMFFSGATYPGTGIYDAAQARGYTKYNLGPCKS
 EYKPSSESMQYDDFRKAVEKLFKAKTDARADERYRMMMDRFDKQQFLRRLKTSNDGSIYHQLHLEE
 LKAIVENQGRFYFPFLKRDADKLVSLVSFRIPYYVGPLSTRNARTDQHGENRFAWSEKPGMQDE
 30 PIFPWNWESIIDRSKSAEKFILRMTGMCTYLQQEPVLPKSSLLYEEFCVLNELNGAHWSIDGDD
 EHRFDAADREGIIEELFRKRRTVSYGDVAGWMERERNQIGAHVCGGQGEKGFESKLGSYIFFCK
 DVFKVERLEQSDYPMIERIILWNTLFEDRKILSQLRKEEYGSRLSAEQIKTICKKRFTGWGRLS
 EKFLTGITVQVDEDSVSIMDVLREGCPVSGKRGRAMVMMEILRDEELGFQKKVDDFNRAFFAEN
 AQALGVNELPGSPAVRRSLNQSIRIVDEIASIAGKAPANIFIEVTRDEDPKKKGRRTKRRYNDL
 35 KDALEAFKKEDPELWRELCEAPNDMDERLSLYFMQRGKCLYSGRAIDIHQLSNAGIYEVDHII
 PRTYVKDDSLNKAALVYREENQRKTDMLLIDPEIRRRMSGYWRMLHEAKLIGDKKFRNLLRSRI
 DDKALKGFIARQLVETGQMVKLVRSLLEARYPETNII SVKASISHDLRTAAELVKCREANDFHH
 AHDAFLACRVGLFIQKRHPCVYENPIGLSQVVRNYVRQQADIFKRCRTIPGSSGFI VNSFMTSG
 FDKETGEIFKDDWDAAEAEVEGIRRLNFRQCFISRMPFEDHGVFWDATIYSPRAKKTAAALPLKQ
 40 GLNPSRYGSFSREQFAYFFIYKARNPRKEQTLFEFAQVPVRLSAQIRQDENALERYARELAKDQ
 GLEFIRIERSKILKNQLIEIDGDRLCITGKEEVNACELAFQAQDEMVRVIRMLVSEKPVSRCEVI
 SLFNRILLHGDQASRLSKQLKLALLSEAFSEASDNVQRNVVLGLIAIFNGSTNMVNLSDIGGS
 KFAGNVRIKYYKKELASPKVNVHLIDQSVTGMFERRTKIGL

SEQ ID NO: 317

MENKQYYIGLDVGTNSVGWAVTDTSYNLLRAKGKDMWGARLFKANTAAERRTKRTSRRRSERE
 KARKAMLKELFADEINRVDPSPFFIRLEESKFFLDDRSENNRQRYTLFNDATFTDKDYEEKYKTI

FHLRSALINSDEKFDVRLVFLAILNLFSHRGHFLNASLKGDGDIQGMDVIFYNDLVESCEYFEIE
 LPRITNIDNFEKILSQKGKSRKILEELSEELSISKKDKSKYNLIKLSGLEASVVELYNIEDI
 QDENKKIKIGFRESDYEESSLKVKEIIGDEYFDLVERAKSVHDMGLLSNIIGNSKYLCEARVEA
 YENHHKDLLKIKELLLKKYDKKAYNDMFRKMTDKNYSAYVGSVNSNIAKERRSVDKRKIEDLYKY
 5 IEDTALKNIPDDNKDKIEILEKIKLGEFLKKQLTASNGVIPNQLQSRELRAILKKAENYLPFLK
 EKGEKNLTVSEMIQQLFEFQIPYYVGPLDKNPCKDNKANSWAKIKQGGRILPWNFEDKVDVKGS
 RKEFIEKMVRKCTYISDEHTLPKQSLLEYKFMVLNEINNIKIDGEEKISVEAKQKIYNDLFLVKGK
 KVSQKDIKKELISLNNIMDKDSVLSGTDVTCNAYLSSIGKFTGVFKEEINKQSIVDMIEDIIFLK
 TVYGDEKRFVKEEIVEKYGDEIDKDKIKRILGKFSNWGNLSKSFLELEGADVGTGEVRSIIQS
 10 LWETNFNLMELLSSRFTYMDELEKRVKKLEKPLSEWTIEDLDDMYLSSPVKRMIVQSMKIVDEI
 QTVIGYAPKRIFVEMTRSEGEKVRTKSRKDRLELYNGIKEDSKQWVKELDSKDESYFRSKMY
 LYYLQKGRCMYSGEVIELDKLMDDNLYDIDHIYPRSFVKDDSLDNLVLVKEINNRRKQNDPITP
 QIQASCQGFWKILHDQGFMSNEKYSRLTRKTQEFSDDEEKLSEFINRQIVETGQATKMAQILQKS
 MGEDVDVVFSEKARLVSEFRHKFELFKSRLINDFHHANDAYLNIIVGNSYFVKFTRNPANFIKDA
 15 RKNPDNPVYKYHMDRFFERDVKSKSEVAWIGQSEGNSTIVIVKKTMAKNSPLITKKVEEGHGS
 ITKETIVGVKEIKFGRNKVEKADKTPKKPNLQAYRPIKTSDERLCNILRYGGRTSISISGYCLV
 EYVKKRKTIRSLEAIPVYLGRKDSLSEEKLLNYFRYNLNDGGKDSVSDIRLCLPFIISTNSLVKI
 DGYLYYLGKNDRIQLYNAYQLKMKKEEVEYIRKIEKAVSMSKFDEIDREKNPVLTEEKNIEL
 YNKIQDKFENTVFSKRMSLVKYNKKDLSFGDFLKNKKSKEEIDLEKQCKVLYNIIIFNLSNLKE
 20 VDLSDIGGSKSTGKCRCKKNITNYKEFKLIQQSITGLYSCEKDLMTI

SEQ ID NO: 318

MKNLKEYYIGLDIGTASVGWAVTDESYNIPKFNGKKMWGVRLFDDAKTAEERRTQRGSRRRLNR
 RKERINLLQDLFATEISKVDPNFFLRDNDSDLYREDKDEKLKSKYTLFNDKDFKDRDYHKKYPT
 25 IHHLIMDLIEDEGKKDIRLLYLACHYLLKNRGHFIFEGQKFDTKNSFDKSINDLKIHLRDEYNI
 DLEFNEDLIEIITDTTLNKTNNKKELKNIVGDTKFLKAISSAIMIGSSQKLVDLFEDGEFEETT
 VKSVDFSTTAFDDKYSEYEEALGDTISLLNILKSIYDSSILENLLKDADKSKDGNKYISKAFVK
 KFNKHGKDLKTLKRIKKYLPSEYANIFRNKSINDNYVAYTKSNITSNKRTKASKFTKQEDFYK
 FIKKHLDTIKETKLNSSSENEIDLKLIDEMLTIDIEFKTFIPKLKSSDNGVIPYQLKLMELKKILDN
 30 QSKYYDFLNESDEYGTVDKVESIMEFRIPYYVGPLNPDSKYAWIKRENTKITPWNFKDIVDLD
 SSREEFIDRLIGRCTYLKEEKVLPKASLIYNEFMVLNELLNKLNEFLITEEMKKAIFEELFKT
 KKKVTLKAVSNLLKKEFNLTGDIILLSGTDGDFKQGLNSYIDFKNIIGDKVDRDDYRIKIEEIIK
 LIVLYEDDKTYLKKKIKSAYKNDFTDDEIKKIAALNYKDWGRLSKRFLTGIEGVDKTTGEKGS
 IYFMREYNLNLMEMLMSGHYTFTEEVEKLNPNVENRELCYEMVDELYLSPSVKRMWLQSLRVVDEI
 35 KRIIGKDPKKIFIEMARAKEAKNSRKESRKNKLEFYKFGKKAFINEIGEERYNYLLNEINSEE
 ESKFRWDNLYLYYTQLGRCMYSLEPIDLADLKSNNIYDQDHIYPKSKIYDDSLNENRVLVKKNLN
 HEKGNQYPIPEKVLNKNAYGFWKILFDKGLIGQKKYTRLTRRTPFEERELAEFIERQIVETRQA
 TKETANLLKNICQDSEIVYSKAENASRFRQEFDIKCRTVNDLHHMHDAYLNIIVGNVYNTKFT
 KNPLNFIKDKDNVRSYNLENMFKYDVVRGSYTAWIADDSEGNVKAATIKKVRELEGKNYRFT
 40 MSYIGTGGLYDQNLMRKGKGQIPQKENTNKSNIKYGGYNKASSAYFALIESDGKAGRETRLET
 IPIMVYNQEKYGNTEAVDKYLKDNLELQDPKILKDKIKINSLIKLDGFLYNIKGKTGDSLSIAG
 SVQLIVNKEEQKLIKMDKFLVKKKDNKDIDKVTSDNIKEEELIKLYKTLSDKLNGIYSNKRN
 NQAKNISEALDKFKEISIEEKIDVLNQIILLFQSYNNGCNLKSIGLSAKTGVPFIPKKLNYKEC
 KLINQSITGLFENEVDLLNL

45

SEQ ID NO: 319

MGKMYYLGLDIGTNSVGAVTDPSTYHLLKFKGEPMWGAHVFAAGNQSAERRSFRTSRRRLDRRQ
 QRVKLVQEIFAPVISPIDPRFFIRLHESALWRDDVAETDKHIFNDPTYTDKEYYSYPTIHL
 IVDLMESSEKHDPRVLVYLAVALVAHRGHFLNEVDKDNIGDVLSDAFYPEFLAFLSDNGVSPW
 VCESKALQATLLSRNSVNDKYKALKSLIFGSQKPEDNFDANISEDGLIQLLAGKKVKVKNLFPQ
 5 ESNDASFTLNDKEDAIEEILGTLPDECEWIAHIRRLFDWAIMKHALKDGRITISESKVKLYEQH
 HHDLTQLKYFVKTYLAKEYDDIFRNVDSSETTKNYVAYSYHVKEVKGTLPKNKATQEEFCKYVLG
 KVKNIECSEADKVDFDEMIQRLTDNSFMPKQVSGENRVIPYQLYYYELKTIILNKAASYLPFLTQ
 CGKDAISNQDKLLSIMTFRIPYFVGPLRKDNSEHAWLERKAGKIYPWNFNDKVDLDKSEEA FIR
 RMTNTCTYYPGEDVLPDLSLIYEKFMILNEINNIRIDGYPI SVDVKQQVFGLEKKRRVTVKDI
 10 QNLLLSLGALDKHGKLTGIDTTIHSNYNTYHHFKSLMERGVLTRDDVERIVERMTYSDDTKRVR
 LWLNNNYGTLTADDVKHISRLRKHDFGRLSKMFLTGLKG VHKETGERASILDFMWNTNDNLMQL
 LSECYTFSDEITKLQEAYYAKAQLSLNDFLDSMYISNAVKRPIYRTLAVVNDIRKACGTAPKRI
 FIEMARDGESKKKRSVTRREQIKNLYRSIRKDFQQEVDVFLEKILENKSDGQLQSDALYLYFAQL
 GRDMYTGDPIKLEHIKDQSFYNIDHIYPQSMVKDDSLDNKVLVQSEINGEKSSRYPLDAAIRNK
 15 MKPLWDAYYNHGLISLKKYQRLTRSTPFTDDEKWDFINRQLVETRQSTKALAILLKRKFPDTEI
 VYSKAGLSSDFRHEFGLVKSRNINDLHAKDAFLAIVTGNVYHERFNRRWFMVNQPYSVKTKTL
 FTHSIKNGNFVAWNGEEDLGRIVKMLKQNKNTIHFTRFSFDRKEGLFDIQPLKASTGLVPRKAG
 LDVVKYGGYDKSTAAYYLLVRFTLEDKKTQHKLMIPVEGLYKARIDHDKEFLT DYAQTTISEI
 LQKDKQKVINIMFPMGTRHIKLSNMISIDGFYLSIGGKSSKGKSVLCHAMVPLIVPHKIECYIK
 20 AMESFARKFKENNKLRIVEKFDKITVEDNLNLYELFLQKLQHNPYNKFFSTQFDVLTNGRSTFT
 KLSPEEQVQTLLNILSIFKTCRSSGCDLKSINGSAQAARIMISADLTGLSKKYSDIRLVEQSAS
 GLFVSKSQNLLEYL

SEQ ID NO: 320

25 MTKKEQPYNIGLDIGTSSVGWAVTNDNYDLLNIKKKNLWGVRLFEEAQTAKETRLNRSTRRRYR
 RRKNRINWLNEIFSEELAKTDPSTLIRLQNSWVSKKDPDRKRDKYNLFIDGPYTDKEYYREFPT
 IFHLRKELILNKDKADIRLIYLAHLNLIKRYGNFTYEHQKFNISNLNNLSKELIELNQQLIKY
 DISFPDDCDWNHISDILIGRGNATQKSSNILKDFTLTKETKLLKEVINLILGNVAHLNTIFKT
 SLTKDEEKLNFSGKDIESKLDDLDSDILDDDQFTVLDAANRIYSTITLNEILNGESYFSMAKVNQ
 30 YENHAIDLCKLRDMWHTTKNEEAVEQSRQAYDDYINKPKYGTKELYTSLKKFLKVALPTNLAKE
 AEEKISKGYLVKPRNSENQGVVQYQNLKIEMEKI IDNQSQYYPFLKENKEKLLSILSFRIPIYV
 GPLQSAEKNPFAWMERKSNGHARPWNFDEIVDREKSSNKFIRRMTVTD SYLVGEPVLPKNSLIY
 QRYEVLNELLNIRITENLKTNP IGSRLTVETKQRIYNELFKKYKKVTVKKLTKWLIAQGGYKNP
 ILIGLSQKDEFNSTLT TYLDMKKIFGSSFMEDNKNYDQIEELIEWLTIFEDKQILNEKLHSSKY
 35 SYTPDQIKKISNMRYKGWGRLSKKILMDITTETNTPQLQLSNYSILDLMWATNNNFISIMSND
 KYDFKNYIENHNLKNEDQNISDLVNDIHVSPALKRGITQS I KIVQEIVKFMGHAPKHIFIEVT
 RETKKSEITTSREKRIKRLQSKLLNKANDFKPQLREYLVPNKKIQEELKKHKNDLSSERIMLYF
 LQNGKSLYSEESLNINKLSDYQVDHILPRTYIPDDSLNKALVLAKENQRKADDLLNSNVIDR
 NLERWTYMLNNMIGLKKFKNLTRRVI TDKDKLGF IHRQLVQTSQMVKGVANILDNMYKNQGT
 40 CIQARANLSTAFRKALSGQDDTYHFKHPELVKNRNVNDFHHAQDAYLASFLGTYRLRRFPTNEM
 LLMNGEYNKFYGVKELYSKKKKLPDSRKNGFI ISPLVNGTTQYDRNTGEI IWNVGFRDKILKI
 FNYHQCNVTRKTEIKTGQFYDQTIYSPKNPKYKKLIAQKKDMDPNYGGFSGDNKSSITIVKID
 NNKIKPVAIPIRLINDLKDKKTLQNWLEENVKHKSIQI IKNNVPIGQIIYSKKVGLLSLNSDR
 EVANRQQILLPPEHSALLRLLQIPDEDLQILAFYDKNILVEILQELITKMKKFYPFYKGEREF
 45 LIANIENFNQATTSEKVN SLEELITLLHANSTSAHLIFNNIEKKAFGRKTHGLTLNNTDFIYQS
 VTGLYETRIHIE

SEQ ID NO: 321

MTKFNKNYSIGLDIGVSSVGYAVVTEDYRVPFAKFKVLGNTEKEKIKKNLIGSTTFVSAQPAKG
 TRVFRVNRRRIDRRNHRITYLRDIFQKEIEKVDKNFYRRLDESFRVLGDKSEDLQIKQPPFGDK
 ELETAYHKKYPTIYHLRKHLADADKNSPVADIREVYMAISHILKYRGHFLTLDKINPNNINMQN
 5 SWIDFIESCQEVFDLEISDESKNIADIFKSSENROEKVKILPYFQQELLKKDKSIFKQLLQLL
 FGLKTKFKDCFELEEEEPDLNFSKENYDENLENFLGSLEEDFSDVFAKLKVLRTDITLLSGMLTYT
 GATHARFSATMVERYEEHRKDLQRFKFFIKQNLSEQDYLDIFGRKTQNGFDVDKETKGYVGYIT
 NKMVLTNPQKQKTIQQNFYDYISGKITGIEGAEYFLNKISDGTFLRKLRTSDNGAIPNQIHAYE
 LEKIIERQGKDYPFLENKDKLLSILTFKIPYYVGPLAKGSNSRFAWIKRATSSDILDDNDEDT
 10 RNGKIRPWNQKQLINMDETRDAFITNLIGNDIILLNEKVLPKRSLIYEEVMLQNELTRVKYKDK
 YGKAHFFDSELRQNIINGLFKNNSKRVNAKSLIKYLSDNHKLNAIEIVSGVEKGKSFNSTLKT
 YNDLKTIFSEELLDESIYQKELEEIIVKITVFDKKSIIKNYLTKEFFGHLEILDEEKINQLSKLR
 YSGWGRYSAKLLLDIRDEDTGFNLLQFLRNDEENRNLTKLISDNTLSFEPKIKDIQSKSTIEDD
 IFDEIKKLAGSPAIRGILNSIKIVDELVQIIGYPPHNIVIEAMARENMTTEEGQKKAKTRKTKL
 15 ESALKNIENSLLENGKVPHSDEQLQSEKLYLYLQNGKDMYTLDKTGSPAPLYLDQLDQYEVHD
 IIPYSFLPIDSIDNKVLTHRENNQQKLNINPDKETVANMKPFWEKLYNAKLISQTKYQRLTTSE
 RTPDGVLTESMKAGFIERQLVETRQIIKHVARILDNRFSDTKIIITLKSQILITNFRNTFHIKIR
 ELNDYHHAHDAYLAVVVGQTLLKVYPKLAPELIYGHHAHFNREHENKATLRKHLYSNIMRFFNN
 PDSKVS KDIDWCNRDLPIIKDVIYNSQINFVKRTMIKKGAFYNQNPVGKFNKQLAANNRYPLKT
 20 KALCLDTSIYGGYGPMNSALSIIIIAERFNEKKGKIETVKEFHDIFIIDYEKFNNNPFQFLNDT
 SENGFLKKNINRVLGFYRIPKYSMLQKIDGTRMLFESKSNLHKATQFKLTQTQNELFFHMKRL
 LTKSNLMDLKS SAIKESQNFILKHKEEFDNISNQLSAFSQKMLGNTTSLKNLIKGYNERKIKE
 IDIRDETIKYFYDNFIKMF S FVKSGAPKDINDFFDNKCTVARMRPKPKDKLLNATLIHQSI TGL
 YETRIDLSKLGED

25
SEQ ID NO: 322

MKQEYFLGLDMGTGSLGWAVTDSTYQVMRKHGKALWGTRLFESASTAEERRMFRTARRRLDRRN
 WRIQVLQEIFSEEISKVDPGFFLRMKESKYYPEDKRDAEGNCPPELPYALFVDDNYTDKNYHKDY
 PTIYHLRKMLMETTEIPDIRLVYLVLHMMKHRGHFLLSGDISQIKEFKSTFEQLIQNIQDEEL
 30 EWHISLDDAAIQFVEHVLKDRNLTRSTKKSRLIKQLNAKSACEKAILNLLSGGTVKLSDIFNNK
 ELDESERP KVSFADSGYDDYIGIVEAEALAEQYYIIASAKAVYDWSVLVEILGNSVSISEAKIKV
 YQKHQADLKTLLKIVRQYMTKEDYKRVFVDTEEKLNNYSAYIGMTKKNGKKVDLKSQCTQADF
 YDFLKKNVIKVIDHKEITQEIESEIEKENFLPKQVTKDNGVIPYQVHDYELKKILDNLGTRMPF
 IKENA EKIQQLF EFRIPIYVGPLNRVDDGKD GFTWSVRKSDARIYPWNFTVIDVEASAEKFI
 35 RRMTNKCTYLVGEDVLPKDSLVS KFMVLNELLNRLNGEKISVELKQRIYEELFCKYRKVTRK
 KLERYLVIEGIAKKGVEITGIDGDFKASLTAYHDFKERLTDVQLSQRAKEAIVLNVVLF GDDKK
 LLKQRLSKMYPNLTTGQLKGICSLSYQGWGRLSKTFLEEITVPAPGTGEVWNIMTALWQTNLNL
 MQLLSRNYGFTNEVEEFNTLKKETDLSYKTVDELYVSPAVKRQIWQTLKVVK EIQKVMGNAPKR
 VFVEMAREKQEGKRS DSRKKQLVELYRACKNEERDWITELNAQSDQQLRSDKFLYYIQKGRCM
 40 YSGETIQ LDELWDNTKYDIDHIYPQSKTMDDSLNNRVLVKKNYNAIKSDTYPLSLDIQKKMMSF
 WKMLQQQGFI TKEKYVRLVRSDEL S ADEL AGFIERQIVETRQSTKAVATILKEALPDTEIVYVK
 AGNVS NFRQTYELLKVREMNDLHHAHAKDAYLNIIVGNAYFVKFTKNAAWFIRNNPGRSYNLKRMF
 EFDIERSGEIAWKAGNKGSI VTVKKVMQKNNILVTRKAYEVKGGLFDQQIMKKGKGQVPIKGND
 ERLADIEKYGGYNKAAGTYFMLVKS LDKKGKEIRTIEFVPLYLKNQIEINHESAIQYLAQERGL
 45 NSPEILL SKIKIDTLFKVDGFKMWLSGRTGNQLIFKGANQLILSHQEAAILKGVVKYVNRKNEN
 KDAKLSERDGMTEEKLLQLYDTFLDKLSNTVYSIRLSAQIKTLTEKRAKFI GLSNEDQCIVLNE
 ILHMFQCQSGSANLKLIGPGSAGILVMNNNITACKQISVINQSPTGIYEKEIDLKIL

SEQ ID NO: 323

MKKPYSIGLDIGTNSVGWAVVTDDYKVPAAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRL
 KRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFHRLLEDSEFLVTEDEKRGHRPIFGNLEEEVKY
 5 HENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEF LAVY
 DNTFENSSLQE QNVQVEEILTDKISKS AKKDRVLKLF PNEKSNGRFAEFLKLIVGNQADFKKH
 ELEEKAPLQFSKDTYEEEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVDVGTAPLSAS
 MIQRYNEHQMDLAQLKQFIRQKLS DKYNEVFS DVSKDGYAGYIDGKTNQEA FYKYLKGLLNKIE
 GSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFY PFLADNQDRIEKL LTFRI
 10 PYYVGPLARGKSDFAWLSRKSADKITPWNFDEIVDKESSAEAFINRMTNYDLYLPNQKVL PKHS
 LLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRI
 VDLTGLDKENKVFNASYGTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRENY
 SDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLI DDGNSNRNFMQLINDDALS
 FKEEIIAKAQVIGETDNLNQVVS DIAGSPAIIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQ
 15 FTNQGRNSQQRLKGLTDSIKEFGSQILKEHPVENSQ LQNDRLFLYYLQNGRDMYTGEELDIDY
 LSQYDIDHII PQA FIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRK
 FDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKIRQVKIVTLKS
 NLVSNFRKEFELYK VREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKA
 TAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQVNIVKKVEEQTGGFSKES
 20 ILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIME
 KMTFERDPVAFLEKRGYRNVQEEENIKLPKYSLFKLENGRKRLLASARELQKGNEIVLPNHLGT
 LLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVS NF SKKYTLAEGNLEKIKELYAQNNGEDLK
 ELASSFINLLTFTAIGAPATFKFFDKNIDRKRYTSTTEILNATLIHQ SITGLYETRIDLNLKLG
 D

25
 SEQ ID NO: 324

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
 KRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAY
 HEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTY
 30 NQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
 DLAEDA KLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSAS
 MIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFYKF IKPILEKMD
 GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRI
 PYYVGPLARGNSRF AWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHS
 35 LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
 SVEISGVEDRFNASLGTYHDL LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
 HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFL KSDGFANRNF MQLIHDDSLTF
 KEDIQKAQVSGQDSLHEHIANLAGSPAIIKKGILQTVKVVD ELVKVMGRHKPENIVIVEMARENQ
 TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
 40 LSDYDV DHIVPQSFLKDDSIDNKVLT RSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRK
 FDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVDVRKMIK
 SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
 45 KSKKLKSVKELLGITIMERS SFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELNGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEK LKGS PEDNEQKQLFVEQHKHYLDEII EQISEFSKR V

ILADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQ SITGLYETRIDLSQLGGD

SEQ ID NO: 325

5 MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRL
KRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSF LVPDDKRDSKYPIFGNLVEEKAY
HDEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNNDIQKNFQDFLDY
NAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCF
NLDEKASLHFSKESYDEDLETLLGYIGDDYSDVFLKAKKLYDAILLSGFLTVDNETEAPLSSA
10 MIKRYNEHKEDLALLKEYIRNISLKYTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKKLLAEFE
GADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDQAKFYFPLAKNKERIEKILTFRI
PYVVGPLARGNSDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLP EEEKVL PKHS
LLYETFN VYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDG
IELKGIEKQFNSSLSTYHDLNLIINDKEFLDDSSNEAIEEIIHTLTIFEDREMIKQRLSKFEN
15 IFDKSVLKKLSRRHYTGWGLSAKLINGIRDEKSGNTILDYLI DDGISNRNFMQLIHDDALSFK
KKIQKAQIIGDEDKGNIEVVKSLPGSPAIIKKGILQSIKIVDELVKVMGGRKPESIVVEMAREN
QYTNQGKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTG
DDLDIDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSDDVPSLEVVKRKT FWYQLLS
KLISQRKFDNLTKAERGGLSPEDKAGFIQRQLVETRQITKHVARLLDEKFNNKKDENNRVRTV
20 KIITLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVVASALLKKYPKLEPEFVYGDYPKYN
SFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATVRRVLS
YPQVNVVKKVEEQNHGLDRGPKGLFNANLSSKPKPNSNENLVGAKEYLDPKKYGGYAGISNSF
TVLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLEKGYKDIELIIELPKYSLFELS
DGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKKEFEEL
25 FYYILEFNENYVGAKKNGKLLNSAFQSWQNHSIDELCSSFIGPTGSEKGLFELTSRGSAA DFE
FLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGEG

SEQ ID NO: 326

30 MKKQKFS DYYLGF DIGTNSVGWCVTDL DYNVLRFNKKDMWGSRLFDEAKTAAERRVQRNSRRRL
KRRKWRLNLLEEIFSDEIMKIDSNFFRRLKESSLWLEDKNSKEKFTLFND DNYKD YDFYKQYPT
IFHLRDEL IKNPEKKDIRLIY LALHSIFKSRGHFLFEGQNLKEIKNFETLYNNLISFLEDNGIN
KSIDKD NIEKLEKIICDSGKGLKDKEKEFKGIFNSDKQLVAIFKLSVGSSVSLNDFDTDEYKK
EEVEKEKISFREQIYEDDKPIIYYSILGEKIELLDIAKSFYDFMVLNNILSDSNYISEAKVKLYE
EHKKDLKNLKYIIRKYNKENYDKLFDKNENNYPAYIGLNKEKDKKEVVEKSRLKIDDLIKVIK
35 GYLPKPERIEEKDKTIFNEILNKIELKTILPKQRISDNGLTPYQIHEVELEKILENQSKYYDFL
NYEENG VSTKDKLLKTFKFRIPYVVGPLNSYHKDKGNSWIVRKEEGKILPWNFEQKV DIEKSA
EEFIKRMTNKCTYLNGEDVIPKDSFLYSEYIILNELNKVQVND EFLNEENKRKIIDELFKENKK
VSEKKFKEYLLVNQIANRTVELKGIKDSFNSNYVSYIKFDIFGEKLNLDIYKEISEKSILWKC
LYGDDKKIFEKKIKNEYGDILNKDEIKKINSFKFNTWGR LSEKLLTGIEFINLETGECYSSVME
40 ALRRTNYNLMELLSSKFTLQESIDNENKEMNEVS YRDLIEESYVSPSLKRAILQTLKIYEEIKK
ITGRVPKKVFIEMARGGDESMKNKKIPARQEQLKKLYDSCGN DIANFSIDIKEMKNSLSSYDNN
SLRQKKLYLYLQFGKCMYTGREIDLDRLLQNNDTYDIDHIYPRSKVIKDDSF DNLVLVLKNEN
AEKSNEYPVKKEIQEKMSFWRFLKEKNFISDEKYKRLTGKDDFELRGFMARQLVNVRQT TKEV
GKILQQIEPEIKIVYSKAEIASSFREMFDFIKVRELNDTHHAKDAYLNIVAGNVYNTKFTEKPY
45 RYLQEIKENYDVKKIYNYDIKNAWDKENSLEIVKKNMEKNTVNITRFIKEEKGELFNLP IKKG
ETSNEIISIKPKLYDGKDNKLNEKYGYTSLKAAYFIYVEHEKKNKKVKTFERITRIDSTLIK N
EKNLIKYLVSQKKLLNPKIIKKIYKEQTLIIDSYPYTFTGVDSNKKVELKNKKQLYLEKKYEQI

LKNALKFVEDNQGETEENYKFIYLKKRNNNEKNETIDAVKERYNIEFNEMYDKFLEKLSSSKDYK
 NYINNKLTYTNFLNSKEKFKKLKLWEKSLILREFLKIFNKNTYGKYEIKDSQTKEKLFSPFEDTG
 RIRLGQSSSLGNNKELLEESVTGLFVKKIKL

5 SEQ ID NO: 327

MKNYTIGLDIGVASVGWVCIDENYKILNYYNNRHAFGVHEFESAESAAGRRLKRGMRRRYNNRRKK
 RLQLLQSLFDSYITDSGFFSKTDSQHFWKNNNEFENRSLTEVLSSLRISSRKYPTIYHLRSDLI
 ESNKKMDLRLVYLALHNLVKYRGHFLQEGNWSEAASAEGMDDQLLELVTRYAELENLSPLDLSE
 SQWKAETLLLNRLTKTDQSKELTAMFGKEYEPFCKLVAGLGVSLHQLFPSSEQALAYKETKT
 10 KVQLSNENVEEVMELLLEESALLEAVQPFYQQVVLVYELLKGETYVAKAKVSAFKQYQKDMASL
 KNLLDKTFGEKVYRSYFISDKNSQREYQKSHKVEVLCKLDQFNKEAKFAETFYKDLKKLLEDKS
 KTSIGTTEKDEMLRIIKAIDSNQFLQKQKGIQNAIPHQNSLYEAEKILRNQQAHYPFITTEWI
 EKVKQILAFRIPPYYIGPLVKDTTQSPFSWVERKGDAPITPWNFDEQIDKAASAEAFISMRKTC
 TYLKGQEVLPKSSLTIERFEVLNENLNGIQLRTTGAESDFRHRLSYEMKCWIIDNVFKQYKTVST
 15 KRLLQELKKSPYADELYDEHTGEIKEVFGTQKENAFATSLSGYISMKSILGAVVDDNPAMTEEL
 IYWIAVFEDREILHLKIQEKYPSITDVQRQKLALVKLPGWGRFSRLIDGLPLDEQGQSVLDHM
 EQYSSVFMEVLKNGKGFLEKKIQKMNQHQVDGTTKIRYEDIEELAGSPALKRGIWRSVKIVEEL
 VSIFGEPANIVLEVAREDEGEKKRTKSRKDQWEELTKTTLNKNDPDLKSFIDGEIKSQGDQRFNEQR
 FWLYVTQQGKCLYTGKALDIQNLSMYEVDHILPQNFVKDDSLDNLALVMPEANQRKNQVGQNK
 20 PLEIIIEANQQYAMRTLWERLHELKLISSGKLGRLLKPSFDEVDKDKFIARQLVETRQIIKHVRD
 LLDERFSKSDIHLVKAGIVSKFRFSEIPKIRDYNNKHAMDALFAAALIQSILGKYGKNFLAF
 DLSKKDRQKQWRSVKGSNKEFFLFKNFGNLRQLSPVTGEEVSGVEYMKHVYFELPWQTTKMTQT
 GDGMFYKESIFSPKVQAKYVSPKTEKFVHDEVKNHSICLVEFTFMKKEKEVQETKFIDLKVIE
 HHQFLKEPESQLAKFLAEKETNSPIIHARIIRTIPKYQKIWIEHFPYYFISTRELHNNARQFEIS
 25 YELMEKVQQLSERSSVEELKIVFGLLIDQMNDNYPITYTKSSIQDRVQKFVDTQLYDFKSFEIGF
 EELKKAVAANAQRSDTFGSRISKKPKPEEVAIGYESITGLKYRKPRSVVGTR

SEQ ID NO: 328

MKKEIKDYFLGLDVGTGSVGWAVTDTDYKLLKANRKDLWGMRCFETAETAEVRLHRGARRRIE
 30 RRKKRIKLLQELFSQEIAKTDEGFFQRMKESPFYAEDKTIHQENTLFNDKDFADKTYHKAYPTI
 NHLIKAWIENKVKPDPRLLYLACHNIIKKRGHFLFEGDFDSENQFDTSIQALFEYLREDMEVDI
 DADSQKVKEILKDSSLKNSKQSRNLKILGLKPSDKQKKAITNLISGNKINFADLYDNPDLKDA
 EKNSISFSKDDFDALSDDLASILGDSFELLKAKAVYNCSVLSKVIQDEQYLSFAKVKIYEKHK
 TDLTKLKNVIKKHFPKDYKKVFGYNKNEKNNNNYSYGVGVCKTKSKKLIINNSVNQEDFYKFLK
 35 TILSAKSEIKEVNDILTEIETGTFLPKQISKSNAEIPYQLRKMELEKILSNAEKHFSFLKQKDE
 KGLSHSEKIIMLLTFKIPYYIGPINDNHKKFFPDRCWVVKKEKSPSGKTTTPWNFFDHIDKEKTA
 EAFITSRTNFCTYLVGESVLPKSSLLYSEYTVLNEINNLIIDGKNICDIKLKQKIYEDLFKK
 YKKITQKQISTFIKHEGICNKTDEVIILGIDKECTSSLKSYIELKNIFGKQVDEISTKNMLEEI
 IRWATIYDEGEGKTILKTKIKAEYGKYCSDEQIKKILNLKFSGWGRLSRKFLFETVTSEMPGFSE
 40 PVNIITAMRETQNNLMELLSSEFTFTENIKKINSGFEDAQKQSYDGLVKPLFLSPSVKKMLWQ
 TLKLVKEISHITQAPPKKIFIEMAKGAELEPARTKTRLKILQDLYNNCKNDADAFSSEIKDLSG
 KIENEDNLRRLRSDKLYLYYTQLGKCMYCGKPIEIGHVFDTSNYDIDHIYPQSKIKDDSI SNRVL
 VCSSCNKNKEDKYPLKSEIQSKQRGFWNFLQRNFIISLEKLNRLTRATPISDDETAKFARQLV
 ETRQATKVAKVLEKMFPEKIVYSKAETVSMFRNKFDIVKCREINDFHHAHDAYLNIVVGNVY
 45 NTKFTNNPWNFIKEKRDNPKIADTYNYYKVFDDYVKKRNNITAWEKGKTIITVKDMLKRNTPITYT
 RQAACKKGELFNQTIMKKGLGQHPLKKEGPF SNISKYGGYNKVSAAYYTLIEYEEKGNKIRSLE
 TIPLYLVKDIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSLLKINGFPCHITGKTNDSFLLRP

AVQFCCSNNEVLYFKKIIRFSEIRSQREKIGKTISPYEDLSFRSYIKENLWKTKNDEIGEKEF
YDLLQKKNLEIYDMLLTCHKDITIYKKRPNSATIDILVKGKEKFKSLIENQFEVILEILKLFSA
TRNVSDLQHIGGSKYSGVAKIGNKISSLDNCILYQSIITGIFEKRIDLKLV

5 SEQ ID NO: 329

MEGQMKNNGNNLQQGNYYLGLDVGTSVGVAVTDTDYNVLKFRGKSMWGARLFDEASTAEERT
HRGNRRRLARRKYRLLEQLFEKEIRKIDDNFFVRLHESNLWADDSKPSKFLFNDTNFTDK
DYLKKYPTIYHLRSDLIHNSTEHDIRLVFLALHHLIKYRGHFIDNSANGDVKTLD EAVSDFEE
YLNENDIEFNIENKKEFINVLSDKHLTKKEKKISLKKLYGDITDSENINISVLIEMLSGSSISL
10 SNLFKDIEFDGKQNLSDSDIEETLNDVVDILGDNIDLLIHAKEVYDIAVLTSSLGKHKYLCA
KVELFEKNKDLMLKKYIKKNHPEDYKKIFSSPTEKKNYAAYSQTNSKNVCSQEEFCLFIKPY
IRDMVKSENEDEVRIAKEVEDKSFLTCLKGTNNSVVPYQIHERELNQILKNIVAYLPFMNDEQE
DISVVDKIKLIFKFKIPYVVGPLNTKSTRSWVYRSDEKIYPWNFSNVIDLDDKTAHEFMNRLIGR
CTYTNDPVLPMDSLLYSKYNVLEINPIKVNKAIPVEVKQAIYTDLFENSKKKVTRKSIYIYL
15 LKNGYIEKEDIVSGIDIEIKSKLKHHDFTQIVQENKCTPEEIERIIGILVYSDDKSMLRRWL
KNNIKGLSENDVKYLAKLNYKEWGRLSKTLTDTIYTINPEDGEACSILDIMWNTNATLMEILSN
EKYQFKQNIENYKAENYDEKQNLHEELDDMYISPAARRSIWQALRIVDEIVDIKKSAPKKIFIE
MAREKKSAMKKKRTESRKDTLLELYKSCKSQADGFYDEELFEKLSNESNSRLRRDQLYLYYTQM
GRSMYTGKRIDFDKILNDKNTYDIDHIYPRSKIKDDSI TNRVLVEKDINGEKTDIYPISEDIRQ
20 KMQPFWKILKEKGLINEEKYKRLTRNYELTDEELSSFVARQLVETQQSTKALATLLKKEYPSAK
IVYSKAGNVSEFRNRKDKELPKFREINDLHAKDAYLNIVVGNVYDTKFTEKFFNNIRNENYSL
KRVFDFSVPGAWDAKGSTFNTIKKYMANNPIIAFAPYEVKGELFDQQIVPKGKGQFPKQKGD
IEKYGGYNKLSSAFLFAVEYKGGKARERSLETVIKDVLYLQDPIKYCESVLGLKEPQIIKPK
ILMGSLSFSINNKKLVVTGRSGKQYVCHHIYQLSINDEDSQYLKNI AKYLQEEP DGNIERQNILN
25 ITSVNNIKLFDVLCTKFNSNTYIEIILNSLKNDVNEGREFSELDILEQCNI LLQLLKAFKCNRE
SSNLEKLNNKKQAGVIVIPHLFTKCSVFKVIHQSIITGLFEKEMDLLK

SEQ ID NO: 330

MGRKPYILSLDIGTGSVGYACMDKGFNVLYHDKDALGVYLF DGALTAQERRQFRTSRRRKNRR
30 IKRLGLLQELLAPLVQNPNFYQFQRQFAWKNDNMDFKNKSLSEVLSFLGYESKKYPTIYHLQEA
LLLKDEKFDPELIYMALYHLVKYRGHFLFDHLKIENLTNNDNMHDFVELIETIYENLNNIKLNLNLD
YEKTKVIYEILKDNEMTKNDRAKRVKNMEKKLEQFSIMLLGLKFNEGKLFNHADNAEELKGANQ
SHTFADNYEENLTPFLTVEQSEFIERANKIYLSLTLQDILKGKKSMMASKVAAYDKFRNELKQV
KDIVYKADSTRTOFKKIFVSSKSLKQYDATPNDQTFSSCLFDQYLIRPKKQYSLLIKELKKI
35 IPQDSELYFEAENDTLLKVLNTTNDASIPMQINLYEAETILRNQQKYHAEITDEMIEKVLSLIQ
FRIPYVVGPLVNDHTASKFGWMERKSNESEIKPWNFDEVVDRSKSATQFIRRM TNKCSYLNEDV
LPKNSLLYQEMEVLNELNATQIRLQTDPKNRKYRMMPQIKLFAVEHIFKKYKTVSHSKFLEIML
NSNHRENFMNHGEKLSIFGTQDDKKFASKLSSYQDMTKIFGDIEGKRAQIEEIIQWITIFEDKK
ILVQKLKECYPELT SKQINQLKKLNYSGWGRLEKLLTHAYQGHSIIELLRHSDENFMEILTND
40 VYGFQNFIKEENQVQSNKIQHDIANLTTSPALKKGIWSTIKLVRELTSIFGEPEKIIMEFATE
DQQKGKKQKSRQLWDDNIKKNLKSVDEYKYIIDVANKLNNEQLQQEKLWLYLSQNGKCMYSG
QSIDLDALLSPNATKHVEVDHIFPRSF IKDDSIDNKVLVIKKMNQTKGDQVPLQFIQQPYERIA
YWKSLNKAGLISDSKLHKLMPKPEFTAMDKEGFIQRQLVETRQISVHVRDFLKEEYPNTKVI PMK
AKMVSEFRKKFDIPKIRQMNDAAHDAIDAYLNGVVYHGAQLAYPNVDLDFDNFKWEKVREKWKAL
45 GEFNTKQKSRELFFFKKLEKMEVSQGERLISKIKLDMNHFKINYSRKLANIPQQFYNQTA VSPK
TAEKYESNKSNEVVYKGLTPYQTYVVAIKSVNKKGKEKMEYQMIDHYVDFYKFQNGNEKELA
LYLAQRENKDEVLDAQIVYSLNKGDLLEYINNHPCYFVSRKEVINAKQFELTVEQQLSLYNVMNN

KETNVEKLLIEYDFIAEKVINEYHHYLSNKLKEKRVRTFFSESNQTHEDFIKALDELFKVVTAS
 ATRSDKIGSRKNSMTHRAFLGKGKDVKIAYTSISGLKTTKPKSLFKLAESRNEL

SEQ ID NO: 331

5 MAKILGLDLGTNSIGWAVVERENIDFSLIDKGVRIFSEGVKSEKGISSRAAERTGYRSARKIK
 YRRKLRYETLKVLSLNMCPLSIEEVEEWKKSFGKDYPLNPEFLKWLSTDEESNVNPFYFRDR
 ASKHVSLFELGRAFYHIAQRRGFLSNRLDQSAEGILEEHCPKIEAIVEDLISIDEISTNITDY
 FFETGILDSNEKNGYAKDLDEGDKKLVSLYKSLAILKKNESDFENCKSEIERLNKKDVLGKV
 KGKIKDISQAMLDGNYKTLGQYFYSLSKEKIRNQYTSREEHYLSEFITICKVQGIDQINEEEK
 10 INEKKFDGLAKDLYKAIFQRP LKSQKGLIGKCSFEKSKSRCAISHPDFEYRMWYTLNTIKIG
 TQSDKKLRFLTQDEKLKLVKPFYRKNDFNFDVLAKELIEKGSSFGFYKSSKKNDDFFYWFNYKPT
 DTVAACQVAASLKNAIGEDWKTFSKYQTINSNKEQVSRTVDYKDLWHLLTVATSDVLYEFAI
 DKLGLDEKNAKAFSKTKLKKDFASLSLSAINKILPYLKEGLLYSHAVFVANINIENIVDENIWKDE
 KQRDYIKTQISEIENYTLKSRFEIINGLLKEYKSENEDEGKRVYYSKEAEQSFENDLKKKLV
 15 FYKSNEIENKEQOETIFNELLPIFIQQLKDYEFIKIQRLDQKVLIFLKGKNETGQIFCTEEKGT
 AEEKEKKIKNRLKKLYHPSDIEKFKKKI KDEFNGNEKIVLGSPLTPSIKNPMAMRALHQLRKVL
 NALILEGQIDEKTIHHEMARELNDANKRKGIDYQNDNKKFREDAIKEIKKLYFEDCKKEVEP
 TEDDILRYQLWMEQNRSEIYEEGKNISICDIIGSNPAYDIEHTIPRSRSQDNSQMNKTLCSQRF
 NREVKKQSMPIELNNHLEILPRIAHWKEEADNLTREIEIISRSIKAAATKEIKDKKIRRRHYLT
 20 LKRDYLQGGKYDRFIWEEPVGFKNSQIPDTGIIITKYAQAYLKS YFKKVESVKGGMVAEFRKIWG
 IQESFIDENGMKHYKVKDRSKHTHTIDAITIACMTKEKYDVLAAHAWTLEDQQNKKEARSIEA
 SKPWKTFKEDLLKIEEILVSHYTPDNVKKQAKKIVRVRGKKQFVAEVERDVNGKAVPKKAASG
 KTIYKLDGEGKKLPRLQQGDTIRGSLHQDSIYGAIKNPLNTDEIKYVIRKDLESIKGSDVESIV
 DEVVKEKIKEAIANKVLLSSNAQQKNKLVGTVWMNEEKRIAINKVRIYANSVKNPLHIKEHSL
 25 LSKSKHVHKQKVYQNDENYAMAIYELDGKRD FELINIFNLAKLIKQGGFYPLHKKKEIKGKI
 VFVPIEKRNKRDVVLKRGQQVVFYDKEVENPKDISEIVDFKGRIYIIEGLSIQRIVRPSGKVDE
 YGVIMLRYFKEARKADDIKQDNFKPDGVFKLGENKPTRKMNHQFTAFVEGIDFKVLPSGKFEKI

SEQ ID NO: 332

30 MEFKKVLGLDIGTNSIGCALLSLPKSIQDYGKGGRLEWLT SRVIPLDADYMKAFIDGKNGLPQV
 ITPAGKRRQKGRSRLKHRYKLRRSRLIRVFKTLNWLPEDFPLDNPKRIKETISTEGKFSFRIS
 DYVPISDESYREFYREFGY PENEIEQVIEEINFRRKTKGKNKNPMIKLLPEDWVVYYLRKKALI
 KPTTKEELIRIIYLFNQRRGFKSSRKDLTETAILDYDEFAKRLAEKEKYS AENYETKFVSITKV
 KEVVELKTDGRKGKRFKVILED SRIEPIEIERKEKPDWEGKEYTFLVTQKLEKGKFKQNKPD
 35 PKEEDWALCTTALDNRMGSKHPGEFFDEL LKAFKEKRGYKIRQYPVNRWRYKKELEFIWTKQC
 QLNPELNNLNINKEILRKLATVLYPSQS KFFGPKIKEFENS DVLHIISEDIIYYQRDLKSQKSL
 ISECRYEKRKGIDGEIYGLKCIPKSSPLYQEFRIWQDIHNIKVIRKESEVNGKKKINIDETQLY
 INENIKEKLFELFNSKDSLSEKDILELISLNIINS GIKISKKEEETHRINLFANRKELKGNET
 KSRYRKVFKKLGFDGEYILNHPSKLNRLWHS DYSDYADKEKTEKSILSSLGWKNRNGKWEKSK
 40 NYDVFNLPLEVAKAIANLPPLKKEYGSYSALAIRKMLVVMRDGKYWQHPDQIAKDQENTSLMLF
 DKNLIQLTNNQRKVLNKYLLTLAEVQKRSTLIKQKLNEIEHNPYKLELVSDQDLEKQVLKSFLE
 KKNESDYLKGLKTYQAGYLIYGKHSEKDVPIVNSPDELGEYIRKKLPNNSLRNPIVEQVIRETI
 FIVRDVWKSFGIIDEIHIELGRELKNNSEERKKTSESQEKNFQEKERARKLLKELLNSSNFEHY
 DENGKIFSSFTVNPNDSPLDIEKFRIWKNQSGLTDEELNKKLKDEKIPTIEIEVKKYILWLTQ
 45 KCRSPYTGKIIPLSKLFD SNVYEIEHIIPRSKMKNDS TNNLVICELGVNKA GDRLAANFISES
 NGKCKFGEVEYTLKYGDYLYCKDTFKYQKAKYKNLLATEPPEDFIERQINDTRYIGRKLAE
 LTPVVKDSKNIIFTIGSITSELKITWGLNGVWKDILRPRFKRLESIINKKLIFQDEDDPNKYHF

DLSINPQLDKEGLKRLDHRHHALDATIIAATTREHVRYLNSLNAADNDEEKREYFLSLCNHKIR
DFKLPWENFTSEVKSLLSCVVSYKESKPILSDPFNKYLKWEYKNGKWQKVFAIQIKNDRWKAV
RRSMFKPIGTVWIKKIKEVSLKEAIKIQAIWEEVKNDPVRKKKEKYIYDDYAQKVIKIVQEL
GLSSMRKQDDEKLNKFINEAKVSAGVNKNLNTTNKTIYNLEGRFYEKIKVAEYVLYKAKRMPL
5 NKKEYIEKLSLQKMFNDLPNFILKESILDNYPEILKELESDNKYIIEPHKKNNPVNRLLEHIL
EYHNNPKEAFSTEGLEKLNKKAINKIGKPIKYITRLDGDINEEIEIFRGAVFETDKGSNVYFVMY
ENNQTKDREFLKPNPSISVLKAIHKNKIDFFAPNRLGFSRIILSPGDLVYVPTNDQYVLKDN
SSNETIINWDDNEFISNRIYQVKKFTGNSCYFLKNDIASLILSYSASNGVGEFGSQNISEYSVD
DPPIRIKDVCIKIRVDRLGNVRPL

SEQ ID NO: 333

MKHILGLDLGTNSIGWALIERNIEEKYGKIIIGMGSRIVPMGAELSKFEQQAQTKNADRRTNRG
ARRLNKRYKQRRNKLIYILQKLDMLPSQIKLKEDFSDPNKIDKITILPISKKEQLTAFDLVSL
RVKALTEKVGLEDLGKIIYKYNQLRGYAGGSLEPEKEDIFDEEQSKDKKNKSFIAFSKIVFLGE
15 PQEEIFKNKKLNRRRAIIVETEEGNFEGSTFLENIKVGDSLELLINISASKSGDTITIKLPNKTN
WRKKMENIENQLKEKSKEMGREFYISEFLELLKENRWAKIRNNTILRARYESEFEAIWNEQVK
HYPFLENLDKKTILIEIVSFIFPGEKESQKKYRELGLEKGLKYIIKNQVVFYQRELKDQSHLISD
CRYEPNEKAIKASHPVFQYKQVWEQINKLIVNTKIEAGTNRKGEKKYKIDRPIPTALKWIFE
ELQNKKEITFSAIFKKLKAFFDLREGIDFLNGMSPKDKLKGNETKLQLQKSLGELWDVLGLDSI
20 NRQIELWNILYNEKGNEYDLTSDRTSKVLEFINKYGNIVDDNAEETAIRISKIKFARAYSSLS
LKAVERILPLVRAGKYFNNDFSQQQLQSKILKLLNENVEDPFAKAAQTYLDNNQSVLSEGGVNS
IATILVYDKHTAKEYSHDELYKSYKEINLLKQGDLRNPLVEQIINEALVLIRDIWKNYGIKPNE
IRVELARDLKNSAKERATIHKRKNKDNTINNLIKETLVKNKKELSLANIEKVKLWEAQRHLSPY
TGQPIPLSDLFDKEKYDVDHIIIPISRYFDDSFNTKVISEKSVNQEKANRTAMEYFEVGSCLKYSI
25 FTKEQFIAHVNEYFSGVKRKNLLATSIPEDPVQRQIKDTQYIAIRVKEELNKIVGNENVKTTTG
SITDYLRNHWGLTDKFKLLKERYEALLESEKFLEAEYDNYKKDFDSRKKEYEEKEVLFEEQEL
TREEFIKEYKENYIRYKKNKLIIGWSKRIDHRHHAIDALIVACTEPAHIKRLNDLNKVLQDWL
VEHKSEFMPNFEGSNSELLEEILSLPENERTEIFTQIEKFRAIEMPWKGFPEQVEQKLKEIIS
HKPKDKLLLQYNKAGDRQIKLRGQLHEGTLYGISQGKEAYRIPLTKFGGSKFATEKNIQKIVSP
30 FLSGFIANHLKEYNNKKEEAFSAEGIMDLNNKLAQYRNEKGELKPHTPISTVKIYYKDPSKNKK
KKDEEDLSLQKLDREKAFNEKLYVKTGDNYLFAVLEGEIKTKKTSQIKRLYDIISFFDATNFLK
EEFRNAPDKKTFDKDLLFRQYFEERNKAKLLFTLKQGDFVYLPNENEEVILDKESPLYNQYWG
LKERGKNIYVVQKFSKKQIYFIKHTIADIKKDVEFGSQNCYETVEGRSIKENCCKLEIDRLGN
IVKVIKR

SEQ ID NO: 334

MHVEIDFPHFSGDShLAMNKNEILRGSSVLYRLGLDLGSNSLWGFVTHLEKRGDRHEPVALGP
GGVRIFPDGRDPQSGTSNAVDRRMARGARKRRDRFVERRKELIAALIKYNLLPDDARERRALEV
LDPYALRKALTDTLPAHHVGRALFHLNQRGRGFQSNRKTDSKQSEDGAIKQAASRLATDKGNET
40 LGVFFADMHLRKSIEDRQTAIRAEVLRLGKDHLTGNAKKKIWAKVRKRLFGDEVLPADAPHG
RARATITGTKASYDYYPTRDMLRDEFNAIWAGQSAHHATITDEARTEIEHIIFYQRPLKPAIVG
KCTLDPATRPFKEDPEGYRAPWSHPLAQRFRILSEARNLEIRDTGKGSRLTKEQSDLVVAALL
ANREVKFDKLRITLLKLPAAEFNLESRRRAALDGDQTAARLSKKGFNKAWRGFPFERQIAIVA
RLEETEDENELIAWLEKECALDGAAAARVANTTLPDGHCRLLGLRAIKKIVPIMQDGLDEDGVAG
45 AGYHIAAKRAGYDHAKLPTGEQLGRLPYQGWLQDAVVGSGDARDQKEKQYGFNPPTVHIGLG
QLRRVVNDLIDKYGPPTIEISIEFTRALKLSEQQKAERQREQRRNQDNKARAEELAKFGRPANP
RNLLKMRLWEELAHDPDRKCVYTGEQISIERLLSDEVDIDHILPVAMTLDDSPANKIICMRYA

NRHKRKQTPSEAFGSSPTLQGHRYNWDDIAARATGLPRNKRWRFDANAREEFDKRGGFLLARQLN
ETGWLARLAKQYLGAVTDPNQIWVVPGRLTSMRLRGKWGLNGLLPSDNYAGVQDKAEFFLASTDD
MEFSGVKNRADHRHHAIDGLVTALTDRSLLWKMAYDEEHEKFVIEPPWPTMRDDLKAALEKM
VVSHKPDHGIEGKLHEDSAYGFVKPLDATGLKEEEAGNLVYRKAIESLNENEVDRIQDLRTI
5 VRDHVNVEKTKGVALADALRQLQAPSDDYPQFKHGLRHVRILKKEKGDYLVPIANRASGVAYKA
YSAGENFCVEVFETAGGKWDGEAVRRFDANKKNAGPKIAHAPQWRDANEGAKLVMRIHKGDLIR
LDHEGRARIMVVHRLDAAAGRFKLADHNETGNLDRHATNNDIDPFRWLMASYNLTKKLAAPV
RVDELGRVWRVMPN

10 SEQ ID NO: 335

METTLGIDLGINSIGLALVDQEEHQILYSGVRIFPEGINKDTIGLGEKEESRNATRRAKRQMR
QYFRKKLRKAKLLELLIAYDMCPLKPEDVRRWKNWDKQQKSTVRQFPDTPAFREWLKQNPYELR
KQAVTEDVTRPELGRILYQMIQRRGFLSSRKGKEEGKIFTGKDRMVGIDETRKNLQKQTLGAYL
YDIAPKNGEKYRFRTERVRARYTLRDMYIREFEIIWQRQAGHLGLAHEQATRKNIFLEGSATN
15 VRNSKLITHLQAKYGRGHVLIEDTRITVTFQLPLKEVLGGKIEIEEEQLKFKSNEVSLFWQRPL
RSQKSLLSKCVFEGRNFDYPVHQWIIAGPTAPLSHPEFEEFRAYQFINNIYKGNEHLTAIQ
REAVFELMCTESKDFNFEEKIPKHLKLFEKFNDDTTKVPACTTISQLRKLFPHPVWEEKREEIW
HCFYFYDDNTLLFEKLQKDYALQTNDEKIKKIRLSESYGNVSLKAIRRNINPYLKKGYAYSTAV
LLGGIRNSFGKRFEYFKEYEPEIEKAVCRILKEKNAEGEVIRKIKDYLVHNRFGFAKNDRAFQK
20 LYHHSQAITTQAQKERLPETGNLRNPVQQLNELRRTVNKLLATCREKYGPSFKFDHIHVEMG
RELRSSKTEREKQSRQIRENEKKNEAAKVLALEYGLKAYRDNIQKYLLEYKEIEEKGGTVCCPYT
GKTLNISHTLGSDNSVQIEHIIPYSISLDDSLANKTLCDATFNREKGELTPYDFYQKDPSPKEW
GASSWEEIEDRAFRLPYAKAQRFIRRKQESNEFISRQLNDTRYISKKAVEYLSAICSDVKAF
PGQLTAELRHLWGLNNILQSAPDITFPLPVSATENHREYVITNEQNEVIRLFPKQGETPRTEK
25 GELLLTGEVERKVFRCKGMQEFQTDVSDGKYWRRIKLSSSVTWSPLFAPKPISADGQIVLKGRI
EKGVFVCNQLKQKLKTGLPDGSYWISLPVISQTFKEGESVNNSKLTSQQVQLFGRVREGIFRCH
NYQCPASGADGNFWCTLDTDTAQPAFTPIKNAPPGVGGGQIILTGDDVDKGFHADDLHYELP
ASLPKGKYYGIFTVESCDPTLPIELSAKPTSKGENLIEGNIWVDEHTGEVRFDPKKNREDQRH
HAIDAIVIALSSQSFLQRLSTYNARRENKKRGLDSTEHFSPWPFGAQDVRQSVVPLLVSYKQN
30 PKTLCKISKTLTKDGKKIHSCGNAVGRQLHKTETVYGQRTAPGATEKSYHIRKDIRELKTSKHIG
KVVDITIRQMLLKLHLENYHIDITQEFNIPSNAFFKEGVYRIFLPNKHGEPVPIKKIRMKEELG
NAERLKDNIQYVNPNNHHVMIYQDADGNLKEEIVSFWSVIERQNQGQPIYQLPREGRNIVSI
LQINDTFLIGLKEEEPEVYRNDLSTLSKHLRYVQKLSGMYTFRHHLASTLNNEEEFRIQSLE
AWKRANPVKVQIDEIGRITFLNGPLC

35 SEQ ID NO: 336

MESSQILSPIGIDLGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRATRHRV
RNKKRNQFVKRVALQLFQHILSRDLNAKEETALCHYLNNGYTYVDTDLDEYIKDETTINLLKE
LLPSESEHNFIDWFLQKMQSSEFRKILVSKVEEKDDKELKNAVKNKNFITGFEKNSVEGHRH
40 RKVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSNLQWKNLHRYLAKNPKQFDEQTFGNE
FLRMLKNFRHLKGSQESLAVRNLIQQLQSQDYISILEKTPPEITIPPYEARTNTGMEKDQSL
LNPEKLNLYPNWRNLIPGIIDAHPFLEKDLEHTKLDRKRISPSKQDEKRDYILQRYLDLN
KKIDKFKIKKQLSFLGQKQLPANLIETQKEMETHFNSSLVSVLIQIASAYNKEREDAAQGIWF
DNAFSLCELSNINPPRKQKILPLLVGAILSEDFINNKKDWAKFKIFWNTHKIGRTSLKSKCKEI
45 EEARKNSGNAFKIDYEEALNHPEHSNNKALIKIIQTIPDIIQAIQSHLGHNDSQLIYHNPFSL
SQLYTILETKRDGFHKNCVAVTCENYWRSQKTEIDPEISYASRLPADSVRPFDGVLARMMQRLA
YEIAMAKWEQIKHIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAIEKQNIQWEEKF

QRIINASMNICPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYL
 LEHLSPLYLKHQFGTDNVSDIKNFISQNVANIKKYISFHLLTPEQQKAARHALFLDYDDEAFKT
 ITKFLMSQQKARVNGTQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSKQEPK
 LVKSRQQSFPSHAIDATLTMSIGLKEFPQFSQELDNSWFINHLMPDEVHLNPVRSKEKYNKPN
 5 SSTPLFKDSL YAERFIPVWVKGETFAIGFSEKDLFEIKPSNKEKLF TLLKTYSTKNPGESLQEL
 QAKSKAKWLYFPINKTLALEFLHHYFHKIEIVTPDDTTVCHFINSRLRYTTKESITVKILKEPMP
 VLSVKFESSKKNVLGSFKHTIALPATKDWERLFNHPNFLALKANPAPNPKEFNEFIRKYFLSDN
 NPNSDIPNNGHNIKPKQKHKAVRKVFSLPVIPGNAGTMMRIRRKDNKGQPLYQLQTIDDTPSMGI
 QINEDRLVKQEVLM DAYKTRNLSTIDGINNSEGQAYATFDNWLTLVPVSTFKPEIIKLEMKPHSK
 10 TRRYIRITQSLADFIKTIDEALMIKPSDSIDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTG
 KIVTYEFESDSTPQWIoTLYVTQLKKQP

SEQ ID NO: 337

MKKIVGLDLGTNSIGWALINAYINKEHLYGIEACGSRIIPMDAAILGNFDKGNSSISQTADRTSY
 15 RGIRRLRERHLLRRERLHRILDLLGFLPKHYSDSLNRYGKFLNDIECKLPWVKDETGSYKFIFQ
 ESFKEMLANFTEHHPILIANNNKKVPYDWTIYYLRKKALTQKISKEELAWILLNFNQKRGYYQLR
 GEEEEPTPNKLVEYYSLKVEKVEDSGERKKGKDTWYNVHLENGMIYRRTSNIPLDWEGKTKEFIVT
 TDLEADGSPKKDKEGNIKRSFRAPKDDDWTLIKKKTEADIDKIKMTVGAYIYDTLLQKPDQKIR
 GKLVRTIERKYYKNELYQILKTQSEFHEELRDKQLYIACLNELYPNNEPRRNSISTRDFCHLFI
 20 EDIIFYQRPLKSKKSLIDNCPYEENRYIDKESGEIKHASIKCIAKSHPLYQEFRLWQFIVNLRI
 YRKETDVDVTQELLPTeadyVTLFEWLNKKEIDQKAFFKYPPFGFKTTSNYRWNVEDKPY
 CNETHAQI IARLGKAHIPKAFLSKEKEETLWHILYSIEDKQEI EKALHSFANKNNLSEEFIEQF
 KNFPFPFKKEYGSYSAKAIAKLLPLMRMGKYWSIENIDNGTRIRINKIIDGEYDENIRERVRQKA
 INLTDITHFRALPLWLACYLVYDRHSEVKDIVKWKTPKDIDLKSLKSFQHSRLNPIVEQVITET
 25 LRTVRDIWQQVGHIDEIHIELGREMKNPADKRARMSQQMIKNENTNLRIKALLTEFLNPEFGIE
 NVRPYSPSQDQLRIYEEGVLNSILELPEDIGIILGKFNQDTLKRPRTRSEILRYKLWLEQKYR
 SPYTGEMIPLSKLF TPAYEIEHIIPQSRFYDDSLSNKVICESEINKLKDRSLGYEFIKNHHGEK
 VELAFDKPVEVLSVEAYEKL VHESYSHNRSKMKLLMEDIPDQFIERQLNDSRYISKVVKSLLS
 NIVRENEQEAI SKNVIPCTGGITDRLKKDWGINDVWNKIVLPRFIRLNELTESTRFTSINTNN
 30 TMIPSMPL ELQKGFNKKRIDHRHHAMD AIIACANRNIVNYLNNVSASKNTKITRRDLQTLCH
 KDKTDNNGNYKWVIDKPWETFTQDTLTALQKITVSFKQNL RVINKTTNHYQHYENGKKIVSNQS
 KGDSWAIRKSMHKETVHGEVNLRMIKTVSFNEALKKPQAIVEMDLKKKILAMLELG YDTKRKN
 YFEENKDTWQDINPSKIKVYYFTKETKDRYFAVRKPIDTSFDKKKIKESITDTGIQQIMLRHLE
 TKDNDPTLAFSPDGIDEMNRNII LNLNGKKHQPIYKVRVYEKA EKFTVGQKGNKRTKFVEAAKG
 35 TNLFFAIYETEEIDKDTKKVIRKRSYSTIPLNVVIERQKQGLSSAPEDENG NLPKYILSPNDLV
 YVPTQEEINKGEVVMPI DRDRIYKMVDSSGITANFIPASTANLIFALPKATAEIIYCNGENCIQN
 EYGIGSPQSKNQKAITGEMVKEICFPKVDRLGNI IQVGSCILT N

SEQ ID NO: 338

MSRSLTFSFDIGYASIGWAVIASASHDDADPSVCGCGTVLFPKDDCQAFKRREYRRLRRNIRSR
 40 RVRIERIGRLLVQAQIIITPEMKETSGHPAPFYLA SEALKGHRTLAPIELWHVLRWYAHNRGYDN
 NASWSNSLSEDDGGNGEDTERVKHAQDLMDKHGTATMAETICRELKLEEGKADAPMEVSTPAYKN
 LNTAFPR LIVEKEVRRILELSAPLIPGLTAEIIELIAQHHP LTTEQRGVLLQHGIKLARRYRGS
 LLFGQLIPRFDNRIISRCPVTWAQVYEAELKKGNSEQSARERA EKLSKVPTANCPEFYEYRMAR
 45 ILCNIRADGEPLSAEIRRELMNQARQEGKLTASLEKAISSRLGKETETNVSNYFTLHPDSEEA
 LYLNP AVEVLQRSIGIGQILSPSVYRIANRLRGKSVTPNYLLNLLKSRGESGEALEKKIEKES
 KKKEADYADTPLPKPYATGRAPYARTVLKKVVEEILDGEDPTRPARGEAHPDGELKAHDGCLYC

LLD TDSSVNQH QKERR LD TMTNNHLVRHRMLILDRL LKDLIQDFADGQKDRI SRVCVEVGKELT
 TFSAMDSKKIQRELTLRQKSH TDAVNRLKRKLPGKALSANLIRKCR IAMDMNWTCPFTGATYGD
 HELENLELEHIVPHSFRQSNALSSVL TWPGVNR MKGQRTGYDFVEQE QENPVDPKPNLHICSL
 NNYRELVEKLDDKKGHEDDRRRKKR KALLMVRGLSHKHQSQNHEAMKEIGMTEGMMTQSSHL M
 5 KLACKSIKTS LPA DAHIDMIPGAVTAEVRKA WDVFGVFKELCPEAADPD SGKILKENLRS LTHLH
 HALDACVLGLIPYIIPAHHNGLLRRVLAMRRIPEKLIPQVRPVANQRHYVLNDDGRMMLRDL SA
 SLKENIREQLMEQRV IQHVPADMGGALLKETMQRVLSVDGSGEDAMVSLSKKDKGKKEKNQVKA
 SKLVGVFPEGPSKLKALKA AIEIDGNYGVALDPKPVVIRHIKVFKRIMALKEQNGGKPVRI LKK
 10 GMLIHLTSSKDPKHAGVWRIESIQDSKGGVKLDLQRAHCAVPKNKTHECNWREVDLISLLKKYQ
 MKRYPTSYTGTPR

SEQ ID NO: 339

MTQKVLGLDLGTNSIGSAVRNLDLSDDLQWQLEFFSSDIFRSSVNKESNGREYSLAAQRSAHR R
 SRGLNEVRRRLWATLNL LKHGFCPMSSES LMRWCTYDKRKGLFREYPIDDKDFNAWILLDFN
 15 GDGRPDYSSPYQLRRELVT RQFDFEQPIERYKLGRALYHIAQHRGFKSSKGETLSQQETNSKPS
 STDEIPDVAGAMKASEEKL SKGLSTYMKEHNLLTVGA AFAQLEDEGVRVRNNNDYRAIRSQFQH
 EIETIFKFQQGLSVESELYERLISEKKNVGTIFYKRPLRSQRGNVGKCTLERSKPRCAIGHPLF
 EKFRATL INNIKVRMSVDTLDEQLPMKLRLDLYNECFLAFVRTEFKFEDIRKYLEKRLGIHFS
 YNDKTINYKDSTSVAGCPITARFRKMLGEEWESFRVEGQKERQAHSKNNISFHRVSYSIEDIWH
 20 FCYDAEEPEAVLAF AQETLRLERKKA EELVRIWSAMPQGYAMLSQKAIRNINKILMLGLKYSDA
 VILAKVPELV DVSDEELLSIAKDYYLVEAQVNYDKRINSIVNGLIAKYKSVSE EYRFADHNYEY
 LLDSEDEKDIIRQIENSLGARRWSLMDANEQTDILQKVRDRYQDFFRSHERKFVESPKLGESFE
 NYLTKKFPMVEREQWKKLYHPSQITIYRPVSVGKDRSVLRLGNPDIGA IKNPTVLRVLNTLRRR
 VNQLLDDGVISPDETRVVETARELNDANRKWALDTYNRIRHDENEKIKKILEEFYPKRDGIST
 25 DDIDKARYVIDQREVDYFTGSKTYNKDIKKYKFWLEQGGQCMYTGR TINLSNLFDPNAFDIEHT
 IPESLSFDSSDMNLTLCDAHYNRFIKKNHIPTDMPNYDKAITIDGKEYPAITSQLQRWVERVER
 LNRNVEYWKGQARRAQNKDRKDQCMREHLWKMELEYWKKKLERFTVTEVTDGFKNSQLVDTRV
 ITRHAVLYLKSIFPHVDVQRGDVTAKFRKILGIQSVDEKKDRSLHSHHAIDATTLTIIPVSAKR
 DRMLELFAKIEEINKMLSFSGSEDRTGLIQELEG LKNKLQMEVKVCRIGHNVSEIGTFINDNII
 30 VNHHIKNQALTPVRRRLRKKGYIVGGVDNPRWQTGDALRGEIHKASYGAI TQFAKDDEGKVL M
 KEGRPQVNPTIKFVIRRELKYKKSAA DSGFASWDDLGAIVDKELFALMKGQFPAETSFKDACE
 QGIYMIKKGKNGMPDIKLHHIRHVRCEAPQSG LKIKEQTYKSEKEYKRYFYAAVGDL YAMCCYT
 NGKIREFRIYSLYDVSCHRKSDIEDIPEF ITDKKGNRLMLDYKLRTGDMILLYKDNPAELYDLD
 NVNLSRRLYKINRFESQSNLVLMT HHLSTSKERGRSLGKTVDYQNL PESIRSSVKS LNFLIMGE
 35 NRDFVIKNGKIIIFNHR

SEQ ID NO: 340

MLVSPISVDLGKNTGFFSFTDSL DNSQSGTVIYDESFVLSQVGRRSKRHSKRNNLRNKL VKRL
 FLLILQEHHGLSIDVLPDEIRGLFNKRGYTYAGFELDEKKKDALES DTLKEFLSEKLQSIDRDS
 40 DVEDFLNQIASNAESFKDYKKGF EAVFASATHSPNKKLELKD ELKSEYGENAKELLAGLRVTKE
 ILDEFDKQENQGNLPRAKYFEELGEYIATNEKVKSFFDSNSLKLTDMTKLIGNISNYQLKELRR
 YFNDKEMEKGDIWIPNKLHKITERFVRSWHPKNDADRQRRAE LMKDLKSKEIMELLTTTEPVM T
 IPPYDDMNNRGAVKCQTLRLNEEYLDKHLPNWRDIAKRLNHGKFND DLADSTVKGYSEDSTLLH
 RLLDTSKEIDIYELRGKKPNELLVKT LGQSDANRLYGFAQNYELIRQKV RAGIWPVKNKDDS
 45 LNLEDNSNMLKRCNHNPPHKKNQIHNLVAGILGVKLDEAKFAEF EKELWSAKVG NKKLSAYCKN
 IEELRKTHGNTFKIDIEELRKKDPAELSKEEKAKLRLTDDVILNEWSQKIANFFDIDDKHRQRF
 NNLFSMAQLHTVIDTPRSGFSSTCKRCTAENRFRSETAFYNDETGEFHKKATATCQRLPADTQR

PFSGKIERIDKLGELAKIKAKELEGMEAKEIKVPIILEQNAFEYEESLRKSKTGSNDRVINS
 KKDRDGKKLAKAKENAEDRLKDKDKRIKAFSSGICPYCGDTIGDDGEIDHILPRSHTLKIYGTV
 FNPEGNLIYVHQCNQAKADSIYKLS DIKAGVSAQWIEEQVANIKGYKTF SVLSAEQQKAFRYA
 LFLQNDNEAYKKVVDWLRTDQSARVNGTQKYLAKKIQEKLTKMLPNKHLSEFELADATEVSEL
 5 RRQYARQNPLLAKEKQAPSSHAIDAVMAFVARYQKVKFDGTPPNADEVAKLAMLD SWNPASNE
 PLTKGLSTNQKIEKMIKSGDYGQKNMREVFSGKIFGENAIGERYKPIVVQEGGYIIGYPATVKK
 GYELKNCKVVT SKNDIAKLEKI IKNQDLISLKENQYIKIFSINKQTI SELSNRYFNMNYKNLVE
 RDKEIVGLLEFIVENC RYYTKKVDVKFAPKYIHETKYPFYDDWRRFDEAWRYLQENQNKTSKD
 RFVIDKSSLNEYYPDKNEYKLDVDTQPIWDDFCRWYFLDRYKTANDKKSIRIKARKTFSLAE
 10 SGVQGKVFRAKRKIPTGYAYQALPMDNNVIAGDYANILLEANSKTL SLVPKSGISIEKQLDKKL
 DVIKKT DVRLAIDNNSFFNADFDT HGIRLIVENTS SVKVGNFPI SAIDKSAKRMIFRALFEKEK
 GKRRKKTTISFKESGPVQDYLKVFLKKIVKIQLR TDGSISNIVVRKNAADFTLSFRSEHIQKLL
 K

15 SEQ ID NO: 341
 MAYRLGLDIGITSVGWAVVALEKDESGLKPVRIQDLGVRIFDKAEDSKTGASLALPRREARSAR
 RRTRRRRHRLWRVKRLLEQHGISMEQIEALYAQRTSSPDVYALRVAGLDRCLIAEEIARVLIH
 IAHRRGFQSNRKSEIKDS DAGKLLKAVQENENLMQSKGYRTVAEMLVSEATKTDAEGKLVHGKK
 HGYVSNVRNKAGEYRHTVSRQAI VDEVKIFAAQRALGNDVMSEELDSYKILCSQRNFDDGP
 20 GGDSFYGHGSPDGVRSIYERMVGSCTFETGEKRAPRSSYSFERFQLLTKVNLRIYRQQED
 GG RYPCELTQTERARVIDCAYEQTKITYGKLRLKLLDMKDTE SFAGLTYGLNRSRNTEDTVFVE
 MKFYHEVRKALQ RAGVFIQDLSIETLDQIGWILSVWKSDDNRRKKLSTLGLSDNVIEELLPLNG
 SKFGHLSLKAIRKILP FLEDGYSYDVACELAGYQFQGKTEYVKQRLPLPLGEGEV TNPVVRAL
 SQAIVVNAVIRKHGSPESIHELARELSKNLDERRKIEKAQKENQKNNEQIKDEIREILGSAH
 25 VTGRDIVKYKLFKQQQEF CMYSGEKLDVTRLFEPGYAEVDHIIPYGISFDDSYDNKVLVKTEQN
 RQKGNRTPLEYLRDKPEQKAKFIALVESIPLSQKKKNHLLMDKRAIDLEQEGFRERNLSDTRYI
 TRALMNHIQAWLLFDETA STRSKRVVCVNGAVTAYMRARWGLTKDRDAGDKHHAADAVVACIG
 DSLIQRVTKYDKFKRNALADRNRVYQQVSKSEGITQYVDKETGEVFTWESFDERKFLPNEPLEP
 WPFPRDEL LARLSDDPSKNIRAI GLLTYSETE QIDPIFVSRMPTRKVTGA AHKETIRSPRIVKV
 30 DDNKGTEIQVVVSKVALTELKLTGDGEIKDYFRPEDDPRLYNTLRERLVQFGGDAKAAFKEPVY
 KISKDGSVRTPVRKVKIQEKLTLGVPVHGGRGIAENGGMVRIDVFAKGGKYFVPIYVADV LKR
 ELPNRLATAHKPYSEWRVVDSDYQFKFSLYPNDAMIKPSREVDITYKDRKEPVGCRIMYFVSA
 NIASASISLRTHDNSGELEGLGIQGLEVF EKYVVGPLGDTHPVYKERRMPPFRVERKMN

35 SEQ ID NO: 342
 MPVLSPLSPNAAQGRRRWSLALDIGEGSIGWAVAEVDAEGRVLQLTGTGVTLFPSAWSNENGTY
 VAHGAADRAVRGQQQRHDSRRRLAGLARLCAPVLERSPEDLKDLTRTPPKADPRAIFFLRADA
 ARRPLDGP ELFRVLHHMAAHRGIRLAELQEVDPPPESDADDAAPAATEDEDGTRRAAADERAFR
 RLMAEHMHRHGTQPTCGEIMAGRLRETPAGAQPVTRARDGLRVGGGVAVPTRALIEQEFD AIRA
 40 IQAPRHPDLPWDSLRLVLVDQAPIAVPPATPCLFLEELRRRGETFQGRITITRE AIDRGLTV DPL
 IQALRIRET VGNLRLHERITEPDGRQRYVPRAMP ELGLSHGELTAPERDTLVRALMHD PDGLAA
 KDGRIPYTRLRKLIGYDN SPVCF AQERDTSGGGITVNP TDPLMARWIDGWVDLPLKARS LYVRD
 VVARGADSAALARLLAEGA HGVPVAAAAVPAATAA ILES DIMQPGRYSVCPWAAEA ILDWAN
 APTEGFYDVTRGLFGFAPGEIVLEDLRRRARGALLAHLPR TMAAARTPNRAAQQRG PLPAYESVI
 45 PSQLITSLRAHKGRAADWSAADPEERNPFLRTWTGNAATDHILNQVRKTANEVITKYGNRRGW
 DPLPSRITVELAREAKHGVI RRNEIAKENRENEGRRKESAA LD TFCQDNTVSWQAGGLPKERA
 ALRLRLAQRQEFFCPYCAERP KLRATDLFSPAETEIDHVIERRMGGD GPDNLVLAHKDCNNAKG

KKTPHEHAGDLLDSPALAAALWQGWRKENADRLKKGKHKARTPREDKDFMDRVGWRFEEDARAKA
 EENQERRGRMLHDTARATRLARLYLAAAVMPEDPAEIGAPPVETPPSPEDPTGYTAIYRTISR
 VQPVNGSVTHMLRQRLQORDKNRDYQTHHAEDACLLLLAGPAVVQAFNTEAAQHGAADAPDDR
 5 DLMPSTDAYHQRRARALGRVPLATVDAALADIVMPESDRQDPETGRVHWRLTRAGRGLKRRID
 DLTRNCVILSRPRPSETGTPGALHNATHYGRREITVDGRTDTVVVTQRMNARDLVALLDNAKIV
 PAARLDAAAPGDTILKEICTEIAADRHDRVVDPEGTHARRWISARLAALVPAHAEAVARDIAELA
 DLDALADADRTPEQEARRSALRQSPYLGRAISAKKADGRARAREQEILTRALLDPHWGPRGLRH
 LIMREARAPSLVRIRANKTDAFGRVPVDAAVVVKTDGNAVSQLWRLTSVVTDDGRRIPLPKPIE
 KRIEISNLEYARLNLGLDEGAGVTGNNAPRPLRQDIDRLTPLWRDHGTAPGGYLGTAVGELEDK
 10 ARSALRGKAMRQTLTDAGITAEAGWRLDSEGAVCDLEVAKGDTVKKDGKTYKVGVITQGIFGMP
 VDAAGSAPRTPEDCEKFEEQYGIKPWKAKGIPLA

SEQ ID NO: 343

MNYTEKEKLFMYILALDIGIASVGWAILDKESETVIEAGSNIFPEASAADNQLRRDMRGAKRN
 15 NRRLKTRINDFIKLWENNLSIPQFKSTEIVGLKVRAITEEITLDELYLILYSYLKHRGISYLE
 DALDDTVSGSSAYANGLKLNALKELETHYPCEIQQERLNTIGKYRGQSQIINENGEVLDLSNVFT
 IGAYRKEIQRVFEIQKKYHPELTDEFCDGYMLIFNRKRKYEGPGNEKSRTDYGRFTTKLDANG
 NYITEDNIFEKLGKCSVYPDELRAAAASYTAQEYNVNLNDLNNLTINGRKLNEEKHEIVERIK
 SSNTINMRKIIISDCMGENIDDFAGARIDKSGKEIFHKFEVYNKMRKALLEIGIDISNYSREELD
 20 EIGYIMTINTDKEAMMEAFQKSWIDLSDDVKQCLINMRKTNGALFNKWQSFSLKIMNELIPEMY
 AQPKEQMTLLTEMGVTKGTQEEFAGLKYIPVDVVSIEDIFNPVRRSVRISFKILNAVLKKYKAL
 DTIVIEMPDRNSEEQKKRINDSQKLNEKEMEYIEKKLAVTYGIKLSPSDFSQKQLSLKLKLW
 NEQDGICLYSGKTIDPNDIINNPFQLEIDHIIPRSISFDDARSNKVLVYRSENQKKGNQTPYYY
 LTHSHSEWSFEQYKATVMNLSKKKEYAISRKKIQNLLYSEDTKMDVLKGFINRNINDTSYASR
 25 LVLNTIQNFFMANEADTKVKVIKGSYTHQMRCNLKLDKNRDESYSHHAVDAMLIGYSELGYEAY
 HKLQGEFIDFETGEILRKDMWDENMSDEVYADYLYGKKWANIRNEVVKAENVKYWHYVMRKS
 RGLCNQTIIRGTREYDGKQYKINKLDIRTKEGIKVFALAFSKKDSDRERLLVYLNDRTFFDLC
 KIYEDYSDAANPFVQYKETGDIIRKYSKKHNGPRIDKLKYKDGEVGACIDISHKYGFEEKGSKK
 VILESLVPYRMDVYYKEENHSYYLVGVKQSDIKFEKGRNVIDEEAYARILVNEKMIQPGQSRAD
 30 LENLGFKFKLSFYKNDIIEYEKDGKIYTERLVSRTPMPQRNYIETKPIDKAKFEKQNLVGLGKT
 KFIKKYRYDILGNKYSCSEEKFTSFC

SEQ ID NO: 344

MLRLYCANNLVNNVQNLWKYLLLLIFDKKIIIFLFKIKVILIRRYMENNNKEKIVIGFDLGVAS
 35 VGWSIVNAETKEVIDLGVRLFSEPEKADYRRAKRTTRRLLRKKFKREKFHKLILKNAEIFGLQ
 SRNEILNVYKDQSSKYRNILKLKINALKEEIKPSELVWILRDYLQNRGYFYKNEKLTDEFVSNS
 FPSKKLHEHYEKYGFFRGSVKLDNKLNDKKDAKEKDEEEESDAKKESEELIFSNKQWINEIVK
 VFENQSYLTESFKEEYLKLFNYVRPFNKGPGSKNSRTAYGVFSTDIDPETNKFKDYSNIWDKTI
 GKCSLFEEEIRAPKNLPSALIFNLQNEICTIKNEFTEFKNWWLNAEQKSEILKFVFTLNFNWKD
 40 KKYSDDKKFNKNLQDKIKKYLLNFALENFNLNNEILKNRDLNDTVLGLKGVKYKESNATADAA
 LEFSSLKPLYVFIKFLKEKKLDLNYLLGLENTTEILYFLDSIYLAISSSDLKERNEWFKKLLKE
 LYPKIKNNLEIIEINVEDIFEITDQEKFSFSKTHSLSREAFNHIIPLLLSNNEGKNYESLKHS
 NEELKKRTEKAELKAQQNQKYLKDNFLKEALVPLSVKTSVLQAIKIFNQIIKNFGKKYEISQVV
 IEMARELTKPNLEKLLNATNSNIKILKEKLDQTEKFDDFTKKKFIDKIENSVVFRNKLFLWFE
 45 QDRKDPYTQLDIKINEIEDETEIDHVIPYSKSADDSWFNKLKLVKKSTNQLKKNKTVWEYYQNES
 DPEAKWNKFVAWAKRIYLVQKSDKESKDNSEKNSIFKNKKPNLKFKNITKKLFDPYKDLGFLAR
 NLNDTRYATKVFRDQLNNYSKHHKSDDENKLFKVVCMNGSITSFLRKSMWRKNEEQVYRFNFWK

KDRDQFFHHAVDASIIAIFSLLTkTLyNKLrvYEsYdVQRREDGVYLINkETGEVKKADkDYWK
 DQHNFLKIRENAIEIKNVLNNVDFQnQVRYSRKANTKLNTQlFNETLYGVKEFENNfYKLEKVN
 LFSRKDLRKFILEDLNEESEKNKKNENGSRKRILTEKYIVDEILQILENEEFKDSKSDINALNK
 YMDSLPSKFSEFFSQDFINkCKKENSILTFDAIKHNDPKKVIKIKNLKFFREDATLKNKQAVH
 5 KDSKNQIKSFYEsYKCVGFIWLNKNNDLEESIFVPINSRVIHFgDKDKDIFDFDSYNKEKLLNE
 INLKRPENKKFNSINEIEFVKFVKPGALLNFENQqIYYISTLESSSLRAKIKLLNKMDKGKAV
 SMKKITNPDEYKIIeHVNPLGINLNWTKKLENNN

SEQ ID NO: 345

10 MLMSKHVLGLDLGVGSIGWCLIALDAQGDPAEILGMGSRVVPLNNATKAIEAFNAGAAFTASQE
 RTARRTMRRGFARYQLRRYRLRRELEKVGMLPDAALIQLPLeLWELRERAATAGRRLTLPELG
 RVLCHINQKRGYRHVKSDAAAIvGDEGEKKKDSNSAYLAGIRANDEKLQAEHKTvGQYFAEQLR
 QNQSESPTGGISYRIKdQIFSRQCYIDEYDQIMAVQRVHYPDILTDEFIRMLRDEVIFMQRPLK
 SCKHLVSLCEFEKQERVMRVQqDDGKGWQLVERRVKFGPKVAPKSSPLFQLCCIEAVNNIRL
 15 TRPNGSPCDITPEERAKIVAhLQSSASLSFAALKKLLKEKALIADQLTSKSGLKGNSTRVALAS
 ALQPYPQYHLLDMELETRMMTVQLTDEETGEVTEREVAVVTDsYVRKPLYRLWHILYSIEERE
 AMRRALITQLGMKEEDLDGGLLDQLYRLDFVKPGYGNKSAKFICKLLPQLQqGLGYSEACAAGV
 YRHSNSPTSEEITERTLLEKIPLLQRNELRQPLVEKILNQMINLVNALKAEYGIDEVRVELARE
 LKMSREERERMARNNKDREERNKGVAAKIRECGLYPTKPRIQKYMLWKEAGRQCLYCGRSIEEE
 20 QCLREGGMEVEHIIPKSVLYDDSYGNKTCACRCNKEKGNRTALEYIRAKGREAEYMKRINDLL
 KEKKISYSKHQRLRWLKEDIPSDFLERQLRLTQYISRQAMAILQqGIRRVsASEGGVTARLRSL
 WGYGKILHTLNLDRYDSMGETERVSREGEATEELHITNWSKRMdHRHHAIDALVVACTRQSYIQ
 RLNRLSSEFGREDKKKEDQEAQEQqATETGRLSNLERWLTQRPHFSVRTVSDKVAEILISYRPG
 QRVVTRGRNIYRKkMADGREVSCVQRGVLVPRGELMEASfYGKILSQGRVRIvKRYPLHDLKGE
 25 VVDPHLRELITTYNqELKSREKGAPIPLCLDKDKKQEVRSVRcYAKTSLDKAIPMcFDEKGE
 PTAfVKSASNNHHLALYRTPKGKLVESIVTFWDaVDRARYGIPLVITHPREVMEQVLQRGDIPeQ
 VLSLLPPSDWVFVDSLQqDEMvVIGLSDEELQRALeAQNYRKISEHLYRVQKMSSSYVfRYHL
 ETSVADDKNTSGRIPKFHRVQSLKAYEERNIRKVRVDLLGRISLL

30 SEQ ID NO: 346

MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRtNRQGRRLARRKKHRRV
 RLNRLFEEsGLITDFTKISINLNPYQLRVKGLTDELSNEELFIAlKNMVKHrgISYLDdASDDG
 NSSVGdYAQIVKENSKQLETKTPGQIQlERYQTYGQLRGDFtVEKDGKKHRLINVFPTsAYRSE
 ALRILQTQQEFNPQITDEFINRYLEILTGKRKYHhGPGNEKSRTDYGRYRTsGETLDNIFGILI
 35 GKCTfYPDEFRAAKASYTAQEFNLLNDLNNLTVPtETKkLSKEQKNQIINyVKNEKAMGPakLF
 KYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKtLETLDIEQMDRETLDKLAyVLTlNTE
 REGIQEALeHEfADGSfSQKQVDELvQFRKANSSIFGKGWHNFsvKLMMELIPeLYETSEEQMT
 ILTRLGKQKTtSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAiKIVNAAIKEYGDFdNIvIEMAR
 ETNEDDEKKAIQKIqKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKiRLWHQqGERC
 40 LYTGKTISIhDLINNSNqFEVDHILPLSITFDDSLANKVLvYATANQEKGQRTPYQALDSMDDA
 WsFRELKAfVRESKTLsNKKKEYLLTEEDISKfDVRKKfIErNLVDTRYASRVVLNALQEHfRA
 HKIDTKVSVVRGQFTSQRRLHWGIEKTRDTYHHhAVDALIIAASSQLNLWKKQKNTLVsYSEDQ
 LLDIETGELISDDEYKESVfKAPYQHfVDTLKSKEfEDSILfSYQVDSKfNRKISDATIYATRQ
 AKVGKDKADETYVLGKIkdIYTQdGYDAFMKIYKKDKSKFLMYRHDPQTfEKVIEPILENYPNK
 45 QINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYyDSKLGNHIDITPKDSNNKVVLQ
 SVSPWRADVfFNKTTGKYEILGLKYADLQfEKGTGTyKISQEKYNDIKKKEGVDSdSEfKfTLY

KNDLLLVKDTETKEQQLFRLSRTMPKQKHVELKPYDKQKFEGGEALIKVLGNVANSQGCKKG
LGKSNISIIYKVRTDVLGNQHIIKNEGDKPKLDF

SEQ ID NO: 347

5 MNAEHGKEGLLIMEENFQYRIGLDIGITSVGWAVLQNNNSQDEPVRIITDLGVRIFDVAENPKNGD
ALAAPRRDARTTRRRLRRRRHRLERIKFLLQENGLIEMDSFMERYYYKGNLPDVYQLRYEGLDRK
LKDEELAQVLIHIAKHGRGFRSTRKAETKEKEGGAVLKATTENQKIMQEKGYRTVGEMLYLDEAF
HTECLWNEKGYVLTTPRNRDDYKHTILRSMLVEEVHAFIAAQRAGHNQKATEGLEEAYVEIMTS
10 QRSFDMGPGLPDQKPSPYAMEGFGDRVGKCTFEKDEYRAPKATYTAELFVALQKINHTKLIDE
FGTGRFFSEEERKTIIGLLSSKELKYGTIRKKLNIDPSLKFNLSNYSAKKEGETEEERVLDTE
KAKFASMFWTYEYSKCLKDRTEEMPVGEKADLDFDRIGEILTAYKNDDSRSSRLKELGLSGEEID
GLLDLSPAKYQRVSLKAMRKMOPYLEDGLIYDKACEAAGYDFRALNDGNKKHLLKGEEINAIVN
DITNPVVKRSVSQTIKVINAI IQKYGSPQAVNIELAREMSKNFQDRTNLEKEMKKRQQENERAK
15 QQI IELGKQNP TGQDILKYRLWNDQGGYCLYSGKKIPLEELFDGGYDIDHILPYSITFDDSYRN
KVLVTAQENRQKGNRTPYEYFGADEKRWEDYEASVRLLRDYKKQKLLKKNFTEERKEFKER
NLNDTKYITRVVNMIRQNLLELEPFNHPEKKKQVWAVNGAVTSYLRKRWGLMQKDRSTDRHAM
DAVVIACCTDGMIIKISRYMQGRELAYS RNFKFPDEETGEILNRDNFTREQWDEKFGVKVPLPW
NSFRDELDIRLLNEDPKNFLTHADVQRELDYPGWMYGEEESPIEEGRYINYIRPLFVSRMPNH
KVTGSAHDATIRSARDYETRGVVITKVPLTDLKLNKDNEIEGYDDKDSDRLLYQALVRQLLLHG
20 NDGKKAFAEDFHKPKADGTEGPVVRVKVIEKKQTS GVMVRGGTGIAANGEMVRIDVFRENGKYY
FVPVYTADVVRKVLNRAATHTKPYSEWRVMDANFVFSLSYRDLIHVSKSKDIKTNLVNGGLL
LQKEIFAYYTGADIATASIAGFANDSNFKFRGLGIQSLEIFEKQCQVDILGNISVVRHENRQEFH

SEQ ID NO: 348

25 MRVLGLDAGIASL GWALIEIEESNRGELSQGTIIAGAGTWMF DAPEEKTQAGAKLKSEQRRTFRG
QRRVVRRRRQRMNEVRRI LSHGLLPSSDRDALKQPGLDPWIRAEALDRLLGPVELAVALGHI
ARHRGFKSNSKGAKTNDPADDTSKMRAVNETREKLARFGSAAKMLVEDESFVLRQTPTKNGAS
EIVRRFRNREGDYSRSLRLRDDLAEMRALFTAQARFQSAIATADLQTAFTKAAFFQRPLQDSEK
LVGPCPFEVDEKRAPKRGYSFELFRFLSRLNHVTLRDGKQERTLTRDELALAAADF GAAAKVSF
30 TALRKKLKL PETTVFVGKADEESKLDVVARSGKAAEGTARLRSVIVDALGELAWGALLCSPEK
LDKIAEVISFRSDIGRISEGLAQAGCNAPLVDAL TAAASDGRFDPFTGAGHISSKAARNILSGL
RQGMTYDKACCAADYDHTASRERGAFDVGHGREALKRILQEERISREL VGSPTARKALIESIK
QVKAIVERYGVPDRIHVELARDVGKSI EEREEITRGIEKRNQKDKLRGLFEKEVGRPPQD GAR
GKEELLRFELWSEQMGRCLYTDDYISPSQLVATDDAVQVDHILPWSRFADDSYANKTLCMAKAN
35 QDKKGRTPYEWFKAEKTDTEWDAFIVRVEALADMKGFKKRN YKL RNAE EAAAKFRNRNLNDTRW
ACRL LAEALKQLYPKGEKDKDGKERRRVFSRPGALTDRLRRAWGLQWMKKSTKGDRIPDDRHA
LDAIVIAATTESLLQRATREVQEIEDKGLHYDLVKNVTPPWPGFREQAVEAVEKVFVARAERRR
ARGKAHDATIRHIAVREGEQRVYERRKVAELKLADLDRVKDAERNARLIEKLRNWIEAGSPKDD
PPLSPKGDP IFKVRLVTKSKVNIALDTGNPKRPGTVDRGEMARVDVFRKASKKGKYEYLVPIY
40 PHDIATMTKTPPIRAVQAYKPEDEWPEMDSSYEFCWSLVPMTYLQVISSKGEIFEGYYRGMNRSV
GAIQLSAHSNSSDVVQGIGARTLTEFKKFNVDRFGRKHEVERELRTWRGETWRGKAYI

SEQ ID NO: 349

45 MGNYLGLDVGIGSIGWAVINIEKKRIEDFNVRIFKSGEIQEKNRNSRASQQCRRSRGLRRLYR
RKSHRKLRLKNYLSII GLTTSEKIDYYYETADNNVIQLRNKGLSEKLTPEEIIAACLIHICNNRG
YKDFYEVNVEDIEDPDERNEYKEEHSIVLISNLMNEGGYCTPAEMICNCREFDEPN SVYRK FH
NSAASKNHYLITRHMLVKEVDLILENQSKYYGILDDKTI AKIKDII FAQRDFEIGPGKNERFR

FTGYLDSIGKCQFFKDQERGSRFTVIADIYAFVNVLSQYTYTNNRGESVFDTSFANDLINSALK
 NGSMKRELKAIKASYHIDISDKNSDTSITKCFKYIKVVKPLFEKYGYDWDKLIENYTDTDNNV
 LNRIGIVLSQAQTPKRRREKLKALNIGLDDGLINELTKLKLSTANVSYKYMQGSIEAFCEGDL
 YGKYQAKFNKEIPDIDENAKPQKLPPFKNEDDCEFFKNPVVFRSINETRKLINAIIDKYGYPA
 5 VNIETADELNKTFEDRAIDTKRNNNDNQKENDRIVKEIIECIKCDEVHARHLIEKYKLWEAQEGK
 CLYSGETITKEDMLRDKDKLFEVDHIVPYSILNDNTINNKAIVYAEENQKKGQRTPLMYMNEAQ
 AADYRVRVNTMFKSKKCSKKKYQYLMLPDLNDQELLGGWRSRNLNDTRYICKYLVNLYRKNLRF
 DRSYESSDEDDLKIRDHYRVFPVKSFRFTSMFRRWWLNEKTWGRYDKAELKKLTYLDHAADAI
 ANCRPEYVVLAGEKLKLNKMYHQAGKRITPEYEQSKKACIDNLYKLFRMDRRTAEKLLSGHGRL
 10 TPIIPNLSEEVDKRLWDKNIYEQFWKDDKDKKSCEELYRENVASLYKGDPKFASSLSMPVISLK
 PDHKYRGITITGEEAIRVKEIDGKLIKLRKKSISEITAESINSIYTDDKILIDSLKTIFEQADYK
 DVG DY LKKTNQHFFTTSSGKRVNKVTVIEKVPSRWLRKEIDDNNFSLNDSSYYCIELYKDSKG
 DNNLQGIAMSDIVHDKTKKLYLKPFDNYPPDDYYTHVMYIFPGDYLRKSTSKKSQGEQLKFEGY
 FISVKNVNENSFRFISDNKPCAKDKRVSITKKDIVIKLAVDLMGKVQGENNGKGISCGEPLSLL
 15 KEKN

SEQ ID NO: 350

MLSRQLLGASHLARPVSYSYNVQDNDVHCSYGERCFMRGKRYRIGIDVGLNSVGLAAVEVSDEN
 SPVRLNNAQSVIHDGGVDPQKNKEAITRKNMSGVARRTRMRRRKRERLHKLDMLLGKFGYPVI
 20 EPESLDKPFEEWHVRAELATRYIEDDELRRRESISIALRHMARHRGWRNPYRQVDSLISDNPYSK
 QYGELKEKAKAYNDDATAEEEESTPAQLVVAMLDAGYAEAPRLRWRTGSKKPDAGYLPVRLMQ
 EDNANELKQIFRVQRPADWKPLFRSVFYAVSPKGSAEQRVGQDPLAPEQARALKASLAFQEY
 RIANVITNLRIKDASAE LRKLTVDEKQSIYDQLVSPSSEDITWSDLCDFLGFKRSQKGVGSLT
 EDGEERISSRPPRLTSVQRIYESDNKIRKPLVAWWKSASDNEHEAMIRLLSNTVDIDKVREDVA
 25 YASAIEFIDGLDDDALTKLDSVDLPSGRAAYSVELTQKLTRQMLTTDDDLHEARKTLFNVTDSW
 RPPADPIGEPLGNPSVDRVLKNVNRVLMNCQQRWGNPVSVNIEHVRSSFSSVAFARKDKREYEK
 NNEKRSIFRSSLSEQLRADEQMEKVRESDLRRLEAIQRQNGQCLYCGRTITFRTCEMDHIVPRK
 GVGSTNTRTNFAAVCAECNRMKSNTPF AIWASEDAQTRGVSLAEAKKRVMTFTFNPKSYAPRE
 VKAFKQAVIARLQQTEDDAIDNRSIESVAWMADDELHRRIDWYFNAKQYVNSASIDDAEAEATMK
 30 TTVSVFQGRVTASARRAAGIEGKIHF IGQQSKTRLDRRHAVDASVIAMMNTAAQTLMERESL
 RESQRLIGLMPGERSWKEYPYEGTSRYESFHLWLDNMDVLELLNDALDNDRIAVMQSQRYVLG
 NSIAHDATIHPLEKVPLGSAMSADLIRRASTPALWCALTRLPDYDEKEGLPEDSHREIRVHDTR
 YSADDEMGGFFASQAAQIAVQEGSADIGSAIHARVYRCWKTNAKGVRKYFYGMIRVFQTDLLRA
 CHDDLFTVPLPPQSISMRYGEPRVVQALQSGNAQYLGLSVVGDEIEMDFSSLDVDGQIGEYLQF
 35 FSQFSGGNLAWKHVVVDGFFNQTLRIRPRYLAAEGLAKAFSDDVVPDGVQKIVTKQGWLPPVN
 TASKTAVRIVRRNAFGEPRLSAHHMPCSWQWRHE

SEQ ID NO: 351

MYSIGLDLGISSVGWSVIDERTGNVIDLGVRLESAKNSEKNLERRTNRGGRRLLIRKTNRLKDA
 40 KKILAAVGFYEDKSLKNSCP YQLRVKGLTEPLSRGEIYKVTLHILKKRGISYLDEVDTAAKES
 QDYKEQVRKNAQLLT KYTPGQIQQLQRLKENNRVKTGINAQGN YQLNVFKVSAYANELATILKTQ
 QAFYPNELTDDWIALFVQPGIAEEAGLIYKRKPYHGGPGNEANNSPYGRWSDFQKTGEPATNIF
 DKLIGKDFQGELRASGLSLSAQQYNLLNDLTNLKIDGEVPLSSEQKEYILTELMTKEFTRFGVN
 DVVKLLGVKKERLSGWRLDKKGKPEIHTLKGYNWRKIFAEAGIDLATLPTETIDCLAKVLTNL
 45 TERE GIENTLAFELPELSESVKLLVLDYKELSQSISTQSWHRFSLKTLHLLIPELMNATSEQN
 TLLEQFQLKSDVRKRYSEYKKLPTKDVLA EIYNPTVNKTVSQAFKVIDALLVKYGKEQIRYITI
 EMPRDDNEEDEKKRIKELHAKNSQRKNDSSQSYFMQKSGWSQEKFQTTIQKNRRFLAKLLYYEYQ

DGICAYTGLPISPPELLVSDSTEIDHIIIPISISLDDSIINNKVVLVLSKANQVKGQQTPYDAWMDGS
 FKKINGKFSNWDDYQKWVESRHF'SHKKENNLETRNIFDSEQVEKFLARNLNDTRYASRLVLNT
 LQSFFTINQETKVRVNGSFTHTLRKKWGADLDKTRETHHHHAVDATLCAVTSFVKVSRHYHAVK
 EETGEKVMREIDFETGEIVNEMSYWEFKKSKKYERKTYQVKWPNFREQLKPVNLHPRIKFSHQV
 5 DRKANRKLSDATIYSVREKTEVKTLKSGKQKITTTDEYTIGKIKDIYTLDGWEAFKKKQDKLLMK
 DLDEKTYERLLSIAETTPDFQEVEEKNGKVVRKRSFPAVYCEENDIPAIQKYAKKNNGLPIRS
 LKYYDGKLNKHINITKDSQGRPVEKTKNGRKVTLQSLKPYRYDIYQDLETKAYYTVQLYYSDLR
 FVEGKYGITEKEYMKKVAEQTKGQVVRFCFSLQKNDGLEIEWKDSQRYDVRFYNFQSANSINFK
 GLEQEMMPAENQFKQKPYNNGAINLNI AKYGKEGKKLRKFNTDILGKKHYLFYEKEPKNI IK

SEQ ID NO: 352

MYFYKNKENKLNKKVVLGLDLGIASVGWCLTDISQKEDNKFPIILHGVRLFETVDDSDDKLLNE
 TRRKKRGQRRNRRLFTRKRDFIKYLIDNNIIIELEFDKNPKILVRNFIEKYINPFSKNLELKYYK
 SVTNLPIGFHNLRKAAINEKYKLDKSELIVLLYFYLSLRGAFFDNPEDTKSKEMNKNEIEIFDK
 15 NESIKNAEFPIDKIIIEFYKISGKIRSTINLKFQGHQDYLKEIKQVFEEKQNIDFMNYEKFAMEEKS
 FFSRIRNYSEPGNEKSF'SKYGLYANENGNPELI INEKGQKIYTKIFKTLWESKIGKCSYDKKL
 YRAPKNSFSKAVFDITNKLTDWKHKNEYISERLKRKILLSRFLNKDSKSAVEKILKEENIKFEN
 LSEIAYNKDDNKINLPIINAYHSLTTIFKKHLINFENYLI SNENDLSKLMSFYKQQSEKLFVPN
 EKGSYEINQNNVLHIFDAISNILNKFSSTIQDRIRILEGYFEFSNLKKDVKSSSEIYSEIAKLRE
 20 FSGTSSLSFGAYYKFIPNLISEGSKNYSTISYEEKALQNQKNFSSHNLFEKTWVEDLIASPTV
 KRSLRQTMNLLKEIFKYSEKNLEIEKIVVEVTRSSNNKHERKKIEGINKYRKEKEYEELKKVYD
 LPNENTTLLKKLWLLRQQQGYDAYSLRKIEANDVINKPWNIDIDHIVPRSISFDDSFNLVIVN
 KLDNAKKSNDLSAKQFIEKIYGIEKLKEAKENWGNWYLRNANGKAFNDKGKFIKLYTIDNDEF
 DNSDFINRNLSDTSYITNALVNHLTFSNSKYKYSVVSVNGKQTSNLRNQIAFVGIKNNKETERE
 25 WKRPEGFKSINSNDFLIREEGKNDVKDDVLIKDRSFNGHHAEDAYFITIIISQYFRSFKRIERLN
 VNYRKETRELDDELEKNKIFKEKASFDNFLINALDELNEKLNQMRFSRMVITKKNQQLFNETL
 YSGKYDKGKNTIKKVEKLNLLDNRTDKIKKIEEFFDEDKLKENELTKLHIFNHDKNLYETLKII
 WNEVKIEIKNKNLNEKNYFKYFVNKKLQEGKISFNEWVPILDNDFKIIRKIRYIKFSSEEKETD
 EIIIFSQSNFLKIDQRQNF'SFHNTLYWVQI WVYKNQKDQYCFISIDARNSKFEEKDEIKINYEKLLK
 30 TQKEKLQIINEEPILKINKGDLFENEKELFYIVGRDEKPQKLEIKYILGKKIKDQKQIQKPVK
 KYFPNWKKVNLT YMGEIFKK

SEQ ID NO: 353

MDNKNYRIGIDVGLNSIGFCAVEVDQHDTPLGFLNLSVYRHDAGIDPNGKKTNTTRLAMSGVAR
 35 RTRRLFRKRKRRLAALDRFIEAQGWTLDPHADYKDPYTPWLVRaelAQTPIRDENDLHEKLAIA
 VRHIARHRGWRSPWPVRS LHVEQPPSDQYLALKERVEAKTLLQMPEGATPAEMVVALDLSVDV
 NLRPKNREKTDTRPENKKPGFLGGKLMQSDNANELRKIAKIQGLDDALLRELIELVFAADSPKG
 ASGELVG YDVLPGQH GKRRAEKAHPAFQRYRIASIVSNLRIRHLGSGADERLDVETQKRVFEYL
 LNAKPTADITWSDVAEEIGVERNLLMGTATQTADGERASAKPPVDVTNVAFAATCKIKPLKEWWL
 40 NADYEARCVMVSALSHAEKLT EGTA AEVEVAEFLQNLSDNEKLD SFSLPIGRAAYSVD SLER
 LTKRMIENGEDLFEARVNEFGVSEDWRPPAEPIGARVGNPAVDRVLKAVNRYLMAAEAEWGAPL
 SVNIEHVREGFISKRQAVEIDRENQKRYQRNQAVRSQIADHINATSGVRGSDVTRYLAIQRQNG
 ECLYCGTAITFVNSEMDHIVPRAGLGSTNTRDNLVATCERCNKSKSNKPFVWAAECGIPGVS
 AEALKRVDFWIADGFASSKEHRELQKGVKDRLKRKVS DPEIDNRSMESVAWMARELAHRVQYYF
 45 DEKHTGTVRVFRGSLTSAARKASGFESRVNF IGNGKTRLD RRHHAMDAATVAML RNSVAKTL
 VLRGNIRASERAIGAAETWKSFRGENVADRQIFESWSENMRVLVEKFNALALYNDEVSIFSSLRL
 QLGNGKAHDDTITKLQMHKVGDAWSLTEIDRASTPALWCALTRQPDFTWKDGLPANEDRTIIVN

GTHYGPLDKVGIFGKAAASLLVRGGSVDIGSAIHHARIYRIAGKKPTYGMVRVAFAPDLLRYRNE
DLFNVELPPQSVSMRYAEPKVREAIREGKAEYLGWL VVGDEL LLDLSSETSGQIAELQQDFPGT
THWTVAGFFSPSRLRLRPVYLAQEGLGEDVSEGSKSIIAGQGWRPAVNKVFGSAMPEVIRRDGL
GRKRRFSYSGLPVSWQG

5

SEQ ID NO: 354

MRLGLDIGTSSIGWWLYETDGAGSDARITGVVDGGVRIFSDGRDPKSGASLAVDRRAARAMRRR
RDRYLRRRATLMKVLAETGLMPADPAEAKALEALDPFALRAAGLDEPLPLPHLGRALFHLNQRR
GFKSNRKTDRGDNESGKIKDATARLDMEMMANGARTYGEFLHKRRQKATDPRHVP SVRTRLSIA
10 NRGGPDGKEEAGYDFYPDRRHLEEEFHKLWAAQGAHHPELTETLRDLLFEKIFFQRPLKEPEVG
LCLFSGHHGVPPKDPRLPKAHPLTQRRVLYETVNQLRVTADGREARPLTREERDQVIHALDNKK
PTKSLSSMVLKLPALAKVLKLRDGERFTLETGVRDAIACDPLRASPAHPDRFGPRWSILDADAQ
WEVISRIRRVQSDAEHAALVDWLTEAHGLDRAHAEATAHAPLPDGYGRLGLTATTRILYQLTAD
VVTYADAVKACGWHHSDGRTGECFDRLPYYGEVLERHVIPGSYHPDDDDITRFGRI TNPTVHIG
15 LNQLRRLVNRI IETHGKPHQIVVELARDLKKSEEQKRADIKRIRDTEAAKKRSEKLEELEIED
NGRNRMLLRLWEDLNPDDAMRRFCPYTGTRISAAMIFDGSCD VDHILPYSRTLDDSFNRTLCL
REANRQKR NQTPWQAWGDTPHWHAI AANLKNLPENKRWRFAPDAMTRFEGENGFLDRALKDTQY
LARISRSYLDTLFTKGGHVWVVPGRFTEMLRRHWGLNSLLSDAGRGAVKAKNRTDHRHHAIDAA
VIAATDPGLLNRI SRAAGQGEAAGQSAELIARDTPPPWEGFRDDLVRVLDRI IVSHRADHGRID
20 HAARKQGRDSTAGQLHQETAYSIVDDIHVASRTDLLSLKPAQLLDEPGRSGQVRDPQLRKALRV
ATGGKTGKDFENALRYFASKPGPYQAIRRVRI IKPLQAQARVPVPAQDPIKAYQGGSNHLFEIW
RLPDGEIEAQVITSFEAHTLEGEKRPHPAAKRLLRVHKGDMVALERDGRRVVGHVQKMDIANG
FIVPHNEANADTRNNDKSDPFKWIQIGARPAIASGIRRVSVDEIGRLRDGGTRPI

25 SEQ ID NO: 355

MLHCIAVIRVPPSEEPGFFETHADSCALCHHGCMTYAANDKAIRYRVGIDVGLRSIGFCAVEVD
DEDHPIRILNSVVHVHDAGTGGPGETESLRKRS GVAARARRRGRAEKQRLKKLDV LLEELGWGV
SSNELLD SHAPWHIRKRLVSEYIEDETERREQCLSVAMAHIARHRGWRNSFSKVDTL LLEQAPSD
RMQGLKERVEDRTGLQFSEEV TQGELVATLLEHDGDVTIRGFVRKGGKATKVHGVLE GKYMQSD
30 LV AELRQICRTQRVSETTFEKLVL SIFHSKEPAPSAARQRERVGLDELQLALDPAAKQPRAERA
HPAFQKFVKVATLANMRIREQSAGERSLTSEELNRVARYLLNHTESSESPTWDDVARKLEVPRHR
LRGSSRASLETGGGLTYPPVDDTTVRVMSAEVDWLADWWD CANDES RGHMIDAI SNGCGSEPDD
VEDEEVNELISSATAEDMLKLELLAKKLPSGRVAYSLKTLREVTA AILETGDDLSQAITRLYGV
DPGWVPTPAPIEAPVGNPSVDRVLKQVARWLKFASKRWGVPQTVNIEHTREGLKSASLLEEERE
35 RWERFEARREIRQKEMYKRLGISGPFRRSDQVRYEILDLDQDCACLYCGNEINFQTFEVDHIIPR
VDASSDSRRTNLA AVCHSCNSAKGGLAFGQWVKRGDCPSGV SLENAIKRVRSWSKDRLGLTEKA
MGKRKSEVISRLKTEMPYEEFDGRSMESVAWMAIELKKRIEGYFN SDRPEGCAAVQVNAYSGRL
TACARRAAHV DKRVR LIRLKGDDGHKKNRFD RNRNHAMDALVIALMTPAIARTIAVREDRREAQQ
LTRAFESWKNFLGSEERMQDRWESWIGDVEYACDRLNELIDADKIPVTENLRLRNSGKLHADQP
40 ESLKKARRGSKRPRPQRYVLGDALPADVINRVTDPGLWTALVRAPGFDSQLGLPADLNRGLKLR
GKRISADFPIDYFPTDSPALAVQGGYVGLEFH HARLYRIIGPKEKVKYALLRVCAIDL CGIDCD
DLFEVELKPSSI SMRTADAKLKEAMNGSAKQIGWLVLGDEIQIDPTKFPKQSIGKFLKECGPV
SSWRVSALDTPSKITLKPRLLSNEPLLKTSRVGGHESDLVVAECVEKIMKKTGWVVEINALCQS
GLIRVIRRNALGEVRTSPKSGLPISLNL R

45

SEQ ID NO: 356

MRYRVGLDLGTASVGA AVFSMDEQGNPMELIWHYERLFSEPLVPDMGQLKPKKAARRLARQQRR
 QIDRRASRLRRIAIVSRRLGIAPGRND SGVHGNDVPTLRAMAVNERIELGQLRAVLLRMGKKRG
 YGGTFKAVRKVGEAGEVASGASRLEEEMVALASVQNKDSVTVGEYLAARVEHGLPSKLKVAANN
 EYYAPEYALFRQYLGLPAIKGRPDCLPNMYALRHQIEHEFERIWATQSQFHDVMKDHGVKEEIR
 5 NAIFFQRPLKSPADKVGRC SLQTNLPRA PRAQIAAQNFRIEKQ MADLRWGMGRRAEMLNDHQKA
 VIRELLNQQKELSF RKIYKELERAGCPGPEGKGLNMDRAALGGRDDLSGNTTLAAWRKLGLED R
 WQELDEVTQIQVINFLADLGSPEQLD TDDWSCR FMGKNRPRNF SDEFVAFMNELRMTDGFDR L
 SKMGFEGRSSYSIKALKALTEWMIAPHWRET PETHRVDEEAAIRECYPESLATPAQGGRQSKL
 EPPPLTGNEVDVALRQVRHTINMMIDDLGSVPAQIVVEMAREMKGGVTRNDIEKQNKRFASE
 10 RKKAAQSIEENGKTPTPARILRYQLWIEQGHQCPYCESNISLEQALSGAYTNFEHILPRTL TQI
 GRKRSELVLAHRECND EKG NRTPYQAFGHDDRRWRIVEQRANALPKKSSRKTRLLLLKDFEGEA
 LTDESIDEFADRQLHESSWLAKVTTQWLSSLSGSDVYVSRGSLTAELRRRWGLDTVIPQVRFESG
 MPVVDEEGAEITPEEF EKFR LQWEGHRVTREMRDTRRPDKRIDHRHHLVDAIVTALT SRSLYQQ
 YAKAWKVADEKQRHGRVDVKVELPMPILTIRDIALEAVRSVRISHKPD RYPDGRFF EATAYGIA
 15 QRLDERSGEKVDWLVSRSKSLTDLAPEKKSIDVDKVRANISRIVGEAIRLHISNIFEKRVSKGMT
 PQQALREPIEFQGNILRKVRCFYSKADDCVRIEHSSRRGHYKMLLNDGFAYMEVPCKEGILYG
 VPNLVRPSEAVGIKRAPESGDFIRFYKGDTVKNIKTGRVYTIKQILGDGGGKLILTPVTETKPA
 DLLSAKWGRLKVGGRNIHLLRLCAE

20 SEQ ID NO: 357
 MIGEHVRGGCLFDDH WTPNWGA FRLPNTV RTFTKAENPKDGSSLAEP RRQARGLRRRLRRKTQR
 LEDLRRL LAKEGVL SLS DLET L FRET PAKDPYQLRAEGLDRPLSFPEWVRVLYHITKHRGFQSN
 RRNPVEDGQERSRQEEEGKLLSGVGENERLLREGGYRTAGEMLARDPKFQDHRNRNAGDYSHTL
 SRSLLEEARRLFQSQRT LGNPHASSNLEEAF LHLVAFQNP FASGEDIRNKAGHCSLEPDQIRA
 25 PRRSASAE TFMLLQKTGNLRLIHRRTGEERPLTDKEREQIHL LAWKQEKVTHKTLRRHLEIPEE
 WLFTGLPYHRSGDKAE EKLFVHLAGIHEIRKALDKGPDPAVWDTLRSRRD LLD SIADTLTFYKN
 EDEILPRLES LGLSPENARALAPLSFSGTAHLSLSALGKLLPHLEEGKSYTQARADAGYAAPP
 DRHPKLPPLEEADWRNPVVFRA LTQTRKVVNALVRRYGPPWC I HLETARELSQPAKVRRIETE
 QQANEKKKQQAEREFLDIVGTAPGPGDLLKMRLWREQGGFCPYCEEYLNPTRLAEPGYAEMDHI
 30 LPYSRSLDNGWHNRVLVHGKDN RDKGNRTPFEAFGGDTARWDR LVAWVQASHLSAPKKRNLLRE
 DFGEAEARELKD RNLT DTRFI TKTAATLLRDRLTFHPEAPKDPVMTLNGRLTAFLRKQWGLHKN
 RKN GDLHHALDAAVLAVASRSFVYRLSSHNAAWGELPRGREANGFSLPYPAFRSEVLARLCPT
 REEILLRLDQGGVGYDEAFRNGLRPV FVSRAPSRRLRGKAHMETLRSPKWKDHPEGPRTASRIP
 LKDLNLEKLERMVGKDRDRKLYEALRERLA AFGGNGKAFVAPFRKPCRSGEGLVRS LRIFDS
 35 GYSGVELRDGGEVYAVADHESMVRVDVYAKKNRFYLVPVYVADVARGIVKNRAIVAHKSEEWD
 LVDGSFDFRFS LFP GD LVEIEKKDGAYLGYYKSCHRGDGRLLLLDRHDRMPRESDCGT FYVSTRK
 DVLSMSKYQVDPLGEIRLVGSEKPPFVL

40 SEQ ID NO: 358
 MEKKRKVTLGFDLGIASVGWAIVDSETNQVYKLGSR LFDAPDTNLERRTQRGTRRLRRRRKYRN
 QKFYNLVKRTEVFGLSSREAIENRFRELSIKYPNIIELKTKALSQEVCPDEIAWILHDYLNKRG
 YFYDEKETKEDFDQQTVESMP SYKLNEFYKKYGYFKGALSQPT ESEMKNKDLKEAFFFD FSNK
 EWLKEINYFFNVQKNILSETFIEEFKKIFSFTRDISKGP GSDNMPSPY GIFGEFGDNGQGGRYE
 HIWDKNIGKCSIFTNEQRAPKYLP SALIFNFLNELANIRLYSTD KKNIQPLWKLSSVDKLNILL
 45 NLFNLP ISEKKKLTSTNINDIVKKESIKSIMISVEDIDMIKDEWAGKEPNVYGVGLSGLNIEE
 SAKENKFKFQDLKILNVLINLLDNVGIKFEFKDRNDI IKNLELLDNLYLFLIYQKESNNKDSSI
 DLFI AKNESLN IENLKLKLKEFLLGAGNEFENHNSKTHSLSKKAIDEILPKLLDNNEGWNLEAI

KNYDEEIKSQIEDNSSLMQDKKYLNDFLKDAILPPNVKVTFFQQAILIFNKIIQKFSKDFEI
 DKVVIELAREMTQDQENDALKGIAKAQSKSKSLVEERLEANNIDKSVFNDKYEKLIYKIFLWIS
 QDFKDPYTGAQISVNEIVNNKVEIDHIIPYSLCFDDSSANKVLVHKQSNQEKSNSLPYEYIKQG
 HSGWNWDEFTKYVKRVFVNNVDSILSKKERLKKSENLLTASYDGYDKLGFLARNLNDTRYATIL
 5 FRDQLNNYAEHHLIDNKKMFKVIAMNGAVTSFIRKNMSYDNKLRLKDRSDFSHHAYDAAIALF
 SNKTKTLYNLIDPSLNGIISKRSEGYWVIEDRYTGEIKELKKEDWTSIKNNVQARKIAKEIEEY
 LIDLDEVEFFSRKTKRKTNRQLYNETIYGIATKTDEDGITNYYKKEKFSILDDKDIYLRLLRER
 EK FVINQSNPEVIDQIIIEIIESYGKENNIPSRDEAINIKYTKNKINYNLYLKQYMRSLTKSLDQ
 FSEEFINQMIANKTFVLYNPTKNTTRKIKFLRLVNDVKINDIRKNQVINKFNGKNNEPKAFYEN
 10 INSLGAIVFKNSANNFKTLSINTQIAIFGDKNWDIEDFKTYNMEKIEKYKEIYGIDKTYNFHSF
 IFPGTILLDKQNKEFYYISSIQTVRDIIEIKFLNKIEFKDENKNQDTSKTPKRLMFGIKSIMNN
 YEQVDISPF GINKKIFE

SEQ ID NO: 359

15 MGYRIGLDVGITSTGYAVLKT DKNGLPYKILTDSVIYPRAENPQTGASLAEPRIKRGLRRRT
 RRTKFRKQRTQQLFHSGLLSKPEIEQILATPQAKYSVYELRVAGLDRRLTNSELFVLYFFIG
 HRGFKSNRKAELNPENEADKKQMGQLLNSIEEIRKAI AEKGYRTVGELYLKDPKYNDHKRNKGY
 IDGYLSTPNRQMLVDEIKQILDKQRELGNEKLTDEFYATYLLGDENRAGIFQAQRDFDEGPGAG
 PYAGDQIKKMVGKDIFEPTEDRAAKATYTFQYFNLLQKMTSLNYQNTTGDTWHTLNLGLDRQAI I
 20 DAVFAKAEKPTKYKPTDFGELRKLKLPDDARFNLVNYGSLQTQKEIETVEKKTRFVDFKAYH
 DLVKVLPEEMWQSRQLLDHIGTALTLYSSDKRRRRYFAEELNLPALIEKLLPLNFSKFGLHSI
 KSMQNIIPYLEMGQVYSEATTNTGYDFRKKQISKDTIREEITNPVVRRAVTKTIKIVEQIIRRY
 GKPDGINIELARELGRNFKERGD IQKRQDKNRQTNDKIAAELTELGPVNGQNIIRYKLHKEQN
 GVDPYTG DQIPFERAFSEG YEVDHIIPY S ISWDDSYTNKVLTSACKNREKGNRIPMVYLANNEQ
 25 RLNALTNIADNIIRNSRKRQKLLKQKLSDEELKDWKQRNINDTRFITRVLYNYFRQAIEFNPEL
 EKKQRVLPLNGEVT SKIRSRWGF LKVREDGDLHHAIDATVIAAITPKFIQQVTKYSQHQEVKNN
 QALWHD AEIKDAEYAAEAQRMDADLFNKIFNGFPLPWPFEFLDELLARISDNPVEMMKSRSWNTY
 TPIEIAKLKPVFVVR LANHKISGPAHLDTIRS AKLFDEKGIVLSRV SITKLKINKKGQVATGDG
 IYDPENSNNGDKVVS AIRQALEAHNGSGELAFPDGYLEYVDHGT KKLVRKVRVAKKVSLPVRL
 30 KNKAAADNGSMVRIDVFNTGKKFVFPVIYIKD TVEQVLPNKAIARGKSLWYQITESDQFCFSLY
 PGDMVHIESKTGIKPKYSNKENNTSVVPIKNFYGYFDGADIATASILVRAHDSSYTARSIGIAG
 LLKFEKYQVDYFGRYHKVHEKKRQLFVKRDE

SEQ ID NO: 360

35 MQKNINTKQNHIIYIKQAQKIKEKLGDKPYRIGLDLGVGSIGFAIVSMEENDGNVLLPKEIIMVG
 SRIFKASAGAADRKL SRGQRNNHRHTRERMRYLWKVLAEQKLALPVPADLDRKENSSEGETSAK
 RFLGDVLQKDIYELRVKSLDERLSLQELGYVLYHIAGHRGSSAIRTFENDSEEAQKENTENKKI
 AGNIKRLMAKKNYRTYGEYLYKEFFENKEKHKREKISNAANNHKFSPTRD LVIKEAEAILKKQA
 GKDG FHKELTEEYIEKLT KAIGYSEKLIPE SGFCPYLKDEKRLPASHKLNEERRLWETLNNAR
 40 YSDPIVDIVTGEITGYYEKQFTKEQKQKLF DYLLTGSELTPAQTKKLLGLKNTNFEDIILQGRD
 KKAQKIKGYKLIKLESMPFWARLSEAQQDSFLYDWN SCPDEKLLTEKLSNEYHLTEEEIDNAFN
 EIVLSSSYAPLGKSAMLIILEKIKNDLSYTEAVEEALKEGKLTKEKQAIKDRLPYYGAVLQEST
 QKIIAKGFSPQFKDKGYKTPHTNKYELEYGRIANPVVHQT LNELRKL VNEIIDILGKKPCEIGL
 ETARELKKS AEDRSKLSREQNDNESNRNRIYEIYIRPQQQVIITRRENPRNYILKFELLEEQKS
 45 QCPFCGGQISPNDIINNQADIEHLFPIAESEDNGRNNLVISHSACNADKAKRSPWA AFASA AKD
 SKYDYNRILSNVKENIPHKAWRFNQGAF EKFIENKPMAARFKTDNSYISKVAHKYLACLFEPKN
 IICVKGSLTAQLRMAWGLQGLMIPFAKQLITEKESSEFNKDVNSNKKIRLDNRHHALDAIVIA Y

ASRGYGNLLNKMAGKDYKINYSERNWLSKILLPPNNIVWENIDADLESFESSVKTALKNAFISV
 KHDHSDNGELVKGTMYKIFYSERGYTLTTYKKLSALKLTDPQKKKTPKDFLETALLKFKGRESE
 MKNEKIKSAIENNKRLFDVIQDNLEKAKKLEENEKSKEGKKEKNINDASIYQKAIISLSDGK
 YVQLSKKEPGKFFAISKPTPTTTGYGYDTGDSLVDLYYDNKGKLCGEIIRKIDAQQKNPLKYK
 5 EQGFTLFEIRIYGGDILEVDFDIHSDKNSFRNNTGSAPENRVFIKVGTFTEITNNNIQIWFNII
 KSTGGQDDSFITNSMQQYNPRKLILSSCGFIKYRSPILKNKEG

SEQ ID NO: 361

MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRFERAEVPKTGDSLAMARRL
 10 ARSVRRLTRRRARHLLRARLLKREGVLQAADFENGLIKSLPNTPWQLRAAALDRKLTPLEWS
 AVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNAHALQTGDFRTPAELALNKFEKESGHI
 RNQRGDYSHTFSRKDLQAEILLLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLG
 HCTFEPAEPKAAKNTYTAERFIWLTCLNNLRILEQGSRPLTDTERATLMDEPYRKSCLTYAQA
 RKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALKKEGLKDKKSPNLSPELQDEIGT
 15 AFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEI
 YGDHYGKKNTEEKIYLPPIPADEIRNPVVLRLALSQARKVINGVVRRYGSPARIHIETAREVGKS
 FKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPSKSKDILKLRLYEQQHKGKCLYSGKEINLG
 RLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVE
 TSRFPRSKKQRILLQKFDEDDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGGKRVFASNGQITN
 20 LLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQ
 KTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLEKLSSRPEAVHEYVTPLFVSR
 APNRKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLKDLKMNVREREPKLYEALKARLEAHK
 DDPKAFAPFYKYDKAGNRTQQVKAVRVEQVQKTGVVVRNHNGIADNATMVRVDVFEKGDKEY
 LVPIYSWQVAKGILPDRAVVQKDEEDWQLIDDSFNFKFSLHPNDLVEVITTKARMEGYFASCH
 25 RGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFOKYQIDELGKEIRPCRLKKRPPVR

SEQ ID NO: 362

MQTTNLSYILGLDLGIASVGWAVVEINENEDPIGLIDVGVRIFERAEVPKTGESLALSRRRLARS
 TRRLIRRRARHLLAKRFLKREGILSTIDLEKGLPNQAWELRVAGLERRLSAIEWGAVLLHLIK
 30 HRGYLSKRKNESQTNKELGALLSGVAQNHLQLQSDDYRTPAELALKKFAKEEGHIRNQRGAYT
 HTFNRLDLLAELNLLFAQQHQFGNPHCKEHIQQYMTPELLMWQKPALSGEAILKMLGKCTHEKNE
 FKAACKHTYSAERFVWLTCLNNLRILEDGAERALNEEERQLLINHPYEKSKLTYAQVRKLLGLSE
 QAIFKHLRYSKENAESATFMELKAWHAIRKALENQGKDTWQDLAKKPDLLDEIGTAFSLYKTD
 EDIQQYLTKVPNSVINALLVSLNFDKFIELSLKSLRKILPLMEQGKRYDQACREIYGHYGEA
 35 NQKTSQLLPAIPAQEIRNPVVLRTLSQARKVINAIIRQYGSPARVHIETGRELGKSFKERREIQ
 KQQEDNRTKRESAVQKFKELFSDFSSEPFSKSKDILKFRLYEQQHKGKCLYSGKEINIHRLNEKGYV
 EIDHALPFSRTWDDSFNNKVLVLASENQNKGNTPYEWLQKINSERWKNFVALVLGSQCSAAK
 KQRLLTQVIDDNKFIDRNLDTRYIARFLSNYIQENLLLVGKNKKNVFTPNGQITALLRSRWGL
 IKARENNNRHHALDAIVVACATPSMQQKITRFRFKEVHPYKIENRYEMVDQESGEIISPHFPE
 40 PWAYFRQEVNIRVFDNHPDVTVLKEMPLPDRPQANHQFVQPLFVSRAPTRKMSGQGHMETIKSAKR
 LAEGISVLRIPLTQLKPNLLENMVNKEREPALYAGLKARLAEFNQDPAKAFATPFYKQGGQQVK
 AIRVEQVQKSGVLVRENNGVADNASIVRTDVFIKNNKFFLVPIYTWQVAKGILPNKAIIVAHKNE
 DEWEEMDEGAKFKFSLFPNDLVELKTKKEYFFGYIIGLDRATGNISLKEHDGEISKGKDGVRV
 45 GVKLALSFEKYQVDELGKNRQICRPQQRQPVR

SEQ ID NO: 363

MGIRFAFDLGTNSIGWAVWRTGPGVFGEDTAASLDGSGVLIFKDGRNPKDGQSLATMRRVPRQS
 RKRRDRFVLRRLDRLAALRKAGLFPVDVEEGRRLAATDPYHLRAKALDESLTPHEMGRVIFHLN
 QRRGFRSNRKADRQDREKKGIAEGSKRLAETLAATNCRTLGEFLWSRHRGTPRTRSPTRIRMEG
 EGAKALYAFYPTREMVRAEFERLWTAQSRFAPDLLTPERHEE IAGILFRQDLAPPKIGCCTFE
 5 PSERRLPRALPSVEARGIYERLAHLRITGTPVSDRGLTRPERDVLASALLAGKSLTFKAVRRTL
 KILPHALVNFEEAGEKGLDGALTAKLLSKPDHYGAAWHGLSFAEKDTFVGKLLDEADEERLIRR
 LVTENRLSEDAARRCASIPLADGYGRLGRTANTEILAALVEETDETGTVVITYAEAVRRAGERTG
 RNWHHSDEDERDGVILDRLPYYGEILQRHVVPGSGEPEEKNEAARWGRLANPTVHIGLNQLRKVVN
 RLIAAHGRPDQIVVELARELKLNRQKERLDRENKRNREENERRTAILAEHGQRDTAENKIRLR
 10 LFEEQARANAGIALCPYTGRAIGIAELFTSEVEIDHILPVSLTLDDSLANRVLCRREANREKRR
 QTPFQAFGATPAWNDIVARAALPPNKRWRFDPAALERFEREGGFLGRQLNETKYLSRLAKIYL
 GKICDPDRVYVTPGTLTGLLRARWGLNSILSDSNFKNRSDHRHHAVDAVVIGVLTRGMIQRIAH
 DAARAEDQDLDRVFRDVPVPFEDFRDHVRERVSTITVAVKPEHGKGGALHEDTSYGLVPDTPDN
 AALGNLVVRKPIRSLTAGEVDRVRDRALRARLGALAAPFRDESGRVRDAKGLAQALEAFGAENG
 15 IRRVRILKPDASVVTIADRRTGVPYRAVAPGENHHVDIVQMRDGSWRGFAASVFEVNRPGWRPE
 WEVKKLGKGLVMRLHKGDMVELSDKDGQRRVKVQVQIEISANRVRLSPHNDGGKLQDRHADADD
 PFRWDLATIPLLKDRGCVAVRVDPIGVVTLRRSNV

SEQ ID NO: 364

20 MMEVFMGRVLVLGLDIGITSVGFGIIDLDESEIVDYGVRLFKEGTAAENETRRTKRGGRRLLKRRR
 VTRREDMLHLLKQAGIISTSFHPLNPNPYDVRVKGLNERLNGEELATALHLLCKHRGSSVETIED
 DEAKAKEAGETKKVLSMNDQLLKSGKYVCEIQKERLRTNGHIRGHENNFKTRAYVDEAFQILSH
 QDLSNELKSAIITIIISRKRMYYDGGGGLSPTPYGRYTYFGQKEPIDLIEKMRGKCSLFPNEPR
 APKLAYSALFNLLNDLNNLSIEGEKLTSEQKAMILKIVHEKGKITPKQLAKEVGVSLQIRGF
 25 RIDTKGSPLLSELTYKMIREVLEKSNDEHLEDHVIFYDEIAEILTKTKDIEGRKKQISELSSDL
 NEESVHQLAGLTFTAYHSLSFKALRLINEEMLKTELNQMQSITLFGKQNNELSVKGMKNIQA
 DDTAILSPVAKRAQRETFKVVRNLREIYGEFDSIVVEMAREKNSEEQRKAIRERQKFFEMRNKQ
 VADIIGDDRKINAKLREKLVLYQEODGKTAYSLEPIDLKLIDDPNAYEVDHIIPISISLDDSI
 TNKVLVTHRENQEKGNLTPISAFVKGRFTKGS LAQYKAYCLKLKEKNIKTNKGYRKKVEQYLLN
 30 ENDIYKYDIQKEFINRNLDVTSYASRVVNLTLTYFKQNEIPTKVFTVKGSLTNAFRRKINLKK
 DRDEDYGHHAIDALI IASMPKMRLSTIFSRYKIEDIYDESTGEVFSSGDDSMYYDDRYFAFIA
 SLKAIKVRKF SHKIDTKPNRSVADETIYSTRVIDGKEKVVKYKDIYDPKFTALAEDILNNAYQ
 EKYLMALHDPQTFDQIVKVNNYFEEMSKSEKYFTKDKKGRIKISGMNPLSLYRDEHGMLKKYS
 KKGDPAITQMKYFDGVLGNHIDISAHYQVRDKKVVLQQISPYRTDFYYSKENGKYKFTIRYKD
 35 VRWSEKKKKYVIDQQDYAMKKAEEKIDDTYEFQFSMHRDELIGITKAEGEALIYPDETWHNFNF
 FFHAGETPEILKFTATNNDKSNKIEVKPIHCYCKMRLMPTISKKIVRIDKYATDVVGNLYKVKK
 NTLKFEFD

SEQ ID NO: 365

40 MKKILGVDLGITSFGYAILQETGKDLYRCLDNSVVMRNNPYDEKSGESSQSIRSTQKSMRRLIE
 KRKKRIRCVAQTMERYGILDYSETMKINDPKNNPIKNRWQLRAVDWKRPLSPQELFAIFAHMA
 KHRGYKSIATEDLIYELELELGLNDPEKESEKKADERRQVYNALRHLEELRKKYGETIAQTIH
 RAVEAGDLRSYRNHDDYEKMIRREDIEEEEIEKVLLRQAEALGALGLPEEQVSELIDELKACITDQ
 EMPTIDESLFGKCTFYKDELAAPAYSILYDLYRLYKKLADLNIDGYEVTQEDREKVIEWVEKKI
 45 AQGKNLKKITHKDLRKILGLAPEQKIFGVEDERIVKGKKEPRTFVPFFFLADIAKFELFASIQ
 KHPDALQIFRELAELQRSKTPQEALDRRLALMAGKGIDTDDRELLELFKNKRSGTREL SHRYI
 LEALPLFLEGYDEKEVQRILGFDDREDYSRYPKSLRHLHLREGNLFKEENPINNHAVKSLASW

ALGLIADLSWRYGPFDEIILETTRDALPEKIRKEIDKAMREREKALDKIIIGKYKKEFPSIDKRL
 ARKIQLWERQKGLDLYSGKVINLSQLLDGSADIEHIVPQSLGGLSTDYNTIVTLKSVNAAGNR
 LPGDWLAGNPDYRERIGMLSEKGLIDWKKRKNLLAQSLDEIYTENTHSGKIRATSYLEALVAQV
 LKRYYPFPDPPELRKNGIGVRMIPGKVTSKTRSLGKIKSRETNFHHAEADALILSTLTRGWQNR
 5 LHRMLRDNYGKSEAELKELWKKYMPHIEGLTLADYIDEAFRRFMSKGEESLFYRDMFDTIRSIS
 YWVDKKPLSASSHKETVYSSRHEVPTLRKNILEAFDSLNVIKDRHKLTTEEFMKRYDKEIRQKL
 WLHRIGNTNDESYRAVEERATQIAQILTRYQLMDAQNDKEIDEKFQQALKELITSPIEVTGKLL
 RKMRFVYDKLNAMQIDRGLVETDKNMLGIHISKGPNEKLIFRRMDVNNAHELQKERSGILCYLN
 EMLFIFNKKGLIHYGCLRSYLEKGQGSKYIALFNPRFPANPKAQPSKFTSDSKIKQVGIGSATG
 10 I IKAHLDDLGHVRSYEVFGTLPEGSIEWFKEESGYGRVEDDPHH

SEQ ID NO: 366

MRPIEPWILGLDIGTDSLGWAVFSCEEKGPPTAKELLGGGVRLFDSGRDAKDHTSRQAERGAFR
 RARRQTRTWPWRRDRLIALFQAAGLTPPAETRQIALALRREAVSRPLAPDALWAALLHLAHR
 15 GFRSNRIDKRERAAAKALAKAKPAKATAKATAPAKEADDEAGFWEGAEALRQMAASGAPT
 ALLADDLDRGQPVMRYNQSDRDGVVAPTRALIAEELAEIVARQSSAYPGLDWPVATRLVLDQR
 PLRSKGAGPCAFLPGEDRALRALPTVQDFIIRQTLANLRLPSTSADEPRPLTDEEHAKALALLS
 TARFVEWPALRRALGLKRGVKFTAETERNGAKQAARGTAGNLTEAILAPLIPGWSGWDLDRKDR
 VFSDLWAARQDRSALLALIGDPRGPTRVTEDETAEAVADAIQIVLPTGRASLSAKAARATAQAM
 20 APGIGYDEAVTLALGLHHSHRPRQERLARLPYAAALPDVGLDGDVPVGPPEADDGAAAEAYYG
 RIGNISVHIALNETRKIVNALLHRHGPIRLVMVETTRELKAGADERKRMIAEQAERERENAEI
 DVELRKSDRWMANARERRQVRRLARRQNNLCPYTSTPIGHADLLGDAYDIDHVIPLARGGRDSL
 DNMLVCQSDANKTKGDKTPWEAFHDKPGWIAQRDDFLARLDPQTAKALAWRFADDAGERVARKS
 AEDEDQGF LPRQLTDTGYIARVALRYLSLVTNEPNAVATNGRLTGLLRLAWDITPGPAPRDL
 25 PTPRDALRDDTAARRFLDGLTPPPLAKAVEGAVQARLAALGRSRVADAGLADALGLTLASLGGG
 GKNRADHRHHFIDAAMIAVTTTRGLINQINQASGAGRILDLRKWPRTNFEPYPYTFRAEVMKQWD
 HIIHPSIRPAHRDGGSLHAATVFGVRNRPDARVLVQRKPVEKLF LDANAKPLPADKIAEIIDGFA
 SPRMAKRFKALLARYQAAHPEVPPALAALAVARDPAFGPRGMTANTVIAGRSDDGEDAGLITP
 FRANPKAAVRTMGNAVYEVWEIQVKGRPRWTHRVLTRFDRTQPAPPPPPENARLVMRLRRGDLV
 30 YWPLESGLRFLVKKMAVDGRLALWPARLATGKATALYAQLSCPNNINLNGDQGYCVQSAEGIRK
 EKIRTTSTALGRLRLSKKAT

SEQ ID NO: 367

MKYTLGLDVGIASVGWAVIDKDNKIIDLGVRFCFKAEESEKTESLATARRIARGMRRRISRRS
 35 QRLRLVKKLFVQYEEIKDSSEFNRIFDTSRDGWKDPWELRYNALSRIKPYELVQVLTHITKRR
 GFKSNRKEDLSTTKEGVVITSIKNNSEMLRTKNYRTIGEMIFMETPENSNNKRNKVDEYIHTIAR
 EDLLNEIKYIFSIQRKLGSPFVTEKLEHDFLNIWEFQRPFASGDSILSKVGKCTLLKEELRAPT
 SCYTSEYFGLLQSINNVLVLEDNNTLTNLNDQRAKII EYAHFKNEIKYSEIRKLLDIEPEILFK
 AHNLTHTKNPSGNNESSKKFYEMKSYHKLKSTLPTDIWGLKHSNKESLDNLFYCLTVYKNDNEIKD
 40 YLQANNLDYLI EYIAKLPTFNKFKHLSLVAMKRIIPFMEKGYKYS DACNMAELDFTGSSKLEKC
 NKLTVETPIIENVTPVIRALTQARKVINAI IQKYGLPYMVNIELAREAGMTRQDRDNLKKEHE
 NNRKAREKISDLIRQNGRVASGLDILKWRLWEDQGGRCAYSGKPIPVCDLLNDSLTDIDHIYPY
 SRSMDDSYMKNVLVLTDENQNKRSYTPYEVWGSTEKWEDEARIYSMHL PQSKEKRLNRFIT
 KDLDSEFISRNLDTRYISRFLKNYIESYLQFSNDSPKSCVVCVNGQCTAQLRSRWGLNKNREES
 45 DLHHALDAAVIACADRKIIKEITNYNERENHNYKVYPLPWH SFRQDLMETLAGVFI SRAPRR
 KITGPAHDETIRSPKHFNKGLTSVKIPLTTVTLEKLETMVKNKTKGGISDKAVYNVLKNRLIEHN
 NKPLKAF AEKIYKPLKNGTNGAIIRSIRVETPSYTG VFRNEGKGISDNSLMVRVDVFKKKDKYY

LVPIYVAHMIKKELPSKAIVPLKPESQWELIDSTHEFLFSLYQNDYLVIKTKKGITEGYRSCH
RGTGSLSLMPHFANNKNVKIDIGVRTAISIEKYNVDILGNKSIVKGEPRRGMEKYNSFKSN

SEQ ID NO: 368

5 MIRTLGIDIGIASIGWAVIEGEYTDKGLNKEIVASGVRVFTKAENPKNKESSLALPRTLARSAR
RRNARKKGRIQQVKHYLSKALGLDLECFVQGEKLATLTFQTSKDFLSPWELRERALLYRVLDKEEL
ARVILHIAKRRGYDDITYGVEDNDSGKIKKAI AENSKRIKEEQCKTIGEMMYKLYFQKSLNVRN
KKESYNRCVGRSELREELKTIFQIQQELKSPWVNEELIYKLLGNPDAQSKQEREGLIFYQRPLK
10 GFGDKIGKCSHIKKGENSEPYRACKHAPSAAEFVALTKSINFLKNLTNRHGLCF SQEDMCVYLK
ILQEAQKNEKGLTYSKLKL LLDLP SDFEFLGLDYS GKNPEKAVFLSLPSTFKLNKITQDRKTQD
KIANILGANKDWEAILKELES LQLSKEQIQTIKDAKLNFSKHINLSLEALYHLLPLMREGKRYD
EGVEILQERGIFSKPQPKNRQLLP LSELAKEESYFDIPNPVLRRALSEFRKVVNALLEKYGGF
HYFHIELTRDVCKAKSARMQLEKINKKNKSENDAASQLLEVLGLPNTYNNRLKCKLWKQQEEYC
15 LYSGEKITIDHLKDQRALQIDHAFPLSRSLDDSQSNKVLCLTSSNQEKS NKTPYEWLGSDEKKW
DMYVGRVYSSNFSPSKKRKL TQKNFKERNEEDFLARNLVD TGYIGRVTKEYIKHSLSFLPLPDG
KKEHIRIISGSMTSTMR SFWGVQEKNRDHHLHHAQDAIIACIEPSMIQKYTTYLKDKETHRLK
SHQKAQILREGDHKLSLRWPM SNFKDKIQESI QNIIPSHHVSHKVTGELHQETVRTKEFYQAF
GGEEGVKKALKFGKIREINQGIVDNGAMVRVDIFKSKDKGKFYAVPIYTYDFAIGKLPNKAIVQ
20 GKKNGI IKDWLEMDENYEFCSLFKNDCIKIQTKEMQEAVLAIYKSTNSAKATIELEHLSKYAL
KNEDEEKMF TDT DKEKNKMTRESCGIQGLKVFQKVKLSVLGEVLEHKPRNRQNIALKTT PKHV

SEQ ID NO: 369

MYSIGLDIGIASVGWSVINKDKERIEDMGVRIFQKAENPKDGSSSLASSRREKRGSRRRNRKK
HRLDRIKNILCESGLVKKNEIEKIYKNAYLKSPWELRAKSLEAKISNKEIAQILLHIAKRRGFK
25 SFRKTRDNADDTGKLLSGIQENKKIMEEKG YLTIGDMVAKDPKFNT HVRNKAGSYLFSFSRKL
EDEVRKIQAKQKELGNTHFTDDVLEKYIEVFNSQRNFDEGPSKPSPYSEIGQIAKMIGNCTFE
SSEKRTAKNTWSGERFVFLQKLNNFRIVGLSGKRPLTEEERDIVEKEVYLKKEVRYEKLKILY
LKEEERFGDLNYSKDEKQDKKTEKTKFISLIGNYTIKKLNLSEK LKSEIEEDKSKLDKII EILT
30 FNKSDKTIESNLKKLELSREDIEILLSEEFSGTLNLSLKAIAKKILPYLEKGLSYNEACEKADYD
YKNNGIKFKRGELLPVVDKDLIANPVVLRAISQTRKVVNAIIRKYGTPHTIHVEVARDLAKSYD
DRQTI IKENKKRELENEKTKKFISEEFGIKNVKGKLL LKYRLYQEQEGRCAYS RKELSLSEVIL
DESMTDIDHIIPYSRSMDDSYSNKVLVLSGENRKKSNLLPKEYFDRQGRDWDTFVLNVKAMKIH
PRKKSNNLLKEKFTREDNKDWKSRALNDTRYISRFVANYLENAL EYRDDSPKKRVFMIPGQLTAQ
35 LRARWRLNKVRENGDLHHALDAAVVAVTDQKAINNISNISRYKELKNCKDVIPSIEYHADEETG
EVYFEEVKDTRFPMPSWGF DLELQKRLESENPREEFYNLLSDKRYLGWFNYEEGFIEKLRPVFV
SRMPNRGBVKGAHQETIRSSKKISNQIAVSKKPLNSIKLDLEKMQGRDTRDKLYEALKNRLEE
YDDKPEKAFAEPFYKPTNSGKRGPLVRGIKVEEKQNVGVYVNGGQASNGSMVRIDVFRKNGKFY
TVPIYVHQ TLLKELPNRAINGKPYKDWDLIDGSFEFLYSFY PNDLIEIEFGKSKSIKNDNKLTK
40 TEIPEVNLSEVLGYRGM DTSTGAATIDTQDGKI QMRIGIKTVKNIKKYQVDVLGNVYKVKREK
RQTF

SEQ ID NO: 370

MSKKVSRRYEEQAQEICQRLGSRPYSIGLDLGVSIGVAVAAAYDPIKKQPSDLVFVSSRIFIPS
TGAAERRQKRGQRNSLRHRANRLKFLWKLLAERNLMLS YSEQDVPDPARLRFEDAVVRANPYEL
45 RLKGLNEQLTLSELGYALYHIANHRGSSSVRTFLDEEKSSDDKLEEQQAMTEQLAKEKGISTF
IEVLTAFTNTGLIGYRNSESVKSKGVVPVTRDIISNEIDVLLQTQKQFYQEILSDEYCDRIVSA
ILFENEKIVPEAGCCPYFPDEKKLPRCHFLNEERRLWEAINNARIKMPMQEGA AKRYQSASFSD

EQRHILFHIARSGTDITPKLVQKEFPALKTSIIVLQGKEKAIQKIAGFRFRRLLEEKSFWKRLSE
 EQKDDFFSAWTNTPDDKRLSKYLMKHLLLTENEVDALKTVSLIGDYGPIGKTATQLLMKHLED
 GLTYTEALERGMETGEFQELSVWEQQSLLPYYGQILTGSTQALMGKYWHSFAFKEKRDSEGFFKP
 NTNSDEEKYGRIANPVVHQTLNELRKL MNELITILGAKPQEITVELARELKVGAEKREDIIKQQ
 5 TKQKEKEAVLAYSKEYCEPNNDKRYIERFRILLEDQAFVPCYCLEHISVADIAAGRADVDFHIFPRD
 DTADNSYGNKVVAHRQCNDIKGKRTPYAAFSTSAWGPIIMHYLDETPGMWRKRKRKFETNEEEYA
 KYLQSKGFVSRFESDNSYIAKAAKEYLRCLFNPNNTAVGSLKGMETSILRKAWNLLQGIDDLLG
 SRHWSKDADTSPTMRKNRDDNRHHGLDAIVALYCSRLVQMINTMSEQ GKRAVEIEAMIPIPGY
 ASEPNLSFEAQRELFRKKILEFMDLHAFVSMKTDNDANGALLKDTVYSILGADTQGEDLVFVVK
 10 KKI KDIGVKIGDYEEVASAIRGRITDKQPKWYPMEMKDKIEQLQSKNEAALQKYKESLVQAAAV
 LEESNRKLIESGKKPIQLSEKTISSKALELVGGYYYLISSNNKRTKTFVVKEPSNEVKGFADFTG
 SNLCLDFYHDAQGKLCGEIIRKIQAMNPSPYKPAYMKQGYSLYVRLYQGDVCELRASDLTEAESN
 LAKTTHVRLPNAKPGRTFVIIITFTEMGSGYQIYFSNLAKSKKGQDTSFTLTITIKNYDVRKVQL
 SSAGLVRYVSPLLVDKIEKDEVALCGE

SEQ ID NO: 371

MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPPEANVENNEGRRSKRGSRRRLKRRRIHRL
 ERVKKLLEDYNLLDQSQIPQSTNPYAIRVKGLSEALSKDELVIALHIAKRRGIHKIDVIDSND
 DVGNELSTKEQLNKN SKLLKDKFVCQIQLERMNEGQVRGEKNRFTADI I KEIIQLLNQKNFH
 20 QLDENFINKYIELVEMRREYFEGPGKGSFYGWEGDPKAWYETLMGHCTYFPDELRSVKYAYSAD
 LFNALNDLNNLVIQRDGLSKLEYHEKYHIIENVFKQKKKPTLKQIANEINVNPEDIKGYRITKS
 GKPQFTEFKLYHDLKSVLFDQSILENEDVLDQIAEILTIYQDKDSIKSKLTELDILLNEEDKEN
 IAQLTGTYTGTHRLSLKCI RLVLEEQWYSSRNQMEIFTHLNIKPKKINLTAANKIPKAMIDEFIL
 SPVVKRTFGQAINLINKIIEKYGVPEDIIEELARENNSKDKQKF INEMQKKNENTRKRINEIIG
 25 KYGNQNAKRLVEKIRLHDEQEGKCLYSLESIPLEDLLNNPNHYEVDHII PRSVSFDNSYHNKVL
 VKQSENSKKSNTLPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFEVQ
 KEFINRNLVDTRYATREL TNYLKAYFSANNMNVKVKTINGSF TDYLRKVWKFKKERNHGYKHA
 EDALIIANADFLFKENKKLKAVNSVLEKPEIESKQLDIQVDSEDNYSEMFIIPKQVQDIKDFRN
 FKYSHRVDKKPNRQLINDTLYSTRKKDNSTYIVQTIKDIYAKDNTTLKKQF DKSPEKFLMYQHD
 30 PRTFEKLEVIMKQYANEKNPLAKYHEETGEYLT KYSKKNNGPIVKS LKYIGNKLGSHLDVTHQF
 KSSTKKLVKLSIKPYRFDVYLT DKG YKFITISYLDVLKKNYIIPEQKYDKLKLGAIDKNAK
 FIASFYKNDLIKLDGEIYKIIIGVNSDTRNMIELDLPDIRYKEYCELNNIKGEPRIKKTIGKKVN
 SIEKLTTDVLGNVFTNTQYTKPQLLFKRGN

SEQ ID NO: 372

MIMKLEKWRLGLDLGTNSIGWSVFSLDKDNSVQDLIDMGVRIFSDGRDPKTKEPLAVARRTARS
 QRKLIYRRKLRRKQVFKFLQEQLFPKTKEECMTLKS LN PYELRIKALDEKLEPYELGRALFNL
 AVRGRGFKSNRKDGSREEVSEKKSPDEIKTQADMQTHLEKAIKENGCRITITEFLYKNQGENGGIR
 FAPGRMTYYPTRKMYEEEFNLIRSKQEKYYPQVDWDDIYKAIFYQRPLKPQQRGYCIYENDKER
 40 TFKAMPCSQKLRLQDIGNLAYYEGGSKKRVELNDNQDKVLYELLNSKDKVTFDQMRKALCLAD
 SNSFNLEENRDFLIGNPTAVKMRSKNRFGKLWDEIPLLEEQDLIIETIITADEDDAVYEVIKKYD
 LTQEQRDFIVKNTILQSGTSM LCKEVSEKLVKRLEEIADLKYHEAVESLGYKFADQTVKEYDLL
 PYYGKVLPGSTMEIDL SAPETNPEKH YGKISNPTVHVALNQTRVVVNALIKY GKP SQIAIELS
 RDLKNNVEKKAEIARKQNQRAKENIAINDTISALYHTAFPGKSFYPNRNDRMKYRLWSELGLGN
 45 KCIYCGKGISGAELFTKEIEIEHILPFSRTLLDAESNLTVAHSSCNAFKAERSPF EAFGTNPSG
 YSWQEIIQ RANQLKNTSKKNKFSNAMDSFEKDSSFIARQLSDNQYIAKAALRYLKCLVENPSD
 VWTNGSMTKLLRDKWEMDSILCRKFTEKEVALLGLKPEQIGNYKKNRFDHRHHAIDAVVIGLT

DRSMVQKLATKNSHKGNRIEIPEFPILRSDLIEKVKNIIVVSFKPDHGAEGKLSKETLLGKIKLH
 GKETFVCRENIVSLSEKNLDDIVDEIKSKVKDYVAKHKGQKIEAVLSDFSKEGKIKKVRVNRV
 QTPIEITSGKISRYLSPEDYFAAVIWEIPGEKKTFAQYIRRNEVEKNSKGLNVVKPAVLENGK
 PHPAAKQVCLLHKDDYLEFSDKGKMYFCRIAGYAATNNKLDIRPVYAVSYCADWINSTNETMLT
 5 GYWKPTPTQNWVSVNVLFDKQKARLVTVSPIGRVFRK

SEQ ID NO: 373

MSSKAIDSLEQLDLFKPQEYTLGLDLGIKSIGWAILSGERIANAGVYLFETAEEELNSTGNKLIS
 KAAERGRKRRIRRM LDRKARRGRHIRYLLEREGLPTDELEE VVHQS NR TLWDVRAE AVERKLT
 10 KQELA AVL FHLVRHRGYFPNTKKLPDDES DSADEEQGKINRATSRLREELKASDCKTIGQFLA
 QNRDRQRNREGDYSNL MARKLVFEEALQILAFQRKQGHEL SKDFEKT YLDVLMGQSRGRSPKLG
 NCSLIPSEL RAPSSAPSTEWFKFLQNLGNLQISNAYREEWSIDAPRAQIIDACSQRSTSSYWQ
 IRRDFQIPDEYRFNLVNYERRDPDVLQEYLLQQQERKTLANFRNWKQLEKIIGTGHP IQTLDEA
 ARLITLIKDDEKLS DQLADLLPEASDKAITQLCELDFTTAAKISLEAMYRILPHMNQGMGFFDA
 15 CQQESLPEIGVPPAGDRVPPFDEMYNPVNRVLSQSRKLINAVIDEYGMPAKIRVELARDLGKG
 RELRERIKLDQLDKSKQNDQRAEDFRAEFQQA PRGDQSLRYRLWKEQNCTCPYSGRMIPVNSVL
 SEDTQIDHILPISQSFDNSLSNKVLCFTEENAQKSNRTPFEYLDAA DFQRLEAISGNWPEAKRN
 KLLHKSFGKVAEEWKSRA LNDTRYLTSALADHLRHHL PDSKIQT VNGRITGYLRKQWGLEKDRD
 KHTHHAVDAIIVVACTTPAIVQQVTLYHQDIRRYKKLGEKRPTPWPETFRQDVL DVEE EIFITRQ
 20 PKKVSGGIQTKD TLRKHSKPDRQRVALTKVKLADLERLVEKDASNRNLYEHLKQCLEESGDQP
 TKAFKAPFYMPSGPEAKQRPILSKVTLLREKPEPPKQLTELSGGRRYDSMAQGRLDIYRYKPGG
 KRKDEYRVVLQRMIDLMRGEENVHVFQKGV PYDQGPEIEQNYTFLFSLYFDDLVEFQRSADSEV
 IRGYRTFN IANGQLKISTYLEGRQDFDFGANRLAHFAKVQVNLLGKVIK

25 SEQ ID NO: 374

MRS LRYRLALDLGSTSLGWALFR LDACNRPTAVIKAGVRIFSDGRNPKDGSS LAVTRRAARAMR
 RRRDRLLKRKTRMQAKLVEHGFFPADAGKRKALEQLNPYALRAKGLQEALLPGEFARALFHINQ
 RRGFKSNRKTDKKDNDSGVLKKAIGQLRQQMAEQGSRTVGEYLLWTRLQQGGVRARYREKPYTT
 EEGKKRIDKSYDLYIDRAMIEQEF DALWAAQA AFNPTLFHEAARADLKD TLLHQ RPLRPVKPGR
 30 CTLLPEEERAPLALPSTQRFRIHQEVNHLRLLDENLREVALTLAQRD AVVTALET KAKLSFEQI
 RKLLKLSGSGVQFNLEDAKRT ELKGNATSAALARKELFGAAWSGFDEALQDEIVWQLVTEEGEGA
 LIAWLQTHTGVD EARAQAIVDVSLPEGYGNLSRKALARIVPALRAAVITYDKAVQAAGFDHHSQ
 LGFEYDASEVEDLVHPETGEIRSVFKQLPYYGKALQRHVAFGSGKPEDPDEKRYGKIANPTVHI
 GLNQVRMVVNALIRRYGRPTEVVIELARDL KQSREQKVEAQRQADNQR RNARIRRSIAEVLGI
 35 GEERVGRSDIQKWICWEE LSFDAADRCPYSGVQISAA MLLSDEVEVEHILPFSKTLDDSLNNR
 TVAMRQANRIKRN RTPWDARAEFEAQGWSYEDILQRAERMPLRKRYRFAPDGYERWLGDDKDFL
 ARALNDTRYLSRVAAEYLR LVCPGTRVIPGQLTALLRGKFG LNDVLGLDGEKNRNDHRHHA VDA
 CVIGVTDQGLMQRFATASAQARGDGLTRLVDGMPMPWPTYRDHVERAVRHIWVSHRPDHGFEGA
 MMEETSYGIRKDGSIKQRRKADGSAGREISNLIRIHEATQPLRHGVSADGQPLAYKGYVGGSNY
 40 CIEITVNDKGKWEGEVISTFRAYGVVRAGGMGR LRNPHEGQNGRKLIMRLVIGDSVRLEVDGAE
 RTMRIVKISGSNGQIFMAPIHEANVDARNTDKQDAFTYTSKYAGSLQAKTRRVTISPIGEVRD
 PGFKG

SEQ ID NO: 375

45 MARPAFRAPRREHVNGWTPDPHRISKPF FILVSWHLLSRVVIDSSSGCFPGTSRDHTDKFAEWE
 CAVQPYRLSFDLGTNSIGWGLLNLD RQGKPREIRALGSRIFSDGRDPQDKASLAVARRLARQMR
 RRRDRYLTRRTRL MGALVRFGLMPADPAARKRLEVAVDPYLARERATRERLEPFEIGRALFHLN

QRRGYKPVRTATKPDDEEAGKVKEAVERLEAAIAAAGAPTLGAWFAWRKTRGETLRARLAGKGKE
 AAYPFYPARRMLEAEFDTLWAEQARHHPDLLTAEAREILRHRIFHQRP LKPPPVGRC TLYPDDG
 RAPRALPSAQRLRLRFQELASLRVIHLDL SERPLTPAERDRIVAFVQGRPPKAGRKPGKVQKSVP
 FEKLRGLLELPPGTGFSLES DKRELLGDET GARIAPAFGPGWTALPLEEQDALVELLLTEAEP
 5 ERAIAALTARWALDEATAAKLAGATLPDFHGRYGRRAVAELLPVLERETR GDDPDGRV RPIRLDE
 AVKLLRGGKDHSDFSREGALLDALPYYGAVLERHVAFGTGNPADPEEKRVGRVANPTVHIALNQ
 LRHLVNAILARHGRPEEIVIELARDLKRS AEDRRREDKRQADNQKRNEERKRLILSLGERPTPR
 NLLKLRLWEEQGPVENRRCPYSGETISMRMLLSEQVDIDHILPFSVSLDDSAANKVVCLREANR
 IKRNRSPWEAFGHDSERWAGILARAEALPKNKRWRFPDALEKLEGEGLRARHLNDTRHLSRL
 10 AVEYLRCVCPKVRVSPGRLTALLRRRWGIDAILAEADGPPPEVPAETLDPSPA EKNRADHRHHA
 LDAVVIGCIDRSMVQRVQLAAASA EREAAA REDNI RRVLEGFKEEPWDGFRAELERRARTIVVS
 HRPEHGIGGALHKETAYGVPDPPEEGFNLVVRKPIDGLSKDEINSVRDPRLRRALIDRLAIRRR
 DANDPATALAKAAEDLAAQPASRGIRRVRLKKE SNPIRVEHGGNPSGPRSGGPFHKL LLLAGEV
 HHVDVALRADGRRWVGHWVTLFEAHGGRGADGAAAPPR LGDGERFLMRLHKGDCLKLEHKGRVR
 15 VMQVVKLEPSSNSVVVVEPHQVKTD RSKHVKISCDQLRARGARRVTVDPLGRVRVHAPGARVGI
 GGDAGRTAMEPAEDIS

SEQ ID NO: 376

MKRTSLRAYRLGVDLGANSLGWFFVWLDDHGQPEGLGPGGVRI FPDGRNPQSKQSNAAGRRLAR
 20 SARRRRDRYLQRRGKLMGLLVKHGLMPADEPARKRLECLDPYGLRAKALDEVLP LHHVGRALFH
 LNQRRGLFANRAIEQGDKDASAIKAAAGRLQTSMQACGARTLGEFLNRRHQ LRATVRARSPVGG
 DVQARYEFYPTRAMVDAEFEAIWAAQAPHHTMTAEAHDTIREAIFSQRAMKRPSIGKCSLDPA
 TSQDDVDGFRCAWSHPLAQRFRIWQDV RNLA VVETGPTSSRLGKEDQDKVARALLQTDQLSFDE
 IRGLLGLPSDARFNLES DRRDHLKG DATGAILSARRHFGPAWHDRSLDRQIDIVALLESALDEA
 25 AIIASLGTTHSLDEAAAQRALSALLPDGYCRLGLRAIKRVLPLMEAGRTYAEASAAGYDHALL
 PGGKLSPTGYLPYYGQWLQNDVVGSDDERDTNERRWGRLPNPTVHIGIGQLRRVVNELIRWHGP
 PAEITVELTRDLKLSPRRLAELEREQAENQRKNDKRTSLLRKLGLPASTHNLLKLRLWDEQGDV
 ASECPYTGEAIGLERLVSDDDVIDHLIPFSISWDDSAANKVVC MRYANREKGNRTPF EAFGHRQ
 GRPYDWADIAERAARLPRGKRWRFGPGARAQFEELGDFQARLLNETSWLARVAKQYLA AVTHPH
 30 RIHVLPGRLTALLRATWELNDLLPGSDDRAAKSRKDHRHHAIDALVAALTDQALLRRMANAHDD
 TRRKIEVLLPWPTFRIDLETRLKAMLVSHKPDHGLQARLHEDTAYGTVEHPETEDGANLVYRKT
 FVDISEKEIDRIRDRLRLDLVRAHVAGERQQGKTLKA AVLSFAQRRDIAGHPNGIRHVRLTKSI
 KPDYLVPIRDKAGRIYKSYNAGENAFVDILQAESGRWIARATTVFQANQANESH DAPAAQPIMR
 VFKGDMRLRIDHAGA EKFKIVRLSPSNNLLYLVEHHQAGVFQTRHDDPEDSFRWLFASF DKLRE
 35 WNAELVRIDTLGQPWRRKRGL ETGSEDATRIGWTRPKKWP

SEQ ID NO: 377

MERIFGFDIGTTSIGFSVIDYSSTQSAGNIQRLGVRI FPEARPDGTPLNQQRRQKRMRRQLR
 RRRIRRKALNETLHEAGFLPAYGSADWPVVMAD EPYELRRRGLEEGLSAYEFGRAIYHLAQHRH
 40 FKGRELEESDTPDPDV DDEKEAANERAATLKALKNEQTTLGAWLARRPPSDRKRGIHAHRNVVA
 EEFERLWEVQSKFHPALKSEEMRARI SDTIFAQRPVFWRKNTLGECRFMPGEPLCPKGSWLSQQ
 RRMLEKLNNLAIAGGNARPLDAEERDAILSKLQQQASMSWPGVRSALKALYKQRGEPGA EKSLK
 FNLELGGESKLLGNAL EAKLADMFGPDWPAHPRKQ EIRHAVHERLWAADYGETPDKKRVIILSE
 KDRKAHREAAANSFVADFGITGEQAAQLQALKLPTGWEPYSIPALNLF LAELEKGERFGALVNG
 45 PDWEGWRRTNFPHRNQPTGEILDKLPSPASKEERERISQLRNPTVVRTQNELRKVVNNLIGLYG
 KPDRIRIEVGRDVGKSKREREEIQSGIRNEKQRKKATEDLIKNGIANPSRDDVEKWILWKEGQ
 ERCPYTG DQIGFNALFREGRYEVEHIWPRSRSFDNSPRNKTLCKDVNIEKGNRMPFEAFGHDE

DRWSAIQIRLQGMVSAKGGTGMSPGKVKRFLAKTMPEDFAARQLNDTRYAAKQILAQLKRLWPD
 MGPEAPVKVEAVTGQVTAQLRKLWTLNNILADDGEKTRADHRHHAIDALTVACTHPGMTNKLRSR
 YWQLRDDPRAEKPALTPPWDTIRADA EKAVSEIVVSHRVRKKVSGPLHKETTYGDTGTDIKTKS
 GTYRQFVTRKKIESLSKGELDEIRDPRIKEIVA AHVAGRGGDPKKAFFPYPCVSPGGPEIRKVR
 5 LTSKQQNLNMAQTGNGYADLGSNNHIAIYRLPDGKADFEIVSLFDASRRLAQRNP IVQRTRADG
 ASFVMSLAAGEAIMIPEGSKKGIWIVQGVWASGQVVLERD TDADHSTTTRPMPNPILKDDAKKV
 SIDPIGRVVRPSND

SEQ ID NO: 378

10 MNKRILGLDTGTNSL GWAVVDWDEHAQSYELIKYGDVIFQEGVKIEKGISSKAAERSGYKAIR
 KQYFRRRLRKIQVLKVLVKYHLCPYLSDDDLRQWHLQKQYPKSDELMLWQRTSDEEGKNPYDR
 HRCLHEKLDLTVEADRYTLGRALYHLTQRRGFLSNRLDTSADNKEDGVVKS GISQLSTEMEEAG
 CEYLG DYFYKLYDAQGNKVRIRQRYTDRNKHYQHEFDAICEKQELSS ELIEDLQRAIFFQLPLK
 15 SQRHGVGRCTFERGKPRCADSHPDYEEFRMLCFVNNIQVKGPHDLELRPLTYEEREKIEPLFFR
 KSKPNDFDFEDIAKALAGKKNYAWIHDKEERAYKFNYRMTQGVPGCPTIAQLKSIFGDDWKTGIA
 ETYT LIQKKNGSKSLQEMVDDVWNVLYSFSSVEKLKEFAHHLQLDEESA EKFAKIKLSHSFAA
 LSLKAIRKFLPFLRKGMYYTHASFFANIPTIVGKEIWNKEQNRKYIMENVGELVFNYQPKHREV
 QGTIEMLIKDFLANNFELPAGATDKLYHPSMIETYPNAQRNEFGILQLGSPRTNAIRNPMAMRS
 LHILRRVNVNQLLKESIIDENTEVHVEYARELNDANKRRAIADRQKEQDKQHKKYGD EIRKLYKE
 20 ETGKDIEPTQTDVLKFQLWEEQNHHCLYTGEQIGITDFIGSNPKFDIEHTIPQSVGGDSTQMNL
 TLCDNRFNREVKKAKLPTEL ANHEEILTRIEPWKNKYEQLVKERDKQRTFAGMDKAVKDIRIQK
 RHKLQMEIDYWRGKYERFTMT EVPEGFSRRQGTGIGLISRYAGLYLKS LFHQADSRNKS NVYVV
 KGVATAEFKRMWGLQSEYEKKCRDNHSHHCDAITIACIGKREYDLMAEYYRMEETF KQGRGSK
 PKFSKPWATFTEDVLNIYKNLLVVDTPNNMPKHTKKYVQTSIGKVLAQGD TARGSLHLD TYYG
 25 AIERDGEIRYVVRRLSSFTKPEELENIVDETVKRTIKEAIADKNFKQAIAEPIYMNEEKGILI
 KKVRCFAKSVKQPINIRQHRDL SKKEYKQYHVMNENNYLLAIYEGLVKNKV VREFEIVSYIEA
 AKYYKRSQDRNIFSSIVPTHSTKYGLPLKTKLLMGQLVLMFEENPDEIQVDNTKDLVKRLYKVV
 GIEKDGRIKFKYHQEARKEGLPIFSTPYKNNDDYAPIFRQSINNINILVDGIDFTIDILGKVTL
 KE

30 SEQ ID NO: 379
 MNYKMGLDIGIASVGWAVINLDLKRIEDLGVRIFDKAEHPQNGESLALPRRIARSARRRLRRRK
 HRLERIRRLVSENVLTKEEMNLLFKQKKQIDVWQLRVDALERKLNDELARVLLHLAKRRGFK
 SNRKSERNSKESSEFLKNIEENQSILAQYRSVGEMIVKDSKFAYHKRNKLD SYSNMIARDDLER
 35 EIKLIFEKQREFNNPVCTERLEEKYLNIWSSQRPFASKEDIEKKVGFC TFEPEKEKRAPKATYTF
 QSFIVWEHINKLRLVSPDETRALTEIERNLLYKQAFSKNKMTYYDIRKLLNLSDDIHFKGLLYD
 PKSSLKQIENIRFLELDSYHKIRKCIENVYGKDGIRMFNETDIDTFGYALTIFKDD EDIVAYLQ
 NEYITKNGKRVSNLANKVYDKSLIDELLNLSFSKFAHLSMKAIRN ILPYMEQGEIYSKACELAG
 YNFTGPKKKEKALLLPVIPNIANPVVMRALTQSRKVVAI IKKYGSPVSIHIELARDLSHSFDE
 40 RKKIQKDQ TENRKKNETAIKQLIEYELTKNPTGLDIVKFKLWSEQQGRCMYSLKPIELERLLEP
 GYVEVDHILPYSRSLDDSYANKVLVLTKENREKGNHTPVEYLG LGSERWKKFEKFVLANKQFSK
 KKKQNLLRLRYEETEEKEFKERNLNDTRYISKFFANFIKEHLKFADGDGGQKVYTINGKITAHL
 RSRWDFNKNREESDLHHA VDAVIVACATQGMIKKITEFYKAREQNKESAKKKEPIFPQPWPHFA
 DELKARLSKFPQESIEAFALGN YDRKKLESRLPVFVSRMPKRSVTGA AHQETLRRCVGIDEQSG
 45 KIQTAVKTKLSDIKLDKDGHPMYQKESDPRTYEAIQRLL EHNNDPKKAFQEPLYKPKKN GEP
 GPVIRTVKIIDTKNKVVHLDGSKTVAYNSNIVRTDVFEKDGKYYCVPVYTMDIMKGTL PNKAIE

ANKPYSEWKEMTEEYTFQFSLFPNDLVRIVLPREKTIKTSTNEEII IKDIFAYYKTIDSATGGL
ELISHDRNFSLRGVGSKTLKRFEKYQVDVLGNIHKVKGEKRVGLAAPTNOQKKGKTVDSLQSVSD

SEQ ID NO: 380

5 MRRLGLDLGTNSIGWCLLDLGDDGEPVSIFRTGARIFSDGRDPKSLGSLKATRREARLTRRRRD
RFIQRQKNLINALVKYGLMPADEIQRQALAYKDPYPIRKALDEAIDPYEMGRAIFHINQRRGF
KSNRKSADNEAGVVKQSIADLEMKLGEAGARTIGEFLADRQATNDTVRARRLSGTNALLYEFYPD
RYMLEQEFDTLWAKQAAFNPPLYIEAARERLKEIVFFQRKLKPQEVGRCIFLSDEDRIKALPS
10 FQRFRIYQELSNLAWIDHDGVAHRITASLALRDHLFDELEHKKKLTFKAMRAILRKQGVVDYPV
GFNLESDNRDHLIGNLTSCIMRDAKKMIGSAWDRLDEEEQDSFILMLQDDQKGDDEVRSILTQQ
YGLSDDVAEDCLDVRLPDGHGSLSKKAIDRILPVLRDQGLIYYDAVKEAGLGEANLYDPYAALS
DKLDYYGKALAGHVMGASGKFEDSDEKRYGTISNPTVHIALNQVRVAVNELIRLHGKPDDEVVIE
IGRDLPMGADGKRELERFQKEGRAKNERARDELKKLGHIDSRESRQKFQLWEQLAKEPVDRCPP
15 FTGKMMSISDLFSDKVEIEHLLPFSLTLDSDMANKTVCFRQANRDKGNRAPFDAFGNSPAGYDW
QEILGRSQNLPHYAKRWRFLPDAMKRFEADGGFLERQLNDTRYISRYTTEYISTIIPKNKIWVVT
GRLTSLLRGFWGLNSILRGHNTDDGTPAKKSRRDHRHHAIDAIIVVGMTSRGLLQKVSKAARRSE
DLDLTRLFEGRIDPWDGFRDEVKKHIDAIIVSHRPRKKSQALHNDTAYGIVEHAENGASTVVH
RVPITSLGKQSDIEKVRDPLIKSALLNETAGLSGKSFENAVQKWCADNSIKSLRIVETVSIPI
TDKEGVAYKGYKGDGNAYMDIYQDPTSSKWKGEIVSRFDANQKGFIPSWQSQFPTARLIMRLRI
20 NDLLKLQDGEIEEIIYRVQRLSGSKILMAPHTEANVDARDRDKNDTFKLTSKSPGKLQSASARKV
HISPTGLIREG

SEQ ID NO: 381

25 MKNILGLDLGLSSIGWSVIRENSEEQELVAMGSRVVS LTAAELSSFTQGNVGSINSQRTQKRTQ
RKGYDRYQLRRTLLRNKLDTLGMLPDDSLSYLPKLQLWGLRAKAVTQRIELNELGRVLLHLNQK
RGYKSIKSDFSGDKKITDYVKTVKTRYDELKEMRLTIGELFFRRLTENAFFRCKEQVYPRQAYV
EEFDCIMNCQRKFYPDILTDETRCIRDEIIYYQRPLKSCKYLVSRCEFEKRFYLNAAAGKKTEA
GPKVSPRTSPLFQVCRLWESINNIVVKDRRNEIVFISAEQRAALFDFLNTHEKLGSDLLKLLG
30 LSKTYGYRLGEQFKTGIQGNKTRVEIERALGNYPDKKRLQLFNLQEESSSMVNTETGEIIPMIS
LSFEQEPLYRLWHVLYSIDDREQLQSVLRQKFGIDDDDEVLERLSAIDLVKAGFGNKSSKAIRRI
LPFLQLGMNYAEACEAAGYNHNNYTKAENEARALLDRLPAIKKNELRQPVVEKILNQMVNVVN
ALMEKYGRFDEIRVELARELKQSKEERSNTYKSINKNQRENEQIAKRIVEYGVPTSRISQKYKM
WEESKHCCICYGQPVDVGDFLRGFDVEVEHIIPKSLYFDDSFANKVCSCRSCKEKNRRTAYDY
MKSKGEKALSDYVERVNTMYTNNQISKTKWQNLTPVDKISIDFIDRQLRESQYIARKAKEILT
35 SICYNVTATSGSVTSFLRHVWGWDTVLHDLNFDYKVKVGLTEVIEVNHRSVIRREQIKDWSKR
FDHRHHAIDALTIACKQAYIQRLNNLRAEEGPDFNKMSLERYIQSQPHFSVAQVREAVDRILV
SFRAGKRAVTPGKRYIRKNRKRISVQSVLIPRGALSEESVYGVIVHWEKDEQGHVIQKQRAVMK
YPITSINREMLDKEKVVDKRIHRILSGRLAQYNDNPKEAFAKPVYIDKECRIPRTVRCFAKPA
INTLVPLKKDDKGNPVAWVNPNNHHVAIYRDEDGKYKERTVTFWEAVDRCRVGIPAIVTQPDT
40 IWDNILQRNDISENVLESPLDVKWQFVLSLQQNEMFILGMNEEDYRYAMDQQDYALLNKYLYRV
QKLSKSDYSFRYHTETSVEDKYDGKPNLKL SMQMGLKRVSIKSLGLNPHKVHISVLGEIKEI
S

SEQ ID NO: 382

45 MAEKQHRWGLDIGTNSIGWAVIALIEGRPAGLVATGSRIFSDGRNPKDGSSLAVERRGPRQMRR
RRDRYLRRRDRFMQALINVGLMPGDAAARKALVTENPYVLRQRGLDQALTLPEFGRALFHLNQR
RGFQSNRKTDRTAKESGKVKNAIAAFRAGMGNARTVGEALARLEDGRPVARMVGQKDEHY

ELYIAREWIAQEFDALWASQQRFHAEVLADAARDRLRAILLFQRKLLPVPVVGKCFLEPNQPRVA
 AALPSAQRFRMLMQELNHLRVMTLADKRERPLSFQERNDLLAQLVARPKCGFDMRLRKIVFGANKE
 AYRFTIESERRKELKGCDTAAKLAKVNALGTRWQALSLEQDRLVCLLLDGENDAVLADALREH
 YGLTDAQIDTLLGLSFEDGHMRLGRSALLRVLDALESGRDEQGLPLSYDKAVVAAGYPAHTADL
 5 ENGERDALPYYGELLWRYTQDAPTAKNDAERKFGKIANPTVHIGLNQLRKLVNALIORYGKPAQ
 IVVELARNLKAGLEEKERIKKQQTANLERNERIRQKLQDAGVPDNRENRLRMRLFEEELGQGNGL
 GTPCIYSGRQISLQRLFSNDVQVDHILPFSKTLDDSFANKVLAQHDANRYKGNRGPFEAFGANR
 DGYAWDDIRARA AVLPRNKRNRFAETAMQDWLHNETDFLARQLTDTAYLSRVARQYLTAICSKD
 DVYVSPGRLTAMLRKWLNRVLDGVMEEQGRP AVKNRDDHRHHAIDAVVIGATDRAMLQQVAT
 10 LAARAREQDAERLIGDMPTPWPNFLEDVRAAVARCVVSHKPDHGPEGGLHNDTAYGIVAGPFED
 GRYRVRHRVSLFDLKPGLSNVRCDAPLQAELEPIFEQDDARAREVALTALAERYRQRKVVLEE
 LMSVLP IIRPRGEDGKTLPD SAPYKAYKGDSNYCYELFINERGRWDGELISTFRANQAAYRRFRN
 DPARFRRYTAGGRPLLMRLCINDYIAVGTA AERTIFRVVKMSENKITLAEHFEGGTLKQRDADK
 DDPFKYLT KSPGALRDLGARRIFVDLIGRVLDPGIKGD

SEQ ID NO: 383

MIERILGVDLGISSLGWAIVEYDKDDEAANRIIDCGVRLFTAETPKKKE SPNKARREARGIRR
 VLNRRRVRMNMIIKKLFLRAGLIQDVDLDGEGGMFYSKANRADVWELRHDGLYRLLKGDELARVL
 IHI AKHRGYKFIGDDEADEESGKVKKAGVVL RQNFEEAAGCRTVGEWLWRER GAN GKRNKHGDY
 20 EISIHRDLLVEEVEAIFVAQQEMRSTIATDALKAAYREIAFFVRPMQRIEKMVGHC TYFPEERR
 APKSAPTA EKFI AISKFFSTVIIDNEGWEQKIIERKTLEELLDFAVSREKVEFRHLRKFLDLSD
 NEIFKGLHYKGPKTAKKREATLFD PNEPTELEFDKVEAEKKAWISLRGA AKLREALGNEFYGR
 FVALGKHAD EATKILTYKDEGQKRRELTKLPLEAEMVERLVKIGFSDFLKL SLKAIRDILPAM
 ESGARYDEAVLMLGVPHKEKSAILPPLNKTDIDILNPTVIRAF AQFRKVANALVRKYGA FDRVH
 25 FELAREINTKGEIEDIKESQRKNEKERKEAADWIAETSFQVPLTRKNILKKRLYIQQDGRCAYT
 GDVIELERLFD EGYCEIDHILPRSRSADDSFANKVLC LARANQQKTDRTPYEWF GHDAARWNAF
 ETRTSAPS NRVRTGKGKIDRL LKKNFDENSEMAFKDRNLNDTRYMARAIKTYCEQYWFVFN SHT
 KAPVQVRSGLT SVLRYQWGLESKDRESHTHHAVDAII IAFSTQGMVQKLSEYYRFKETHREKE
 RPKLAVPLANFRDAVEEATRIENTETVKEGVEVKRL LISRPPRARVTGQAHEQTAKPYPRIKQV
 30 KNKKKWRLAPIDEEK FESFKADRVASANQKNFYETSTIPRVDVYHKKGKFHLVPIYLHEMVLNE
 LPNLSLGTNPEAMDENFFKFSIFKDDLISIQTQGT PKKPAKIIMGYFKNMHGANMVLSSINNSP
 CEGFTCTPVSMDKKHKDKCKLCPEENRIAGRCLQGFLDYWSQEGLRPPRKEFECDQGVKFALDV
 KKYQIDPLGYYYEVKQEKRLGTIPQMRS AKKLVKK

SEQ ID NO: 384

MNNSIKSKPEVTIGLDLGVGSVGWAIVDNETNIIHHLGSRLFSQAKTAEDRRSFRGVRR LIRR
 KYKLKRFVNL IWKYNSYFGFNKEDI LN NYQEQQKLHNTVLNLKSEALNAKIDPKALSWILHDY
 LKNRGHFYEDNRDFNVYPTKELAKYFDKYGYKGIIDSKEDNDNKLEELTKYKFSNKH WLEEV
 KKVLSNQTGLPEKFKEEYESLFSYVRNYSEGPGSINSVSPYGIYHLDEKEGKV VQYNNIWDKT
 40 IGKCNIFPDEYRAPKNSPIAMIFNEINELSTIRSYSIYLTGWFINQEFKKAYLNKLLDLLIKTN
 GEKPIDARQFKKLREETIAESIGKETLKDVENEEKLEKEDHKWKLKGLKLNTNGKI QYNDLSSL
 AKFVHKLKQHLKLDLLEDQYATLDKINFLQSLFVYLGKHLRYSNRVDSANLKEFSDSNKL FER
 ILQKQKDG LFKLFEQTDK DDEKILAQTHSLSTKAMLLAITRMTNLDNDEDNQKNNDKGWNFEAI
 KNFDQKFIDITKKNNNLSLKQNKRYLDDRFINDAILSPGVKRILREATKVFNAILKQFSEEYDV
 45 TKVVIELARELSEEKELENTKNYKKLIKNGDKISEGLKALGISEDEIKDILKSPTKSYKFLW
 LQQDHIDPYS LKEIAFDDIFTKTEKFEIDHIIPYSISFDDSSSNKLLVLAESNQAKSNQTPYEF
 ISSGNAGIKWEDYEAYCRKFKDGDSSLLDSTQRSKKFAKMMKTD TSSKYDIGFLARNLNDTRYA

TIVFRDALEDYANNHLVEDKPMFKVVCINGSVTSFLRKNFDDSSYAKKDRDKNIHHAVDASIIIS
 IFSNETKTLFNQLTQFADYKLFKNTDGSWKKIDPKTGVVTEVTDENWKQIRVRNQVSEIAKVIE
 KYIQDSNIERKARYSRKIENKTNISLFNDTVYSAKKVGIEDQIKRKNLKTLDIHESAKENKNSK
 VKRQFVYRKLVNVSLNNDKLADLFAEKEDILMYRANPWVINLAEQIFNEYTENKKIKSQNVFE
 5 KYMLDLTKEFPEKFSEFLVKSMRLNKTAIYDDKKNIHVRIKRLKMLSSELKENKLSNVIIRSK
 NQSGTKLSYQDTINSLALMIMRSIDPTAKKQYIRVPLNTLNLHLGDHDFDLHNMDAYLKKPKFV
 KYLKANEIGDEYKPWRVLTSGTLLIHKKDKKLMYISSFQNLNDVIEIKNLIETEKENDDDSDSK
 KKKKANRFLMTLSTILNDYILLDAKDNFDILGLSKNRIDEILNSKLGLDKIVK

10 SEQ ID NO: 385

MGGSEVGTVPVTWRLGVDVGERSIGLAAVSYEEDKPKEILAAVSWIHDGGVGDERSGASRLALR
 GMARRARLRRFRRARLRDLMLLSELGWTPLPDKNVSPVDAWLARKRLAEYVVDETERRLL
 GYAVSHMARHRGWRNPWTTIKDLKNLPQPSDSWERTRESLEARYSVSLEPGTVGQWAGYLLQRA
 PGIRLNPTQQSAGRAELSNATAFETRLRQEDVLWELRCIADVQGLPEDVVSVIDAVFCQKRP
 15 SVPAERIGRDPLDPSQLRASRACLEFQEYRIVA AVANLRIRDGSGSRPLSLEERNAVIEALLAQ
 TERSLTWSDIALEILKLPNESDLTSVPEEDGPSSLAYSQFAPFDETSARIAEFIKNRRKIPTF
 AQWWQEQRDTSRSDLVAALADNSIAGEEEQELLVHLPDAELEALEGLALPSGRVAYSRLTSLGL
 TRVMRDDGVDVHNARKTCFGVDDNWRPPLPALHEATGHPVVDRLAILRKFLSSATMRWGPPQS
 IVVELARGASESRERQAEAAAARRAHRKANDRIRAE LRASGLSDPSPADLVRARLLELYDCHCM
 20 YCGAPISWENSELDHIVPRTDGGSNRHENLAITCGACNKEKGRRPFASWAETSNRVQLRDVIDR
 VQKLKYSNGMYWTRDEFSRYKKS VVARLKRRTSDPEVIQSIESTGYAAVALRDRLLSYGEKNGV
 AQVAVFRGGVTAEARRWLDISIERLFSRVAIFAQSTSTKRLDRRHHAVDVAVLTTLTPGVAKTL
 ADARSRRVSAEFWRPSPDVNRHSTEPPQSPAYRQWKESCSGLGDLIISTAARDSIAVAAPLRLR
 PTGALHEETLRAFSEHTVGAAWKGAE LRRIVEPEVYAAFLALTDPGGRFLKVSPSEDVLPADEN
 25 RHIVLSDRVLGPRDRVKLFPDDRGSIRVRGGAAYIASFHARVFRWGSSHSPSFALLRVSLADL
 AVAGLLRDGVDVFTAELPPWTPAWRYASIALVKAVESGDAKQVGWLVPGEDELDFGPEGVTTAAG
 DL SMFLKYFPERHWVVTGFEDDKRINLKPAFLSAEQAEVLRTERS DRPDTLTEAGEILAQFFPR
 CWRATVAKVLCHPGLTVIRRTALGQPRWRRGHL PYSWRPWSADPWSSGGTP

30 SEQ ID NO: 386

MHNKKNITIGFDLGIASIGWAIIDSTTSKILDWGTTRTFEERKTANERRAFRSTRNRIRRKAYRN
 QRFINLILKYKDLFELKNISDIQRANKKD TENYEKIIISFFTEIYKKCAAKHSNILEVKVKALDS
 KIEKLDLIWILHDYLENRGFFYDLEEENVADKYEGIEHPSILLYDFFKKNNGFFKSNSSIPKDLG
 GYSF SNLQWVNEIKKLF EVQEINPEFSEKFLNLFTSVRDYAKGPGSEHSASEYGIFQKDEKGV
 35 FKKYDNIWDKTIGKCSFFVEENRSPVNYPSYEIFNLLNQLINLSTDLKTTNKKIWQLSSNDRNE
 LLDELLKVKEKAKIISISLKKNEIKKIILKDFGFEKSDIDDQDTIEGRKIIKEEPTTKLEVTKH
 LLATIYSHSSDSNWININNILEFLPYLDAICII LDREKSRGQDEV LKKLTEKNIFEVLKIDREK
 QLDFVKSIFSN TKFNFKKIGNFSLKAIREFLPKMF EQNKNSEY LKWKDEEIRRKWEEQKSKLGK
 TDKKTKYLNPRIFQDEIISP GTKNTFEQAVLV LNQIIKKYSKENIIDAI IIESPREKNDKKTIE
 40 EIKKR NKKGKGTLEKLFQILNLNENKGYKLS DLETKPAKLLDRLRFYHQQDGIDLYTLDKINID
 QLINGSQKYEIEHIIPYSMSYDNSQANKILTEKAENLKKGKLIASEYIKRNGDEFYNKYEEKAK
 ELFINKYKKNKKLDSYVDLDEDSAKNRFRFLTLDYDEFQVEFLARNLNDTRYSTKLFYHALVE
 HFENNEFFTYIDENSSKHVKIISTIKGHVTKYFRAKPVQKNNGPNENLNNNKPEKIEKNRENNE
 HHAVDAAIVAIIGNKNPQIANLLTLADNKT DKKFLLHDENYKENIETGELVKIPKFEVDKLAKV
 45 EDLKKIIQEKYEEAKKHTAIKFSRKTRTILNGLSDETLYGFKYDEKEDKYFKI IKKKLVT SKN
 EELKKYFENPF GKADGKSEYTVLMAQSHLSEFNKLKEIFEKYNGFSNKTGN AFVEYMNDLALK
 EPTLKAEIESAKSVEKLLY NFKPSDQFTYHDNINNKSFKR FYKNIRIIEYKSIPIKFKILSKH

DGGKSFKDTLFLSLVYKVYENGKESYKSIPVTSQMRNFGIDEFDFLDENLYNKEKLDIYKSD
 FAKPIPVNCKPVPFVLKKGSILKKKSLDIDDFKETKETEEGNYYFISTISKRFNRDTAYGLKPLK
 LSVVKPVAEPSTNPIFKEYIPIHLDELGNEYFPVKIKEHTDDEKLMCTIK

5

Nucleic Acids Encoding Cas9 Molecules

Nucleic acids encoding the Cas9 molecules or Cas9 polypeptides, e.g., an eaCas9 molecule or eaCas9 polypeptides are provided herein.

Exemplary nucleic acids encoding Cas9 molecules or Cas9 polypeptides are described in
 10 Cong *et al.*, SCIENCE 2013, 399(6121):819-823; Wang *et al.*, CELL 2013, 153(4):910-918; Mali *et al.*, SCIENCE 2013, 399(6121):823-826; Jinek *et al.*, SCIENCE 2012, 337(6096):816-821. Another exemplary nucleic acid encoding a Cas9 molecule or Cas9 polypeptide is shown in **Fig. 8**.

In an embodiment, a nucleic acid encoding a Cas9 molecule or Cas9 polypeptide can be a
 synthetic nucleic acid sequence. For example, the synthetic nucleic acid molecule can be
 15 chemically modified, e.g., as described in Section VIII. In an embodiment, the Cas9 mRNA has one or more (e.g., all of the following properties: it is capped, polyadenylated, substituted with 5-methylcytidine and/or pseudouridine).

In addition, or alternatively, the synthetic nucleic acid sequence can be codon optimized,
 e.g., at least one non-common codon or less-common codon has been replaced by a common
 20 codon. For example, the synthetic nucleic acid can direct the synthesis of an optimized messenger mRNA, e.g., optimized for expression in a mammalian expression system, e.g., described herein.

In addition, or alternatively, a nucleic acid encoding a Cas9 molecule or Cas9
 polypeptide may comprise a nuclear localization sequence (NLS). Nuclear localization
 25 sequences are known in the art.

Provided below is an exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. pyogenes*.

ATGGATAAAA AGTACAGCAT CGGGCTGGAC ATCGGTACAA ACTCAGTGGG
 GTGGGCCGTG ATTACGGACG AGTACAAGGT ACCCTCCAAA AAATTTAAAG
 30 TGCTGGGTAA CACGGACAGA CACTCTATAA AGAAAAATCT TATTGGAGCC
 TTGCTGTTTC ACTCAGGCGA GACAGCCGAA GCCACAAGGT TGAAGCGGAC
 CGCCAGGAGG CGGTATACCA GGAGAAAGAA CCGCATATGC TACCTGCAAG
 AAATCTTCAG TAACGAGATG GCAAAGGTTG ACGATAGCTT TTTCCATCGC
 CTGGAAGAAT CCTTTCTTGT TGAGGAAGAC AAGAAGCACG AACGGCACCC

	CATCTTTGGC	AATATTGTCG	ACGAAGTGGC	ATATCACGAA	AAGTACCCGA
	CTATCTACCA	CCTCAGGAAG	AAGCTGGTGG	ACTCTACCGA	TAAGGCGGAC
	CTCAGACTTA	TTTATTTGGC	ACTCGCCCAC	ATGATTAAAT	TTAGAGGACA
	TTTCTTGATC	GAGGGCGACC	TGAACCCGGA	CAACAGTGAC	GTCGATAAGC
5	TGTTTCATCCA	ACTTGTGCAG	ACCTACAATC	AACTGTTCTGA	AGAAAACCCCT
	ATAAATGCTT	CAGGAGTCGA	CGCTAAAGCA	ATCCTGTCCG	CGCGCCTCTC
	AAAATCTAGA	AGACTTGAGA	ATCTGATTGC	TCAGTTGCCC	GGGGAAAAGA
	AAAATGGATT	GTTTGGCAAC	CTGATCGCCC	TCAGTCTCGG	ACTGACCCCA
	AATTTCAAAA	GTAACCTCGA	CCTGGCCGAA	GACGCTAAGC	TCCAGCTGTC
10	CAAGGACACA	TACGATGACG	ACCTCGACAA	TCTGCTGGCC	CAGATTGGGG
	ATCAGTACGC	CGATCTCTTT	TTGGCAGCAA	AGAACCTGTC	CGACGCCATC
	CTGTTGAGCG	ATATCTTGAG	AGTGAACACC	GAAATTACTA	AAGCACCCCT
	TAGCGCATCT	ATGATCAAGC	GGTACGACGA	GCATCATCAG	GATCTGACCC
	TGCTGAAGGC	TCTTGTGAGG	CAACAGCTCC	CCGAAAAATA	CAAGGAAATC
15	TTCTTTGACC	AGAGCAAAAA	CGGCTACGCT	GGCTATATAG	ATGGTGGGGC
	CAGTCAGGAG	GAATTCTATA	AATTCATCAA	GCCCATTTCTC	GAGAAAATGG
	ACGGCACAGA	GGAGTTGCTG	GTCAAACCTTA	ACAGGGAGGA	CCTGCTGCGG
	AAGCAGCGGA	CCTTTGACAA	CGGGTCTATC	CCCCACCAGA	TTCATCTGGG
	CGAACTGCAC	GCAATCCTGA	GGAGGCAGGA	GGATTTTTAT	CCTTTTCTTA
20	AAGATAACCG	CGAGAAAATA	GAAAAGATTG	TTACATTTCAG	GATCCCGTAC
	TACGTGGGAC	CTCTCGCCCC	GGGCAATTCA	CGGTTTGCCT	GGATGACAAG
	GAAGTCAGAG	GAGACTATTA	CACCTTGGA	CTTCGAAGAA	GTGGTGGACA
	AGGGTGCATC	TGCCCAGTCT	TTCATCGAGC	GGATGACAAA	TTTTGACAAG
	AACCTCCCTA	ATGAGAAGGT	GCTGCCCAAA	CATTCTCTGC	TCTACGAGTA
25	CTTTACCGTC	TACAATGAAC	TGACTAAAGT	CAAGTACGTC	ACCGAGGGAA
	TGAGGAAGCC	GGCATTTCCT	AGTGGAGAAC	AGAAGAAGGC	GATTGTAGAC
	CTGTTGTTCA	AGACCAACAG	GAAGGTGACT	GTGAAGCAAC	TTAAAGAAGA
	CTACTTTAAG	AAGATCGAAT	GTTTTGACAG	TGTGGAAATT	TCAGGGGTTG
	AAGACCGCTT	CAATGCGTCA	TTGGGGACTT	ACCATGATCT	TCTCAAGATC
30	ATAAAGGACA	AAGACTTCCT	GGACAACGAA	GAAAATGAGG	ATATTCTCGA
	AGACATCGTC	CTCACCTTGA	CCCTGTTCGA	AGACAGGGAA	ATGATAGAAG
	AGCGCTTGAA	AACCTATGCC	CACCTCTTCG	ACGATAAAGT	TATGAAGCAG
	CTGAAGCGCA	GGAGATACAC	AGGATGGGGA	AGATTGTCAA	GGAAGCTGAT
	CAATGGAATT	AGGGATAAAC	AGAGTGGCAA	GACCATACTG	GATTTCTCTA
35	AATCTGATGG	CTTCGCCAAT	AGGAACTTCA	TGCAACTGAT	TCACGATGAC
	TCTCTTACCT	TCAAGGAGGA	CATTCAAAAG	GCTCAGGTGA	GCGGGCAGGG
	AGACTCCCTT	CATGAACACA	TCGCGAATTT	GGCAGGTTCC	CCCGCTATTA
	AAAAGGGCAT	CCTTCAAACCT	GTCAAGGTGG	TGGATGAATT	GGTCAAGGTA
	ATGGGCAGAC	ATAAGCCAGA	AAATATTGTG	ATCGAGATGG	CCCGCGAAAA
40	CCAGACCACA	CAGAAGGGCC	AGAAAAATAG	TAGAGAGCGG	ATGAAGAGGA
	TCGAGGAGGG	CATCAAAGAG	CTGGGATCTC	AGATTCTCAA	AGAACACCCC
	GTAGAAAACA	CACAGCTGCA	GAACGAAAAA	TTGTACTTGT	ACTATCTGCA
	GAACGGCAGA	GACATGTACG	TCGACCAAGA	ACTTGATATT	AATAGACTGT
	CCGACTATGA	CGTAGACCAT	ATCGTGCCCC	AGTCCTTCCT	GAAGGACGAC
45	TCCATTGATA	ACAAAGTCTT	GACAAGAAGC	GACAAGAACA	GGGGTAAAAG
	TGATAATGTG	CCTAGCGAGG	AGGTGGTGAA	AAAAATGAAG	AACTACTGGC
	GACAGCTGCT	TAATGCAAAG	CTCATTACAC	AACGGAAGTT	CGATAATCTG

```

ACGAAAGCAG AGAGAGGTGG CTTGTCTGAG TTGGACAAGG CAGGGTTTAT
TAAGCGGCAG CTGGTGGAAA CTAGGCAGAT CACAAAGCAC GTGGCGCAGA
TTTTGGACAG CCGGATGAAC ACAAATACG ACGAAATGA TAAACTGATA
CGAGAGGTCA AAGTTATCAC GCTGAAAAGC AAGCTGGTGT CCGATTTTCG
5 GAAAGACTTC CAGTTCTACA AAGTTCGCGA GATTAATAAC TACCATCATG
CTCACGATGC GTACCTGAAC GCTGTTGTCG GGACCGCCTT GATAAAGAAG
TACCCAAAGC TGGAATCCGA GTTCGTATAC GGGGATTACA AAGTGTACGA
TGTGAGGAAA ATGATAGCCA AGTCCGAGCA GGAGATTGGA AAGGCCACAG
CTAAGTACTT CTTTTATTCT AACATCATGA ATTTTTTTTAA GACGGAAATT
10 ACCCTGGCCA ACGGAGAGAT CAGAAAGCGG CCCCTTATAG AGACAAATGG
TGAAACAGGT GAAATCGTCT GGGATAAGGG CAGGGATTTC GCTACTGTGA
GGAAGGTGCT GAGTATGCCA CAGGTAAATA TCGTGAAAAA AACC GAAGTA
CAGACCGGAG GATTTTCCAA GGAAAGCATT TTGCCTAAAA GAAACTCAGA
CAAGCTCATC GCCCGCAAGA AAGATTGGGA CCCTAAGAAA TACGGGGGAT
15 TTGACTCACC CACCGTAGCC TATTCTGTGC TGGTGGTAGC TAAGGTGGAA
AAAGGAAAGT CTAAGAAGCT GAAGTCCGTG AAGGAACTCT TGGGAATCAC
TATCATGGAA AGATCATCCT TTGAAAAGAA CCCTATCGAT TTCCTGGAGG
CTAAGGGTTA CAAGGAGGTC AAGAAAGACC TCATCATTAA ACTGCCAAAA
TACTCTCTCT TCGAGCTGGA AAATGGCAGG AAGAGAATGT TGGCCAGCGC
20 CGGAGAGCTG CAAAAGGGAA ACGAGCTTGC TCTGCCCTCC AAATATGTTA
ATTTTCTCTA TCTCGCTTCC CACTATGAAA AGCTGAAAGG GTCTCCCGAA
GATAACGAGC AGAAGCAGCT GTTCGTGCGA CAGCACAAGC ACTATCTGGA
TGAAATAATC GAACAAATAA GCGAGTTCAG CAAAAGGGTT ATCCTGGCGG
ATGCTAATTT GGACAAAGTA CTGTCTGCTT ATAACAAGCA CCGGGATAAG
25 CCTATTAGGG AACAAGCCGA GAATATAATT CACCTCTTTA CACTCACGAA
TCTCGGAGCC CCCGCCGCCT TCAAATACTT TGATACGACT ATCGACCGGA
AACGGTATAC CAGTACCAAA GAGGTCCTCG ATGCCACCCT CATCCACCAG
TCAATTACTG GCCTGTACGA AACACGGATC GACCTCTCTC AACTGGGCGG
CGACTAG

```

30 (SEQ ID NO: 22)

Provided below is the corresponding amino acid sequence of a *S. pyogenes* Cas9 molecule.

```

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
35 HEKYPTIYHLRKKLVDSTDKADLRLLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTY
NQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF
DLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD
GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
40 PYYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKS
LLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTQKQKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTF
KEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQ
45 TTQKGQKNSRERMKRIIEGIELGSQLKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR
LSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK

```

FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIAS
 SEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
 MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVG
 5 KSKKLKSVKELLGITIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLAS
 AGELQKGNELALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRV
 ILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
 ATLIHQ SITGLYETRIDLSQLGGD*
 (SEQ ID NO: 23)

10

Provided below is an exemplary codon optimized nucleic acid sequence encoding a Cas9
 molecule of *N. meningitidis*.

ATGGCCGCCTTCAAGCCCAACCCCATCAACTACATCCTGGGCCTGGACATCGGCATCGCCAGCG
 TGGGCTGGGCCATGGTGGAGATCGACGAGGACGAGAACCCCATCTGCCTGATCGACCTGGGTGT
 15 GCGCGTGTTCGAGCGCGCTGAGGTGCCCAAGACTGGTGACAGTCTGGCTATGGCTCGCCGGCTT
 GCTCGCTCTGTTTCGGCGCCTTACTCGCCGGCGCGCTCACCGCCTTCTGCGCGCTCGCCGCCTGC
 TGAAGCGCGAGGGTGTGCTGCAGGCTGCCGACTTCGACGAGAACGGCCTGATCAAGAGCCTGCC
 CAACACTCCTTGGCAGCTGCGCGCTGCCGCTCTGGACCGCAAGCTGACTCCTCTGGAGTGGAGC
 GCCGTGCTGCTGCACCTGATCAAGCACCGCGGCTACCTGAGCCAGCGCAAGAACGAGGGCGAGA
 20 CCGCCGACAAGGAGCTGGGTGCTCTGCTGAAGGGCGTGGCCGACAACGCCACGCCCTGCAGAC
 TGGTGACTTCCGCACTCCTGCTGAGCTGGCCCTGAACAAGTTCGAGAAGGAGAGCGGCCACATC
 CGCAACCAGCGCGGCGACTACAGCCACACCTTCAGCCGCAAGGACCTGCAGGCCGAGCTGATCC
 TGCTGTTTCGAGAAGCAGAAGGAGTTCGGCAACCCCCACGTGAGCGGCGGCCCTGAAGGAGGGCAT
 CGAGACCCTGCTGATGACCCAGCGCCCCGCCCTGAGCGGCGACGCCGTGCAGAAGATGCTGGGC
 25 CACTGCACCTTCGAGCCAGCCGAGCCCAAGGCCGCCAAGAACACCTACACCGCCGAGCGCTTCA
 TCTGGCTGACCAAGCTGAACAACCTGCGCATCCTGGAGCAGGGCAGCGAGCGCCCCCTGACCGA
 CACCGAGCGCGCCACCCTGATGGACGAGCCCTACCGCAAGAGCAAGCTGACCTACGCCACGGCC
 CGCAAGCTGCTGGGTCTGGAGGACACCGCCTTCTTCAAGGGCCTGCGCTACGGCAAGGACAACG
 CCGAGGCCAGCACCTGATGGAGATGAAGGCCTACCACGCCATCAGCCGCGCCCTGGAGAAGGA
 30 GGGCCTGAAGGACAAGAAGAGTCCTCTGAACCTGAGCCCCGAGCTGCAGGACGAGATCGGCACC
 GCCTTCAGCCTGTTCAAGACCGACGAGGACATCACCGGCCCGCTGAAGGACCGCATCCAGCCCG
 AGATCCTGGAGGCCCTGCTGAAGCACATCAGCTTCGACAAGTTCGTGCAGATCAGCCTGAAGGC
 CCTGCGCCGCATCGTGCCCTGATGGAGCAGGGCAAGCGCTACGACGAGGCCTGCGCCGAGATC
 TACGGCGACCACTACGGCAAGAAGAACACCGAGGAGAAGATCTACCTGCCTCCTATCCCCGCCG
 35 ACGAGATCCGCAACCCCGTGGTGCTGCGCGCCCTGAGCCAGGCCCGCAAGGTGATCAACGGCGT
 GGTGCGCCGCTACGGCAGCCCCGCCGCTCCACATCGAGACCGCCCGCAGGTGGGCAAGAGC
 TTCAAGGACCGCAAGGAGATCGAGAAGCGCCAGGAGGAGAACCGCAAGGACCGCGAGAAGGCCG
 CCGCCAAGTTCCGCGAGTACTTCCCCAACTTCGTGGGCGAGCCCAAGAGCAAGGACATCCTGAA
 GCTGCGCCTGTACGAGCAGCAGCACGGCAAGTGCTGTACAGCGGCAAGGAGATCAACCTGGGC
 40 CGCCTGAACGAGAAGGGCTACGTGGAGATCGACCACGCCCTGCCCTTCAGCCGCACCTGGGACG
 ACAGCTTCAACAACAAGGTGCTGGTGCTGGGCAGCGAGAACCAGAACAAGGGCAACCAGACCCC
 CTACGAGTACTTCAACGGCAAGGACAACAGCCGCGAGTGGCAGGAGTTCAAGGCCCGCGTGAG
 ACCAGCCGCTTCCCCCGCAGCAAGAAGCAGCGCATCCTGCTGCAGAAGTTCGACGAGGACGGCT
 TCAAGGAGCGCAACCTGAACGACACCCGCTACGTGAACCGCTTCCTGTGCCAGTTTCGTGGCCGA
 45 CCGCATGCGCCTGACCGGCAAGGGCAAGAAGCGCGTGTTCCGCCAGCAACGGCCAGATCACCAAC

CTGCTGCGCGGCTTCTGGGGCCTGCGCAAGGTGCGCGCCGAGAACGACCGCCACCACGCCCTGG
 ACGCCGTGGTGGTGGCCTGCAGCACCGTGGCCATGCAGCAGAAGATCACCCGCTTCGTGCGCTA
 CAAGGAGATGAACGCCTTCGACGGTAAAACCATCGACAAGGAGACCGGCGAGGTGCTGCACCAG
 AAGACCCACTTCCCCCAGCCCTGGGAGTTCTTCGCCCAGGAGGTGATGATCCGCGTGTTTCGGCA
 5 AGCCCGACGGCAAGCCCGAGTTCGAGGAGGCCGACACCCCCGAGAAGCTGCGCACCCCTGCTGGC
 CGAGAAGCTGAGCAGCCGCCCTGAGGCCGTGCACGAGTACGTGACTCCTCTGTTCGTGAGCCGC
 GCCCCCAACCGCAAGATGAGCGGTGAGGGTCACATGGAGACCGTGAAGAGCGCCAAGCGCCTGG
 ACGAGGGCGTGAGCGTGCTGCGCGTGCCCTGACCCAGCTGAAGCTGAAGGACCTGGAGAAGAT
 GGTGAACCGCGAGCGCGAGCCCAAGCTGTACGAGGCCCTGAAGGCCCGCCTGGAGGCCCAAG
 10 GACGACCCCGCCAAGGCCCTTCGCCCAGGCCCTTCTACAAGTACGACAAGGCCGGAACCGCACCC
 AGCAGGTGAAGGCCGTGCGCGTGGAGCAGGTGCAGAAGACCGGCGTGTGGGTGCGCAACCACAA
 CGGCATCGCCGACAACGCCACCATGGTGCGCGTGGACGTGTTTCGAGAAGGGCGACAAGTACTAC
 CTGGTGCCCATCTACAGCTGGCAGGTGGCCAAGGGCATCCTGCCCCGACCGCGCCGTGGTGAGG
 GCAAGGACGAGGAGGACTGGCAGCTGATCGACGACAGCTTCAACTTCAAGTTCAGCCTGCACCC
 15 CAACGACCTGGTGGAGGTGATCACCAAGAAGGCCCGCATGTTTCGGCTACTTCGCCAGCTGCCAC
 CGCGGCACCGGCAACATCAACATCCGCATCCACGACCTGGACCACAAGATCGGCAAGAACGGCA
 TCCTGGAGGGCATCGGCGTGAAGACCGCCCTGAGCTTCCAGAAGTACCAGATCGACGAGCTGGG
 CAAGGAGATCCGCCCCTGCCGCCTGAAGAAGCGCCCTCCTGTGCGCTAA
 (SEQ ID NO: 24)

20 Provided below is the corresponding amino acid sequence of a *N. meningitidis* Cas9 molecule.

MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRFERAEVPKTGDSLAMARRL
 ARSVRRLTRRRHRLRLRARRLLKREGVLQAADFENGLIKSLPNTPWQLRAAALDRKLTPLEWS
 AVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNAHALQTGDFRTPAELALNKFESGHI
 25 RNQRGDYSHTFSRKDLQAEILLLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLG
 HCTFEPAPKAAKNTYTAERFIWLTKLNNLRILEQGSRPLTDTERATLMDEPYRKSCLTYAQA
 RKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRLEKEGLKDKKSPLNLSPELQDEIGT
 AFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKVFQISLKALRRIVPLMEQGKRYDEACAEI
 YGDHYGKKNTTEEKIYLPIPADEIRNPVVLRLALSQARKVINGVVRRYGSPARIHIETAREVGKS
 30 FKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPSKDIKLRLYEQQHKGKCLYSKEINLG
 RLNEKGVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVE
 TSRFRPSKKQRILLQKFDEDFGFKERNLNDTRYVNRFLCQFVADRMRLTGKGGKRVFASNGQITN
 LLRGFWGLRKVRAENDRRHALDAVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQ
 KTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRLLAEKLSSRPEAVHEYVTPLFVSR
 35 APNRKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLKDLKMNVREREPKLYEALKARLEAHK
 DDPKAFAPFYPKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYY
 LVPIYSWQVAKGILPDRAVVQKDEEDWQLIDDSFNFKFSLHPNDLVEVITTKARMEGYFASCH
 RGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFKYQIDELGKEIRPCRLKKRPPVR*
 (SEQ ID NO: 25)

40 Provided below is an amino acid sequence of a *S. aureus* Cas9 molecule.

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGRARRLKRRRRHRI
 QRVKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDT
 GNELSTKEQISRNSKALEEKYVAELQLERLKKDGVRGSINRFKTSYVKEAKQLLKVQKAYHQ
 LDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLY
 45 NALNDLNNLVITRDENEKLEYEYEFQI IENVFKQKKKPTLQKIAKEILVNEEDIKGYRVTSTGK

PEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQIS
 NLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSP
 VVKRSFIQSIKVINAI IKKYGLPNDII IELAREKNSKDAQKMINEMQKRNRQTNERIEE IIRTT
 GKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVK
 5 QEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLA KGKGRISKTKEYLLEERDINRFSVQKD
 FINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAD
 ALI IANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKD
 YKYSHRVDKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHH
 DPQTYQKLKLIMEQYGDEKNPLYKYEEETGNLYTKYSKKDNGPVIKKIKYYGNKLNALHLDITDD
 10 YPNSRNKVVKLSLKPYPYRFDVYLDNGVYKFVTVKNLDVIKKENYEVNSKCYEEAKKLKKISNQA
 EFIA SFYNNDLIKINGELYRVIGVNNDDLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKT
 QSIKKYSTDILGNLYEVKSKKHPQIIKKG*
 (SEQ ID NO: 26)

15 Provided below is an exemplary codon optimized nucleic acid sequence encoding a Cas9
 molecule of *S. aureus* Cas9.

ATGAAAAGGAACTACATTCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTG
 ACTATGAAACAAGGGACGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAA
 CAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGAAACGACGGAGAAGGCACAGAATC
 20 CAGAGGGTGAAGAACTGCTGTTTCGATTACAACCTGCTGACCGACCATTCTGAGCTGAGTGGAA
 TTAATCCTTATGAAGCCAGGGTGAAAGGCCTGAGTCAGAAGCTGTCAGAGGAAGAGTTTTCCGC
 AGCTCTGCTGCACCTGGCTAAGCGCCGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACC
 GGCAACGAGCTGTCTACAAAGGAACAGATCTCACGCAATAGCAAAGCTCTGGAAGAGAAGTATG
 TCGCAGAGCTGCAGCTGGAACGGCTGAAGAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTT
 25 CAAGACAAGCGACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAGTGCAGAAGGCTTACCACCAG
 CTGGATCAGAGCTTCATCGATACTTATATCGACCTGCTGGAGACTCGGAGAACCTACTATGAGG
 GACCAGGAGAAGGGAGCCCCCTTCGGATGGAAGACATCAAGGAATGGTACGAGATGCTGATGGG
 ACATTGCACCTATTTTCCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTGTAC
 AACGCCCTGAATGACCTGAACAACCTGGTCATCACCAGGGATGAAAACGAGAACTGGAATACT
 30 ATGAGAAGTTCCAGATCATCGAAAACGTGTTTAAGCAGAAGAAAAAGCCTACACTGAAACAGAT
 TGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAAGGGCTACCGGGTGACAAGCACTGGAAAA
 CCAGAGTTACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAAGAAATCA
 TTGAGAACGCCGAACCTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGA
 CATCCAGGAAGAGCTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGT
 35 AATCTGAAGGGGTACACCGGAACACACAACCTGTCCCTGAAAGCTATCAATCTGATTCTGGATG
 AGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGCTGGTCCCCAAAAA
 GGTGGACCTGAGTCAGCAGAAAGAGATCCCAACCACACTGGTGGACGATTTTATTCTGTCACCC
 GTGGTCAAGCGGAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCC
 TGCCCAATGATATCATTATCGAGCTGGCTAGGGAGAAGAACAGCAAGGACGCACAGAAGATGAT
 40 CAATGAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTGAAGAGATTATCCGAACCTACC
 GGGAAAGAGAACGCAAAGTACCTGATTGAAAAAATCAAGCTGCACGATATGCAGGAGGGGAAAGT
 GTCTGTATTCTCTGGAGGCCATCCCCCTGGAGGACCTGCTGAACAATCCATTCAACTACGAGGT
 CGATCATATTATCCCCAGAAGCGTGTCTTTCGACAATTCCTTTAACAACAAGGTGCTGGTCAAG
 CAGGAAGAGA ACTCTAAAAAGGGCAATAGGACTCCTTTCCAGTACCTGTCTAGTTTCAATTCCA
 45 AGATCTCTTACGAAACCTTTAAAAAGCACATTCTGAATCTGGCCAAAGGAAAGGGCCGCATCAG

CAAGACCAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGAT
 TTTATTAACCGGAATCTGGTGGACACAAGATACGCTACTCGCGGCCTGATGAATCTGCTGCGAT
 CCTATTTCCGGGTGAACAATCTGGATGTGAAAGTCAAGTCCATCAACGGCGGGTTCACATCTTT
 TCTGAGGCGCAAATGGAAGTTTAAAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGAAGAT
 5 GCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAGCCAAGA
 AAGTGATGGAGAACCAGATGTTTGAAGAGAAGCAGGCCGAATCTATGCCCCGAAATCGAGACAGA
 ACAGGAGTACAAGGAGATTTTCATCACTCCTCACCAGATCAAGCATATCAAGGATTTCAAGGAC
 TACAAGTACTCTCACCGGGTGGATAAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTATA
 GTACAAGAAAAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAA
 10 AGATAATGACAAGCTGAAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCAT
 GATCCTCAGACATATCAGAACTGAAGCTGATTATGGAGCAGTACGGCGACGAGAAGAACCAC
 TGTATAAGTACTATGAAGAGACTGGGAACCTACCTGACCAAGTATAGCAAAAAGGATAATGGCCC
 CGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGCCCATCTGGACATCACAGACGAT
 TACCCTAACAGTCGCAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGATTTCGATGTCTATC
 15 TGGACAACGGCGTGTATAAATTTGTGACTGTCAAGAATCTGGATGTCATCAAAAAGGAGAACTA
 CTATGAAGTGAATAGCAAGTGCTACGAAGAGGCTAAAAAGCTGAAAAAGATTAGCAACCAGGCA
 GAGTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAACTGTATAGGGTCA
 TCGGGGTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTGACATCACTTACCGAGA
 GTATCTGGAAAACATGAATGATAAGCGCCCCCTCGAATTATCAAAAACAAATTGCCTCTAAGACT
 20 CAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAACCTGTATGAGGTGAAGAGCAAAAAGC
 ACCCTCAGATTATCAAAAAGGGC
 (SEQ ID NO: 39)

If any of the above Cas9 sequences are fused with a peptide or polypeptide at the C-
 terminus, it is understood that the stop codon will be removed.

Other Cas Molecules and Cas Polypeptides

Various types of Cas molecules or Cas polypeptides can be used to practice the
 inventions disclosed herein. In some embodiments, Cas molecules of Type II Cas systems are
 30 used. In other embodiments, Cas molecules of other Cas systems are used. For example, Type I
 or Type III Cas molecules may be used. Exemplary Cas molecules (and Cas systems) are
 described, e.g., in Haft *et al.*, PLOS COMPUTATIONAL BIOLOGY 2005, 1(6): e60 and Makarova *et*
al., NATURE REVIEW MICROBIOLOGY 2011, 9:467-477, the contents of both references are
 incorporated herein by reference in their entirety. Exemplary Cas molecules (and Cas systems)
 35 are also shown in **Table 12**.

Table 12: Cas Systems

Gene name [‡]	System type or subtype	Name from Haft <i>et al.</i> [§]	Structure of encoded protein (PDB accessions) [¶]	Families (and superfamily) of encoded protein ^{***}	Representatives
<i>cas1</i>	• Type I • Type II • Type III	<i>cas1</i>	3GOD, 3LFX and 2YZS	COG1518	SERP2463, SPy1047 and <i>ygbT</i>
<i>cas2</i>	• Type I • Type II • Type III	<i>cas2</i>	2IVY, 2I8E and 3EXC	COG1343 and COG3512	SERP2462, SPy1048, SPy1723 (N-terminal domain) and <i>ygbF</i>
<i>cas3'</i>	• Type I ^{‡‡}	<i>cas3</i>	NA	COG1203	APE1232 and <i>ygcB</i>
<i>cas3''</i>	• Subtype I-A • Subtype I-B	NA	NA	COG2254	APE1231 and BH0336
<i>cas4</i>	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-D • Subtype II-B	<i>cas4</i> and <i>csa1</i>	NA	COG1468	APE1239 and BH0340
<i>cas5</i>	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-E	<i>cas5a</i> , <i>cas5d</i> , <i>cas5e</i> , <i>cas5h</i> , <i>cas5p</i> , <i>cas5t</i> and <i>cmx5</i>	3KG4	COG1688 (RAMP)	APE1234, BH0337, <i>devS</i> and <i>ygcI</i>
<i>cas6</i>	• Subtype I-A • Subtype I-B • Subtype I-D • Subtype III-A • Subtype III-B	<i>cas6</i> and <i>cmx6</i>	3I4H	COG1583 and COG5551 (RAMP)	PF1131 and slr7014
<i>cas6e</i>	• Subtype I-E	<i>cse3</i>	1WJ9	(RAMP)	<i>ygcH</i>
<i>cas6f</i>	• Subtype I-F	<i>csy4</i>	2XLJ	(RAMP)	<i>y1727</i>
<i>cas7</i>	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-E	<i>csa2</i> , <i>csd2</i> , <i>cse4</i> , <i>csh2</i> , <i>csp1</i> and <i>cst2</i>	NA	COG1857 and COG3649 (RAMP)	<i>devR</i> and <i>ygcJ</i>
<i>cas8a1</i>	• Subtype I-A ^{‡‡}	<i>cmx1</i> , <i>cst1</i> , <i>csx8</i> , <i>csx13</i> and CXXC-CXXC	NA	BH0338-like	LA3191 ^{§§} and PG2018 ^{§§}
<i>cas8a2</i>	• Subtype I-A ^{‡‡}	<i>csa4</i> and <i>csx9</i>	NA	PH0918	AF0070, AF1873, MJ0385, PF0637, PH0918 and SSO1401
<i>cas8b</i>	• Subtype I-B ^{‡‡}	<i>csh1</i> and TM1802	NA	BH0338-like	MTH1090 and TM1802
<i>cas8c</i>	• Subtype I-C ^{‡‡}	<i>csd1</i> and <i>csp2</i>	NA	BH0338-like	BH0338
<i>cas9</i>	• Type II ^{‡‡}	<i>csn1</i> and <i>csx12</i>	NA	COG3513	FTN_0757 and SPy1046

Table 12: Cas Systems

Gene name [‡]	System type or subtype	Name from Haft <i>et al.</i> [§]	Structure of encoded protein (PDB accessions) [¶]	Families (and superfamily) of encoded protein ^{***}	Representatives
<i>cas10</i>	• Type III ^{‡‡}	<i>cmr2</i> , <i>csm1</i> and <i>csx11</i>	NA	COG1353	MTH326, Rv2823c ^{§§} and TM1794 ^{§§}
<i>cas10d</i>	• Subtype I-D ^{‡‡}	<i>csc3</i>	NA	COG1353	slr7011
<i>csy1</i>	• Subtype I-P ^{‡‡}	<i>csy1</i>	NA	y1724-like	y1724
<i>csy2</i>	• Subtype I-F	<i>csy2</i>	NA	(RAMP)	y1725
<i>csy3</i>	• Subtype I-F	<i>csy3</i>	NA	(RAMP)	y1726
<i>cse1</i>	• Subtype I-E ^{‡‡}	<i>cse1</i>	NA	YgcL-like	<i>ygcL</i>
<i>cse2</i>	• Subtype I-E	<i>cse2</i>	2ZCA	YgcK-like	<i>ygcK</i>
<i>csc1</i>	• Subtype I-D	<i>csc1</i>	NA	alr1563-like (RAMP)	alr1563
<i>csc2</i>	• Subtype I-D	<i>csc1</i> and <i>csc2</i>	NA	COG1337 (RAMP)	slr7012
<i>csa5</i>	• Subtype I-A	<i>csa5</i>	NA	AF1870	AF1870, MJ0380, PF0643 and SSO1398
<i>csn2</i>	• Subtype II-A	<i>csn2</i>	NA	SPy1049-like	SPy1049
<i>csm2</i>	• Subtype III-A ^{‡‡}	<i>csm2</i>	NA	COG1421	MTH1081 and SERP2460
<i>csm3</i>	• Subtype III-A	<i>csc2</i> and <i>csm3</i>	NA	COG1337 (RAMP)	MTH1080 and SERP2459
<i>csm4</i>	• Subtype III-A	<i>csm4</i>	NA	COG1567 (RAMP)	MTH1079 and SERP2458
<i>csm5</i>	• Subtype III-A	<i>csm5</i>	NA	COG1332 (RAMP)	MTH1078 and SERP2457
<i>csm6</i>	• Subtype III-A	APE2256 and <i>csm6</i>	2WTE	COG1517	APE2256 and SSO1445
<i>cmr1</i>	• Subtype III-B	<i>cmr1</i>	NA	COG1367 (RAMP)	PF1130
<i>cmr3</i>	• Subtype III-B	<i>cmr3</i>	NA	COG1769 (RAMP)	PF1128
<i>cmr4</i>	• Subtype III-B	<i>cmr4</i>	NA	COG1336 (RAMP)	PF1126
<i>cmr5</i>	• Subtype III-B ^{‡‡}	<i>cmr5</i>	2ZOP and 2OEB	COG3337	MTH324 and PF1125
<i>cmr6</i>	• Subtype III-B	<i>cmr6</i>	NA	COG1604 (RAMP)	PF1124

Table 12: Cas Systems

Gene name [‡]	System type or subtype	Name from Haft <i>et al.</i> [§]	Structure of encoded protein (PDB accessions) [¶]	Families (and superfamily) of encoded protein ^{***}	Representatives
<i>csb1</i>	• Subtype I-U	GSU0053	NA	(RAMP)	Balac_1306 and GSU0053
<i>csb2</i>	• Subtype I-U ^{§§}	NA	NA	(RAMP)	Balac_1305 and GSU0054
<i>csb3</i>	• Subtype I-U	NA	NA	(RAMP)	Balac_1303 ^{§§}
<i>csx17</i>	• Subtype I-U	NA	NA	NA	Btus_2683
<i>csx14</i>	• Subtype I-U	NA	NA	NA	GSU0052
<i>csx10</i>	• Subtype I-U	<i>csx10</i>	NA	(RAMP)	Caur_2274
<i>csx16</i>	• Subtype III-U	VVA1548	NA	NA	VVA1548
<i>csaX</i>	• Subtype III-U	<i>csaX</i>	NA	NA	SSO1438
<i>csx3</i>	• Subtype III-U	<i>csx3</i>	NA	NA	AF1864
<i>csx1</i>	• Subtype III-U	<i>csa3</i> , <i>csx1</i> , <i>csx2</i> , DXTHG, NE0113 and TIGR02710	1XMX and 2I71	COG1517 and COG4006	MJ1666, NE0113, PF1127 and TM1812
<i>csx15</i>	• Unknown	NA	NA	TTE2665	TTE2665
<i>csf1</i>	• Type U	<i>csf1</i>	NA	NA	AFE_1038
<i>csf2</i>	• Type U	<i>csf2</i>	NA	(RAMP)	AFE_1039
<i>csf3</i>	• Type U	<i>csf3</i>	NA	(RAMP)	AFE_1040
<i>csf4</i>	• Type U	<i>csf4</i>	NA	NA	AFE_1037

IV. Functional Analysis of Candidate Molecules

Candidate Cas9 molecules, candidate gRNA molecules, candidate Cas9 molecule/gRNA molecule complexes, can be evaluated by art-known methods or as described herein. For example, exemplary methods for evaluating the endonuclease activity of Cas9 molecule are described, e.g., in Jinek *et al.*, SCIENCE 2012, 337(6096):816-821.

Binding and Cleavage Assay: Testing the endonuclease activity of Cas9 molecule

The ability of a Cas9 molecule/gRNA molecule complex to bind to and cleave a target nucleic acid can be evaluated in a plasmid cleavage assay. In this assay, synthetic or *in vitro*-transcribed gRNA molecule is pre-annealed prior to the reaction by heating to 95°C and slowly cooling down to room temperature. Native or restriction digest-linearized plasmid DNA (300 ng

(~8 nM)) is incubated for 60 min at 37°C with purified Cas9 protein molecule (50-500 nM) and gRNA (50-500 nM, 1:1) in a Cas9 plasmid cleavage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA) with or without 10 mM MgCl₂. The reactions are stopped with 5X DNA loading buffer (30% glycerol, 1.2% SDS, 250 mM EDTA), resolved by a 0.8 or 1% agarose gel electrophoresis and visualized by ethidium bromide staining. The resulting cleavage products indicate whether the Cas9 molecule cleaves both DNA strands, or only one of the two strands. For example, linear DNA products indicate the cleavage of both DNA strands. Nicked open circular products indicate that only one of the two strands is cleaved.

Alternatively, the ability of a Cas9 molecule/gRNA molecule complex to bind to and cleave a target nucleic acid can be evaluated in an oligonucleotide DNA cleavage assay. In this assay, DNA oligonucleotides (10 pmol) are radiolabeled by incubating with 5 units T4 polynucleotide kinase and ~3–6 pmol (~20–40 mCi) [γ -³²P]-ATP in 1X T4 polynucleotide kinase reaction buffer at 37°C for 30 min, in a 50 μ L reaction. After heat inactivation (65°C for 20 min), reactions are purified through a column to remove unincorporated label. Duplex substrates (100 nM) are generated by annealing labeled oligonucleotides with equimolar amounts of unlabeled complementary oligonucleotide at 95°C for 3 min, followed by slow cooling to room temperature. For cleavage assays, gRNA molecules are annealed by heating to 95°C for 30 s, followed by slow cooling to room temperature. Cas9 (500 nM final concentration) is pre-incubated with the annealed gRNA molecules (500 nM) in cleavage assay buffer (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, 5% glycerol) in a total volume of 9 μ L. Reactions are initiated by the addition of 1 μ L target DNA (10 nM) and incubated for 1 h at 37°C. Reactions are quenched by the addition of 20 μ L of loading dye (5 mM EDTA, 0.025% SDS, 5% glycerol in formamide) and heated to 95°C for 5 min. Cleavage products are resolved on 12% denaturing polyacrylamide gels containing 7 M urea and visualized by phosphorimaging. The resulting cleavage products indicate that whether the complementary strand, the non-complementary strand, or both, are cleaved.

One or both of these assays can be used to evaluate the suitability of a candidate gRNA molecule or candidate Cas9 molecule.

Binding Assay: Testing the binding of Cas9 molecule to target DNA

Exemplary methods for evaluating the binding of Cas9 molecule to target DNA are described, e.g., in Jinek *et al.*, SCIENCE 2012; 337(6096):816-821.

For example, in an electrophoretic mobility shift assay, target DNA duplexes are formed by mixing of each strand (10 nmol) in deionized water, heating to 95°C for 3 min and slow
5 cooling to room temperature. All DNAs are purified on 8% native gels containing 1X TBE. DNA bands are visualized by UV shadowing, excised, and eluted by soaking gel pieces in DEPC-treated H₂O. Eluted DNA is ethanol precipitated and dissolved in DEPC-treated H₂O. DNA samples are 5' end labeled with [γ -³²P]-ATP using T4 polynucleotide kinase for 30 min at 37°C. Polynucleotide kinase is heat denatured at 65°C for 20 min, and unincorporated radiolabel
10 is removed using a column. Binding assays are performed in buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT and 10% glycerol in a total volume of 10 μ l. Cas9 protein molecule is programmed with equimolar amounts of pre-annealed gRNA molecule and titrated from 100 pM to 1 μ M. Radiolabeled DNA is added to a final concentration of 20 pM. Samples are incubated for 1 h at 37°C and resolved at 4°C on an 8% native polyacrylamide
15 gel containing 1X TBE and 5 mM MgCl₂. Gels are dried and DNA visualized by phosphorimaging.

Differential Scanning Fluorimetry (DSF)

The thermostability of Cas9-gRNA ribonucleoprotein (RNP) complexes can be measured
20 via DSF. This technique measures the thermostability of a protein, which can increase under favorable conditions such as the addition of a binding RNA molecule, e.g., a gRNA.

The assay is performed using two different protocols, one to test the best stoichiometric ratio of gRNA:Cas9 protein and another to determine the best solution conditions for RNP formation.

25 To determine the best solution to form RNP complexes, a 2 μ M solution of Cas9 in water+10x SYPRO Orange® (Life Technologies cat#S-6650) and dispensed into a 384 well plate. An equimolar amount of gRNA diluted in solutions with varied pH and salt is then added. After incubating at room temperature for 10' and brief centrifugation to remove any bubbles, a Bio-Rad CFX384™ Real-Time System C1000 Touch™ Thermal Cycler with the Bio-Rad CFX
30 Manager software is used to run a gradient from 20°C to 90°C with a 1° increase in temperature every 10seconds.

The second assay consists of mixing various concentrations of gRNA with 2uM Cas9 in optimal buffer from assay 1 above and incubating at RT for 10' in a 384 well plate. An equal volume of optimal buffer + 10x SYPRO Orange® (Life Techonologies cat#S-6650) is added and the plate sealed with Microseal® B adhesive (MSB-1001). Following brief centrifugation to
5 remove any bubbles, a Bio-Rad CFX384™ Real-Time System C1000 Touch™ Thermal Cycler with the Bio-Rad CFX Manager software is used to run a gradient from 20°C to 90°C with a 1° increase in temperature every 10seconds.

V. Genome Editing Approaches

10 Mutations in the *USH2A* gene may be corrected using one of the approaches discussed herein. In an embodiment, a mutation in the *USH2A* gene is corrected by homology directed repair (HDR) using an exogenously provided template nucleic acid (see Section V.1).

V.1 HDR Repair and Template Nucleic Acids

15 The donor template or template nucleic acid provides for alteration of the target sequence. While not wishing to be bound by theory, it is believed that alteration of the target sequence occurs by homology-directed repair (HDR) with the donor template. While not wishing to be bound by theory, it is believed that plasmid donors serve as templates for homologous recombination and it is believed that single stranded donor templates provide for alteration of the
20 target sequence potentially by alternate methods of homology directed repair (e.g., single strand annealing) between the target sequence and the donor template. Donor template-effected alteration of a target sequence depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a double strand break or two single strand breaks.

Double strand break mediated correction

25 In an embodiment, double strand cleavage is effected by a Cas9 molecule having cleavage activity associated with an HNH-like domain and cleavage activity associated with anRuvC-like domain, e.g., an N-terminal RuvC-like domain, e.g., a wildtype Cas9. Such embodiments require only a single gRNA.

Single strand break mediated correction

30 In other embodiments, two single strand breaks, or nicks, are effected by a Cas9 molecule having nickase activity, e.g., cleavage activity associated with an HNH-like domain or cleavage

activity associated with an N-terminal RuvC-like domain. Such embodiments require two gRNAs, one for placement of each single strand break. In an embodiment, the Cas9 molecule having nickase activity cleaves the strand to which the gRNA hybridizes but not the strand that is complementary to the strand to which the gRNA hybridizes. In an embodiment, the Cas9 molecule having nickase activity does not cleave the strand to which the gRNA hybridizes but rather cleaves the strand that is complementary to the strand to which the gRNA hybridizes.

In an embodiment, the nickase has HNH activity, e.g., a Cas9 molecule having the RuvC activity inactivated, e.g., a Cas9 molecule having a mutation at D10, e.g., the D10A mutation. D10A inactivates RuvC therefore the Cas9 nickase has (only) HNH activity and will cut on the strand to which the gRNA hybridizes (the complementary strand, which does not have the NGG PAM on it). In other embodiments, a Cas9 molecule having an H840, e.g., an H840A, mutation can be used as a nickase. H840A inactivates HNH therefore the Cas9 nickase has (only) RuvC activity and cuts on the non-complementary strand (the strand that has the NGG PAM and whose sequence is identical to the gRNA). In other embodiments, a Cas9 molecule having an H863, e.g., an H863A, mutation can be used as a nickase. H863A inactivates HNH therefore the Cas9 nickase has (only) RuvC activity and cuts on the non-complementary strand (the strand that has the NGG PAM and whose sequence is identical to the gRNA).

In an embodiment, in which a nickase and two gRNAs are used to position two single strand nicks, one nick is on the + strand and one nick is on the – strand of the target nucleic acid. The PAMs can be outwardly facing. The gRNAs can be selected such that the gRNAs are separated by, from 0-50, 0-100, or 0-200 nucleotides. In an embodiment, there is no overlap between the target sequences that are complementary to the targeting domains of the two gRNAs. In an embodiment, the gRNAs do not overlap and are separated by as much as 50, 100, or 200 nucleotides. In an embodiment, the use of two gRNAs can increase specificity, e.g., by decreasing off-target binding (Ran *et al.*, Cell 2013; 154(6):1380-1389).

In an embodiment, a single nick can be used to induce HDR. In an embodiment, using a single nick to induce HDR is less efficient and has a lower on-target activity than is seen with a double nickase approach.

Placement of double strand or single strand breaks relative to the target position

The double strand break or single strand break in one of the strands should be sufficiently close to the target sequence or signature such that correction occurs. In an embodiment, the

distance is not more than 50, 100, 200, 300, 350 or 400 nucleotides. While not wishing to be bound by theory, it is believed that the break should be sufficiently close to the target sequence such that the break is within the region that is subject to exonuclease-mediated removal during end resection. If the distance between the target sequence and a break is too great, the mutation may not be included in the end resection and, therefore, may not be corrected, as donor sequence may only be used to correct sequence within the end resection region.

In an embodiment, a gRNA, e.g., a unimolecular (or chimeric) or modular gRNA molecule, is configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site is between 0-40 bp away from the target position (e.g., less than 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position).

In an embodiment, two gRNAs, e.g., independently, unimolecular (or chimeric) or modular gRNA, are configured to position two single-strand breaks. In an embodiment, the gRNAs are configured to position cuts at the same position, or within a few nucleotides of one another, on different strands, essentially mimicking a double strand break. In an embodiment, the two nicks are between 0-40 bp away from the target position (e.g., less than 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position) respectively, and the two single strand breaks are within 25-55 bp of each other (e.g., between 25 to 50, 25 to 45, 25 to 40, 25 to 35, 25 to 30, 50 to 55, 45 to 55, 40 to 55, 35 to 55, 30 to 55, 30 to 50, 35 to 50, 40 to 50, 45 to 50, 35 to 45, or 40 to 45 bp) and no more than 100 bp away from each other (e.g., no more than 90, 80, 70, 60, 50, 40, 30, 20 or 10 bp). In an embodiment, the gRNAs are configured to place a single strand break on either side of the target position. In an embodiment, the gRNAs are configured to place a single strand break on the same side (either 5' or 3') of the target position.

Regardless of whether a break is a double strand or a single strand break, the gRNA should be configured to avoid unwanted target chromosome elements, such as repeated elements, e.g., an *Alu* repeat, in the target domain. In addition, a break, whether a double strand or a single strand break, should be sufficiently distant from any sequence that should not be altered. For example, cleavage sites positioned within introns should be sufficiently distant from any intron/exon border, or naturally occurring splice signal, to avoid alteration of the exonic sequence or unwanted splicing events.

Length of the homology arms

The homology arm should extend at least as far as the region in which end resection may occur, e.g., in order to allow the resected single stranded overhang to find a complementary region within the donor template. The overall length could be limited by parameters such as plasmid size or viral packaging limits. In an embodiment, a homology arm does not extend into repeated elements, e.g., Alu repeats.

Exemplary homology arm lengths include a least 50, 100, 250, 500, 750 or 1000 nucleotides.

Target position, as used herein, refers to a site on a target nucleic acid (e.g., the chromosome) that is modified by a Cas9 molecule-dependent process. For example, the target position can be a modified Cas9 molecule cleavage of the target nucleic acid and template nucleic acid directed modification, e.g., correction, of the target position. In an embodiment, a target position can be a site between two nucleotides, e.g., adjacent nucleotides, on the target nucleic acid into which one or more nucleotides is added. The target position may comprise one or more nucleotides that are altered, e.g., corrected, by a template nucleic acid. In an embodiment, the target position is within a target sequence (e.g., the sequence to which the gRNA binds). In an embodiment, a target position is upstream or downstream of a target sequence (e.g., the sequence to which the gRNA binds).

A template nucleic acid, as that term is used herein, refers to a nucleic acid sequence which can be used in conjunction with a Cas9 molecule and a gRNA molecule to alter the structure of a target position. In an embodiment, the target nucleic acid is modified to have the some or all of the sequence of the template nucleic acid, typically at or near cleavage site(s). Target position, as used herein, refers to a nucleotide or nucleotides that are altered by the template nucleic acid, e.g., by altering, e.g., by recombination, e.g., homologous recombination or by homology directed repair. In an embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In an embodiment, the template nucleic acid is DNA, e.g., double stranded DNA. In an alternate embodiment, the template nucleic acid is single stranded DNA. In an embodiment, the template nucleic acid is encoded on the same vector backbone, e.g. AAV genome, plasmid DNA, as the Cas9 and gRNA. In an embodiment, the template nucleic acid is excised from this backbone *in vivo*, e.g. is flanked by gRNA recognition sequences.

In an embodiment, the template nucleic acid alters the structure of the target position by participating in a homology directed repair event. In an embodiment, the template nucleic acid alters the sequence of the target position. In an embodiment, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

5 Typically, the template sequence undergoes a breakage mediated or catalyzed recombination with the target sequence. In an embodiment, the template nucleic acid includes sequence that corresponds to a site on the target sequence that is cleaved by an eaCas9 mediated cleavage event. In an embodiment, the template nucleic acid includes sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Cas9 mediated event, and a
10 second site on the target sequence that is cleaved in a second Cas9 mediated event.

In an embodiment, the template nucleic acid can include sequence which results in an alteration in the coding sequence of a translated sequence, e.g., one which results in the substitution of one amino acid for another in a protein product, e.g., transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introducing a
15 stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation.

In other embodiments, the template nucleic acid can include sequence which results in an alteration in a non-coding sequence, e.g., an alteration in an exon or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control element, e.g., a
20 promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

A template nucleic acid having homology with a target position in the *USH2A* gene from can be used to alter the structure of a target sequence. The template sequence can be used to alter an unwanted structure, e.g., an unwanted or mutant nucleotide.

A template nucleic acid comprises the following components:
25 [5' homology arm]-[replacement sequence]-[3' homology arm].

The homology arms provide for recombination into the chromosome, thus replacing the undesired element, e.g., a mutation or signature, with the replacement sequence. In an embodiment, the homology arms flank the most distal cleavage sites.

In an embodiment, the 3' end of the 5' homology arm is the position next to the 5' end of
30 the replacement sequence. In an embodiment, the 5' homology arm can extend at least 10, 20,

30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' from the 5' end of the replacement sequence.

In an embodiment, the 5' end of the 3' homology arm is the position next to the 3' end of the replacement sequence. In an embodiment, the 3' homology arm can extend at least 10, 20,
 5 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 3' from the 3' end of the replacement sequence.

Exemplary Template Nucleic Acids

Exemplary template nucleic acids (also referred to herein as donor constructs) to
 correction a mutation, e.g., a deletion of guanine at nucleotide position 2299 (2299delG) in the
 10 *USH2A* gene, are provided.

Suitable sequence for the 5' homology arm can be selected from (e.g., includes a portion of) or include the following sequence:

AAAACATTTTCTTCATTCGTAAAATGTATATGTGTACTCCTTTAAATAGAAGTAATA
 TAAAAACAGAATTTACTTAGTGTTTAAAGAGGTATGTTCTGAGTCACACAAGATGA
 15 CAAGCAATGTGATTGCTTTATGAGCCAAGGAGAGCATGATTTATATTAATTGAAAAT
 GATAAAATAGAGGAGCATACAAAAGGATTAAACCAAAAATTGCCCTGGATAAGTTT
 TATTTATATTAATTACTTAAATGTGTGGATTCAGAAATAAGTGTATATGCTGTTTTCA
 CAAAAATAGTTATCAGCTGACATTTTTTTCTTTTTTCCCAGCTTCACGAAGGTATAAT
 TAAATAAAAATTGTATATATTTATGGCAGACAACATGATGTTTTTGATATATGTACAC
 20 ATTATAAAATGATTAATTCCAGCTAATTAATGTATCCATCACCTCATGTACTTATCAT
 GTTTTTGGGGTGAGAACATTTAAGATCTAATCTCTTAGCAATTTTCAAGTATACAAT
 ACATTATTATTAAGTATAGTCACCATGCTGTACAATAGAGCTCCAGAACTTATTCAT
 TCTGTCTAGCTGAACTTTGTACTCAGCTTAACCTTTTATTAAACATCTTTAGAGATT
 TCTTATCTTTAGAAAAACAATAATTTGTTATATGTAATTCTACTATAATTTTAAATG
 25 AGCACATTTGTAAAAATAGTTTTTAAGATTTGTTAAAGAGAAAAAGAGCTCCAGCAT
 ATGTAACAGAAACAACATTTGCATTAAGCATTTTTCTTTGCATTAAGTAATAATTAA
 AAATTTATGAAGTTCATCGCAAACAGTTGTATATTAAGCTAAATTAAATATTGTCA
 TTGAATTTTGAGAGTAAGATTGGCCCCCTATGGCATTGCTTGTGAGAAAACACTCAA
 TATTTTGTGTTTCGTATCATCTGCAGTAGCATTGTTTGTGTCTCGTCTATCTTGAATGA
 30 AATCATTTTCCCATCCTCACCTTTTAAATATATTTTATCTTTAGGGCTTAGGTGTGAT
 CATTGCAATTTTGGATTTAAATTTCTCCGAAGCTTTAATGATGTTGGATGTGAGCCCT

GCCAGTGTAACCTCCATGGCTCAGTGAACAAATTCTGCAATCCTCACTCTGGGCAGT
GT

(SEQ ID NO: 387) (5'H arm for 2299delG correction)

Suitable sequence for the 3' homology arm can be selected from (e.g., includes a portion
5 of) or include the following sequence:

AGTGCAAAAAGAAGCCAAAGGACTTCAGTGTGACACCTGCAGAGAAAACCTTTTAT
GGGTTAGATGTCACCAATTGTAAGGCCTGTGACTGTGACACAGCTGGATCCCTCCCT
GGGACTGTCTGTAATGCTAAGACAGGGCAGTGCATCTGCAAGCCCAATGTTGAAGG
GAGACAGTGAATAAATGTTTGGAGGGAACTTCTACCTACGGCAAAATAATTCTTT
10 CCTCTGTCTGCCTTGCAACTGTGATAAGACTGGGACAATAAATGGCTCTCTGCTGTG
TAACAAATCAACAGGACAATGTCCTTGCAAATTAGGGGTAACAGGTCTTCGCTGTAA
TCAGTGTGAGCCTCACAGGTACAATTTGACCATTGACAATTTTCAACACTGCCAGAT
GTGTGAGTGTGATTCCCTTGGGGACATTACCTGGGACCATTGTGACCCAATCAGTGG
CCAGTGCCTGTGTGTGCCTAATCGTCAAGGAAGAAGGTGTAATCAGTGTCAACCAG
15 GTAAGAAAGAAATGTATTACATTTTCAGTGCACAATGACATTCCTTTTGTTAACTTA
GGTAACTTCTCCCTGTTTCTGGTTTGTGGCTTCTACAAATTTTATTTCCAAAATCATT
ACTGTATTTATATCATTATCCAACACATATATAACTATTTAACTTATTCAAAATTATC
TGCATATTTATGTTACTATTTTGAGAGGATACTTTAGATAAAACTCAGCCGATCGGA
TTTATTTTCATAATTGAGACTCAATTTCTACACTTGAAGTAAATCTCCTTTTTAACAGT
20 TTTTAAAAATCAGATCAACAAGAGTCAATTTTATTTTCCAGAGAAAGGAAAATTTG
AGTTGAATATCCATAACAATGCCAAATATTCAAATGATGAACTAAATCTCTGAATAAA
GCTGGCTAAATGTTTTTGCTGAAGAGGCTATATGTTCTAGTTTTATATAGAAATACCT
AGAATTGTTTCCACATGCCATCAAATTAATAAAATAGGCCACTGTTTAATCTCATTA
TATACAAACTTATCTTTCCATCTCTTTCCCAATTGGGAGAGGGATAGACCCCATCTAT
25 GGCTCTCCTTACATTTAAGATTTTAACTAAAATACTATACCTTCTTTACAATAAATTC
ATTATGA

(SEQ ID NO: 388) (3'H arm for 2299delG correction)

In an embodiment, the replacement sequence comprises or consists of a guanine (G)
residue.

30 In an embodiment, to correct a deletion of guanine at nucleotide position 2299
(2299delG) in the *USH2A* gene, the homology arms, e.g., the 5' and 3' homology arms, may

each comprise about 1000 base pairs (bp) of sequence flanking the most distal gRNAs (e.g., 1150bp of sequence on either side of the mutation). The 5' homology arm is shown as bold sequence, the inserted base to correct the guanine deletion is shown as non-bold and boxed sequence, and the 3' homology arm is shown as underlined sequence.

5 **AAAACATTTTCTTCATTCGTAAAATGTATATGTGTACTCCTTTAAATAGAAGTA**
ATATAAAAAACAGAATTTACTTAGTGTTTAAAGAGGTATGTTCTGAGTCACACA
AGATGACAAGCAATGTGATTGCTTTATGAGCCAAGGAGAGCATGATTTATATTA
ATTGAAAATGATAAAATAGAGGAGCATACAAAAGGATTAAACCAAAAATTGCCC
TGGATAAGTTTTATTTATATTAATTACTTAAATGTGTGGATTCAGAAATAAGTGT
10 **ATATGCTGTTTTACAAAAATAGTTATCAGCTGACATTTTTTTCTTTTTTCCCAG**
CTTCACGAAGGTATAATTAAATAAAAATTGTATATATTTATGGCAGACAACATG
ATGTTTTTGATATATGTACACATTATAAAATGATTAATTCCAGCTAATTAATGTAT
CCATCACCTCATGTACTTATCATGTTTTTGGGGTGAGAACATTTAAGATCTAAT
CTCTTAGCAATTTTCAAGTATACAATACATTATTATTAAGTATAGTCACCATGCT
15 **GTACAATAGAGCTCCAGAACTTATTCATTCTGTCTAGCTGAACTTTGTACTCA**
GCTTAACCTTTTATTAAACATCTTTAGAGATTTCTTATCTTTAGAAAAACAATA
ATTTGTTATATGTAATTCTACTATAATTTTAAATGAGCACATTTGTTAAAATAGT
TTTTAAGATTTGTTAAAGAGAAAAAGAGCTCCAGCATATGTAACAGAAACAACA
TTTGCAATTAAGCATTTTTCTTTGCATTAAGTAATAATTAATAAATTTATGAAGTTC
20 **ATCGCAAACAGTTGTATATTAAAGCTAAATTAATATTGTCATTGAATTTTGAG**
AGTAAGATTGGCCCCCTATGGCATTGCTTGTGAGAAAACACTCAATATTTTGTG
TTCGTATCATCTGCAGTAGCATTGTTTGTGTCTCGTCTATCTTGAATGAAATCA
TTTTCCCATCCTCACCTTTTAAATATATTTTATCTTTAGGGCTTAGGTGTGATCA
TTGCAATTTTGGATTTAAATTTCTCCGAAGCTTTAATGATGTTGGATGTGAGCC
25 **CTGCCAGTGTAACCTCCATGGCTCAGTGAACAAATTCTGCAATCCTCACTCTGG**
GCAGTGTG**AGTGCAAAAAAGAAGCCAAAGGACTTCAGTGTGACACCTGCAGAGAA**
AACTTTTATGGGTTAGATGTCACCAATTGTAAGGCCTGTGACTGTGACACAGCTGGA
TCCCTCCCTGGGACTGTCTGTAATGCTAAGACAGGGCAGTGCATCTGCAAGCCCAAT
GTTGAAGGGAGACAGTGCAATAAATGTTTGGAGGGAACTTCTACCTACGGCAAAA
30 **TAATTCCTTCCTCTGTCTGCCTTGCAACTGTGATAAGACTGGGACAATAAATGGCTCT**
CTGCTGTGTAACAAATCAACAGGACAATGTCCTTGCAAATTAGGGGTAACAGGTCTT

CGCTGTAATCAGTGTGAGCCTCACAGGTACAATTTGACCATTGACAATTTTCAACAC
TGCCAGATGTGTGAGTGTGATTCCTTGGGGACATTACCTGGGACCATTTGTGACCCA
ATCAGTGGCCAGTGCCTGTGTGTGCCTAATCGTCAAGGAAGAAGGTGTAATCAGTGT
CAACCAGGTAAGAAAGAAATGTATTACATTTTCAGTGCACAATGACATTCCTTTTGT
5 TAACTTAGGTAAGTCTCCCTGTTTCTGGTTTGTGGCTTCTACAAATTTTATTTCCAA
AATCATTACTGTATTTATATCATTATCCAACACATATATAACTATTTAACTTATTCAA
AATTATCTGCATATTTATGTTACTATTTTGAGAGGATACTTTAGATAAAACTCAGCCG
ATCGGATTTATTTTCATAATTGAGACTCAATTTCTACACTTGAAGTAAATCTCCTTTTT
AACAGTTTTTTTAAAAATCAGATCAACAAGAGTCAATTTTATTTTCCAGAGAAAGGAA
10 AATTTGAGTTGAATATCCATACAATGCCAAATATTCAAATGATGAACTAAATCTCTG
AATAAAGCTGGCTAAATGTTTTTGCTGAAGAGGCTATATGTTCTAGTTTTATATAGA
AATACCTAGAATTGTTTCCACATGCCATCAAATTAATAAAATAGGCCACTGTTTAAT
CTCATTATATACAACTTATCTTTCCATCTCTTTCCCAATTGGGAGAGGGATAGACCC
CATCTATGGCTCTCCTTACATTTAAGATTTTAACTAAAATACTATACCTTCTTTACAA
15 TAAATTCATTATGA (Template Construct 1; SEQ ID NO: 389)

As described below in **Table 13**, shorter homology arms, e.g., 5' and/or 3' homology arms may be used.

20 It is contemplated herein that one or both homology arms may be shortened to avoid including certain sequence repeat elements, e.g., Alu repeats, LINE elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm may be shortened to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

25 In an embodiment, to correct a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene (i.e., insert the missing guanine at position 2299), the 5' homology arm may be shortened less than 600 nucleotides, e.g., approximately 550 nucleotides, i.g., 552 nucleotides, to avoid inclusion of a LINE repeat element in the 5' homology arm. An exemplary 5' homology arm is shown as bold sequence, the inserted base to correct the guanine
30 deletion is shown as non-bold and boxed sequence, and an exemplary 3' homology arm is shown as underlined sequence.

AGCTTAACCTTTTATTAAACATCTTTAGAGATTTCTTATCTTTAGAAAAACAAC
AATTTGTTATATGTAATTCTACTATAATTTTAAATGAGCACATTTGTTAAAATAG
TTTTTAAGATTTGTTAAAGAGAAAAAGAGCTCCAGCATATGTAACAGAAACAAC
ATTTGCATTAAGCATTTTTCTTTGCATTAAGTAATAATTA AAAATTTATGAAGTT
5 CATCGCAAACAGTTGTATATTAAAGCTAAATTAATATTGTCATTGAATTTTGA
GAGTAAGATTGGCCCCCTATGGCATTGCTTGTGAGAAAACACTCAATATTTTGT
GTTTCGTATCATCTGCAGTAGCATTGTTTGTGTCTCGTCTATCTTGAATGAAATC
ATTTTCCCATCCTCACCTTTTAAATATATTTTATCTTTAGGGCTTAGGTGTGATC
ATTGCAATTTTGGATTTAAATTTCTCCGAAGCTTTAATGATGTTGGATGTGAGC
10 CCTGCCAGTGTAACCTCCATGGCTCAGTGAACAAATTCTGCAATCCTCACTCTG
GGCAGTGTGAGTGCAAAAAAGAAGCCAAAGGACTTCAGTGTGACACCTGCAGAGA
AAACTTTTATGGGTTAGATGTCACCAATTGTAAGGCCTGTGACTGTGACACAGCTGG
ATCCCTCCCTGGGACTGTCTGTAATGCTAAGACAGGGCAGTGCATCTGCAAGCCCAA
TGTTGAAGGGAGACAGTGCAATAAATGTTTGGAGGGAACTTCTACCTACGGCAAA
15 ATAATTCTTTCCTCTGTCTGCCTTGCAACTGTGATAAGACTGGGACAATAAATGGCT
CTCTGCTGTGTAACAAATCAACAGGACAATGTCCTTGCAAATTAGGGGTAACAGGTC
TTCGCTGTAATCAGTGTGAGCCTCACAGGTACAATTTGACCATTGACAATTTTCAAC
ACTGCCAGATGTGTGAGTGTGATTCTTGGGGACATTACCTGGGACCATTGTGACC
CAATCAGTGGCCAGTGCCTGTGTGTGCCTAATCGTCAAGGAAGAAGGTGTAATCAGT
20 GTCAACCAGGTAAGAAAGAAATGTATTACATTTTCAGTGCACAATGACATTCCTTTT
GTAACTTAGGTAACCTTCTCCCTGTTTCTGGTTTGTGGCTTCTACAAATTTTATTTCCA
AAATCATTACTGTATTTATATCATTATCCAACACATATATAACTATTTAACTTATTCA
AAATTATCTGCATATTTATGTTACTATTTTGAGAGGATACTTTAGATAAAACTCAGCC
GATCGGATTTATTTTCATAATTGAGACTCAATTTCTACACTTGAAGTAAATCTCCTTTT
25 TAACAGTTTTTTAAAAATCAGATCAACAAGAGTCAATTTTATTTTCCAGAGAAAGGA
AAATTTGAGTTGAATATCCATACAATGCCAAATATTCAAATGATGAACTAAATCTCT
GAATAAAGCTGGCTAAATGTTTTTGCTGAAGAGGCTATATGTTCTAGTTTTATATAG
AAATACCTAGAATTGTTTCCACATGCCATCAAATTAATAAAATAGGCCACTGTTTAA
TCTCATTATATACAACTTATCTTTCCATCTCTTTCCCAATTGGGAGAGGGATAGACC
30 CCATCTATGGCTCTCCTTACATTTAAGATTTTAACTAAAATACTATACCTTCTTTACA
ATAAATTCATTATGA (Template Construct 2; SEQ ID NO: 390)

It is contemplated herein that, in an embodiment, template nucleic acids for correcting a mutation may designed for use as a single-stranded oligonucleotide (ssODN). When using a ssODN, 5' and 3' homology arms may range up to about 200 base pairs (bp) in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length. Longer homology arms are also contemplated for ssODNs as improvements in oligonucleotide synthesis continue to be made.

In an embodiment, an ssODN may be used to correct a deletion of guanine at nucleotide position 2299 (2299delG) in the *USH2A* gene (i.e., insert the missing guanine at position 2299). For example, the ssODN may include 50 bp 5' and 3' homology arms as shown below. The 5' homology arm is shown as bold sequence, the inserted base to correct the guanine deletion is shown as non-bold and boxed sequence, and the 3' homology arm is shown as underlined sequence.

ACCTCCATGGCTCAGTGAACAAATTCTGCAATCCTCACTCTGGGCAGTGTG**AGT**
GCAAAAAAGAAGCCAAAGGACTTCAGTGTGACACCTGCAGAGAAAAC (Template

Construct 3; SEQ ID NO: 391)

The table below provides exemplary template nucleotides. In an embodiment, the template nucleotide includes the 5' homology arm and the 3' homology arm of a row from this **Table 13**. In other embodiments, a 5' homology arm from the first column can be combined with a 3' homology arm from this Table. In each embodiment, the combination of the 5' and 3' homology arms include the replacement sequence, a guanine residue to correct the guanine deletion at position 2299 of *USH2A*.

It is contemplated herein that, in an embodiment, Cas9 could potentially cleave donor constructs either prior to or following homology directed repair (e.g., homologous recombination), resulting in a possible non-homologous-end-joining event and further DNA sequence mutation at the chromosomal locus of interest. Therefore, to avoid cleavage of the donor sequence before and/or after Cas9-mediated homology directed repair, alternate versions of the donor sequence may be used where silent mutations are introduced. These silent mutations may disrupt Cas9 binding and cleavage, but not disrupt the amino acid sequence of the repaired gene. For example, mutations may include those made to a donor sequence to repair the *USH2A* gene, the mutant form which can cause Usher Syndrome. If gRNA *USH2A*-179 with the 20-base target sequence GTTAGATGTCACCAATTGTA is used with a donor construct to

correct the 2299G deletion and the donor construct contains the sequence

ACTTTTATGGGTTAGATGTCACCAATTGTAAGGCCTGTGACTG, the donor sequence

may be changed to ACTTTTATGGGTTAGATGTCACCAATTGTAAAGCCTGTGACTG,

where the bold A has been changed from a G at that position so that codon 793 still codes for the

- 5 amino acid lysine, but the PAM sequence AGG has been modified to AAG to reduce or eliminate Cas9 cleavage at that locus.

Table 13

5' homology arm (the number of nucleotides from SEQ ID NO: 5'H, beginning at the 3' end of SEQ ID NO: 5'H)	Replacement Sequence=G	3' homology arm (the number of nucleotides from SEQ ID NO: 3'H, beginning at the 5' end of SEQ ID NO: 3'H)
10 or more		10 or more
20 or more		20 or more
50 or more		50 or more
100 or more		100 or more
150 or more		150 or more
200 or more		200 or more
250 or more		250 or more
300 or more		300 or more
350 or more		350 or more
400 or more		400 or more
450 or more		450 or more
500 or more		500 or more
550 or more		550 or more
600 or more		600 or more
650 or more		650 or more
700 or more		700 or more
750 or more		750 or more
800 or more		800 or more
850 or more		850 or more
900 or more		900 or more
1000 or more		1000 or more
1100 or more		1100 or more
1200 or more		1200 or more
1300 or more		1300 or more
1400 or more		1400 or more

1500 or more		1500 or more
1600 or more		1600 or more
1700 or more		1700 or more
1800 or more		1800 or more
1900 or more		1900 or more
1200 or more		1200 or more
At least 50 but not long enough to include a repeated element.		At least 50 but not long enough to include a repeated element.
At least 100 but not long enough to include a repeated element.		At least 100 but not long enough to include a repeated element.
At least 150 but not long enough to include a repeated element.		At least 150 but not long enough to include a repeated element.
5 to 100 nucleotides		5 to 100 nucleotides
10 to 150 nucleotides		10 to 150 nucleotides
20 to 150 nucleotides		20 to 150 nucleotides
Template Construct No. 1		
Template Construct No. 2		
Template Construct No. 3		

In an embodiment, a single or dual nickase eaCas9 is used to cleave the target DNA near the site of the mutation, or signature, to be modified, e.g., replaced. While not wishing to be bound by theory, in an embodiment, it is believed that the Cas9 mediated break induces HDR
5 with the template nucleic acid to replace the target DNA sequence with the template sequence.

V.2 NHEJ Approaches for Gene Targeting

As described herein, nuclease-induced non-homologous end-joining (NHEJ) can be used to target gene-specific knockouts. Nuclease-induced NHEJ can also be used to remove (e.g.,
10 delete) sequences in a gene of interest.

While not wishing to be bound by theory, it is believed that, in an embodiment, the genomic alterations associated with the methods described herein rely on nuclease-induced NHEJ and the error-prone nature of the NHEJ repair pathway. NHEJ repairs a double-strand

break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. Two-thirds of these mutations typically alter the reading frame and, therefore, produce a non-functional protein. Additionally, mutations that maintain the reading frame, but which insert or delete a significant amount of sequence, can destroy functionality of the protein. This is locus dependent as mutations in critical functional domains are likely less tolerable than mutations in non-critical regions of the protein.

The indel mutations generated by NHEJ are unpredictable in nature; however, at a given break site certain indel sequences are favored and are over represented in the population, likely due to small regions of microhomology. The lengths of deletions can vary widely; most commonly in the 1-50 bp range, but they can reach greater than 100-200 bp. Insertions tend to be shorter and often include short duplications of the sequence immediately surrounding the break site. However, it is possible to obtain large insertions, and in these cases, the inserted sequence has often been traced to other regions of the genome or to plasmid DNA present in the cells.

Because NHEJ is a mutagenic process, it can also be used to delete small sequence motifs (e.g., motifs less than or equal to 50 nucleotides in length) as long as the generation of a specific final sequence is not required. If a double-strand break is targeted near to a target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. In this way, DNA segments as large as several hundred kilobases can be deleted. Both of these approaches can be used to delete specific DNA sequences; however, the error-prone nature of NHEJ may still produce indel mutations at the site of repair.

Both double strand cleaving eaCas9 molecules and single strand, or nickase, eaCas9 molecules can be used in the methods and compositions described herein to generate NHEJ-mediated indels. NHEJ-mediated indels targeted to the gene, e.g., a coding region, e.g., an

early coding region of a gene, of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a start codon, within a first exon of the coding sequence, or within 500 bp of the start codon (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp).

5 Placement of double strand or single strand breaks relative to the target position

In an embodiment, in which a gRNA and Cas9 nuclease generate a double strand break for the purpose of inducing NHEJ-mediated indels, a gRNA, e.g., a unimolecular (or chimeric) or modular gRNA molecule, is configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site is between 0-30 bp
10 away from the target position (e.g., less than 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position).

In an embodiment, in which two gRNAs complexing with Cas9 nickases induce two single strand breaks for the purpose of inducing NHEJ-mediated indels, two gRNAs, e.g., independently, unimolecular (or chimeric) or modular gRNA, are configured to position two
15 single-strand breaks to provide for NHEJ repair a nucleotide of the target position. In an embodiment, the gRNAs are configured to position cuts at the same position, or within a few nucleotides of one another, on different strands, essentially mimicking a double strand break. In an embodiment, the closer nick is between 0-30 bp away from the target position (e.g., less than 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position), and the two nicks are
20 within 25-55 bp of each other (e.g., between 25 to 50, 25 to 45, 25 to 40, 25 to 35, 25 to 30, 50 to 55, 45 to 55, 40 to 55, 35 to 55, 30 to 55, 30 to 50, 35 to 50, 40 to 50, 45 to 50, 35 to 45, or 40 to 45 bp) and no more than 100 bp away from each other (e.g., no more than 90, 80, 70, 60, 50, 40, 30, 20 or 10 bp). In an embodiment, the gRNAs are configured to place a single strand break on either side of a nucleotide of the target position.

25 Both double strand cleaving eaCas9 molecules and single strand, or nickase, eaCas9 molecules can be used in the methods and compositions described herein to generate breaks both sides of a target position. Double strand or paired single strand breaks may be generated on both sides of a target position to remove the nucleic acid sequence between the two cuts (e.g., the region between the two breaks is deleted). In one embodiment, two gRNAs, e.g.,
30 independently, unimolecular (or chimeric) or modular gRNA, are configured to position a double-strand break on both sides of a target position. In an alternate embodiment, three gRNAs,

e.g., independently, unimolecular (or chimeric) or modular gRNA, are configured to position a double strand break (i.e., one gRNA complexes with a cas9 nuclease) and two single strand breaks or paired single strand breaks (i.e., two gRNAs complex with Cas9 nickases) on either side of the target position. In another embodiment, four gRNAs, e.g., independently,

5 unimolecular (or chimeric) or modular gRNA, are configured to generate two pairs of single strand breaks (i.e., two pairs of two gRNAs complex with Cas9 nickases) on either side of the target position. The double strand break(s) or the closer of the two single strand nicks in a pair will ideally be within 0-500 bp of the target position (e.g., no more than 450, 400, 350, 300, 250, 200, 150, 100, 50 or 25 bp from the target position). When nickases are used, the two nicks in a pair are within 25-55 bp of each other (e.g., between 25 to 50, 25 to 45, 25 to 40, 25 to 35, 25 to 30, 50 to 55, 45 to 55, 40 to 55, 35 to 55, 30 to 55, 30 to 50, 35 to 50, 40 to 50, 45 to 50, 35 to 45, or 40 to 45 bp) and no more than 100 bp away from each other (e.g., no more than 90, 80, 70, 60, 50, 40, 30, 20 or 10 bp).

15 **V.3 Single-Strand Annealing**

Single strand annealing (SSA) is another DNA repair process that repairs a double-strand break between two repeat sequences present in a target nucleic acid. Repeat sequences utilized by the SSA pathway are generally greater than 30 nucleotides in length. Resection at the break ends occurs to reveal repeat sequences on both strands of the target nucleic acid. After resection, 20 single strand overhangs containing the repeat sequences are coated with RPA protein to prevent the repeats sequences from inappropriate annealing, e.g., to themselves. RAD52 binds to and each of the repeat sequences on the overhangs and aligns the sequences to enable the annealing of the complementary repeat sequences. After annealing, the single-strand flaps of the overhangs are cleaved. New DNA synthesis fills in any gaps, and ligation restores the DNA duplex. As a 25 result of the processing, the DNA sequence between the two repeats is deleted. The length of the deletion can depend on many factors including the location of the two repeats utilized, and the pathway or processivity of the resection.

In contrast to HDR pathways, SSA does not require a template nucleic acid to alter or correct a target nucleic acid sequence. Instead, the complementary repeat sequence is utilized.

V. 4 Other DNA Repair Pathways

SSBR (single strand break repair)

Single-stranded breaks (SSB) in the genome are repaired by the SSBR pathway, which is a distinct mechanism from the DSB repair mechanisms discussed above. The SSBR pathway has four major stages: SSB detection, DNA end processing, DNA gap filling, and DNA ligation. A more detailed explanation is given in Caldecott, Nature Reviews Genetics 9, 619-631 (August 2008), and a summary is given here.

In the first stage, when a SSB forms, PARP1 and/or PARP2 recognize the break and recruit repair machinery. The binding and activity of PARP1 at DNA breaks is transient and it seems to accelerate SSBr by promoting the focal accumulation or stability of SSBr protein complexes at the lesion. Arguably the most important of these SSBr proteins is XRCC1, which functions as a molecular scaffold that interacts with, stabilizes, and stimulates multiple enzymatic components of the SSBr process including the protein responsible for cleaning the DNA 3' and 5' ends. For instance, XRCC1 interacts with several proteins (DNA polymerase beta, PNK, and three nucleases, APE1, APTX, and APLF) that promote end processing. APE1 has endonuclease activity. APLF exhibits endonuclease and 3' to 5' exonuclease activities. APTX has endonuclease and 3' to 5' exonuclease activity.

This end processing is an important stage of SSBR since the 3'- and/or 5'-termini of most, if not all, SSBs are 'damaged'. End processing generally involves restoring a damaged 3'-end to a hydroxylated state and and/or a damaged 5' end to a phosphate moiety, so that the ends become ligation-competent. Enzymes that can process damaged 3' termini include PNKP, APE1, and TDP1. Enzymes that can process damaged 5' termini include PNKP, DNA polymerase beta, and APTX. LIG3 (DNA ligase III) can also participate in end processing. Once the ends are cleaned, gap filling can occur.

At the DNA gap filling stage, the proteins typically present are PARP1, DNA polymerase beta, XRCC1, FEN1 (flap endonuclease 1), DNA polymerase delta/epsilon, PCNA, and LIG1. There are two ways of gap filling, the short patch repair and the long patch repair. Short patch repair involves the insertion of a single nucleotide that is missing. At some SSBs, "gap filling" might continue displacing two or more nucleotides (displacement of up to 12 bases have been reported). FEN1 is an endonuclease that removes the displaced 5'-residues. Multiple DNA

polymerases, including Pol β , are involved in the repair of SSBs, with the choice of DNA polymerase influenced by the source and type of SSB.

In the fourth stage, a DNA ligase such as LIG1 (Ligase I) or LIG3 (Ligase III) catalyzes joining of the ends. Short patch repair uses Ligase III and long patch repair uses Ligase I.

5 Sometimes, SSBR is replication-coupled. This pathway can involve one or more of CtIP, MRN, ERCC1, and FEN1. Additional factors that may promote SSBR include: aPARP, PARP1, PARP2, PARG, XRCC1, DNA polymerase β , DNA polymerase δ , DNA polymerase ϵ , PCNA, LIG1, PNK, PNKP, APE1, APTX, APLF, TDP1, LIG3, FEN1, CtIP, MRN, and ERCC1.

10 MMR (mismatch repair)

Cells contain three excision repair pathways: MMR, BER, and NER. The excision repair pathways have a common feature in that they typically recognize a lesion on one strand of the DNA, then exo/endonucleases remove the lesion and leave a 1-30 nucleotide gap that is subsequently filled in by DNA polymerase and finally sealed with ligase. A more complete picture is given in Li, Cell Research (2008) 18:85–98, and a summary is provided here.

Mismatch repair (MMR) operates on mispaired DNA bases.

The MSH2/6 or MSH2/3 complexes both have ATPases activity that plays an important role in mismatch recognition and the initiation of repair. MSH2/6 preferentially recognizes base-base mismatches and identifies mispairs of 1 or 2 nucleotides, while MSH2/3 preferentially recognizes larger ID mispairs.

hMLH1 heterodimerizes with hPMS2 to form hMutL α which possesses an ATPase activity and is important for multiple steps of MMR. It possesses a PCNA/replication factor C (RFC)-dependent endonuclease activity which plays an important role in 3' nick-directed MMR involving EXO1. (EXO1 is a participant in both HR and MMR.) It regulates termination of mismatch-provoked excision. Ligase I is the relevant ligase for this pathway. Additional factors that may promote MMR include: EXO1, MSH2, MSH3, MSH6, MLH1, PMS2, MLH3, DNA Pol δ , RPA, HMGB1, RFC, and DNA ligase I.

30 Base excision repair (BER)

The base excision repair (BER) pathway is active throughout the cell cycle; it is responsible primarily for removing small, non-helix-distorting base lesions from the genome. In

contrast, the related Nucleotide Excision Repair pathway (discussed in the next section) repairs bulky helix-distorting lesions. A more detailed explanation is given in Caldecott, Nature Reviews Genetics 9, 619-631 (August 2008), and a summary is given here.

Upon DNA base damage, base excision repair (BER) is initiated and the process can be simplified into five major steps: (a) removal of the damaged DNA base; (b) incision of the subsequent a basic site; (c) clean-up of the DNA ends; (d) insertion of the correct nucleotide into the repair gap; and (e) ligation of the remaining nick in the DNA backbone. These last steps are similar to the SSBR.

In the first step, a damage-specific DNA glycosylase excises the damaged base through cleavage of the N-glycosidic bond linking the base to the sugar phosphate backbone. Then AP endonuclease-1 (APE1) or bifunctional DNA glycosylases with an associated lyase activity incised the phosphodiester backbone to create a DNA single strand break (SSB). The third step of BER involves cleaning-up of the DNA ends. The fourth step in BER is conducted by Pol β that adds a new complementary nucleotide into the repair gap and in the final step

XRCC1/Ligase III seals the remaining nick in the DNA backbone. This completes the short-patch BER pathway in which the majority (~80%) of damaged DNA bases are repaired.

However, if the 5' -ends in step 3 are resistant to end processing activity, following one nucleotide insertion by Pol β there is then a polymerase switch to the replicative DNA

polymerases, Pol δ/ϵ , which then add ~2–8 more nucleotides into the DNA repair gap. This creates a 5' -flap structure, which is recognized and excised by flap endonuclease-1 (FEN-1) in association with the processivity factor proliferating cell nuclear antigen (PCNA). DNA ligase I then seals the remaining nick in the DNA backbone and completes long-patch BER. Additional factors that may promote the BER pathway include: DNA glycosylase, APE1, Polb, Pold, Pole, XRCC1, Ligase III, FEN-1, PCNA, RECQL4, WRN, MYH, PNKP, and APTX.

Nucleotide excision repair (NER)

Nucleotide excision repair (NER) is an important excision mechanism that removes bulky helix-distorting lesions from DNA. Additional details about NER are given in Marteijn et al., Nature Reviews Molecular Cell Biology 15, 465–481 (2014), and a summary is given here.

NER a broad pathway encompassing two smaller pathways: global genomic NER (GG-NER) and transcription coupled repair NER (TC-NER). GG-NER and TC-NER use different factors

for recognizing DNA damage. However, they utilize the same machinery for lesion incision, repair, and ligation.

Once damage is recognized, the cell removes a short single-stranded DNA segment that contains the lesion. Endonucleases XPF/ERCC1 and XPG (encoded by ERCC5) remove the lesion by cutting the damaged strand on either side of the lesion, resulting in a single-strand gap of 22–30 nucleotides. Next, the cell performs DNA gap filling synthesis and ligation. Involved in this process are: PCNA, RFC, DNA Pol δ , DNA Pol ϵ or DNA Pol κ , and DNA ligase I or XRCC1/Ligase III. Replicating cells tend to use DNA pol ϵ and DNA ligase I, while non-replicating cells tend to use DNA Pol δ , DNA Pol κ , and the XRCC1/ Ligase III complex to perform the ligation step.

NER can involve the following factors: XPA-G, POLH, XPF, ERCC1, XPA-G, and LIG1. Transcription-coupled NER (TC-NER) can involve the following factors: CSA, CSB, XPB, XPD, XPG, ERCC1, and TTDA. Additional factors that may promote the NER repair pathway include XPA-G, POLH, XPF, ERCC1, XPA-G, LIG1, CSA, CSB, XPA, XPB, XPC, XPD, XPF, XPG, TTDA, UVSSA, USP7, CETN2, RAD23B, UV-DDB, CAK subcomplex, RPA, and PCNA.

Interstrand Crosslink (ICL)

A dedicated pathway called the ICL repair pathway repairs interstrand crosslinks. Interstrand crosslinks, or covalent crosslinks between bases in different DNA strand, can occur during replication or transcription. ICL repair involves the coordination of multiple repair processes, in particular, nucleolytic activity, translesion synthesis (TLS), and HDR. Nucleases are recruited to excise the ICL on either side of the crosslinked bases, while TLS and HDR are coordinated to repair the cut strands. ICL repair can involve the following factors: endonucleases, e.g., XPF and RAD51C, endonucleases such as RAD51, translesion polymerases, e.g., DNA polymerase zeta and Rev1), and the Fanconi anemia (FA) proteins, e.g., FancJ.

Other pathways

Several other DNA repair pathways exist in mammals. Translesion synthesis (TLS) is a pathway for repairing a single stranded break left after a defective replication event and involves translesion polymerases, e.g., DNA pol ζ and Rev1..

Error-free postreplication repair (PRR) is another pathway for repairing a single stranded break left after a defective replication event.

V.5 Examples of gRNAs in Genome Editing Methods

gRNAs as described herein can be used with a Cas9 molecule that cleaves both or a single strand and a template nucleic acid to alter the sequence of a target nucleic acid, e.g., at a target position or a target genetic signature. gRNAs useful in these method are described below.

In an embodiment, the gRNA, e.g., a chimeric gRNA, is configured such that it comprises one or more of the following properties;

a) it can position, e.g., when targeting a Cas9 molecule that makes double strand breaks, a double strand break (i) within 50, 100, 150 or 200 nucleotides of a target position, or (ii) sufficiently close that the target position is within the region of end resection;

b) it has a targeting domain of at least 17 nucleotides, e.g., a targeting domain of (i) 17, (ii) 18, or (iii) 20 nucleotides; and

c) the tail domain is (i) at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length or (ii) the tail domain comprises 15, 20, 25, 30, 35, 40 nucleotides or all of the corresponding portions of a naturally occurring tail domain, e.g., a naturally occurring *S. pyogenes* or *S. thermophilus* tail domain.

In an embodiment, the gRNA is configured such that it comprises properties: a and b(i).

In an embodiment, the gRNA is configured such that it comprises properties: a and b(ii).

In an embodiment, the gRNA is configured such that it comprises properties: a and b(iii).

In an embodiment, the gRNA is configured such that it comprises properties: a and c.

In an embodiment, the gRNA is configured such that in comprises properties: a, b, and c.

In an embodiment, the gRNA is configured such that in comprises properties: a(i), b(i), and c(i).

In an embodiment, the gRNA is configured such that in comprises properties: a(i), b(i), and c(ii).

In an embodiment, the gRNA is configured such that in comprises properties: a(i), b(iii), and c(i).

In an embodiment, the gRNA is configured such that in comprises properties: a(i), b(iii), and c(ii).

In an embodiment, the gRNA, e.g., a chimeric gRNA, is configured such that it comprises one or more of the following properties;

a) it can position, e.g., when targeting a Cas9 molecule that makes single strand breaks, a single strand break (i) within 50, 100, 150 or 200 nucleotides of a target position, or (ii)

5 sufficiently close that the target position is within the region of end resection;

b) it has a targeting domain of at least 17 nucleotides, e.g., a targeting domain of (i) 17, (ii) 18, or (iii) 20 nucleotides; and

c) the tail domain is (i) at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length, or (ii) the tail domain comprises 15, 20, 25, 30, 35, 40 nucleotides or all of the corresponding portions
10 of a naturally occurring tail domain, e.g., a naturally occurring *S. pyogenes* or *S. thermophilus* tail domain.

In an embodiment, the gRNA is configured such that it comprises properties: a and b(i).

In an embodiment, the gRNA is configured such that it comprises properties: a and b(ii).

In an embodiment, the gRNA is configured such that it comprises properties: a and b(iii).

15 In an embodiment, the gRNA is configured such that it comprises properties: a and c.

In an embodiment, the gRNA is configured such that in comprises properties: a, b, and c.

In an embodiment, the gRNA is configured such that in comprises properties: a(i), b(i), and c(i).

20 In an embodiment, the gRNA is configured such that in comprises properties: a(i), b(i), and c(ii).

In an embodiment, the gRNA is configured such that in comprises properties: a(i), b(iii), and c(i).

In an embodiment, the gRNA is configured such that in comprises properties: a(i), b(iii), and c(ii).

25 In an embodiment, the gRNA is used with a Cas9 nickase molecule having HNH activity, e.g., a Cas9 molecule having the RuvC activity inactivated, e.g., a Cas9 molecule having a mutation at D10, e.g., the D10A mutation.

In an embodiment, the gRNA is used with a Cas9 nickase molecule having RuvC activity, e.g., a Cas9 molecule having the HNH activity inactivated, e.g., a Cas9 molecule having
30 a mutation at H840, e.g., a H840A.

In an embodiment, a pair of gRNAs, e.g., a pair of chimeric gRNAs, comprising a first and a second gRNA, is configured such that they comprises one or more of the following properties;

a) one or both of the gRNAs can position, e.g., when targeting a Cas9 molecule that makes single strand breaks, a single strand break within (i) 50, 100, 150 or 200 nucleotides of a target position, or (ii) sufficiently close that the target position is within the region of end resection;

b) one or both have a targeting domain of at least 17 nucleotides, e.g., a targeting domain of (i) 17 or (ii) 18 nucleotides; and

c) the tail domain of one or both is (i) at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length of (ii) comprises, 15, 20, 25, 30, 35, 40, or all of the corresponding portions of a naturally occurring tail domain, e.g., a naturally occurring *S. pyogenes*, *S. aureus* or *S. thermophilus* tail domain.

d) the gRNAs are configured such that, when hybridized to target nucleic acid, they are separated by 0-50, 0-100, 0-200, at least 10, at least 20, at least 30 or at least 50 nucleotides;

e) the breaks made by the first gRNA and second gRNA are on different strands; and

f) the PAMs are facing outwards.

In an embodiment, one or both of the gRNAs is configured such that it comprises properties: a and b(i).

In an embodiment, one or both of the gRNAs is configured such that it comprises properties: a and b(ii).

In an embodiment, one or both of the gRNAs is configured such that it comprises properties: a and b(iii).

In an embodiment, one or both of the gRNAs configured such that it comprises properties: a and c.

In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a, b, and c.

In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(i), and c(i).

In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(i), and c(ii).

In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(i), c, and d.

In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(i), c, and e.

5 In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(i), c, d, and e.

In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(iii), and c(i).

10 In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(iii), and c(ii).

In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(iii), c, and d.

In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(iii), c, and e.

15 In an embodiment, one or both of the gRNAs is configured such that in comprises properties: a(i), b(iii), c, d, and e.

In an embodiment, the gRNAs are used with a Cas9 nickase molecule having HNH activity, e.g., a Cas9 molecule having the RuvC activity inactivated, e.g., a Cas9 molecule having a mutation at D10, e.g., the D10A mutation.

20 In an embodiment, the gRNAs are used with a Cas9 nickase molecule having RuvC activity, e.g., a Cas9 molecule having the HNH activity inactivated, e.g., a Cas9 molecule having a mutation at H840, e.g., a H840A.

In an embodiment, the gRNAs are used with a Cas9 nickase molecule having RuvC activity, e.g., a Cas9 molecule having the HNH activity inactivated, e.g., a Cas9 molecule having a mutation at H863, e.g., a H863A.

25

VI. Target Cells

Cas9 molecules and gRNA molecules, e.g., a Cas9 molecule/gRNA molecule complex, can be used to manipulate a cell, e.g., to edit a target nucleic acid, in a wide variety of cells.

In some embodiments, a cell is manipulated by editing (e.g., correcting) one or more target genes, e.g., as described herein. In some embodiments, the expression of one or more target genes (e.g., one or more target genes described herein) is modulated, e.g., *in vivo*.

5 In an embodiment, the target cell is a retinal cell, e.g., a cell of the retinal pigment epithelium or a photoreceptor cell. In an embodiment, the target cell is a cone photoreceptor cell or cone cell, a rod photoreceptor cell or rod cell, or a macular cone photoreceptor cell. Cone photoreceptor cells in the macula are the first to demonstrate cell death in Usher Syndrome and in cone-rod dystrophies in general (this is the opposite of rod-cone dystrophies). In an exemplary embodiment, cone photoreceptors in the macular are targeted, i.e., cone
10 photoreceptors in the macular are the target cells. In an embodiment, the target cell is a cochlear cell, e.g. an inner hair cell or an outer hair cell.

In an embodiment, the target cell is removed from the subject, the mutation corrected *ex vivo*, and the cell returned to the subject. In an embodiment, a photoreceptor cell is removed from the subject, the mutation corrected *ex vivo*, and the photoreceptor cell returned to the
15 subject. In an embodiment, a cone photoreceptor cell is removed from the subject, the mutation corrected *ex vivo*, and the cone photoreceptor cell returned to the subject. In an embodiment, an inner or outer hair cell is removed from the subject, the mutation corrected *ex vivo*, and the inner or outer hair cell returned to the subject.

In an embodiment, the cells are induced pluripotent stem cells (iPS) cells or cells derived
20 from iPS cells, e.g., iPS cells from the subject, modified to alter the gene and differentiated into retinal progenitor cells or retinal cells, e.g., retinal photoreceptors, and injected into the eye of the subject, e.g., subretinally, e.g., in the submacular region of the retina.

In an embodiment, the cells are induced pluripotent stem cells (iPS) cells or cells derived from iPS cells, e.g., iPS cells from the subject, modified to alter the gene and differentiated into
25 cochlear cells, e.g., inner or outer hair cells, and injected into the cochlea of the subject.

In an embodiment, the cells are targeted *in vivo*, e.g., by delivery of the components, e.g., a Cas9 molecule and gRNA molecules, or a Cas9 molecule, gRNA molecules and donor template, to the target cells. In an embodiment, the target cells are retinal pigment epithelium or photoreceptor cells. In an embodiment, the target cells are inner or outer hair cells of the
30 cochlea. In an embodiment, AAV is used to transduce the target cells.

VII. Delivery, Formulations and Routes of Administration

The components, e.g., a Cas9 molecule, gRNA molecule or template construct molecule, or all three, can be delivered, formulated, or administered in a variety of forms, see, e.g., **Tables 14 and 15**. When a Cas9 or gRNA component is delivered encoded in DNA the DNA will typically include a control region, e.g., comprising a promoter, to effect expression. Useful promoters for Cas9 molecule sequences include CMV, EF-1a, MSCV, PGK, CAG control promoters. Useful promoters for gRNAs include H1, EF-1a and U6 promoters. Promoters with similar or dissimilar strengths can be selected to tune the expression of components. Sequences encoding a Cas9 molecule can comprise a nuclear localization signal (NLS), e.g., an SV40 NLS.

In an embodiment, a promoter for a Cas9 molecule or a gRNA molecule can be, independently, inducible, tissue specific, or cell specific.

Table 14 provides examples of how the components can be formulated, delivered, or administered.

Table 14

Elements			
Cas9 Molecule(s)	gRNA Molecule(s)	Donor Template Nucleic Acid	Comments
DNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
DNA		DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA, here from a single molecule. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	<u>DNA</u>	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this

			embodiment, the donor template is provided on the same DNA molecule that encodes the Cas9.
DNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the Cas9.
mRNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this embodiment, the donor template is provided as a DNA molecule.
mRNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided as a separate DNA molecule.
mRNA	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
Protein	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is provided as a protein, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided as a separate DNA molecule.
Protein	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is provided as a protein, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
Protein	RNA	DNA	In this embodiment, an eaCas9 molecule is provided as a protein, and a gRNA is provided as transcribed or synthesized RNA. In this embodiment, the donor template is provided as a DNA molecule.

Elements			
Cas9 Molecule(s)	gRNA Molecule(s)	Donor Template Nucleic Acid	Comments
DNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
DNA		DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA, here from a single molecule. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided on the same DNA molecule that encodes the Cas9.
DNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the Cas9.
mRNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this

			embodiment, the donor template is provided as a DNA molecule.
mRNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided as a separate DNA molecule.
mRNA	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
Protein	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is provided as a protein, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided as a separate DNA molecule.
Protein	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is provided as a protein, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
Protein	RNA	DNA	In this embodiment, an eaCas9 molecule is provided as a protein, and a gRNA is provided as transcribed or synthesized RNA. In this embodiment, the donor template is provided as a DNA molecule.

Table 15 summarizes various delivery methods for the components of a Cas system, e.g., the Cas9 molecule component and the gRNA molecule component, as described herein.

Table 15

Delivery Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical (eg, electroporation, particle gun, Calcium Phosphate transfection)	YES	Transient	NO	Nucleic Acids and Proteins

<i>Viral</i>	Retrovirus	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modifications	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
<i>Non-Viral</i>	Cationic Liposomes	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
	Polymeric Nanoparticles	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
<i>Biological Non-Viral Delivery Vehicles</i>	Attenuated Bacteria	YES	Transient	NO	Nucleic Acids
	Engineered Bacteriophages	YES	Transient	NO	Nucleic Acids
	Mammalian Virus-like Particles	YES	Transient	NO	Nucleic Acids
	Biological liposomes: Erythrocyte Ghosts and Exosomes	YES	Transient	NO	Nucleic Acids

DNA-based Delivery of a Cas9 molecule and or one or more gRNA molecules

Nucleic acids encoding Cas9 molecules (e.g., eaCas9 molecules) and/or gRNA molecules, can be administered to subjects or delivered into cells by art-known methods or as described herein. For example, Cas9-encoding and/or gRNA-encoding DNA can be delivered, e.g., by vectors (e.g., viral or non-viral vectors), non-vector based methods (e.g., using naked DNA or DNA complexes), or a combination thereof.

DNA encoding Cas9 molecules (e.g., eaCas9 molecules) and/or gRNA molecules can be conjugated to molecules (e.g., N-acetylgalactosamine) promoting uptake by the target cells (e.g.,

hepatocytes). Donor template molecules can be conjugated to molecules (e.g., N-acetylgalactosamine) promoting uptake by the target cells (e.g., hepatocytes).

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a vector (e.g., viral vector/virus or plasmid).

5 Vectors can comprise a sequence that encodes a Cas9 molecule and/or a gRNA molecule.

A vectors can also comprise a sequence encoding a signal peptide (e.g., for nuclear localization, nucleolar localization, mitochondrial localization), fused, e.g., to a Cas9 molecule sequence. For example, the vectors can comprise a nuclear localization sequence (e.g., from SV40) fused to the sequence encoding the Cas9 molecule.

10 One or more regulatory/control elements, e.g., promoters, enhancers, introns, polyadenylation signals, Kozak consensus sequences, and internal ribosome entry sites (IRES), can be included in the vectors. In some embodiments, the promoter is recognized by RNA polymerase II (e.g., a CMV promoter). In other embodiments, the promoter is recognized by RNA polymerase III (e.g., a U6 promoter). In some embodiments, the promoter is a regulated
15 promoter (e.g., inducible promoter). In other embodiments, the promoter is a constitutive promoter. In some embodiments, the promoter is a tissue specific promoter. In some embodiments, the promoter is a viral promoter. In other embodiments, the promoter is a non-viral promoter.

In some embodiments, the vector is a viral vector (e.g., for generation of recombinant
20 viruses). In some embodiments, the virus is a DNA virus (e.g., dsDNA or ssDNA virus). In other embodiments, the virus is an RNA virus (e.g., an ssRNA virus). In some embodiments, the virus infects dividing cells. In other embodiments, the virus infects non-dividing cells. Exemplary viral vectors/viruses include, e.g., retroviruses, lentiviruses, adenovirus, adeno-associated virus (AAV), vaccinia viruses, poxviruses, and herpes simplex viruses.

25 In some embodiments, the virus infects both dividing and non-dividing cells. In some embodiments, the virus can integrate into the host genome. In some embodiments, the virus is engineered to have reduced immunity, e.g., in human. In some embodiments, the virus is replication-competent. In other embodiments, the virus is replication-defective, e.g., having one or more coding regions for the genes necessary for additional rounds of virion replication and/or
30 packaging replaced with other genes or deleted. In some embodiments, the virus causes transient expression of the Cas9 molecule and/or the gRNA molecule. In other embodiments, the virus

causes long-lasting, e.g., at least 1 week, 2 weeks, 1 month, 2 months, 3 months, 6 months, 9 months, 1 year, 2 years, or permanent expression, of the Cas9 molecule and/or the gRNA molecule. The packaging capacity of the viruses may vary, e.g., from at least about 4 kb to at least about 30 kb, e.g., at least about 5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 30 kb, 35 kb, 40 kb, 45 kb, or 50 kb.

Exemplary viral vectors/viruses include, e.g., retroviruses, lentiviruses, adenovirus, adeno-associated virus (AAV), vaccinia viruses, poxviruses, and herpes simplex viruses.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a recombinant retrovirus. In some embodiments, the retrovirus (e.g., Moloney murine leukemia virus) comprises a reverse transcriptase, e.g., that allows integration into the host genome. In some embodiments, the retrovirus is replication-competent. In other embodiments, the retrovirus is replication-defective, e.g., having one of more coding regions for the genes necessary for additional rounds of virion replication and packaging replaced with other genes, or deleted.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a recombinant lentivirus. For example, the lentivirus is replication-defective, e.g., does not comprise one or more genes required for viral replication.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a recombinant adenovirus. In some embodiments, the adenovirus is engineered to have reduced immunity in human.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a recombinant AAV. In some embodiments, the AAV can incorporate its genome into that of the host cell. In some embodiments, the AAV is a self-complementary adeno-associated virus (scAAV), e.g., a scAAV that packages both strands which anneal together to form double stranded DNA.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a hybrid virus, e.g., a hybrid of one or more of the viruses described herein.

A Packaging cell is used to form a virus particle that is capable of infecting a target cell. Such a cell includes a 293 cell, which can package adenovirus, and a ψ 2 cell or a PA317 cell, which can package retrovirus. A viral vector used in gene therapy is usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vector typically contains the minimal viral sequences required for packaging and subsequent integration into a

host or target cell (if applicable), with other viral sequences being replaced by an expression cassette encoding the protein to be expressed, eg. Cas9. For example, an AAV vector used in gene therapy typically only possesses inverted terminal repeat (ITR) sequences from the AAV genome which are required for packaging and gene expression in the host or target cell. The missing viral functions are supplied in *trans* by the packaging cell line. Henceforth, the viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

In an embodiment, the viral vector has the ability of cell type and/or tissue type recognition. For example, the viral vector can be pseudotyped with a different/alternative viral envelope glycoprotein; engineered with a cell type-specific receptor (e.g., genetic modification of the viral envelope glycoproteins to incorporate targeting ligands such as a peptide ligand, a single chain antibody, a growth factor); and/or engineered to have a molecular bridge with dual specificities with one end recognizing a viral glycoprotein and the other end recognizing a moiety of the target cell surface (e.g., ligand-receptor, monoclonal antibody, avidin-biotin and chemical conjugation).

In an embodiment, the viral vector achieves cell type specific expression. For example, a tissue-specific promoter can be constructed to restrict expression of the transgene (Cas 9 and gRNA) in only the target cell. The specificity of the vector can also be mediated by microRNA-dependent control of transgene expression. In an embodiment, the viral vector has increased efficiency of fusion of the viral vector and a target cell membrane. For example, a fusion protein such as fusion-competent hemagglutinin (HA) can be incorporated to increase viral uptake into cells. In an embodiment, the viral vector has the ability of nuclear localization. For example, a virus that requires the breakdown of the cell wall (during cell division) and therefore will not infect a non-dividing cell can be altered to incorporate a nuclear localization peptide in the matrix protein of the virus thereby enabling the transduction of non-proliferating cells.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a non-vector based method (e.g., using naked DNA or DNA complexes). For example, the DNA can

be delivered, e.g., by organically modified silica or silicate (Ormosil), electroporation, gene gun, sonoporation, magnetofection, lipid-mediated transfection, dendrimers, inorganic nanoparticles, calcium phosphates, or a combination thereof.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a combination of a vector and a non-vector based method. For example, virosomes combine liposomes with an inactivated virus (e.g., HIV or influenza virus), which can result in more efficient gene transfer, e.g., in respiratory epithelial cells than either viral or liposomal methods alone.

In an embodiment, the delivery vehicle is a non-viral vector. In an embodiment, the non-viral vector is an inorganic nanoparticle. Exemplary inorganic nanoparticles include, e.g., magnetic nanoparticles (e.g., Fe_3MnO_2) or silica. The outer surface of the nanoparticle can be conjugated with a positively charged polymer (e.g., polyethylenimine, polylysine, polyserine) which allows for attachment (e.g., conjugation or entrapment) of payload. In an embodiment, the non-viral vector is an organic nanoparticle (e.g., entrapment of the payload inside the nanoparticle). Exemplary organic nanoparticles include, e.g., SNALP liposomes that contain cationic lipids together with neutral helper lipids which are coated with polyethylene glycol (PEG) and protamine and nucleic acid complex coated with lipid coating.

Exemplary lipids for gene transfer are shown below in **Table 16**.

Table 16: Lipids Used for Gene Transfer

Lipid	Abbreviation	Feature
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
<i>N</i> -[1-(2,3-Dioleoyloxy)propyl] <i>N,N,N</i> -trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Diocadecylamidoglycylspermine	DOGS	Cationic
<i>N</i> -(3-Aminopropyl)- <i>N,N</i> -dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleoyloxy- <i>N</i> -[2(sperminecarboxamido-ethyl)- <i>N,N</i> -dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleoyl-3-trimethylammonium-propane	DOPA	Cationic
<i>N</i> -(2-Hydroxyethyl)- <i>N,N</i> -dimethyl-2,3-bis(tetradecyloxy)-1-	MDRIE	Cationic

propanaminium bromide		
Dimyristoxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3 β -[<i>N</i> -(<i>N</i> ', <i>N</i> '-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Diodeoxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Diocadecylamidoglycylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammonium bromide	CLIP-6	Cationic
Ethyl dimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearoyloxy- <i>N,N</i> -dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
<i>O,O'</i> -Dimyristyl- <i>N</i> -lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylphosphocholine	DSEPC	Cationic
<i>N</i> -Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
<i>N</i> - <i>t</i> -Butyl- <i>N</i> O-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl]imidazolium chloride	DOTIM	Cationic
<i>N</i> 1-Cholesteryloxy carbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)- <i>N</i> -ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleyloxy-3- dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]- dioxolane	DLin-KC2-DMA	Cationic
dilinoleyl- methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Exemplary polymers for gene transfer are shown below in **Table 17**.

Table 17: Polymers Used for Gene Transfer

Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis(succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimide	DTBP
Poly(ethylene imine) biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(<i>N</i> -vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM
Poly(amido ethylenimine)	SS-PAEI
Triethylenetetramine	TETA

Poly(β -aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	
Poly(α -[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(N-ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(N-2-hydroxypropylmethacrylamide)	pHPMA
Poly(2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
N-Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

In an embodiment, the vehicle has targeting modifications to increase target cell uptake of nanoparticles and liposomes, e.g., cell specific antigens, monoclonal antibodies, single chain antibodies, aptamers, polymers, sugars (e.g., N-acetylgalactosamine (GalNAc)), and cell penetrating peptides. In an embodiment, the vehicle uses fusogenic and endosome-destabilizing peptides/polymers. In an embodiment, the vehicle undergoes acid-triggered conformational changes (e.g., to accelerate endosomal escape of the cargo). In an embodiment, a stimuli-cleavable polymer is used, e.g., for release in a cellular compartment. For example, disulfide-based cationic polymers that are cleaved in the reducing cellular environment can be used.

In an embodiment, the delivery vehicle is a biological non-viral delivery vehicle. In an embodiment, the vehicle is an attenuated bacterium (e.g., naturally or artificially engineered to be invasive but attenuated to prevent pathogenesis and expressing the transgene (e.g., *Listeria monocytogenes*, certain *Salmonella strains*, *Bifidobacterium longum*, and modified *Escherichia coli*), bacteria having nutritional and tissue-specific tropism to target specific tissues, bacteria having modified surface proteins to alter target tissue specificity). In an embodiment, the vehicle is a genetically modified bacteriophage (e.g., engineered phages having large packaging capacity, less immunogenic, containing mammalian plasmid maintenance sequences and having incorporated targeting ligands). In an embodiment, the vehicle is a mammalian virus-like particle. For example, modified viral particles can be generated (e.g., by purification of the

“empty” particles followed by *ex vivo* assembly of the virus with the desired cargo). The vehicle can also be engineered to incorporate targeting ligands to alter target tissue specificity. In an embodiment, the vehicle is a biological liposome. For example, the biological liposome is a phospholipid-based particle derived from human cells (e.g., erythrocyte ghosts, which are red blood cells broken down into spherical structures derived from the subject (e.g., tissue targeting can be achieved by attachment of various tissue or cell-specific ligands), or secretory exosomes – subject (i.e., patient) derived membrane-bound nanovesicle (30 -100 nm) of endocytic origin (e.g., can be produced from various cell types and can therefore be taken up by cells without the need of for targeting ligands).

In an embodiment, one or more nucleic acid molecules (e.g., DNA molecules) other than the components of a Cas system, e.g., the Cas9 molecule component and/or the gRNA molecule component described herein, are delivered. In an embodiment, the nucleic acid molecule is delivered at the same time as one or more of the components of the Cas system are delivered. In an embodiment, the nucleic acid molecule is delivered before or after (e.g., less than about 30 minutes, 1 hour, 2 hours, 3 hours, 6 hours, 9 hours, 12 hours, 1 day, 2 days, 3 days, 1 week, 2 weeks, or 4 weeks) one or more of the components of the Cas system are delivered. In an embodiment, the nucleic acid molecule is delivered by a different means than one or more of the components of the Cas system, e.g., the Cas9 molecule component and/or the gRNA molecule component, are delivered. The nucleic acid molecule can be delivered by any of the delivery methods described herein. For example, the nucleic acid molecule can be delivered by a viral vector, e.g., an integration-deficient lentivirus, and the Cas9 molecule component and/or the gRNA molecule component can be delivered by electroporation, e.g., such that the toxicity caused by nucleic acids (e.g., DNAs) can be reduced. In an embodiment, the nucleic acid molecule encodes a therapeutic protein, e.g., a protein described herein. In an embodiment, the nucleic acid molecule encodes an RNA molecule, e.g., an RNA molecule described herein.

Delivery of RNA encoding a Cas9 molecule

RNA encoding Cas9 molecules (e.g., eaCas9 molecules, eiCas9 molecules or eiCas9 fusion proteins) and/or gRNA molecules, can be delivered into cells by art-known methods or as described herein. For example, Cas9-encoding and/or gRNA-encoding RNA can be delivered, e.g., by microinjection, electroporation, lipid-mediated transfection, peptide-mediated delivery,

or a combination thereof. Cas9-encoding and/or gRNA-encoding RNA can be conjugated to molecules (e.g., GalNAc) promoting uptake by the target cells (e.g., target cells described herein).

5 Delivery Cas9 molecule protein

Cas9 molecules (e.g., eaCas9 molecules, eiCas9 molecules or eiCas9 fusion proteins) can be delivered into cells by art-known methods or as described herein. For example, Cas9 protein molecules can be delivered, e.g., by microinjection, electroporation, lipid-mediated transfection, peptide-mediated delivery, or a combination thereof. Delivery can be accompanied by DNA
10 encoding a gRNA or by a gRNA. Delivery can be accompanied by a donor template. Cas9 protein can be conjugated to molecules (e.g., GalNAc) promoting uptake by the target cells (e.g., target cells described herein).

Route of administration

15 Systemic modes of administration include oral and parenteral routes. Parenteral routes include, by way of example, intravenous, intrarterial, intramuscular, intradermal, subcutaneous, intranasal, and intraperitoneal routes. Components administered systemically may be modified or formulated to target the components to the eye or inner ear.

Local modes of administration include, by way of example, intraocular, intraorbital,
20 subconjunctival, intravitreal, subretinal, transscleral or introcochlear routes. In an embodiment, significantly smaller amounts of the components (compared with systemic approaches) may exert an effect when administered locally (for example, intravitreally) compared to when administered systemically (for example, intravenously). Local modes of administration can reduce or eliminate the incidence of potentially toxic side effects that may occur when
25 therapeutically effective amounts of a component are administered systemically.

In an embodiment, components described herein are delivered subretinally, e.g., by subretinal injection. Subretinal injections may be made directly into the macular, e.g., submacular injection.

In an embodiment, components described herein are delivered by intravitreal injection.
30 Intravitreal injection has a relatively low risk of retinal detachment. In an embodiment, nanoparticle or viral, e.g., AAV vector, is delivered intravitreally.

In an embodiment, components described herein are delivered into the inner ear, e.g., by intracochlear injection. Intracochlear injections may be made in the vicinity of inner and/or outer hair cells.

Methods for administration of agents to the eye and inner ear are known in the medical arts and can be used to administer components described herein. Exemplary methods include intraocular injection (e.g., retrobulbar, subretinal, submacular, intravitreal and intrachoroidal), iontophoresis, eye drops, intraocular implantation (e.g., intravitreal, sub-Tenons and sub-conjunctival) and intracochlear injection.

Administration may be provided as a periodic bolus (for example, subretinally, intravenously, intravitreally or by intracochlear injection) or as continuous infusion from an internal reservoir (for example, from an implant disposed at an intra- or extra-ocular location (see, U.S. Pat. Nos. 5,443,505 and 5,766,242)) or from an external reservoir (for example, from an intravenous bag). Components may be administered locally, for example, by continuous release from a sustained release drug delivery device immobilized to an inner wall of the eye or via targeted transscleral controlled release into the choroid (see, for example, PCT/US00/00207, PCT/US02/14279, Ambati et al. (2000) INVEST. OPHTHALMOL. VIS. SCI.41:1181-1185, and Ambati et al. (2000) INVEST. OPHTHALMOL. VIS. SCI.41:1186-1191). A variety of devices suitable for administering components locally to the inside of the eye are known in the art. See, for example, U.S. Pat. Nos. 6,251,090, 6,299,895, 6,416,777, 6,413,540, and PCT/US00/28187.

In addition, components may be formulated to permit release over a prolonged period of time. A release system can include a matrix of a biodegradable material or a material which releases the incorporated components by diffusion. The components can be homogeneously or heterogeneously distributed within the release system. A variety of release systems may be useful, however, the choice of the appropriate system will depend upon rate of release required by a particular application. Both non-degradable and degradable release systems can be used. Suitable release systems include polymers and polymeric matrices, non-polymeric matrices, or inorganic and organic excipients and diluents such as, but not limited to, calcium carbonate and sugar (for example, trehalose). Release systems may be natural or synthetic. However, synthetic release systems are preferred because generally they are more reliable, more reproducible and produce more defined release profiles. The release system material can be selected so that components having different molecular weights are released by diffusion through or degradation

of the material.

Representative synthetic, biodegradable polymers include, for example: polyamides such as poly(amino acids) and poly(peptides); polyesters such as poly(lactic acid), poly(glycolic acid), poly(lactic-co-glycolic acid), and poly(caprolactone); poly(anhydrides); polyorthoesters;

5 polycarbonates; and chemical derivatives thereof (substitutions, additions of chemical groups, for example, alkyl, alkylene, hydroxylations, oxidations, and other modifications routinely made by those skilled in the art), copolymers and mixtures thereof. Representative synthetic, non-

degradable polymers include, for example: polyethers such as poly(ethylene oxide), poly(ethylene glycol), and poly(tetramethylene oxide); vinyl polymers-polyacrylates and

10 polymethacrylates such as methyl, ethyl, other alkyl, hydroxyethyl methacrylate, acrylic and methacrylic acids, and others such as poly(vinyl alcohol), poly(vinyl pyrrolidone), and poly(vinyl acetate); poly(urethanes); cellulose and its derivatives such as alkyl, hydroxyalkyl, ethers, esters, nitrocellulose, and various cellulose acetates; polysiloxanes; and any chemical derivatives thereof (substitutions, additions of chemical groups, for example, alkyl, alkylene, hydroxylations, 15 oxidations, and other modifications routinely made by those skilled in the art), copolymers and mixtures thereof.

Poly(lactide-co-glycolide) microsphere can also be used for intraocular injection. Typically the microspheres are composed of a polymer of lactic acid and glycolic acid, which are structured to form hollow spheres. The spheres can be approximately 15-30 microns in diameter 20 and can be loaded with components described herein.

Bi-Modal or Differential Delivery of Components

Separate delivery of the components of a Cas system, e.g., the Cas9 molecule component and the gRNA molecule component, and more particularly, delivery of the components by differing modes, can enhance performance, e.g., by improving tissue specificity and safety.

25 In an embodiment, the Cas9 molecule and the gRNA molecule are delivered by different modes, or as sometimes referred to herein as differential modes. Different or differential modes, as used herein, refer modes of delivery that confer different pharmacodynamic or pharmacokinetic properties on the subject component molecule, e.g., a Cas9 molecule, gRNA molecule, template nucleic acid, or payload. For example, the modes of delivery can result in 30 different tissue distribution, different half-life, or different temporal distribution, e.g., in a selected compartment, tissue, or organ.

Some modes of delivery, e.g., delivery by a nucleic acid vector that persists in a cell, or in progeny of a cell, e.g., by autonomous replication or insertion into cellular nucleic acid, result in more persistent expression of and presence of a component. Examples include viral, e.g., adeno associated virus or lentivirus, delivery.

5 By way of example, the components, e.g., a Cas9 molecule and a gRNA molecule, can be delivered by modes that differ in terms of resulting half-life or persistence of the delivered component in the body, or in a particular compartment, tissue or organ. In an embodiment, a gRNA molecule can be delivered by such modes. The Cas9 molecule component can be delivered by a mode which results in less persistence or less exposure to the body or a particular
10 compartment or tissue or organ.

More generally, in an embodiment, a first mode of delivery is used to deliver a first component and a second mode of delivery is used to deliver a second component. The first mode of delivery confers a first pharmacodynamic or pharmacokinetic property. The first pharmacodynamic property can be, e.g., distribution, persistence, or exposure, of the component,
15 or of a nucleic acid that encodes the component, in the body, a compartment, tissue or organ. The second mode of delivery confers a second pharmacodynamic or pharmacokinetic property. The second pharmacodynamic property can be, e.g., distribution, persistence, or exposure, of the component, or of a nucleic acid that encodes the component, in the body, a compartment, tissue or organ.

20 In an embodiment, the first pharmacodynamic or pharmacokinetic property, e.g., distribution, persistence or exposure, is more limited than the second pharmacodynamic or pharmacokinetic property.

In an embodiment, the first mode of delivery is selected to optimize, e.g., minimize, a pharmacodynamic or pharmacokinetic property, e.g., distribution, persistence or exposure.

25 In an embodiment, the second mode of delivery is selected to optimize, e.g., maximize, a pharmacodynamic or pharmacokinetic property, e.g., distribution, persistence or exposure.

In an embodiment, the first mode of delivery comprises the use of a relatively persistent element, e.g., a nucleic acid, e.g., a plasmid or viral vector, e.g., an AAV or lentivirus. As such vectors are relatively persistent product transcribed from them would be relatively persistent.

30 In an embodiment, the second mode of delivery comprises a relatively transient element, e.g., an RNA or protein.

In an embodiment, the first component comprises gRNA, and the delivery mode is relatively persistent, e.g., the gRNA is transcribed from a plasmid or viral vector, e.g., an AAV or lentivirus. Transcription of these genes would be of little physiological consequence because the genes do not encode for a protein product, and the gRNAs are incapable of acting in isolation. The second component, a Cas9 molecule, is delivered in a transient manner, for example as mRNA or as protein, ensuring that the full Cas9 molecule/gRNA molecule complex is only present and active for a short period of time.

Furthermore, the components can be delivered in different molecular form or with different delivery vectors that complement one another to enhance safety and tissue specificity.

Use of differential delivery modes can enhance performance, safety and efficacy. E.g., the likelihood of an eventual off-target modification can be reduced. Delivery of immunogenic components, e.g., Cas9 molecules, by less persistent modes can reduce immunogenicity, as peptides from the bacterially-derived Cas enzyme are displayed on the surface of the cell by MHC molecules. A two-part delivery system can alleviate these drawbacks.

Differential delivery modes can be used to deliver components to different, but overlapping target regions. The formation active complex is minimized outside the overlap of the target regions. Thus, in an embodiment, a first component, e.g., a gRNA molecule is delivered by a first delivery mode that results in a first spatial, e.g., tissue, distribution. A second component, e.g., a Cas9 molecule is delivered by a second delivery mode that results in a second spatial, e.g., tissue, distribution. In an embodiment the first mode comprises a first element selected from a liposome, nanoparticle, e.g., polymeric nanoparticle, and a nucleic acid, e.g., viral vector. The second mode comprises a second element selected from the group. In an embodiment, the first mode of delivery comprises a first targeting element, e.g., a cell specific receptor or an antibody, and the second mode of delivery does not include that element. In an embodiment, the second mode of delivery comprises a second targeting element, e.g., a second cell specific receptor or second antibody.

When the Cas9 molecule is delivered in a virus delivery vector, a liposome, or polymeric nanoparticle, there is the potential for delivery to and therapeutic activity in multiple tissues, when it may be desirable to only target a single tissue. A two-part delivery system can resolve this challenge and enhance tissue specificity. If the gRNA molecule and the Cas9 molecule are

packaged in separated delivery vehicles with distinct but overlapping tissue tropism, the fully functional complex is only be formed in the tissue that is targeted by both vectors.

Ex vivo delivery

In some embodiments, components described in **Table 14** are introduced into cells which are then introduced into the subject. Methods of introducing the components can include, e.g., any of the delivery methods described in **Table 15**.

VIII. Modified Nucleosides, Nucleotides, and Nucleic Acids

Modified nucleosides and modified nucleotides can be present in nucleic acids, e.g., particularly gRNA, but also other forms of RNA, e.g., mRNA, RNAi, or siRNA. As described herein, “nucleoside” is defined as a compound containing a five-carbon sugar molecule (a pentose or ribose) or derivative thereof, and an organic base, purine or pyrimidine, or a derivative thereof. As described herein, “nucleotide” is defined as a nucleoside further comprising a phosphate group.

Modified nucleosides and nucleotides can include one or more of:

- (i) alteration, e.g., replacement, of one or both of the non-linking phosphate oxygens and/or of one or more of the linking phosphate oxygens in the phosphodiester backbone linkage;
- (ii) alteration, e.g., replacement, of a constituent of the ribose sugar, e.g., of the 2' hydroxyl on the ribose sugar;
- (iii) wholesale replacement of the phosphate moiety with “dephospho” linkers;
- (iv) modification or replacement of a naturally occurring nucleobase;
- (v) replacement or modification of the ribose-phosphate backbone;
- (vi) modification of the 3' end or 5' end of the oligonucleotide, e.g., removal, modification or replacement of a terminal phosphate group or conjugation of a moiety; and
- (vii) modification of the sugar.

The modifications listed above can be combined to provide modified nucleosides and nucleotides that can have two, three, four, or more modifications. For example, a modified nucleoside or nucleotide can have a modified sugar and a modified nucleobase. In an embodiment, every base of a gRNA is modified, e.g., all bases have a modified phosphate group, e.g., all are phosphorothioate groups. In an embodiment, all, or substantially all, of the

phosphate groups of a unimolecular or modular gRNA molecule are replaced with phosphorothioate groups.

In an embodiment, modified nucleotides, e.g., nucleotides having modifications as described herein, can be incorporated into a nucleic acid, e.g., a “modified nucleic acid.” In some embodiments, the modified nucleic acids comprise one, two, three or more modified nucleotides. In some embodiments, at least 5% (e.g., at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or about 100%) of the positions in a modified nucleic acid are a modified nucleotides.

Unmodified nucleic acids can be prone to degradation by, e.g., cellular nucleases. For example, nucleases can hydrolyze nucleic acid phosphodiester bonds. Accordingly, in one aspect the modified nucleic acids described herein can contain one or more modified nucleosides or nucleotides, e.g., to introduce stability toward nucleases.

In some embodiments, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can exhibit a reduced innate immune response when introduced into a population of cells, both *in vivo* and *ex vivo*. The term “innate immune response” includes a cellular response to exogenous nucleic acids, including single stranded nucleic acids, generally of viral or bacterial origin, which involves the induction of cytokine expression and release, particularly the interferons, and cell death. In some embodiments, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can disrupt binding of a major groove interacting partner with the nucleic acid. In some embodiments, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can exhibit a reduced innate immune response when introduced into a population of cells, both *in vivo* and *ex vivo*, and also disrupt binding of a major groove interacting partner with the nucleic acid.

Definitions of Chemical Groups

As used herein, “alkyl” is meant to refer to a saturated hydrocarbon group which is straight-chained or branched. Example alkyl groups include methyl (Me), ethyl (Et), propyl (e.g., n-propyl and isopropyl), butyl (e.g., n-butyl, isobutyl, t-butyl), pentyl (e.g., n-pentyl, isopentyl, neopentyl), and the like. An alkyl group can contain from 1 to about 20, from 2 to

about 20, from 1 to about 12, from 1 to about 8, from 1 to about 6, from 1 to about 4, or from 1 to about 3 carbon atoms.

As used herein, “aryl” refers to monocyclic or polycyclic (*e.g.*, having 2, 3 or 4 fused rings) aromatic hydrocarbons such as, for example, phenyl, naphthyl, anthracenyl, phenanthrenyl, indanyl, indenyl, and the like. In some embodiments, aryl groups have from 6 to about 20 carbon atoms.

As used herein, “alkenyl” refers to an aliphatic group containing at least one double bond.

As used herein, “alkynyl” refers to a straight or branched hydrocarbon chain containing 2-12 carbon atoms and characterized in having one or more triple bonds. Examples of alkynyl groups include, but are not limited to, ethynyl, propargyl, and 3-hexynyl.

As used herein, “arylalkyl” or “aralkyl” refers to an alkyl moiety in which an alkyl hydrogen atom is replaced by an aryl group. Aralkyl includes groups in which more than one hydrogen atom has been replaced by an aryl group. Examples of “arylalkyl” or “aralkyl” include benzyl, 2-phenylethyl, 3-phenylpropyl, 9-fluorenyl, benzhydryl, and trityl groups.

As used herein, “cycloalkyl” refers to a cyclic, bicyclic, tricyclic, or polycyclic non-aromatic hydrocarbon groups having 3 to 12 carbons. Examples of cycloalkyl moieties include, but are not limited to, cyclopropyl, cyclopentyl, and cyclohexyl.

As used herein, “heterocyclyl” refers to a monovalent radical of a heterocyclic ring system. Representative heterocyclyls include, without limitation, tetrahydrofuranyl, tetrahydrothienyl, pyrrolidinyl, pyrrolidonyl, piperidinyl, pyrrolinyl, piperazinyl, dioxanyl, dioxolanyl, diazepinyl, oxazepinyl, thiazepinyl, and morpholinyl.

As used herein, “heteroaryl” refers to a monovalent radical of a heteroaromatic ring system. Examples of heteroaryl moieties include, but are not limited to, imidazolyl, oxazolyl, thiazolyl, triazolyl, pyrrolyl, furanyl, indolyl, thiophenyl pyrazolyl, pyridinyl, pyrazinyl, pyridazinyl, pyrimidinyl, indolizinyl, purinyl, naphthyridinyl, quinolyl, and pteridinyl.

Phosphate Backbone Modifications

The Phosphate Group

In some embodiments, the phosphate group of a modified nucleotide can be modified by replacing one or more of the oxygens with a different substituent. Further, the modified nucleotide, *e.g.*, modified nucleotide present in a modified nucleic acid, can include the

wholesale replacement of an unmodified phosphate moiety with a modified phosphate as described herein. In some embodiments, the modification of the phosphate backbone can include alterations that result in either an uncharged linker or a charged linker with unsymmetrical charge distribution.

5 Examples of modified phosphate groups include, phosphorothioate, phosphoroselenates, borano phosphates, borano phosphate esters, hydrogen phosphonates, phosphoroamidates, alkyl or aryl phosphonates and phosphotriesters. In some embodiments, one of the non-bridging phosphate oxygen atoms in the phosphate backbone moiety can be replaced by any of the following groups: sulfur (S), selenium (Se), BR_3 (wherein R can be, e.g., hydrogen, alkyl, or
10 aryl), C (e.g., an alkyl group, an aryl group, and the like), H, NR_2 (wherein R can be, e.g., hydrogen, alkyl, or aryl), or OR (wherein R can be, e.g., alkyl or aryl). The phosphorous atom in an unmodified phosphate group is achiral. However, replacement of one of the non-bridging oxygens with one of the above atoms or groups of atoms can render the phosphorous atom chiral; that is to say that a phosphorous atom in a phosphate group modified in this way is a stereogenic
15 center. The stereogenic phosphorous atom can possess either the “R” configuration (herein R_p) or the “S” configuration (herein S_p).

Phosphorodithioates have both non-bridging oxygens replaced by sulfur. The phosphorus center in the phosphorodithioates is achiral which precludes the formation of oligoribonucleotide diastereomers. In some embodiments, modifications to one or both non-bridging oxygens can
20 also include the replacement of the non-bridging oxygens with a group independently selected from S, Se, B, C, H, N, and OR (R can be, e.g., alkyl or aryl).

The phosphate linker can also be modified by replacement of a bridging oxygen, (i.e., the oxygen that links the phosphate to the nucleoside), with nitrogen (bridged phosphoroamidates), sulfur (bridged phosphorothioates) and carbon (bridged methylenephosphonates). The
25 replacement can occur at either linking oxygen or at both of the linking oxygens.

Replacement of the Phosphate Group

The phosphate group can be replaced by non-phosphorus containing connectors. In some embodiments, the charged phosphate group can be replaced by a neutral moiety.

30 Examples of moieties which can replace the phosphate group can include, without limitation, e.g., methyl phosphonate, hydroxylamino, siloxane, carbonate, carboxymethyl, carbamate, amide, thioether, ethylene oxide linker, sulfonate, sulfonamide, thioformacetal,

formacetal, oxime, methyleneimino, methylenemethylimino, methylenehydrazo, methylenedimethylhydrazo and methyleneoxymethylimino.

Replacement of the Ribophosphate Backbone

Scaffolds that can mimic nucleic acids can also be constructed wherein the phosphate linker and ribose sugar are replaced by nuclease resistant nucleoside or nucleotide surrogates. In some embodiments, the nucleobases can be tethered by a surrogate backbone. Examples can include, without limitation, the morpholino, cyclobutyl, pyrrolidine and peptide nucleic acid (PNA) nucleoside surrogates.

Sugar Modifications

The modified nucleosides and modified nucleotides can include one or more modifications to the sugar group. For example, the 2' hydroxyl group (OH) can be modified or replaced with a number of different "oxy" or "deoxy" substituents. In some embodiments, modifications to the 2' hydroxyl group can enhance the stability of the nucleic acid since the hydroxyl can no longer be deprotonated to form a 2'-alkoxide ion. The 2'-alkoxide can catalyze degradation by intramolecular nucleophilic attack on the linker phosphorus atom.

Examples of "oxy"-2' hydroxyl group modifications can include alkoxy or aryloxy (OR, wherein "R" can be, e.g., alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or a sugar); polyethyleneglycols (PEG), $O(CH_2CH_2O)_nCH_2CH_2OR$ wherein R can be, e.g., H or optionally substituted alkyl, and n can be an integer from 0 to 20 (e.g., from 0 to 4, from 0 to 8, from 0 to 10, from 0 to 16, from 1 to 4, from 1 to 8, from 1 to 10, from 1 to 16, from 1 to 20, from 2 to 4, from 2 to 8, from 2 to 10, from 2 to 16, from 2 to 20, from 4 to 8, from 4 to 10, from 4 to 16, and from 4 to 20). In some embodiments, the "oxy"-2' hydroxyl group modification can include "locked" nucleic acids (LNA) in which the 2' hydroxyl can be connected, e.g., by a C_{1-6} alkylene or C_{1-6} heteroalkylene bridge, to the 4' carbon of the same ribose sugar, where exemplary bridges can include methylene, propylene, ether, or amino bridges; O-amino (wherein amino can be, e.g., NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino) and aminoalkoxy, $O(CH_2)_n$ -amino, (wherein amino can be, e.g., NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino). In some embodiments, the "oxy"-2' hydroxyl group modification can include the methoxyethyl group (MOE), $(OCH_2CH_2OCH_3)$, e.g., a PEG derivative).

“Deoxy” modifications can include hydrogen (i.e. deoxyribose sugars, e.g., at the overhang portions of partially ds RNA); halo (e.g., bromo, chloro, fluoro, or iodo); amino (wherein amino can be, e.g., NH_2 ; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroaryl amino, diheteroaryl amino, or amino acid); $\text{NH}(\text{CH}_2\text{CH}_2\text{NH})_n\text{CH}_2\text{CH}_2$ -
 5 amino (wherein amino can be, e.g., as described herein), $-\text{NHC}(\text{O})\text{R}$ (wherein R can be, e.g., alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar), cyano; mercapto; alkyl-thio-alkyl; thioalkoxy; and alkyl, cycloalkyl, aryl, alkenyl and alkynyl, which may be optionally substituted with e.g., an amino as described herein.

The sugar group can also contain one or more carbons that possess the opposite
 10 stereochemical configuration than that of the corresponding carbon in ribose. Thus, a modified nucleic acid can include nucleotides containing e.g., arabinose, as the sugar. The nucleotide “monomer” can have an alpha linkage at the 1' position on the sugar, e.g., alpha-nucleosides. The modified nucleic acids can also include “abasic” sugars, which lack a nucleobase at C-1'. These abasic sugars can also be further modified at one or more of the constituent sugar atoms.
 15 The modified nucleic acids can also include one or more sugars that are in the L form, e.g. L-nucleosides.

Generally, RNA includes the sugar group ribose, which is a 5-membered ring having an oxygen. Exemplary modified nucleosides and modified nucleotides can include, without
 20 limitation, replacement of the oxygen in ribose (e.g., with sulfur (S), selenium (Se), or alkylene, such as, e.g., methylene or ethylene); addition of a double bond (e.g., to replace ribose with cyclopentenyl or cyclohexenyl); ring contraction of ribose (e.g., to form a 4-membered ring of cyclobutane or oxetane); ring expansion of ribose (e.g., to form a 6- or 7-membered ring having an additional carbon or heteroatom, such as for example, anhydrohexitol, altritol, mannitol, cyclohexanyl, cyclohexenyl, and morpholino that also has a phosphoramidate backbone). In
 25 some embodiments, the modified nucleotides can include multicyclic forms (e.g., tricyclo; and “unlocked” forms, such as glycol nucleic acid (GNA) (e.g., R-GNA or S-GNA, where ribose is replaced by glycol units attached to phosphodiester bonds), threose nucleic acid (TNA, where ribose is replaced with α -L-threofuranosyl-(3' \rightarrow 2')).

Modifications on the Nucleobase

30 The modified nucleosides and modified nucleotides described herein, which can be incorporated into a modified nucleic acid, can include a modified nucleobase. Examples of

nucleobases include, but are not limited to, adenine (A), guanine (G), cytosine (C), and uracil (U). These nucleobases can be modified or wholly replaced to provide modified nucleosides and modified nucleotides that can be incorporated into modified nucleic acids. The nucleobase of the nucleotide can be independently selected from a purine, a pyrimidine, a purine or pyrimidine analog. In some embodiments, the nucleobase can include, for example, naturally-occurring and synthetic derivatives of a base.

Uracil

In some embodiments, the modified nucleobase is a modified uracil. Exemplary nucleobases and nucleosides having a modified uracil include without limitation pseudouridine (ψ), pyridin-4-one ribonucleoside, 5-aza-uridine, 6-aza-uridine, 2-thio-5-aza-uridine, 2-thio-uridine ($s2U$), 4-thio-uridine ($s4U$), 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxy-uridine (ho^5U), 5-aminoallyl-uridine, 5-halo-uridine (e.g., 5-iodo-uridine or 5-bromo-uridine), 3-methyl-uridine (m^3U), 5-methoxy-uridine (mo^5U), uridine 5-oxyacetic acid (cmo^5U), uridine 5-oxyacetic acid methyl ester ($mcmo^5U$), 5-carboxymethyl-uridine (cm^5U), 1-carboxymethyl-pseudouridine, 5-carboxyhydroxymethyl-uridine (chm^5U), 5-carboxyhydroxymethyl-uridine methyl ester ($mchm^5U$), 5-methoxycarbonylmethyl-uridine (mcm^5U), 5-methoxycarbonylmethyl-2-thio-uridine (mcm^5s2U), 5-aminomethyl-2-thio-uridine (nm^5s2U), 5-methylaminomethyl-uridine (mnm^5U), 5-methylaminomethyl-2-thio-uridine (mnm^5s2U), 5-methylaminomethyl-2-seleno-uridine (mnm^5se^2U), 5-carbamoylmethyl-uridine (ncm^5U), 5-carboxymethylaminomethyl-uridine ($cmnm^5U$), 5-carboxymethylaminomethyl-2-thio-uridine ($cmnm^5s2U$), 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-aurinomethyl-uridine (τcm^5U), 1-aurinomethyl-pseudouridine, 5-aurinomethyl-2-thio-uridine (τm^5s2U), 1-aurinomethyl-4-thio-pseudouridine, 5-methyl-uridine (m^5U , i.e., having the nucleobase deoxythymine), 1-methyl-pseudouridine ($m^1\psi$), 5-methyl-2-thio-uridine (m^5s2U), 1-methyl-4-thio-pseudouridine ($m^1s^4\psi$), 4-thio-1-methyl-pseudouridine, 3-methyl-pseudouridine ($m^3\psi$), 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine (D), dihydropseudouridine, 5,6-dihydrouridine, 5-methyl-dihydrouridine (m^5D), 2-thio-dihydrouridine, 2-thio-dihydropseudouridine, 2-methoxy-uridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, 4-methoxy-2-thio-pseudouridine, N1-methyl-pseudouridine, 3-(3-amino-3-carboxypropyl)uridine (acp^3U), 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine ($acp^3\psi$), 5-(isopentenylaminomethyl)uridine (inm^5U), 5-

(isopentenylaminomethyl)-2-thio-uridine (inm⁵s2U), α -thio-uridine, 2'-O-methyl-uridine (Um), 5,2'-O-dimethyl-uridine (m⁵Um), 2'-O-methyl-pseudouridine (ψ m), 2-thio-2'-O-methyl-uridine (s2Um), 5-methoxycarbonylmethyl-2'-O-methyl-uridine (mcm⁵Um), 5-carbamoylmethyl-2'-O-methyl-uridine (ncm⁵Um), 5-carboxymethylaminomethyl-2'-O-methyl-uridine (cmnm⁵Um),
 5 3,2'-O-dimethyl-uridine (m³Um), 5-(isopentenylaminomethyl)-2'-O-methyl-uridine (inm⁵Um), 1-thio-uridine, deoxythymidine, 2'-F-ara-uridine, 2'-F-uridine, 2'-OH-ara-uridine, 5-(2-carbomethoxyvinyl) uridine, 5-[3-(1-E-propenylamino)uridine, pyrazolo[3,4-d]pyrimidines, xanthine, and hypoxanthine.

Cytosine

10 In some embodiments, the modified nucleobase is a modified cytosine. Exemplary nucleobases and nucleosides having a modified cytosine include without limitation 5-aza-cytidine, 6-aza-cytidine, pseudoisocytidine, 3-methyl-cytidine (m³C), N4-acetyl-cytidine (act), 5-formyl-cytidine (f⁵C), N4-methyl-cytidine (m⁴C), 5-methyl-cytidine (m⁵C), 5-halo-cytidine (e.g., 5-iodo-cytidine), 5-hydroxymethyl-cytidine (hm⁵C), 1-methyl-pseudoisocytidine, pyrrolo-
 15 cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine (s2C), 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxy-cytidine, 2-methoxy-5-methyl-cytidine, 4-methoxy-pseudoisocytidine, 4-methoxy-1-methyl-pseudoisocytidine, lysidine (k²C),
 20 α -thio-cytidine, 2'-O-methyl-cytidine (Cm), 5,2'-O-dimethyl-cytidine (m⁵Cm), N4-acetyl-2'-O-methyl-cytidine (ac⁴Cm), N4,2'-O-dimethyl-cytidine (m⁴Cm), 5-formyl-2'-O-methyl-cytidine (f⁵Cm), N4,N4,2'-O-trimethyl-cytidine (m⁴₂Cm), 1-thio-cytidine, 2'-F-ara-cytidine, 2'-F-cytidine, and 2'-OH-ara-cytidine.

Adenine

25 In some embodiments, the modified nucleobase is a modified adenine. Exemplary nucleobases and nucleosides having a modified adenine include without limitation 2-amino-purine, 2,6-diaminopurine, 2-amino-6-halo-purine (e.g., 2-amino-6-chloro-purine), 6-halo-purine (e.g., 6-chloro-purine), 2-amino-6-methyl-purine, 8-azido-adenosine, 7-deaza-adenine, 7-deaza-8-aza-adenine, 7-deaza-2-amino-purine, 7-deaza-8-aza-2-amino-purine, 7-deaza-2,6-
 30 diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyl-adenosine (m¹A), 2-methyl-adenine (m²A), N6-methyl-adenosine (m⁶A), 2-methylthio-N6-methyl-adenosine (ms2m⁶A), N6-

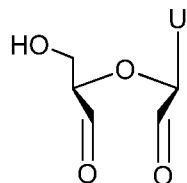
isopentenyl-adenosine (i^6A), 2-methylthio-N6-isopentenyl-adenosine (ms^2i^6A), N6-(cis-hydroxyisopentenyl)adenosine (io^6A), 2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine ($ms2io^6A$), N6-glycinylicarbamoyl-adenosine (g^6A), N6-threonylicarbamoyl-adenosine (t^6A), N6-methyl-N6-threonylicarbamoyl-adenosine (m^6t^6A), 2-methylthio-N6-threonylicarbamoyl-adenosine (ms^2g^6A), N6,N6-dimethyl-adenosine (m^6_2A), N6-hydroxynorvalylcarbamoyl-adenosine (hn^6A), 2-methylthio-N6-hydroxynorvalylcarbamoyl-adenosine ($ms2hn^6A$), N6-acetyl-adenosine (ac^6A), 7-methyl-adenine, 2-methylthio-adenine, 2-methoxy-adenine, α -thio-adenosine, 2'-O-methyl-adenosine (Am), N⁶,2'-O-dimethyl-adenosine (m^6Am), N⁶-Methyl-2'-deoxyadenosine, N6,N6,2'-O-trimethyl-adenosine (m^6_2Am), 1,2'-O-dimethyl-adenosine (m^1Am), 2'-O-ribosyladenosine (phosphate) (Ar(p)), 2-amino-N6-methyl-purine, 1-thio-adenosine, 8-azido-adenosine, 2'-F-ara-adenosine, 2'-F-adenosine, 2'-OH-ara-adenosine, and N6-(19-amino-pentaoxanonadecyl)-adenosine.

Guanine

In some embodiments, the modified nucleobase is a modified guanine. Exemplary nucleobases and nucleosides having a modified guanine include without limitation inosine (I), 1-methyl-inosine (m^1I), wyosine (imG), methylwyosine (mimG), 4-demethyl-wyosine (imG-14), isowyosine (imG2), wybutosine (yW), peroxywybutosine (o_2yW), hydroxywybutosine (OH_yW), undermodified hydroxywybutosine (OH_yW*), 7-deaza-guanosine, queuosine (Q), epoxyqueuosine (oQ), galactosyl-queuosine (galQ), mannosyl-queuosine (manQ), 7-cyano-7-deaza-guanosine (preQ₀), 7-aminomethyl-7-deaza-guanosine (preQ₁), archaeosine (G^+), 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine (m^7G), 6-thio-7-methyl-guanosine, 7-methyl-inosine, 6-methoxy-guanosine, 1-methyl-guanosine (m^1G), N2-methyl-guanosine (m^2G), N2,N2-dimethyl-guanosine (m^2_2G), N2,7-dimethyl-guanosine ($m^2,7G$), N2, N2,7-dimethyl-guanosine ($m^2,2,7G$), 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-meth thio-guanosine, N2-methyl-6-thio-guanosine, N2,N2-dimethyl-6-thio-guanosine, α -thio-guanosine, 2'-O-methyl-guanosine (Gm), N2-methyl-2'-O-methyl-guanosine (m^2Gm), N2,N2-dimethyl-2'-O-methyl-guanosine (m^2_2Gm), 1-methyl-2'-O-methyl-guanosine (m^1Gm), N2,7-dimethyl-2'-O-methyl-guanosine ($m^2,7Gm$), 2'-O-methyl-inosine (Im), 1,2'-O-dimethyl-inosine (m^1Im), O⁶-phenyl-2'-deoxyinosine, 2'-O-ribosylguanosine (phosphate) (Gr(p)), 1-thio-guanosine, O⁶-methyl-guanosine, O⁶-Methyl-2'-deoxyguanosine, 2'-F-ara-guanosine, and 2'-F-guanosine.

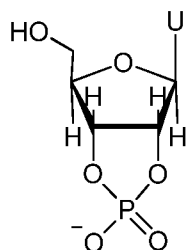
Modified gRNAs

In some embodiments, the modified nucleic acids can be modified gRNAs. In some
 embodiments, gRNAs can be modified at the 3' end. In this embodiment, the gRNAs can be
 modified at the 3' terminal U ribose. For example, the two terminal hydroxyl groups of the U
 5 ribose can be oxidized to aldehyde groups and a concomitant opening of the ribose ring to afford
 a modified nucleoside as shown below:



wherein "U" can be an unmodified or modified uridine.

In another embodiment, the 3' terminal U can be modified with a 2'3' cyclic phosphate
 10 as shown below:



wherein "U" can be an unmodified or modified uridine.

In some embodiments, the gRNA molecules may contain 3' nucleotides which can be
 stabilized against degradation, e.g., by incorporating one or more of the modified nucleotides
 15 described herein. In this embodiment, e.g., uridines can be replaced with modified uridines, e.g.,
 5-(2-amino)propyl uridine, and 5-bromo uridine, or with any of the modified uridines described
 herein; adenosines and guanosines can be replaced with modified adenosines and guanosines,
 e.g., with modifications at the 8-position, e.g., 8-bromo guanosine, or with any of the modified
 adenosines or guanosines described herein. In some embodiments, deaza nucleotides, e.g., 7-
 20 deaza-adenosine, can be incorporated into the gRNA. In some embodiments, O- and N-alkylated
 nucleotides, e.g., N6-methyl adenosine, can be incorporated into the gRNA. In some
 embodiments, sugar-modified ribonucleotides can be incorporated, e.g., wherein the 2' OH-
 group is replaced by a group selected from H, -OR, -R (wherein R can be, e.g., alkyl, cycloalkyl,
 aryl, aralkyl, heteroaryl or sugar), halo, -SH, -SR (wherein R can be, e.g., alkyl, cycloalkyl, aryl,
 25 aralkyl, heteroaryl or sugar), amino (wherein amino can be, e.g., NH₂; alkylamino, dialkylamino,

heterocyclyl, arylamino, diarylamino, heteroaryl amino, diheteroaryl amino, or amino acid); or cyano (-CN). In some embodiments, the phosphate backbone can be modified as described herein, e.g., with a phosphothioate group. In some embodiments, the nucleotides in the overhang region of the gRNA can each independently be a modified or unmodified nucleotide including, but not limited to 2'-sugar modified, such as, 2-F 2'-O-methyl, thymidine (T), 2'-O-methoxyethyl-5-methyluridine (Teo), 2'-O-methoxyethyladenosine (Aeo), 2'-O-methoxyethyl-5-methylcytidine (m5Ceo), and any combinations thereof.

In an embodiment, one or more or all of the nucleotides in single stranded RNA molecule, e.g., a gRNA molecule, are deoxynucleotides.

miRNA binding sites

microRNAs (or miRNAs) are naturally occurring cellular 19-25 nucleotide long noncoding RNAs. They bind to nucleic acid molecules having an appropriate miRNA binding site, e.g., in the 3' UTR of a mRNA, and down-regulate gene expression. While not wishing to be bound by theory it is believed that the down regulation is either by reducing nucleic acid molecule stability or by inhibiting translation. An RNA species disclosed herein, e.g., an mRNA encoding Cas9 can comprise an miRNA binding site, e.g., in its 3'UTR. The miRNA binding site can be selected to promote down regulation of expression in a selected cell type. By way of example, the incorporation of a binding site for miR-122, a microRNA abundant in liver, can inhibit the expression of the gene of interest in the liver.

Examples

The following Examples are merely illustrative and are not intended to limit the scope or content of the invention in any way.

5 Example 1: Evaluation of candidate guide RNAs

The suitability of candidate gRNAs can be evaluated as described in this example. Although described for a chimeric gRNA, the approach can also be used to evaluate modular gRNAs.

Cloning gRNAs into Vectors

10 For each gRNA, a pair of overlapping oligonucleotides is designed and obtained. Oligonucleotides are annealed and ligated into a digested vector backbone containing an upstream U6 promoter and the remaining sequence of a long chimeric gRNA. Plasmid is sequence-verified and prepped to generate sufficient amounts of transfection-quality DNA. Alternate promoters may be used to drive in vivo transcription (e.g., H1 promoter) or for in vitro
15 transcription (e.g., T7 promoter).

Initial gRNA Screen

Each gRNA to be tested is transfected, along with a plasmid expressing Cas9 and a small amount of a GFP-expressing plasmid into human cells. In preliminary experiments, these cells can be immortalized human cell lines such as 293T, K562 or U2OS. Alternatively, primary
20 human cells may be used. In this case, cells may be relevant to the eventual therapeutic cell target (for example, photoreceptor cells). The use of primary cells similar to the potential therapeutic target cell population may provide important information on gene targeting rates in the context of endogenous chromatin and gene expression.

Transfection may be performed using lipid transfection (such as Lipofectamine or
25 Fugene) or by electroporation. Following transfection, GFP expression can be determined either by fluorescence microscopy or by flow cytometry to confirm consistent and high levels of transfection. These preliminary transfections can comprise different gRNAs and different targeting approaches (17-mers, 20-mers, nuclease, dual-nickase, etc) to determine which gRNAs/combinations of gRNAs give the greatest activity.

Efficiency of cleavage with each gRNA may be assessed by measuring NHEJ-induced indel formation at the target locus by a T7E1-type assay or by sequencing. Alternatively, other mismatch-sensitive enzymes, such as Celi/Surveyor nuclease, may also be used.

For the T7E1 assay, PCR amplicons are approximately 500-700bp with the intended cut site placed asymmetrically in the amplicon. Following amplification, purification and size-
5 verification of PCR products, DNA is denatured and re-hybridized by heating to 95°C and then slowly cooling. Hybridized PCR products are then digested with T7 Endonuclease I (or other mismatch-sensitive enzyme) which recognizes and cleaves non-perfectly matched DNA. If indels are present in the original template DNA, when the amplicons are denatured and re-
10 annealed, this results in the hybridization of DNA strands harboring different indels and therefore lead to double-stranded DNA that is not perfectly matched. Digestion products may be visualized by gel electrophoresis or by capillary electrophoresis. The fraction of DNA that is cleaved (density of cleavage products divided by the density of cleaved and uncleaved) may be used to estimate a percent NHEJ using the following equation: %NHEJ = $(1 - (1 - \text{fraction cleaved})^{1/2})$. The T7E1 assay is sensitive down to about 2-5% NHEJ.
15

Sequencing may be used instead of, or in addition to, the T7E1 assay. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed, miniprep and sequenced with a single primer. For large sequencing numbers, Sanger sequencing may be used for determining the exact nature of indels after determining the NHEJ
20 rate by T7E1.

Sequencing may also be performed using next generation sequencing techniques. When using next generation sequencing, amplicons may be 300-500bp with the intended cut site placed asymmetrically. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon,
25 e.g., for use in high throughput sequencing (for example on an Illumina MiSeq). This method allows for detection of very low NHEJ rates.

Example 2: Assessment of Gene Targeting by HDR

The gRNAs that induce the greatest levels of NHEJ in initial tests can be selected for
30 further evaluation of gene targeting efficiency. In this case, cells are derived from disease subjects and, therefore, harbor the relevant mutation.

Following transfection (usually 2-3 days post-transfection,) genomic DNA may be isolated from a bulk population of transfected cells and PCR may be used to amplify the target region. Following PCR, gene targeting efficiency can be determined by several methods.

Determination of gene targeting frequency involves measuring the percentage of alleles
5 that have undergone homologous directed repair (HDR) with the exogenously provided donor template or endogenous genomic donor sequence and which therefore have incorporated the desired correction (e.g., the missing G nucleotide at position 2299). If the desired HDR event creates or destroys a restriction enzyme site, the frequency of gene targeting may be determined by a RFLP assay. If no restriction site is created or destroyed, sequencing may be used to
10 determine gene targeting frequency. If a RFLP assay is used, sequencing may still be used to verify the desired HDR event and ensure that no other mutations are present. If an exogenously provided donor template is employed, at least one of the primers is placed in the endogenous gene sequence outside of the region included in the homology arms, which prevents amplification of donor template still present in the cells. Therefore, the length of the homology
15 arms present in the donor template may affect the length of the PCR amplicon. PCR amplicons can either span the entire donor region (both primers placed outside the homology arms) or they can span only part of the donor region and a single junction between donor and endogenous DNA (one internal and one external primer). If the amplicons span less than the entire donor region, two different PCRs should be used to amplify and sequence both the 5' and the 3'
20 junction.

If the PCR amplicon is short (less than 600bp) it is possible to use next generation sequencing. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, e.g., for use
25 in high throughput sequencing (for example on an Illumina MiSeq). This method allows for detection of very low gene targeting rates.

If the PCR amplicon is too long for next generation sequencing, Sanger sequencing can be performed. For Sanger sequencing, purified PCR amplicons will be cloned into a plasmid backbone (for example, TOPO cloned using the LifeTech Zero Blunt® TOPO® cloning kit), transformed, minipreped and sequenced.

The same or similar assays described above can be used to measure the percentage of alleles that have undergone HDR with endogenous genomic donor sequence and which therefore have incorporated the desired correction.

Incorporation by Reference

All publications, patents, and patent applications mentioned herein are hereby incorporated by reference in their entirety as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference. In case
5 of conflict, the present application, including any definitions herein, will control.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described
10 herein. Such equivalents are intended to be encompassed by the following claims.

Other embodiments are within the following claims.

What is claimed is:

1. A gRNA molecule comprising a targeting domain which is complementary with a target
5 domain from the *USH2A* gene.
2. The gRNA molecule of claim 1, wherein said targeting domain is configured to provide a cleavage event selected from a double strand break and a single strand break, within 200 nucleotides of a target position of a guanine deletion at nucleotide position 2299 (2299delG) in the *USH2A* gene.
- 10 3. The gRNA molecule of claim 1 or 2, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Table 1.
4. The gRNA molecule of any of claims 1-3, wherein said targeting domain is selected from those in Table 1.
- 15 5. The gRNA molecule of any of claims 1-4, wherein said targeting domain is GAGUGCAAAAAAGAAGCCAA.
6. The gRNA molecule of any of claims 1-4, wherein said targeting domain is GUUAGAUGUCACCAAUUGUA.
7. The gRNA molecule of any of claims 1-4, wherein said targeting domain is
20 GGUGUCACACUGAAGUCCUU.
8. The gRNA molecule of any of claims 1-4, wherein said targeting domain is GCCAUGGAGGUACACUGGC.
9. The gRNA molecule of any of claims 1-4, wherein said targeting domain is GUCACAGGCCUUACAAU.
- 25 10. The gRNA molecule of any of claims 1-4, wherein said targeting domain is GUCACACUGAAGUCCUU.
11. The gRNA molecule of any of claims 1-4, wherein said targeting domain is UGCAAAAAAGAAGCCAA.
12. The gRNA molecule of any of claims 1-4, wherein said targeting domain is
30 UGCAGAGAAAACUUUUA.

13. The gRNA molecule of any of claims 1-4, wherein said targeting domain is UGUUCACUGAGCCAUGG.

14. The gRNA molecule of any of claims 1-4, wherein said targeting domain is AUGGAGGUUACACUGGC.

5 15. The gRNA molecule of claim 1 or 2, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Table 2.

16. The gRNA molecule of any of claims 1, 2 or 15, wherein said targeting domain is selected from Table 2.

10 17. The gRNA molecule of any of claim 1 or 2, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Table 3.

18. The gRNA molecule of claim 1, wherein said targeting domain is selected from Table 3.

15 19. The gRNA molecule of claim 1 or 2, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Tables 4A-4E.

20. The gRNA molecule of any of claims 1, 2 or 19, wherein said targeting domain is selected from Tables 4A-4E.

20 21. The gRNA molecule of claim 1 or 2, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Tables 5A-5F.

22. The gRNA molecule of any of claims 1, 2 or 21, wherein said targeting domain is selected from Tables 5A-5F.

25 23. The gRNA molecule of claim 1 or 2, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Tables 6A-6D.

24. The gRNA molecule of any of claims 1, 2 or 23, wherein said targeting domain is selected from Tables 6A-6D.

25. The gRNA molecule of any of claims 1-24, wherein said gRNA is a modular gRNA.

30 26. The gRNA molecule of any of claims 1-24, wherein said gRNA is a chimeric gRNA.

27. The gRNA molecule of any of claims 1-26, wherein said targeting domain is 16 nucleotides or more in length.

28. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 16 nucleotides in length.

5 29. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 17 nucleotides in length.

30. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 18 nucleotides in length.

10 31. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 19 nucleotides in length.

32. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 20 nucleotides in length.

33. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 21 nucleotides in length.

15 34. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 22 nucleotides in length.

35. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 23 nucleotides in length.

20 36. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 24 nucleotides in length.

37. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 25 nucleotides in length.

38. The gRNA molecule of any of claims 1-27, wherein said targeting domain is 26 nucleotides in length.

25 39. The gRNA molecule of any of claims 1-38, comprising from 5' to 3':

 a targeting domain;

 a first complementarity domain;

 a linking domain;

 a second complementarity domain;

30 a proximal domain; and

 a tail domain.

40. The gRNA molecule of any of claims 1-39, comprising:

- a linking domain of no more than 25 nucleotides in length;
- a proximal and tail domain, that taken together, are at least 20 nucleotides in length;
- a targeting domain of 17 or 18 nucleotides in length.

5 41. The gRNA molecule of any of claims 1-40, comprising:

- a linking domain of no more than 25 nucleotides in length;
- a proximal and tail domain, that taken together, are at least 30 nucleotides in length;
- a targeting domain of 17 or 18 nucleotides in length.

42. The gRNA molecule of any of claims 1-41, comprising:

- 10 a linking domain of no more than 25 nucleotides in length;
- a proximal and tail domain, that taken together, are at least 30 nucleotides in length;
- a targeting domain of 17 nucleotides in length.

43. The gRNA molecule of any of claims 1-42, comprising:

- 15 a linking domain of no more than 25 nucleotides in length;
- a proximal and tail domain, that taken together, are at least 40 nucleotides in length;
- a targeting domain of 17 nucleotides in length.

44. A nucleic acid comprising a sequence encoding (a) a gRNA molecule comprising a targeting domain that is complementary with a target domain in a *USH2A* gene.

20 45. The nucleic acid of claim 44, wherein said gRNA molecule is a gRNA molecule of any of claims 1-43.

46. The nucleic acid of claim 44 or 45, wherein said targeting domain is configured to provide a cleavage event selected from a double strand break and a single strand break, within 200 nucleotides of a target position of a guanine deletion at nucleotide position 2299 (2299delG) in a *USH2A* gene.

25 47. The nucleic acid of any of claims 44-46, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Table 1.

48. The nucleic acid of any of claims 44-47, wherein said targeting domain is selected from those in Table 1.

30 49. The nucleic acid of any of claims 44-48, wherein said targeting domain is:
GAGUGCAAAAAGAAGCCAA.

50. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
GUUAGAUGUCACCAAUUGUA.

51. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
GGUGUCACACUGAAGUCCUU.

5 52. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
GCCAUGGAGGUUACACUGGC.

53. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
GUCACAGGCCUUACAAU.

54. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
10 GUCACACUGAAGUCCUU.

55. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
UGCAAAAAAGAAGCCAA.

56. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
UGCAGAGAAAACUUUUA.

15 57. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
UGUUCACUGAGCCAUGG.

58. The nucleic acid of any of claims 44-49, wherein said targeting domain is:
AUGGAGGUUACACUGGC.

59. The nucleic acid of any of claims 44-46, wherein said targeting domain comprises a
20 sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain
sequence from Table 2.

60. The nucleic acid of any of claims 44-46 or 59, wherein said targeting domain is selected
from Table 2.

61. The nucleic acid of any of claims 44-46, wherein said targeting domain comprises a
25 sequence that is the same as, or differs by no more than nucleotides from, a targeting domain
sequence from Table 3.

62. The nucleic acid of any of claims 44-46 or 61, wherein said targeting domain is selected
from Table 3.

63. The nucleic acid of any of claims 44-46, wherein said targeting domain comprises a
30 sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting
domain sequence from Tables 4A-4E.

64. The nucleic acid of any of claims 44-46 or 63, wherein said targeting domain is selected from Tables 4A-4E.

65. The nucleic acid of any of claims 44-46, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Tables 5A-5F.

66. The nucleic acid of any of claims 44-46 or 65, wherein said targeting domain is selected from Tables 5A-5F.

67. The nucleic acid of any of claims 44-46, wherein said targeting domain comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Tables 6A-6D.

68. The nucleic acid of any of claims 44-46 or 67, wherein said targeting domain is selected from Tables 6A-6D.

69. The nucleic acid of any of claims 44-68, wherein said gRNA is a modular gRNA.

70. The nucleic acid of any of claims 44-68, wherein said gRNA is a chimeric gRNA.

71. The nucleic acid of any of claims 44-70, wherein said targeting domain is 16 nucleotides or more in length.

72. The nucleic acid of any of claims 44-71, wherein said targeting domain is 16 nucleotides in length.

73. The nucleic acid of any of claims 44-71, wherein said targeting domain is 17 nucleotides in length.

74. The nucleic acid of any of claims 44-71, wherein said targeting domain is 18 nucleotides in length.

75. The nucleic acid of any of claims 44-71, wherein said targeting domain is 19 nucleotides in length.

76. The nucleic acid of any of claims 44-71, wherein said targeting domain is 20 nucleotides in length.

77. The nucleic acid of any of claims 44-71, wherein said targeting domain is 21 nucleotides in length.

78. The nucleic acid of any of claims 44-71, wherein said targeting domain is 22 nucleotides in length.

79. The nucleic acid of any of claims 44-71, wherein said targeting domain is 23 nucleotides in length.

80. The nucleic acid of any of claims 44-71, wherein said targeting domain is 24 nucleotides in length.

5 81. The nucleic acid of any of claims 44-71, wherein said targeting domain is 25 nucleotides in length.

82. The nucleic acid of any of claims 44-71, wherein said targeting domain is 26 nucleotides in length.

83. The nucleic acid of any of claims 44-82, comprising from 5' to 3':

- 10 a targeting domain;
 a first complementarity domain;
 a linking domain;
 a second complementarity domain;
 a proximal domain; and
15 a tail domain.

84. The nucleic acid of any of claims 44-83, comprising:

- a linking domain of no more than 25 nucleotides in length;
 a proximal and tail domain, that taken together, are at least 20 nucleotides in length;
 a targeting domain of 17 or 18 nucleotides in length.

20 85. The nucleic acid of any of claims 44-84, comprising:

- a linking domain of no more than 25 nucleotides in length;
 a proximal and tail domain, that taken together, are at least 30 nucleotides in length;
 a targeting domain of 17 or 18 nucleotides in length.

86. The nucleic acid of any of claims 44-85, comprising:

- 25 a linking domain of no more than 25 nucleotides in length;
 a proximal and tail domain, that taken together, are at least 30 nucleotides in length;
 a targeting domain of 17 nucleotides in length.

87. The nucleic acid of any of claims 44-86, comprising:

- a linking domain of no more than 25 nucleotides in length;
30 a proximal and tail domain, that taken together, are at least 40 nucleotides in length;
 a targeting domain of 17 nucleotides in length.

88. The nucleic acid of any of claims 44-87, further comprising: (b) a sequence that encodes a Cas9 molecule.

89. The nucleic acid of claim 88, wherein said Cas9 molecule comprises a nickase molecule.

90. The nucleic acid of claim 88 or 89, wherein said Cas9 molecule is an eaCas9.

5 91. The nucleic acid of claim 90, wherein said eaCas9 forms a double strand break in a target nucleic acid.

92. The nucleic acid of claim 90, wherein said eaCas9 molecule forms a single strand break in a target nucleic acid.

10 93. The nucleic acid of claim 92, wherein said single strand break is formed in the strand of the target nucleic acid to which the targeting domain of said gRNA is complementary.

94. The nucleic acid of claim 92, wherein said single strand break is formed in the strand of the target nucleic acid other than the strand to which the targeting domain of said gRNA is complementary.

15 95. The nucleic acid of any of claims 90, 92 or 93, wherein said eaCas9 molecule comprises HNH-like domain cleavage activity but has no, or no significant, N-terminal RuvC-like domain cleavage activity.

96. The nucleic acid of any of claims 90, 92, 93 or 95, wherein said eaCas9 molecule is an HNH-like domain nickase.

20 97. The nucleic acid of any of claims 90, 92, 93, 95 or 96, wherein said eaCas9 molecule comprises a mutation at D10.

98. The nucleic acid of any of claims 90, 92 or 94, wherein said eaCas9 molecule comprises N-terminal RuvC-like domain cleavage activity but has no, or no significant, HNH-like domain cleavage activity.

25 99. The nucleic acid of any of claims 90, 92, 94 or 98, wherein said eaCas9 molecule is an N-terminal RuvC-like domain nickase.

100. The nucleic acid of claim 90, 92, 94, 98 or 99, wherein said eaCas9 molecule comprises a mutation at H840.

30 101. The nucleic acid of any of claims 44-100, further comprising: (c) a sequence that encodes a second gRNA molecule described herein having a targeting domain that is complementary to a second target domain of the *USH2A* gene.

102. The nucleic acid of claim 101, wherein said second gRNA is a gRNA molecule of any of claims 1-43.
103. The nucleic acid of claim 102, wherein said targeting domain of said second gRNA is configured to provide a cleavage event selected from a double strand break and a single strand break, within 200 nucleotides of a guanine deletion at nucleotide position 2299 (2299delG) in the *USH2A* gene.
104. The nucleic acid of any of claims 101-103, wherein said targeting domain of said second gRNA comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Table 1.
105. The nucleic acid of any of claims 101-104, wherein said targeting domain of said second gRNA is selected from those in Table 1.
106. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is GAGUGCAAAAAGAAGCCAA.
107. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is GUUAGAUGUCACCAAUUGUA.
108. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is GGUGUCACACUGAAGUCCUU.
109. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is GCCAUGGAGGUUACACUGGC.
110. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is GUCACAGGCCUUACAAU.
111. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is GUCACACUGAAGUCCUU.
112. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is UGCAAAAAGAAGCCAA.
113. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is UGCAGAGAAAACUUUUA.
114. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is UGUUCACUGAGCCAUGG.
115. The nucleic acid of any of claims 101-105, wherein said targeting domain of said second gRNA is AUGGAGGUUACACUGGC.

116. The nucleic acid of any of claims 101-103, wherein said targeting domain of said second gRNA comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Table 2.

117. The nucleic acid of any of claims 101-103 or 116, wherein said targeting domain of said second gRNA is selected from Table 2.

118. The nucleic acid of any of claims 101-103, wherein said targeting domain of said second gRNA comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Table 3.

119. The nucleic acid of any of claims 101-103 or 118, wherein said targeting domain of said second gRNA is selected from Table 3.

120. The nucleic acid of any of claims 101-103, wherein said targeting domain of said gRNA comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Tables 4A-4E.

121. The nucleic acid of any of claims 101-103 or 120, wherein said targeting domain of said gRNA is selected from Tables 4A-4E.

122. The nucleic acid of any of claims 101-103, wherein said targeting domain of said gRNA comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Tables 5A-5F.

123. The nucleic acid of any of claims 101-103 or 122, wherein said targeting domain of said gRNA is selected from Tables 5A-5F.

124. The nucleic acid of any of claims 101-103, wherein said targeting domain of said gRNA comprises a sequence that is the same as, or differs by no more than 3 nucleotides from, a targeting domain sequence from Tables 6A-6D.

125. The nucleic acid of any of claims 101-103 or 124, wherein said targeting domain of said gRNA is selected from Tables 6A-6D.

126. The nucleic acid of any of claims 101-125, wherein said second gRNA is a modular gRNA.

127. The nucleic acid of any of claims 101-125, wherein said second gRNA is a chimeric gRNA.

128. The nucleic acid of any of claims 101-127, wherein said targeting domain is 16 nucleotides or more in length.

129. The nucleic acid of any of claims 101-128, wherein said targeting domain is 16 nucleotides in length.

130. The nucleic acid of any of claims 101-128, wherein said targeting domain is 17 nucleotides in length.

5 131. The nucleic acid of any of claims 101-128, wherein said targeting domain is 18 nucleotides in length.

132. The nucleic acid of any of claims 101-128, wherein said targeting domain is 19 nucleotides in length.

10 133. The nucleic acid of any of claims 101-128, wherein said targeting domain is 20 nucleotides in length.

134. The nucleic acid of any of claims 101-128, wherein said targeting domain is 21 nucleotides in length.

135. The nucleic acid of any of claims 101-128, wherein said targeting domain is 22 nucleotides in length.

15 136. The nucleic acid of any of claims 101-128, wherein said targeting domain is 23 nucleotides in length.

137. The nucleic acid of any of claims 101-128, wherein said targeting domain is 24 nucleotides in length.

20 138. The nucleic acid of any of claims 101-128, wherein said targeting domain is 25 nucleotides in length.

139. The nucleic acid of any of claims 101-128, wherein said targeting domain is 26 nucleotides in length.

140. The nucleic acid of any of claims 101-139, wherein said second gRNA comprises from 5' to 3':

25 a targeting domain;
a first complementarity domain;
a linking domain;
a second complementarity domain;
a proximal domain; and
30 a tail domain.

141. The nucleic acid of any of claims 101-140, wherein said second gRNA comprises:

a linking domain of no more than 25 nucleotides in length;
a proximal and tail domain, that taken together, are at least 20 nucleotides in length;
a targeting domain of 17 or 18 nucleotides in length.

5 142. The nucleic acid of any of claims 101-141, wherein said second gRNA comprises:
a linking domain of no more than 25 nucleotides in length;
a proximal and tail domain, that taken together, are at least 30 nucleotides in length;
a targeting domain of 17 or 18 nucleotides in length.

10 143. The nucleic acid of any of claims 101-142, wherein said second gRNA comprises:
a linking domain of no more than 25 nucleotides in length;
a proximal and tail domain, that taken together, are at least 30 nucleotides in length;
a targeting domain of 17 nucleotides in length.

15 144. The nucleic acid of any of claims 101-143, wherein said second gRNA comprises:
a linking domain of no more than 25 nucleotides in length;
a proximal and tail domain, that taken together, are at least 40 nucleotides in length;
a targeting domain of 17 nucleotides in length.

20 145. The nucleic acid of any of claims 101-144, wherein the targeting domain of said gRNA molecule and the targeting domain of said second gRNA molecules are complementary to opposite strands of the target nucleic acid molecule.

146. The nucleic acid of any of claims 101-145, wherein said gRNA molecule and said second gRNA molecule are configured such that the PAMs are oriented outward.

25 147. The nucleic acid of any of claims 101-146, wherein said gRNA molecule and said second gRNA molecule are configured such that they do not overlap and are separated by as much as 50, 100, or 200 nucleotides.

148. The nucleic acid of any of claims 101-147, wherein said gRNA and second gRNA are configured such that single strand breaks are formed on each strand of the target nucleic acid.

149. The nucleic acid of any of claims 101-148, wherein said gRNA and second gRNA are configured such that single strand breaks are formed on each strand of the target nucleic acid and the single strand breaks are within 50-100 nucleotides of one another.

150. The nucleic acid of any of claims 101-149, wherein said gRNA molecule and said
5 second gRNA molecule are configured such that the first and second breaks are 5' to the guanine deletion at nucleotide position 2299 in the *USH2A* gene.

151. The nucleic acid of any of claims 101-149, wherein said gRNA molecule and said second gRNA molecule are configured such that the first and second breaks are 3' to the guanine deletion at nucleotide position 2299 in the *USH2A* gene.

10 152. The nucleic acid of any of claims 101-149, wherein said gRNA molecule and said second gRNA molecule are configured such that the first and second breaks flank the guanine deletion at nucleotide position 2299 in the *USH2A* gene.

153. The nucleic acid of any of any of claims 44-152, further comprising: (d) a template nucleic acid.

15 154. The nucleic acid of claim 153, wherein the template nucleic acid is a single stranded nucleic acid.

155. The nucleic acid of claim 153, wherein said template nucleic acid is a double stranded nucleic acid.

156. The nucleic acid of any of claims 153-155, wherein said template nucleic acid
20 comprises a nucleotide sequence insertion or change in the target nucleic acid.

157. The nucleic acid of any of claims 153-156, wherein said template nucleic acid comprises a nucleotide sequence that is used to modify the target position.

158. The nucleic acid of any of claims 153-157, wherein said template nucleic acid comprises a nucleotide sequence that corresponds to wildtype sequence of the target position.

25 159. The nucleic acid of any of claims 153-158, wherein said template nucleic acid comprises a guanine to replace the deleted guanine at position 2299 in the *USH2A* gene.

160. The nucleic acid of any of claims 153-159, wherein said template nucleic acid comprises a 5' homology arm.

161. The nucleic acid of any of claims 153-160, wherein said template nucleic acid
30 comprises a 5' homology arm from Table 13.

162. The nucleic acid of any of claims 153-161, wherein the template nucleic acid comprises a 3' homology arm.

163. The nucleic acid of any of claim 153-162, wherein the template nucleic acid comprises a 3' homology arm from Table 13.

5 164. The nucleic acid of any of claims 88-163, wherein each of (a) and (b) is present on the same nucleic acid molecule.

165. The nucleic acid of claim 164, wherein said nucleic acid molecule is an AAV vector.

10 166. The nucleic acid of any of claims 88-163, wherein: (a) is present on a first nucleic acid molecule; and (b) is present on a second nucleic acid molecule.

167. The nucleic acid of claim 166, wherein said first and second nucleic acid molecules are AAV vectors.

168. The nucleic acid of any of claims 164-167, wherein said nucleic acid does not comprise (c) a sequence that encodes a second gRNA molecule.

15 169. The nucleic acid of any of claims 101-163, wherein each of (a) and (c) is present on the same nucleic acid molecule.

170. The nucleic acid of claim 169, wherein said nucleic acid molecule is an AAV vector.

20 171. The nucleic acid of any of claims 101-163, wherein (a) is present on a first nucleic acid molecule; and (c) is present on a second nucleic acid molecule.

172. The nucleic acid of claim 171, wherein said first and second nucleic acid molecules are AAV vectors.

173. The nucleic acid of any of claims 169-172, wherein said nucleic acid does not comprise (d) a template nucleic acid.

25 174. The nucleic acid of any of claims 101-163, wherein each of (a), (b), and (c) are present on the same nucleic acid molecule.

175. The nucleic acid of claim 174, wherein said nucleic acid molecule is an AAV vector.

30 176. The nucleic acid of any of claims 101-163, wherein: one of (a), (b), and (c) is encoded on a first nucleic acid molecule; and a second and third of (a), (b), and (c) is encoded on a second nucleic acid molecule.

177. The nucleic acid of claim 176, wherein said first and second nucleic acid molecules are AAV vectors.

178. The nucleic acid of any of claims 101-163, wherein: (a) is present on a first nucleic acid molecule; and (b) and (c) are present on a second nucleic acid molecule.

5 179. The nucleic acid of claim 178, wherein said first and second nucleic acid molecules are AAV vectors.

180. The nucleic acid of any of claims 101-163, wherein: (b) is present on a first nucleic acid molecule; and (a) and (c) are present on a second nucleic acid molecule.

10 181. The nucleic acid of claim 180, wherein said first and second nucleic acid molecules are AAV vectors.

182. The nucleic acid of any of claims 101-163, wherein: (c) is present on a first nucleic acid molecule; and (a) and (b) are present on a second nucleic acid molecule.

183. The nucleic acid of claim 182, wherein said first and second nucleic acid molecules are AAV vectors.

15 184. The nucleic acid of any of claims 153-163, wherein each of (a), (b), (c) and (d) are present on the same nucleic acid molecule.

185. The nucleic acid of claim 184, wherein said nucleic acid molecule is an AAV vector.

20 186. The nucleic acid of any of claims 153-163, wherein: one of (a), (b), (c) and (d) is encoded on a first nucleic acid molecule; and a second, third, and fourth of (a), (b), (c) and (d) is encoded on a second nucleic acid molecule.

187. The nucleic acid of claim 186, wherein said first and second nucleic acid molecules are AAV vectors.

25 188. The nucleic acid of any of claims 153-163, wherein: (a) is present on a first nucleic acid molecule; and (b), (c), and (d) are present on a second nucleic acid molecule.

189. The nucleic acid of claim 188, wherein said first and second nucleic acid molecules are AAV vectors.

190. The nucleic acid of any of claims 153-163, wherein: (b) is present on a first nucleic acid molecule; and (a), (c), and (d) are present on a second nucleic acid molecule.

30 191. The nucleic acid of claim 190, wherein said first and second nucleic acid molecules are AAV vectors.

192. The nucleic acid of any of claims 153-163, wherein: (c) is present on a first nucleic acid molecule; and (a), (b), and (d) are present on a second nucleic acid molecule.

193. The nucleic acid of claim 192, wherein said first and second nucleic acid molecules are AAV vectors.

5 194. The nucleic acid of any of claims 153-163, wherein: (d) is present on a first nucleic acid molecule; and (a), (b), and (c) are present on a second nucleic acid molecule.

195. The nucleic acid of claim 194, wherein said first and second nucleic acid molecules are AAV vectors.

10 196. The nucleic acid of any of claims 153-163, wherein: a first and second of (a), (b), (c) and (d) is encoded on a first nucleic acid molecule; and a third and fourth of (a), (b), (c) and (d) is encoded on a second nucleic acid molecule.

197. The nucleic acid of claim 196, wherein said first and second nucleic acid molecules are AAV vectors.

15 198. The nucleic acid of any of claims 153-163, wherein: (a) and (b) are present on a first nucleic acid molecule; and (c) and (d) are present on a second nucleic acid molecule.

199. The nucleic acid of claim 198, wherein said first and second nucleic acid molecules are AAV vectors.

200. The nucleic acid of any of claims 153-163, wherein (a) and (c) are present on a first nucleic acid molecule; and (b) and (d) are present on a second nucleic acid molecule.

20 201. The nucleic acid of claim 200, wherein said first and second nucleic acid molecules are AAV vectors.

202. The nucleic acid of any of claims 153-163, wherein (a) and (d) are present on a first nucleic acid molecule; and (b) and (c) are present on a second nucleic acid molecule.

25 203. The nucleic acid of claim 202, wherein said first and second nucleic acid molecules are AAV vectors.

204. The nucleic acid of any of claims 153-163, wherein: (b) and (d) are present on a first nucleic acid molecule; and (a) and (c) are present on a second nucleic acid molecule.

205. The nucleic acid of claim 204, wherein said first and second nucleic acid molecules are AAV vectors.

206. The nucleic acid of any of claims 166, 168, 171, 173, 176, 178, 180, 182, 186, 188, 190, 192, 194, 196, 198, 200, 202 or 204, wherein said first nucleic acid molecule is other than an AAV vector and said second nucleic acid molecule is an AAV vector.

207. The nucleic acid of any of claims 44-206, wherein said nucleic acid comprises a promoter operably linked to the sequence that encodes said gRNA molecule of (a).

208. The nucleic acid of any of claims 101-167 or 169-207, wherein said nucleic acid comprises a second promoter operably linked to the sequence that encodes the second gRNA molecule of (c).

209. The nucleic acid of claim 208, wherein the promoter and second promoter differ from one another.

210. The nucleic acid of claim 208, wherein the promoter and second promoter are the same.

211. The nucleic acid of any of claims 88-210, wherein said nucleic acid comprises a promoter operably linked to the sequence that encodes the Cas9 molecule of (b).

212. A composition comprising the (a) gRNA molecule of any of claims 1-43.

213. The composition of claim 212, further comprising (b) a Cas9 molecule of any of claims 88-100.

214. The composition of any of claims 211 or 213, further comprising (c) a second gRNA molecule of any of claims 1-43 or 101-152.

215. The composition of any of claims 212-214, further comprising: (d) a template nucleic acid of any of claims 153-163.

216. A method of altering a cell comprising contacting said cell with: (a) a gRNA of any of claims 1-43; (b) a Cas9 molecule of any of claims 88-100; optionally, (c) a second gRNA of any of claims 1-43 or 101-152; and (d) a template nucleic acid of any of claims 153-163.

217. The method of claim 216, comprising contacting said cell with (a), (b), (c), and (d).

218. The method of claim 216 or 217, wherein said cell is from a subject suffering from or likely to develop Usher Syndrome or retinitis pigmentosa-39.

219. The method of any of claims 216-218, wherein said cell is from a subject having a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG).

220. The method of any of claims 216-219, wherein said cell is a photoreceptor cell.

221. The method of any of claims 216-220, wherein said contacting is performed *ex vivo*.

222. The method of claim 221, wherein said contacted cell is returned to said subject's body.

5 223. The method of any of claims 216-220, wherein said contacting is performed *in vivo*.

224. The method of any of claims 216-223, comprising acquiring knowledge of the presence of a guanine deletion at nucleotide position 2299 in the *USH2A* gene in said cell.

10 225. The method of claim 224, comprising acquiring knowledge of the presence of a guanine deletion at nucleotide position 2299 in the *USH2A* gene in said cell by sequencing a portion of the *USH2A* gene.

226. The method of any of claims 216-225, comprising, based on the presence of a guanine deletion at nucleotide position 2299 in the *USH2A* gene, selecting a template nucleic acid.

15 227. The method of any of claims 216-226, comprising correcting the guanine deletion at nucleotide position 2299 in the *USH2A* gene.

228. The method of any of claims 216-227, wherein contacting comprises contacting said cell with a nucleic acid that expresses at least one of (a), (b), and (c).

20 229. The method of any of claims 216-228, wherein contacting comprises contacting the cell with a nucleic acid of any of claims 44-211.

230. The method of any of claims 216-229, wherein contacting comprises delivering to said cell said Cas9 molecule of (b) and a nucleic acid which encodes and (a) and optionally (c).

231. The method of any of claims 216-229, wherein contacting comprises delivering to said cell said Cas9 molecule of (b), said gRNA of (a) and optionally said second gRNA of (c).

25 232. The method of any of claims 216-229, wherein contacting comprises delivering to said cell said gRNA of (a), optionally said second gRNA of (c) and a nucleic acid that encodes the Cas9 molecule of (b).

30 233. A method of treating a subject having or likely to develop Usher Syndrome or retinitis pigmentosa 39, comprising contacting said subject (or a cell from said subject) with: (a) a gRNA of any of claims 1-43; (b) a Cas9 molecule of any of claims 88-100; optionally, (c) a

second gRNA of any of claims 1-43 or 101-152; and (d) a template nucleic acid of any of claims 153-163.

234. The method of claim 233, further comprising contacting said subject with (a), (b), (c), and (d).

5 235. The method of claim 233 or 234, wherein said subject has a guanine deletion at nucleotide position 2299 in the *USH2A* gene.

236. The method of any of claims 233-235, comprising acquiring knowledge of the presence of a guanine deletion at nucleotide position 2299 in the *USH2A* gene in said subject.

10 237. The method of claim 236, comprising acquiring knowledge of the presence of a guanine deletion at nucleotide position 2299 in the *USH2A* gene in said subject by sequencing a portion of the *USH2A* gene.

238. The method of any of claims 233-237, comprising, based on the presence of a guanine deletion at nucleotide position 2299 in the *USH2A* gene in said subject, selecting a template nucleic acid.

15 239. The method of any of claims 233-238, comprising correcting the guanine deletion at nucleotide position 2299 in the *USH2A* gene.

240. The method of any of claims 233-239, wherein a cell of said subject is contacted *ex vivo* with (a), (b), (d) and optionally (c).

241. The method of claim 240, wherein said cell is returned to the subject's body.

20 242. The method of any of claims 233- 241, wherein treatment comprises introducing a cell into said subject's body, wherein said cell subject was contacted *ex vivo* with (a), (b), (d) and optionally (c).

243. The method of any of claims 233-239, wherein said contacting is performed *in vivo*.

25 244. The method of claim 243, wherein said contacting comprises subretinal delivery.

245. The method of claim 244, wherein said contacting comprises subretinal injection.

246. The method of any of claims 233-245, wherein contacting comprises contacting said subject with a nucleic acid that expresses at least one of (a), (b), and (c).

30 247. The method of any of claims 233-246, wherein contacting comprises contacting said subject with a nucleic acid of any of claims 44-211.

248. The method of any of claims 233-247, wherein contacting comprises delivering to said subject said Cas9 molecule of (b) and a nucleic acid which encodes and (a) and optionally (c).

249. The method of any of claims 233-247, wherein contacting comprises delivering to said subject said Cas9 molecule of (b), said gRNA of (a) and optionally said second gRNA of (c).

250. The method of any of claims 233-247, wherein contacting comprises delivering to said subject said gRNA of (a), optionally said second gRNA of (c) and a nucleic acid that encodes the Cas9 molecule of (b).

251. A gRNA molecule of any of claims 1-43 for use in treating Usher Syndrome or retinitis pigmentosa 39 in a subject.

252. The gRNA molecule of claim 252, wherein the gRNA molecule is used in combination with (b) a Cas9 molecule of any of claims 88-100.

253. The gRNA molecule of claim 251 or 252, wherein the gRNA molecule is used in combination with (c) a second gRNA molecule of any of claims 1-43 or 101-152.

254. The gRNA molecule of any of claims 251-253, wherein the gRNA molecule is used in combination with (d) a template nucleic acid of any of claims 153-163.

255. Use of a gRNA molecule of any of claims 1-43 in the manufacture of a medicament for treating Usher Syndrome or retinitis pigmentosa 39 in a subject.

256. The use of claim 255, wherein the medicament further comprises (b) a Cas9 molecule of any of claims 88-100.

257. The use of claim 255 or 256, wherein the medicament further comprises (c) a second gRNA molecule of any of claims 1-43 or 101-152.

258. The use of any of claims 255-257, wherein the medicament further comprises (d) a template nucleic acid of any of claims 153-163.

259. A composition of any of claim 212-215 for use in treating Usher Syndrome or retinitis pigmentosa 39 in a subject.

260. A reaction mixture comprising a gRNA, a nucleic acid, or a composition described herein, and a cell from a subject having or likely to develop Usher Syndrome or retinitis pigmentosa-39, or a subject having a mutation in the *USH2A* gene, e.g., a deletion of guanine at nucleotide position 2299 (2299delG).

261. A kit comprising, (a) gRNA molecule of any of claims 1-43, or nucleic acid that encodes said gRNA, and one or more of the following: (b) a Cas9 molecule of any of claims 88-100; (c) a second gRNA molecule of any of claims 1-43 or 101-152; (d) a template nucleic acid of any of claims 153-163; and (e) nucleic acid that encodes one or more of (b), (c), or (d).

5 262. The kit of claim 261, comprising nucleic acid that encodes one or more of (a), (b) and (c).

263. The kit of claim 261 or 262, further comprising a template nucleic acid that is a single strand DNA.

264. A non-naturally occurring template nucleic acid of any of claims 153-163.

10

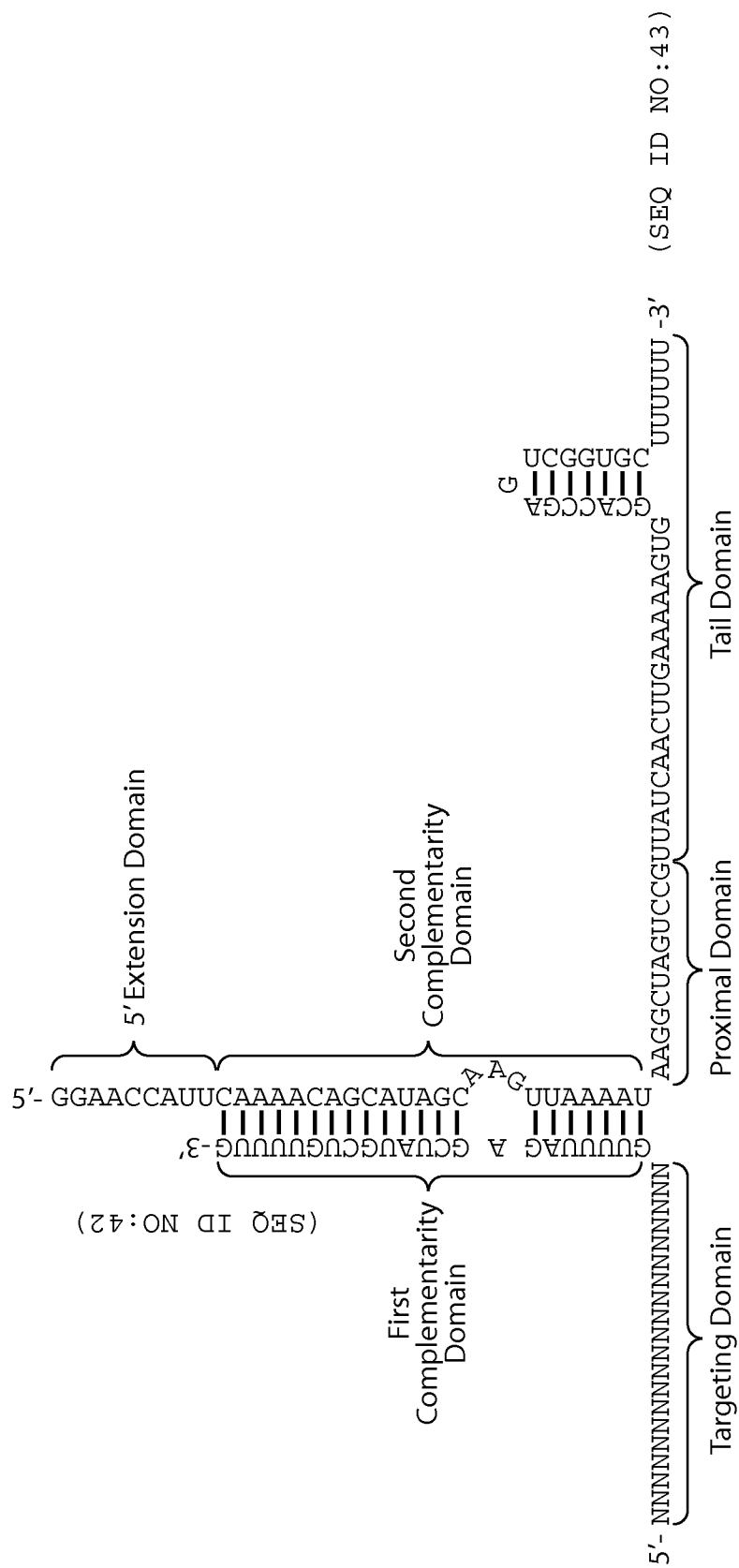


Fig. 1A

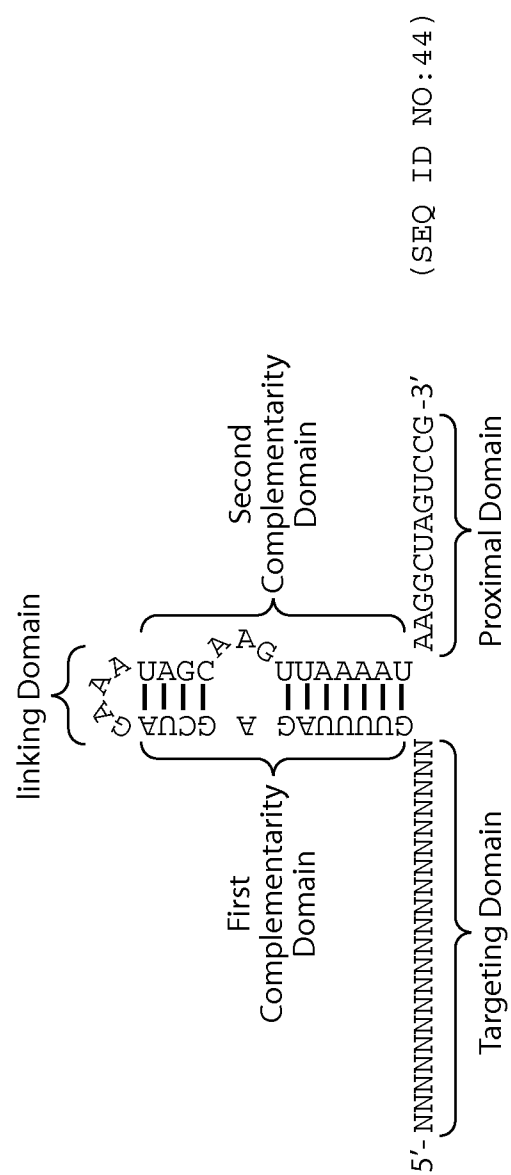


Fig. 1B

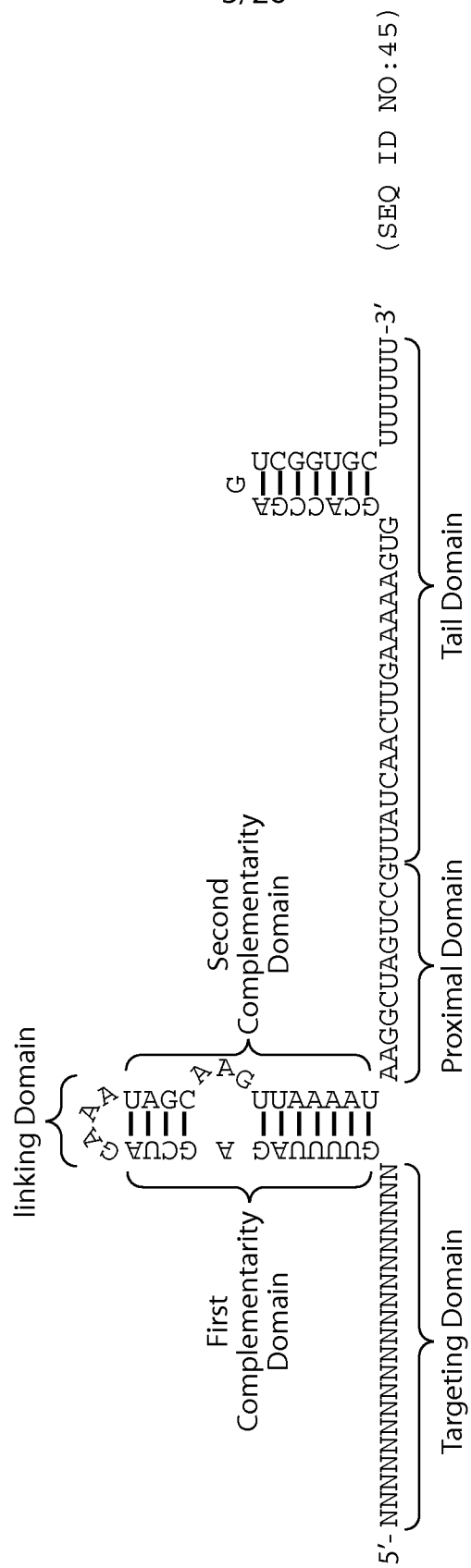


Fig. 1C

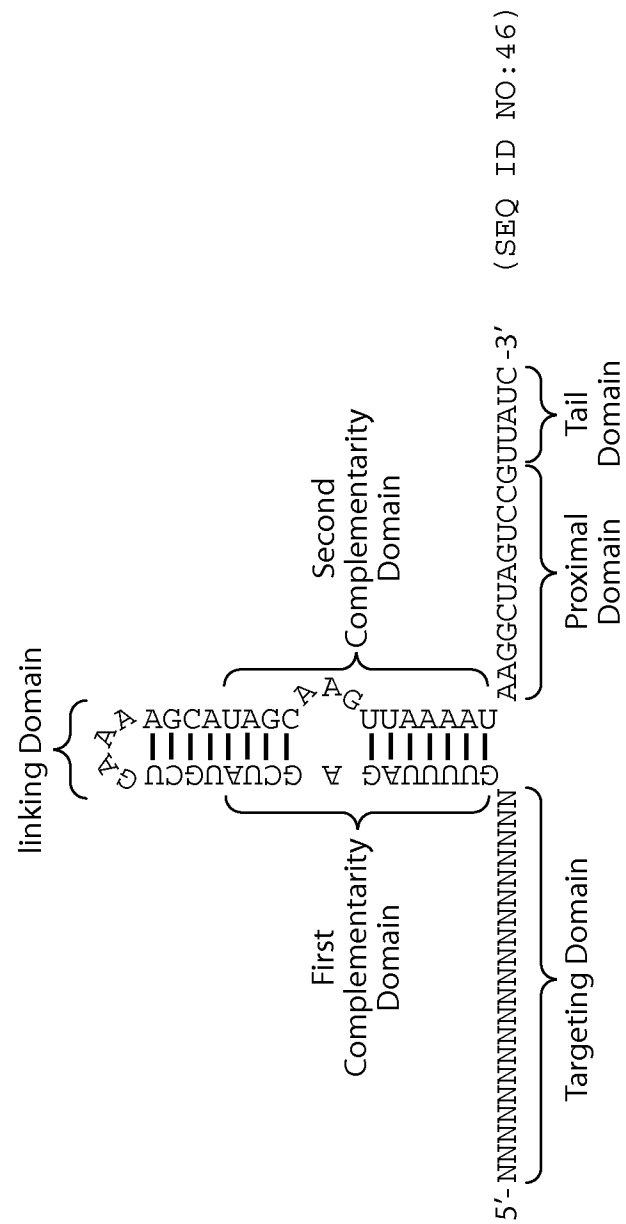


Fig. 1D

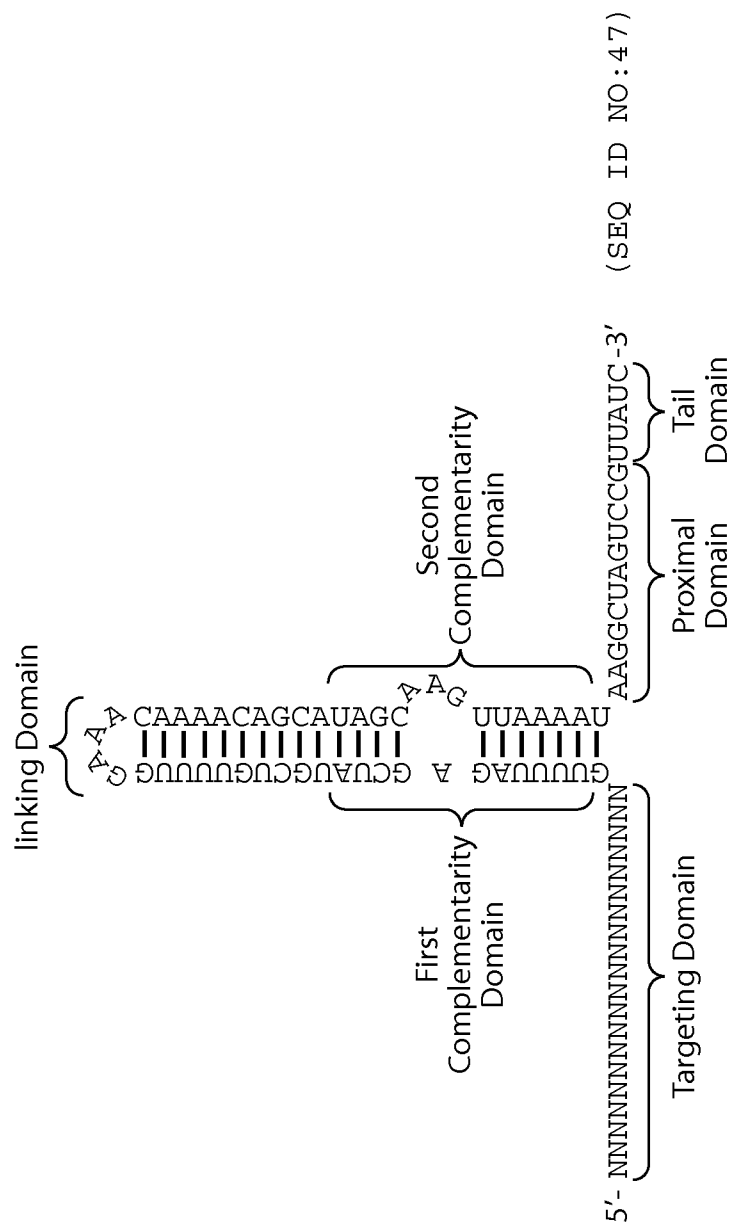


Fig. 1E

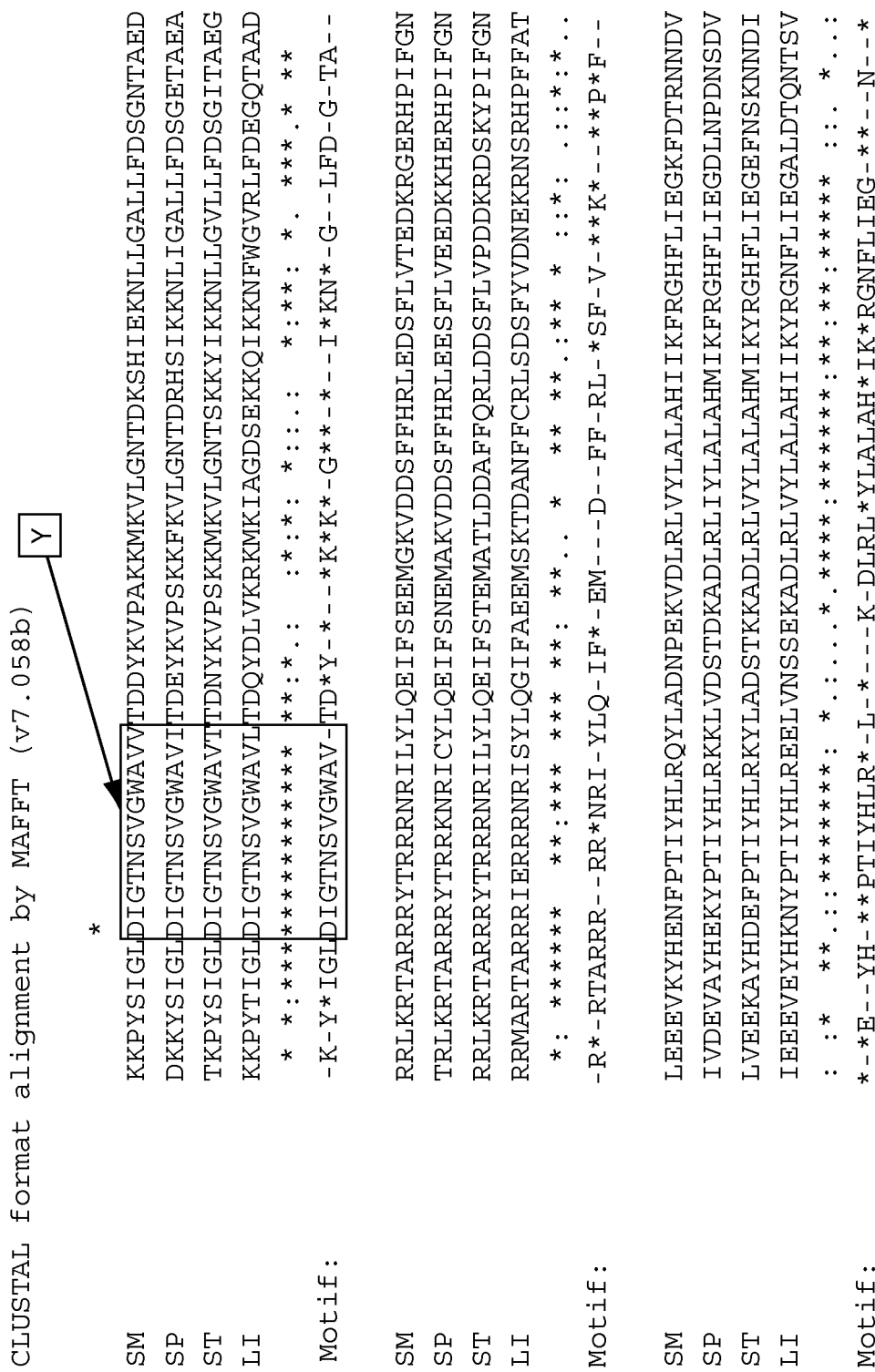


Fig. 2A

[illegible]

Fig. 2B

SM LQEMRAIIRRQAEFYPFLADNQDRIEKLTLTFRIPYYVGPLARGKSDFAWLSRKSADKITP
SP LGELHAILRRQEDFYPPFLKDNREKIEKILTFRIPYYVGPLARGNSRFFAWMTRKSEETITP
ST LQEMRAILDQAKFYPFLAKNERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITP
LI LEELEAILHQAKYYPFLKENYDKISLVTFRIPYFVGPLANGQSEFFAWLTRKADGEIRP
* *:***: :* :***** * :*:*****:*****:*****:*****:***** * *
Motif: L-E*-AI*-Q--*YPFL--N-**-I*-**TFRIPY*VGPLA-G*S-FAW--RK-----I-P

SM WNFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTE-QG
SP WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMR
ST WNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPHKSLLYETFNVYNELTKVRFIAESMR
LI WNIEEKVDFGKSAVDFIEKMTNKDITYLPKENVLPKHSLCYQKYL VYNELTKVRYIND-QG
*:***: :* ** *:***: * *:***:***** *: : *****: :
Motif: WN***-D---SA--FI**MT--D--LP***VLPKHSL-Y*-*-VYNELTKV***-*-

SM KTAFFDANMKQEIFDGVFKVYRKVTCKDKLMDFLEKEFFDEFRIVDLTGLDKENKVFNASYG
SP KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR---FNASLG
ST DYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYL-HAIYGDGIELKGIEKQ---FNSSLS
LI KTSYFSGQEKEQIFNDLFFKQKRKVKKKDLLELFL-RNMHVESPTIEGLEDS---FNSSYS
* :*:***: :* ** *:***: * *:***: : * :*:***: * :*:***: * :
Motif: ----**--*-K*-I-----FK--RKV-----*-*****-G*****-FN*S--

SM TYHDLCKIL-DKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLNYSDDLTKEQVKKLE
SP TYHDLCLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYYAHLFDDKVMKQLK
ST TYHDLJNIINDKEFLDDSSNEAIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLS
LI TYHDLCLKVGIKQEIILDNPVNTEMLENIVKILTVFEDKRMKEQLQFSDVLDGVVLKKLE
***** : : :*:***: * :*:***: * :*:***:*****:*****: : : :*:***: * :
Motif: TYHDL-*****LD*-N--**E*I*-LT**FED*-MI-**-L-----**--**K*L-

Fig. 2C

SM
SP
ST
LI

KLKSAKLITQPKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETD
QLLNAKLITQPKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD
QLLKSLISQPKFDNLTKAERGGLSPEDKAGFIQRQLVETRQITKHVARLLDEKFNKKD
KLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNIEKD
:*...*:;:*****;*****:*...*:;*:
*L---*L***RKFD-LTKAERGGL*--DKA-FI*ROLVETRQITK*VA-*L---**N-*D

Motif:

Motif:

[illegible]

Motif:

SM EPEFVYGDYPHFHGHKE-----NK-ATAKKFFYSNIMNFFFKKDDVRTD-----
 SP ESEFVYGDYKVYDVRKMIAKSEQEIGK-ATAKYFFYSNIMNFFFKTEITLANGEIRKRPLI
 ST EPEFVYGDYPKYN SFRE-----RKSATEKVFFYSNIMNI FFKSISLADGRVIERPLI
 LI EPEFVYGDYHQFDWFKA-----NK-ATAKKQFYTNIMLFFAQKDRIID-----
 *.***** :. : * * * * * : * * * * * :
 Motif: E-EFVYGDY--*---*-----K-AT-K--FY*NIM-*F-----*

Motif:

SM
---KNGEIIWKKDEHISNIKKVLSYPQVNIVKKVEEQTGFSKE-----SILPK

SP
ETNGETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGFSKE-----SILPK

ST
EVNEETGESVMNKESDLATVRRVL SYPQNVVVKKVEEQNHGLDRCKPGLFNANLSSPKP

LI
---ENGEILWDK-KYLDTVKKVMSYRQMNI VKKTEIQKGEFSA-----TIKPK

: ** : . * : : : : * : * : * : * : :

Motif:
---*-GE-*W-K-***V*M-Q*N*VKK-E-Q-***-----*--PK

Motif:

Fig. 2E

SM	GNSDK-LIPR	TKKFFYWD	TKKYGGF	DSPIV	AYSIL	VIADIE	KGSKKL	KT	VKAL	VGVTIM
SP	RNSDK-LI	ARKD---	WDP	KYGGF	DSPTV	AYSVL	VVAK	VEKG	SKKL	KSVKELLGITIM
ST	PNSNENLV	GAKEY---	LDP	KYGGY	AGISN	FTVL	VKG	TIEK	GAKKI	TNVLEFQGISIL
LI	GNSSK-LI	PRKTN---	WDP	MYGGL	DSPNM	AYAVI	-EY	AKGN	-KL	VFEKKIIRVTIM
Motif:	**:	*	*	*	*	*	*	*	*	*
	-NS-	*	L	*	-K-	-D-	-KYGG-	*****	-KG-	-K*-----*
SM	EKMTF	ERDP	VAF	LER	KGYR	NVQE	ENI	I	KLP	KYSLF
SP	ERSF	EKNP	ID	F	LEAK	GYKE	VKKDL	I	I	KLP
ST	DRIN	YKDK	KL	N	F	LEK	GYKDI	-E	L	I
LI	ERKAF	EKDE	KA	F	LEE	QGYR	QP-	-K	V	LAK
Motif:	::	:::	*	*	*	:::	*	*	*	*
	**	----	FL-	*	GY	**	-----*	*LP	KY*L	**--*G-*R*IAS-----E-*K
SM	GN	EIV	PN	HL	G	TLL	YHAK	NI	HKV	-----
SP	GNEL	AL	PS	KY	VN	FL	YLASH	YE	KLK	GSPED
ST	GNQI	FL	SQ	KF	V	KL	LYHAK	RI	SNT	-----
LI	GNQV	LP	NHL	VT	LL	HHAA	NCE	VS	-----	DG
Motif:	**:	*	:::	*	*	*	*	*	*	*
	GN*	---	L	---	*	L*	-A	-----*	*	-----E*
SM	LA	E	N	L	E	K	I	K	E	L
SP	LADAN	L	D	K	V	L	S	A	N	K
ST	GAKK	N	G	L	L	N	S	A	F	Q
LI	LA	E	A	N	L	N	K	I	N	Q
Motif:	*	*	*	*	*	*	*	*	*	*
	-A	-N	---	*	-----*	*	-----*	*	-----L	*

Fig. 2F

SM	KR-YTSTTEILNATLIHQSI	TGLYETRIDLNKLGGD	(SEQ ID NO:1)
SP	KR-YTSTKEVLDATLIHQSI	TGLYETRIDLSQLGGD	(SEQ ID NO:2)
ST	YRDYTPSLLKDATLIHQSV	TGLYETRIDLAKLGEG	(SEQ ID NO:3)
LI	KR-YNNLKELLNSTIIYQSI	TGLYESRKRID----	(SEQ ID NO:4)
	* * . . : : * : * : * : * : * : *		
Motif:	-R-Y-----*--**T*I*QS*	TGLYE*R--L-----	

Fig. 2G

14/26

Alignment of the N terminal RucV-like Domains disclosed in Chylinski et al.
(excluding sequence outliers).

(CLUSTAL format alignment by MAFFT (v7.058b))

```

1 DIGTNSVGWAVT (SEQ ID NO:54)
12 DIGTNSVGWAVT (SEQ ID NO:55)
3 DVG TNSVGWAVT (SEQ ID NO:56)
20 DVG TNSVGWAVT (SEQ ID NO:57)
15 DMGTNSVGWAVT (SEQ ID NO:58)
4 DVG TSSVGWAVT (SEQ ID NO:59)
7 DIGTASVGWAVT (SEQ ID NO:60)
6 DVG TGSVGWAVT (SEQ ID NO:61)
9 DIGTNSVGWAVV (SEQ ID NO:62)
10 DIGTNSVGWAVI (SEQ ID NO:63)
11 DIGTNSVGWAVL (SEQ ID NO:64)
42 DLGTNSIGWAVV (SEQ ID NO:65)
48 DLGTNSIGWAI - (SEQ ID NO:66)
43 DLGTNSIGWALV (SEQ ID NO:67)
2 DIGTNSVGWCVT (SEQ ID NO:68)
14 DIGTNSVGYAVT (SEQ ID NO:69)
5 DMGTGSLGWAVT (SEQ ID NO:70)
16 DIGTSSVGWAAI (SEQ ID NO:71)
8 DLGTGSVGWAVV (SEQ ID NO:72)
22 DLGVGSVGWAIV (SEQ ID NO:73)
23 DLGIASIGWAI (SEQ ID NO:74)
24 DLGIASVGWAIV (SEQ ID NO:75)
25 DLGVASVGWSIV (SEQ ID NO:76)
26 DIGIASVGWAIL (SEQ ID NO:77)
28 DLGISSVGWSVI (SEQ ID NO:78)
32 DIGIASVGWSVI (SEQ ID NO:79)
33 DVGIGSIGWAVI (SEQ ID NO:80)
39 DLGVGSIGFAIV (SEQ ID NO:81)
34 DIGYASIGWAVI (SEQ ID NO:82)
47 DTGTNSLGWAIV (SEQ ID NO:83)
50 DLGTNSIGWCLL (SEQ ID NO:84)
49 DIGTDSLGWAVF (SEQ ID NO:85)
18 DIGSNSIGFAVV (SEQ ID NO:86)
41 DLGVGSIGVAVA (SEQ ID NO:87)
45 DLGIASCGWGVV (SEQ ID NO:88)

```

Fig. 3A

21	DLGIASVGWCLT	(SEQ	ID NO: 89)
27	DIGIGSVGVGIL	(SEQ	ID NO: 90)
29	DIGITSVGYGLI	(SEQ	ID NO: 91)
30	DIGITSVGFGII	(SEQ	ID NO: 92)
31	DVGITSTGYAVL	(SEQ	ID NO: 93)
40	DLGITSFGYAIL	(SEQ	ID NO: 94)
17	DIGNASVGWSAF	(SEQ	ID NO: 95)
19	DVGTNSCGWVAM	(SEQ	ID NO: 96)
35	DVGERSIGLAAV	(SEQ	ID NO: 97)
36	DVGLNSVGLAAV	(SEQ	ID NO: 98)
37	DVGLMSVGLAAI	(SEQ	ID NO: 99)
38	DVGTFSVGLAAI	(SEQ	ID NO: 100)
13	DIGTGSVGYACM	(SEQ	ID NO: 101)
44	DLGTTSIGFAHI	(SEQ	ID NO: 102)
46	DLGTNSIGSSVR	(SEQ	ID NO: 103)
	* * * *		

Fig. 3B

Alignment of the N terminal RucV-like Domains disclosed in Chylinski et al.
(CLUSTAL format alignment by MAFFT (v7.058b))

1	D-----IGTNSVGWAVT	(SEQ	ID NO:104)
12	D-----IGTNSVGWAVT	(SEQ	ID NO:105)
3	D-----VGTNSVGWAVT	(SEQ	ID NO:106)
20	D-----VGTNSVGWAVT	(SEQ	ID NO:107)
15	D-----MGTNSVGWAVT	(SEQ	ID NO:108)
4	D-----VGTSSVGWAVT	(SEQ	ID NO:109)
7	D-----IGTASVGWAVT	(SEQ	ID NO:110)
6	D-----VGTGSVGWAVT	(SEQ	ID NO:111)
9	D-----IGTNSVGWAVV	(SEQ	ID NO:112)
10	D-----IGTNSVGWAVI	(SEQ	ID NO:113)
52	D-----IGTNSIGWAVI	(SEQ	ID NO:114)
11	D-----IGTNSVGWAVL	(SEQ	ID NO:115)
42	D-----LGTNSIGWAVV	(SEQ	ID NO:116)
48	D-----LGTNSIGWAI -	(SEQ	ID NO:117)
43	D-----LGTNSIGWALV	(SEQ	ID NO:118)
2	D-----IGTNSVGWCVT	(SEQ	ID NO:119)
14	D-----IGTNSVGYAVT	(SEQ	ID NO:120)
5	D-----MGTGSLGWAVT	(SEQ	ID NO:121)
16	D-----IGTSSVGWAAI	(SEQ	ID NO:122)
8	D-----LGTGSVGWAVV	(SEQ	ID NO:123)
22	D-----LGVGSVGWAI V	(SEQ	ID NO:124)
23	D-----LGIASIGWAI I	(SEQ	ID NO:125)
24	D-----LGIASVGWAI V	(SEQ	ID NO:126)
68	D-----LGIASVGWAVV	(SEQ	ID NO:127)
25	D-----LGVASVGWSIV	(SEQ	ID NO:128)
26	D-----IGIASVGWAIL	(SEQ	ID NO:129)
66	D-----IGIASVGWAVL	(SEQ	ID NO:130)
59	D-----IGIASIGWAVI	(SEQ	ID NO:131)
61	D-----IGIASVGWAI I	(SEQ	ID NO:132)
64	D-----VGIASVGWAVI	(SEQ	ID NO:133)
62	D-----IGIASVGWAL -	(SEQ	ID NO:134)
67	D-----IGIASVGWAMV	(SEQ	ID NO:135)
32	D-----IGIASVGWSVI	(SEQ	ID NO:136)
28	D-----LGISSVGWSVI	(SEQ	ID NO:137)
63	D-----IGITSVGWAVI	(SEQ	ID NO:138)

Fig. 4A

17/26

33	D-----VGIGSIGWAVI	(SEQ	ID	NO:139)
57	D-----LGISSLGWAIV	(SEQ	ID	NO:140)
39	D-----LGVGSIGFAIV	(SEQ	ID	NO:141)
34	D-----IGYASIGWAVI	(SEQ	ID	NO:142)
50	D-----LGTNSIGWCLL	(SEQ	ID	NO:143)
54	D-----LGTNSIGWGLL	(SEQ	ID	NO:144)
47	D-----TGTNSLGAIV	(SEQ	ID	NO:145)
49	D-----IGTDSLGAIVF	(SEQ	ID	NO:146)
51	D-----LGSTSLGWAIF	(SEQ	ID	NO:147)
58	D-----IGISSIGWAFS	(SEQ	ID	NO:148)
21	D-----LGIASVGWCLT	(SEQ	ID	NO:149)
45	D-----LGIASCGWGV	(SEQ	ID	NO:150)
18	D-----IGSNSIGFAVV	(SEQ	ID	NO:151)
65	D-----IGTTSIGFSVI	(SEQ	ID	NO:152)
29	D-----GITSVGYGLI	(SEQ	ID	NO:153)
30	D-----GITSVGFGLI	(SEQ	ID	NO:154)
44	D-----LGTTSIGFAHI	(SEQ	ID	NO:155)
27	D-----IGIGSVGVGIL	(SEQ	ID	NO:156)
41	D-----LGVGSIGGAVA	(SEQ	ID	NO:157)
31	D-----VGITSTGYAVL	(SEQ	ID	NO:158)
40	D-----LGITSFYAIL	(SEQ	ID	NO:159)
53	D-----IGTSSIGWVLY	(SEQ	ID	NO:160)
55	D-----LGSNSLGFVVT	(SEQ	ID	NO:161)
56	D-----LGANSLGWFFV	(SEQ	ID	NO:162)
17	D-----IGNASVGWSAF	(SEQ	ID	NO:163)
19	D-----VGTNSCGWVAM	(SEQ	ID	NO:164)
35	D-----VGERSIGLAAV	(SEQ	ID	NO:165)
36	D-----VGLNSVGLAAV	(SEQ	ID	NO:166)
37	D-----VGLMSVGLAAI	(SEQ	ID	NO:167)
38	D-----VGTFSVGLAAI	(SEQ	ID	NO:168)
13	D-----IGTGSVGYACM	(SEQ	ID	NO:169)
46	D-----LGTNSIGSSVR	(SEQ	ID	NO:170)
60	DIGLRIGITSCGWSI -	(SEQ	ID	NO:171)
69	D-----MGAKYTGFFYA	(SEQ	ID	NO:172)
73	D-----LGGKNTGFFSF	(SEQ	ID	NO:173)
74	D-----LGVKNTGVFSA	(SEQ	ID	NO:174)
70	D-----LGAKFTGVVALY	(SEQ	ID	NO:175)
71	D-----LGGKFTGVCLS	(SEQ	ID	NO:176)
72	D-----LGGTYTGTFTIT	(SEQ	ID	NO:177)

Fig. 4B

18/26

Alignment of the HNH-like Domains disclosed in Chylinski et al.
(CLUSTAL format alignment by MAFFT (v7.058b))

```

1 YDIDHIYPRS-LTKD-----DSF-DNLVLCERTAN      (SEQ ID NO: 178)
2 -DIDHIYPRSKVIK-----DSF-DNLVLVLKNEN      (SEQ ID NO: 179)
3 -DRDHIYPQS-KIKD-----DSI-DNLVLVNKTYN      (SEQ ID NO: 180)
4 -DIDHIYPRS-KIKD-----DSI-TNRVLVEKDIN      (SEQ ID NO: 181)
6 -DIDHIYPQS-KIKD-----DSI-SNRVLVCSSCN      (SEQ ID NO: 182)
5 -DIDHIYPQS-KTMD-----DSL-NNRVLVKKNYN      (SEQ ID NO: 183)
7 -DQDHIYPKS-KIYD-----DSL-ENRVLVKKNLN      (SEQ ID NO: 184)
8 -QIDHIYPQS-LVKD-----DSF-DNRVLVVPSEN      (SEQ ID NO: 185)
9 -DIDHIIPQA-FIKD-----NSI-DNRVLTSSKEN      (SEQ ID NO: 186)
12 -DIDHIIPQA-FLKD-----NSI-DNKVLVSSASN      (SEQ ID NO: 187)
16 -DIDHIIPQA-YTKD-----NSL-DNRVLVSNITN      (SEQ ID NO: 188)
11 -DIDHIYPQS-FITD-----NSI-DNLVLTSSAGN      (SEQ ID NO: 189)
10 -DVDHIYPQS-FLKD-----DSI-DNKVLTRSDKN      (SEQ ID NO: 190)
14 -NIDHIYPQS-MVKD-----DSL-DNKVLVQSEIN      (SEQ ID NO: 191)
18 -DIDHILPQS-LIKD-----DSL-DNRVLVNATIN      (SEQ ID NO: 192)
19 -DIDHILPQS-FIKD-----DSL-ENRVLVKKAVN      (SEQ ID NO: 193)
13 -EVDHIFPRS-FIKD-----DSI-DNKVLVIKKMN      (SEQ ID NO: 194)
15 -EVDHIIPRS-YIKD-----DSF-ENKVLVYREEN      (SEQ ID NO: 195)
17 -DIDHIIPQA-VTON-----DSI-DNRVLVARAEN      (SEQ ID NO: 196)
22 -EIDHIIPYS-ISFD-----DSS-SNKLLVLAESN      (SEQ ID NO: 197)
24 -EIDHIIPYS-LCFD-----DSS-ANKVLVHKQSN      (SEQ ID NO: 198)
32 -DIDHIIPYS-RSMD-----DSY-SNKVLVLGSEN      (SEQ ID NO: 199)
63 -DIDHIIPYS-KSMD-----DSF-SNKVLCLEEN      (SEQ ID NO: 200)
59 -EIDHIYPYS-RSFD-----DSY-MNKVLVFTKQN      (SEQ ID NO: 201)
65 -QIDHIYPYS-RSMD-----DSY-MNKVLVLTIDEN      (SEQ ID NO: 202)
64 -EIDHIIPFS-RSFD-----DSL-SNKILVLGSEN      (SEQ ID NO: 203)
68 -EIDHALPFS-RTWD-----DSF-SNKVLVLGSEN      (SEQ ID NO: 204)
69 -EIDHALPFS-RTWD-----DSF-SNKVLVLASEN      (SEQ ID NO: 205)
28 -EIDHIIPIS-ISLD-----DSI-SNKVLVLSEAN      (SEQ ID NO: 206)
30 -EVDHIIPIS-ISLD-----DSI-TNKVLVTHREN      (SEQ ID NO: 207)
62 -QVDHALPYS-RSYD-----DSK-SNKVLVLTTHEN      (SEQ ID NO: 208)
27 -EVDHILPLS-ITFD-----DSL-SNKVLVYATAN      (SEQ ID NO: 209)
26 -EIDHIIPRS-ISFD-----DAR-SNKVLVYRSEN      (SEQ ID NO: 210)

```

Fig. 5A

19/26

29 -EVDHIIPRS-VSFD-----NSY-HNKVLVKQSEN
 67 -DIDHILPYS-ITFD-----DSF-RNKVLVTSQEN
 58 -EIDHILPRS-RSAD-----DSF-ANKVLCCLARAN
 51 -EIEHLLPFS-LTLD-----DSM-ANKTVCFRQAN
 55 -DIDHILPFS-VSLD-----DSA-ANKVVCLREAN
 57 -DIDHILPFS-ISWD-----DSA-ANKVVCMRYAN
 56 -DIDHILPVA-MTLD-----DSP-ANKIICMRYAN
 54 -DVDHILPYS-RTLD-----DSF-PNRTLCLREAN
 52 -EIEHILPFS-RTLD-----DSL-NNRTVAMRRAN
 31 -EVDHIIPYS-ISWD-----DSY-TNKVLTSACN
 45 -QVDHILPWS-RFGD-----DSY-LNKTLCARSN
 53 -QVDHILPFS-KTLD-----DSF-ANKVLAQHDAN
 60 -QIDHAFPLS-RSLD-----DSQ-SNKVLCCLTSSN
 21 -DIDHIVPRS-ISFD-----DSF-SNLVIVNKLIDN
 23 -EIEHIIIPYS-MSYD-----NSQ-ANKILTEKAEN
 25 -EIDHVIPYS-KSAD-----DSW-FNKLLVKKSTN
 49 -EMDHILPYS-RSLD-----NGW-HNRVLVHGKDN
 33 -EVDHIVPYS-LILD-----NTI-NNKALVYAEEN
 42 -EIEHVIPQS-LYFD-----DSF-SNKVICEAEVN
 43 -DIEHIIPQA-RLFD-----DSF-SNKTLARSVN
 44 -EIEHIVPKA-RVFD-----DSF-SNKTLTFHRIN
 20 -DKDHIIPQS-MKCD-----DSI-INNLVLVNKNAN
 66 -EVEHIIWPRS-RSFD-----NSP-RNKTLCKRQDVN
 61 -IVNHIIPYN-RSFD-----DTY-HNRVLTLTETK
 46 -DMEHTIPKS-ISFD-----NSD-QNLTLCESYYN
 47 -DIEHTIPRS-AGGD-----STK-MNLTLCSSRFN
 48 -DIEHTIPRS-ISQD-----NSQ-MNKTLCSLKFN
 50 -DIDHVIPLA-RGGR-----DSL-DNMVLCQSDAN
 39 -DIEHLFPIA-ESED-----NGR-NNLVI SHSACN
 41 -DVDHIFPRD-DTAD-----NSY-GNKVVAHRQCN
 40 -DIEHIVPQS-LGGL-----STD-YNTIVTLKSVN
 35 -ELDHIVPRT-DGGS-----NRH-ENLAITCGACN
 36 -EMDHIVPRKGVGST-----NTR-TNFAAVCAECN
 37 -EMDHIVPRKGVGST-----NTR-VNLAACAACN
 38 -EMDHIVPRAGQGST-----NTR-ENLVAVCHRCN
 70 -EIDHILPRS-LIKDARGIVFNAE-PNLIYASSRGN
 71 -EIDHIIPRS-LTGRTKKTVFNSE-ANLIYCSSKGN
 73 -EIDHIIPRS-LTLKKSESIYNSE-VNLI FVSAQGN

(SEQ ID NO:211)
 (SEQ ID NO:212)
 (SEQ ID NO:213)
 (SEQ ID NO:214)
 (SEQ ID NO:215)
 (SEQ ID NO:216)
 (SEQ ID NO:217)
 (SEQ ID NO:218)
 (SEQ ID NO:219)
 (SEQ ID NO:220)
 (SEQ ID NO:221)
 (SEQ ID NO:222)
 (SEQ ID NO:223)
 (SEQ ID NO:224)
 (SEQ ID NO:225)
 (SEQ ID NO:226)
 (SEQ ID NO:227)
 (SEQ ID NO:228)
 (SEQ ID NO:229)
 (SEQ ID NO:230)
 (SEQ ID NO:231)
 (SEQ ID NO:232)
 (SEQ ID NO:233)
 (SEQ ID NO:234)
 (SEQ ID NO:235)
 (SEQ ID NO:236)
 (SEQ ID NO:237)
 (SEQ ID NO:238)
 (SEQ ID NO:239)
 (SEQ ID NO:240)
 (SEQ ID NO:241)
 (SEQ ID NO:242)
 (SEQ ID NO:243)
 (SEQ ID NO:244)
 (SEQ ID NO:245)
 (SEQ ID NO:246)
 (SEQ ID NO:247)
 (SEQ ID NO:248)

Fig. 5B

72	-EIDHIYPRS-LSKKHFGVIFNSE-VNLIYCSSQGN	(SEQ ID NO:249)
74	-EIDHILPRS-HTLKIYGTVFNPE-GNLIYVHQKCN	(SEQ ID NO:250)
75	-ELDHIIPRS-HKKY---GTLNDE-ANLICVTRGDN	(SEQ ID NO:251)
34	-ELEHIVPHS-FRQS-----NAL-SSLVLTWPGVN	(SEQ ID NO:252)
	:* * . . :	

Fig. 5C

Alignment of the HNH-like Domains disclosed in Chylinski et al. (excluding sequence outliers).
(CLUSTAL format alignment by MAFFT (v7.058b))

```

1  YDIDHIYPRS-LTKDDS-FDNLVLCERTAN (SEQ ID NO:253)
2  -DIDHIYPRSKVIKDDS-FDNLVLVLKNEN (SEQ ID NO:254)
3  -DRDHIYPQS-KIKDDS-IDNLVLVNKTYN (SEQ ID NO:255)
4  -DIDHIYPRS-KIKDDS-ITNRVLVEKDIN (SEQ ID NO:256)
6  -DIDHIYPQS-KIKDDS-ISNRVLCSSCN (SEQ ID NO:257)
5  -DIDHIYPQS-KTMDDS-LNRRVLVKKNYN (SEQ ID NO:258)
7  -DQDHIYPKS-KIYDDS-LENRVLVKKNLN (SEQ ID NO:259)
8  -QIDHIYPQS-LVKDDS-FDNRVLVPSSEN (SEQ ID NO:260)
9  -DIDHIIPQA-FIKDNS-IDNRVLTSSKEN (SEQ ID NO:261)
12 -DIDHIIPQA-FLKDNS-IDNKVLVSSASN (SEQ ID NO:262)
16 -DIDHIIPQA-YTKDNS-LDNRVLVSNI TN (SEQ ID NO:263)
11 -DIDHIYPQS-FITDNS-IDNLVLTSSAGN (SEQ ID NO:264)
10 -DVDHIYPQS-FLKDDS-IDNKVLTRSDKN (SEQ ID NO:265)
14 -NIDHIYPQS-MVKDDS-LDNKVLVQSEIN (SEQ ID NO:266)
18 -DIDHILPQS-LIKDDS-LDNRVLVNATIN (SEQ ID NO:267)
19 -DIDHILPQS-FIKDDS-LENRVLVKKAVN (SEQ ID NO:268)
13 -EVDHIFPRS-FIKDDS-IDNKVLVIKMMN (SEQ ID NO:269)
15 -EVDHIIPRS-YIKDDS-FENKVLVYREEN (SEQ ID NO:270)
17 -DIDHIIPQA-VTQNDN-IDNRVLVARAEN (SEQ ID NO:271)
21 -DIDHIVPRS-ISFDDS-FSNLVI VNKLDN (SEQ ID NO:272)
22 -EIDHIIPYS-ISFDDS-SSNKLLVLAESN (SEQ ID NO:273)
24 -EIDHIIPYS-LCFDDS-SANKVLVHKQSN (SEQ ID NO:274)
28 -EIDHIIPIS-ISLDDS-INNKVLVLSKAN (SEQ ID NO:275)
30 -EVDHIIPIS-ISLDDS-ITNKVLVTHREN (SEQ ID NO:276)
27 -EVDHILPLS-ITFDDS-LANKVLVYATAN (SEQ ID NO:277)
26 -EIDHIIPRS-ISFDDA-RSNKVLVYRSEN (SEQ ID NO:278)
29 -EVDHIIPRS-VSFDNS-YHNKVLVKQSEN (SEQ ID NO:279)
31 -EVDHIIPYS-ISWDDS-YTNKVL TSAKCN (SEQ ID NO:280)
32 -DIDHIIPYS-RSMDDN-YSNKVLVLSGEN (SEQ ID NO:281)
23 -EIEHIIPYS-MSYDNS-QANKILTEKAEN (SEQ ID NO:282)
33 -EVDHIVPYS-LILDNT-INNKALVYAEEN (SEQ ID NO:283)
25 -EIDHVI PYS-KSADDS-WFNKLLVKKSTN (SEQ ID NO:284)
49 -EMDHILPYS-RSLDNG-WHNRVLVHGKDN (SEQ ID NO:285)
42 -EIEHVI PYS-LYFDDS-FSNKVI CEAEVN (SEQ ID NO:286)
43 -DIEHIIPQA-RLFDDS-FSNKTL EARSVN (SEQ ID NO:287)

```

Fig. 6A

44	-EIEHIVPKA-RVFDDSD-FSNKTLTFHRIN	(SEQ ID NO: 288)
20	-DKDHIIPQS-MKKDDSIINNLLVLVNKNAN	(SEQ ID NO: 289)
45	-QVDHILPWS-RFGDDSD-YLNKTLCTARSN	(SEQ ID NO: 290)
50	-DIDHVIPLA-RGGRDS-LDNMVLCSQSDAN	(SEQ ID NO: 291)
46	-DMEHTIPKS-ISFDNS-DQNLTLCESYYN	(SEQ ID NO: 292)
47	-DIEHTIPRS-AGGDST-KMNLTLCSSRFN	(SEQ ID NO: 293)
48	-DIEHTIPRS-ISQDNS-QMNKTLCSLKFN	(SEQ ID NO: 294)
39	-DIEHLFPIA-ESEDNG-RNNLVISHSACN	(SEQ ID NO: 295)
41	-DVDHIFPRD-DTADNS-YGNKVVAHRQCN	(SEQ ID NO: 296)
40	-DIEHIVPQS-LGGLST-DYNTIVTLKSVN	(SEQ ID NO: 297)
35	-ELDHIVPRT-DGGSNR-HENLAITCGACN	(SEQ ID NO: 298)
36	-EMDHIVPRKGVGSTNT-RTNFAAVCAECN	(SEQ ID NO: 299)
37	-EMDHIVPRKGVGSTNT-RVNLAACAACN	(SEQ ID NO: 300)
38	-EMDHIVPRAGQGSTNT-RENLVAVCHRCN	(SEQ ID NO: 301)
34	-ELEHIVPHS-FRQSNALSSSLVLTWPGVN	(SEQ ID NO: 302)
	: ; * *	

Fig. 6B

-Y

B.

Fig. 7A

24/26

			B
NmCas9	DTRYVNRFLCQFVADRMRLTGKGKKRVF-----ASNGQITNLLRGFWGLRKVRAENDRH		
SpCas9	ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYH		
	*TR-***-Q**--RM-----*--***-----*--***-----*R--***-KVR--N*-H		
NmCas9	HALDAVVVACSTVAMQQKI---TRFVRYKEMNAFDGKTID---KETGEVLHQKTHFPQP		
SpCas9	B HAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI		
	HH-AD-*--A---A*-K-----Y-***-D-***-----*E-G*---*---*--		
NmCas9	WEFFAQEVMIRVFGKPDGKPE-----FEEADTLEKLRTLLAEKLSSRPEAVHEY		
SpCas9	MNFFKTEITLA-NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL-----MPQ-----		
	-*FF--E*-***-G*-**--*P-----***--*--*R-*L*-----P*-----		
NmCas9	VTPLFVSRAPNRKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLDLEKVMN--REREP		
SpCas9	-----VNIVKKTEVQTGGFSKES----ILPKRNSDKLIARKKDWD		
	-----**--VK-***--G-S-----L-***-K*-***--*P		
NmCas9	KLYEALKARLEAHKDDPAKAFAPFYKYDKAGNRTQQVKAVR---VEQVQKTGVWVRNH-		
SpCas9	KKYGGFD-----SPTVAYSVLVAKVEKGK-SKKLKSVELLGITIMERSSSFENPI		
	K-Y--*-***--P*-A**-----*G-*****K*V*-----*--*---*N--		
NmCas9	-----NGIAD-----NATMVRVDVFEKGDKYLVPIY-----		
SpCas9	DFLEAKGYKEVKKDLIIKLPKYSLEFLENGRKRMLASAGELQKGNELALPSKYVNFLYLA		
	-----*G--*-----*--*-----**KG**--L---Y-----		
NmCas9	-SWQVAKGILPDRAVVQGKDEEDWQLIDDS-----FNFKFSLHPNDLVEVI-----		
SpCas9	SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHRD		
	-***--KG---D---Q---E*-***-D*-----F---*--L---*L-*V*-----		
NmCas9	-----TKKARMFYGFASCHRGTGNINIRIHDLDHKIGKNGILEGIGV		
SpCas9	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS-TKEVL DATLIHQSI-----		
	-----F-YF-*-----LD--*--*--I-----		
NmCas9	KTALSFQKYQIDELGKEIRPCRLKKRPPVR	(SEQ ID NO:6)	
SpCas9	-TGLYETRIDLSQLGGD-----	(SEQ ID NO:7)	
	-T-L---*--*--*LG-*-----		

Percent Identity Matrix - created by Clustal2.1

Fig. 7B

25/26

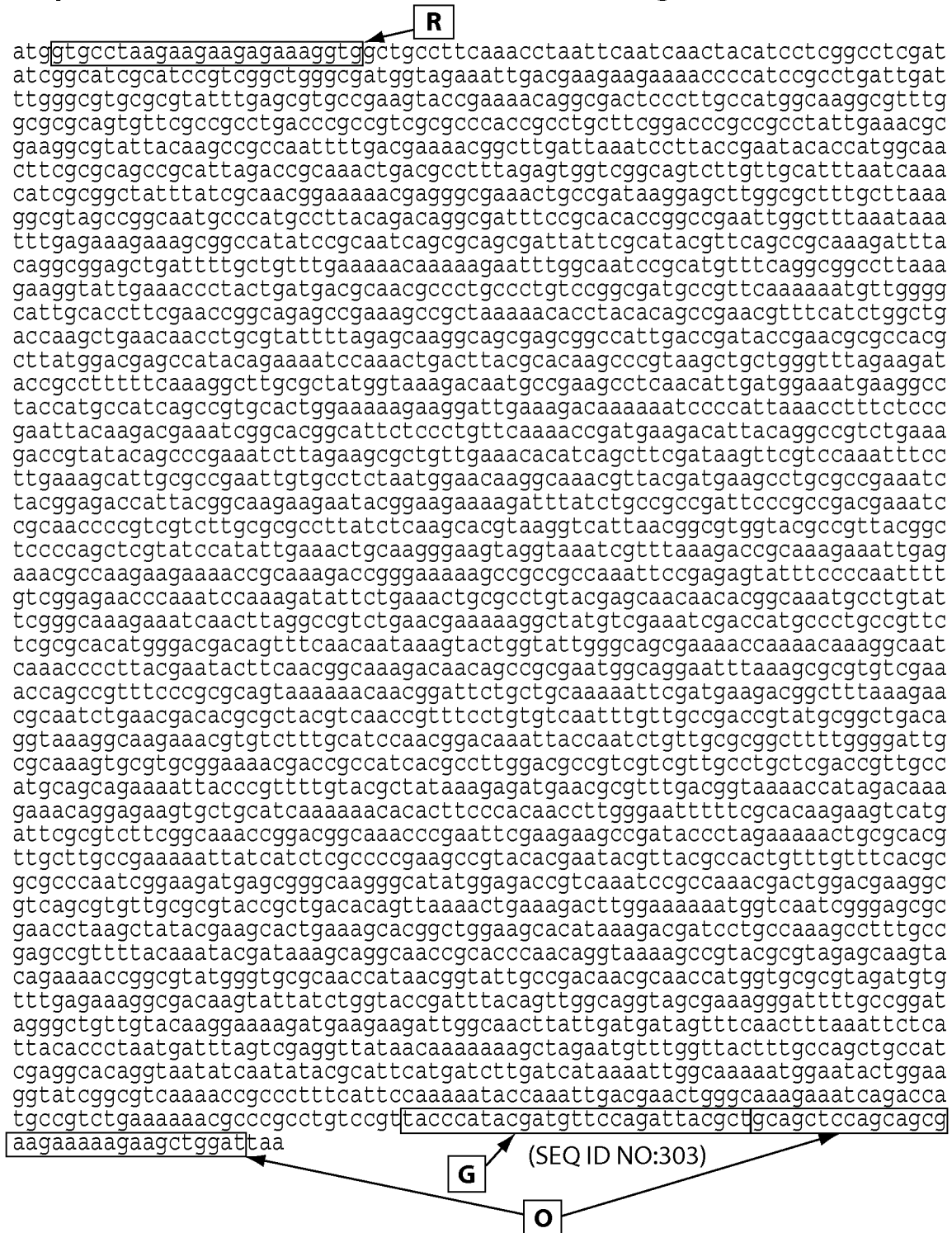
Sequence of the NmCas9 ORF with dual NLS and HA tags

Fig. 8

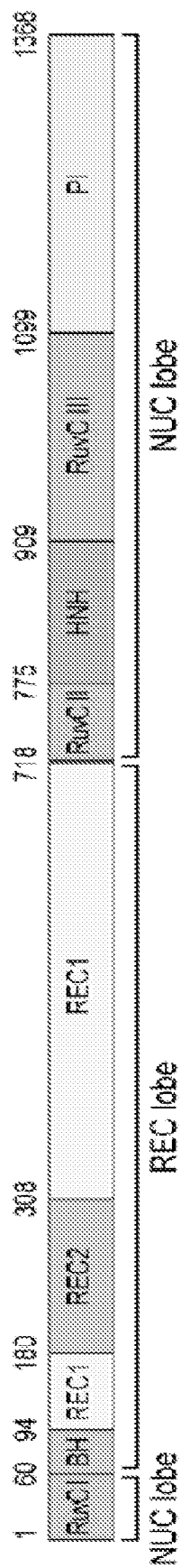


Fig. 9A

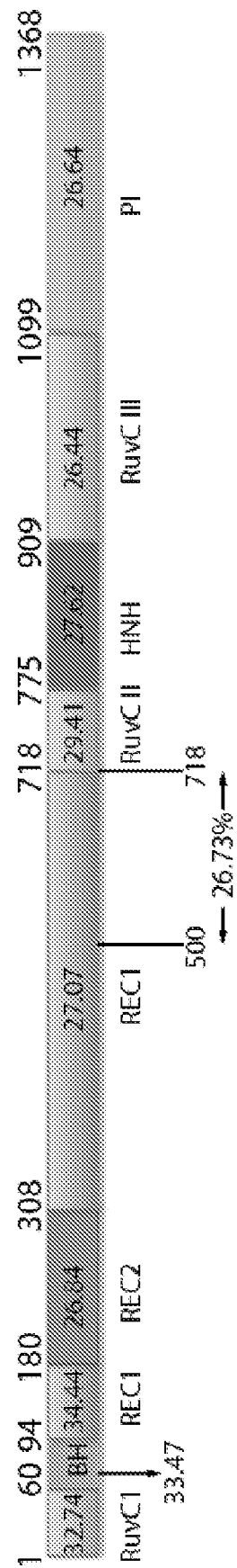


Fig. 9B

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/019064

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/113 A61K31/7088 C12N9/22
 ADD. A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	YANG JUN ET AL: "Current understanding of usher syndrome type II", FRONTIERS IN BIOSCIENCE, vol. 17, January 2012 (2012-01), pages 1165-1183, XP008176648, the whole document	1,2, 25-46, 69-103, 126-160, 164-263
Y	JEFFRY D SANDER ET AL: "CRISPR-Cas systems for editing, regulating and targeting genomes", NATURE BIOTECHNOLOGY, vol. 32, no. 4, 2 March 2014 (2014-03-02), pages 347-355, XP055172520, ISSN: 1087-0156, DOI: 10.1038/nbt.2842 the whole document	1,2, 25-46, 69-103, 126-160, 164-263
	----- -/-	



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

16 June 2015

Date of mailing of the international search report

24/06/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040,
 Fax: (+31-70) 340-3016

Authorized officer

Andres, Serge

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/019064

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>CHRISTEL VACHÉ ET AL: "Usher syndrome type 2 caused by activation of an USH2A pseudoexon: Implications for diagnosis and therapy", HUMAN MUTATION, vol. 33, no. 1, 16 January 2012 (2012-01-16), pages 104-108, XP055140082, ISSN: 1059-7794, DOI: 10.1002/humu.21634 the whole document</p> <p>-----</p>	1-264
A	<p>BONNET CRYSTEL ET AL: "Usher syndrome (sensorineural deafness and retinitis pigmentosa): pathogenesis, molecular diagnosis and therapeutic approaches", CURRENT OPINION IN NEUROLOGY, vol. 25, no. 1, February 2012 (2012-02), pages 42-49, XP008176647, the whole document</p> <p>-----</p>	1-264
A	<p>RAN F ANN ET AL: "Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity", CELL, vol. 154, no. 6, September 2013 (2013-09), pages 1380-1389, XP028716272, ISSN: 0092-8674, DOI: 10.1016/J.CELL.2013.08.021 cited in the application the whole document</p> <p>-----</p>	1-264
A	<p>WESTON M D ET AL: "Genomic structure and identification of novel mutations in usherin, the gene responsible for Usher syndrome type IIa", AMERICAN JOURNAL OF HUMAN GENETICS, vol. 66, no. 4, 1 April 2000 (2000-04-01), pages 1199-1210, XP002454426, ISSN: 0002-9297, DOI: 10.1086/302855 cited in the application the whole document</p> <p>-----</p>	1-264
A	<p>YANG HUI ET AL: "One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering", CELL, vol. 154, no. 6, September 2013 (2013-09), pages 1370-1379, XP028716273, ISSN: 0092-8674, DOI: 10.1016/J.CELL.2013.08.022 the whole document</p> <p>-----</p> <p style="text-align: center;">-/--</p>	1-264

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/019064

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>YANFANG FU ET AL: "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs", NATURE BIOTECHNOLOGY, vol. 32, no. 3, 26 January 2014 (2014-01-26), pages 279-284, XP055194360, ISSN: 1087-0156, DOI: 10.1038/nbt.2808 the whole document</p> <p>-----</p>	1-264
A	<p>PATRICK D HSU ET AL: "DNA targeting specificity of RNA-guided Cas9 nucleases", NATURE BIOTECHNOLOGY, vol. 31, no. 9, 1 September 2013 (2013-09-01), pages 827-832, XP002718604, ISSN: 1546-1696, DOI: 10.1038/NBT.2647 the whole document</p> <p>-----</p>	1-264
T	<p>ZHENG ANDREW ET AL: "Personalized therapeutic strategies for patients with retinitis pigmentosa.", EXPERT OPINION ON BIOLOGICAL THERAPY, vol. 15, no. 3, March 2015 (2015-03), pages 391-402, XP008176606, ISSN: 1744-7682</p> <p>-----</p>	